

ENHANCING THE USER SUPPORT PROCESS IN FEDERATED E-SCIENCE

Dissertation with the aim of achieving a doctoral degree

at the Faculty of Mathematics, Informatics and Natural Sciences

Department of Informatics

of Universität Hamburg

Submitted by Hashim Chunpir

2015, Hamburg

Date of oral defence: 19.05.2015

The following evaluators recommend the admission of the dissertation:

First Reviewer: Professor Dr. Thomas Ludwig

Second Reviewer: Professor Dr. Markus Nüttgens

Dedication

This work is dedicated to the orphans and poor people around the world!

Acknowledgement

I am indebted to Prof. Dr. Thomas Ludwig for accepting me as his doctoral candidate in his scientific computing research group, which is a part of the University of Hamburg. Prof. Dr. Thomas Ludwig was instrumental in motivating me to reach my research goals that I have outlined earlier. Moreover, he helped me in every step of my research work and administrative issues that I faced from time to time. He supervised my research by providing valuable tips from his experience and motivated me to collaborate with other researchers at different universities. By his continuous encouragement, I never gave up in my research despite facing difficult times and circumstances.

I would also like to show my greatest appreciation to Prof. Dr. Markus Nüttgens for providing his opinions that provided invaluable guidance in collation of this study. I cherish cheerful discussions with Prof. Dr. Paul Drews, especially on the research methodologies and research paradigm that provided me insight to organise my study design. I am thankful to Dr. Stephan Kindermann for his advice and contribution towards development of my dissertation. His thoughts at the beginning about my dissertation encouraged me to continue my dissertation in this direction. I am also obliged to Prof. Dr. Kari Kuutti whose comments and suggestions were of inestimable value for my study. Furthermore, I appreciate the collaboration of Prof. Dr. Helena Karasti for dedicating her valuable time to comment on my research activities. I also appreciate comments of Mr. Badewi and his cooperation during this work.

I am very thankful to my family, friends, parents and my brother for their kind support and encouragement throughout my studies. I am thankful to Dr. M. Lautenschlager, Dr. Torsten Rathmann, Mr. Hinrich Reichardt and all the members of the data management group for their full support. I am thankful to all the members of the DKRZ for their strong support, especially for the nice small talks and German to English translations. I got a key-note speaker opportunity; I thank Mr. Dean N. Williams for his invitation. I would like to express my gratitude to A. Thiede, F. Schulte, E. Betke, E. Curri, Prof. M. C. Kindsmüller, Prof. Preßmar and J. Twiefel for their continuous support and encouragement as well. I am very thankful to the German Climate Computing Centre (DKRZ) for employing me and during my employment I had the opportunity to discover a practical problem and a knowledge gap that resulted in the genesis of this thesis. I am thankful to all the stakeholders of the Earth System Grid Federation (ESGF) for their cooperation to conduct various studies and participating whole heartedly in the studies of this research. I express my affection and appreciation to all those people who helped me in any way in achieving this task. I am terribly sorry, if I have not written name of any of those. I owe you a drink!

Abstract

The maturity of grid technology has enabled inter-connectivity of data centres and supercomputing (HPC) instruments, which gave rise to e-Science infrastructures. E-Science infrastructures have reached their production level as far as their capability to serve and connect data is concerned. However, researchers face difficulties in their quest to perform complex operations, using the e-Science infrastructures, for big data analysis and research. These operations include; accessing data as well as performing other operations such as ultra-scale visualization of data and data processing via high performance computation. Unfortunately, there are fewer studies conducted so far that examine the user process, address stakeholder concerns as well as user-oriented service aspects including analysis of technology used in e-Research. In this study, the author has initiated a movement in the form of an effort to enhance and improve the user support in order to acquire the services of e-Science infrastructures to facilitate research activities of users. In this research, qualitative and quantitative methods such as interviews, questionnaires and participatory observations are selected to acquire the current state of user support process in e-Science infrastructure. In this work, Earth System Grid Federation, which is a popular distributed e-Science infrastructure to host climate data to assess the global warming, was chosen as a case study. The results indicate that the current user support process needs to be re-structured, in order to provide better service possibilities to researchers. Moreover, it is further emphasised in this study that enormous amount of resources can be saved by altering the management and governance of an e-Science infrastructure. A stimulating outcome of this research is a Federated e-Science User Support Enhancement (FeUSE) framework, which is recommended and proposed to improve the user support services of e-Science facilities such as federated e-Science infrastructures like Earth System Grid Federation.

Table of contents

Dedication	I
Acknowledgement	III
Abstract	V
Table of contents.....	VII
List of figures.....	XII
List of tables.....	XVIII
1 Introduction to the dissertation	20
1.1 Introduction	20
1.2 Introduction to the research area	20
1.3 Main focus of existing research	21
1.4 Objectives of the project	22
1.5 Extensions and contribution to the other fields.....	23
1.6 The structure of the thesis	23
1.7 Chapter summary	25
2 Literature review and state of the art.....	26
2.1 Introduction	26
2.2 An overview of e-Science and associated concepts.....	27
2.3 Overview of user-support.....	30
2.4 State of the art of end-user support in e-Science infrastructures	35
2.5 Models used in structuring user-support communication	42
2.6 Overview of techniques, frameworks and technologies for improving user-support	46
2.7 Tools that can be used in modelling user-support processes	55
2.8 Soft systems methodology	55

2.9	Knowledge gap analysis	56
2.10	Chapter summary.....	56
3	Research aims and objectives	58
3.1	Introduction	58
3.2	Research objectives and the research question.....	58
3.3	Study design	63
3.4	The methodological domain	65
3.5	Reasons for the choice of the case study method	67
3.6	Earth System Grid Federation as a platform for case study	69
3.7	Research methods	73
3.8	Chapter summary.....	80
4	ESGF - case study profile.....	81
4.1	Introduction	81
4.2	History of ESGF	82
4.3	Projects supported by ESGF.....	88
4.4	Interaction of a user with ESGF: An example.....	90
4.5	ESGF system architecture	92
4.6	Future expansion of ESGF data archive system.....	102
4.7	Chapter summary.....	104
5	Governance and organisation of ESGF	105
5.1	Introduction	105
5.2	ESGF current governance scheme.....	106
5.3	Organisation of ESGF P2P network, the global picture.....	111
5.4	Challenges faced by ESGF as a diverse networked organisation.....	118
5.5	ESGF development philosophy	124
5.6	Chapter summary.....	126

6	Presentation of empirical findings – (As Is)	128
6.1	Introduction	128
6.2	Data sources and methods used to collect data	130
6.3	Use of soft systems thinking methodology in modelling the current ESGF user support process	133
6.4	Current perception about the end users	143
6.5	Attributes of people supporting users of ESGF	148
6.6	Nature of user support provided currently in ESGF	155
6.7	The structure and model of user support in ESGF	156
6.8	Current practices of user support	160
6.9	Satisfaction level	179
6.10	Service level agreements	181
6.11	Criticism of the current state of user support	181
6.12	Chapter summary	186
7	Analysis of empirical findings	188
7.1	Introduction	188
7.2	Analysis of the current empirical results	189
7.3	Analysis of historical aspects of the user support process	212
7.4	The core problems identified in the current ESGF user support process	224
7.5	Chapter summary	229
8	The solution (To Be) – FeUSE framework	231
8.1	Introduction	231
8.2	Federated e-Science infrastructure user support enhancement (FeUSE) framework	231
8.3	Vision	232
8.4	Strategy	234

8.5	Communicate the strategy	259
8.6	Agreement and assignment.....	259
8.7	Support operation	271
8.8	Monitoring	283
8.9	Institutionalization	289
8.10	Implementation of FeUSE in ESGF	289
8.11	The future vision of this research	292
8.12	Chapter summary.....	299
9	Conclusion and future work	301
9.1	Introduction	301
9.2	Conclusion of the empirical findings.....	301
9.3	Conclusion and summary of the FeUSE framework.....	306
9.4	Limitations of this study and future works.....	309
10	Bibliography.....	312
11	Appendix A	326
	Declaration on Oath	329

List of figures

Figure 1.1 The overall structure of the dissertation.	24
Figure 2.1 Graphical representation of the arrangement of the chapter.....	27
Figure 2.2 Different channels to submit user-requests in in GGUS system, taken from (David, 2005).....	37
Figure 2.3 The organisation of support in GGUS system, taken from (David, 2005). ...	38
Figure 2.4 Previous user-support process with the mailing-list in ESGF and C3Grid-INAD e-Science infrastructure projects. This figure is redrawn and originally taken from (Hashim Chunpir, 2013).	39
Figure 2.5 Current user support process in C3Grid-INAD e-Science infrastructure project. This figure is redrawn and originally taken from (Hashim Chunpir, 2013).	39
Figure 2.6 Representation of user and support staffs interactions in e-Science infrastructure projects based on RT system.....	40
Figure 2.7 Two-level end-user support structure (left) and N-Level end-user support structure (right), adapted and redrawn from (N. Leung & Lau, 2007).....	45
Figure 3.1 The research design of this study with different phases and each phase having its own aim and outcomes as planned by the author.	65
Figure 3.2 The architecture of Earth System Grid Federation (ESGF) taken from (Dean N. Williams et al., 2013).....	70
Figure 3.3 The relationship of CMIP5 to organizations established to coordinate climate research activities internationally and to the IPCC, the modelling canters, and the climate research community, taken from (Taylor et al., 2012).	72
Figure 4.1 The distribution of registered users of in current phase of ESGF.	87
Figure 4.2 The steps that users do to get a data file from ESGF system and to visualize it using UV-CDAT visualization tool, adapted from (Dean N. Williams et al., 2013).	90
Figure 4.3 The peer-to-peer architecture of ESGF. The users can access ESGF data-holdings after authentication, either via web-browser or a desktop client, taken from (ESGF, 2013).....	93
Figure 4.4 The P2P federated, distributed and dynamic architecture of ESGF with each node configurable in any of different flavours or roles. Each node can have 1 to 4 roles. Redrawn from (Dean N. Williams et al., 2013).....	94

Figure 4.5 Software components that may be deployed onto an ESGF Node, grouped by “flavour” or area of functionality, taken from (Cinquini et al., 2012).	96
Figure 4.6 The components of node architecture coordinated by node manager, taken from (ESGF, 2013).....	96
Figure 4.7 Data search services architecture in ESGF, taken from (Cinquini et al., 2012).	99
Figure 4.8 The architecture of ESGF Security Services framework (also known as identity provider “IdP”) taken from (Cinquini et al., 2012).....	100
Figure 4.9 Screenshot of ESGF UI, NASA gateway.....	101
Figure 4.10 The screenshot of UI of dashboard.....	102
Figure 4.11 Future directions, the computer scientists are working on and the future science domains that ESGF may expand to.	103
Figure 5.1 The governance scheme of ESGF.	107
Figure 5.2 Various global organisations, geographically distributed, forming an ESGF consortium by connecting physical and human resources. The different colours show different entities, the diagram is represented using the notation coined by the author.....	115
Figure 5.3 Visualisation of UV-CDAT’s physical and human resources in the form of a chart in federated ESGF consortium.	116
Figure 5.4 Depiction of a new working group called the International Climate Network Working Group (ICNWG), recently started in ESGF.....	117
Figure 5.5 The constituents of the ESGF organisational chart highlighted from Figure 5.2, Figure 5.3 and Figure 5.4.	118
Figure 5.6 Federation of ESGF showing collaborations between a few of its remote data centres, data archives, and potential data transfers between sites taken from (Dean N. Williams et al., 2013).	120
Figure 5.7 Three layered quality assurance concept of ESGF.....	122
Figure 6.1 The contents of the chapter with main topics and sub-topics covered in the chapter.	129
Figure 6.2 The transformation process in ESGF user support.....	134
Figure 6.3 The conceptual model of the user support system in ESGF e-Science infrastructure.	136
Figure 6.4 The figure shows the user support process in climate e-Science infrastructures in the form of rich picture mind map.	137

Figure 6.5 Asynchronous and synchronous type of communication channels in the user support process of ESGF.	138
Figure 6.6 The time issues in the user support process.	139
Figure 6.7 The incoming incidents from users and the categorization of user requests.	140
Figure 6.8 The usability, understandability and frequency of update of the online help webpages provided by ESGF.	140
Figure 6.9 The types of users interacting with ESGF.	141
Figure 6.10 The operation of user support in ESGF is based on the following factors: Attitude, satisfaction, analytical skills and technical knowledge of support staffs and users.	142
Figure 6.11 The dependency of user support process on the structure, function and UI of ESGF e-Science infrastructure.	142
Figure 6.12 The administration, structure and model of ESGF.	143
Figure 6.13 The workflow sequence of the users with ESGF.	147
Figure 6.14 The main activities of the employees who support users (other than support activity).	150
Figure 6.15 The responses of the respondents to the question what do they consider them in a support process.	153
Figure 6.16 Respondents (user-support staff) engaged in different projects. Some respondents are working on different projects in parallel.	153
Figure 6.17 The time dedicated by the support staff in climate e-infrastructures support frameworks.	154
Figure 6.18 Average percentages of total incoming user-requests handled by ESGF human resources (support staff) in average.	155
Figure 6.19 The different user-support models in use in climate e-Science infrastructure projects as marked by percentage of support staff of ESGF.	157
Figure 6.20 The percentages of responses belonging to the structure of support in climate e-Science infrastructure support.	158
Figure 6.21 Distribution of support activities represented in the form of FLS, SLS and others, expressed in percentage.	159
Figure 6.22 The user support in ESGF e-Science infrastructure is mainly “On-demand”.	160

Figure 6.23 The proportion of incoming user-requests that is understandable.	163
Figure 6.24 Number of times support staffs have to write back to a user to fully understand the user request.	164
Figure 6.25 The utilization of user-request patterns by the end users.	164
Figure 6.26 Percentage of the user-requests on the average that remain to be unresolved in a year.	165
Figure 6.27 The average percentage of support staffs who delegate user-requests to multiple peers, simultaneously.	167
Figure 6.28 Category-mix of the user-requests that are received by support staff member in the current support process in climate e-Science, depicted in percentage.	170
Figure 6.29 The popularity of medium used (in percentage) for external communication with users.	171
Figure 6.30 Modes of internal communication used (in percentage) within e-Science projects.	172
Figure 6.31 Ranking of e-mail leading to Request Tracking Software as a (1 st popular) medium.	173
Figure 6.32 Ranking of mailing-lists as a (2 nd preferred) medium.	173
Figure 6.33 The percentage of (average) automatic reply mechanism present in the current support framework in e-Science.	175
Figure 6.34 The average response time to a user-request by the support staff in climate e-infrastructures support frameworks.	175
Figure 6.35 Support staff's (average personal) reply time.	176
Figure 6.36 The percentage of request resolution time of incoming user-problems with respect to time frame.	177
Figure 6.37 Support staffs (shown in percentage) who handle incoming user-requests in a non-documented way.	178
Figure 6.38 Frequency of user requests handled in a non-documented way in a month.	178
Figure 6.39 Communication media used (given in percentage in this figure) for undocumented user requests.	179
Figure 6.40 The responses of the support staff as far as satisfaction factor of the support staff is concerned.	180
Figure 6.41 The satisfaction level of end users as estimated by the support staff.	181

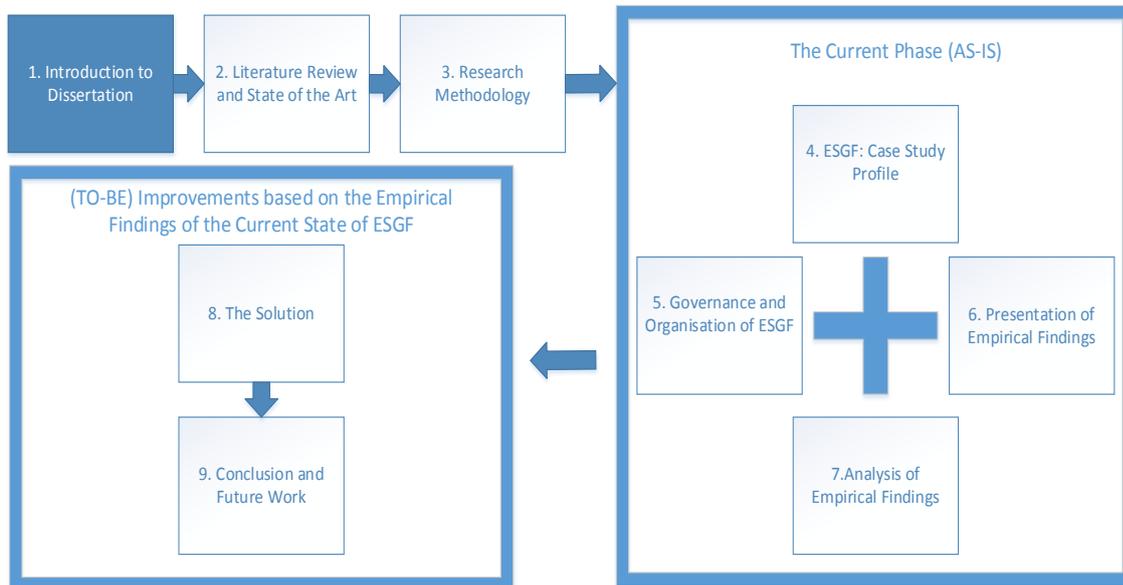
Figure 7.1 The conceptual dependency of ESGF User Support System.....	210
Figure 7.2 Current ESGF user support process with ML.....	214
Figure 7.3 Integrating help desk in ESGF user support process.	216
Figure 7.4 Intermediate phase with coexisting mailing list (ML) and help desk (HD) in ESGF user support process.....	217
Figure 7.5 Intermediate phase with coexisting mailing list (ML) and Askbot user forum in ESGF user support process.....	219
Figure 7.6 The next phase without coexisting mailing list (ML) that ESGF staffs want to achieve as an ideal ESGF user support process.....	220
Figure 7.7 Using user’s forum and RTS system in combination as a solution.	221
Figure 7.8 The screen shot of the tutorial in the form of video webcast on how to log on in the ESGF federation gateway to access climate data.	223
Figure 8.1 The first phase of FeUSE Framework.	233
Figure 8.2 Main points of the strategy of the FeUSE framework to enhance the user support services of ESGF.	237
Figure 8.3 The sub-phase of the second step of the FeUSE framework.	238
Figure 8.4 Classification of ESGF with respect to data download and access via ESGF system.	240
Figure 8.5 Classification of ESGF users with respect to technical affinity and skill sets.	241
Figure 8.6 The new "to be" governance structure of ESGF.....	245
Figure 8.7 The sub-phase “form governance scheme” of the second step of FeUSE framework.....	246
Figure 8.8 User support employee roles in ESGF. USPO and FLS are the full-time user support employees. SLS and USC (represented in red) do user support tasks on demand as the need arises. However, USC monitors the user support process.	247
Figure 8.9 The sub-phase of the step two of the FeUSE Framework.	254
Figure 8.10 The major steps in the resource planning sub-phase of the second step of the FeUSE framework.	257
Figure 8.11 The main output of the user-stakeholder relationship management sub-phase.	258
Figure 8.12 Step three of the FeUSE framework; communicating the strategy to all the stakeholders and collecting feedback.	259

Figure 8.13 Agreement of the stakeholders and assignment of responsibilities to all concerned stakeholders for the user support process in federated infrastructures.	260
Figure 8.14 The fifth phase of the FeUSE framework that addresses the division of incoming issues into different types depending on the nature of origin of issues.	272
Figure 8.15 The workflow of events created by the federated components of the e-Science infrastructure, adapted from OGC (Office of Government Commerce, 2011).	275
Figure 8.16 The workflow of the problem management in FeUSE framework (Office of Government Commerce, 2011).	280
Figure 8.17 The access management workflow in the FeUSE framework adapted from OGC (Office of Government Commerce, 2011).	282
Figure 8.18 The workflow of activities in RFC adapted from OGC (Office of Government Commerce, 2011).	285
Figure 8.19 The graph shows the number of policy makers of ESGF and its associated projects that are in favour of introducing FeUSE framework to enhance user support service in ESGF.	290
Figure 8.20 Time to implement the FeUSE framework in ESGF.	291
Figure 8.21 The technologies currently being deployed in the ESGF e-Science infrastructure, which provide intangible benefits to an organisation.	293
Figure 8.22 The web 2.0 tools that contribute to the various concepts that can bring value to the stakeholders of ESGF and other e-Science infrastructures taken from (Andriole, 2010).	294
Figure 8.23 The abilities of the different concepts that the web 2.0 tools can support taken from (Andriole, 2010).	295
Figure 8.24 The sketch of e-Research 2.0 platform.	296
Figure 8.25 Functionality aspects of the eResearch platform.	298

List of tables

Table 4.1 The projects and models supported by the ESGF data archive system.....	89
Table 6.1 The table demonstrates the interviewees, their roles, their location of job and the mode of the conducted interview.....	130
Table 6.2 CATWOE analysis of user support system in ESGF e-Science infrastructure.	134
Table 6.3 The frequency of updating the user-support webpages as recorded according to the responses in survey questionnaire.	151
Table 6.4 The average number of user-requests received by the support staff per working day.	161
Table 6.5 Percentage of categories deemed relevant by the support staff, shown in percentage, in ESGF and associated projects support.	168
Table 6.6 Percentage of sub-categories recorded under “1. General information ...” first category of Table 6.5 considered relevant by the support staff members in ESGF and associated projects support, shown in percentage.	169
Table 7.1 The comparison of advantages and disadvantages of the mailing-lists and the request tracking software.....	212
Table 7.2 The table shows the number of problems and their description deduced from interviews.....	225
Table 8.1 Table shows the ideal target MPRT and MTPW as well as MTPS as recommended by the FeUSE framework.....	263
Table 8.2 Table shows the target MTPUF, recommended by FeUSE.	263
Table 8.3 Table shows the ideal target RTBS as recommended by the FeUSE framework.....	264
Table 8.4 The description of the components of e-Research 2.0 social collaboration and support platform.	297
Table 11.1 The summary of the development of ESGF since its inception.	326

1 Introduction to the dissertation



From this chapter the following paper were published:

- *Hashim Iqbal Chunpir, Thomas Ludwig and Endrit Curri: Improving Processes for User Support in e-Science. IEEE e-Science Conference 2014: 87-90*

The paper provides an overview of the project. The paper was published in 11th eScience conference at Guarujá, Brasil.

1.1 Introduction

Chapter 1 introduces the reader to the main research area of this study in Section 1.2, which is about e-Science and e-Science infrastructures. In Section 1.3, the author provides the current focus of an existing research in the field of e-Science as well as e-Science infrastructures. Moreover, the author highlights the area of user support in e-Science that has not been completely covered yet including the knowledge gap. The objectives of this dissertation including the main research question of this study have been highlighted in Section 1.4. The contributions and extensions of this research are emphasised in Section 1.5. Furthermore, the scope of this doctoral project is also given in Section 1.5. The organisation of the thesis and overview of the structure of this doctoral project is illustrated in Section 1.6. Finally, the summary of this chapter is given in Section 1.7.

1.2 Introduction to the research area

E-Science and e-Science infrastructures have been widely deployed to access and share the knowledge, data, computing and even human resources to facilitate intra-disciplinary and inter-disciplinary research also known as e-research. Cyber-

infrastructures, collaboratories and e-Research are the different names of the same concept popularly known as “e-Science” in Europe and “cyber-infrastructures” in the US (Jirotko et. al. 2012). Prof. Jan Gulliksen of Royal Institute of Technology, Sweden said in the NordiCHI conference at Copenhagen in 2012 that the e-Science infrastructures support the concept of Science 2.0. There is a tremendous shift from the traditional Science activities which are known under the term “Science 1.0” to “Science 2.0.” Science 2.0 means that the researcher use internet to get the research data stored at data centres to perform research rather than using a physical lab environment. It is based on integrated case studies, validity of scientific data and collaboration of people by using e-Science infrastructures.

“E-Science denotes the systematic development of research methods that exploit advanced computational thinking.” Professor Malcolm Atkinson, an e-Science Envoy

With the inception of e-Science in 1999 by John Taylor, new methods started to develop to support scientists in conducting scientific research. The aim of e-Science is to come up with new scientific discoveries mainly by analysing vast amounts of data accessible over the internet, using diverse high-performance computing (HPC), data resources and other instruments. These resources could be geographically widely-dispersed.

E-Science includes medicine, earth sciences, climate sciences, particle physics, bio-informatics, social sciences and other fields. Many of these fields have well-developed e-Science infrastructures. An example of a well-developed e-Science infrastructure in the field of particle physics, comprises of computing resources that compute the data origination from CERN; Large Hadron Collider (LHC). The recent discovery of Higgs Boson Particle was also based on e-Science infrastructure (reported by BBC on 1st August 2012). According to a report on “E-Science and Data Support Services” a study conducted by Association of Research Libraries (ARL) member institutions, it is a fact that investments in e-science activities are being made even during difficult budget times.

According to Research Council UK:

“E-Science infrastructure denotes the digital equipment, software, tools, portals, deployments, operational teams, support services and training that provide computational services to researchers.”

1.3 Main focus of existing research

E-Science has mainly focused on the following few areas; the technical aspects of e-Science e.g. anatomy of data grid (Buyya et al. 2005), development of middleware, storage of data in grid (Hey, T., & Trefethen 2005, pages 817-821) and socio-structural aspects of e-Science. The examples of socio-structural aspects of e-Science include; virtual organisations (VOs), collaborative work environments (CWE) and virtual

research environments (VRE). The current research in the field of e-Science improved the technical considerations and scientific research methods. However, little studies have been conducted to analyse and evaluate the user support process in e-Science, empirically. Since e-Science is still a nascent field there are fewer readily usable tools available to support researchers in their work. In addition, the support process from support staffs depends on diversity of socio-technical factors and has not been highlighted at a micro-level.

1.4 Objectives of the project

Support process plays an important role to facilitate researchers i.e. end-users of e-Science infrastructures to accomplish their research activities, with the help of e-Science infrastructures. But how the support staff processes user-requests, as far as the support is concerned? How to facilitate the researchers in order to service their requests properly? How to model and define the current e-Science supporting framework? This research project investigates the current state of user support process in e-Science infrastructures such as Earth System Grid Federation (ESGF) and its associated projects like Collaborative Climate Community data and processing grid infrastructure (C3Grid) and possibly proposes a model of the current community support process to help users of E-Science infrastructures in the field of climate science. The focus of this research is studying the user support process by observing the support staffs interaction with the end-user requests.

In this research, empirical methods such as interviews, thematic surveys and field study are applied to find out the user support process mainly from the perspective of the support staffs. The user support community comprises of support staffs including first level support, second level support personnel and the users. The user support staffs handle user-requests themselves but at the same time they are also the users of the support system. The support system comprises of set of tools, techniques and methods that facilitate user-request resolution. Studying support staffs interaction with the users and within themselves via the support tools provides insight into the salient features and factors that affect the support process. Furthermore, common practices of the user support community and the problems with the support process in e-Science are identified in this project. In addition, common mistakes within the user support process are uncovered to correct the process.

It is generally accepted that e-Science is a relatively new field in science (Jirotko et. al. 2013). Therefore, looking from the perspectives of human computer interaction (HCI), computer supported cooperative work (CSCW), IT service management (ITSM) and organization, this research on modelling support process is expected to form a viable user support structure. This user support structure is thought to be easy to use for the support staff and efficient in supporting users to understand the e-Science infrastructure.

The outcomes of this research project about the user support process can be applied to the other fields e.g. user-support in federated IT and other industries.

1.5 Extensions and contribution to the other fields

This research helps in reducing human errors, redundancy of work, wastage of resources and time by structuring the user support process in e-Science infrastructures. Examining the user support process of e-Science infrastructures in climate science reveals the structure and function of user-support service. Furthermore, the aspects of the user support system can be represented and such knowledge can be useful in defining a user support model in other e-Science infrastructures as well. This in turn can be a particularly rich source for generating practical implications for system design, user modelling and user instruction.

Once, the problems of the support process in climate e-Science infrastructures have been identified, it might help in identifying similar sorts of problems in other fields. As an inter-disciplinary study, the research will contribute to IT service management (ITSM), IT governance and organization, human-computer interaction (HCI), computer-supported cooperative work (CSCW) and e-Science. For IT organization, exploring the current support process in e-Science is useful for understanding the organizational structure of support staff in e-Science, as well as their role in problem solving. For HCI, the exploration could provide new system and user interaction (UI) techniques involved in supporting staff members and end users` communication in e-Science infrastructures. In CSCW the exploration of support process of climate e-Science infrastructures shall facilitate understanding support process from a micro perspective within the science—technology—sociality triangle, across different placements, on the spectrum from development to theory and also across different orientations.

For e-Science, investigating a user support process in climate science will help to apply the similar user support practices and models to other domains such as biology, astronomy, social sciences etc. However, applying support practices and models to other fields are out of scope of this study.

1.6 The structure of the thesis

The structure of the thesis can be visualised with the help of following diagram shown in Figure 1.1. The structure of the thesis is divided into three main parts. The first part comprises of three chapters from Chapter 1 to 3. This chapter; Chapter 1 is a brief introduction to the dissertation, already presented in the last sections of this chapter. Chapter 2 discusses the critical review, theoretical background and state of the art of empirical studies and other writings in the area of e-Science along with associated concepts such as process and user support theory that is relevant to this research project. In Chapter 3, the author describes the research objectives including the main research

question of this study, the process to analyse empirical findings in the light of the research methodology adapted to achieve the research goals.

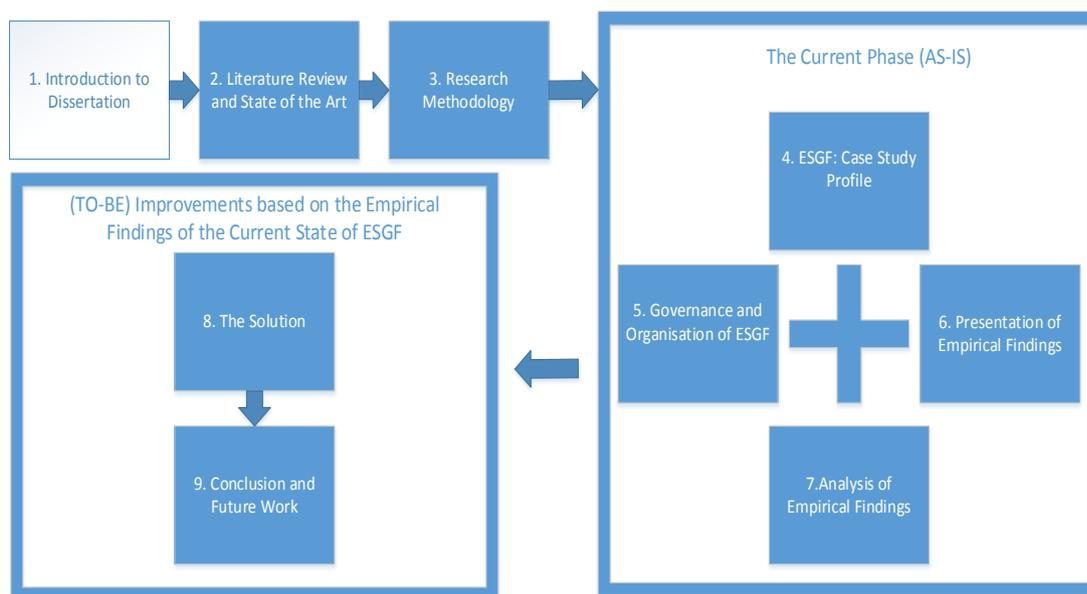


Figure 1.1 The overall structure of the dissertation.

The second part comprises of four chapters; from Chapter 4 to 7 and illustrates the “as-is” i.e. the current status of the user support process in e-Science infrastructures based on the single case study of ESGF. Chapter 4 illustrates the profile of the chosen case study: Earth System Grid Federation (ESGF) and explores the history, composition of the features that constitutes ESGF. In Chapter 5, the author does a brief analysis of the structure, governance and organisation of ESGF. The empirical findings about the current user support process in ESGF from multiple sources such as interviews, surveys and others are presented in Chapter 6. The detailed analysis of the whole empirical investigation is detailed in Chapter 7. Moreover, the key problems of the user support process are highlighted in Chapter 7 too.

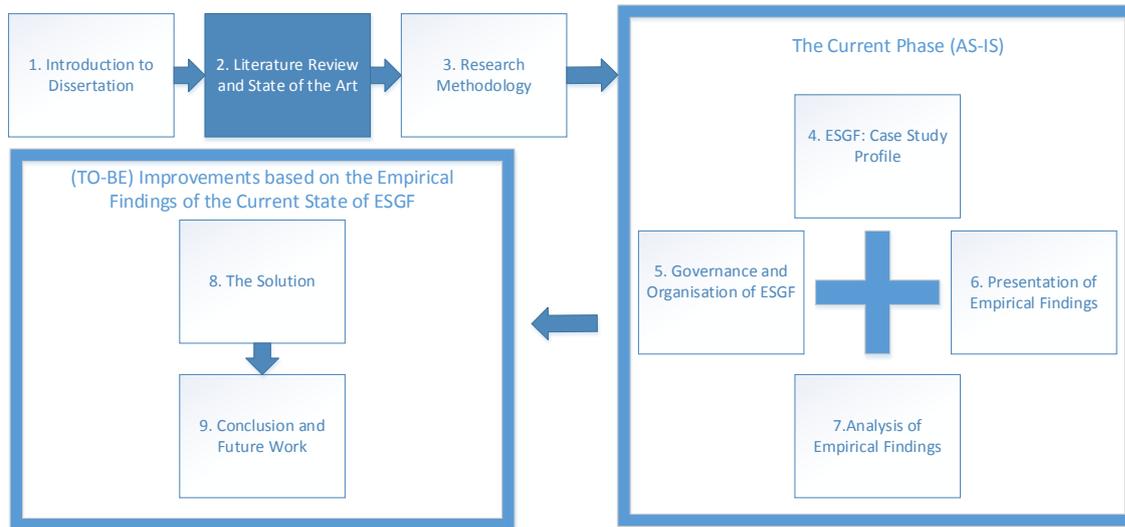
The third and the last part of this thesis, comprises of two chapters; from Chapter 8 to 9 and constitutes “to-be” part. This part provides guidelines how to solve the problems that were identified in the as-is part to improve the current status of the user support process in ESGF. In Chapter 8, Federated e-Science User Support Enhancement (FeUSE) framework is suggested to improve the current state of the user support process as well as service of ESGF and other similar e-Science infrastructures. FeUSE is a theoretical framework that provides best practice guidelines and recommendation that an e-Science infrastructure such as ESGF can follow to structure use support service in federated environment where federating partners are scattered all over the world at different geographic locations. Finally, conclusions of this study are drawn in Chapter 9, indicating the future directions of the research.

In this dissertation, American Psychological Association (APA), version six, citation format is followed. Moreover, in this thesis, English from the United Kingdom is chosen as a language standard. Most of the illustrations included in this thesis have been coined by the author himself and therefore no source is provided. The figures from other sources are adapted or re-drawn by the author and are cited accordingly.

1.7 Chapter summary

In this chapter, the main research area of this study which is about e-Science, e-Science infrastructures coupled with the user support is introduced. Furthermore, the author indicates at the lack of research in the area of user support in e-Science. Besides, the author defines the scope of the study keeping in focus the research goals of this thesis. The possible contributions and extensions of this research are given in Section 1.5. An introduction to some important terms that are used in the doctoral thesis is also given. Finally, the structure of the thesis is represented by a diagram in Figure 1.1. The overview of the structure of this doctoral project is explained in the concluding section, Section 1.6.

2 Literature review and state of the art



2.1 Introduction

This chapter gives analytical background to the terms and concepts along with their inter-connection with each other, relevant to the topic of this dissertation. It further demonstrates the state of the art in the field of end-user support, especially in the field of e-Science infrastructures. It is important to note that since this dissertation revolves around end-user support in climate e-Science infrastructures, the user-support and its associated concepts in other fields such as in industry though consulted are not the core part of this literature review. The arrangement of this chapter is represented in Figure 2.1.

The chapter starts with an overview of grid-computing, data-grid network, e-research, e-Science and other related concepts in Section 2.2. In Section 2.3, an overview of end-user support and other related concepts in this domain are delineated. In Section 2.5, state of the art of end user practices in different e-Science infrastructure projects, within the domain of climate science and within non-climate science domain are described. Support models used in structuring user-support communications are given in Section 2.6. An overview of frameworks, concepts, techniques using different technologies employed in end user support to improve it, is briefly reviewed in Section 2.4. A short overview of the modelling tools that can be used to depict user support process in different e-Science infrastructure projects is given in Section 2.10. A precise description of soft systems methodology (SSM) and knowledge gap analysis is presented in Section 2.8 and Section 2.9, respectively. Finally, the summary of the chapter is given in Section 2.10.

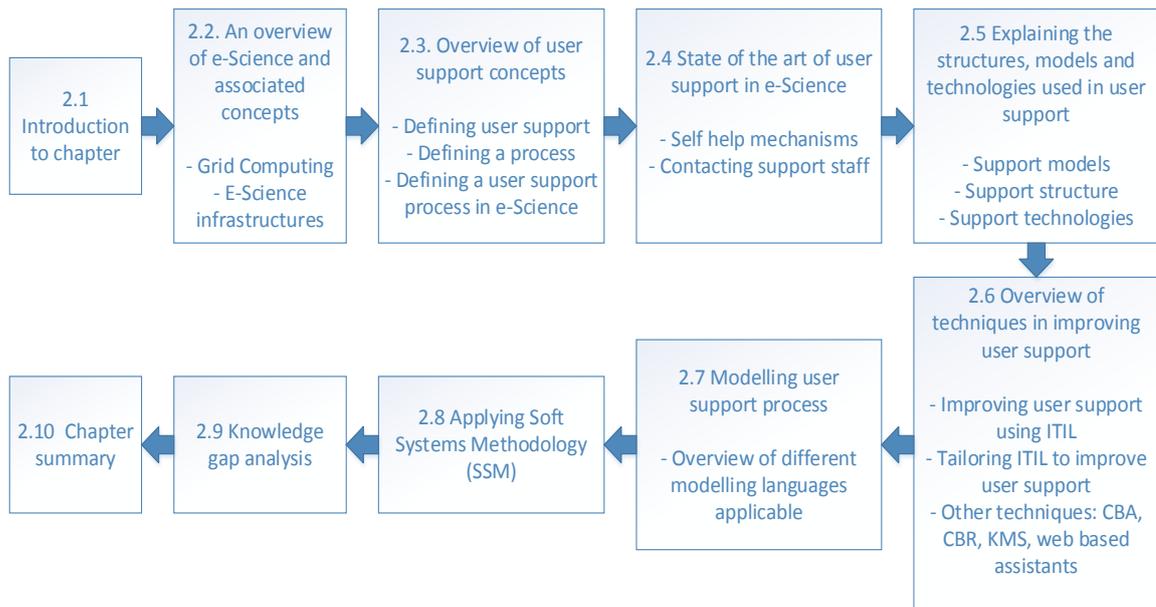


Figure 2.1 Graphical representation of the arrangement of the chapter.

2.2 An overview of e-Science and associated concepts

The industrial and scientific communities have been always looking forward for more computational power and to share data among each other. To achieve this goal researchers have studied multiple solutions for interconnecting organizations in order to share computational and data resources. This gave birth to grid computing networks, clusters, peer-to-peer (P2P) networks and other associated concepts that form the basis of e-Science infrastructures also called cyber-infrastructures (Buyya & Murshed, 2002). Hence the field of e-Science was born.

There exist multiple definitions of a grid computing network some of them are:

- Ian Foster defined grid computing as “A computational grid is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high-end computational capabilities” (Foster & Kesselman, 1998).
- To address social and policy issues he modified the definition and stated that grid computing is concerned with “coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations” (Foster, Kesselmann, & Tuecke, 2001).
- Krauter et al. defined grid computing as “A distributed network computing (NC) system is a virtual computer formed by a networked set of heterogeneous machines that agree to share their local resources with each other. A grid is a very large scale, generalized distributed NC system that can scale to Internet-size environments with machines distributed across multiple organizations and administrative domains” (Krauter, Buyya, & Maheswaran, 2002).

- Franco Travostino et al. defined grid computing in the book *Grid Networks* as “The grid is a flexible, distributed, information technology environment that enables multiple services to be created with a significant degree of independence from the specific attributes of underlying support infrastructure” (Travostino, Mambrett, & Karmous-Edwards, 2006).
- Frederic Magoules et al. defined the grid computing in the book “*Grid Resource Management*” as “A hardware and software infrastructure that provides transparent, dependable, pervasive and consistent access to large-scale distributed resources owned and shared by multiple administrative organizations in order to deliver support for a wide range of applications with the desired qualities of service. These applications can perform high throughput computing, on-demand computing, data intensive computing or collaborative computing” (Magoules, Nguyen, & Yu, 2008).

From all these definitions and statements, the author extracted properties that are central from the perspective of supporting end-users in e-Science infrastructures. The current e-Science infrastructures are based on grid computing network environments to connect geographically scattered data centres and high performance computing (HPC) facilities to achieve specific scientific goals such as research on climate change. The author therefore proposes the following definition of an e-Science infrastructure that supports end-user’s perspective; “a combination of multiple administrative domains (MADs) where each administrative domain (AD) consists of multiple local and remote sites and resources based on grid computing network to facilitate users to access the resource primarily for research purposes.”

These multiple administrative domains (MADs) share resources such as storage, computation and other services dynamically with each other for a purpose of achieving a common research or probably a commercial goal. To achieve this common goal the staffs of administrative domain support the concerns of end-users in achieving their research or commercial goals. The support provided can be technical or scientific in nature. Thus, this network intrinsically encompasses an explicit or implicit end user support framework.

Grid computing networks that constitute an e-Science infrastructure are complex networks; users need an interface to access the resources offered by e-Science infrastructure. This includes command line tools, web portals, different application interfaces and graphical user interface (GUI). Mostly, a grid computing network is interfaced with its users through a middleware. There exist many different grid middlewares, the most commonly used ones are: gLite (EGEE & EMI, 2011), Unicore (Unicore Forum, 2013), ARC (Nordugrid, 2012), and Unibus (Kurzyniec, Sławinska, Sławinski, & Sunderam, 2007). The largest grid service in Europe is the Enabling Grid for EsienceE (EGEE) project which is based on gLite software (EGEE, 2010). EGEE

connects many local grids from different countries for example, Germany's NGI-DE-Grid (Juling, 2010), Netherlands BIG-Grid (BiGgrid, Nikhef, & NBIC, 2010), UK National Grid Service (NGS, 2010), and Belgium BE-grid (The Belgian Grid For Research, 2011).

Some of the application areas of grid computing are in the fields of astronomy (AstroGrid & Institute for Astronomy University of Edinburgh, 2012), medicine (BIRN, 2012), climate (Earth System Grid Federation, 2010), economics (Buyya, Abramson, & Giddy, 2007), geoscience (NEES & NSF, 2012), fusion energy (FusionGRID, 2007), neuroscience (neuGRID, 2011) and volunteer computing (SETI@home, 2011). However, over the last couple of years the trend has been towards social sciences and to humanities, to the extent that many now argue that the term *e-Research* should replace e-Science for representing grid computing networks (Jirotko, Procter, Rodden, & Bowker, 2006). An example of grid project in humanities is TextGrid (TextGrid, 2013). A comprehensive and classified list of different grid projects is available at "Enterthegrid", which is the largest directory on grid computing (Bv, 2010).

An *administrative domain* (AD) is an entity which follows homogeneous policy within all its local and remote sites. In contrast, grid computing networks are composed of different or multiple administrative domains that are often located in different countries or even continents. They have different organizational policies and they must respect different laws in each country they are located in. This heterogeneity of policies and resources arise issues in the end-user support management of grid computing networks. These issues are currently not being undertaken in a serious manner by organizational managements of the administrative domains as well as virtual organizations (VO). A *virtual organisation* in the jargon of grid computing is set of organisations that share resources amongst each other. The institutions constituting a virtual organisation share common concerns, requirements, goals but may vary in resources, policies, organisational practices, scope and structure.

The reason that the end-user support did not get an attention is because the e-Science infrastructures are still passing through their evolutionary phase. The organizational management of e-Science infrastructures is dealing with resource and policy heterogeneity issues. These issues include a number of resource management and application scheduling challenges in the domain of security, handling resources, fault tolerance, continuously changing resource conditions and politics (Buyya & Murshed, 2002). Therefore, the researchers are occupied in further development of different technical components that are important for the evolution of e-Science infrastructure technologies. These components range from middle-wares, data-archives to security management software, identity management (authentication and authorization). All these issues also have an impact on end-user support.

In order to minimize the cost and maximize the efficiency, large corporates and research institutions have to collaborate with each other. There exists a substantial need to form a collective end-user support policy as well as a collective end-user support process, because the end-users are having direct access to each-others resources forming a large computational and data-grid network. This large computational and data-grid network is also known as “grid-federation” (Earth System Grid Federation, 2010). There are a number of challenges that e-Science infrastructures face at the moment, for example technical and design challenges, however providing adequate and usable support for collaboration across geographical, institutional and disciplinary boundaries is one of the major challenges (Jirotko et al., 2006).

One solution to this problem can be overcome by forming virtual organizations and defining certain common sharing agreements for user-support e.g. signing service level agreements (SLA) or operation level agreements (OLA). *SLA* is an agreement between the users and the service providers and *OLA* is an agreement between the service providers or the units of a service provider. However, this solution is just a beginning in the direction of defining user-support framework and does not guarantee smooth user-support process as deep synergizing efforts are required by the support staff members belonging to each of the collaborating organizations. The best solution is to monitor, analyse, evaluate end-user support activities, practices and support solutions that are currently in practice, not only in the field of e-Science but also outside e-Science. The idea is to borrow end-user support good practices from the fields outside e-Science also.

2.3 Overview of user-support

User-support also known as end-user support is found in different areas in industry as well in academia and it has a diversity of meanings depending on the nature of business and the environment where support function is carried out. Support services provide information and help to the users of some system or a product. The aim of support is to increase understanding in resolving problems that may arise during interacting of a user or a customer with a particular product, technology or even a system. In literature, “end-user support” has a variety of synonyms in different contexts such as; customer support, help desk¹ (HD), customer service, technical support, technical support services, client support services, customer support services, computer assistance, computer help hotline, call center and information center (Beisse, 2010). Moreover, the word “support” with

¹ A help desk system or a service desk system (also known as: tracking system, issue tracking system, service ticket system, support ticketing system, task tracking system, ticketing system, trouble ticketing system) is a software for the management of issues such as error messages and user requests. The main purpose is to manage and resolve incidents quickly and effectively, and to make sure that all requests are followed up. It ideally handles incidents (hardware and software failures), service requests, complaints, feedbacks and requests for change reported by users or customers. Help desk systems offer the possibility of assigning a ticket to a function body called, a support unit. A ticket can be passed directly to an employee within or outside a function body a support unit.

any modifier such as “user,” “product,” “hardware,” “software,” “network,” or any other similar word can be used (Beisse, 2010). Most companies offer end-user support as a free or as a paid service (Beisse, 2010). In this thesis, the term end user support and user support has been used interchangeably.

An end user, a customer or a client is a person who uses a product, service or a system. Support services attempt to help the end user to solve a particular problem that one encounters while using or interacting with a product, service or a system. In an organization that provides end-user support, there could be human resources or automated resources allocated to provide end-user support via different media, for instance; telephone e-mail, live, or a website. The human resources that are employed by the organization to provide end-user support are known as support staff, help desk staff, human support agent or support specialists. “User-support is an important element in reaching the goal of universal usability of a system” (Aberg & Shahmehri, 2001). “Bill Gates of Microsoft said that he believes that customer service is destined to become the primary value added function in every business” (Negash, Ryan, & Igbaria, 2003).

2.3.1 What is user-support?

From the literature review, I attempt to provide a definition for the end user support: End user support provides necessary information and services to the clients or end users to help them use a system, a service or a product more productively to attain their goal or solve their problem.

In the jargon of e-Science, end-user support can be considered as the strategy that a particular organization selects to provide user support depending on the organization size, type, location, financial situation and goals for support services as well as level of skills and support needs of clients.

The common ways that an organization provides support are mentioned as follows (Beisse, 2010):

- Peer support; which is an informal network of peer workers who provide support to their colleagues.
 - Part-time user support; it is a formal support function where the employees of an organization are doing support along with other activities. This can make significant demand on their time and can interrupt or compete with other assigned tasks. To mitigate these interruptions, sometimes an organization can assign all the support tasks to user support workers or a work team.
 - Help desk (HD) support provides single point of contact (SPOC) for users in need of technical support whether internal workers or external clients. It is generally accepted that a formal help desk was first established about
-

twenty-six years ago (N. Leung & Lau, 2007). It is also known as Service Desk (SD) (Jäntti, 2012a).

- User support centre also known as information centre and provides wider range of services than a HD. End-user training, information material, online-information, trouble-shooting and assistance are part of it.
- User support as an IT responsibility; an IT team is responsible for supporting users in IT relevant matters.
- User support outsourced to a vendor; in this case a vendor is contracted to perform user support.

2.3.2 What is a process?

The term “process” is found in different literature in many fields. In this dissertation, the author uses the term process in connotation of supporting end-users in e-Science infrastructures. End user support activities form a process of supporting end-users. In the user-support process in e-Science infrastructure projects there are elements of business-process as well as engineering-process. Therefore, it is vital to examine the notion of process in these fields. However, there are a number of definitions from researchers of different backgrounds who define the concept of “process”. The process definition as according to Johansson et al. is described as:

It is “a set of linked activities that take an input and transform it to create an output. Ideally, the transformation that occurs in the process should add value to the input and create an output that is more useful and effective to the recipient either upstream or downstream” (Johansson, McHugh, Pendlebury, & Wheeler, 1994).

It is interesting to note that mentioned the upstream part of the value chain as a possible recipient of the process output. This observation is relevant to e-Science user support process because the upstream recipients are the designers and developers of the e-Science infrastructures (Johansson et al., 1994).

Hammer and Champy define process from the customer’s or end-user’s perspective as: “A collection of activities that takes one or more kinds of inputs and creates an output that is of value to the customer” (Hammer & Champy, 1993).

In the field of engineering, a process is a set of inter-related activities, tasks that transforms inputs into outputs (ANSI, 1998). These tasks can be carried by human nature or machines, which are agents in accomplishing a process (Gilb & Brodie, 2005).

A process is defined by Davenport as:

“A structured, measured set of activities designed to produce a specific output for a particular customer or market. It implies a strong emphasis on how work is done within an organization, in contrast to a product focus’s emphasis on what. A process is thus a specific ordering of work activities across time and space, with a beginning and an end,

and clearly defined inputs and outputs: a structure for action. ... Taking a process approach implies adopting the customer's point of view. Processes are the structure by which an organization does what is necessary to produce value for its customers" (Davenport, 1993).

Rummler and Brache examine process from the business organisation's perspective and define business process with a focus on the organization's external customers or users:

"A business process is a series of steps designed to produce a product or service. Most processes (...) are cross-functional, spanning the 'white space' between the boxes on the organization chart. Some processes result in a product or service that is received by an organization's external customer. We call these primary processes. Other processes produce products that are invisible to the external customer but essential to the effective management of the business. We call these support processes" (Rummler & Brache, 1995).

In the definition above two types of processes distinguished from each other as far as the concept of customer value is concerned. If a process is directly involved in the creation of customer value then it is a primary process. If a process is concerned with the organization's internal activities then it is a supporting process. These processes are embedded in some form of organizational structure. A process can be cross-functional, i.e. it ranges over several business functions.

From the definitions explored above, it is observed that there is a link between all activities that take place and there is a transformation within a process. A (business) process has the following characteristics as according to (Johansson et al., 1994):

1. **Definability:** A process must have clearly defined boundaries, input and output. E.g. user-support process in e-Science infrastructures has clear boundaries.
 2. **Order:** A process must comprise of activities that are ordered according to their position in time and space. E.g. Priority of incoming user-requests, Solution time of incoming user-requests.
 3. **Customer or User:** There must be a recipient or audience of the process' outcome, a user or a customer. E.g. users of e-science infrastructures.
 4. **Value-adding:** The transformation taking place within the process must add value to the recipient, either upstream or downstream. Values added to down-stream are solving end user problems and to the up-stream is betterment in terms of design and development of the e-science infrastructure.
 5. **Embeddedness:** A process cannot exist in itself; it must be embedded in an organizational structure. The process of end-user support in e-Science infrastructures is part of the organisations that form a federation.
-

6. Cross-functionality: A process regularly may span several functions. The end-user support is cross functional as it spans various levels and units of organisation. A cross-functional team is involved with different functional expertise working towards common goals i.e. servicing users (Krajewski & Ritzman, 2005).

In business process literature; a process manager or process owner is required who is responsible for the performance and continuous improvement of the process (Hammer & Champy, 1993). In the end-user support process in e-Science there is no practice or practical assignment of some person as a process owner. In a business process, there can be sub-processes, procedures, activities or tasks with certain conditions to fulfil the ultimate objective or to produce an output (Anderson, 2010). The inputs and outputs can be received or sent to other business processes, organizational units or internal or external stakeholders (Hammer & Champy, 1993). Thus there is a process chain. The activities or tasks of a business process can either be carried out manually by human resources or by using automated systems.

According to the Association Française d'Ingénierie Système; a process is a set of transformations of inputs into outputs (products or services) by consuming resources while respecting all sorts of applicable constraints that meet a defined mission corresponding to a specific purpose in a specific environment. This definition is open to all domains though originally coming from the Systems engineering branch. Thus from this definition Constraints, Products, Resources, Environment, Transformation (CPRET) perspective is important in each process.

2.3.3 User-support process in e-Science infrastructures

After examining all the definitions of “process” or “business process”, the user-support process in e-Science infrastructure can be defined by the author as follows:

“The user-support process in e-Science is an operational process that serves the end-users of e-Science infrastructure in achieving their goals. The user-requests initiated by users are the inputs to this process and these user-requests are processed or transformed by user support staff with the help of tools and methods to provide solutions, hence meeting the user support needs. This process is iterative in nature as the nature of support in e-Science projects.”

The end-user support process is an example of a *process*. The *environment* is an e-Science infrastructure. The *mission* of this process is servicing and satisfying end-users` of e-Science infrastructures incoming queries. The *constraints* are user-support times, support staffs, financial resources and supporting technologies. The *resources* are Support staffs team and the support tools. The *inputs* are user requests and problems. The *transformation* is to understand the user-requests and provide a solution. The *product* is an end-user support framework that provides solutions to the users.

A process model describes the way processes normally should work or roughly how a process may look like. Processes are modelled by using modelling language. Processes are interdependent on modelling languages but are entirely different concepts. The goals of the process model is to describe i.e. track what actually happens during a process, to prescribe i.e. how a process might be conducted or performed in a desired manner, and to explain and evaluate the possible courses of actions based on rational arguments (Rolland, 1998).

2.4 State of the art of end-user support in e-Science infrastructures

This section provides an overview of the state-of-the-art support technologies and concepts employed to support end users in e-Science infrastructures, especially in the field of Earth Science Modeling (ESM). At the moment, the support efforts in e-Science are being pursued on intuitive basis. However, there is eloquent use of web portals for information dissemination, web-based communication technologies like mailing-lists, e-mails, HD systems, user-forums, user-forms, some social media and telephone assistance. These state of the art support technologies and concepts are used for communication in between e-Science infrastructure developers as well as in the user communities of e-Science infrastructures.

In e-Science infrastructures or grid communities, mostly support facility is used by grid-users, administrators and developers (Rathmann & Stöckle, 2012). In e-Science infrastructure user-support most of the times, there are two possibilities to get help:

1. To get self-help by browsing through documents and information available online.
2. To ask experts for help.

2.4.1 Self-help mechanisms for users

For end-users to get self-help most of the e-Science infrastructures and data-grids have a web presence for dissemination of basic information about the projects with some basic frequently asked questions (FAQs), How To's and Wikis. Most of the project websites² contain user manuals, brochures, installation guides in the form of instructions, grid architecture documentation, software as well as hardware documentation, integration instructions, grid access certificate instructions, project deliverables and even screencasts. The layout and information architecture of the e-Science infrastructure project websites are mostly based on content management system (CMS) templates and usually contain quick-links and news. For instance, C3Grid and IS-ENES are hosted by

² Some examples of e-Science infrastructure websites include C3Grid (www.c3grid.de), ESGF (www.esgf.org), IS-ENES (www.verc.enes.org), Open Science Grid (www.opensciencegrid.org/documentation/), WissGrid (www.wissgrid.de), TextGrid (www.textgrid.de), EGI (www.egi.eu) and BalticGrid (www.balticgrid.org).

Plone CMS (Hashim Chunpir, 2013; ENES, 2013). All of these web-based information dissemination and information search tools have been combined together to offer self-help for the end-user support.

After observing the support portals of various e-Science infrastructure projects, one can note that scientific or domain specific user-support information is relatively minimal or missing on websites of most of these e-Science infrastructure projects. However, information retrieval oriented user self-help based on keyword tagging or category classification to get answers to technical and domain-specific questions is missing. Application of searches based on metadata or on full-text (or other content-based) indexing is also limited. The usability of information architecture (IA) of websites containing all project-related, stake-holder relation (especially user-related) dissemination-material is not assessed and remains under-studied in e-Science infrastructure projects. The use of adaptive user-interfaces (UI) is not seen so far in the e-Science infrastructure websites. The user information in the websites is usually updated by human support agents irregularly.

2.4.2 Contacting support staff via help desk

End-users also have possibility to contact experts and ask them about particular problems, mostly technical problems via mailing-lists, user-forms, forums, blogs, web-based and e-mail based help desks¹. Examples of e-mail and mailing list based user support centres (especially for technical queries) are found in AstroGrid (AstroGrid & Institute for Astronomy University of Edinburgh, 2012), Crossgrid (Gomes et al., 2004, pp. 76–77), C3Grid, Medigrid and HEP-Grid (Rathmann & Stöckle, 2012). Sometimes support centres combine help desk, mailing lists with blogs and even social media to get help e.g. (“Open Science Grid,” 2013) in Open Science Grid project. In rare case telephone numbers are also provided to contact a support employee within specific support timings e.g. (“Open Science Grid,” 2013). Domain specific scientific support is not as common as technical support in e-Science infrastructures support centers; therefore there is a need to introduce domain specific scientific support.

Web-based help desk systems are also deployed for end-user support; for instance in case of European Grid Infrastructure (EGI) project, the follow-up project of Enabling Grids for E-Science (EGEE). National Grid User Support (NGUS) and Global Grid User Support (GGUS)³ are examples of web-based end-user support help desks in e-

³ During Data-grid project (“The DataGrid Project,” 2004), the predecessor project of EGEE, an effort to build a global grid user support started from 2001 onwards and during EGEE project a global user-support system known as Global Grid User Support (GGUS) system was developed which is based on XGUS template maintained by Karlsruhe Institute of Technology (KIT). GGUS development continued further during EGI (European Grid Infrastructure), the follow-up project of EGEE.

Science infrastructures that require web-based access the system to insert or track an issue. Substantial amount of effort and resources have been invested to build, deploy and operate these web-based help desks. XGUS template (where “X” can be adapted to a particular body for instance National or Global) is a system which is based on the BMC Remedy Action Request system with an Oracle database for tickets, news, portal and user administration (“xGUS User’s Manual,” 2013). XGUS template is actually a customized help desk system made available for user support for grid (“XGUS website,” 2013). XGUS has at the moment four instances namely NGI-DE help desk, NGI-CH help desk, NGI_AEGIS help desk and EUMEDGrid help desk and is maintained by Karlsruhe Institute of Technology (KIT) (“XGUS website,” 2013).

Global Grid User Support (GGUS)⁴ system is the main mechanism by which users can send request for support if they encounter any issue while using the grid. The GGUS system creates a trouble ticket to record the request and tracks the ticket from creation until solution is found (EGI, 2015). The GGUS system is central help desk application that interfaces with many regional or local help desks known as regional operations centres (ROC), see Figure 2.2 (EGI contributors, 2013).

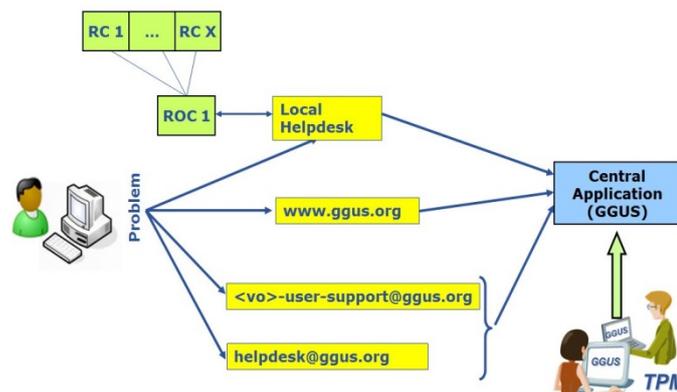


Figure 2.2 Different channels to submit user-requests in in GGUS system, taken from (David, 2005).

There are different ways by which a user can submit request via e-mail. Users can submit a support request via their regional help desk, regional operations' centre (ROC), their virtual organisation (VO) or directly an e-mail to GGUS central help desk, as shown in Figure 2.2. Within GGUS there is an internal support structure for all support requests. The ticket processing manager (TPM)⁵ monitors every ticket and provides solution to close the ticket. There are two types of ticket processing managers (TPMs): Generic TPMs and VO TPMs. Generic TPMs can provide solution to the generic problems of grid and VO TPMs provide VO related software and hardware support.

⁴ The GGUS system is the main support access point for the EGI project.

⁵ TPM (Ticket Processing Manager) is the term used for an FLS (First Level Support) in GGUS system.

Apart from TPMs there are other specialized support units that provide support. The organisation and hierarchy of GGUS support centre is shown in Figure 2.3.

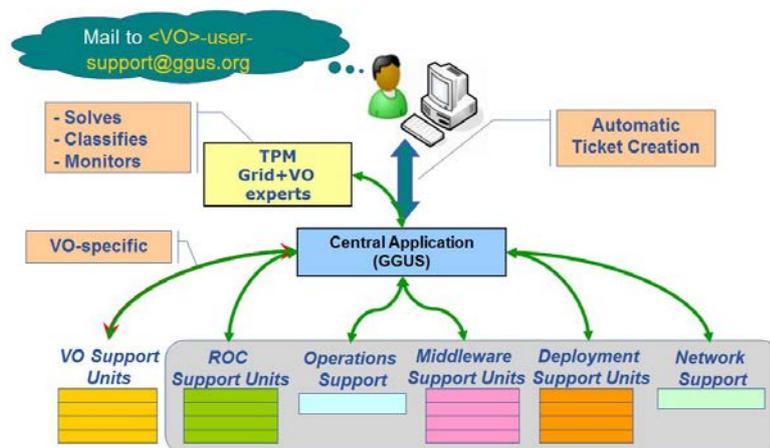


Figure 2.3 The organisation of support in GGUS system, taken from (David, 2005).

The users can login to the web-based GGUS help desk system and can track the progress of their tickets. The urgency of the tickets, the complete treatment of tickets and the detailed duties of TPMS is described in GGUS user guide (EGI, 2015).

In the user support centre C3Grid-INAD (Climate Collaborative Grid), an e-mail based help desk system based on RT Best practical issue tracking system was recently deployed that replaced the traditional mailing-list system (Hashim Chunpir, 2013). The following Figure 2.4 and Figure 2.5 show the former user-support process in ESGF as well as C3Grid-INAD and the new user-support process in C3Grid-INAD support with request-tracking (RT) help desk software that replaced the mailing-list. In Figure 2.4, a user sends a user-query via a mailing-list and the support experts monitors the mailing-list and responds until the user-query is resolved. The problem with mailing-list is that all support experts spend a lot of time monitoring the whole list. There can be a possibility that some of the user-queries are ignored. There is no query record or tracking mechanisms available with the mailing-list.

In Figure 2.5 the request tracking system replaced the mailing-list and every incoming user-request is monitored by first level support (FLS)⁶ thus releasing the second level support (SLS)⁷ to take care of each and every incoming request. This step is depicted with the step 1 in Figure 2.5. The FLS can respond to the user's request by send a solution or requesting further information about user's request, depicted as step 4 in Figure 2.5. If FLS cannot solve the problem then the user request is assigned to

⁶ FLS is also known as Front Line Support, human support agent. It is known as Ticket Processing Manger (TPM) in case of "GGUS" system.

⁷ SLS is also known as Second Line Support, specialized human support agent. In "GGUS" system it is known as support unit specialist of a respective support unit e.g. middleware support unit.

concerned SLS or multiple SLS, depicted as step 2 in Figure 2.5. SLS can now directly solve user’s request or demand more information about user request.

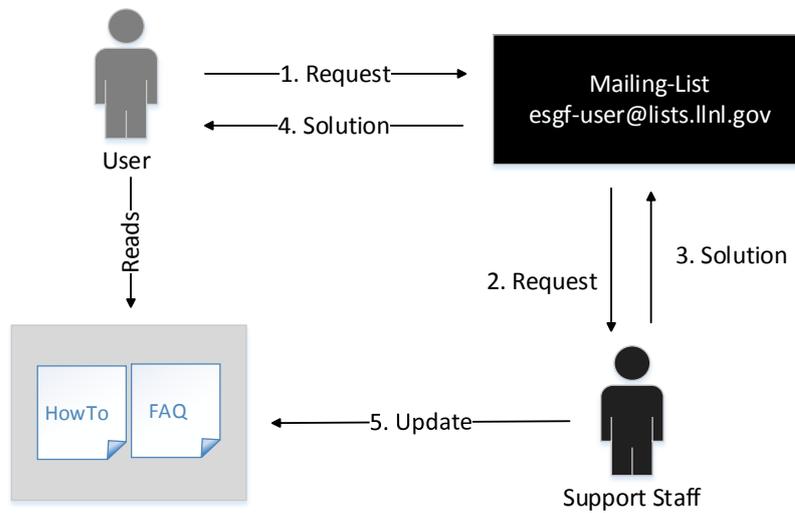


Figure 2.4 Previous user-support process with the mailing-list in ESGF and C3Grid-INAD e-Science infrastructure projects. This figure is redrawn and originally taken from (Hashim Chunpir, 2013).

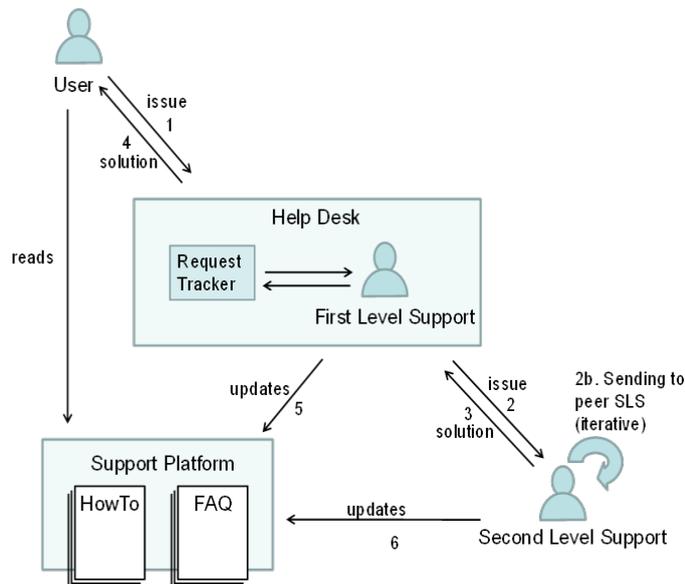


Figure 2.5 Current user support process in C3Grid-INAD e-Science infrastructure project. This figure is redrawn and originally taken from (Hashim Chunpir, 2013).

There are different e-Science infrastructure projects where the RT based help desk structure is similar as shown in Figure 2.5. This user support procedure e-Science infrastructure projects can be generalized in a diagram shown in in Figure 2.6.

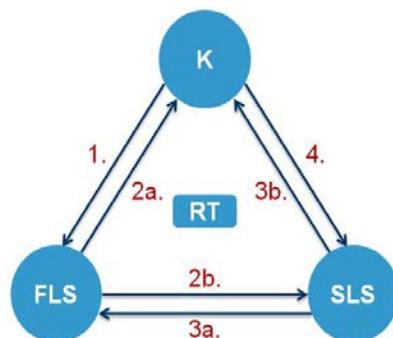


Figure 2.6 Representation of user and support staffs interactions in e-Science infrastructure projects based on RT system.

The following abbreviations used in the Figure 2.6, describe following concepts:

K = a User “K”

FLS = First Level Support

SLS = Second Level Support: The SLS comprises of technical specialists who are able to answer the technical nature of queries.

RT = Request Tracker (software)⁸

The steps of communication between end-user “k” and support staff (both FLS and SLS via e-mail) through request tracking software based help desk are as follows:

1. User “K” sends an E-Mail, which is sent directly to FLS via RT system.
2.
 - a. FLS can send a reply to the user “K” directly. The FLS personnel look into the matter and answer the query depending on the nature of the support required. The FLS can directly respond to the request initiated by the C3Grid-INAD portal users. In case FLS needs more information; it can be requested by FLS.
 - b. If the nature of support is not handled or addressed by the FLS, the query is forwarded to C3Grid Second Level Support (SLS) personnel. The SLS comprises of technical specialists who are able to answer the technical nature of queries. It can be an individual SLS or a group of individuals.
3.
 - a. SLS can reply to FLS.
 - b. Or SLS can now get directly in touch with the user “K” which is the case most of the times.

⁸ Here request tracking software can be any one.

4. K can now get in touch with SLS as well and provide more information required by SLS to provide solution. “K” can also provide feedback to the solution suggested by SLS.

Despite benefits of request tracking software, nevertheless there are some shortcomings of request tracking software based help desk (Rathmann & Stöckle, 2012, p. 12): Is the community ready to take the responsibility to monitor and operate its own help desk? Administrative aspects, complicated right management might cause some impediments for the smooth organisation and operation of help desk. Financial aspects and human resource aspects such as operational costs for instance daily ongoing costs of server administrator’s time to maintain and analyze HD software etc. are important to cater with. This factor will be even more complicated in terms of personnel and organizational commitment especially in international communities where a 24/7 operation is almost inevitable. Data protection and security issues are also needed to be taken care of.

In web-based help desks where the tickets can be submitted via forms, the complexity of input forms (where ticket info can be entered) might be scary to the end-users especially the users who do not have IT affinity (or are not familiar with complex UD). Tickets submitted via e-mails (which is also a practice e.g. in case of C3Grid-INAD help desk) simplify this problem. Yet, there is a price to that which means that the tickets generated via e-mails might have authentication problems (i.e. a user or member of the community authentication). The e-mail addresses are often not institutional, which complicates the mapping. For community-driven help desk it is a serious problem. At the same time, it opens the possibility of spam attacks.

It is indeed challenging to provide end-user support in e-Science infrastructures especially in Earth System Modelling (ESM) community because the nature of end-user support is dynamic due to ever-changing software applications and hardware replacements. In order to meet demands of the researcher communities for substantial increase in the computing power and data production at a faster scale, there is ever increasing need to compare the data produced against all kinds of other data for research purposes (Hiller & Budich, 2013). It is important that the data produced at different data centres all over the world should be provided to massively growing communities of researchers globally. Lack of continuity of funding opportunities for end-user support makes it even more difficult.

In almost all e-Science end-user support frameworks, there is a gap of knowledge about the effectiveness of end-user support as there are relatively fewer or no studies conducted to measure the quality, efficiency and usability of the end-user support. There is none user-studies or user experience studies undertaken by the researchers so far that would prove that the above approaches fulfil the requirements of end-users as well as the expectation of supporters (support staff). The support of e-Science projects

is normally short-lived only until particular projects are under work. Once the project funding is over the user support becomes inactive or even completely disappears for instance in case of “Crossgrid” project (Gomes et al., 2004). The application of service delivery frameworks or good practice guidelines (e.g. Six sigma technique, ITIL, MOF Microsoft Operations Framework) to improve end-user support from the commercial IT organisations has not yet been borrowed and applied to the e-Science end-user support world. Neither, any studies have been conducted so far that suggest adaptations or modifications of the popular end user-support frameworks in the existing end-user support of e-Science. Therefore, there is a great potential to assess the efficiency, quality, organisation, user-support culture of end-user support systems and explore the possible solutions that might be applicable to make the end-user support more satisfying within the resource limits for the end-users as well as the end-user support staff.

2.5 Models used in structuring user-support communication

This section provides an insight into support tools and technologies that are used in organizations especially IT organizations as well as grid computing projects as a communication medium between end-user of a service or a product and a support staff. In better organized organizations that provide services or a product support, a user has the possibility either to find the required information via self-help or to ask some relevant person for support.

2.5.1 Support models

In order to solve business and organizational problems, investments are being made in information management systems and IT for the last seven decades. Software, hardware, networks and other IT related products are now part of almost every organization. IT support therefore has become a common phenomenon for organization-internal users as well as organizational-external users or clients. To meet the support needs of clients a formal help desk emerged. It is not known when exactly a help desk was first formally put into operation, however it was first established or at least came to be known around 26 years ago in 1987 (Kendall, 2002).

Before help desks were formally established, the technical support in the field of IT was done more informally, where users or clients contacted experts in the IT area (N. Leung & Lau, 2007). However, as the organisations expanded, in order to keep business operations running, there was an immediate need to service users in their IT-related problems e.g. hardware, Software and telecommunication, promptly. The reasons for establishing dedicated help desks were discussed by (N. Leung & Lau, 2007): The first reason that gave an impetus to setup a dedicated help desk was that IT staffs were not reachable for assistance or were engaged in other activities. Furthermore, if a support person is engaged with tasks other than user-support then support duty can lead to frustration and might interrupt completion of other tasks (Mckoen, 2000). Other issues

observed by Peters are contacting wrong persons to solve problems, thus as an aftermath wasting user's time (Peters, 1993). The help desk (HD) concept evolved to resolve technical problems within less time (N. Leung & Lau, 2007). Consequently, HD provided an access point to users so that they can get support on the fly. The role of HD as a strategic proactive incident support and daily communication system based on the data collected from the users was highlighted by (Marcella & Middleton, 1996).

There are different HD models which are in practice to solve the incoming user requests. In 1980's a decentralized model was quite popular (K. Y. Leung & Lau, 2005). For instance, in Western Kentucky University there were nine HDs, each serviced specialized category of problems (Kirchmeyer, 2002). In a decentralized model there are multiple points of contacts, each of the contact point dealing with a special area of a problem category. A user can contact a particular contact point based on the type of problem. The decentralized model worked very well initially, however, it caused problems because of expansion and complexity in organisational structures, IT infrastructures and other related products (N. Leung & Lau, 2007). The major problem noticed was that a user had to switch contact from one HD to another HD himself or herself, thus, making user irritated and wasting user's time (Kirchmeyer, 2002).

In order to mitigate the drawbacks of decentralized HDs, other HD models evolved. Organisations changed to centralized HDs i.e. a Single Point Of Contact (SPOC) model. This model is currently in use at an extensive scale. The main feature of a centralized HD was to combine all decentralized HDs and give them a single point of access. Therefore, the user does not need to switch contact from one HD to another HD. This model consolidated not only the points of contacts but also policies, standards, quality of service and service level agreements (SLA)⁹. Other advantages include inter-communication amongst the support staffs, better resource allocation and improving rate of user-request resolution (N. Leung & Lau, 2007).

An extension to the centralized user-support HD was made by the multinational organisations, thus forming the concept of distributed and centralized HD. For illustration, Morgan Stanley, a large investment and financial services bank adapted this model and connected four HDs situated at the US, the UK, Japan and Hong Kong (Heckman & Guksey, 1998) and (K. Y. Leung & Lau, 2005). This support model is currently getting popular with the expansion of businesses in different locations nationwide or across the nation. The global corporations use this model because it enables

⁹ An SLA is a service contract where nature and definition of services is formally defined normally between customer and service provider (Service Level Agreement Zone, 2013). It may define an agreement between stakeholders about for instance performance measurement, problem management, customer duties, warranties, disaster recovery and termination of agreement (Mavrogeorgi et al., 2012). Examples may include solution delivery time, uptime and average speed to answer (ASA). It is interesting to note that an OLAs (Operation Level Agreements) which are similar in nature to SLAs is used amongst groups within the service provider to support SLAs.

them to operate HDs 24 hours a day, 7 days a week with a single point of contact for users, serviced by HD teams at multiple locations around the globe. For the customers or the users who contact HDs via phone call, a modern call routing technology is used for the users to route their calls from one HD to another (N. Leung & Lau, 2007). This model is also known as virtual HD (N. Leung & Lau, 2006).

Outsourcing end-user support is another HD model widely in practice. Outsourcing can be done partly or fully, temporarily or permanent, onshore or offshore, to single or multiple vendors as according to the needs and culture of an organisation. The reasons that lead to outsourcing HD as cited by Kettler and Willems are: cost effectiveness, specialization of skill by the people where HD is outsourced to a vendor, no handling and recruitment of support or HD human resources, no training worries, in house people can concentrate on infrastructural development and not on daily troubleshooting (Kettler, K. and Willems, 1999). However, the outsource HD model is not applicable to every project. Normally, IT managers outsource functions that are not critical to the core business operations. It depends on the mutual trust, transparency, reliability and string alliance with the outsource service provider (K. Y. Leung & Lau, 2005).

E-support is another model which is getting popular amongst organisations day by day. According to Broome and Streitweiser all support concepts in e-support model include use of internet or the web as a primary communication channel (Broome, C. and Streitwieser, 2002). All web portals, web-forms, e-mails and other online resources including Wikis, FAQs, online documents, and knowledge bases help users to constitute e-support. It helps to provide better and cheaper service for an organisation. Web-training to users and remote control technologies to take control of user's desktops are part of e-support. The adaptation of e-support as a support model depends on the nature of supported business, organisational culture, resources available and technology available for deployment.

2.5.2 Support structure

The support structure is unique to organisational strategic goals, type of businesses, target audience (i.e. customers and users) and the user support doctrine. Czegel indicated three different user support structures for user support namely one-level, two-level, three-level and n-level, see Figure 2.7 (Czegel, 1998). These three levels of user support seem to dominate the business (Mira Kajko-Mattsson, 2004).

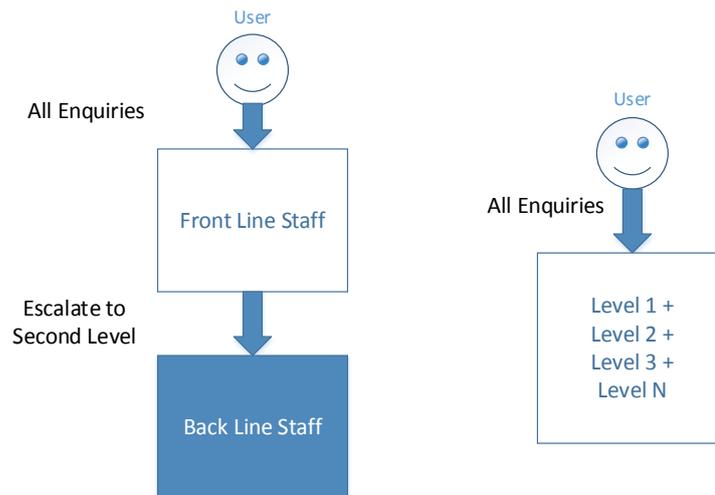


Figure 2.7 Two-level end-user support structure (left) and N-Level end-user support structure (right), adapted and redrawn from (N. Leung & Lau, 2007).

2.5.3 Technologies used in help desk and user support communication

In order to support end-users, appropriate technologies should be chosen within the financial constraints that would suite the underlying user-support doctrine. This would ensure efficient and effective end user support (N. Leung & Lau, 2007). Support process evolves with the passage of time, needs and requirements of users and so do the technologies used in supporting the underlying support philosophy of helping end users. The transformation from the initial ad-hoc user support process to a mature and well developed user support process takes time, effort and lot of other resources (K. Y. Leung & Lau, 2005). The pre-historic help desk constituted simply a desk, telephone and a paper notebook (Kendall, 2002). With the passage of time other sophisticated technologies became available to solve end-user problems. Technologies also played a pivotal role in making user support centralized or SPOC.

Call centres that support users have only one telephone number to let their clients reach support consultants. Most of the call centres use Automatic Call Distribution system (ACD). ACD can handle large number of calls on a single system. It can manage flow of calls, record historical data and generate call statistic report (Underwood, Hegdahl, & Gimbel, 2003). If all operators are busy and cannot take a call then the call is put in a queue until the operator gets free to take a call. The first person waiting in the call queue gets the opportunity to talk to operator first. Interactive voice response system (IVR) is usually put in front of ACD to get options from the user to get information or perform a specific function via telephone pad (Czegel, 1998). ACD also has a capability to generate call statistics e.g. average total incoming calls, average + time of incoming calls, abandoned calls, and average hold time (Czegel, 1998; K. Y. Leung & Lau, 2005; N. Leung & Lau, 2007; Underwood et al., 2003). This statistics collected can be

monitored by the management of an organisation and based on this statistics the overall user support process can be improved. Calls can be recorded to improve quality of support and the support employees can be trained to improve the user issue handling and showing empathy to clients.

Help desk management systems (HDMS) and Request Tracking (RT) software are the technologies that automated the user-support (I. A. Middleton & Marcella, 1997; I. Middleton, 1999). Recording call information, ticket escalation, storing ticket and reporting are the main functions pointed out by Czegel (Czegel, 1998). There are many HD management systems which are used in different organisations. For example in academia at the University of Colorado: WREQ, Best Practical RT; Request Tracker (RT), Open Technology Real Services (OTRS) and ACCORD5 Trellis Desk (Trellis) were tried to manage incoming user requests (Herrick, Metz, & Crane, 2012). At the end of evaluation of all these software, Trellis was selected because of its high flexibility and high customization capability that suited the requirements of user-support department. Therefore, Trellis was tailored to the needs of user-support at University of Colorado (Herrick et al., 2012).

HD expert systems is an effective technology that helps the user him or herself or first level support staff to handle most of the queries because of the capability to imitate the reasoning and intelligence of human experts (K. Y. Leung & Lau, 2005). Embedded inference engine is the major component in expert systems that diagnoses a user query and presents a solution from a knowledge-based system. However, it is resource intensive and not an easy task to develop HD expert systems, especially the inference engine (Czegel, 1998). That is why HD expert systems are not widely used or implemented (I. Middleton, 1999). So far no expert system for user-support has been applied in the field of e-Science.

In order to support end-users remote control technologies are used to troubleshoot user problems by viewing, accessing or even taking control of user's computer (K. Y. Leung & Lau, 2005). In the field of academia (Scullen, 2001) reported that almost 80% of the user problems at Griffith University, Australia were solved using remote control technologies. However, security and privacy is always an issue in using remote control technologies. There are web-based as well as client-based remote control technologies available. Web-based support technologies can be connected via internet while client-based technologies require a programme to be installed on the desktop.

2.6 Overview of techniques, frameworks and technologies for improving user-support

In this section various state of the art techniques, frameworks and technologies to improve end user-support are discussed. There are user-support improvement strategies that were applied to improve business process behind end-user support in commerce

and academic arena. These strategies employ state of the art technologies like knowledge management approaches, knowledge centric approaches, expert system-like techniques and end-user support improvement with the help of artificial intelligence based Case-Based Reasoning (CBR) and corpus analysis techniques. All these frameworks have been applied in the commercial and academic fields; however the effectiveness and measurability of their efficiency in these fields still remain unanswered.

2.6.1 Using models and frameworks to improve service desks

There are numerous techniques, concepts and frameworks that can be employed to improve process of end user support in IT organizations in commerce. There is relevant work already done to improve customer or end user services especially in the fields of IT Strategic Management (ITSM) also known as IT support service management (ITSSM), Business Service Management (BSM) and process management studies. ITSM deals with people, process and technology to implement IT services. It focuses on structuring IT activities to service customers and the interactions of IT technical personnel with business customers and users (Jan Van Bon, Pondman, & Kemmerling, 2002). In ITSM there are diverse techniques, models and frameworks which are already in practice to improve IT related processes including support services process. The most popular framework used is IT Infrastructure Library (ITIL) (Jäntti, 2012a, 2012b).

ITIL describes the best practices and guidelines that can be used to improve user-support, especially Service Desk (SD) and other associated processes in IT companies that deal with various types of IT services. ITIL's version 2, version 3 and edition 2011 of ITIL describe the SD function and other IT service management processes. ITIL's version 2 service management service comprise of two parts: Service delivery and service support. The "service support" part provides guidelines about SD function, incident management, problem management, change management, configuration management and release management. In ITIL version 3, the complete service life-cycle thinking is emphasized in five core books: Service Strategy, Service Design, Service Transition, Service Operation and Continual Service Improvement. "Service Operation" part deals with Incident management, request fulfillment, problem management and SD, which are directly concerned with end-user support process. According to ITIL, in service support process, SD extends the services of a traditional HD (Office of Government Commerce, 2011).

Recommendations of the SD function of ITIL can be applied to e-Science infrastructure user support process. However, ITIL is not completely applicable to user support process of e-infrastructures because of the number of reasons given by (Jäntti, 2012a):

- Firstly, it is important according to ITIL framework that the employees in an organization receive training from external ITIL consultants (Jäntti, 2012a). It
-

indeed takes lot of time and financial resources to hire external ITIL consultants in e-Science.

- Secondly, the ITIL consultants do not know much about the local processes and methods applied in a particular organization, therefore they need time to familiarize with these local procedures.
- Thirdly, there is lack of process culture and process phenomenon not only in e-Science infrastructures but also in other IT organisations.
- Fourthly, some procedures and tools recommended in ITIL are too complex and inadequate and may slow down the end-user service process.
- Moreover, it is required that the employees of an organization should be well-trained in ITIL framework and possess excellent process management, improvement and change management skills.
- Finally, management of an organisations or the stakeholders of the e-Science initiative should back the allotment of all the resources in order to implement an ITIL SD refinement process and on top of it motivate and reward people for passing ITIL certificate and participate in IT service manage work in addition to their own indigenous activities (Jäntti, 2012a).

Sharifi et al. investigated the reasons of failure of ITIL implementations in organisations (Sharifi, Ayat, Rahman, & Sahibudin, 2008). According to them, ITIL framework implementation has following drawbacks: Too much time spent on creating complicated process diagrams, concentrating too much on performance, being over-ambitious, allowing demarcation of departments, no work instructions and lack of assignment of process owners. In e-Science assigning a process owner is almost impossible (Jäntti, 2012a)(Jäntti, Cater-Steel, Shrestha, & Campus, 2012). A process owner is a person responsible and accountable for the structure, function, operation, input and output of a process. However, there are certain adaptations that are made to ITIL to tailor it to needs of an organization. For example: Queensland University of Technology (QUT) adapted its own version of best practices based on the ITIL reference model. It was the strategic intent of QUT to implement this model at a university-wide level across all its faculties. The QUT-ITIL project reviewed the ITIL standard process and created a version that better suited university needs (Arora & Bandara, 2006). In this way recommendations can be derived or adapted from such frameworks and applied to new fields, in our case to enhance the process of end-user support in the nascent field of e-Science.

Other frameworks similar to ITIL for improving IT service management process include: Control Objectives for IT and related Technology (COBIT) framework (The IT Governance Institute, 2007), Capability Maturity Model Integration (CMMI) for Services (CMMI, 2009), IT Service Capability Maturity Model (Niessinka, F., Clerca, V., Tijdinka, T., van Vliet, 2005), ISO/IEC 20000-1:2005 Part 1: Specification for

service management (ISO/IEC, 2005a), ISO/IEC 20000-1:2005 Part 2: Code of practice for service management (ISO/IEC, 2005b), ISO/IEC FDIS 20000-1:2010 Part 1: Service management system requirements (ISO/IEC, 2010a), ISO/IEC TR 20000-3:2009 Part 3: Guidance for the scoping and applicability of ISO/IEC 20000-1 (ISO/IEC, 2010b), Microsoft Operations Framework (MOF) (Microsoft, 2009)(Microsoft, 2015), Kapella's Framework for Incident Management and Problem Management (Kapella, 2003) and FITS. But all of them are almost impossible to implement in e-Science completely because the context of the application of these frameworks is mainly commerce and industry. It is important to note that not enough studies have been carried out that could justify the application of the above mentioned frameworks.

Apart from the existing frameworks there are related studies which focus on service management processes in industry. For instance, improvement of incident management processes based on ITIL practices was explored by (Jäntti, 2009). An attempt to create a mature problem management process was suggested by (Mira Kajko-Mattsson, 2002). The support and its associated services were tested by (Jäntti & Kujala, 2011). Studies on service level agreement and service level management were done by (M. Kajko-Mattsson, Ahnlund, & Lundberg, 2004). Studies related to change and configuration management were conducted by (Ward, Aggarwal, Bucu, Olsson, & Weinberger, 2007). Surprisingly, few of ITSM studies have provided details how ITIL-based process implementation has been done in the big data field, e-Science infrastructure and HPC-based science to streamline user support process.

2.6.2 Tailoring ITIL to improve user-support in academia

Arora et al. performed a case study about how an IT faculty within a leading Australian University re-engineered its Service Desk (SD) processes to fulfil the goal of providing more efficient and effective SD services to its students and staffs (Arora & Bandara, 2006). The motivation of this case study was that in 2005, 65% students drop in comparison to the statistics of its peak level in 2001 led to the cost cutting. To reduce expenditure, SD was identified as one of the candidates for potential cost savings. The key goal was increasing efficiency of a SD by applying Business Process Management concepts (BPM) for the process improvement but remaining within the limited budget (Arora & Bandara, 2006).

This case study project had the following objectives to achieve by applying a combination of BPM and other analytical methods (Arora & Bandara, 2006):

1. To move away from the current informal method of completing SD jobs, to a consistent means of managing them.
 2. To improve the SD operation so that the incidents are resolved within the time limit or the threshold.
-

3. Third objective was to reduce the occurrence of incidents being mismanaged.
4. To implement QUT-ITIL guideline and best practices and finally change from informal problem management to a formal problem management process.
5. These objectives are similar to the goals of improving the user-support process in e-Science apart from the fourth objective.

Arora et al. analyzed the sub-processes or sub-phases of incident management by applying techniques of 'Business Process Life Cycle' framework (Rosemann, 2001) and found out the issues that are needed to be streamlined in following sub-processes of the user support system i.e. SD: Reporting of an incident or a request, acknowledgement of incoming request, responding to the incoming incident, incident resolution and finally closure of an incident. The other techniques applied in combination were making a "task force" that included project members, mentors, key process owners and a client to identify key issues in the process combination of techniques to achieve process improvement. A task force is a unit organized for a special task. The authors used extended Event-driven Process Chain (eEPC) modeling norm and used Architecture of Integrated Information Systems (ARIS) tool to depict the current situation of the SD at QUT. On top of it, they conducted a survey with the employees (staffs) as well as the users (students) to identify the issues (Arora & Bandara, 2006).

As a result, they suggested four main recommendations to counter the problems faced by the QUT SD. First of all, they closed multiple points and introduced quasi SPOC of requests entry, they promoted client self service via web and integrating a knowledge-based system where the support staff can look up a solution repository regarding an incoming incident. They also used QUT-ITIL to improve quality of service and establish consistency in the process. Therefore the areas of improvement done by this case study were at organizational, technological and process level. At the process level a mature and consistent client quality service was established. At the organizational level it can reduce costs, almost \$150,000 per year as a net benefit (Arora & Bandara, 2006, p. 1283). At the technological level, it improved client self-service technology and integrated a solution data-base.

The study carried out at the University deals with an academic improvement of a SD at a single location. It does not include diversity of multiple organisations and big data concerns. Secondly, it is based on purely technical oriented incoming issues and does not take into account scientific issues. Thirdly, the university has a range of full-time (dedicated) support employees which is not the case in e-Science infrastructures (especially in climate science e-Science infrastructure use-case: ESGF), where mostly employees are engaged with multiple tasks other than supporting users. Therefore, the organizational, financial, social, political and technological settings in the case study discussed above are different from the nature of e-Science support. However, applying

process exploration techniques used to analyze the end-user support mechanism in QUT can be an interesting application for analyze e-Science user-support scenario.

2.6.3 Techniques used in improving traditional help desk function

Help desk (HD) studies have focused, for example, in examining the role of help desk in the strategic management of information systems (Marcella & Middleton, 1996), knowledge-management-centric help desk (González, Giachetti, & Ramirez, 2005), evolution of a knowledge-management system in the HD (Halverson, Erickson, & Ackerman, 2004) and building IT HD in the university environment (Evans & Jones, 2005).

User support in e-Science especially needs attention because there is plethora of software, network and hardware components which create:

- Diversity of technologies used in e-Science
- Diversity of organisations in terms of personnel, working culture, standards
- No proactive support activity
- No proactive training for users

The basis of needs for improvement in HD in commercial set-up includes the following arguments. It does not include the arguments which were taken into account for e-Science set-up (as stated above):

- Longer waiting times for the users
- Downsizing of staff
- Application of BPR (management methodology) at the initiation of management
- Ever increasing incidents
- Misinterpretation and miscommunication
- Indirect support

Research conducted to relieve above problems, many models, structure, technologies and methodologies have evolved but all failed because human always use reflective design concept as a method to learn from the past experience but lack of local adaptations. If problem domains and user's needs are not investigated thoroughly then there adoption of past experience can cause trouble (González et al., 2005).

2.6.3.1 Self-help approach combined with knowledge management system via software agents

According to (Knapp & Woch, 2002; K. Y. Leung & Lau, 2005) and a survey conducted by (N. Leung & Lau, 2007) that majority of the incoming user-support

queries in IT organisation are “simple and routine”¹⁰. Furthermore, Leung and Lau proposed using self-help approach to tackle simple and routine queries by the users themselves (K. Y. Leung & Lau, 2005). They also used KMS techniques and software agent technology to provide self-help to users (K. Y. Leung & Lau, 2005). The authors identified the simple queries based on the reports generated by HDMS and ACD. The identification of the simple and routine queries is based on the reports generated by HDMS and ACD and filtered by support staff to finally store into knowledge base environment. The skills and experience of support employees is referred to “tacit knowledge” by the authors. Contrary to tacit knowledge, “explicit knowledge” resides in the documents or guidelines. Externalization is a process done to convert skills, techniques experiences and perceptions of experts from various sources into explicit knowledge. All explicit and tacit knowledge is proposed to be put in the knowledge base.

The proposed self-help KMS is composed of five basic constituents: browser, interface agent, search agent, knowledge base that stores the simple and routine technical enquiries and interface database that stores information to facilitate end user communication. Leung et al. applied interface agent to deliver a dynamic user interface to facilitate user communication based on info stored at interface database and search agent as software agents between interface database and browser (K. Y. Leung & Lau, 2005). Similarly, there is a search agent between browser and a knowledge base.

The proposed benefits of this approach is that HD support staff gets freed from the handling of routine queries, efficient and effective support, low cost support service and users depending on themselves (K. Y. Leung & Lau, 2005). There are certain aspects that this research fails to answer are: Firstly it is proposed to introduce a mix of software agent and KMS but this has not been implemented in the real scenario and hence no evaluation of the proposed HD system took place. Therefore, it is not sure whether these proposed benefits can be garnered. Secondly, the programming of creating software agents need involvement of programming specialists and making a KMS need human resources who would update the KMS database regularly, therefore, this approach may not necessarily be cost efficient. Thirdly, the user might not be satisfied with an entire self-help system due to interface design constraints therefore there is a high possibility that it may require adapting user-interface of the self-help system to the needs of users. It could definitely involve more financial and time resources to complete this. Fourthly, users might still ask certain questions about simple and routine queries especially the users with less computer experience.

¹⁰ Simple and routine queries are technical queries that can be solved by a user if relevant information is provided to the user; e.g. password queries, account queries, hardware and software related enquiries.

In e-Science, introducing software agents and KMS is a time and human resource intensive task and therefore it might not be feasible to implement the approach which is mentioned above. Secondly, in e-Science infrastructures, there should be a unanimous shared-agreement of the stakeholders to apply the approach mentioned above and it should be made sure that the funding or financial resources are available. However, a simple self-help system can be introduced initially based on tagging mechanism and faceted search in e-Science infrastructure support to locate relevant information (FAQs and how-to's). A central repository of important information for users can be created where all important information is present in the form of documents, web-pages and other resources. This can be a first step and feasible goal to achieve within the constraints.

2.6.3.2 Using case based reasoning

Another technique to improve end user-support is case based reasoning (CBR). CBR was applied at Daimler Chrysler to improve user support in order to support help desk and change management operations (Roth-Berghofer, 2004). In case based (CB) decision support, particularly HD support experience is reused. CBR technique was applied to improve Daimler Chrysler's internal computer aided design (CAD) and computer aided manufacturing (CAM) help desk which was being used by 1400 engineers to solve more than 600 problems each month (on the average) in a project called HOMER. 17000 trouble tickets addressing whole range of IT related problems of CAD and CAM was the main information source to start modelling with. HOMER helped in diagnosing problems by using a classical CBR approach that comprised four main steps as described by (Aamodt & Plaza, 1994). These four main steps include: Retrieve similar cases that match the target problem, reuse the solutions that may match the target problem, revise the solution applied, make it applicable to the new problem environment or domain and retain the new solution.

Going beyond classical trouble-ticket system such as CBR - computer support was necessary in case of Daimler Chrysler internal help desk system to support 1400 engineering employees who work in a time critical environment. In this case: HD personnel were limited and the environment was very much time critical. It is vital to note that the users of HOMER were highly qualified engineers. However, in e-Science CBR might not be suitable because the environment is not as time critical as in Daimler Chrysler. Secondly, CBR systems are resource intensive especially huge amount of financial resources and willingness of the support employees and other stakeholders is necessary to start modeling of the problem domain and then developing a CBR system. Thirdly, constant monitoring and improvement should be done to the CBR system as recommended by (Roth-Berghofer, 2004). In CBR, cases are the only source of knowledge alone.

2.6.3.3 Using corpus-based approach to improve user support

Marom and Zukerman presented a corpus-based approach for analysis and synthesis of user help desk responses (Marom & Zukerman, 2005). The purpose approach is to automatically resolve the simple and low technical user queries by possibly providing an automatic response based on extracting features from incoming user request. The previous HD responses to the similar problems are saved and sentences are extracted from the previous response to similar problems and then sentences are put together as clusters that finally form end user response. The core idea of this system is to free HD staffs from answering simple and less technical questions. The proposed system was able to generate 14% of the complete responses and the rest of 86% were incomplete responses, which needed monitoring of HD support staff.

Corpus-based approach (CBA) can be used to analyze and synthesize help desk responses but one major drawback is that it is simply a technology oriented solution and does not consider re-engineering business process in an application domain of user-support or help desk. Secondly, the ratio of complete responses generated by the system (only 14%) does not justify the huge amount of effort and time invested in building the system (Zukerman & Marom, 2006). Zukerman and Marom did not present the evaluation and degree of user satisfaction with the system generated responses and the degree of the usefulness of automatically generated incomplete responses for the support staff.

2.6.3.4 Improving user-support using web assistants

Aberg et al. conducted an empirical investigation via a field study to find out the role of human web assistants who provided support to the end-users of Web-Information-Systems (WIS) such as e-commerce websites (Aberg & Shahmehri, 2001). They found out that human assistants have a positive influence on user attitudes towards a website. Furthermore, they found out that human assistance plays a vital role in providing support to users for achieving usability of WIS. They proposed a model that comprises the characteristics of computer-based support and human assistance. The model proposes a flexible UI where users can choose how they want to interact with the support system. For instance, users can look at FAQs or can search, but on top of it, the user also has an opportunity to have textual chat and voice chat with human web assistant.

The study has shown technical feasibility, however, economic feasibility to use these methods is not evaluated (Aberg & Shahmehri, 2001). The study also had other limitations such as the recruitment of web assistants and users was not justified because Aberg et al. arbitrarily assigned volunteers as web-assistants without adequate domain knowledge of WIS. The users were also arbitrarily assigned. It is almost impossible to utilize the human web assistants for a whole working day in the e-Science infrastructure

support because support staffs that are supporting users are responsible for other tasks too such as programming, system administration and other activities. However, there is a potential of a study to be undertaken in e-Science infrastructure support to assign for instance students as user support web-assistants and measure the effectiveness and efficiency of a user-support in e-Science infrastructure project support as well as satisfaction of end-users.

2.7 Tools that can be used in modelling user-support processes

For mapping or modeling a process there are more than seventy languages or notations available to represent a process in different domains and to make it understandable to others. These modeling notations are also known as modeling methods (Cempel, 2010). According to Pacholski et al. (Pacholski, Cempel, & Pawlewski, 2009) there are 70 process modeling methods that fulfill the best 18 criteria. The best 18 criteria include the following 18 characteristics of formality; information expressiveness, expressiveness of actors, expressiveness-dynamics, expressiveness-process, expressiveness-real time, graphical, textual, abstraction and modularization, extendibility, executability, analyzability, evolutionability, multiple conceptual views, computer support, availability, maintainability and standardization. These process modelling methods include: Soft Systems Methodology (SSM), ARIS, Business Process Modelling Language (BPML), Unified Modelling Language (UML) and others.

There are numerous modeling tools but so far few studies has modelled the user-support process in e-Science infrastructures by using any of the modelling language mentioned above. It is important to model user support process in order to understand the dynamics of the process, its understandability, modularity and analyse different relations with the interacting entities. From all these modelling languages and notations, SSM can be chosen to model end user support in e-Science because of its wide usage and understandability by all of the stakeholders (P Checkland & Winter, 2006).

2.8 Soft systems methodology

Soft systems methodology (SSM) is based on systems theory. Systems theory attempts to study a holistic picture. SSM is a general problem solving tool. To improve a system or to develop an effective system to support users of an e-Science infrastructure, it is essential for all stakeholders to understand how support employees perform the tasks of helping users along with performing their other core operational tasks of e-Science infrastructure development. For this purpose, SSM is chosen in this research to portray the current user support practices that constitute user support system in Earth System Grid Federation (ESGF). Though various approaches have been used to model user support systems in industry (Hess, Reuter, Pipek, & Wulf, 2012; Jäntti, 2009) and education sector (Arora & Bandara, 2006), hardly any study has been done using SSM

approach. Moreover, few studies have been conducted that have investigated user support practices in e-Science infrastructures so far (Soehner, Steeves, & Ward, 2010).

2.9 Knowledge gap analysis

From the literature review conducted in this chapter, it is evident that there are numerous studies that took place in the past about exploring the techniques and technologies used in the customer as well as user support. However, there are very few studies where the scientific methods to study the user support process were applied, especially in the field of federated e-Science. Moreover, the user support process and the nature of user support service which is based on the underlying user support process in e-Research and e-Science infrastructure were hardly studied in the past in a scientific manner, using SSM. The literature review done in this chapter indicates a research gap in this direction and this research is aimed to fill this gap.

The scientific study of the process of user support is aimed to provide the current status of the user support process prevalent in an e-Science infrastructure; ESGF. From studying the current practice of user support i.e. “as-is” state, the “to-be” state can be derived and recommendations can be made to achieve the to-be state in ESGF. The successful implementation and adoption of the to-be state of a user support process and service in an e-Science infrastructure like ESGF shall create a role model for other e-Science infrastructures and e-Research facilities to enhance their user support practices. Once the user support process is streamlined, the research activities via an e-Science infrastructure can easily flourish. The (emotional, practical, social) aspects of this research can enrich the field of e-Science (Science 2.0) as well as wellbeing of researchers.

2.10 Chapter summary

User-support technologies and processes have evolved with the passage of time. The first help desk in the 80’s had only a desk, pen and a telephone used by human support agent. The traditional help desk then had gone through different levels of evolution with the change in the commercial organizational set-up and needs of customers to employ techniques like automatic call distributions systems, interactive voice response systems, help desk management system along with associated reporting tools, help desk expert systems, knowledge-management centric help desks, embedding case-based reasoning engine in help desk, help desks based on corpus-based analysis mechanisms, use of remote control technologies to support end-users and web based e-support techniques with and without human support agents.

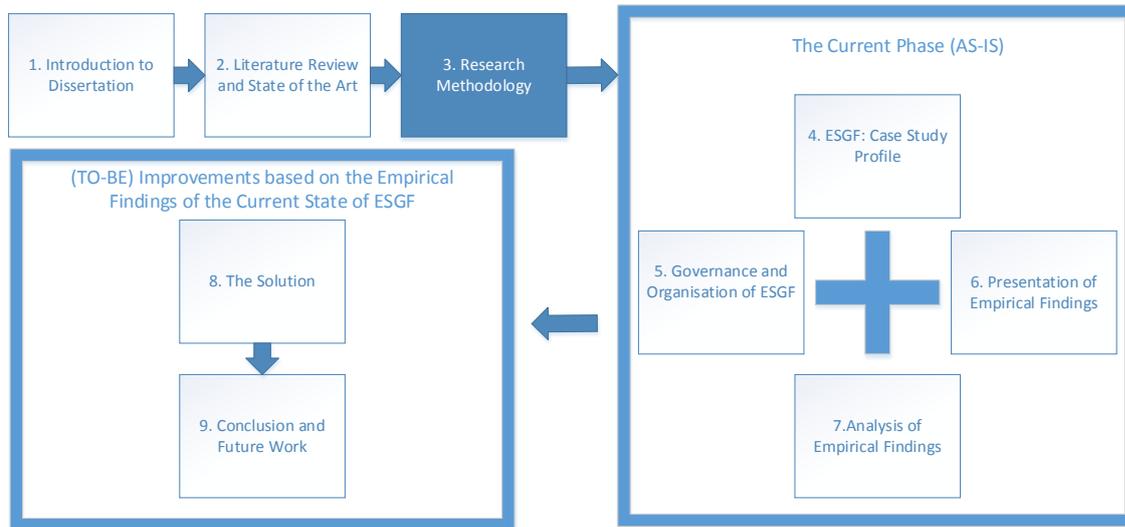
Different techniques, frameworks and models started developing to improve the user-support further with cutting user-support process cost and downsizing support staffs. User support has been always seen as a subsidiary or additional function to the core services of corporation until start of 2000s, when it was realized that customer support

should be made better with the application of business process frameworks to improve service quality and provide customer satisfaction. Since then different support models and structures have been tried to suit the corporation business model of servicing customer and end-user concerns.

Service desk is the concept that combined service management studies with the traditional customer service studies that used the term “help desk”. Until now there are different versions and variety of business process and business service frameworks such as ITIL that provide best practice guidelines for servicing end-users and customers especially in the commercial corporations and companies locally and globally. Some of these frameworks have been modified and adapted to academic setups such as universities or to governmental administrative bodies. However, with the emergence of grid-computing technologies and e-Science infrastructures, it is imperative to study application of these frameworks in servicing users of e-Science infrastructures. Few studies have been conducted to improve or even assess the service quality of users of e-Science infrastructures, keeping economic and human resource factors under control as well as feasible for the stakeholders.

In e-Science infrastructures, as more and more effort is being invested in improving the grid-based technologies, cloud computing, end-user support is being offered either on intuitive basis and technology oriented methods are being applied based on past experience without studying the requirements and the nature of e-Science infrastructural domains, data application and consideration of end-user requirements. In my dissertation, my study will fill this gap by analyzing the supporters or human support agents’ experiences empirically in a case-study of Earth system modeling domain of climate science. In this study, the current support practices will be investigated to find out the requirements of end-users from support staffs’ perspective and the best practices from the commercial domain can be borrowed to streamline end user support process in e-Sciences.

3 Research aims and objectives



3.1 Introduction

This chapter introduces the aims and objectives of this research along with the research question. In this chapter, the research methodology applied to carry out this research, to provide plausible answer and evidence to the research question is discussed. The research design and research methods guide this study with the help of empirical investigation to find out the current user-support framework in e-Science and suggest guidelines to improve it. The research design is based on a real-time scenario in climate science domain of Earth Science Modeling (ESM); a leading e-Science infrastructure named Earth System Grid Federation (ESGF).

Aims and objectives of this research are partly based on the studies reviewed in the previous chapter; literature review and the state of the art. This chapter starts with a discussion on the research objectives and the research question in Section 3.2, followed by study design in Section 3.3, the methodological domain in Section 3.4, the case study approach in Section 3.5 and choice of ESGF as a research platform in Section 3.6. The details about research methods used in this study are provided in Section 3.7 and finally the chapter summary is given in Section 3.8.

3.2 Research objectives and the research question

As stated earlier in the previous chapter, user support process has been studied in different fields, for instance; IT-related, education, automobile and others, yet it remains under-studied in the field of e-Science. Therefore, the goal of this empirical research is to explore the current user support process in e-Science and highlight its salient features.

3.2.1 Research question

In order to answer the main research question: “*What are the user support practices that can potentially be applied to improve the user support process in (climate) e-Science infrastructures?*” it is important to investigate and have a clear understanding of: The current status of user support in climate science e-science infrastructures i.e. *as-is*. This can be found out by exploring the user-support process in an e-Science infrastructure from different sources, for instance; from the perspective of support staff members and the way user support is being practiced in practical settings.

Likewise, it is vital to study a practical example of user-support process in climate e-Science in the form of a case study to find out, what lessons can be drawn from this particular case, to improve the current user support scenario i.e. *to-be*. The lessons learnt from this particular user-support process can then be generalized in the form of best practices and then can be applied to other domains in e-Science in future. In addition, a clear picture of interaction of the support staff with the current user support tools, techniques, methods and procedures within the support process is needed to be captured to improve the current state of user support service. Moreover, an insight into support staff’s organization, user support governance, management, handling and categorization of incoming user’s requests is also significant.

3.2.2 Sub-research areas

The research problem may be further divided into following sub-research areas that interlink with the main research question. Furthermore, these sub-research areas provide a roadmap for addressing the main research goal as well as open up discussions in the related fields:

- i. *What is the current state of user-support in climate e-Science?*

This includes finding out the strengths and limitations of the current user-support practices in ESGF. ESGF is taken as a use case in this study. This sub-research question intends to explore the current user-support practices in e-Science infrastructure especially in the domain of climate science. Improvement of a user-support process cannot be made without having a clear understanding of user-support practices in a working environment of an organizational setting in an e-Science infrastructure. It is vital to study how user-support is managed and carried out by the support personnel.

Discovering the strengths and weaknesses (and the reasons behind it) of the current user support process in practice, in a climate e-Science infrastructure case study, is the first step towards improvement. The important aspect in this research question is to find out why a particular feature is seen as strength in the current environment setting and how this strength can be further improved. Similarly, determining the cause(s) behind the weaknesses of the current user-support process in practice can definitely help

overcoming the weaknesses and thus improving the user-support framework to some extent. Survey-questionnaire, interviews, observations and document analysis are conducted in this doctoral project to find out the current state of user-support and the user-request management process in a climate e-Science infrastructure.

ii. What tools do end-users and support staffs use to communicate to each other?

There are varieties of user-support tools with diverse Graphical User interfaces (GUI) that have been in place for communication amongst support staffs and end-users, mainly to mitigate problems. But how can they be applied to a climate e-Science infrastructure? Are these tools appropriate? What sort of interfaces or support tools, such as request tracking software, does support staff prefer and why? It is important to find a suitable tool, if available in the user-support arena, for the user-support process in e-Science. Or a special user-support tool is needed to be designed, that would suite the end-users, support staffs and in turn makes the whole user-support process efficient. The interviews with support staffs as well as field-observations recorded and analyzed in this research, reveal the lessons learnt from the current media in use, the needs of the preferred media and features of user-support tools. However, this research sub-area is not the main part of this study.

iii. What are the characteristics of e-Science infrastructures that may have an effect on a user-support process (particularly in climate science)?

Infrastructure is a basic physical and organisational structure needed for the operation of a society (Oxford dictionary, 2009). The term generally refers to technical structures such as bridges, electrical grids, telecommunications, sewerage infrastructure and many other infrastructures around us. Fulmer defines infrastructure as “the physical components of inter-related systems that provide some services to society” (Fulmer, 2009). Infrastructures are further categorized into *soft* and *hard* infrastructures. The hard infrastructures include: Transportation infrastructure, energy infrastructure, communication infrastructures and others. The soft infrastructures include: Social infrastructure, governance infrastructure and others.

Most of the e-Science infrastructures have users who benefit from these infrastructures. The users are mainly researchers and they utilize computation and data download capabilities of e-Science infrastructures. This sub-research question intends to describe and explore the features of e-Science infrastructures that might affect the user support structure, organization, behaviour and culture. In this sub-research question, it is important to find out the characteristics of e-Science infrastructures that have an effect on the user-support process and user support service. One can hypothesize that there is a list of factors that might affect the user-support. One of the factors is the *distributed nature of users* present all over the world. 27,000 researchers from 2,700 sites in 19

different countries on six continents are sharing data and tools via the Earth System Grid Federation (ESGF) (Vu, 2013). Another factor is the *geographically dispersed components of an e-Science infrastructure*, with their components controlled by organisations in different parts of the world. The organization structure of e-Science infrastructures has *global* and *local elements*, which is another factor that might affect the user-support. The local elements include for instance; local policies, a set of responsibilities and a local corporate identity that cannot be ignored in an institute, e.g. German Climate Computing Centre (DKRZ). At the same time, DKRZ is participating in global projects such as ESGF and there is a requirement to follow global standards as well.

Budget of e-Science infrastructure projects that is dedicated for support activities is limited, which is yet another significant factor influencing the user support service. In e-Science infrastructure the money is normally allocated to the development of the components of the infrastructure. Little investment is made to improve the usability aspects of the infrastructure as well as improvement in delivering the user support service. This is because the service aspect of an e-Science infrastructure has not been highlighted and there are very few initiatives in this direction.

Yet another factor is that *the usability of user-interface (UI) of gateways*¹¹ i.e. portals may affect the user support process. The *number of human support agents* available to support users, their experience, background, training, qualification, friendliness and analytical skills to support end-users in an e-Science infrastructure is also an important factor. The *global time-zones covered and the hours of operation of the user support centres* also affect the user support process.

The *software and hardware technologies employed* by the e-Science infrastructures to serve end user request may also affect the user support process. For instance, in ESGF, the software technologies are bundled together in the form of a *software stack* that includes the node manager, peer-to-peer protocol, data preparation and publishing software, indexing and search services, security infrastructure, web portal and user-interface (UI), data access software and the underlying protocols for data-access, monitoring system known as dashboard and data visualization tools. This includes the tools used for e-Science infrastructure to support end-users. The *reliability, availability, stability and average down-time* of e-Science infrastructure have definitely an effect on the user support process. The *features* that an e-Science infrastructure offer to the users such as data-download, data-search, available meta-data, the capacity of data (in Terabytes, Petabytes), visualizations, workflows and others have an effect on the user support process.

¹¹ Gateways in e-Science infrastructures (particularly in ESGF) are the portals (websites) via which one can access and download the raw data i.e. Climate data.

E-infrastructures are studied most of the time from a technical perspective to further develop these infrastructures technically, but the social impacts of characteristics of e-Science infrastructures have not been studied so far (Jirotko et al., 2006). The answer to this sub-research area is extracted from the interviews conducted from user support staff, infrastructure developers, principal investigators, project managers, users-support staff communication via mailing-list as well as all the related, relevant and available literature and other resources according to the best of knowledge of the author. Once these features are made known, one can foresee the structure of support in e-Science infrastructures of similar nature.

- iv. *What is the common user-support procedures recommended in literature from other domains that might be adaptable to climate e-Science infrastructures?*

In this sub-research question, it is important to follow the general user-support service guidelines from other domains e.g. in IT infrastructure industry and other industries that are already described in some service frameworks such as ITIL that might be applicable to improve user support in the e-Science infrastructure projects. Apart from service frameworks there are other techniques and methods that have been employed by other IT-related commercial corporations to provide effective end-user support. Some of these improvement recommendations and methodologies can be applied to enhance e-Science infrastructures user support.

This research question intends to explore the list of existing user-support techniques in other industries which are considered by the support staffs to be appropriate and effective when combined with the current and new user-support procedures in e-Science that make user-support more efficient. New user-support guidelines are developed after the empirical investigations on the current user-support process in ESGF as well as from the discussions and input from the support staffs. The answer for this research question is extracted from the interviews and the survey in the e-Science industry which gives us the insight about where and why particular user-support tools are considered suitable and unsuitable.

- v. *What types of organizational, operational and financial challenges are related to user-support operations in climate e-Science?*

This sub-research question leads to identification of some of the challenges that make this user-support scenario different and difficult from other user-support scenarios in other domains. These challenges are peculiar in framing the user-support process of e-Science and are unique in nature due to the environmental setting of the e-Science domain. It is important to differentiate the different practices which are prevalent in other domains that could not be adapted to e-Science and to find out the reasons why they cannot be adapted. The distributed nature of e-Science causes organisational, structural, technical, scientific, cultural and social challenges. All these challenges are

needed to be addressed. To what extent can they be addressed is the question that is needed to be answered in this doctoral project.

vi. *How user-support requests should be classified in climate e-Science Infrastructures?*

First of all, whether the user-support requests should be classified in different categories is an open question that is needed to be discussed with the support staffs. If they are needed to be classified then what may be the classification scheme? These questions can only be answered after analyzing user and user-support staffs' communication. The classification of incoming inquiries is important in terms of archiving user-requests in order to find solutions for future reference and for routing the categories to a particular responsible support staff. The user-support requests can be prioritized and classified in different ways but which classification scheme is the best that can be applied to the other e-Science infrastructure projects for classification of incoming queries, is the outcome of this research question.

3.3 Study design

This doctoral research is an analytical synthesis of both descriptive and explorative scientific study techniques directed towards a *design science research* approach. Design science is based on systematic body of evidence and is aimed at understanding and improving human performance (van Aken & Romme, 2009). According to Hevner, the major purpose of design science research is achieving knowledge to understand the problem domain by designing an artefact (Hevner, March, Park, & Ram, 2004). The evidence is collected from multiple sources. Descriptive investigations focus on what exactly is happening in particular environmental settings such as climate e-Science infrastructure user-support; they are conducted with the help of surveys, observations and interviewing focus groups (Lazar, Feng, & Hochheiser, 2007, p. 21). The research plan of this study comprises of six-phases (see Figure 3.1):

Phase 1: A research problem was defined based on the prior knowledge of the author, his participative observations, weekly meetings, informal coffee table discussions, e-mail discussions, literature and other assumptions.

Phase 2: In the second phase the following steps were executed in parallel:

- a. A web-based survey questionnaire was designed, conducted, the data was collected and then analysed. The respondents of the survey questionnaire were support staff members of ESGF and ESGF-related projects from all over the world.
 - b. Semi-structured interviews were conducted with users, support staff members, managerial staffs including principal investigators, infrastructural developers
-

and others to validate the data collected via questionnaire and to capture a more insight into the user support process.

- c. Field study was conducted to observe the user-support activity especially interaction of support-staff with the support tools (that included user request tracking system). Observations were made from the point when the incoming user requests arrive to a support staff's help desk (via request tracking software, mailing-list or any other means) and the way they are processed. The results of this step are however not part of this thesis.
- d. A mailing-list comprising user-requests and user-support answers was observed to find common patterns, classification of incoming user-requests, solutions, time-delays and other factors. However, the complete outcomes of these observations are also not part of this dissertation.

Phase 3: All data gathered from the various sources of the phase 2 (including the sub-phases; a and b), were analysed and an overall-picture of the current user-support process in ESGF was portrayed in the form of a report describing major features, strengths, weaknesses and the problems found in the current user-support process, in Chapter 7. Based on the analysis, recommendations were then made that shall be implemented by the ESGF management to improve the current user-support process, in Chapter 8.

Phase 4: The suggested recommendations were presented to the ESGF stakeholder committee to put into practice by applying them to the user-support process in ESGF, where ever possible. This step was conducted by ESGF management.

Phase 5: Elements within the user-support process of ESGF are highlighted that can be generalized in form of a Federated e-Science infrastructure User Support Enhancement - FeUSE framework and hence this proposed user-support framework may be adopted to improve other e-Science infrastructures support service, particularly the e-Science infrastructures having similar characteristics as the ESGF.

Phase 6: A validation and evaluation of the new user-support process was made with the help of focused group survey conducted at the 4th ESGF Face-to-Face conference held at Livermore, California, where the stakeholders mainly policy-makers of ESGF participated. The results of the survey were partially included in this dissertation, in Chapter 8. Complete evaluation of the FeUSE was thus not made part of this study.

Applying recommendations of FeUSE to other e-Science infrastructures and other fields is out of scope of this study. Applying and implementing the recommendations (of the author) embodied in the FeUSE framework to the ESGF consortium is the prerogative of ESGF top management and other stakeholders and therefore also out scope of this study.

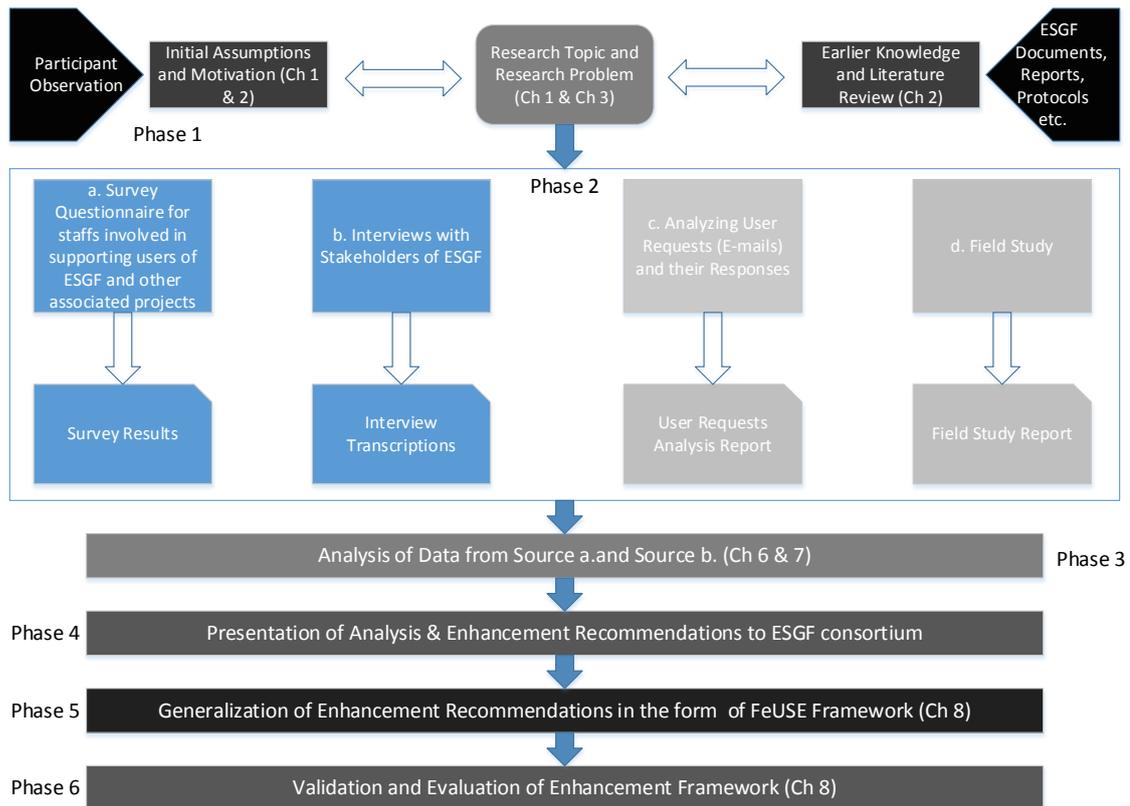


Figure 3.1 The research design of this study with different phases and each phase having its own aim and outcomes as planned by the author.

3.4 The methodological domain

The methodological domain provides insights into the research methods employed to explore user-support in climate e-Science infrastructure domain. In this research qualitative inquiry is mixed with quantitative approach.

3.4.1 Applicability of a qualitative approach

Research on the user-support set-up in e-Science infrastructures in the field of climate science seems to be more consonant with a qualitative approach for a number of reasons. This study is of an *exploratory nature* since the notion of user-support in e-Science infrastructures has been slightly touched in academic literature and to the best of the researcher's knowledge, no empirical evidence on the association of user-support and e-Science infrastructures has been found. Exploratory research is based mainly on the collection of qualitative data and aims at clarifying poorly understood concepts (Lazar, Feng, & Hochheiser, 2010, p. 150; Patton, 1988). Qualitative data is the most efficient and effective way of getting insights into the phenomenon under observation (Markopoulos, 2013).

The *process focus* of the research topic justified the choice of qualitative methodology. Both user-support and usage of e-Science infrastructure appear to have a processual character: User-support and customer-support is a practice of the firm (Srivastava, Shrevani, & Fahey, 1999) and it is seen as a "a social process associated with a unit in which members or users share a common set of elements" (Rousseau, 1990, p. 160). In this regard, the researcher sets out to understand and document the day-to-day reality of user-support in its natural setting i.e. the e-Science infrastructure federation and its interaction with its users. An e-Science infrastructure federation is a global organisation made up of smaller organisations at different geographical locations that form an e-Science infrastructure.

There seems to be *a lack of proven quantitative instrumentation*¹² for the user-support in e-Science Infrastructures (in climate science), since it appears to lack valid and reliable quantitative measures in a particular environmental setting. Rousseau suggests that the uniqueness of a firm's culture restrain the development of a priori measures and quantitative instrumentation (Rousseau, 1990).

As stated by Miles and Huberman, qualitative enquiry contributes to the development of a *holistic* (systemic, encompassing, integrated) *view of the topic under investigation*, i.e. its logic, activities, and explicit and implicit rules (Miles & Huberman, 1994). Similarly, Patton suggests that in qualitative inquiry the researcher attempts to reach a spherical perspective of the phenomenon of interest in its context (Patton, 1988). In qualitative research the study is not reduced to single variables but studied in their entirety (Markopoulos, 2013).

According to Sutton qualitative inquiry is well suited for *theory building purposes* (Sutton, 1997). This is in accordance with the exploratory nature of the study, the emerging interest and the need to promote research in the area of user-support in e-Science infrastructures.

Qualitative research (QR) helps to study rapidly changing dynamic phenomena like user-support in climate e-Science infrastructures (Sofaer, 1999). "QR tries to uncover how individuals arrange themselves, their physical and technical environment and how people make sense of surroundings through symbols, rituals, social structures, social roles, etc." (Flick, 2004).

QR methods are "...a specific and systematic way of discovering and understanding how social realities arise, operate and impact on individuals, organizations and fields" (Neuman, 2006). Design of interactive products e.g. UIs and user support services is based on understanding needs, values, aspirations and capabilities of people who use them. Thus, the foundation of designing a service or an interaction is understanding

¹² Instrumentation: A tool used to collect data.

people by applying qualitative research methods. In this endeavor: Research is a necessary component of design. Design needs not to be based on certainty but on a nuanced understanding of situations.

Rather than counts and measures, qualitative research aims to understand the *what, how, when, meanings, concepts, definitions, symbols, etc.*, that capture the *essence* and *ambience* of a phenomenon (Markopoulos, 2013). In this study this phenomenon is end-user support in climate e-Science infrastructures.

3.4.2 Applicability of quantitative approach

This research study, though mainly qualitative, does contain a quantitative approach with the basic use of statistics and mathematics to explain the current state of the user-support process in climate e-Science infrastructure via survey questionnaire. Mostly closed-ended questions were used in the questionnaire to evaluate the user support setting. The main aim of the quantitative approach is to look at the breadth of the phenomenon under investigation i.e. user-support in climate e-Science in this study (Rocco, Bliss, Pérez-Prado, & Gallagher, 2003; Tashakkori & Teddlie, 2000).

3.5 Reasons for the choice of the case study method

The case study method was selected to explore the user support which is an interface between end-users and support staff in ESGF project. A case study is “a research strategy which focuses on understanding the dynamics present with single settings” (Yin, 2013). It can also be defined as “an empirical inquiry that investigates a contemporary phenomenon within its real-life context” (Eisenhardt, 1989). A case study is an in-depth study of a specific instance with-in a specific real life context (Lazar et al., 2010).

This study is based on a single, in-depth, synchronic case in the e-Science infrastructure user support services sector. Case studies offer an in-depth view of the components and levels of organisational practices such as user support, and provide additional insights regarding the context underlying these practices (Pettigrew, 1997; van de ven, 1992). The *unit of analysis* that determined the focus of the case study method (Miles & Huberman, 1994), was the support activity for users (i.e. researchers and policy makers) in the multi-organisational environment that constitute an e-Science infrastructure in climate science.

The reasons for the choice of this case study approach as a research method are the following:

- It suits well for studying service processes that are linked to a complex organizational context, as cases offer in-depth view of processes in organisations (Buchanan, 2012), the components and levels of the investigated process and the
-

context underlying this process (Greetz, 1973; Pettigrew, 1997; Plakoyiannaki, 2002).

- The use of the case study is appropriate for this dissertation, since e-Science infrastructure support involves a large number of actors in the form of distributed user and support teams and practices where the boundaries between these constituents are not easily distinguishable. Similarly, the complexity of links between actors (users and support staffs) and e-Science infrastructures support practices require the collection of rich data which is facilitated by the case study method (Markopoulos, 2013).
- Case studies focus on “*how*” and “*what*” research questions (Markopoulos, 2013; Yin, 2013) such as: “What are the industrial user support practices that can potentially be applied to climate e-Science infrastructures user support process to improve the current state of support?”, “How user-support requests should be classified in (climate) e- Science Infrastructures?” and “What is the current state of user-support in climate e-Science?” The case study method contributes significantly to the scrutiny and development of contemporary research areas, namely user support in e-Science infrastructures (Jäntti, 2012b).
- Case studies facilitate the exploration of relatively under-investigated topics such as quantitative analysis of the role of end-user support in e-Science infrastructure (Jäntti, 2012b; Plakoyiannaki, 2002). Case studies are valuable in situations where existing knowledge is limited (Plakoyiannaki, 2002) and contribute significantly to the study of contemporary research areas, such as user support in climate e-Science infrastructures (Jirotko, Lee, & Olson, 2013; Jirotko et al., 2006).

The case study findings discussed in this study have been emerged from multiple sources of evidence in the data collection process. These are *archival data*, *interviews*, *survey-questionnaire* and *observations including participant observations*. First, the examination of archival data such as end-user and support staff communication via mailing list, project reports, deliverables, project web portals and documents, offered insights and antecedent conditions to the history of the ESGF, directly associated with the practice of user-support (Pettigrew, 1997).

Second, *a total of 12 detailed, in-depth interviews were conducted with a variety of employees and users from different levels of organisational hierarchy and research units*. These interviews are relevant because they were a rich source of information as it allowed capturing different perspectives and perceptions of the current user support process in ESGF. The composition of the sample of the interviewees allows us to gain an insight into the matter with different perspectives. In addition, they had an overview of different situations in the current user support process as well as the over-all management of ESGF. Likewise, the interviewees who were users of ESGF had

experience in conducting their climate research activities using ESGF and other e-Science the infrastructures.

Third, keeping with Kunda and Watson, detailed participant observations were undertaken including attendance at meetings, internal presentations and user workshops (Kunda, 1992; Watson, 1995). Fourthly, field study was conducted at DKRZ¹³ that gave an insight into support-staffs treatment and interaction with incoming user-requests using the interfaces of request tracking and processing software. However, the results of this field study are not included in this dissertation.

3.6 Earth System Grid Federation as a platform for case study

An important practical use-case in e-Science in the field of climate science cyber-infrastructure is ESGF project. ESGF facilitates to study climate change and impact of climate change on human society and Earth's eco system (Cinquini et al., 2012). Since physical phenomena that govern Earth's climate are so complex and diverse, it is the most important scientific challenges of our time to undergo sophisticated model simulations that generate a huge amount of data, collect observational data from various sources and share that data at a global scale. This is made possible by ESGF to discover, analyse and access the climate data sets which are stored at multiple geographic locations across the globe (Cinquini et al., 2012; Earth System Grid Federation, 2010; ENES, 2013; Vu, 2013). The infrastructural architecture developed by ESGF allows the world to better organise and integrate all climate knowledge via a cooperative federation (see Figure 3.2).

“Hence a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it” Herbert Simon

¹³ German Climate Computing Centre, where the author is an employee

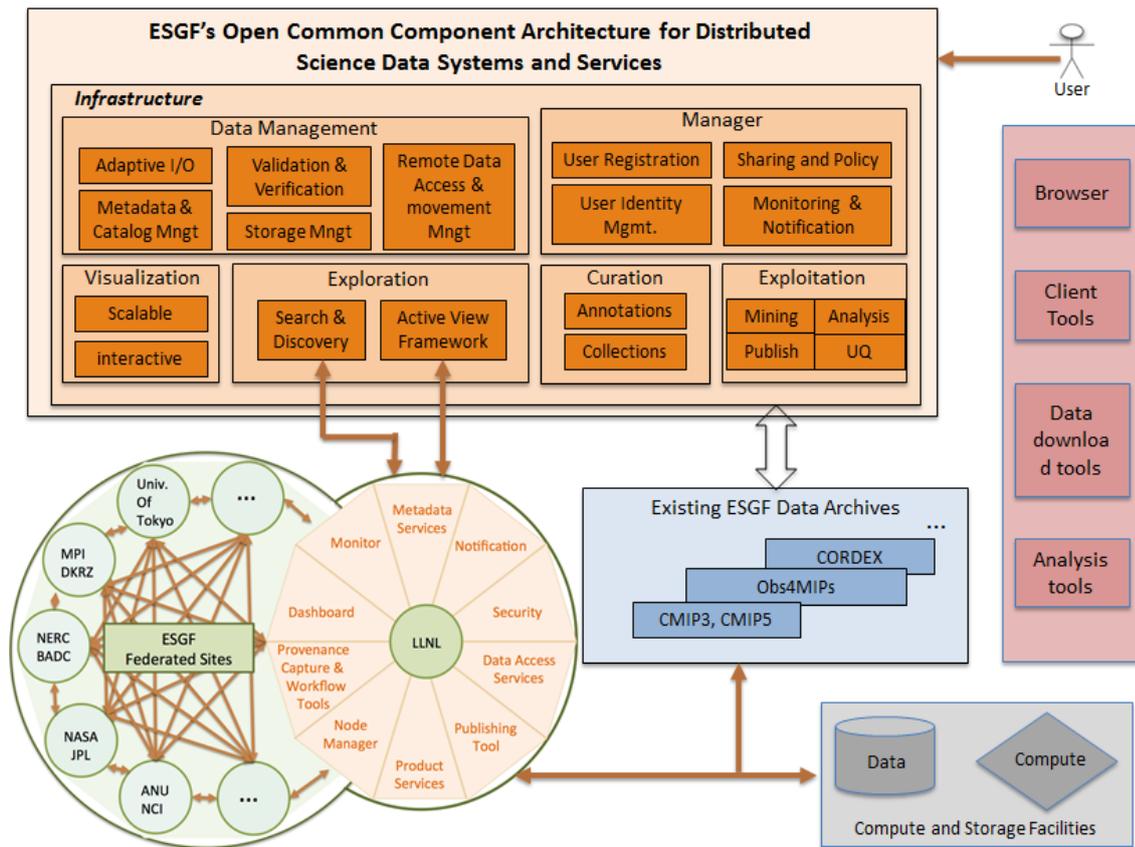


Figure 3.2 The architecture of Earth System Grid Federation (ESGF) taken from (Dean N. Williams et al., 2013).

The ESGF Peer-to-Peer (P2P) enterprise system is “key data-dissemination infrastructure and resource for climate simulation, observation and reanalysis data, expressly designed by the team of computer scientists and climate scientists to handle large-scale data management for worldwide distribution” (Hiller & Budich, 2013, p. 77). ESGF is an open source effort providing a robust, distributed data and computation platform, enabling world wide access to Peta/Exa-scale scientific data (Earth System Grid Federation, 2010). ESGF is an interagency and international effort led by:

- Department of Energy (DOE),
- co-funded by National Aeronautics and Space Administration (NASA), National Oceanic and Atmospheric Administration (NOAA), National Science Foundation (NSF)
- and international laboratories such as the Max Planck Institute for Meteorology (MPI-M), German Climate Computing Centre (DKRZ), the Australian National University (ANU), National Computational Infrastructure (NCI), and the British Atmospheric Data Center (BADC) (Cinquini et al., 2012; Earth System Grid Federation, 2010; Vu, 2013).

ESGF hosts different climate modelling projects. Coupled Model Inter-comparison Project, phase 5 (CMIP5) is the most notable one, which is a global effort coordinated by World Climate Research Program (WCRP) (Cinquini et al., 2012; Taylor, Stouffer, & Meehl, 2012). Twenty climate modeling groups from all over the world met at a meeting on September 2008, forming the Working Group on Coupled Modelling (WGCM) and agreed to promote a new set of coordinated climate model experiments with input from the International Geosphere Biosphere Programme (IGBP) and Analysis Integration Modelling of Earth System (AIMES) project. These experiments comprise the fifth phase of the CMIP5 (Taylor et al., 2012). “CMIP5 will notably provide a multi-model context for: 1) Assessing the mechanisms responsible for model differences in poorly understood feedbacks associated with the carbon cycle and with clouds, 2) examining climate “predictability” and exploring the ability of models to predict climate on decadal time scales, and, more generally and 3) determining why similarly forced models produce a range of responses” (PCMDI, 2013) (Taylor et al., 2012).

The Inter-governmental Panel for Climate Change (IPCC¹⁴)’s Fifth Assessment Report (AR5) is based on CMIP5 (see Figure 3.3). The Fifth Assessment Report (AR5) will provide a clear view of the current state of scientific knowledge relevant to climate change and global warming. CMIP5 promotes (Guilyardi et al., 2011) a standard set of model simulations in order to:

- “Evaluate how realistic the models are in simulating the recent past,
- provide projections of future climate change on two time scales, near term (out to about 2035) and long term (out to 2100 and beyond) and
- understand some of the factors responsible for differences in model projections, including quantifying some key feedbacks such as those involving clouds and the carbon cycle” (PCMDI, 2013).

¹⁴ “The Intergovernmental Panel on Climate Change (IPCC) is the leading international body for the assessment of climate change. It was established by the United Nations Environment Programme (UNEP) and the World Meteorological Organization (WMO) in 1988 to provide the world with a clear scientific view on the current state of knowledge in climate change and its potential environmental and socio-economic impacts. In the same year, the UN General Assembly endorsed the action by WMO and UNEP in jointly establishing the IPCC.”

The IPCC is a scientific body under the auspices of the United Nations (UN). It reviews and assesses the most recent scientific, technical and socio-economic information produced worldwide relevant to the understanding of climate change.” (World Meteorological Organization & UNEP, 2015)

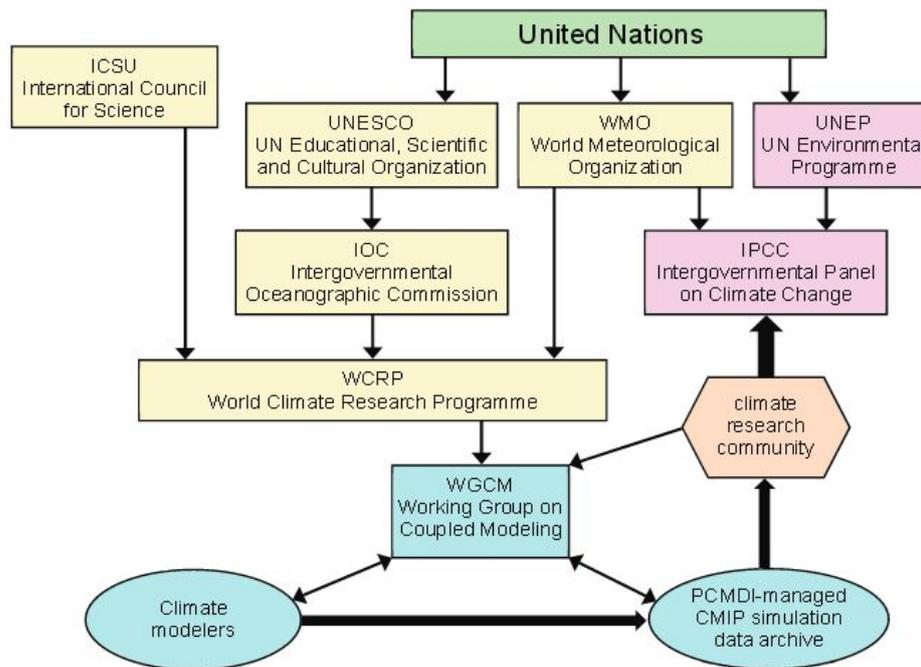


Figure 3.3 The relationship of CMIP5 to organizations established to coordinate climate research activities internationally and to the IPCC, the modelling centers, and the climate research community, taken from (Taylor et al., 2012).

According to (Cinquini et al., 2012; Earth System Grid Federation, 2010; Taylor et al., 2012), the ESGF mission is to:

- Create software infrastructure and tools that facilitate scientific advancements for ESM (Earth System Modelling)
- Enhance and improve current climate research infrastructure capabilities through involvement of the software development community and through adherence to sound software principles
- Integrate and interoperate with other software designed to meet the objectives of ESGF: e.g., software developed by NASA, NOAA, ESIP, and the European IS-ENES
- Foster collaboration across agency and political boundaries
- Support current CMIP5 activities, and prepare for future assessments
- Develop data and metadata facilities for inclusion of observations and reanalysis products for CMIP5 use

The reasons for choosing ESGF:

ESGF is a well-developed e-Science infrastructure in the field of climate science and Earth Science Modelling (ESM) and also it is the first inter-agency and international effort in the arena of e-Science infrastructures in climate science (Hey & Trefethen,

2005). ESGF is popular amongst climate scientists, students and policy making communities as it is servicing more than five thousand researchers from different continents who are accessing huge amount of climate data for climate-model inter-comparison purposes from widely distributed data-archives all over the world (ESGF, 2013; Hey & Trefethen, 2005).

Besides, it was found out by the author with a close interaction with ESGF support staffs that ESGF is currently in a process to organize its geographically distributed support-units. Thus, the need to refurbish and streamline user support was timely observed. The flexible and scalable infrastructural architecture offered by ESGF allows any institution to share or publish data sets, establish data and compute nodes. Even a node present in the federation can assume different responsibilities thus allowing a new organisation to become part of this federation (Cinquini et al., 2012).

Moreover, ESGF data is connected to other projects like; European Network for Earth System modelling (IS-ENES) and Climate Collaborative Community Grid (C3Grid) thus ESGF data-sets are made available by other climate e-Science infrastructures. Another reason to take ESGF as a case study is that the author is a user support manager of C3Grid and therefore has access to the infrastructure resources (technical and human resources) and has developed rapport with ESGF support staffs. This led the author to perform participatory observation in ESGF which is a mini-ethnography (Markopoulos, 2013).

Yet another reason to study ESGF is that it is one of many systems that offer web-based access to scientific data globally and it “stands out for its federation capabilities, size of data holdings, broad developer base, extended adoption, and most notable data holdings i.e., CMIP5, obs4MIPs, ANA4MIPs” (Taylor et al., 2012). Also ESGF shares similarities with other grid infrastructures of other domains, other than Climate Sciences. The similarities include usage of grid infrastructure, data grid, compute grid, web-access, and distributed-architecture.

Furthermore, ESGF is a unique system as the development of ESGF P2P system was done from scratch, without looking at other existing systems. The author could not find any other system, at least in the field of climate science, which is distributed and federated like the ESGF is, with searching capabilities via web that span multiple archives, which are under separate administrative control. All other systems are either centralized, or they may be distributed, but they are administered by one single organization.

3.7 Research methods

The detailed research methods in this research study are survey-questionnaire, interviews, participant observation and archival analysis. They are described below along with subjects sampling, data analysis techniques and validation of results.

3.7.1 Subjects sampling

The subjects of this study constitute stakeholders, including subject matter experts i.e. computer scientists and climate scientists. They are also infrastructure developers, users of the e-Science infrastructure system (ESGF) and social scientists, who basically are observers of the social settings and behaviours of employees as well as end-users. The role of subjects in the study context matters. The subjects were carefully chosen for survey-questionnaire, field-study and interviews, keeping in view the directions provided by Lazar and Morgan (Lazar et al., 2010; Morgan, 2008). A total of 26 subjects (24 males and 2 females) took the survey-questionnaire online, using Qualtrics suite (Qualtrics, 2013). Out of 26 responses, 25 responses were usable and one response was not included in the analysis, as not a minimum number of questions were answered, thus the required data was simply missing. 22 subjects completed 95% of the total questions of the survey. The fielding period was from 20th March 2013 till 30th April 2013. The average duration time of answering the survey was one hour. The link of the survey-questionnaire was sent to the subjects via e-mail. The e-mail was sent to all of the support staffs both FLS and SLS that are doing support of ESGF or ESGF-related climate science e-Science infrastructure projects.

The total of twelve interviewees was selected for interview. Four of the interviewees were developers, technical experts and system administrators of the ESGF e-Science infrastructure. Two of the interviewees were the policy makers of ESGF. Two of the interviewees were the principal investigators of ESGF project. Two of the interviewees were the social scientists who were observers of ESGF-related and non-Climate e-Science infrastructures. The interviewees were selected as per observation of the authors and recommendations were made by the first interviewees using a snow-ball method (Morgan, 2008).

The field study was conducted with the end-user support staffs that were responsible for user support at DKRZ. Bachelor students from the Human Computer Interaction (HCI) department came over to DKRZ, observed 4 employees (who were doing support activities) of DKRZ for one and a half hours, noted their observations and asked some questions. The students were given credits for conducting field-study in their HCI module. The employees at DKRZ were chosen following the criteria that the employees should be doing user-support activities.

3.7.2 Survey-questionnaire

The total of 45 questions was included in the web-based online survey-questionnaire. The suggestions from Couper about web-based questionnaire design were applied (Couper, 2000). From 45 questions, 41 questions were closed-ended, with the first question about the consent to use the system and the rest 3 were open-ended questions. The target population of the survey was user support employees and recruitment

guidelines were followed as provided by Lazar et al. (Lazar et al., 2010). The potential survey respondents “population frame” were chosen carefully. The population frame was defined keeping in view the goals of this research study and based on practical experience and observation of the author (Lazar et al., 2010, p. 103). The aim of the survey-questionnaire was to get to know the current user-support practices in ESGF and its associated projects. The respondents of the questionnaire were present at different locations all over the world. The survey-questionnaire was put online using Qualtrics tool (Qualtrics, 2013). Using survey-questionnaire is an easy and effective way to collect data from different locations. It helps to get a ‘snap-shot’ of the target population. Survey-questionnaire is a relatively unobtrusive method. Survey responses focus on a specific population which is being studied (Lazar et al., 2010).

Before putting the survey online, it was tested with colleagues (potential respondents) of the author who are supporting end-users in e-Science infrastructure as well as survey-questionnaire analysts (two experts both employed by Google for developing and analyzing surveys). The survey tool was reviewed to find out if it can serve the intended survey-questionnaire design. The potential respondents evaluated cognitive and motivational qualities of the survey tool. A pilot study was conducted with the survey tool and implementation procedures. Finally the online survey was taken by two potential respondents (colleagues of authors) with actual environment settings. These stages of pre-testing are suggested by Dillmann (Dillmann, 2000). A special method for measuring the internal reliability of the information captured from this survey was implemented by having the same question asked in different ways. Therefore it was possible to compare the validity and variation of responses to similar questions which were presented in different ways (Babbie, 1990; Ozok, 2007).

An e-mail, introducing the research topic and the importance of research was sent from the author using the DKRZ e-mail account. The author, being the support coordinator of ESGF associated project like C3Grid-INAD, already was in touch with the support staffs of ESGF and other associated projects. Thus, he already built a rapport with most of the potential respondents to the survey-questionnaire. Therefore, the support staffs of ESGF and other associated projects got motivation to take part in the questionnaire. A link to the potential respondents (40 people) was sent via an e-mail that included the aims of questionnaire and the expected time-frame to take the questionnaire. Since the questionnaire was originated from DKRZ, i.e. one of the partner laboratories in ESGF, this established authority, hence motivating the potential respondents to fill out the questionnaire. These instructions to contact potential respondents were also taken from Dillmann (Dillmann, 2000).

The responses were recorded automatically in the Qualtrics data-base with the help of Qualtrics online portal. The recorded data was then “cleaned” i.e. making sure that each response is valid as according to the guidelines of Lazar (Lazar et al., 2010). The

quantitative data was coded and was then analyzed. The qualitative data helped to form relevant graphics to represent the data responses, which were then prepared for interpretation. Some of the interview questions were based on the output of survey-questionnaire.

3.7.3 Interviews

A total of 12 detailed, in-depth semi-structured interviews were conducted with a variety of employees as well as users from different levels of organizational hierarchy and business units of ESGF and its associated projects. Including users and other roles is a vital aspect of this research because in other studies conducted in IT related organisations, the interviews were mainly depended on managers, as interview respondents (Plakoyiannaki, 2002; Siehl, 1985). Moreover, this study tackles a weakness of prior implementation of user support tools and techniques, which were implemented without research in the user support process in e-Science infrastructures. Furthermore, it offers a holistic perspective of the practice of user support which spans not only throughout the organizations but also includes multiple organisations dealing with e-Research practices integrated into a federated consortium (Srivastava et al., 1999).

Each interview was individually conducted and lasted between one and a half hours and sometimes up to two hours. All case-specific interviews were digitally-recorded with the permission of the respondents before they were transcribed. The transcriptions were conducted by hearing the recorded voice and typing words into an editing programme (MS Word) then saved as files and were made available digitally. Two of the transcriptions were done on paper by pen. Most interviews took place in the interviewees' office, although they sometimes took place in a quiet corner of a café or in a free conference room. Few interviews were done via telephone. The purpose of interviews in exploratory case studies is to develop ideas relevant to the concepts under investigation (Oppenheim, 2000). The interviews were a combination of semi-structured and unstructured styles involving open-ended questions and based on themes relevant to the objectives of the research. As far as the *research objective* is concerned, (i.e. to identify employees' and managers' perceptions about the current support system in e-Science infrastructural domain and application of user support practices from the non-e-Research domain to e-Science infrastructural domain), the interviews sought to explore the involvement of the respondents in the current user support practice, their understanding of the notion of user-support and their view of the firm's (their local organization) regarding user-support process and user-support activities.

In the interview process, the “user support process construct” was additionally explored through the narration of stories by organizational members (Boje, 2001). These stories and anecdotes were most of the times linked to the user support operations of the firm

and the interaction with the users. This facilitated the exploration of tacit aspects of the *firm's user support process* such as assumptions, operations and value of support (Kunda, 1992). Following the recommendations by Campbell and Tawadey, the interview also included some additional questions relevant to manifestations of user support in e-Science such as the mission statement, leadership and organizational symbols (Campbell & Tawadey, 1990). Follow up discussions with the informants illuminated how the practice of user support evolved over time (van de ven, 1992) and offered complimentary data that facilitated the better understanding and validation of interviews. The interview notes and question guide has been pre-tested prior to data collection with students and support experts with IT support experience.

The identification of the respondents was based on the author's personal project experience, consultation with other experts (human support agents) and often snowballing method was used to identify another interviewee. Keeping the directions of Lazar (Lazar et al., 2010), interviews with representatives of all stakeholder groups were conducted to provide the richer picture of the situation. With the snowballing method, two primary respondents were identified (a high level manager and an employee, who was providing support extensively) and interviewed; they were then asked to identify other people (secondary respondents) involved in the practice of user-support in e-Science (Morgan, 2008). Each of these secondary respondents is in turn interviewed and asked to identify other members involved in user-support initiatives. The snowballing and stakeholder representation method facilitated the collection of data from multiple participants involved in the practice of user support (Lazar et al., 2010; Morgan, 2008).

Morgan (Morgan, 2008, pp. 816–817) asserts that more functional areas and organisational levels are represented with the snowball technique than with other traditional methods of approaching respondents. The adding of new respondents stopped when repeating patterns of data started to emerge. Table 6.1, page 130 in Chapter 6, provides a list of the respondents of the interviews. Names are not provided in order to protect the interviewees' identity. The recorded conversations and transcriptions are however available with the author. Relevant excerpts of the interviews are italicized and are used in this dissertation with the relevant concepts discussed in forthcoming chapters. Sometimes, important utterances of interviewees are highlighted too in the interview excerpts. The quotes of interviewees are selective and most of the times the repetitive quotes of interviewees are not cited in the thesis.

3.7.4 Field study

A one day field study was conducted on 29th April 2013, with the students from Human Computer Interaction (in German "Mensch Computer Interaktion") department from the University of Hamburg, who came to observe the user support tools and practices of

user support staffs employed at DKRZ. Twelve students from the HCI department divided themselves into three groups of four students each group. They observed three different support employees for an average time of one and half hours each employee. The students observed the way the supporters interacted with the user-support tools (in this case RT system) and asked specific questions to the human support agents about the detailed problems that the staffs face while supporting the end-users and answering the end-user queries. Furthermore, they inquired about the features or characteristics of support tools the human support agents wish to have. In the end, the students proposed their own solutions in the form of paper prototypes for the betterment of interfaces of support tools used to handle and answer user requests. The solutions were presented in the form of paper prototypes. The further discussion on this topic is not a formal part of this dissertation.

3.7.5 Participant observation

The author himself undertook the user support staff (both first level and second level support) duty of a climate e-Science infrastructure project which was directly connected to the ESGF-project (especially ESGF datasets). This gave the author an opportunity to be part of the “tribe” and embed in the local culture of the climate e-Science infrastructures (Markopoulos, 2013). Detailed observations were undertaken, as suggested by Kunda and Watson (Kunda, 1992; Watson, 1995), during the plenary meetings, internal presentations, user-workshops, interaction with other user-support staffs, project leads and policy makers. The participant observation gave an opportunity to the author:

- to get to know the organizational structure and culture of the climate e-Science infrastructures that may not be achieved by observing conscious participants via formal observations, experiments or interviews (Kunda, 1992) and
- to get direct experience with user-support in climate e-Science infrastructures, allowing the researcher to grasp the phenomenon of research, holistically (Patton, 1988).

The value of observational data is of substantial importance since the researcher can learn aspects about user-support and climate e-Science infrastructures that respondents may feel uncomfortable to share in interviews and discussions (Patton, 1988). The participant observation period spans even before the beginning of the study, the time the author joined DKRZ till data collection for the study, in this way, the researcher familiarised himself with the organisational setting, structure and jargon. The impressions and the insights acquired were put in the form of field notes on the same day following the 24-hours principle (Eisenhardt, 1989). The participant observations offered inputs to the structure interviews, questionnaires and identification of potential respondents. Secondly, these observations provided a basis for the collection of archival

data, ESGF documentation and even relevant literature. Thirdly, events, and activities were the focus of the observations and field notes that the researcher prepared 'onsite' with the purpose to understand sufficiently the practice of user support process and its interaction with the stakeholders, especially support staffs and the end-users.

3.7.6 Analysis of mailing-list in ESGF support

The end-users support mailing list was the major communication channel via which end-users submitted their requests to the human support agents in ESGF. The ESGF mailing list was replaced later by the ESGF support platform based on "Askbot" which is a "Stackoverflow-like" clone system, at least partly in November, 2013. The communication between end-users and human support agents (up to 5000 e-mails from the period starting from December, 2013 until September, 2010) was observed by the author and a protocol was developed. Apart from this mailing-list, other documents (both technical and scientific) available online, publications and ESGF reports were analyzed. The examination of archival data offered insights and antecedent conditions to the history of the ESGF climate e-Science infrastructure and its associated support practices focusing end-users and human support agents (Pettigrew, 1997). The outcomes of the archival analysis are not a formal part of this dissertation.

3.7.7 Data collection and content analysis

The data from multiple sources that contained around 20 hours of digitally recorded interviews, their transcriptions, field study notes, (participant) observations in a diary, protocol summary of communication between users and user-support staffs in the form of e-mails and survey-questionnaire results were all collected. Data collection and analysis are iterative and purposeful (Markopoulos, 2013). The analysis of the whole data was done based on content analysis techniques with the help of clustering information into chunks and card sorting techniques. On top of it, NVivo, qualitative data analysis (QDA) computer software package was also initially utilized to perform deep-level data analysis (QSR, 2013). NVivo is reincarnation of NUD* IST (Non-numerical Unstructured Data Indexing, Searching and Theorizing) software index (QSR, 2013). Later on the content analysis was manually by the author. The use of content analysis is of value to this study for a variety of reasons:

- Its appropriateness for exploratory research and theory building purposes (Lazar et al., 2010; Markopoulos, 2013),
 - detecting common patterns that emerge from verbal data (Holsti, 1969; Plakoyiannaki, 2002),
 - elaboration of verbal data (interviews and communication material) (Carney, 1972; Krippendorff, 1980; Plakoyiannaki, 2002) and
 - process of making inferences and analysis of intrinsic meaning in the data (Holsti, 1969).
-

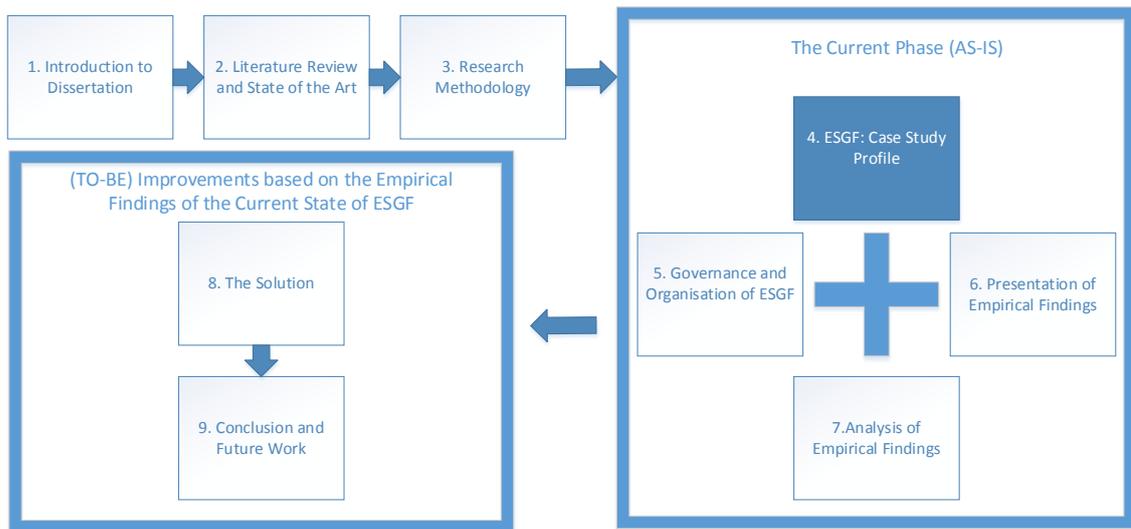
3.7.8 Validation and verification of results

In the context of this research study, the researcher (author) applied the methods of *data, theory, procedural* and *methodological* triangulation to assure validity of research findings as suggested in (Denzin & Lincoln, 1994; Eisenhardt, 1989; Plakoyiannaki, 2002; Yin, 2013); hence increasing the quality of the research findings. Data triangulation refers to the use of different data sources in a study e.g. interviewing people in different status positions. Theory of triangulation refers to approaching data with multiple perspectives and theories in mind. The use of theory triangulation is exercised to extend the possibilities for producing knowledge. Methodological triangulation refers to between-method triangulation and within-method triangulation. The former method illustrates the use of multiple methods to study a single problem such as interviews, observation, archival data and documentation. The latter method pertains to the use of several sub-scales for measuring items in a questionnaire. Triangulation of procedures refers to the use of qualitative and quantitative methods in a single research project. The triangulation helps to achieve multiple perceptions of a single reality (Denzin & Lincoln, 1994; Plakoyiannaki, 2002; Yin, 2013). The “methodological trustworthiness and verification” is the extent to which the research can be audited (Morse et al., 2002). The use of interviews-quotations in the findings and discussion, questionnaires, case-study database that contains all the records allow the research to be audited, thus fulfilling the methodological trustworthiness and verification (Yin, 2013).

3.8 Chapter summary

Chapter 3 illustrates the main research question and the methodology adopted in the research process of this study to address the research question. The research combines mixed strategy i.e. both qualitative and quantitative methodological approach to reach the research goals of this study. However, the explorative nature of the research puts more emphasis on the usage of qualitative methods. This thesis is based on an exploratory single-case study of user-support process of ESGF, a leading e-Science infrastructure in the climate science sector. The duration of the case study was 38 months. The researcher has used multiple sources of evidence to facilitate the triangulation of data procedure. A survey-questionnaire with 45 questions was carefully designed, conducted and data was recorded in a database. Other qualitative research methods included 12 personal interviews, participant observation and archival data analysis. The reliability and validity of the findings has been judged taking into consideration the criteria that the pragmatic paradigm sets.

4 ESGF - case study profile



From this chapter the following peer-reviewed paper was accepted:

- *Hashim Iqbal Chunpir and Thomas Ludwig: When the technology becomes mother of invention. Abstract accepted in HCI 2015*

4.1 Introduction

Chapter 4 provides an overview of the current profile of the chosen case study i.e. ESGF, for this research. The information provided in this chapter aims at enhancing the reader's understanding on the profile of the case study and its setting. Moreover, some technical details about the case study, such as architecture of ESGF are also provided. The information provided in this chapter mainly stems from the analysis of the relevant documents from the ESGF such as publications, technical reports, deliverables and other documents. However, some information is also derived from the interviews conducted with the stakeholders of ESGF. Consequently, this chapter provides a reflection on the evolution, mission and operations of the case study.

In this chapter, relevant information about the historical background of the case study organisation; ESGF, is provided to the reader in Section 4.2. The list of data projects served by the ESGF e-Science infrastructure is given in Section 4.3, the way a user interacts with the ESGF e-Science infrastructure is provided in Section 4.4 and the basic architecture of the ESGF system is discussed in Section 4.5. Finally, the planned further expansion of the ESGF from the domain of climate science to other domains such as astrophysics, biology, energy, chemistry and others is given in Section 4.5, followed by a brief summary of the chapter in Section 4.6.

4.2 History of ESGF

In this section the history of development of the ESGF along with the significant changes in the infrastructure and organisation structure through time are described. It is important to see the developmental steps that the ESGF has been undergone. At first ESGF (then known as ESG) project was initiated as a grid computing research case, just to test the ability of grid computing and its associated technologies. Overtime, the technology matured enough to enable hosting data for research purposes, initially chosen to host climate data. Another important aspect to notice is the dynamic and ever changing structure of ESGF with time, if we observe the history of ESGF.

The history of ESGF is divided into four phases; phase 1 from 1999 to 2001 when it was called ESG-I, phase 2 from 2001 to 2006 when it was called ESG-II, phase 3 from 2006 to 2011 when it was called ESG-CET and finally the current phase: ESGF which is from 2011 onwards. These phases are discussed in Sections 4.1.1 to 4.1.4. The summary of salient features of the historical development of ESGF is also given in Table 11.1, page 326 in Appendix A. The summary of the complete developmental history of ESGF include; the versions of ESGF, geographical range, data-holdings (models / projects) supported, total number of registered users, active users, scalability of architecture, salient features, funders, contributing institutes and salient features of user support provided to use the versions of ESGF system from 1999 up until now.

4.2.1 ESG-I phase 1 (1999-2001)

There are varieties of problems faced by climate scientists, one of them is the need to efficiently access and manipulate climate data for research purposes. Climate scientists must collect number of datasets and analyse them, but these datasets are scattered and are accessible via different platforms using different tools which indeed are time consuming and inefficient in many cases. Therefore, in order to combat this problem a need was felt to create a common environment which could provide a common platform to not only access climate data sets but also analyze those using analysis tools. Consequently, an initiative began in 1999 with the name of “Prototyping an Earth System Grid” (ESG-I) funded under the auspices of DOE’s Next Generation Internet programme (NGI) to cater the needs of climate scientists and to fulfil the emerging challenge of climate data (Bernholdt et al., 2005). The contributing institutes in ESG were Argonne National Laboratory (ANL), Los Alamos National Laboratory (LANL), Lawrence Berkley National Laboratory (LBNL), Lawrence Livermore National Laboratory (LLNL), National Centre for Academic Research (NCAR) and University of Southern California’s Information Sciences Institute (USC/ISI).

In this initial phase, ESG was able to achieve not only the goals of large data-set movement and replication between participating institutes via data grid technologies developed by ESG, but also ESG was able to develop a prototype of climate data

browser. As a result of this achievement ESG got the hottest infrastructure award at a Supercomputing Conference (SC) in year 2000. ESG though demonstrated the potential for remotely accessing and analyzing climate data scattered across different sites within a country with the data transfer rate of 500 MB per second; however, it was still a prototype with few real users. Therefore, it was a technical demonstration of future “*to be*” collective data platform for climate researchers. It is important to note that before the initiation of ESG-I, there was no central archive system to serve the stored climate data. At this stage since the system was a prototype, user support considerations were not made. The success of ESG opened ways to start ESG-II described in the following sub-section.

4.2.2 ESG-II phase 2 (2001-2006)

The success of ESG prototype encouraged DOE to fund another phase of ESG project, known as ESG II whose major aim was to “turn the climate data sets into community resources” under Scientific Discovery through Advanced Computing (SciDAC) programme. Since the ESG prototype was ready, it was important to put it into practice to encourage users to use the system by offering some data holdings to the users. Therefore, ESG II started to dispense Community Climate System Model (CCSM), the Parallel Climate Model (PCM) and the phase 3 of the Coupled Model Intercomparison Project (CMIP3) model data archived at PCMDI. “This first production system led to major advances in model archiving, data management and sharing of distributed data” (D. N. Williams et al., 2009).

Subsequently, ESG II efforts focused on developing technologies to offer the user access to the ESG II system through a web-based security for user registration via a web portal. In addition, the technologies included extracting meta-data from catalogue files and distributed data transport capabilities via OPeNDAP-g¹⁵ protocol. As a result, the system started supporting 10,000 registered international users and managed some 200 terra-bytes (TB) of data (D. N. Williams et al., 2009). At this point, at least an informal user support need came into being to serve the registered users to cater their technical needs.

In this phase, one can observe that the e-Science infrastructure *prototype* engineered in the first phase of ESG I, was evolving with the inclusion of end-users and an addition of another participating institution i.e. a stakeholder, Oak Ridge National Laboratory (ORNL). The product was the ESG II e-Science infrastructure with more data sets added to PCM data model archive and the inclusion of two new data model archives namely: CCSM and CMIP3. Therefore, ESG II was the service provider of the above

¹⁵ Open-source Project for a Network Data Access Protocol, an architecture for data transport including standards for encapsulating structured data and describing data attributes implemented in a distributed computing environment (www.opendap.org/)

mentioned products which are the scientific products in this case. CMIP3 was used to produce IPCC's AR4 report. In this regard, CMIP3 data users used ESG II communication channels to provide suggestions to enhance portability, accuracy and performance issues about climate models. This was the first instance where end-users of the product interacted with the developers; it is interesting to note that there were no considerations of formal user support about the usage of the ESG II. The further evolution of ESG is given in the next section.

4.2.3 ESG-CET phase 3 (2006-2011)

ESG entered a new structural and organizational form with the name of Earth System Grid - Centre for Enabling Technologies (ESG-CET) phase 3, after funding for another phase from DOE's Offices of Advanced Scientific Computing Research (OASCR) and Biological Environmental Research (OBER). The primary goal was to extrapolate the existing system to be compatible to incorporate more data types and data archives at different sites that are further distributed and diverse in nature, even beyond national boundaries (D. N. Williams et al., 2009). Hence, this phase was geared towards fulfilling demands of users i.e. climate researchers, around the globe, to provide them access to: data, information, models, analysis tools, and computational resources required making sense of enormous climate simulation and observational data sets for their research. Another challenge of ESG-CET was to extend the capabilities of the infrastructure, so that a user can conduct initial data analysis where data physically resides, thus reducing network over-head to transfer data. As a result, the extension of ESG-CET e-Science infrastructure to slot in these additional features was commended by American Meteorological Society (AMS) for leadership, which led to a new era in climate system analysis and understanding.

ESG-CET joined the Global Organization for Earth System Science Portals (GO-ESSP) consortium to have collaboration with other institutions. All institutions in GO-ESSP share common data-management interests thus building a community. Another institute, Pacific Marine Environmental Laboratory (PMEL), which is part of National Oceanic and Atmospheric Administration (NOAA), joined in 2010. Therefore, there was also a surge of data-holdings from different climate institutions, offered by ESG-CET to users. It included phases 3 and 5 of the Coupled Model Inter-comparison (CMIP3, CMIP5), Climate Science for a Sustainable Energy Future (CSSEF), Community Climate System Model (CCSM), Parallel Ocean Program (POP), North American Regional Climate Change Assessment Program (NARCCAP), Carbon Land Model Inter-comparison Project (C-LAMP), Atmospheric Infrared Sounder (AIS), Microwave Limb Sounder (MLS), Cloudsat and others in ESG-CET data archive system. Thus, ESG-CET data archive system got bigger and it served over 1 Peta-Bytes (PB) of climate data to 25,000 registered users with 500 users active per month.

As a consequence, this was a gigantic development that pulled users to use ESG-CET infrastructure, to get access to the data-holdings, especially for the generation of IPCC AR4 and IPCC AR5 reports. The interaction of users with the ESG-CET system to access data-holdings led to the necessity of user support. The users were beyond national boundaries thus user support was needed round the clock. For the users to get their problems solved, an effective and efficient user support system was needed, which was not formally present. Keeping this in view, communication channels between users, the developers of the ESG-CET technical system and data managers of concerned data projects were established. The most used channel of communication was via e-mail. In later years, multiple mailing-lists were established to cater the needs of different stakeholders.

In this phase, the main problem was a lack of stakeholders to realize the set-up of formalized user support system. Since the development and evolution of ESG-CET e-Science infrastructure was the primary concern, direct funding was not dedicated for the development of user support activities. It was in 2011 that an initiative was taken by one of the researchers i.e. the author, who was working on C3Grid, a collaborative project of the ESG; to investigate the user support process in e-Science infrastructure and ESG-CET, now ESGF, was chosen as a case study.

4.2.4 ESGF P2P: The current phase (from 2011 onwards)

“ESG has been around for over a decade, started in 1999, started with ESG-1, ESG-2, ESG-CET and now its ESGF. With the passage of time we transformed from client-server to peer to peer system” Interviewee L, Chair of ESGF.

The developments in the previous phase of ESG-CET continued with most of the funding under the DOE’s OBER. Additional funding institutes within the US included NASA, NOAA and NSF; most of them are maintaining and taking care of their concerned administrative jurisdictions including node(s). In the European Union (EU), large funding for ESGF is being provided by IS-ENES project (ENES, 2013). This phase was formally initiated in 2011 and is the current phase. Since then, ESGF-P2P has become an open consortium of institutions, laboratories and centers around the world, that are dedicated to supporting research of climate change, and its environmental and societal impact. With the inclusion of international institutions on board as stakeholders and inclusion of even more data-holdings (see Table 11.1, page 326 and Table 4.1 for a detailed view), the need was felt to generalize the system in the form of a federation to encourage and attract climate data providers worldwide. Consequently, the system architecture of ESGF P2P data archive system (which is detailed in Section 4.5) evolved in the form how the current ESGF peer-to-peer (P2P) looks like.

The federation includes: multiple universities and institutional partners in the US, Europe, Asia, and Australia, thus making it one of the outstanding e-Science infrastructures in the domain of climate science. This was the reason that during this phase, ESGF grew out of the larger Global Organization for Earth System Science Portals (GO-ESSP) community (ESGF, 2010). It now reflects a broad array of contributions from the collaborating partners. The complete list of collaborating institutions is summarized in Table 11.1 at page 326 in Appendix A (Earth System Grid Federation, 2010).

It is interesting to note that with the enormous growth in the organisational structure of ESGF, an upward trend of registered users was recorded (ESGF, 2015). The registered users reached almost 27,000 in number, from different parts of the world as shown in Figure 4.1, with almost 700 to 800 active users per month. With this rapid development in the ESGF organisation, there is an ever-increasing need to meet long-term user-support requirements, as the number of data-holdings and number of users rise. The prominent data holdings are described in Table 4.1. From Table 4.1 and Table 11.1 (page 326 at Appendix A), one can infer that the development of ESGF P2P network is a trend setter for open data sharing in an environment of multi-institute and global collaboration, where beneficiaries of the whole set-up are users i.e. climate researchers. Therefore, user support services cannot be ignored.

The developmental collaboration of various institutes around the globe has contributed socially, technically and politically to introduce a global data connectivity for the users. However, though certain improvements were made with the passage of time to service users by introducing an ad-hoc user support system. Yet, the full potential in delivering user support services was not achieved as the current user-support system lacks a directed and dedicated effort, as well as funding, to develop a long-term user support process (as it is evident from interviews with the stakeholders). Looking at the history of ESGF data archive system, the need to have a long-term, robust and scalable data archive system was sensed and fulfilled to some extent. However, the need to have long-term user support services was not highlighted in the policy of ESGF consortium. This was the reason that the funding was more or less oriented towards developing ESGF technically (to serve data holdings to the users), and efforts to support users were though present but insignificant. From the history of ESGF, it is evident that the data and computational resources are always increasing with the passage of time. There have always been new ways of organizing the ESGF system, i.e. revising the system architecture following new collaborative and organizational reforms to develop new methods of data access and discovery for users. This implies that ESGF is going through a continuous evolution of social, cultural, organizational, legal, institutional and technical re-structuring. Consequently, the users need a dynamic user support process which is adaptable to the changing needs of the system.

Looking at the history of ESGF, one can conclude that the architecture and organisation of the e-Science infrastructure has a well-set trajectory or momentum of offering more scalability, more data holdings, international collaborations of institutions and more users. Historians who have studied e-Science infrastructures have referred to this as “momentum” of an infrastructure and argued that once a particular “path” or momentum has established and tend to continue in a particular direction, making reversals or alteration become costly, difficult and in some cases impossible (Jackson, Edwards, Bowker, & Knobel, 2007). Therefore, it is the right time to study processes behind user support services as at this stage ESGF has achieved maturity in its infrastructural trend. If this is not done at the earliest, efforts of the developers and scientists may get wasted. Thus, as a consequence the full potential of this e-Science infrastructure and collaboration of global institutes may not be fully realized until or unless user support services are not streamlined. In a nutshell, ESGF has matured and became instrumental in contributing to the concept of Science 2.0. With the improvement of support services provided by ESGF, especially to users, this practice of using ESGF will become even more instrumental in flourishing e-Science practices. Offering Science 2.0 is the primary vision of ESGF, as evident from the following interview excerpt:

“ESGF is home of all structured data. We don’t care about free text in Science. Free text is limited to certain area than structured information. We want to capitalize on structured data; this is the vision of ESGF” Interviewee C, ESGF Developer.

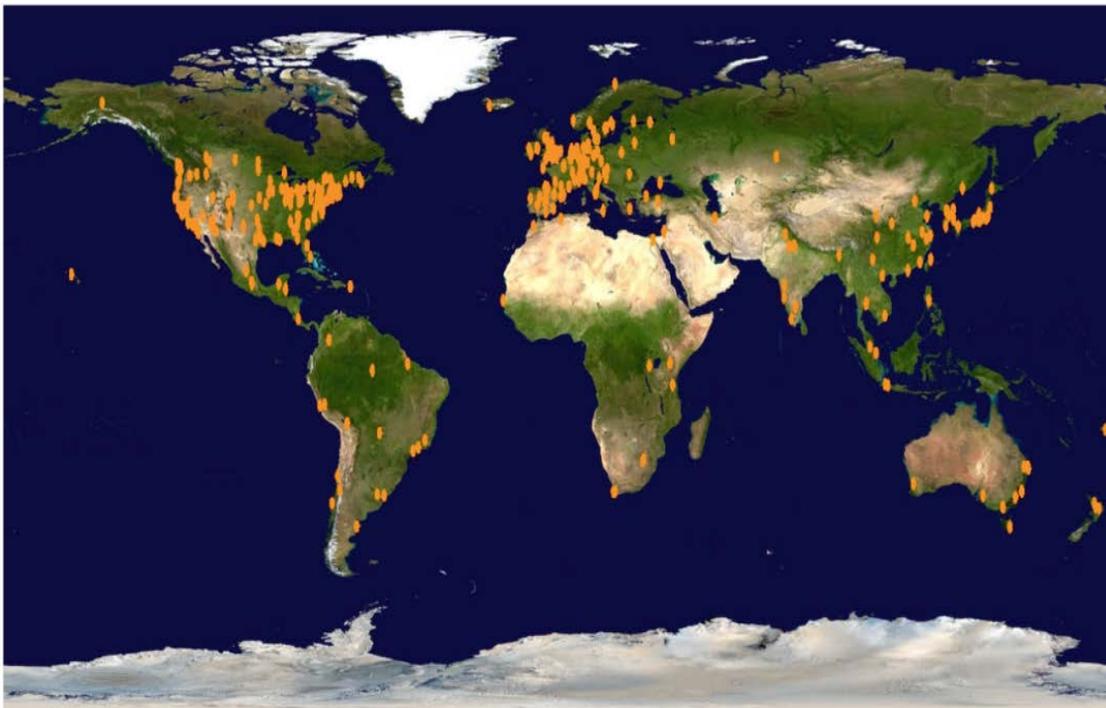


Figure 4.1 The distribution of registered users of in current phase of ESGF.

A social scientist who observed over the last three years the work of scientists, who developed ESGF and its associated projects, was of the opinion, that at times the technologies used in these projects were not stable enough. Moreover, she was of the opinion, that the developers of these infrastructures did a good job.

“There is funding up to the span of three years to develop a quite complex system. In e-Science infrastructures components or technologies are used e.g. Globus which themselves are not stable enough or stabilized. Having these conditions all workers did a good job in C3Grid and ESGF project” Interviewee A, Social Scientist.

Over the years, the transition of ESGF from research environment to production environment took place. With the current production environment of ESGF, supporting users has become a vital goal, as evident from the following quote of the chair of ESGF:

“Basically, we were funded in the US to do research and not to develop a production tool. In 2002, when I saw CMIP then I thought it can be possible to transit from the research environment to a production environment. Once we did that then we thought of how to support the users, so the user support environment became important” Interviewee L, ESGF Chair.

4.3 Projects supported by ESGF

World Climate Research Program (WCRP) has recommended that ESGF infrastructure be supported operationally, and that all future Model Intercomparison Projects (MIPs) follow the CMIP5 process and standards. The MIPs projects and others are given in Table 4.1. Since the beginning, ESGF has focussed on supporting these projects. The projects depicted with (*) at the end of their description, in Table 4.1, are all regional models.

The data projects are generated by climate scientists who are data project modellers from different data centres worldwide. In case the users are having certain queries or questions regarding the modelling background and other scientific queries for research purposes, the users access data via ESGF data archive system. Users should be able to contact the modelling domain experts to have a required scientific understanding of the research problem they are working on. At the moment via ESGF user support, there is no such sort of scientific user support provided, which can be integrated into it.

Table 4.1 The projects and models supported by the ESGF data archive system.

	Domain	Full model name
Model Intercomparison Projects (MIPs)	Model Intercomparison Project	Coupled Model Intercomparison Project (CMIP version 3 and 5)
	Model Intercomparison Project + Geo-engineering	Geo-engineering Model Intercomparison Project (GeoMIP)
	Model Intercomparison Project + Paleoclimate	Paleoclimate Modeling Intercomparison Project (PMIP)
	Model Intercomparison Project + Atmosphere	Transpose-Atmospheric Model Intercomparison Project (TAMIP)
	Model Intercomparison Project + Observational products	Observational products more accessible for coupled model intercomparison (obs4MIPs)
	Model Intercomparison Project + Reanalysis	Reanalysis for the coupled model intercomparison (ANA4MIPs)
	Model Intercomparison Project + Dynamical Core	Dynamical Core Model Intercomparison Project (DCMIP)
	Model Intercomparison Project + Carbon Land	Carbon Land Model Intercomparison Project (C-LAMP)
	Regional Projects	Regional climate Downscaling Experiment
Regional Climate Change Assessment Program for North America		North American Regional Climate Change Assessment Program (NARCCAP)*
Sounder Projects		Atmospheric Infrared Sounder (AIS)*
		Microwave Limb Sounder (MLS)*
Cloud Projects		Clouds and Cryosphere (cloud-cryo)
		European Union Cloud Intercomparison, Process Study & Evaluation Project (EUCLIPSE)*
Community Climate Projects		Community Climate System Model (CCSM)*
Ocean Projects		Parallel Ocean Program (POP)*
Sustainable Energy Projects		Climate Science for a Sustainable Energy Future (CSSEF)*

4.4 Interaction of a user with ESGF: An example

The normal workflow of a user, a scientist or a policy maker is elaborated as follows in the form of steps (also see Figure 4.2):

1. A user having a particular question e.g. “*What is the predicted change of sea surface temperature under different emission scenarios over the next 20 years?*” logs on into an ESGF gateway and searches for required data.
2. A user then requests data for download at a local storage facility or a desktop system.
3. The identity of the user is checked after the user provides authentication information.
4. The data sets are then located from the particular nodes by the ESGF system.
5. The data is retrieved and delivered by the system as a data file.
6. Or it is visualized using a data visualisation software package e.g. UV-CDAT.

These all steps are elaborated in Figure 4.2.

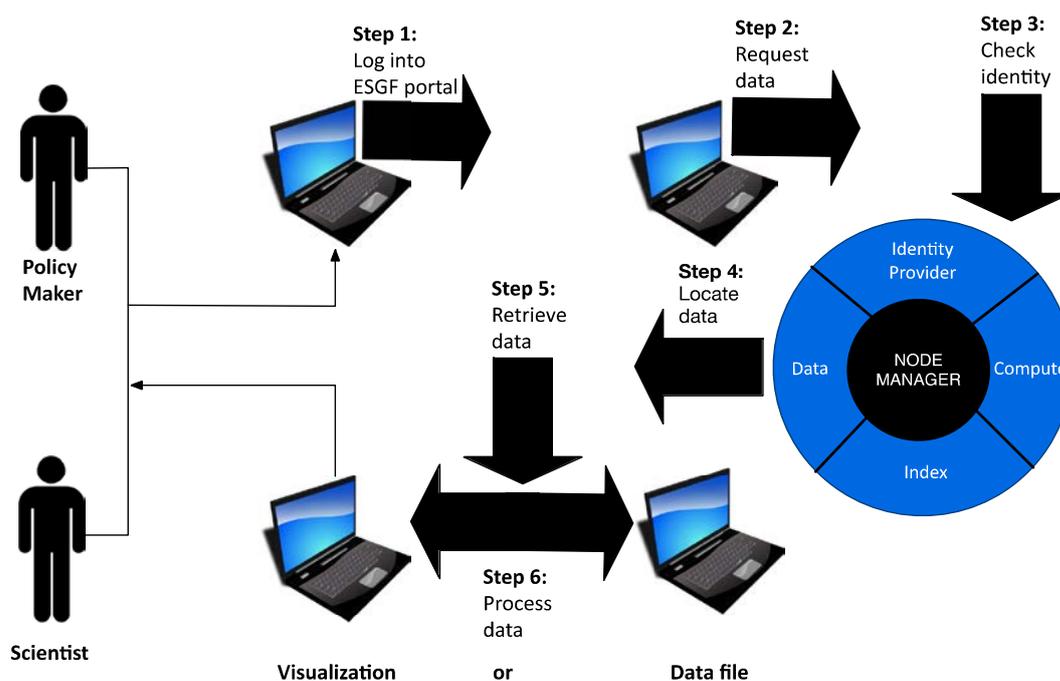


Figure 4.2 The steps that users do to get a data file from ESGF system and to visualize it using UV-CDAT visualization tool, adapted from (Dean N. Williams et al., 2013).

After explaining the steps which are shown in the Figure 4.2, let us consider a scientific use case where a user of the ESGF system is looking for some particular data sets from ESGF data archive with a following problem statement (that is also already described above):

“What is the predicted change of sea surface temperature under different emission scenarios over the next 20 years?”

A climate researcher would do the following steps to achieve the required information:

- The user logs in, in any of the gateway and via a faceted search he or she selects the following details: Project=CMIP5, frequency=monthly, variable=Sea Surface Temperature. With these variables she or he gets more or less ~5000 results. If the results are not displayed due to some error e.g. in UI, a user may contact user support.
- The user selects more constraints: Model=GFDL-ESM2M, experiment=1pctCO2. After selecting more constraints she or he gets 2 results. In this step, if a user has some query she or he might contact a modeller (a climate scientist) via support.
- Now the user does analysis with Lightweight Access Server (LAS) (lat/lon plot, time series, etc.). If there is a problem with LAS, the user may again contact user support.
- User inspects detailed metadata for model and experiment. At this step, metadata might be missing or incorrect, pushing the user to contact user support.
- After inspecting the metadata user may download individual files via browser (one by one, which is seldom the case), if there is problem with downloading the files the user may again contact user support.
- Or the user decides to download all files via Wget script¹⁶ instead of downloading individual files via browser.
- The user may use a desktop tool (e.g. UV-CDAT) for in-depth analysis over multiple files. However, in case of problems with the desktop tool functionality, a user may send user requests or may require to access “visualization help” information from online portal. In this case visualization experts about UV-CDAT can be consulted.
- The user may need to compare to observations MIPs: Project=obs4MIPs, may seek professional help from a modeller (a climate scientist). This help can also be provided via user support system.

While talking to one developer of ESGF, he indicated that in the future, computing the data facility offshore will be provided as well.

“At the moment it is search, find, fetch then it is going to be search, find and compute”
Interviewee C, ESGF Developer.

¹⁶ Wget script is a script via which a user can download bulk of datasets at once.

4.5 ESGF system architecture

In order to understand the ESGF user support, it is relevant to understand how the ESGF system is orchestrated. The architecture of ESGF was mainly driven by politics, as said by one of the developers of ESGF:

“Much of the architecture was driven from politics where people wanted to touch, hold and hug their data” Interviewee C, ESGF Developer.

ESGF evolved into a peer-to-peer (P2P) architecture to connect data from data centres worldwide mainly as a result of a political consensus. Data and metadata are published, stored and served from these multiple data centres around the globe called *nodes*. Nodes interoperate because of the adoption of common services, protocols and APIs, thus establishing mutual trust relationships, consequently, creating a federated structure in ESGF. The nodes are operated by administrative bodies i.e. federating partner institutions present at different locations, thus creating a distributed nature of ESGF. Subsequently, the data and metadata are accessed by ESGF users around the globe. The main driving force to have P2P architecture in the form of a federation was initially because of CMIP5 data project.

“The reason why ESGF exists was to collect and disseminate CMIP5 information. As the data got bigger, physically it was prohibitive to have it all at one place (i.e. at the LLNL-PCMDI institute). It was not feasible to put it at one place at the international level. We built the system in such a way that no one is the boss. There is no central thing (authority). We have peers in the form of nodes” Interviewee C, ESGF Developer.

In case of a problem, a user asks a question via a mailing-list and any data centre’s “*user support unit*” may answer the query, thus creating a distributed user support. ESGF architecture is designed in such a way that the nodes participating in the ESGF federation can join or leave dynamically, thus global data and services will change accordingly, which in turn will be impacting user support, hence making it dynamic in its nature too.

The architecture of ESGF is based on a higher performance, more reliability, and scalability, focused on federation services. ESGF is a system of *distributed* and *federated* nodes that interact *dynamically* through a P2P paradigm, see figures: Figure 4.3, Figure 4.4, Figure 4.5 and Figure 4.6.

“ESGF P2P system was built completely new, from scratch to achieve scalability, composability into set of pieces. Cardinality of a set does not have to be equivalent. The problem with the previous system was: It didn’t scale” Interviewee C, ESGF Developer.

The dynamic, distributed and federated characteristics of the system architecture of the ESGF are reflected in the organisation structure of ESGF. ESGF is a collection of administrative bodies (federating partner institutes) joined together in a similar manner,

forming a dynamic and distributed federation. The topology of ESGF organisation that is based on ESGF P2P data archive network, makes the user support also distributed and federated, (as it is also evident from the survey-questionnaire) because both are related, and both affect each other. That is why the user support units are distributed and connected via mailing-lists and request tracking software (RT). Normally, staffs present in these administrative bodies address incoming user requests, hence forming a federated user support unit, though formally there is no user support unit named, recognised or defined in the governance structure of ESGF.

The distributed, federated and dynamic properties of this organisation are very important in controlling the complete operations and services of ESGF P2P network. Dynamics, distributed nature and federation is also present in other e-Science infrastructures (Brasileiro, Andrade, Lopes, & Sampaio, 2011; Krauter, Buyya, & Maheswaran, 2002; Venugopal, Buyya, & Ramamohanarao, 2006). Therefore, if a user support process is designed for ESGF, it can be applied to other e-Science infrastructures with or without amendments.

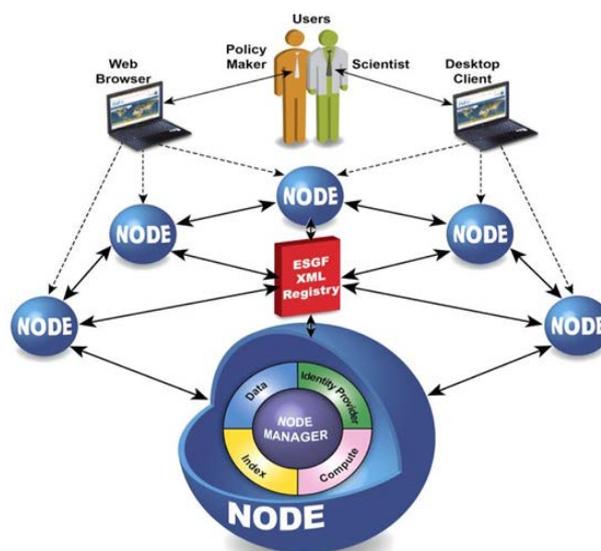


Figure 4.3 The peer-to-peer architecture of ESGF. The users can access ESGF data-holdings after authentication, either via web-browser or a desktop client, taken from (ESGF, 2013).

In ESGF, a client (browser or program) can start from any node in the federation and discover, download and analyse data from multiple locations as if they were stored in a single central archive. Different nodes of ESGF are operated by different institutes around the world and they can have one to four different flavours or functionalities in them (see Figure 4.4), as already been described earlier. The complete node architecture is described in the next Section 4.5.1.

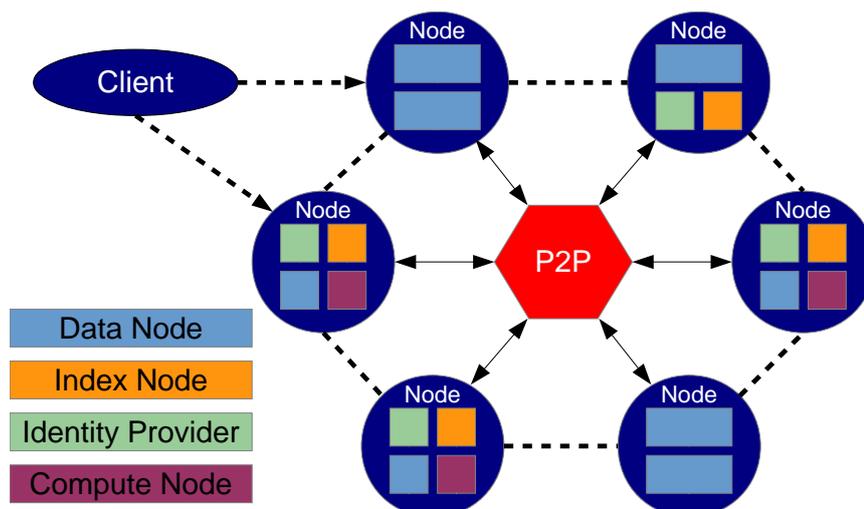


Figure 4.4 The P2P federated, distributed and dynamic architecture of ESGF with each node configurable in any of different flavours or roles. Each node can have 1 to 4 roles. Redrawn from (Dean N. Williams et al., 2013).

4.5.1 Node architecture

Internally, each ESGF node is composed of services and applications that collectively enable data and metadata access as well as user management. These nodes are administered by administrative bodies that are data centres at different locations. The staffs at those centres maintain the nodes, do further development and do any possible changes, if a need arises. In case the functionality of particular components of an ESGF node are affected, if intimated by users via user support system or known otherwise, then the concerned staff tracks the malfunction, solves the problem and communicates the solution to the user.

“There are four flavors in the nodes (i.e. four parts also known as four blocks):

1. *Data node*
2. *Security node or identity provider (IdP)*
3. *Quasi computational section or compute node*
4. *Index node or search*

Every node has the capability of becoming any one of these. For every single node that exists you do not need an IdP for it” Interviewee C, ESGF Developer.

Software components are grouped into four areas of functionality (also known as “flavours”). These four flavours of node architecture are linked to the diagnosis of a problem that leads to user requests. If a problem is encountered in any node then there can be a malfunction in any of the four blocks. However, if a node at a particular data centre location offers only one flavour, then at that location, the specialist (second level support agent) who may be responsible for the maintenance of node, can only check the

status of his or her node. Then she or he may answer the user requests if contacted by the first level support staff. For instance; let us imagine that a user-request comes from Beijing, China where a climate scientist (a user of ESGF) cannot access the ESGF data via gateway offered by DKRZ, then the user may report a problem via user support system and a concerned staff fixes a problem at the site (DKRZ). The flavours of nodes are as follows:

- **Data node** flavour; it enables secure data publication and access (depicted in blue colour in Figure 4.5 and Figure 4.6.)

“A data node exists at a place where data sits” Interviewee C, ESGF Developer.

- **Index node also known as “gateway” or search node** flavour, depicted in orange colour in Figure 4.5, has the following characteristics:

- metadata indexing and searching (with Apache Solr¹⁷ technology)
- web portal UI to drive human interaction
- dashboard suite of admin applications
- model metadata viewer plugin

“There are three index nodes that store index of all the metadata. From the index, one can get to the file” Interviewee C, ESGF Developer.

- **Security node also known as identity provider (IdP)** flavour is depicted in green colour in the Figure 4.5 and Figure 4.6. It is used for user authentication and group membership.

“In order to download the data or to do manipulation you need to go through the security system (i.e. IdP) for validation and authentication” Interviewee C, ESGF Developer.

- **Compute node** flavour is depicted in pink colour in Figure 4.5. Analysis and visualization of data is done via compute node.

“It is easy to put the megabytes of code to the data to do computation. We have a workflow engine and compute is we are working on now. We are enabling compute in one place and federating that across the federation. You can then get the results or the different parts of the results. Computations done and caching can be done. Popular computations are cached. Unpopular ones go to a tape” Interviewee C, ESGF Developer.

¹⁷ **Solr** (pronounced "solar") is an open source enterprise search platform from the **Apache Lucene** project. Its major features include full-text search, hit highlighting, faceted search, dynamic clustering, database integration, and rich document (e.g., Word, PDF) handling. Providing distributed search and index replication, Solr is highly scalable. Solr is the most popular enterprise search engine.

“Until December 2013 Compute part will be done till Dec 2013. All compute tools can be put together until end of the year” Interviewee C, ESGF Developer.

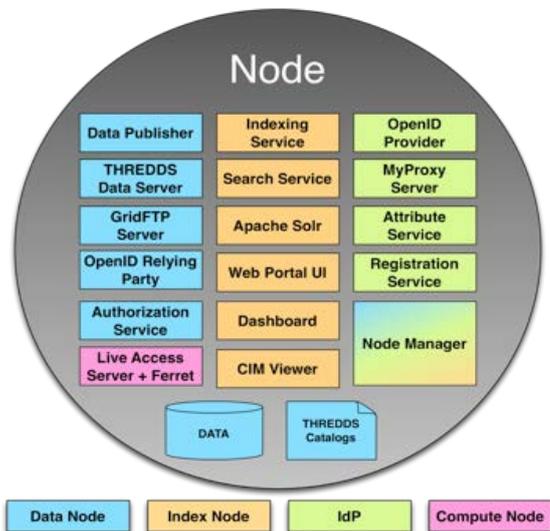


Figure 4.5 Software components that may be deployed onto an ESGF Node, grouped by “flavour” or area of functionality, taken from (Cinquini et al., 2012).

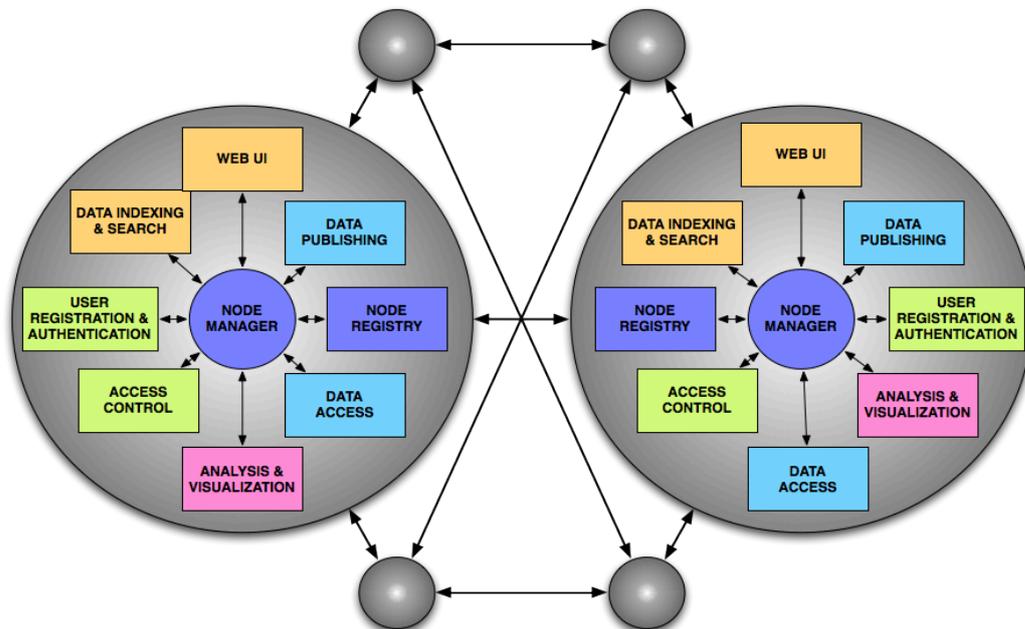


Figure 4.6 The components of node architecture coordinated by node manager, taken from (ESGF, 2013).

The user queries may be about any flavour or multiple flavours of the node depending on the part where malfunction is. Sometimes, it is hard and time consuming to diagnose a problem by the staffs and there are cases where multiple staffs collaborate to detect the problem. The different roles within a node make its behaviour dynamic, because a node can be configured in various combinations depending on site needs, or to achieve higher performance and scalability. Node manager is a central component and is installed for *all* node flavours to exchange services as well as state information amongst other nodes. ESGF data archive system, the data holdings and nodes including node roles are tied to the user support system and its interface with user. If these elements change, this may directly affect the number of incoming user support queries or incidents.

4.5.2 ESGF software

The software components used in ESGF combined together in the form of a bundle are known as “*software stack*”. The software stack combines custom software components developed by ESGF team with other freely available applications from e-Commerce solutions like Apache Tomcat¹⁸, Solr, Postgres¹⁹ and solutions used in geo-informatics branch including Thredds Data Server (TDS)²⁰ as well as Live Access Server (LAS)²¹. The complete documentation of the software components can be found at: esgf.org/wiki. The software stack comprises of the following sub-systems or components:

4.5.2.1 Node manager

Every node in ESGF has one node manager and a node registry. The node registry is an entry in XML file, which is part of the collective ESGF XML Registry file that contains information about all nodes and other attributes present in ESGF P2P network (see Figure 4.3). The functions of node manager are as follows:

- It enables continuous exchange of service and state information among nodes
- Internally, it collects node health information and metrics such as CPU, disk usage, etc. using Peer-To-Peer (P2P) protocol

¹⁸ Apache Tomcat (or simply Tomcat, formerly also *Jakarta Tomcat*) is an open source web-server and servlet container developed by the Apache Software Foundation (ASF).

¹⁹ Postgres is an open-source object-relational database management system (ORDBMS) with an emphasis on extensibility and standards-compliance.

²⁰ The THREDDS Data Server (TDS) is a web server that provides metadata and data access for scientific datasets, using a variety of remote data access protocols. (Source: <http://www.unidata.ucar.edu/software/thredds/current/tds/>)

²¹ The Live Access Server (LAS) is a highly configurable web server designed to provide flexible access to geo-referenced scientific data. (Source: <http://ferret.pmel.noaa.gov/LAS/>)

- Information among peer nodes is exchanged randomly using gossip protocol²². Each node receives information from one node, merges it with its own information and propagates it to two other nodes at random. Therefore, there is no central coordination and no single point of failure
- Nodes can join or leave the federation dynamically
- Each node is bootstrapped with knowledge of one default peer
- Each node can belong to one or more peer groups within which information is exchanged

“There are efforts going on in ESGF to make node discovery even faster. It should be fast to discover if someone joins. Less than 7 seconds in peer to peer network is the target. Get consistency across the federation is important.”

“Not every node can see every other node. The technical capabilities and policies are two different things. For instance; there is a policy, e.g. NASA cannot see China which is a policy decision from China” Interviewee C, ESGF Developer.

4.5.2.2 Node registry

It is an XML document that is payload of P2P protocol. It contains service endpoints²³ and SSL public keys for all nodes in the federation. Its derived products e.g. list of search shards, trusted IdPs and location of Attribute Services are used by federation wide services.

4.5.2.3 Software stack: Data search services (gateway or index node)

The main purpose of ESGF data archive is to make data holdings discoverable and accessible for end-users. This is achieved by ESGF search service, exclusively made for users or clients. The search service is implemented in ESGF software stack by indexing associated metadata. Indexing and searching, facilitates real-time distributed search with the help of Apache Solr²⁴. ESGF search architecture is shown in the Figure 4.7. It includes the following steps:

²² Gossip Protocol is a type of communication protocol used mostly in large scale distributed systems. It is based upon gossip conversations which is common in many social circles. Gossip Protocols are becoming increasingly popular in distributed systems due to its simplicity, scalability and high reliability even in constantly changing environments. Gossip Protocol involves periodic message exchanges between node pairs, which eventually results in information being spread throughout the system which is similar to human gossiping. (Source: <http://nufailm.blogspot.de/2012/02/gossip-protocol-introduction.html>)

²³ Service endpoint is a term used especially in web-services. The endpoint for a service allows clients of the service to find and communicate with the service.

²⁴ The implementation of ESGF data search service is based on Apache Solr because it is a popular open source search engine.

1. Metadata is harvested into master Solr using ESGF indexing service
2. Metadata is replicated to slave Solr
3. Search clients use REST API to query slave Solr
4. Slave Solr distributes query to other slave Solrs to return federation wide results

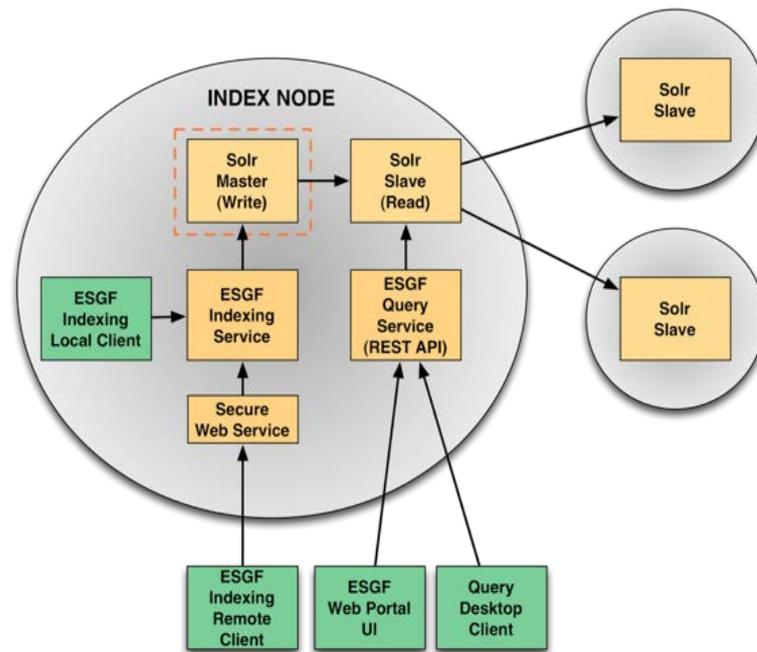


Figure 4.7 Data search services architecture in ESGF, taken from (Cinquini et al., 2012).

4.5.2.4 Software stack: Security infrastructure (identity provider-IdP)

The purpose of security infrastructure (also known as identity provider “IdP”) in ESGF is to enable access to distributed data resources by implementing local authentication and authorization policies. If a user faces some problems with registration, authentication and authorization then security infrastructure can be checked for the diagnosis of a problem by the concerned staffs that do user support activities.

The security infrastructure in ESGF provides federated security services, like; possibility of single-sign-on (SSO). SSO means that once users are registered at a site, they get authenticated at any other site in ESGF to access climate data using “OpenID,” in case of browsers and “X509,”²⁵ certificates for desktop clients. Moreover, security infrastructure in ESGF ensures distributed access control i.e. resources are controlled by

²⁵ An X.509 certificate binds a name to a public key value. The role of the certificate is to associate a public key with the identity contained in the X.509 certificate.

local policies as designed by a particular administrative body. XML registry tracks security services endpoints. User attributes are propagated as digital SAML²⁶ statements.

In a nutshell, ESGF security infrastructure is based on authorisation standards: OpenID, SAML, SSL, X509 (see Figure 4.8). ESGF security infrastructure is language neutral, currently implemented using Java and Python. ESGF security infrastructure comprises of non-intrusive filters and libraries. In order to communicate with data node the ESGF security infrastructure requires digital trust relationships among nodes.

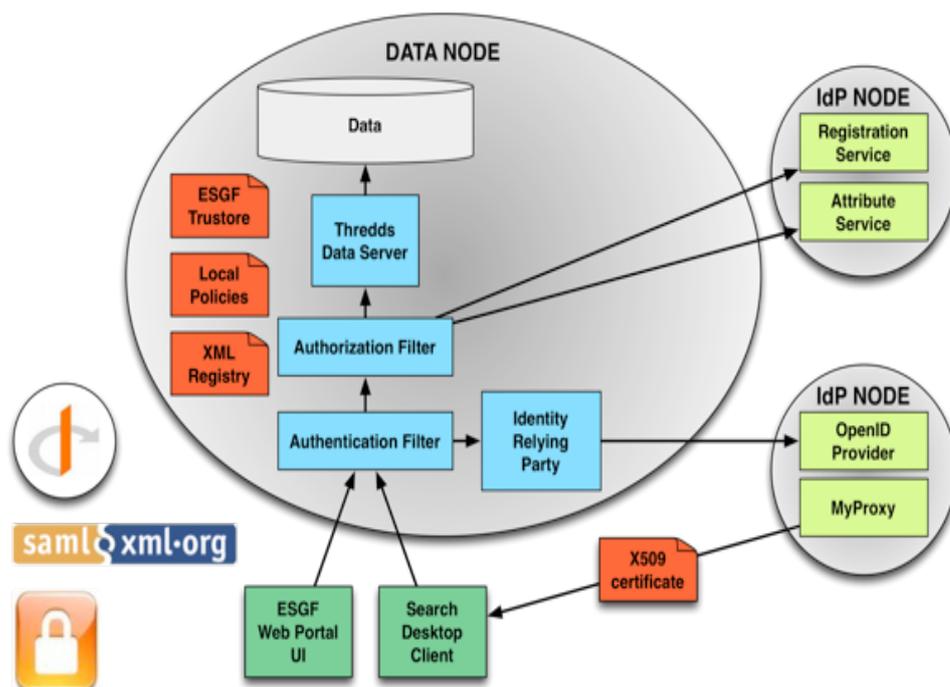


Figure 4.8 The architecture of ESGF Security Services framework (also known as identity provider “IdP”) taken from (Cinquini et al., 2012).

4.5.2.5 Software stack: Web portal user interface (UI) gateway

The purpose of web-based UI is to provide browser based user interface for users to interact with the ESGF data archive system to access data holdings. Users can register and have group membership via web portal UI. Users can also authenticate themselves by providing required fields. ESGF web portal UI offers administrative pages for user management. Furthermore, search UI is offered by ESGF web portal, UI where a user can do free text, facets, geospatial and temporal queries. Metadata inspection (collection-level data info, detailed modelling metadata) and hyperlinks to data

²⁶ Security Assertion Mark-up Language (SAML) is an XML-based open standard data format for exchanging authorization and authentication data between parties, in particular, between an identity provider and a service provider.

download (via HTTP, WGET²⁷, OpenDAP) and visualization (via LAS) is offered via ESGF web portal UI (also known as gateway). There are thirteen gateways at the moment offered by different institution participating in ESGF. All offer a uniform UI with a consistent design and function. A user can log on to any gateway and can perform search queries, using any gateway. A gateway offered by NASA is shown in Figure 4.9.

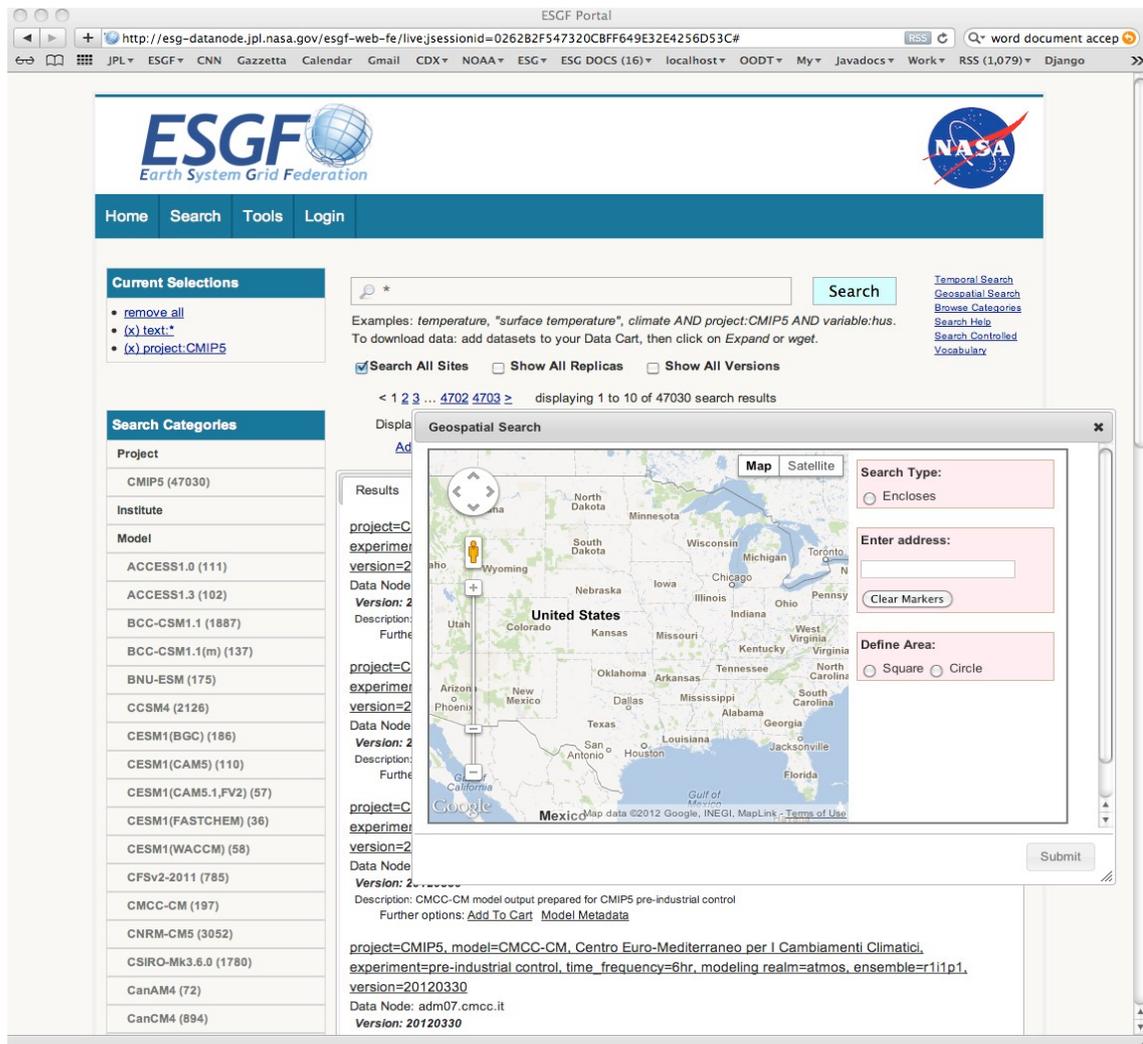


Figure 4.9 Screenshot of ESGF UI, NASA gateway.

4.5.2.6 Software stack: Dashboard (part of data node)

Dashboard is an administrative application for monitoring and reporting metrics for a single node and the whole federation. The following metrics are collected: Network topology (a.k.a. peer groups composition), node types (a.k.a. services for each host), registered users, downloaded data and system metrics like CPU, disk usage, service up time etc. All metrics are collected and stored in a relational data-base (RDB) called

²⁷ Wget is script via which ESGF can download climate data-sets.

dashboard catalogue. These metrics are then displayed via a UI based on Web 2.0 applications that display customizable charts and reports. The screenshot of a dashboard is shown in Figure 4.10.



Figure 4.10 The screenshot of UI of dashboard.

In an interview, the social scientist who observed the development and the social impacts of e-Science infrastructures like ESGF said, answering to the author's question:

“Did the (ESGF) components fulfil the ultimate goal?”

“Well we cannot measure it or judge, but do something, they are in a state of running, of course they do produce errors. You need to put some measures to judge that”
Interviewee A, Social Scientist.

The dashboard is one of the components that monitor the components to some extent which are in the state of running.

4.6 Future expansion of ESGF data archive system

ESGF needs to incorporate the following future challenges in order to expand the ESGF systems capability: ESGF needs to scale with “Big Data” produced by higher resolution models, satellites and instruments. The server side functionality is expected to be expanded. Server side processing through WPS (climate Indexes, custom algorithms) is needed and computer scientists are working on that as well as expanding the GIS mapping services (for climate change impact studies at regional and local scale). ESGF is making advances to facilitate a model to observations, intercomparisons and expand direct client access capabilities. The need for increased support for OpenDAP based

access is felt by the ESGF development team and it is expected to meet this need in the next few years.

Tracking provenance of complex processing workflows for reproducibility as well as the origin of data sets is a major challenge and ESGF has some top scientists working on that. At the same time, ESGF intends to package VMs for cloud deployment and instantiate ESGF nodes on demand for short life-time projects. As the data demands increase, ESGF is creating an environment with elastic allocation of back-end storage and computing resources. In the future, ESGF will cover other scientific domains such as health sciences, biology, chemistry, energy as well radio-astronomy, see Figure 4.11.

A comprehensive ESGF user support process is needed to be formally defined and then put into practice in the arena of ESGF data archive system. The current informal user support process is needed to be investigated. Based on the findings of the investigation, recommendations can be made in future by the investigators.

ESGF Roadmap: Vision for the Future

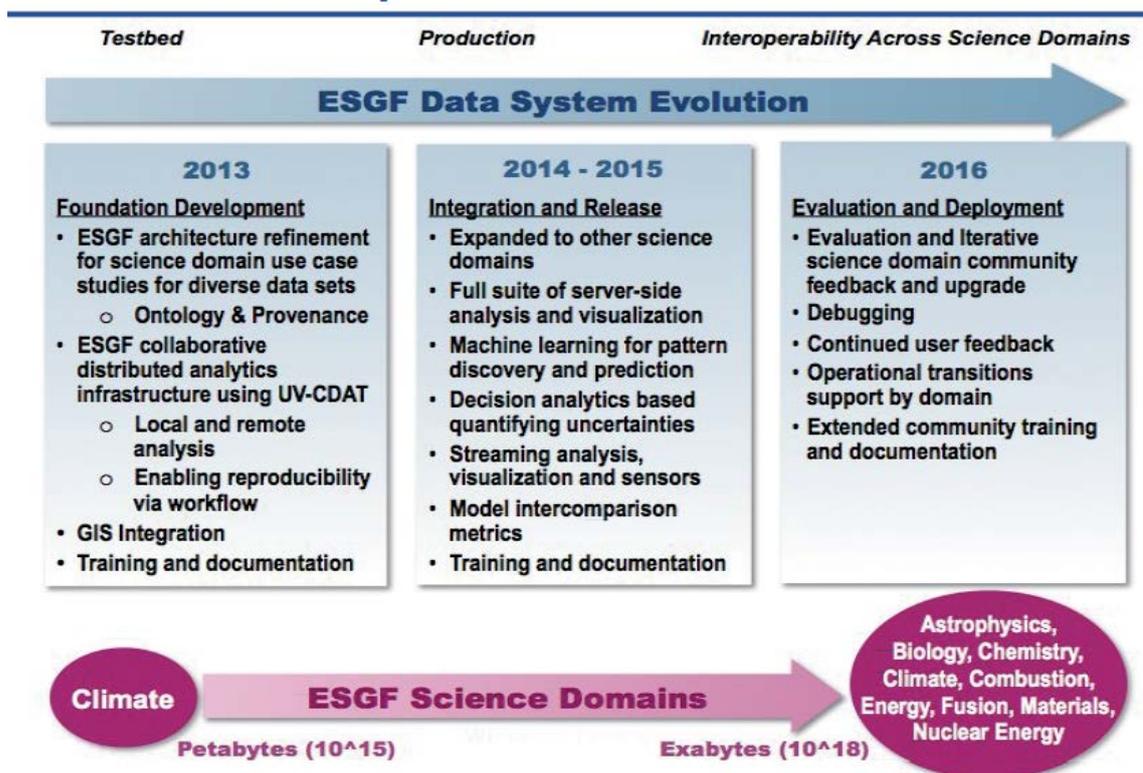


Figure 4.11 Future directions, the computer scientists are working on and the future science domains that ESGF may expand to.

“If data is very big then the solution is not cost effective” Interviewee E, Principal Investigator.

We have succeeded in the initial mission. We need to constantly anticipate the community needs. If documentation is written together with the collaboration of the staffs, we can mitigate a lot of support” Interviewee C, ESGF Developer.

“ESGF is a very huge and distributed e-infrastructure which makes the user support pretty challenging” Interviewee D, ESGF Developer and System Administrator.

“We are trying to extend the system to serve energy, biology, chemistry and domains of other data. We have sent out a proposal for biology; let us see how it turns out with the National Institutes of Health (NIH) in the U.S. We have also written a proposal for the energy domain” Interviewee L, ESGF Chair.

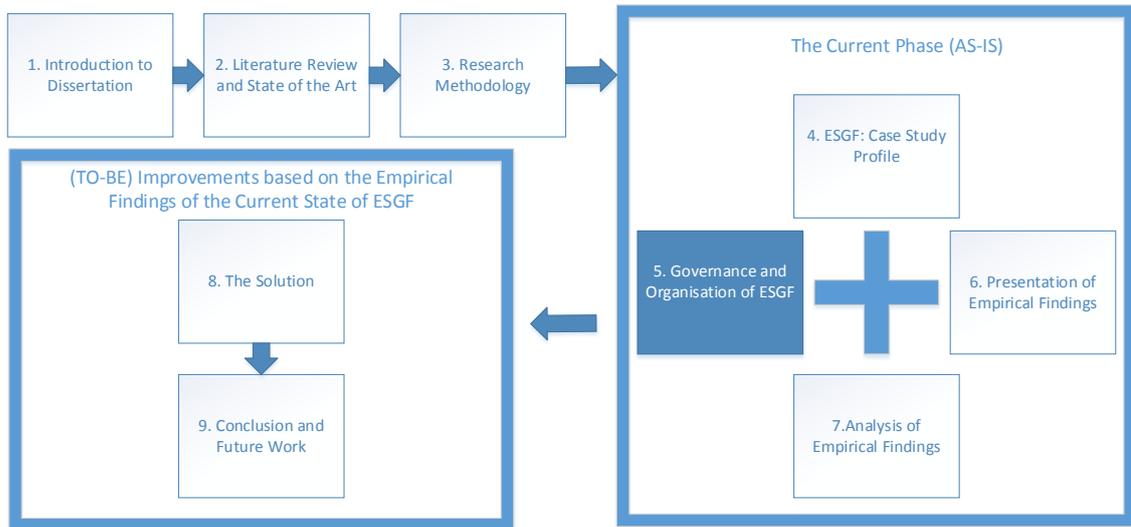
4.7 Chapter summary

In this chapter, the evolution of ESGF infrastructure and organisation was illustrated. ESGF, then known as ESG-I started as a grid technology test-bed to connect and serve climate data. With the passage of time, the funding from agencies such as DOE and other agencies continued to finance further phases of ESG, namely; ESG II, ESG CET and ESGF. The ESGF system reached its production level and served multiple climate data projects to the users who are doing research activities based on climate data around the globe.

The architecture of ESGF is based on peer-to-peer network of nodes, where a node can join or leave the network at any time, thus making ESGF dynamic and flexible. The nodes mainly comprise of data, published by data centres. The nodes can have four flavours namely; data, security also known as identity provider (IdP), index or search facility and finally compute possibility. A single node can have any of these four flavours, but data access facility or data flavour is a must.

The dynamic and evolving architecture of ESGF has an impact on the organisational features that make up this federation. Currently, ESGF is servicing the climate domain; however, in future it will be servicing biology, energy and other domains too. Thus, more users will be attracted to ESGF to get the data provided by this infrastructure and ask questions or seek solutions to their problems while using ESGF infrastructure. This makes the user support process of ESGF further challenging, significant, dynamic and ever changing.

5 Governance and organisation of ESGF



From this chapter the following peer-reviewed paper was published:

- Hashim Iqbal Chunpir, Thomas Ludwig: *Reviewing the Governance Structure of End-User Support in e-Science Infrastructures. GI-Jahrestagung 2014: 1687-1697*

The paper provides an over-view of the governance of ESGF and possible recommended changes to the governance structure. The paper was published in Informatik 2014 conference held at Stuttgart, Germany.

5.1 Introduction

In the previous chapter (Chapter 4), the historical development of the case study: ESGF, the data projects supported by ESGF, the planned future expansion and the system architecture were analysed. This chapter provides insights into the governance behind the ESGF consortium as well as the organisation of ESGF. This study has pointed at the current governance structure of ESGF and recommended changes in it. Moreover, for the first time the complete organisational view of the ESGF was visualized in the form of an organisational chart of a federated e-Science infrastructure. This chart is depicted in Figure 5.2.

The details about the current governance scheme of the ESGF federation, as extracted from the relevant documents as well as interviews are given in Section 5.2. The global picture of ESGF organisation in the form of a network of partner organisations is presented in in Section 5.3. The challenges that ESGF faces as a diverse networked organisation are highlighted in Section 5.4. The development philosophy of the ESGF

infrastructure is given in Section 5.5 and finally the summary of the chapter is given in Section 5.6.

5.2 ESGF current governance scheme

ESGF consortium is an international body of collaborating institutions, with every institute having its own norms, culture, community, specialization, hardware as well as software components, human resources, standards, reputation, goals and expectations.

“Computer scientists have their own goal, for example; goal to do research using the grid technology or the climate research centre wanted to improve their services for the climate researchers. Then the scientists who have the goal to better use the data and the tools. All these goals at some point accumulate in a shared goal. But the negotiating process in the development of an infrastructure tries to align these goals” Interviewee A, Social Scientist.

Therefore, a common governance structure is pivotal to determine the scope, common goals and resolution of conflicting interests. That is why, from the beginning, ESGF has its own governance set-up and policy based on the principles of “governance” developed over years. Refer to Figure 5.1 to have an overview of the current governance scheme. The reason why the governance structure is important for this study is that the user support in ESGF needs an explicit governance strategy. This governance structure in ESGF is comparable to general principles also called the “constitution” found in arena of political science, adhering to the principles of: Critical ideas of fairness, transparency, measurability, responsibility, accountability and performance. ESGF governance structure is composed of the following constituents, see Figure 5.1:

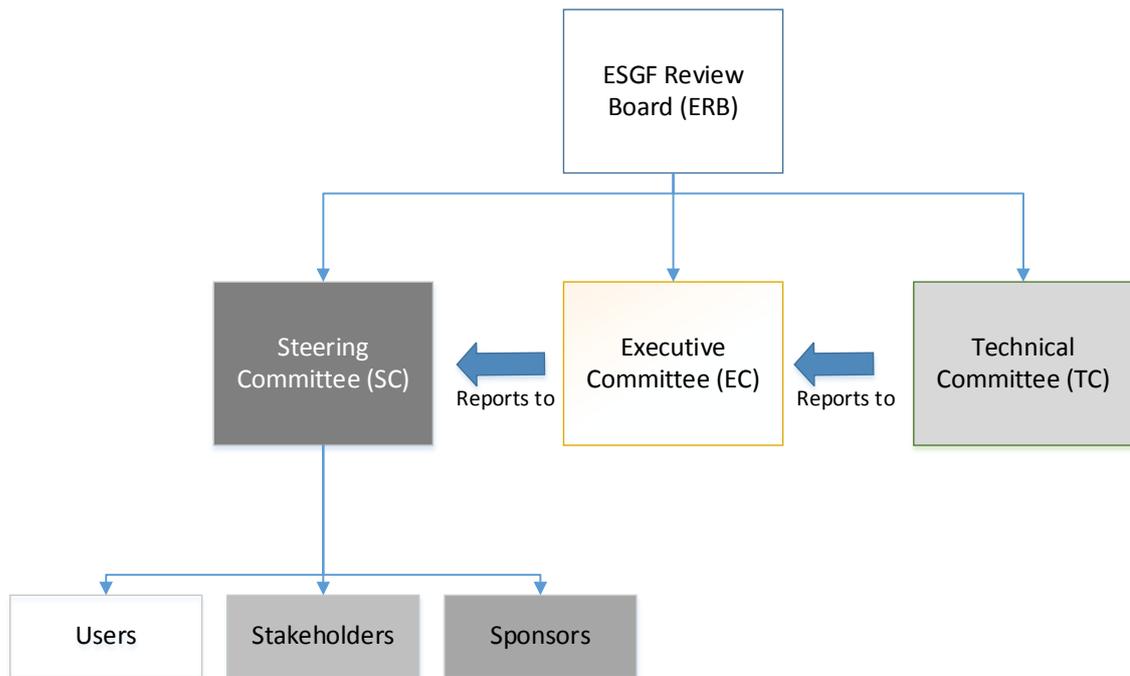


Figure 5.1 The governance scheme of ESGF.

5.2.1 The ESGF review board

ESGF has a review board known as “the ESGF Review Board (ERB)” which closely coordinates with and monitors the three committees: Executive Committee (EC), Steering Committee (SC) and the Technical Committee (TC), please refer to Figure 5.1. The aim of ERB is to protect the interests of users and developers of ESGF by reviewing the code changes that may create a high impact on users and developers (Dean N. Williams & ESGF, 2013). ERB reviews whether the changes done in ESGF archive system will significantly affect backwards compatibility. Consequently, ERB ensures backward compatibility in ESGF archive system. Furthermore, ERB prevents that any changes may affect users, managers or developers of ESGF. Moreover, ERB measures the magnitude of change; i.e. whether a change significantly shifts the functionality and scope of ESGF system. Finally, ERB takes care of any licensing issues and security concerns that are needed to be addressed (Dean N. Williams & ESGF, 2013).

ERB committees meet at least 4 times a year, with additional meetings that may be organised as necessary. There is no fixed schedule to conduct meeting. It is important to observe that the ERB allows debate to settle issues regarding backwards compatibility of the system and protecting adverse changes to take place as far as user’s usability and developer’s development of the system is concerned. However, from the user support services point of view, there is no clause about protection of user support services, which may be added in the agenda of ERB. It might also be useful to fix the schedule of ERB committee meetings. It is pertinent to observe that ERB sets the direction of ESGF

manifesto and allows a common platform to all other committees to interact with one another.

5.2.2 Executive committee

Executive committee (EC) interacts with steering committee (SC) and technical committee (TC), simultaneously. EC is responsible for setting the strategic direction of the ESGF project. It oversees the activities of technical committee and other committees as defined by the EC. Subsequently, EC reviews the requirements of ESGF regarding its architecture, developmental plan, setting priorities and analysing user requirements as reported by the TC to the EC on a defined basis. Finally, EC gives its approval to major architectural changes to be undertaken in ESGF by the TC. EC meets at least once in a month which is sufficient. EC is responsible for organizing all major activities and events of ESGF (Dean N. Williams & ESGF, 2013). EC does not have an explicit agenda of user support in its manifesto. It is recommended by the author to include an explicit agenda about user support operations.

5.2.3 Technical committee

The main activities of the technical committee (TC) are: Capturing and analysing user requirements. Consequently, based on the user requirements TC defines and designs the technical architecture of ESGF. In addition TC organizes ESGF software developmental activities. For instance; design and development of ESGF software stack, testing and releasing component of ESGF software and reporting to the executive committee (EC). TC meets at least once in a week and the communication amongst all members and participants takes place via tele-conferencing tools. Meeting once in a week to discuss important points is sufficient. The chairs of EC and TC communicate weekly to monthly; therefore there is no regular meeting plan, which may be made and the results of discussions may be made public to all members so that the members of EC and TC know about the important ongoing issues and their results. TC has been playing a role in solving user requests and conducting user support activities in ESGF but so far did not investigate the current user support process thoroughly.

5.2.4 Steering committee

The steering committee (SC) comprises of mainly sponsors and other stakeholders. SC provides input to the EC on the entire ESGF project and the path forward from the funding perspective. SC communicates the funding goal to EC, advises on funding issues, represents the sponsors and reviews the overall progress. So far, there is no explicit funding dedicated to user support by SC.

5.2.5 Annual face to face meeting and other roles in ESGF

Face-to-face meeting amongst all members and committees takes place after every 18 months. The objective of face-to-face meeting is to detail progress and future community requirements. However, 18 months is quite a long time. Face-to-face meeting should take place at least twice a year after every 6 months so that there is a frequent exchange about the details of the progress of ESGF. In addition, the schedule of all the concerned meetings should be made available earlier to the stakeholders and the results must be communicated to the stakeholders along with the users. Up until now the issue of long term user support strategy was not discussed during the face-to-face meeting. Moreover, users are not explicitly invited to the annual face-to-face meeting. In the meeting mainly sponsors and developers of the ESGF infrastructure participate for 3 to 4 days. Users can be made part of ESGF face-to-face meeting and one to two days can be dedicated to them. Users can also take advantage of this forum and can present their point of views on specific issues they face. At the same time, the users can get required information delivered in the form of presentations from computer scientists, climate scientist and data mangers.

What is missing in the current governance scheme is that there is no indication about governance of user support, which can be included. A user task force (UTF) can be formed to cater the support requirements of end-users of the ESGF system. Another important development can be to form a user services committee (USC) that would include representation of all types of users along with other committee. The user's committee may comprise of users and ESGF developers who closely interact with users to define user requirements and to fulfil the user requirements by managing issues and implementing the features that facilitate users to take advantage of ESGF P2P e-Science infrastructure to meet their research goals. User surveys should be conducted on regular terms so that the fulfilment of user requirements and their satisfaction can be measured.

5.2.6 ERB committee roles

In the ERB committees there are following roles:

5.2.6.1 Chair and co-chairs

The chair (or co-chairs) organizes meeting agendas and maintains the roadmap and the list of outstanding proposals that require ERB intervention. They may invite individuals or groups who have submitted proposals to present their plans at ERB meetings. They are responsible for facilitating discussions and consensus building in their perspective committees. Therefore, chair or co-chairs are the organisers of the ESGF events and perform as an arbitrator in unresolved matters.

5.2.6.2 Voting members

For each committee, there can be only one voting member per institution. Voting members candidates, one per institution, must be vetted and approved by the respective committees and Chairs or co-Chairs. Voting member candidates must be participant members for at least one year before being nominated as a voting member. However, it is important to take care that if voting members from a particular country or continent are in majority they may politicize or monopolize the direction of development of this infrastructure.

Conflicts are resolved by forming a consensus and discussion amongst the voting members. If agreement is not possible then voting may be held. In case of a tie, chair or co-chair may break the tie.

5.2.6.3 Non-voting members

Non-voting members on the executive committee must be voting members on the technical committee and non-voting members on the technical committee must be voting members on the executive committee.

5.2.6.4 Participatory members

Each committee can invite participants to join meetings. Participants are guests and hold no voting privileges. They may contribute to the code base and participate in meetings on limited bases. Users must be invited and encouraged to join the meetings and they may get a proportion of representation and voting rights to contribute their requirements to the ESGF and influence the development of ESGF according to their wishes.

5.2.6.5 Membership of new members

Executive members are responsible for nominating new members, who are elected by consensus or majority vote, with the chair breaking any tie. ERB technical members are also responsible for nominating new members with final approval of the EC who are elected by consensus or majority vote, with the co-chairs breaking any tie in the case of a conflict.

It is important to make sure that no monopoly should be developed that would be unhealthy to the future direction and development of the ESGF goals and may cause trouble or dissatisfaction amongst not only other members but also the users of ESGF. From the governance of ESGF point of view; one can say that ESGF system builders i.e. infrastructural designers, developers and administrators are “heterogeneous engineers” working together not only with technology but also working on organisational, social, institutional structure, working values and managing expenditure (Jackson et al., 2007). ESGF governance mechanism is a central monitoring, coordination and development authority whose procedures and strategies are set by

members representing the local institute which is part of internetworks of other institutes, geographically distributed and complex in nature. ESGF governance mechanism must include user support services strategies as part and parcel of the overall ESGF development process because user support services facilitate the utility of the system providing data search services to not only the various sorts of users but also developers of ESGF.

To summarize the whole discussion, one can claim that the collaborative governance is necessary in which the contributors to the ESGF can also be able to fulfil their goals as well the goals set by the ESGF governance. From the interviews, a social scientist indicated that there is a risk that personal goals of a worker might be sacrificed because at times the goals of infrastructure are different:

“To achieve the shared goal of an infrastructure, personal goals of people working for the infrastructure might get butchered” Interviewee A, Social Scientist.

A system that supports collaborative governance that outfits to highlight the personal goals of people might be worthy to align and synergise common goals and the personal goals. Collaborative governance shall minimise the risk of personal goals of the employees to get sacrificed.

5.3 Organisation of ESGF P2P network, the global picture

The current organisational structure of ESGF had no formal structure, no formal hierarchy and there is a high level of sovereignty that every federating partner of ESGF consortium enjoys.

“I don’t think there is a formal structure; it’s a consortium of institutions in EU, US and Australia and in Egypt. Common data archive structure for data delivery.

There is not really a formal hierarchy, here people get together because they want to and they respect certain level of sovereignty but they do not have a defined structure” Interviewee D, ESGF Developer and System Administrator.

With LLNL as a leading partner, international institutions such as the British Atmospheric Data Centre (BADC), German Climate Computing Centre (DKRZ), and the Australia National University (ANU) are undertaking the software development and project management of ESGF. For UV-CDAT²⁸, most of the software developer and

²⁸ Ultra scale Visualization-Climate Data Analysis Tool is visualization and a simulation tool which is joint effort to visualize climate model outputs. “It is a powerful and complete front-end to a rich set of visual-data exploration and analysis capabilities, well suited for climate-data analysis problems. UV-CDAT is an open source, easy-to-use application that links together disparate software subsystems and packages to form an integrated environment for analysis and visualization. This project seeks to advance climate science by fulfilling computational and diagnostic as well as visualization capabilities needed for DOE's climate research.”

(Source: <http://uvcdat.llnl.gov/>)

project management teams reside in the United States at DOE/LLNL, New York University (NYU), NASA, NOAA and many other institutions. The sponsors of ESGF and UV-CDAT are represented at the top of Figure 5.2 and Figure 5.3 and include DOE (the primary funder), NASA, NOAA, the European Commission, the Australian government and others. For ESGF and UV-CDAT, there are dozens of data centres around the globe that use these products for data projects; these projects are represented in the figures as blue and green circles. In the Figure 5.2 and Figure 5.3, CMIP5 is shown as a green circle to show that it is the primary project servicing data and analysis for the community. ESGF provided CMIP5 data to the climate experts that generated the 2013 Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report (AR5), a document on the basis of which political decisions are made. Given the past growth of ESGF and UV-CDAT, it is reasonable to assume that the number of future data projects (depicted as blue circles) will increase as time goes on. Likewise, these projects will continue to deliver more climate and weather data for important reports, which will help our understanding of future climate change.

Users with specific queries about these data holdings can consult climate specialists. Entertaining scientific queries along with technical issues is currently a large part of the user support services provided by ESGF and UV-CDAT. Depicted as an elongated disk in Figure 5.2 and Figure 5.3, the data holdings, analyses, and visualizations are made available to users worldwide. For more information on ESGF, visit the reference (Cinquini et al., 2012) and for more information on UV-CDAT, visit the reference (Dean N. Williams et al., 2013).

The executive components of ESGF and UV-CDAT are represented in the figures as rectangles. The ESGF and UV-CDAT development and maintenance teams, also known as administrative bodies, are depicted as diamonds. Development and maintenance teams, led by team heads, currently exist in Asia, Australia, Europe, and North America, with South American and African teams joining soon. Each of these teams manages and maintains the nodes within their region. These nodes collaborate with other nodes to form a system that is not only technically federated but also socially, institutionally and administratively. UV-CDAT is planned to be an integral part of the compute component of the ESGF software stack (H Chunpir, Williams, Cinquini, & Kindermann, 2013).

It is worth noting that the labels attached to each team or administrative body in Figure 5.2 with letters D, G, S and C indicate that a node may have four different roles or flavours within itself (Cinquini, 2013). The letter “D” stands for data node, which means that data holdings are hosted by an administrative body, while the associated number shows how many data nodes are being hosted and maintained by a single administrative body. For example, a label “3D” would indicate that there are three data nodes, which are managed by a particular administrative body. Every node has at minimum a data node and additionally may have a gateway (G), security set-up (S)

and/or compute facility (C) node. A gateway node, also known as an index node, is responsible for representing the data sets available in ESGF system to a user via user interface (UI).

A security node, also known as an identity provider (IdP), is responsible for identification and authorization of a user to ensure that a user is a valid entity who is entitled to access data sets available in ESGF system via its UI gateway from any ESGF node via single sign-on (SSO). A compute node is responsible for computing and visualization in ESGF. It is also where the UV-CDAT software stack is integrated for analysis, diagnostics, model metrics, and visualization.

ESGF subsystems or system components can be anywhere in the world. In addition, the network topology of the ESGF node network directly influences its management and administration structure. Figure 5.4 shows the international network governance of ESGF, known as the International Climate Network Working Group (ICNWG). This working group is dedicated to helping all ESGF climate data sites set up, optimize, and troubleshoot their network infrastructures for international climate data transfers, so that petabytes of data can traverse international networks from end-to-end at the high performance levels required for large-scale data analysis. By 2016, ICNWG will aim to achieve 2 GB/sec for data transfer throughput between at least five ESGF sites.

The number of administrative bodies, data holdings, users, and staffs participating in these systems are likely to increase. Additionally, the role of an administrative body or a node itself may change, and thus the whole arrangement is complex and dynamic. In these systems, demand continuous architecture redesigns activities, software development, hardware changes, data publishing, data curation, data quality check activities, analysis, diagnostics, model metrics and visualization. A dynamic and ever-evolving infrastructure also needs a dynamic user support “service desk.” Computationally intensive sciences carried out in highly distributed network environments, and as in our case, climate science that uses immense data sets requiring grid computing, are likely to be confronted with demanding user support issues that are quite similar to data curation and software development requirements.

The total institutions behind ESGF are reported to be twenty that are responsible for the development and operation of ESGF.

“We have 20 institutions and in each institution there are 5 to 10 people on the average serving for the ESGF consortium. We have ten core institutions and ten other related institutions that install ESGF software stack, running nodes as well as do maintaining activities.” Interviewee D, ESGF Developer and System Administrator.

The comments about the chart from an interviewer:

“In case of distributed organisation it is very important to have a P2P organisational chart of ESGF. What I like about it is a great view; it is nice to have one organisational chart” Interviewee L, ESGF Chair.

“The governance of the project ESGF is broken into three categories: The 1st one is the stakeholders and these are the projects that you listed right. The second part I have it there is the executive committee. EC are the folks that you listed the PI. And then there is the third one which is the team lead. Under the PI you have the team leads like Gavin, Stephen Pescoe and Luca” Interviewee L, ESGF Chair.

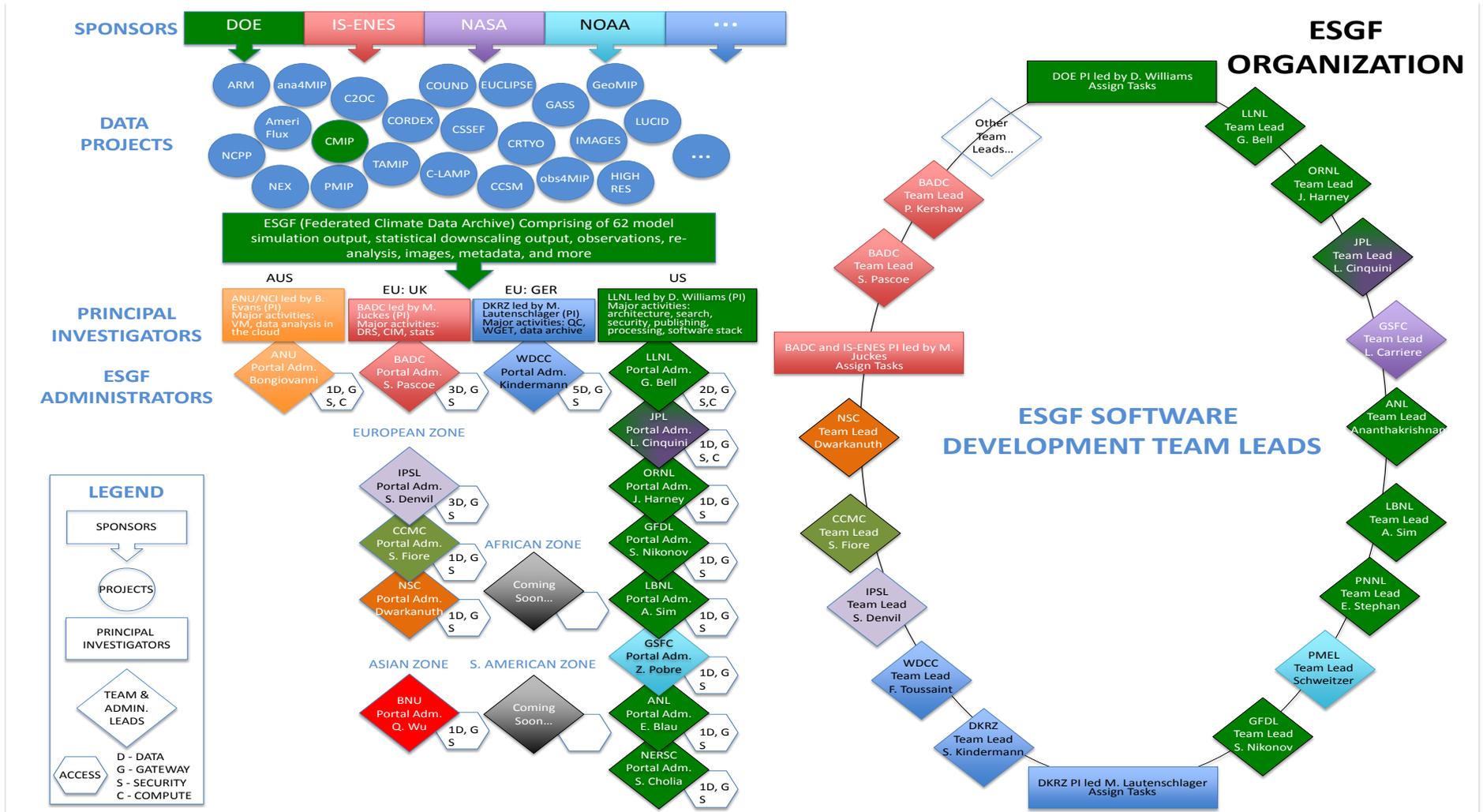


Figure 5.2 Various global organisations, geographically distributed, forming an ESGF consortium by connecting physical and human resources. The different colours show different entities, the diagram is represented using the notation coined by the author.

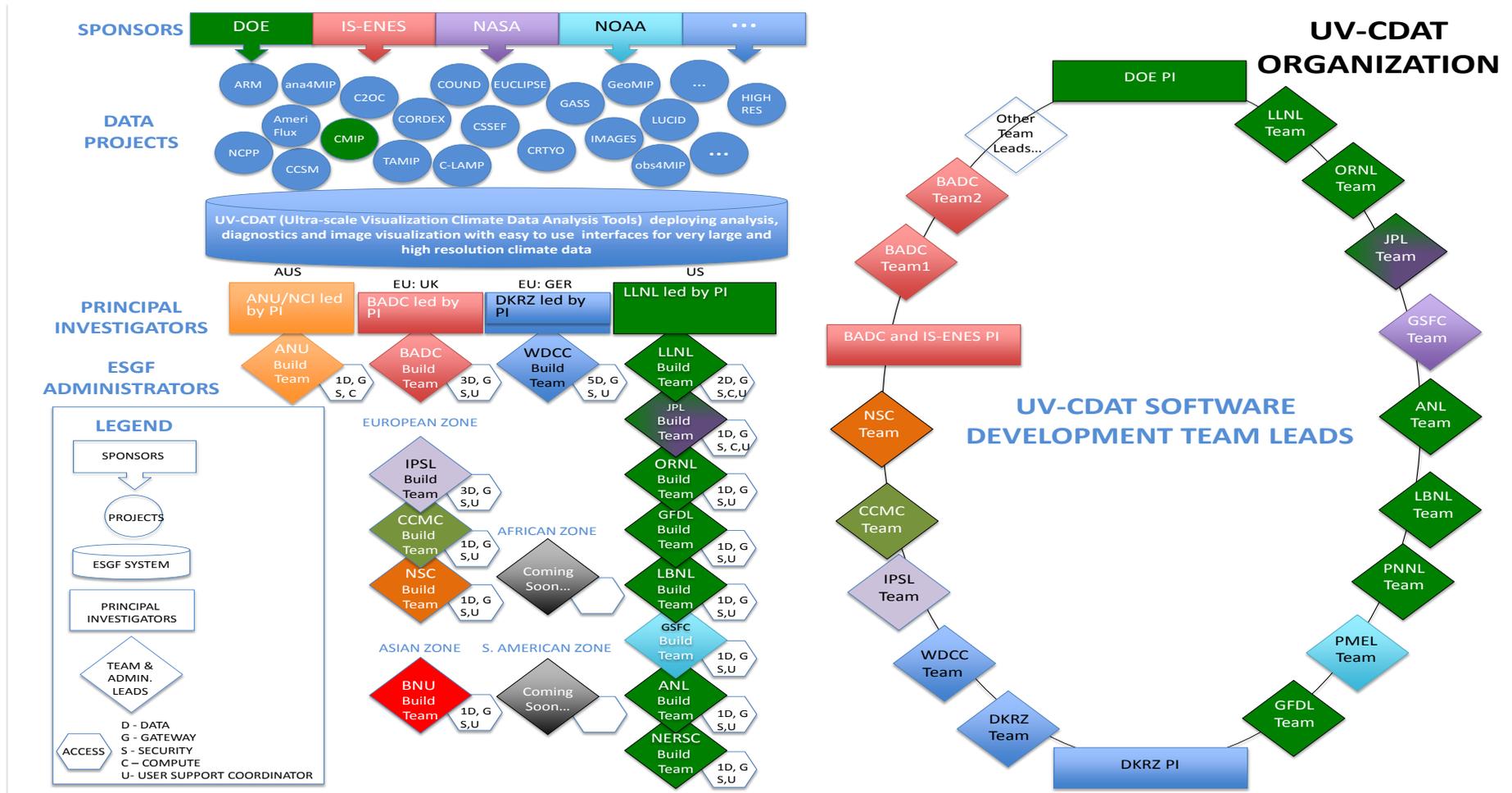


Figure 5.3 Visualisation of UV-CDAT’s physical and human resources in the form of a chart in federated ESGF consortium.

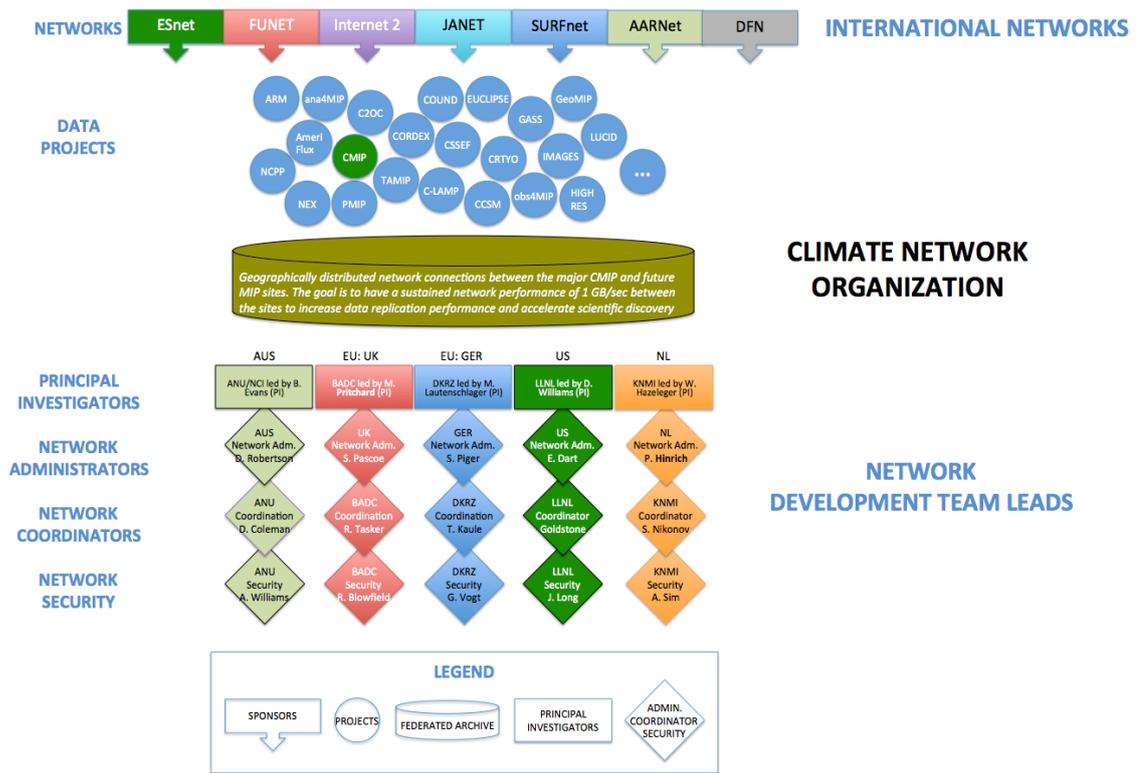


Figure 5.4 Depiction of a new working group called the International Climate Network Working Group (ICNWG), recently started in ESGF.

“We just won the international network award so we’ll be able to make the better connection, hopefully 4 GB/ sec. This network is also led by ESGF. One thing that we found out is that the network speed between the organisations (in ESGF) is far too slow” Interviewee L, ESGF Chair.

From the Figure 5.2 (page 115), one can extract the generic organisational constituents of an ESGF e-Science infrastructure and these constituents can be depicted in Figure 5.5. Other ESGF-like e-Science infrastructures might have a similar composition. In ESGF, the organisational charts for the user support process and other process can be made accessible to the staffs involved in user-support. Interactive charts are also proposed for e-learning and request tracking purposes. Similarly, another view of the ESGF organisational chart for the users is also possible to even track the user queries and learn about the specialization areas of support available at each institute whether technical or scientific specialization.

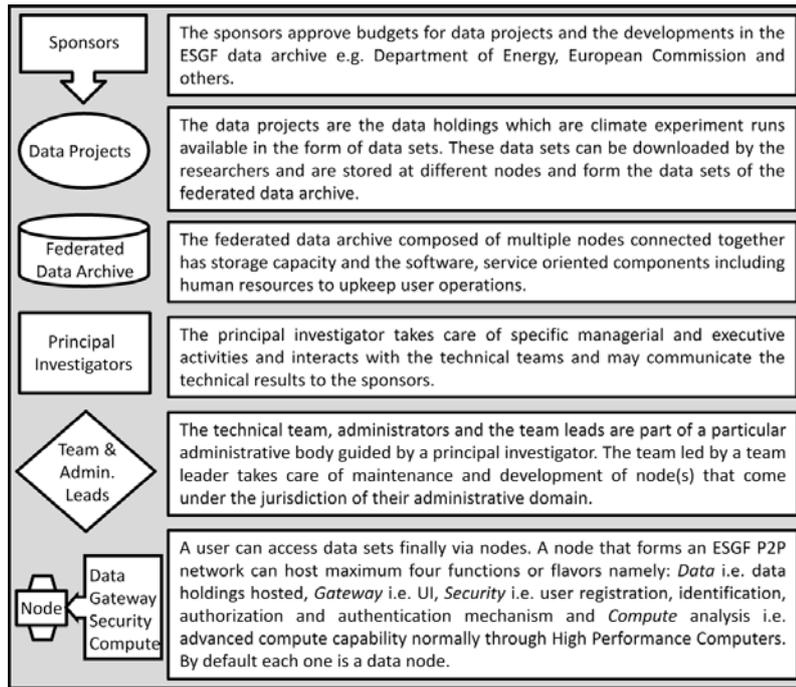


Figure 5.5 The constituents of the ESGF organisational chart highlighted from Figure 5.2, Figure 5.3 and Figure 5.4.

5.4 Challenges faced by ESGF as a diverse networked organisation

Supporting scientific research on climate change needs not only a well-structured and effective system but also a user support that involves non-trivial IT challenges. These challenges create a dynamic, complex and ever-changing environment in which ESGF organisation operates and fulfils the demands of the stakeholders. At the moment, ESGF is facing the following challenges as far as the system's functionality is concerned, tied together with the user support services offered:

- **Multiple data sources:** There are heterogeneous observations collected via institutes, satellites, vertical profiles, etc., including reanalysis data. They are all needed to be stored, updated and standardized in the ESGF data archive. Consequently, particular information may be needed by a user about data sources in ESGF; in that case a user may need to consult a specialist via ESGF support system. At the moment the information about data sources is not covered in the user support system of ESGF.
- **Different dataset versions and data formats:** Data of geospatial and temporal coverage as well as metadata content are in different formats. In ESGF, even though there is a consensus on standards on metadata and data formats, but at times, there is a problem in some nodes (administered under a particular administrative body). The metadata and data formats are sometimes not available at all, or simply not updated; unless detected by a user, who accesses a

particular metadata or a data set. From the analysis of user requests in the past years, this anomaly is usually reported by a user to the ESGF staffs. After receiving this information, the concerned ESGF staff members (in this case especially a data manager) re-checks it and updates the metadata or the data set and makes them available in case they are not available. Similarly, from time to time some data sets are not available in a particular required data format as required by the user, in that case again the user indicates this to ESGF data manager. After getting the feedback from the user, the concerned data manager performs the appropriate action of converting data format or otherwise informs the user that it is not possible to convert.

- **Ever increasing data volumes:** The current data-archives are in peta-bytes (PB), but for example: CMIP5 data project is expected to increase into exa-bytes (EB) in the next 5 years (Taylor et al., 2012). The next generation satellites will collect many terra-bytes (TB) per day. So, there is an ever increasing amount of data that is needed to be stored. Therefore, the data hardware, storage technology and software are always subject to upgrade. This upgrading can cause surge in user requests due to maintenance, thus, causing downtimes.
 - **Data storage and management:** Data are stored and managed at many sites around the world, as there are many administrative sites that are dealing with data and publishing it (see Figure 5.2 and Figure 5.3). However, sometimes there is a problem with the update of dataset versions (including their formats). Even if the data are published by the data managers and quality is assured, yet there are instances as per user experience that the data set may be missing or contain other anomalies, leading to a user to contact user support. Collaborations between a few of remote sites of federation of ESGF such as data centres, data archives, and potential data transfers between sites, are shown in Figure 5.6. For example, the US DOE/LLNL portal (at the top of Figure 5.6) harvests IPCC/CMIP5 data from 10 countries. That is, the original data resides at the data centres, but subsets of the data are replicated at LLNL for backup, better access and use. The DOE/LLNL portal URL is <http://pcmdi9.llnl.gov> (Dean N. Williams et al., 2013).
-

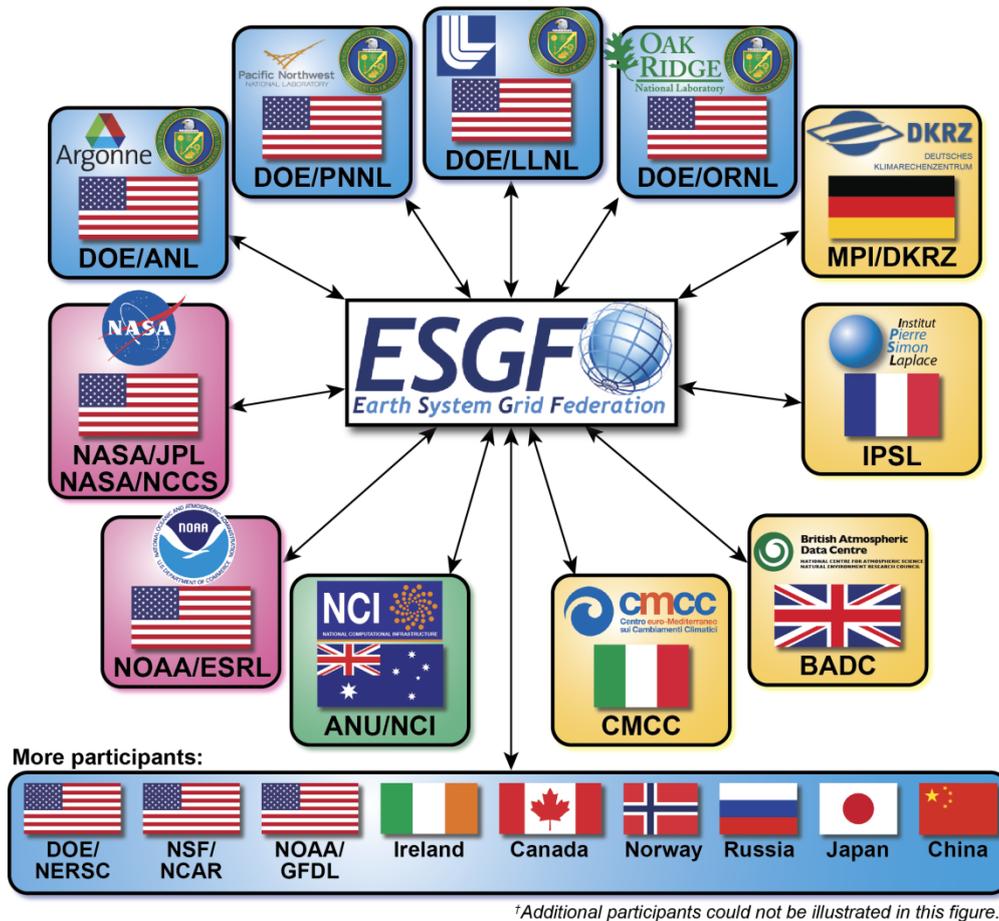


Figure 5.6 Federation of ESGF showing collaborations between a few of its remote data centres, data archives, and potential data transfers between sites taken from (Dean N. Williams et al., 2013).

- Serving multiple communities:** ESGF is made to serve multiple categories of scientists, politicians, educators in public or private sector. Each community has its specific research concerns and needs for specific support. If the scope of ESGF widens to serve other domains as already anticipated by ESGF consortium such as biology, chemistry, astronomy and others, which is already expected by the ESGF management. Therefore, there will be other communities joining ESGF, thus, adding a diverse number of users. As a consequence, the nature of operational service, data access service and user support of ESGF depends on the type of user communities that are interacting with the ESGF system.
- Data analyses at multiple scales:** In ESGF there is diversity of data sets with different attributes, originating from diverse experiments from various sources and geographically distributed institutes. Data must be analysed at multiple scales; global, regional and local. Data analyses and re-analyses are made available to the ESGF archive so that a user can access them. There are different

scientists involved in these steps from multiple institutes. They may be needed to be contacted by the users in case of concerned queries.

- **No complete administrative control of a single organisation:** In ESGF, there is less total administrative control of a single organisation on the ESGF data archives. Distributed operations and services of ESGF create complexity and chaos in detection of the right components that were responsible for malfunction in the whole system as well as finding the right person i.e. staff to diagnose it and correct it. If there is a malfunction in the functionality of ESGF, the correct component under the jurisdiction of a particular administrative body needs to be fixed. Therefore, occasionally there is a bigger time-factor involved in solving complicated problems due to geographic distribution of components. This can cause delay in providing user support services. In operating user support, at multiple sites by the support staffs (first level or second level staff), currently there is little collaboration in handling user queries between e-Science infrastructure staffs of collaborating sites. Therefore, most of ESGF staffs solve the queries individually without the shared opinion of co-workers. Collaborative tools are missing at the moment in the current user support process. The use of collaborative tools may ease administrative control.
 - **Downtimes, outages of multiple centres, software version changes and hardware maintenance:** At times the users may not get required access to data and metadata from the ESGF nodes due to outages, downtimes and deterioration in network performance that often occur at multiple centres because of software version changes and hardware maintenance. Therefore, user queries may dramatically increase within a short time span. Tackling these challenges may decrease user queries.
 - **Voluntary work by the ESGF staffs:** There is a substantial voluntary work undertaken by the data node administrators, who are not explicitly funded to support the system and users. Handling incoming user requests and preparing user support documents are the activities which are not explicitly paid and are most of the times not part of their core job description. Therefore, the amount of time invested by ESGF staffs, especially to support a user, varies dramatically from staff to staff, as it is evident from the survey-questionnaire (details are mentioned in the next chapter). Some of the staffs support users by dedicating more time than others whereas some other staffs do not bother about supporting users at all. In fact, there is no standard or regulation defined by the ESGF consortium, in aspect of dedicating time to support users. Consequently, the voluntary action of staffs to support users may help users but at the same time it may affect their core tasks that they intend to perform as their primary duty. Moreover, handling petty issues that can be handled for instance by a
-

sophomore, is definitely a waste of time for the experts i.e. top computer- or climate scientists.

- **Technical know-how:** A non-researcher might say that ESGF is only for a user having a technical hindsight as the ESGF data-archive system was first and foremost designed for the climate scientists with adequate technical skills. ESGF does not provide suitable online help for non-researchers or non-technical users. So, an interactive e-learning platform and a better designed interface for non-researchers with standardized guidelines are needed. Additionally, information material and web-site information might ease the accessibility and knowledge of users with less technical skills.
- **Quality control of all the data:** Quality control (QC) and quality assurance of climate data in ESGF is cumbersome and not completely an automatic process, which is prone to errors. Though a three layered quality assurance concept of QC has been implemented in ESGF (see Figure 5.7), there are still many cases where the versions of the data are found not up-to-date at various sites. Consequently, creating an upsurge of user queries.

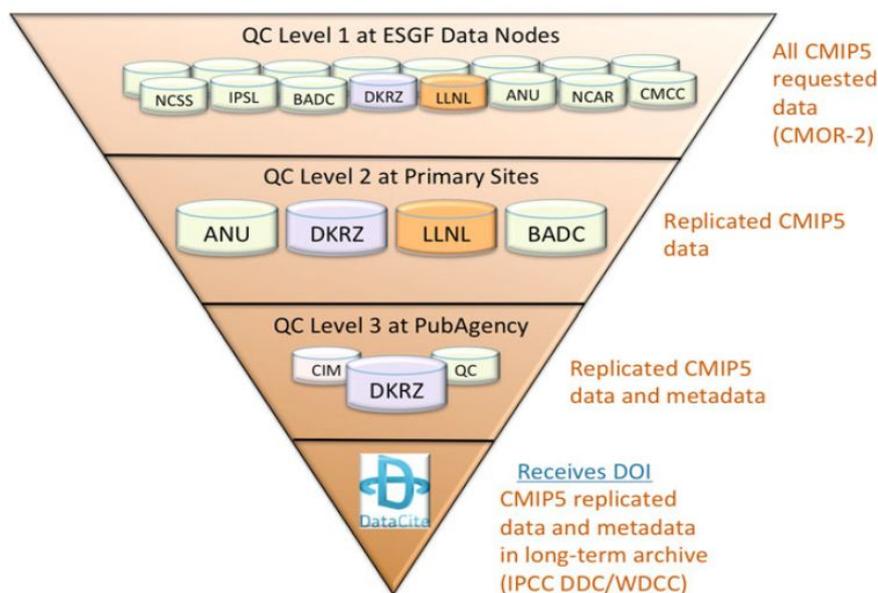


Figure 5.7 Three layered quality assurance concept of ESGF.

- **Multiple user access tools and modes especially for non-researchers:** ESGF is working with international private and non-private institutions and universities to build multiple user access tools and modes that are useful for policy makers and other non-researchers. The primary goal is to create mechanisms to make extremely valuable data available to all, including enabling the commercial use of the data. Facilitating widespread accessibility of these data products can hugely augment the importance of ESGF data holdings. As a result, the broad

community of researchers and non-researchers will be able to efficiently access the most popular data products and analysis for planning and better decision-making (Earth System Grid Federation, 2010). During this process the non-researcher users and other users need strong user support and learning materials and consultations, particularly for access tools, especially in case of accessibility problems.

- **Replicate highly valuable resources and data products at multiple sites:** ESGF development team has enabled the capability to replicate highly valuable resources and data products at multiple sites (nodes). Because climate data is so valuable, the WCRP's Working Group on Coupled Modelling is working to fund ESGF administrators, so centres can be held more accountable for outage of their data holdings (Dean N. Williams, 2015). Currently, as per the experience of users of ESGF data archive system; data products and other resources at a replicated site are sometimes not functional or up-to-date thus leading a user to initiate a query to user support. ESGF is augmenting its technology to allow smart caching, dynamic data replication, and publication of data at multiple index sites (or nodes). As a consequence, ESGF is ensuring more confidence in the availability of data and metadata throughout the P2P network. This development is a step to reduce user queries.

ESGF facilitates climate change research by providing a better accessibility, interoperability and service of data archive system to manage all challenges listed above (Dean N Williams, 2012). A reliable system with less outage with scalable architecture is required in the first place to meet the current challenges. The scalability has been achieved in ESGF to some extent but reliability is still an issue. All sub-systems of ESGF have inter-dependency on each which is in line with the principle of SSM. Keeping in view Checkland's theory of SSM, a user support system of ESGF has also an inter-dependency on other systems that constitute ESGF data archive system and its services. For instance; if there are more outages in the nodes of ESGF data archive system or more UI problems encountered, then it is likely that at that particular point more user queries may originate. The software architecture of ESGF P2P system is described in the Section 4.6.

To cope up with the challenges already pointed out, changes may occur in the system architecture over time in technologies like hardware components, software components, data formats, versioning systems etc. Consequently, evolution is so far the constant part of ESGF. Since the parts of infrastructure will be changing due to changes in technology, data sources, community needs, policies of administrative domains, it is essential to foresee these challenges and develop a dynamic and scalable user support system. A user support system that is compatible with the dynamic and complex

behaviour of ESGF e-Science infrastructure based on a well thought user support process.

A user request about a particular data set or ESGF data archive system malfunction, can be archived, once reported, for future reference. This can avoid re-inventing the wheel, so that the solution process may not be repeated again by the ESGF staffs. In the current user support process in ESGF, the user requests (i.e. communication between the users and the support staff,) are not archived by a central authority. Only a few institutions participating in ESGF, archive user requests internally that are unfortunately not visible to all ESGF staffs and users, as per sources (interviews, participant observations and survey-questionnaire of support staff). Even at the moment, some organisations and their members participating in ESGF user support are not aware about internal archival of user requests. User support archive facility is actually a collection of experience of users in their intercourse with the ESGF system which should be published so that the other users and ESGF staff may take advantage of this. In most of cases, user queries point to the anomalies in the system, some information not well understood by the users and the features that users recommend to have in the ESGF. Storing information about these incidents contribute to organisational memory (OM) of ESGF. Organisational memory is the organisational history.

5.5 ESGF development philosophy

In order to overcome the development challenges faced by ESGF, as already discussed in the section above i.e. Section 4.4; the ESGF management and development team adopted the following strategy, as far as philosophy of software development is concerned:

- **Collaborative development:** Development teams formed worldwide across institutions, frequently collaborate with each other to put forward combined software development strategy based on joint effort of infrastructural developers at various institutions participating in ESGF.

“We believe that infrastructure development is not like you go and develop, like you write a specification and you go and code but it is much more an issue of negotiating a specification, negotiating the goals, negotiating the limits and possibilities of development. That’s why it is important to focus on how do people behave during development of infrastructure, what kind of options do they have, when they have to develop something” Interviewee A, Social Scientist.

The collaboration between ESGF development team takes place via e-meetings and via mailing lists²⁹. For this collaborative work to take place, the infrastructure needs a platform and communication channels to support each

²⁹ The most common mailing-list amongst others is esgf-devel@lists.llnl.gov, used specifically.

other in this collaborative work. This collaboration creates dynamics of development practices and dynamics of doing infrastructure.

“The actors who are doing something they have concerns, they have interests. Each of them, of course their motivation is to build the system because it is good for some reason. But with these goals each of the partner or an actor has related his own goals” Interviewee A, Social Scientist.

- **Open Source, Berkley Software Distribution (BSD) license³⁰:** The ESGF software is open-source that attracts many contributors across the world to the ESGF development and improves its quality. Other institutions can also use parts of ESGF software and at the same time contribute to further evolution of the software. This may lead to some specific queries by the users and contributors of the ESGF software. Consequently, providing guidelines to software contributors and users at the software development level comes under the focus of developers’ support, which can be amalgamated into ESGF user support system sooner or later.

“Projects such as IS-ENES are adding on additional features on top of ESGF. ESGF is open source” Interviewee C, ESGF Developer.

- **Adoption of standards for interoperability across systems, technologies and languages:** The development trend in ESGF data archive system is towards achieving interoperability and standardization of the whole system. This is based on collaboration and exchange of communication between infrastructure developers of different institutes that are part of the ESGF organisation. To achieve interoperability in ESGF; standardized protocols and technologies have been borrowed from industry (i.e. open source) as well as from the geoinformatics domain. From the industry, for instance; Representational State Transfer (REST³¹) and OpenID; an open standard that allows users to get authenticated from other cooperating sites, have been borrowed. A user may get an OpenID from an identity provider (IdP) i.e. ESGF’s security set-up of node and use it to get access to other sites (nodes). Security Assertion Markup Language (SAML) and Secure Sockets Layer (SSL³²) are used in ESGF to

³⁰ **BSD licenses** are a family of permissive free software licenses, imposing minimal restrictions on the redistribution of covered software.

³¹ REST is an architectural style consisting of a coordinated set of constraints applied to components, connectors, and data elements, within a distributed system. REST ignores the details of component implementation and protocol syntax in order to focus on the roles of components. (source: Wikipedia)

³² **Secure Sockets Layer (SSL)**, are cryptographic protocols which are designed to provide communication security over the Internet. They use X.509 certificates and hence asymmetric cryptography to assure the counterparty whom they are talking with, and to exchange a symmetric key. This session key is then used to encrypt data flowing between the parties. This allows for data/message confidentiality and message authentication codes for message integrity.

authenticate between identity provider and service provider, which is an industry standard. From geo-informatics domain: NetCDF, CF and an Open source project for a Network Access Data Protocol (OpenDAP) are borrowed, which are used by Earth scientists.

- **Well-defined APIs:** The ESGF consortium has established well defined APIs (search, security, data access and others) along with the documentation of the APIs available at ESGF Wiki. The APIs are a key to extend the ESGF software part of the infrastructure. Developers and sometimes advanced users want to access API documentation and ask specialists about the details which can be supported by an internal or external user support system.
- **Development of modular components:** In ESGF, each software module is independent of each other thus changes in one module seldom affect the other one.
- **Plug and play, swap and replace:** Since ESGF components and sub-components are continuously in the state of evolution; they are designed to be compatible to each other, thus complying with the principles of plug and play, swap or replace.

ESGF's real distinction from other infrastructures is; that it is a concert of many technologies integrated such that the whole is truly greater than the sum of its parts; while enabling next-level climate insight and experimentation. The development philosophy of ESGF infrastructure is modular, open source and service-oriented. The ESGF development philosophy must include the development of dynamic user support setup to support collaborative development for advanced users and developers of ESGF.

“Well we cannot measure it or judge, but to do something, they are in a state of running, of course they do produce errors. You need to put some measures to judge that” Interviewee L, ESGF Chair.

5.6 Chapter summary

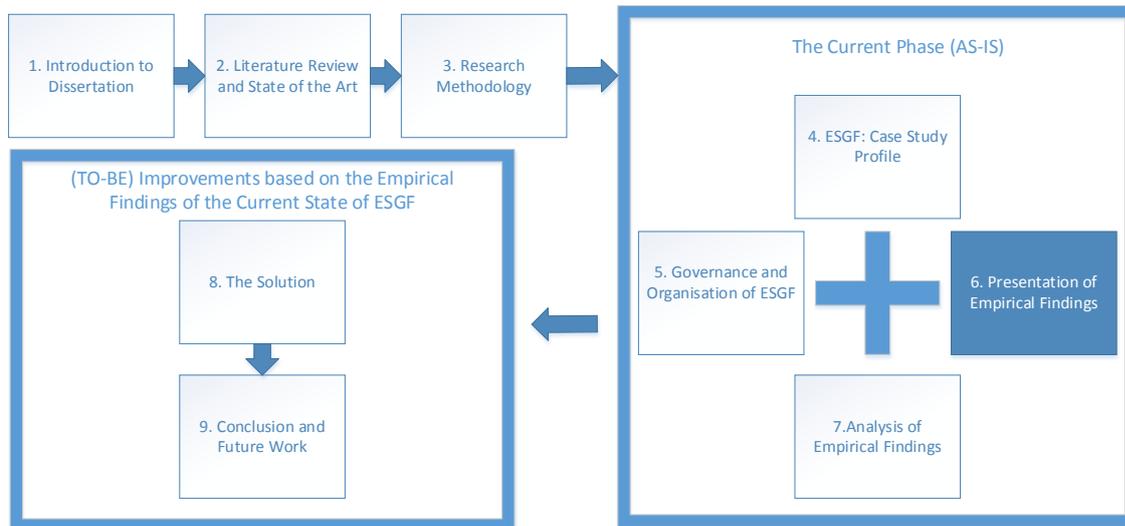
In this chapter, the governance of ESGF infrastructure and organisation was elucidated by the author after finding out the salient features extracted from the relevant documents, observations as well as relevant interviews from the stakeholders. ESGF governance structure comprises of ESGF Review Board (ERB) that closely coordinates with and monitors the three committees: Executive Committee (EC), Steering Committee (SC) and the Technical Committee (TC). The aim of ERB is to protect the interests of users and developers of ESGF by reviewing the code changes that may create a high impact on users and developers (Dean N. Williams & ESGF, 2013).

The main activities of the technical committee (TC) are: Capturing and analysing requirements for the technical development of the infrastructure. Consequently, based on the user requirements TC defines and designs the technical architecture of ESGF. EC is responsible for setting the strategic direction of the ESGF project. It oversees the activities of technical committee and other committees as defined by EC. SC communicates the funding goal to EC, advises on funding issues, represents the sponsors and reviews the overall progress. So far, there is no explicit funding dedicated to user support as well as other user aspects such as user requirements by SC.

The organisation structure of ESGF is visualized in this chapter with all the collaborating federated partners and the nodes that are maintained by the partner institutions along with principal investigators that are leading the development of the e-Science infrastructure. Moreover, the data projects and the sponsors are also visualized in the organisation chart. Similarly, the charts for Ultrascale Visualization – Climate Data Analysis Tools (UV-CDAT) and climate network organisation are also made on the same pattern. The visualisation of the organisational structure of ESGF in the form of a chart is a remarkable milestone in showing the overview of the e-Science infrastructure along with all the stakeholders. Furthermore, these visualization charts of ESGF and UV-CDAT can be used to visualize the support team, especially user support team. Likewise, the user requests can be visualized and tracked with the help of these visualization charts.

ESGF as a diverse peer-to-peer networked organisation faces a number of challenges. These challenges include different data sets and data formats available from diverse resources that are scattered around different geographic locations. Data are stored and managed at many sites around the world, as there are many administrative sites that are dealing with data and publishing it. Moreover, the ever increasing volume of data creates another challenge. ESGF is made to serve multiple categories of scientists, politicians, educators in public and private sector. Each community has its specific research concerns and needs for specific support. At times the users may not get required access to data and metadata from the ESGF nodes due to outages, downtimes and deterioration in network performance that often occur at multiple centres because of software version changes and hardware maintenance. These all challenges faced by ESGF create a need for support especially user support and governance of the user-oriented services.

6 Presentation of empirical findings – (As Is)



From this chapter the following peer-reviewed papers were published:

- Hashim Iqbal Chunpir, Thomas Ludwig, and Amgad Badewi: *Using Soft Systems Methodology (SSM) in Understanding Current User Support Scenario in the Climate Science Domain of Cyber-Infrastructures*. *HCI (10) 2014: 495-506*
- Hashim Iqbal Chunpir, Amgad Badewi, and Thomas Ludwig: *Using Support System in Complex Environment*. *HCI (11) 2014: 392-402*
- Hashim Iqbal Chunpir, Thomas Ludwig, and Amgad Badewi: *A Snap-Shot of User Support Services in ESGF: Advances in the Human Side of Service Engineering. Volume 1, 2014*

These papers provide an over-view of the user support process of ESGF.

6.1 Introduction

The contents of this chapter and the significance of the study results are given in the following sub-sections:

6.1.1 Contents of this chapter

This chapter illustrates the outcomes of the qualitative empirical study that was conducted to capture the current snap-shot of the user support process in climate e-Science infrastructures, especially ESGF. Different data sources were used to collect the relevant information about the different perspectives as well as perceptions of the stakeholders of the current user support process and the ingredients of the process. The analysis and interpretation of the data collected from different sources is not part of this chapter and is covered in the next chapter i.e. Chapter 7. The main aim of this chapter is

to communicate results with the help of graphics, tables, charts, text and explain them to the stakeholders and all interested parties. The results show the main constituents of the current user support process in ESGF. After the investigation of the current user support process, the problems faced by the stakeholders were identified, as the stakeholders interact with it in the user support facility of ESGF. Criticism of the current user support process is given at the end of this chapter in Section 6.11 (page. 181).

This empirical (qualitative and quantitative) investigation revealed a number of issues where attention of the ESGF executive team is needed, in order to improve the existing user-support process in climate e-Science infrastructure projects. The issues about the existing user-support process in climate e-Science infrastructure projects include allocation of human resources, time spent by the human resources to perform user support activities, time to solve the user problems, characteristics of user requests, support tools, support structure and others. The structure of this chapter is summarized visually in the following diagram, see Figure 6.1:

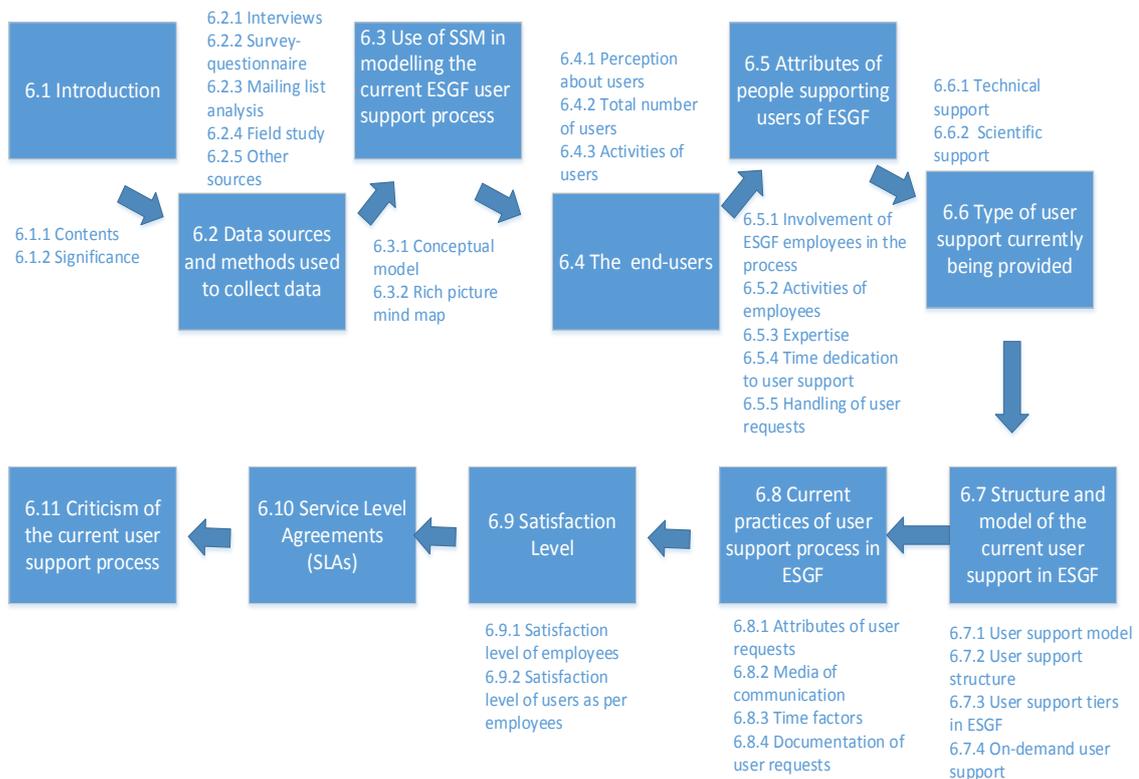


Figure 6.1 The contents of the chapter with main topics and sub-topics covered in the chapter.

6.1.2 Significance of the study and the results

The results collected in this study enable us to understand the current process from different angles, dimensions of subject matter as well perception of the user support practices and process prevailing amongst the stakeholders. For instance, ESGF as a

climate e-Science infrastructure encompasses the elements of management of IT infrastructure, computer science, system administration, data management business processes and climate science. All these elements are integrated into a single whole infrastructure. These elements change with the passage of time depending on the situation, thus making the climate e-Science infrastructure very dynamic. In order to suggest the recommendations to improve the user support process in climate e-Science infrastructures, it is mandatory to find the current elements of the user support process and understand them well. It is important to note that in the user support process, the users, who are the eventual beneficiaries and the main audience of the system, have not been involved in the design of the process.

6.2 Data sources and methods used to collect data

This section illustrates the details about the data sources that are used to extract the current user support practices in ESGF and associated cyber-infrastructures. The snapshot of the current user support practices in e-Science infrastructures was captured in this study via different sources to facilitate data triangulation. These sources are: Survey-questionnaire, interviews with the ESGF stakeholders, analysis of user and employee communication via support mailing-list and participatory observations.

6.2.1 Interviews

For this study, 12 semi-structured narrative interviews were conducted, in-person and via teleconferencing tools. These interviews were conducted with different stakeholders of the ESGF e-Science infrastructure including users, software developers, system administrators, principal investigators, a data scientist and a social scientist to get different perspectives on the current user support process. Interviewees from Germany, USA, Australia and Spain participated in this study. The interview guidelines were divided into three parts: Introduction, entry questions, story like narration of situation guided by directional questions and reflection. The interviews were digitally recorded, fully transcribed and analysed with regard to categories used in guidelines as according to methods of qualitative content analysis with application of deductive categories.

Table 6.1 The table demonstrates the interviewees, their roles, their location of job and the mode of the conducted interview.

Interviewee	Interviewee's Attributes		
	Role	Location	Mode
A	Social Scientist ^a	Germany	In person
B	Climate Scientist, contact of ESGF to Data Modelling institutes, Member of Executive Committee of ESGF	USA	Online
C	Developer, member of technical committee of ESGF	USA	In person

Interviewee	Interviewee's Attributes		
	<i>Role</i>	<i>Location</i>	<i>Mode</i>
D	System administrator, member of technical Committee of ESGF	USA	Online
E	Principal investigator, member of executive committee	Australia	Online
F	Master user ^b at COLA (institute)	USA	Online
G	Developer of C3Grid web services and web portal	Germany	In person
H	Climate research scientist at the Spanish institute / User of ESGF	Spain	Online
I	Climate research scientist at MPI-M / User of ESGF	Germany	In person
J	Climate research scientist at MPI-M / User of ESGF	Germany	In person
K	Climate research scientist at MPI-M / User of ESGF	Germany	In person
L	Chair of executive committee of ESGF	USA	Online

^a. Social Scientist was observing ESGF and related projects for three years.

^b. Master user is the one who collects data-sets from ESGF at a particular research institute so that the researchers can access the data-sets locally.

The interviewees were stakeholders of ESGF comprising mainly: Climate scientists; i.e. the researchers who are the users of the ESGF e-Science infrastructure, technical staff members and executive staff members of the ESGF e-Science infrastructure and its associated projects (as shown in Table 6.1). The composition of the sample of the interviewees allows us to gain an insight into the matter with different perspectives. The interviewees were a rich source of information as it allowed capturing different perspectives and perceptions of the current user support process in ESGF. The employees of the ESGF who participated in the interviews were experienced. In addition, they had an overview of different situations in the current user support process as well as the over-all management of ESGF. Likewise, the interviewees who were users of ESGF had experience in conducting their climate research activities, using ESGF and other e-Science infrastructures. They had been using ESGF for a minimum of 6 months. Furthermore, the users knew well about the usability part of the ESGF. The subsequent sections of this chapter present the brief outcomes of the interviews.

6.2.2 Survey-questionnaire

The goal of the survey-questionnaire was to investigate the user-support process currently in practice in the main projects of e-Science infrastructures related to climate science, especially from the employees' standpoint. There are different methods available to conduct survey for instance: by post, telephone and others. However, online

questionnaire was the method chosen for this survey because it is quicker, automated and supports complete anonymity of participants (Lazar et al., 2010).

As this research focusses on improvement of the user support process in climate e-Science infrastructures, therefore conducting survey with the staffs of climate e-Science infrastructures is a vital step. The reason of conducting a survey with the staffs of e-Science infrastructures as the “target population” (Couper, 2000), is because these staffs though maintaining and further developing e-Science infrastructure are also partly involved in supporting users. Therefore, these staffs can be partially referred to as user support staffs. However, the level of involvement of the user support staffs varies drastically. Some of the staffs are very much involved in supporting users and others are not involved at all. The questionnaire was divided into five different sections to get information about the current user support process including its actors i.e. users and supporting staffs. The five sections included the following: Structure of support, time related factors in user-support process, communication factors in user-support process, user requests related factors in user-support process, users related factors in user-support process and collecting data relating to the *sample description* thus taking basic social biographic data of the staffs with their self-rating about their expertise.

The survey-questionnaire was conducted from March to May 2013 with 26 respondents from different part of the world. The total questions asked in the survey-questionnaire were 45 (41 closed ended questions, 3 open ended questions and 1 question was about seeking the consent of the respondent to participate in the study). Some questions of the similar nature were repeated with different question description in the survey questionnaire to test the validity of responses.

6.2.3 Analysis of mails via the mailing list

E-mail, though not popularly a data source in traditional research methods, but has been recognized as a valid evidence in personal communications (Gupta, Mazumdar, & Rao, 2004). In the ESGF user support process the major medium of correspondence between the users and the ESGF staffs as well as within ESGF staffs is e-mail based. Consequently, in order to improve the user support process, it is imperative to use e-mail as a data source. The importance of e-mails coming to the ESGF user mailing list: “esgf-user@lists.llnl.gov” has been realized by the author and therefore analysis of this mailing list was done. The requests sent by the users were recorded from December 2013 until September 2014. Within this period more than 600 e-mail cases i.e. the user requests and responses of the users have been analysed. An e-mail case is initiated, when a user sends a request and the user support staff replies and the communication continues until the conversation is not followed further or abandoned. The total cases of e-mails which were 600 comprised analysis of more than 1800 e-mails.

6.2.4 Field study

A field study was conducted in May 2013. 16 students of M.Sc. in HCI from the HCI department at the University of Hamburg took part in this study. The students were divided into 4 groups of 4 students each and on 30th May observed the staffs of German Climate Computing Centre (DKRZ), which is a partner institution of the ESGF consortium in Germany. The staffs were observed by each group of four students in their activity of handling the incoming user requests and interaction with the request tracking software up to three hours. The observations were then recorded. The results i.e. the observations were then communicated to the audience at the university in the form of presentations. The field study approach enabled the author to add some new insights about the ESGF user support process in this study.

6.2.5 Other sources

Other sources of information such as documents, participant observation adds to the observation taken from sources mentioned in the previous section. More than 300 relevant documents such as meeting protocols, reports, websites and articles have been reviewed by the author to extract the information about the current user support process in ESGF and its associated e-Science infrastructure projects. Additionally, the author has been working on the C3Grid project which is an associated project of the ESGF initiative. Thus, C3Grid interface was used to connect to the ESGF data nodes. This provided an opportunity to the author to observe the developments in ESGF particularly in the user support process in the form of participant observations while interacting with the staffs and users of the ESGF.

6.3 Use of soft systems thinking methodology in modelling the current ESGF user support process

ESGF is a decentralized, complex and dynamic system. The knowledge flow within the ESGF partner organisations is not transparent. For transferring the knowledge between the partner organisations and the individuals, the best practices are needed to be defined. Before the good practices can be defined, the current user support process in ESGF and its associated e-Science infrastructures are needed to be analysed in such a way that all stakeholders can be able to express their views and perceptions, in some common language which is understood by all stakeholders. Soft systems methodology (SSM) enables the stakeholders to communicate to each other by establishing a platform for common knowledge (P Checkland & Winter, 2006).

The dynamics and complexity of ESGF operations influences the user support process, thus making it a dynamic process as well. A user request from any part of the world can emanate and is handled by a person in a participating institute of the ESGF federation. In order to save time, supporting efforts of human resources i.e. ESGF e-Science

infrastructure staffs and to ensure user satisfaction, it is pivotal to investigate user support process using SSM, so that the process may be made efficient in the near future.

According to SSM; there is a *transformation process* in each conceptual system having a purpose, where an input is transformed into an output. The transformation depicted as “T” is a *Weltanschauung*, a German word equivalent to *worldview* in English. *Weltanschauung* “W” is a very dominant concept in SSM that determines the belief or point of view that makes transformation “T” rational. W and T form the core of a mnemonic CATWOE³³. CATWOE analysis in SSM is used to create a root definition which is the third stage in Checkland’s seven stage of SSM investigative process which has come to be known as “mode 1” SSM. ESGF user support system is a system that has a purpose (or purposes), it exists for a reason and achieves some change, or 'transformation'. ESGF promotes user problem solving; in the long run it educates users and promotes learning about its sub-systems. It 'transforms' unresolved user problems into solutions.

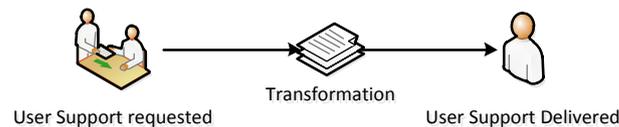


Figure 6.2 The transformation process in ESGF user support.

Using SSM mode 1, CATWOE analysis in the context of ESGF user support system can be stated as in the following Table 6.2:

Table 6.2 CATWOE analysis of user support system in ESGF e-Science infrastructure.

Mnemonic	Description
C= Customers	users of ESGF system (victims or beneficiaries)
A= Actors	ESGF staff (Support staff and developers)
T=Transformation process	Transformation of user incidents into solutions monitored by E1 ³⁴ E2 ³⁵ E3 ³⁶
W=“Weltanschauung” or Worlds perspective	the belief that providing user support will benefit users in their research activities and interaction with the ESGF system monitored by E1 E2 E3

³³ In CATWOE, C stands for *customers*, A for *actors*, T for *transformation process*, W for *worldview*, O for *owners* and E for *environmental constraints*

³⁴ Efficacy: Does the system work? Is transformation achieved?

³⁵ Efficiency: A comparison of value of output versus value of input- is the system worthwhile?

³⁶ Effectiveness: Does the system achieve its longer term goals?

O=Owners	All stakeholders of ESGF
E=Environmental Constraints	Geographically distributed environment with components under control of different authorizes operated by different human resources

E1, E2, E3 can be defined in terms of ESGF user support system as:

E1: Are user support requests answered properly? E2: How many user support requests are answered keeping what standard and how many resources consumed? E3: Do users find user requests solved by employees and UI for self-help a useful way of reaching the research goals of users and interacting with cyber-infrastructures?

An ESGF user support system, which is part of the ESGF system in the wider context can be defined (by the author) in the form of SSM root definition as: *An ESGF user support system owned by ESGF-stakeholders (investors), operated by ESGF staffs (partly staff from node administrative bodies), to support users of ESGF by fulfilling their information needs in order to get information to achieve their research-oriented goals while constrained by ESGF financial, technology, human resources, cultural norms, geographic administrative and general policies.*

6.3.1 Drawing the conceptual model of the current user support system in ESGF e-Science infrastructure

From the root definition, a conceptual model of the current user support system is derived in Figure 6.3. In Figure 6.3, the user inputs to the system are the incoming incidents to the user support system. A user interacts with the system via a GUI or command line interface (CLI) of an ESGF system. Now in case of a user encountering a problem with the ESGF e-Science infrastructure or an information need; a user sends a request via some communication channel to ESGF user support, as shown in the Figure 6.3. The (ESGF) employee services the user-request, thus making a complete user support cycle. It is important to note that E1, E2 and E3 are not included in the diagram because E1, E2, E3 and control actions are not part of the current user support system in ESGF. All activities of all sub-systems in the user support system in ESGF cyber-infrastructure work in complete synergy.

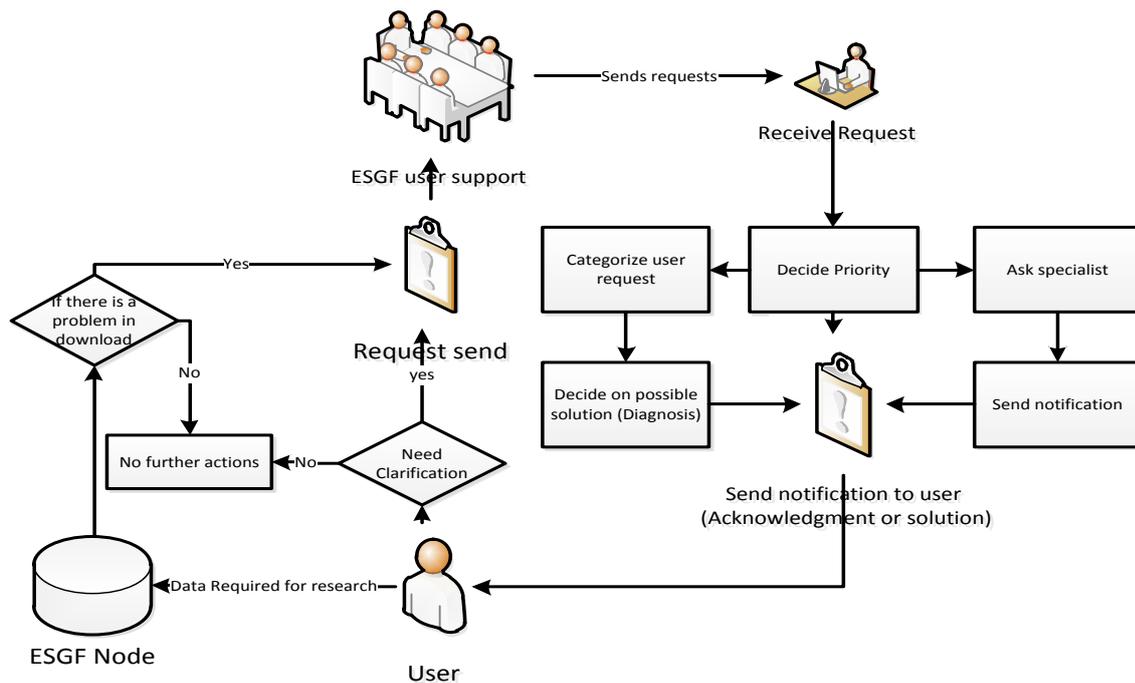


Figure 6.3 The conceptual model of the user support system in ESGF e-Science infrastructure.

ESGF user support exists as part of a wider system or systems. The boundary; which defines what is, and what is not part of the system of user support system is subject to change as the nodes jump in and out of the whole scenario in the ESGF peer-to-peer network. The user support system may be extended, thus servicing other similar e-Science infrastructure systems e.g. C3Grid climate community.

6.3.2 Presenting the findings of current user support system in ESGF e-Science infrastructure in the form of a rich picture mind map

The author being an analyst found the current situation of user support problematic and expressed the situation in the form of what is called a *rich picture mind map*. Following Checkland's guidelines, the current user-support situation in climate e-Science infrastructure projects, especially ESGF and its associated projects, is expressed in the form of *rich picture mind map*, as shown in Figure 6.4 (P Checkland & Winter, 2006). Pictures can provide an excellent way of sorting out and prioritizing complex problem areas and therefore are used in SSM. In a traditional SSM approach, rich pictures are normally hand drawn and they describe elements of structure, process, issues, concerns or developments (Peter Checkland, 1985). There are no rules used in SSM rich pictures approach though matchstick people and bubbles coming out of people are common.

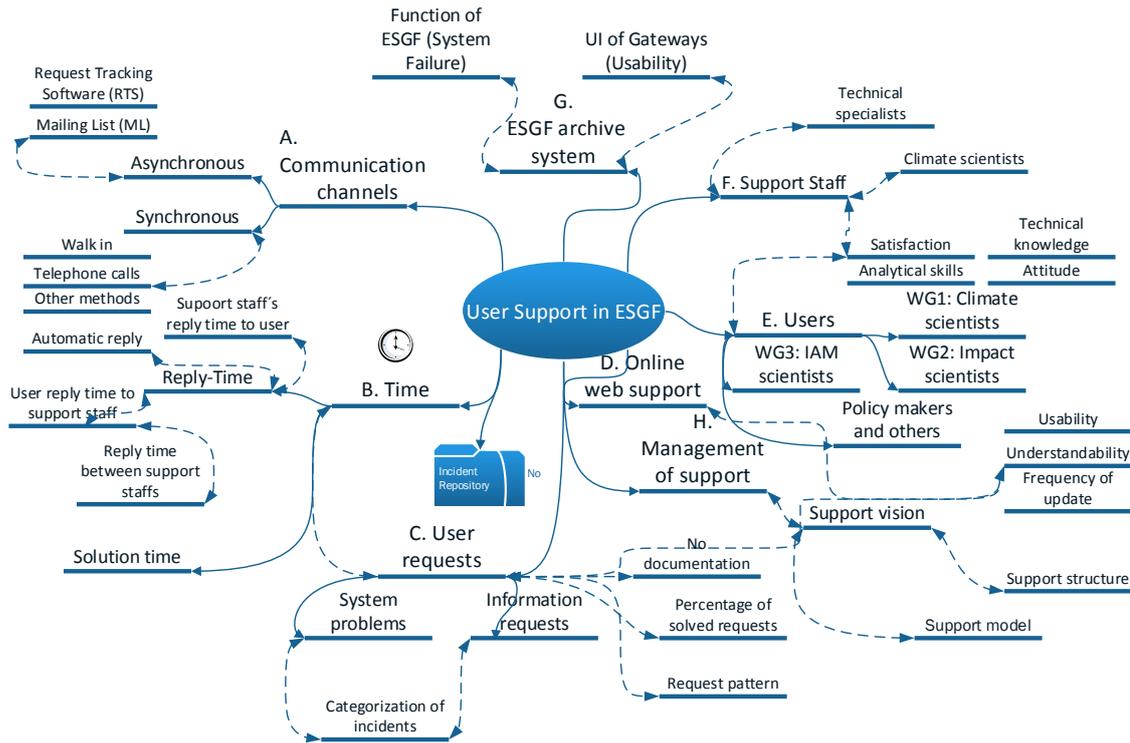


Figure 6.4 The figure shows the user support process in climate e-Science infrastructures in the form of rich picture mind map.

In Figure 6.4, the current user support system scenario is described using thick lines that describe a particular concept associated to the user support system within ESGF. The lined arrows originating from the main oval shaped system describe that the system contains the specified elements and is dependent on these various concepts depicted via a thick line and a textual description on top of them. The dashed-lined arrows depict that a particular concept might have different attributes depicted by these lines.

6.3.2.1 Communication channels

For example, if we look at the Figure 6.4, user support system in ESGF e-Science infrastructure has communication channels (labelled as “A” on top left of Figure 6.4) via which users and staffs communicate in case of a problem. These communication channels are divided into asynchronous and synchronous type of communication channels, currently present in user support process in ESGF e-Science infrastructure, see Figure 6.5 (also shown in Figure 6.4, top left). The asynchronous communication channels in user support of ESGF are mailings lists (ML) and request tracking software (RT), highlighted in Figure 6.5 top left. Most of the communication between users and climate e-Science infrastructure employees is via e-mail (through ML and RT). This result is in accordance to the distributed model and global nature of e-Science infrastructure.

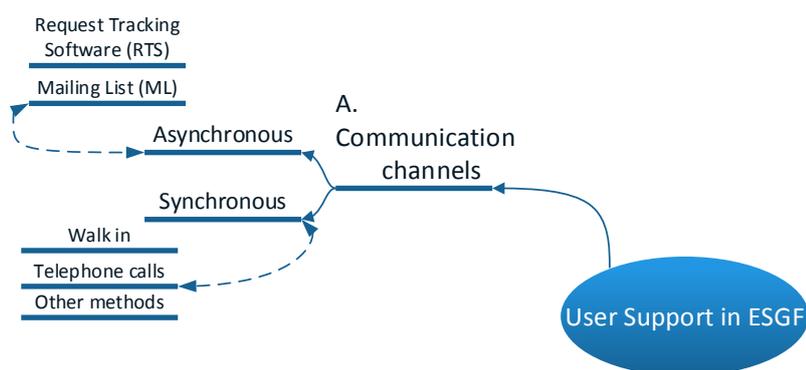


Figure 6.5 Asynchronous and synchronous type of communication channels in the user support process of ESGF.

There are multiple channels, which are used for user and employee communication (shown at the left hand side of the Figure 6.5). These channels are mainly e-mails, telephone, walk-in and skype. The key channel of user and employee communication is e-mail, as according to the survey-questionnaire; 78% of the incoming user requests are received as e-mails. The e-mail communication can be further separated into e-mail delivered via Request Tracking System (RTS) and via mailing-lists (ML). E-mails delivered to employees via Request Tracking System (RTS) are 43% of the total incoming user requests, which is comparatively higher than the e-mail delivered via mailing-lists, which were recorded 35%. Other communication channels like telephone, walk-in and skype show an extremely lower trend of usage, which is recorded 7%, 5% and 4%, respectively. This implies that there is a huge difference between incidents received via e-mails versus other modes of communication.

6.3.2.2 Time issues

Time is an important issue in user support in e-Science infrastructures. Staff's response time and solution time to a user request are important elements of an effective user support (see Figure 6.6, labelled as "B"). Response time is further categorized into *support staff's reply time to a user* shown in Figure 6.6, top left and *reply time between support staffs*. The reply time between support staffs is a response time of a support staff, if a user request is escalated by another support staff, who is the receiver of the user request at first (shown as *reply time between support staffs* in Figure 6.6, bottom left). Finally, user response time to the support staffs response (shown as *user-reply time to support staffs* in Figure 6.6, left) is also a sub-category of reply-time (see Figure 6.6). For the time being, response time and solution time though not systemized with the help of a service level agreement (SLA), works currently well for user support in ESGF e-Science infrastructure.

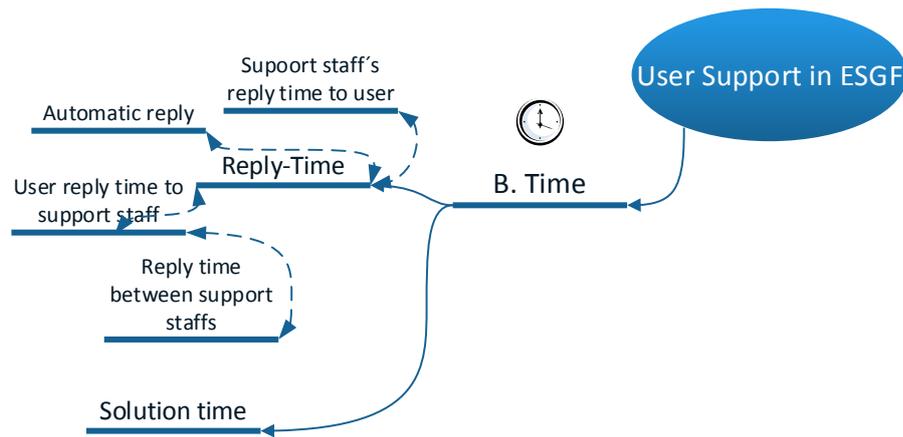


Figure 6.6 The time issues in the user support process.

6.3.2.3 Incoming user requests

The user requests, also known as incoming incidents, can either be due to a problem in the e-Science infrastructure; e.g. outages of nodes, or a user requires information about a particular phenomenon; e.g. how a user shall register etc. (see Figure 6.7, bottom left side, labelled as “C”). There is no user request pattern used currently in the user support process of ESGF (see Figure 6.7, right and bottom side). The user request pattern can be used in future. The advantage of using a user request pattern is to make the incoming user requests uniform. Moreover, using a user request pattern will help in automatically filtering the user requests.

All of these incidents or user requests can be categorized into respective categories (see Figure 6.7, bottom with the bulb symbol). After analysing the results of the survey-questionnaire, these categories can be cited as: *Data access* and *data download problems*, *user authorization*, and *authentication* as well *registration problems*. These are the most common problems encountered in the current user support of ESGF.

Unfortunately, there is no central repository maintained by the current user support system in ESGF (see Figure 6.7, middle), where the information about user requests can be stored and redundant user support enquires can be triggered. Moreover, currently there is no information retrieval system where users or support staffs can search the relevant problem cases. Furthermore, the complete documentation of the user problems are not maintained and shared amongst the partner institutions of ESGF. However, some user support staffs do update the information partly that is useful for staff and users of ESGF system. The percentage of solved requests is around 77% as according to the survey-questionnaire and the mailing-list analysis. Thus, 13% of the incoming requests are abandoned or not pursued by the ESGF staffs.

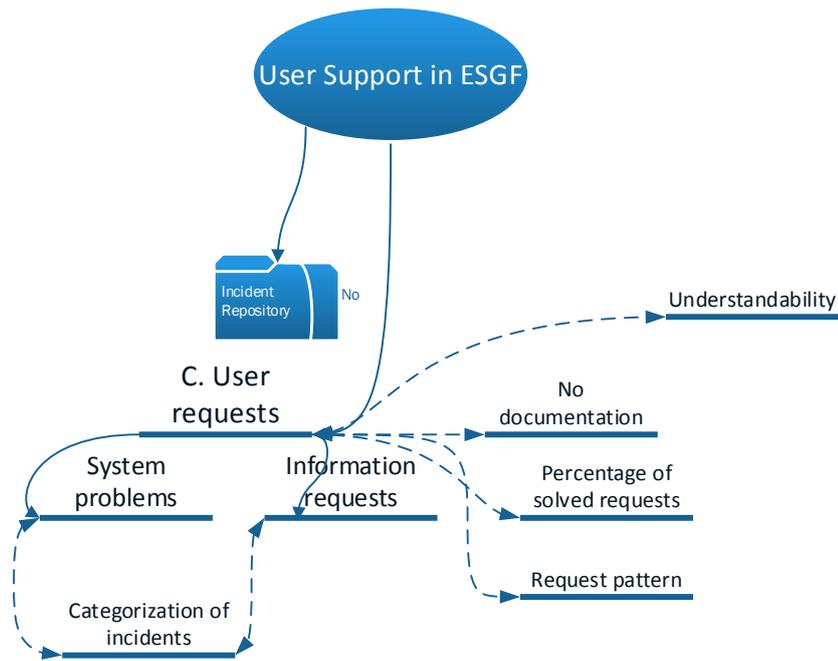


Figure 6.7 The incoming incidents from users and the categorization of user requests.

6.3.2.4 Online support material

There is online support material available on the webpages administered by ESGF. The usability and understandability of the online help resources in ESGF is not determined yet. Experiments with the users of the data projects hosted by ESGF can be designed to gauge the usability and understandability of the online help resources in ESGF. It is however, proposed as a future work. Update of the online help websites is usually done, if there is a new version or release of a software component of an ESGF e-Science infrastructure. However, these updates are not regular. This whole situation is depicted in Figure 6.8 (labelled as “D”).

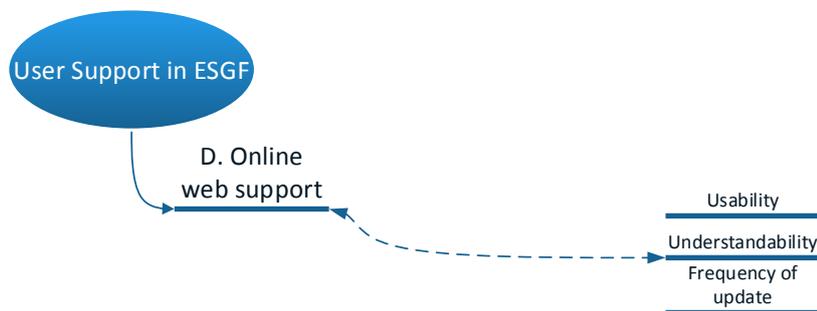


Figure 6.8 The usability, understandability and frequency of update of the online help webpages provided by ESGF.

6.3.2.5 Types of users

The users of ESGF and ESGF-like e-Science infrastructures are divided into four main categories (see Figure 6.9, labelled as “E”), *working group 1*: Advanced core climate scientists, *working group 2*: Impact scientists, *working group 3*: Integrated Assessment Modelling (IAM) scientists and finally *non-climate scientists* such as; policy makers, journalists and anyone who is interested in climate science.

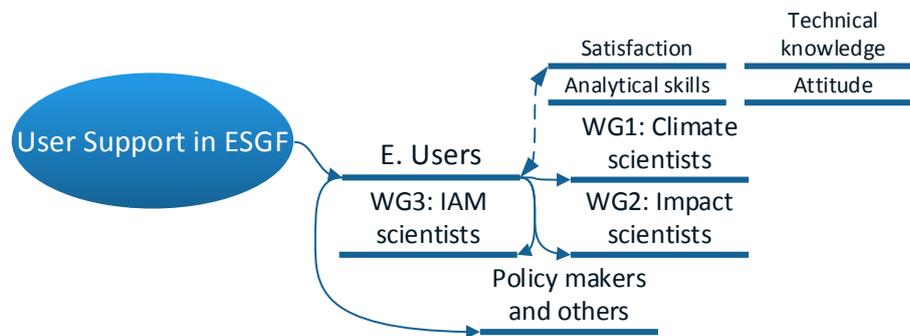


Figure 6.9 The types of users interacting with ESGF.

6.3.2.6 Participation of employees

The employees of ESGF and ESGF-like climate science e-Science infrastructures are technical experts such as computer scientists, network administrators, data curators and climate scientists. The roles amongst the staffs of climate science e-Science infrastructures (ESGF and ESGF-like) are not completely specified. Therefore, there is no formal assigned role of *user support manager* in e-Science infrastructure organization of ESGF. Any climate science e-Science infrastructure employee from any part of the world can jump in and answer a user request and provide solution to user’s problem. Answering a user request or providing a solution to user’s problem is an initiative of an employee (at least in ESGF). There are no explicit long-term support positions financed by the ESGF sponsors.

6.3.2.7 Skills of support staff

User support in ESGF e-Science infrastructure is present to facilitate ESGF users and is operated by its employees. The operation of user support is based on employees’ attitude, knowledge; analytical skills and satisfaction level with the support process that provide support to users (see Figure 6.10, labelled as “F”). Similarly, users in their interaction with the e-Science infrastructure depend on their behaviour of interaction, attitude towards system as well as their knowledge, analytical skills and the level of satisfaction, see Figure 6.10.

According to survey results, the employees who support users are skilful and qualified. Moreover, both users and employees are only satisfied to some extent with the current user support facilities in ESGF and ESGF-like e-Science infrastructures.

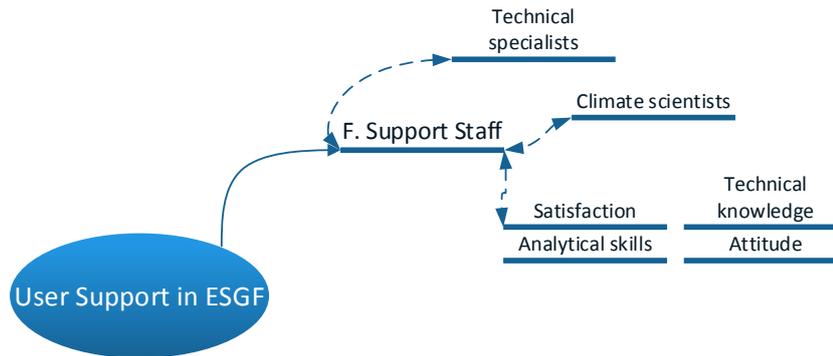


Figure 6.10 The operation of user support in ESGF is based on the following factors: Attitude, satisfaction, analytical skills and technical knowledge of support staffs and users.

6.3.2.8 Dependency on ESGF data archive system

The ESGF data archive system (shown in Figure 6.11, labelled as “G”) and its sub-components such as authorization and registration sub-system, UI of gateways (portals) available to users to browse and access climate data-sets and others, affect the user support process. For instance; if there is any disturbance in a function of a geographically distributed component of the ESGF archive system, then the users experience it (see Figure 6.11, left part). This results in an upsurge of user requests which are entertained by ESGF employees.

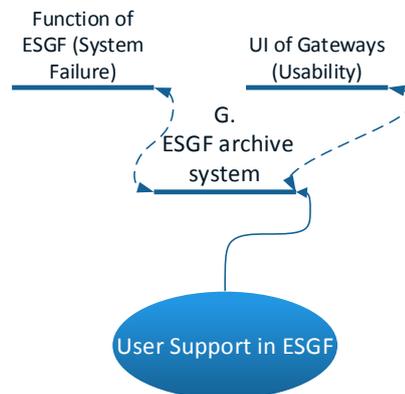


Figure 6.11 The dependency of user support process on the structure, function and UI of ESGF e-Science infrastructure.

6.3.2.9 Administration, structure and model

The administration, structure and model of the user support process in ESGF are depicted in Figure 6.12, labelled as “G”. Unfortunately, in the current ESGF user support process, there is no clear and explicit user support policy as well as vision.

Similarly, there is no clear and explicit user support governance and organisation scheme in ESGF user support process. The user support process in ESGF is mainly based on single point of contact (SPOC) model. However, there are other channels to contact with the user support employees to seek help. There is no split of user support into user support levels such as first level support (FLS) or second level support (SLS). So the structure of the user support process in ESGF has not yet been defined concretely.

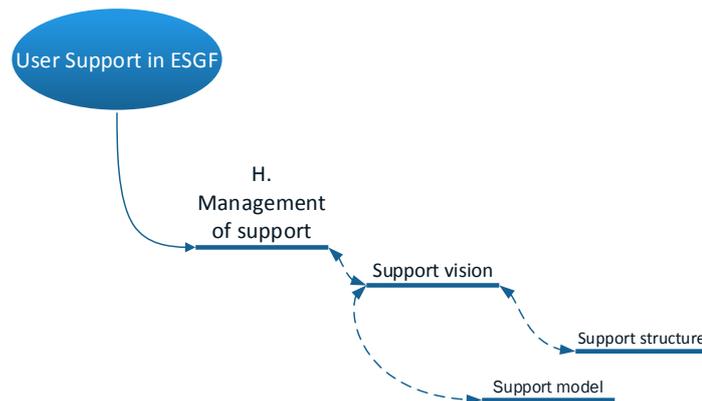


Figure 6.12 The administration, structure and model of ESGF.

6.4 Current perception about the end users

ESGF was initiated as a purely research based project to test the capabilities of grid technologies to connect data archives. At that time, the interaction of users with the system was not present. At present, however, ESGF aims to serve users and provide climate data to them as a main service. Therefore, users of the ESGF are main actors to the ESGF e-Science infrastructure and its user support process. The target audience of ESGF are the beneficiaries of the services provided by this e-Science infrastructure. This section explains the current understanding of the ESGF staffs about the users of ESGF and its associated e-Science projects. It is important to understand the perception of the stakeholders of ESGF about the users of ESGF, in order to understand the behaviour of the users with the system. A social scientist who studied ESGF e-Science infrastructure said in an interview:

“It is interesting to evaluate who the users are, who are currently using the system and the ones who do not use the system. It is important to gain clarity how the potential user might behave. One needs to look at how he works, how he thinks. This is quite a big task” Interviewee A, Social Scientist.

Understanding the behaviour of users with the ESGF system aims to provide input to modelling the system according to the user needs. This phenomenon is similar to the user centred design. User centred design follows a philosophy that addresses user needs.

Individual needs of the users can be addressed from the technology in different ways. Users of ESGF might assume many roles (Reed, 2014). The following sub-sections explain about the role of end users and the total number of users using the ESGF e-Science infrastructure:

6.4.1 Users of the ESGF

The perception amongst the members of ESGF about the users of ESGF is currently manifold. However, it is evident from the interviews as well as survey questionnaire that there is a consensus amongst at least two points that the users of ESGF are climate scientists and non-climate scientists. Once asked to a social scientist, who observed ESGF climate e-Science infrastructure for the last three years that whether the management of ESGF achieved the perception of end-users of the ESGF and its associated e-Science infrastructures, she replied:

“There were some discussions in the beginning about end users, but it seems that the perception (of employees working for ESGF) of what and who an end-user really is, these perceptions are quite different” Interviewee A, Social Scientist.

“For the computer scientists there is a perception that a user is a climate researcher and has particular attributes and needs data access to do research. However, within the climate scientist community there are heterogeneous scientists and it is not easy to know, who does what, because their goals are very different. There are climate modellers, the one working with the climate models, then there are diagnostic people and they work differently. They use different tools, they use different parts of the infrastructure and they have different workflows” Interviewee A, Social Scientist.

Once inquired about the target audience of ESGF, i.e. the end-users, one of the main technical staff involved in technical research and development of ESGF e-Science infrastructure at Lawrence Livermore National Laboratory (LLNL), interviewee C replied:

*“The end-users of ESGF are the target audience of ESGF and are divided into **climate scientists and non-climate scientists**. Climate scientists want to achieve goals for them. They do their science via ESGF. And get their science translated, so that policy makers can see and get the results. These scientists generate IPCC report and they provide it as a translation to policy makers. If one can do it in a dynamic way, then you are creating value, do it faster, you do not need CMIP. The aim is to make for CMIP5 dataset in ESGF. We are going to tune it, get bugs fixed and roll out new features that the audience wants. Make new releases. Hopefully we’ll have a system that works.*

Non-Climate scientists are doing some educating. The nomenclature in scientific community is cryptic. How do you turn it to an understandable thing? That makes it easier. They are probably the policy makers” Interviewee C, ESGF Developer.

Moreover, the interviewee C further explained that there is an apprehension among developers that users are tech savvy and they know how to behave with the ESGF e-Science infrastructure to do science and to get desired results.

“Climate Scientists who use ESGF know what to do. They are extremely educated users. Once they get the data they know what to do. They search, find and get the data. Once they get the data they do not need any help” Interviewee C, ESGF Developer.

However, this perception is not correct for all users as according to observation of the author because many users find it difficult to interact with the GUI of ESGF. Similarly, once asked a system administrator, an ESGF staff located at a different location about who the users are; he replied:

“Looking at e-mails that we get there are big power users in the business since last 10 years. They know how to write script they need massive amount of data for climate analyses etc.

Then we have simple graduate students who need particular data for their research. Then there are people from the private sector, who need to forecast climate for insurance companies” Interviewee D, System Administrator and ESGF Developer.

Looking at the quote of interviewee D, he perceived that there are skilled users, what he calls power users, who know how to write scripts thus tech savvy and probably they have got more experience with the ESGF system as it is their daily routine to download large sets of data. At the same time there are graduate students who need data from ESGF for their research but might require help to get familiar to the system. Apart from the research oriented users there are some users coming from the private sectors as well, for instance; insurance companies, they come under the category of commercial users. It was then asked that whether the commercial users pay for the services, the interviewee replied:

“No they do not pay. It is free. You don’t need to pay for the data but some of the data sites are reserved just for university or non-commercial users. So all the data is available to non-commercial users and only a portion of data is available for commercial users. It depends on modelling groups what do they decide what do they want to do with the data. I mean to make available for commercial purposes or not. In some cases they decided not to” Interviewee D, System Administrator and ESGF Developer.

From the discussion in the form of interviews with the stakeholders of ESGF and the observation it is clear that in the current user support process, the notion of end-users or the main audience of the user was not clear. The notion of end-users and their types is needed to be clear amongst all the stakeholders of ESGF and its associated projects.

6.4.2 Number of users using ESGF worldwide

There is mainly a perception amongst most of the ESGF stakeholders that the total number of users of ESGF ranges from 25,000 to 30,000 as according to interviews. In ESGF literature 27,000 users have been cited (Vu, 2013; Dean N Williams, 2012). However, due to lack of collection of statistics and metrics by the ESGF management, one cannot be completely sure about the number of users. In an interview, once asked about the total number of users of the ESGF e-Science infrastructure, an interviewee stated:

“Globally we have tens of thousands of users may be 20 to 30, 000. People who are really engaged in this community are may be few thousand who are active in the past let’s say 12 months” Interviewee D, System Administrator and ESGF Developer.

Moreover, the interviewee said that every day there are more or less 100 users from different parts of the world on the average who download climate data sets from ESGF e-Science infrastructure.

“We do not have metrics but every day there are almost up to 100 users who are downloading data throughout the world every day” Interviewee D, System Administrator and ESGF Developer.

Similarly, the contact point of ESGF to the climate data modelling projects that are served by the ESGF e-Science system claimed that he does not maintain the list of users, therefore the concrete number might not be known:

“We don’t have a single list of users” Interviewee L, ESGF Chair.

However, from the total number of users who download the data from ESGF at different intervals in time, some of these users may become dormant at some point. Thus, it is said that the active users pulling the data from ESGF may range from few hundreds to few thousands, depending on the time and situation. For instance, one interviewee said:

“There are total users of almost 3000 scientists who are actively pulling data from ESGF. CMIP5 is the main product offered in ESGF” Interviewee C, ESGF Developer.

There is a tendency that the people who are once active may become inactive at another point:

“Some are active and inactive e.g. researchers who are doing PhD they get the data and then they become inactive” Interviewee D, System Administrator and ESGF Developer.

It is, however, important to track the record of users and number of requests asked per user. From recording and tracking records, one can find the areas of climate science, the users are interested in. At the same time, it is important to track hot categories of

questions that they ask can be identified. The ratio of users and the user requests can be identified as well.

“Yes it is very important, to track user requests and to figure out that whether the system is working or not. Compare the number of requests to the number of users but you need to make sure that the ratio is not too high” Interviewee D, System Administrator and ESGF Developer.

6.4.3 Main activities of ESGF users

The main activities of users of the ESGF that they perform once interacting with the ESGF system are; *registering* for an ESGF account, *logging into* ESGF gateway, *searching* for data-sets, *downloading* data, *performing compute functions* on the data using HPC and *visualization of climate data* via visualization applications such as UV-CDAT, see Figure 6.13 (H Chunpir et al., 2013). Users might require help or support in performing all of the main activities that are highlighted. It is also evident from the interviews that these activities are performed by the users of ESGF.

“The main user activities performed using the ESGF e-Science infrastructure is searching, finding, fetching, computing and visualization of climate data especially from CMIP5 project. Searching, finding and fetching. Augment to add some compute capability. So that there is less fetching. The purpose of fetching data is to do computation. The data is too big to move so the solution is to do the computation there” Interviewee C, ESGF Developer.

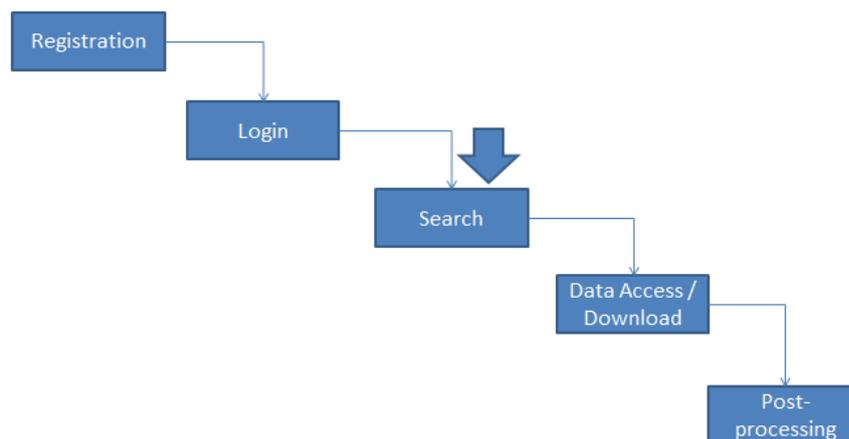


Figure 6.13 The workflow sequence of the users with ESGF.

It is important to note that currently the users do not support other users of ESGF. As per observations of the authors and his colleagues, at least one can say that their contribution to support other users is minimal.

6.5 Attributes of people supporting users of ESGF

The users of ESGF e-Science infrastructure and the employees of ESGF, both are the actors to the user support process of ESGF. Since they are part of the process, it is also crucial to manage their needs and make their tasks of supporting end-users interesting. The employees of ESGF and its associated projects are mainly involved in supporting users of ESGF at different levels. It is significant to know the people who support ESGF users and what exactly their expertise is. What activities do they perform? How much time do they dedicate to the user support activity? How many people are involved in supporting users? How often do the employees answer users' queries? In this section these all issues are dealt with, based on data collected from different data sources (mainly interviews and survey-questionnaire).

6.5.1 Involvement of ESGF employees in the user support process

The users of ESGF are mainly supported by the ESGF e-Science infrastructure employees. The employees are part of all the organisations that are partners of ESGF, distributed all over the world, since it is evident from the structure and organisation of the ESGF e-Science infrastructure (as already described in the previous chapter, Section 5.3, page 111). Therefore, the employees who support users can be in any part of the world, at different time zones. In order to support users they are required to coordinate to each other in user support activities as well as infrastructural development activities. The supporters of ESGF users are mainly top quality computer scientists as well as climate scientists. However, this is not quite the best use of their talents as they often answer routine and petty queries. From the survey-questionnaire, it is evident that the staffs that support users are all having an academic background. Most of them are even having PhD in climate science, computer science and other natural sciences. One of the interviewee has pointed out on this aspect:

“We do have people under our funding who respond to those things (user queries) and they are required to do that, that’s part of their job. But it is not necessarily a good use of their talents because these guys are really top quality computer scientists, who really should be spending time developing ESGF software and not telling people if you want to download that data set quicker, you do this. It is not really good use of this. What I am saying is, it is not really scalable solution to the future to rely on these guys to be on the help desk. They are contributing to the help desk...” Interviewee B, Climate Scientist.

The total number of employees of ESGF who are involved in supporting users of ESGF range from 30 to 38 at maximum as estimated by the author as well as the colleagues working for ESGF. There are different degrees of being involved in the user support activities. This view is evident from the following comment of an interviewee.

*“There are **different degrees of being involved in user support** and in our Face to Face (F2F) meeting there were like 50 or 60 people and lot of these people are not involved in ESG(F) user support, so they won’t be appropriate as help desk candidates anyway.” Interviewee B, Climate Scientist.*

Once asked to an interviewee that how many people are actively involved in the user support activities, the answer was:

“...may be 10. But there were more employees from all 20 institutions who are off and on taking part in user support. If someone answers one question he might be not the active support guy” Interviewee D, System Administrator and ESGF Developer.

It is interesting to note that the users of ESGF are not involved in helping other users at the moment in the current user support process.

6.5.2 Activities of people supporting users of ESGF

The employees of ESGF have several activities that they perform and providing user support is not the part of their main activity. In this sub-section, the details about the main activities performed by the staffs of ESGF and its associated projects along with the nature of user support activities that they cover is provided.

6.5.2.1 Key activities

ESGF employees, who are scattered worldwide, are actually performing multiple tasks. Their core tasks are software development, system administration, documentation i.e. writing project deliverables, research work as depicted in the Figure 6.14. Engagement in user support activities is not the formal official task of ESGF employees. Nonetheless, the ESGF employees do provide user support services at their own initiative i.e. it is not mandatory for them to support users. The incentives of supporting users are not currently set by the e-Science infrastructure management.

The categories mentioned by the support staff under category “others” are: Software use, data management, public relations and guidance.

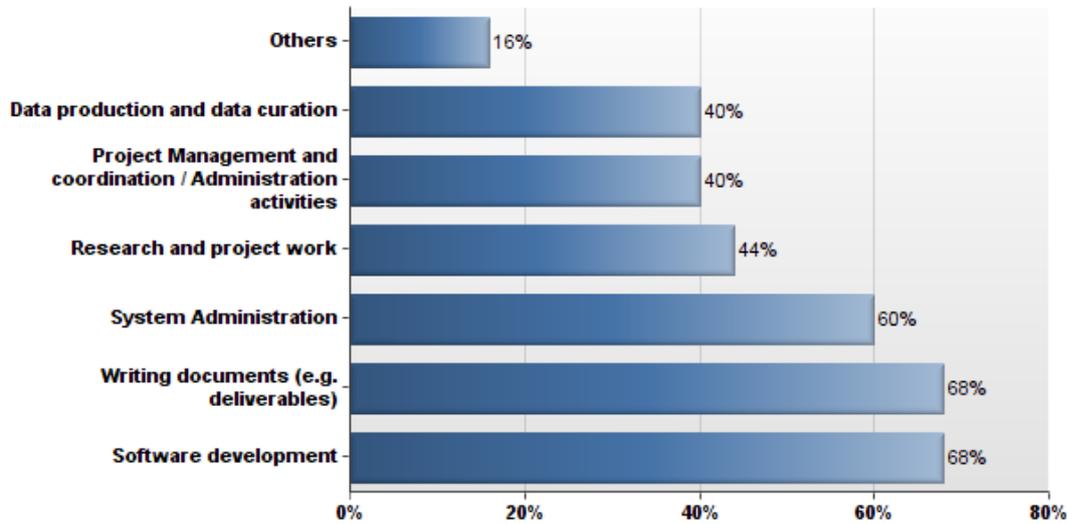


Figure 6.14 The main activities of the employees who support users (other than support activity).

6.5.2.2 User support activities

It was already stated in the previous section that the user support activities are not formally part of the ESGF employees. Nevertheless, they are still being performed in the form of answering questions to users, mainly via e-mail. In addition to the activity of answering user queries, the update of user support information on the user support portals (web-pages) is also an important user support activity. Normally, the employees of ESGF e-Science infrastructure perform these user support activities. There are no policies or standards set by the ESGF management about updating information on the webpages to support users. However, the information recorded from the survey questionnaire about updating information on the webpages can be seen in the Table 6.5. Updating information on user support web-pages is not regular. The update is mostly done when new releases of the software in ESGF are out.

Table 6.3 The frequency of updating the user-support webpages as recorded according to the responses in survey questionnaire.

#	Answer Options	Bar Chart	Response(s)	%
1	Daily		1	4%
2	Weekly		5	22%
3	Monthly		7	30%
4	Quarterly		3	13%
5	Others		7	30%
	Total		23	100%

The comments of the respondents recorded under the "Others" category in the questionnaire-survey were as follows:

- I did not, it was an obligation of the host of the help desk
- Never
- Per release
- When there is something new to be documented
- On demand, whenever new information is needed in case of new systems or versions are installed

Documentation is one of the activities associated with the user support; there are documents, web-pages that are created to share information not only with the users but also with the other ESGF staffs such as developers, data curators, climate scientists etc. However, it has been observed that there is *no namespace set for the documents*. At the same time these *documents are hard to find*. The versioning of the documents is also not done. This is because, not enough resources have been allocated and no standards are set so far. Therefore, documentation in ESGF partner institution is needed to be improved as well as norms of documentation is needed to be set. An example of setting norms of documentation includes for instance; namespaces which is followed by all partner organization of ESGF.

From the interviews, it was noticed that the technical staffs do not like to write documentation in general and documentation of code in particular. From the following interview excerpts, it is clear that there is a resistance to write detailed documentation among the developers working for ESGF:

“Documentation especially in the context of code, I don’t do that, no resources for it. I write self-describing code. Variables are descriptive enough. Too much documentation

scares people away. Tons of instructions may lead to destruction” Interviewee C, ESGF Developer.

“...the documentation for developers is not up to date. There is also like developer support and data manager support which is more technical but we really do not spend time documenting, most of time is spent developing our software. Documentation is done very rarely. Other workers at ESGF just want to code and they do not really care about documentation. This is a common problem in Science, in Software Engineering in general” Interviewee D, System Administrator and ESGF Developer.

Moreover, it is further evident from the following interview excerpt that the documentation in ESGF is not complete:

“In general in ESGF the documentation is not complete as it should be. And it is a matter of resources and they spend all of the resources trying to get things to work and basically didn’t have enough to cover other things. This is because we needed all money to make the system robust and make sure to simplify the things at the node. The people who are managing the nodes, they were successful in doing that although there was a bit of delay” Interviewee B, Climate Scientist.

Once asked explicitly in an interview that whether all information or documents uploaded at the ESGF are up to date. The interviewee replied:

“The FAQs for users are certainly up to date. But there is certainly much to do. The documentation, also for developers is not up to date” Interviewee D, System Administrator and ESGF Developer.

“It would be effective to improve user interface (UI) and the documentation, so that one can use online support” Interviewee D, System Administrator and ESGF Developer.

6.5.3 Expertise

According to the survey-questionnaire, most of the employees who support users are mainly technical experts, 68%, see Figure 6.15. Almost 32% of the employees consider themselves to have both technical and scientific expertise. None of the employees consider themselves as a scientific expert alone.

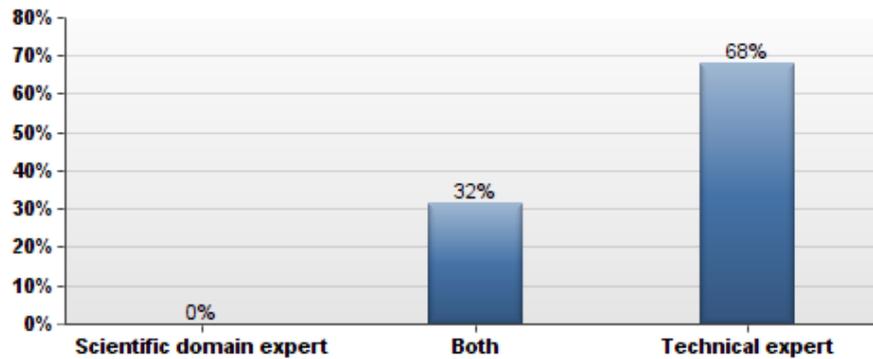


Figure 6.15 The responses of the respondents to the question what do they consider them in a support process.

The employees, who support users, are involved in multiple ESGF core and affiliated projects (or institutions) as shown in the Figure 6.16:

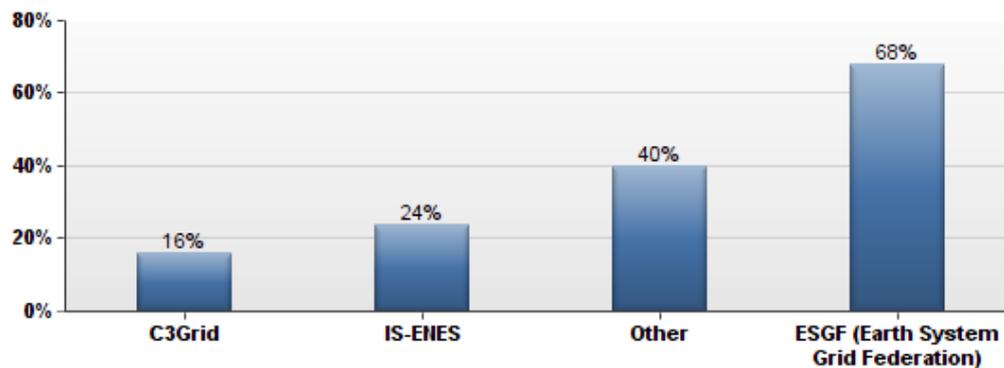


Figure 6.16 Respondents (user-support staff) engaged in different projects. Some respondents are working on different projects in parallel.

Employees, i.e. staffs involved in the user support process are engaged in different projects associated to ESGF in parallel to the core ESGF project work. The main projects other than ESGF include; C3Grid and IS-ENES. However, in addition to ESGF related activities, these projects may have their own indigenous activities. Under the “Other” category, the other projects and the partner institutions associated to the ESGF include European Union Cloud Inter-comparison Process Study & Evaluation project (EUCLIPSE), *Centre national de la recherche scientifique* French National Centre for Scientific Research / Institut Pierre Simon Laplace (CNRS/IPSL), World Data Climate Center (WDCC); 3-responses, German Climate Computing Centre (DKRZ); 4

responses, High Performance Computing (HPC); 2- responses and data formatting of Coupled Model Inter-comparison Project Phase 5 (CMIP5), as stated by the respondents of the survey-questionnaire.

6.5.4 Time dedication to user support activities

The employees associated to ESGF and ESGF-like e-Science infrastructures dedicate time for the user support activities ranging from 1 to 3 hours daily, on the average, as according to 85% of responses, see Figure 6.17. 1 to 3 hours daily is almost 20% to 40% chunk of their working time of the employees.

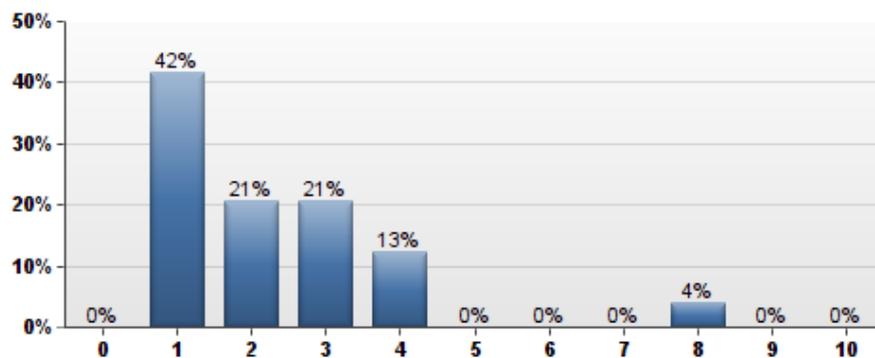


Figure 6.17 The time dedicated by the support staff in climate e-infrastructures support frameworks.

6.5.5 ESGF staffs involved in handling user requests

This sub-section depicts the percentage of incoming user queries handled by the staffs of ESGF, captured through the survey-questionnaire. The responses reveal the percentage of user requests handled by total number of staff members who support users (or total levels of delegation of user requests from one support staff to another) on the average.

Most of the incoming requests could be handled by a single support staff member of ESGF himself or herself, up to 62%, see Figure 6.18. One can infer from this finding, that 62% of incoming queries can be handled by a single person without interchange of information with any other support staff members. Out of all incoming user requests, 28% of the requests are solved with the help of collaboration of two staff members of ESGF. In this case, one can claim that the incoming queries are further delegated to the Second Level Support (SLS). Likewise, around 6% of the user-requests are solved by the collaboration of three staff members of ESGF, see Figure 6.18.

This implies that very fewer requests are further delegated to third level support. Rarely, more than 4 people are involved in solving a request, 4%, as shown in Figure 6.18. So, most of the queries are handled by a single staff in ESGF user support process. These numbers reveal that most of the incoming user requests are probably relatively easy to solve and do not require involvement of more than one staffs.

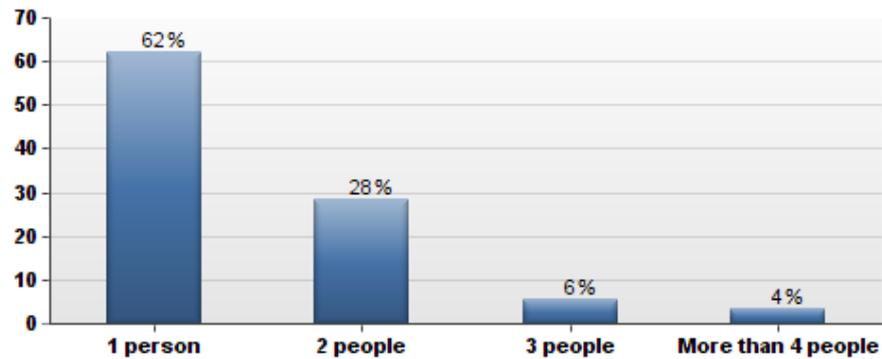


Figure 6.18 Average percentages of total incoming user-requests handled by ESGF human resources (support staff) in average.

6.6 Nature of user support provided currently in ESGF

The types of user support currently provided in ESGF are technical as well as scientific in nature i.e. technical and scientific questions, from the users of ESGF, are answered. Nonetheless, they are not integrated into a single entity. The scientific user support is completely separate and is a disparate unit and is not part of the technical user support unit. The details are given as follows:

6.6.1 Help desk dealing with technical questions

The type of user support currently provided in ESGF is mainly technical user support. In case of a problem that a user of ESGF e-Science infrastructure may face due to an anomaly in its operations or an information need, the user sends a query and waits for ESGF staff to solve problem. Answers to particular questions are provided via e-mail based mailing-list which is part of the help desk. However, in addition to answering the user queries the information for users of ESGF is also provided in the form of user support web pages. This enables a user to get information and help him or herself. This is a known as self-support.

6.6.2 Help desk dealing with scientific questions

Scientific support is currently not formal part of ESGF user support help desk. However, some of the users interviewed want it to be integrated into the ESGF user support. Thus, ESGF user support shall not only entertain technical user queries but also scientific queries. It is better to have a linkage and connection of both. Currently, the scientific queries are directly sent to scientific people to the associated scientific institutes. There is a perception amongst the technical staff that the scientific user support is working very well and that there are not many scientific questions.

“Currently the scientific user support is working well. The (computer and climate) scientists, who are employees of ESGF, who support users, they can refer the user queries to other people who produce the data. They can refer to the places like British Atmospheric Data Center (BADC)... No one has science questions so far” Interviewee C, ESGF Developer.

The perception of technical employees that there are no scientific queries coming to ESGF help desk is, however, challenged. Actually, the users of ESGF can find the contact person in the meta-data of a data-set, that they download and in case of question can directly contact them. Besides meta-data is nicely described.

“Models, meta-data is nicely and meticulously described, written in a way that scientists know” Interviewee C, ESGF Developer.

From the technical staff, there was a suggestion to improve the cooperation with the scientific people by offering collaboration via a user support platform.

“Among the scientific community cooperation among different modelling centers can be made. Probably via support system the integrated collaboration between modelling centers can be made possible” Interviewee C, ESGF Developer.

6.7 The structure and model of user support in ESGF

The structure of user support describes the tiers that the help desk or user support may have. The user support model identifies key elements required to provide user support. The following sub-sections describe the details about the structure and model of user support in ESGF.

6.7.1 User support model

According to the data captured from the survey questionnaire, the current user support model used by the respondents is quite a mix. 50% of the respondents claim that they use single point of contact (SPOC) i.e. centralized service desk model, as shown in the pie chart in Figure 6.19. 41% of the respondents claim that decentralized service desks are being used. It is important to note that centralized and decentralized service desk models can be distributed, i.e. present at different locations. 36% of the respondents

claim that their support desks are also distributed. Only 18% of the respondents are of the opinion that e-support is in place. 14% of the respondents claim other categories, which include telephone, skype and personal contact.

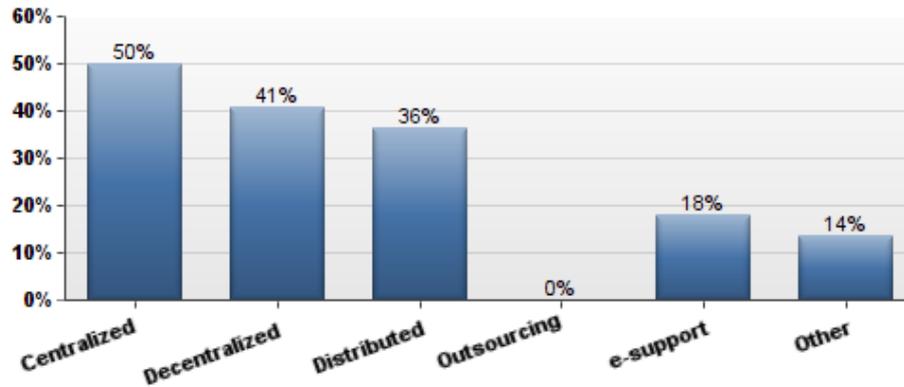


Figure 6.19 The different user-support models in use in climate e-Science infrastructure projects as marked by percentage of support staff of ESGF.

6.7.2 User support structure

From the data collected via the survey-questionnaire, around 41% of the respondents claim that there is a practice of two-tier user support system in ESGF and its associated climate e-Science infrastructures. Similarly, almost 41% of the respondents are of the opinion that there is just a single-tier system in place, in the current ESGF e-Science support process and its associated climate e-Science infrastructures, as shown in Figure 6.20. Under the others category, (see Figure 6.20), there were three replies by the respondents. One respondent claimed that sometimes Second Level Support (SLS) is getting questions directly from the users. Another reply claimed that there are three-tiers at his institution, thus making the three-tier support structure around 6%. Yet, another respondent claimed that there is only one person at their institute who is tacking the user queries.

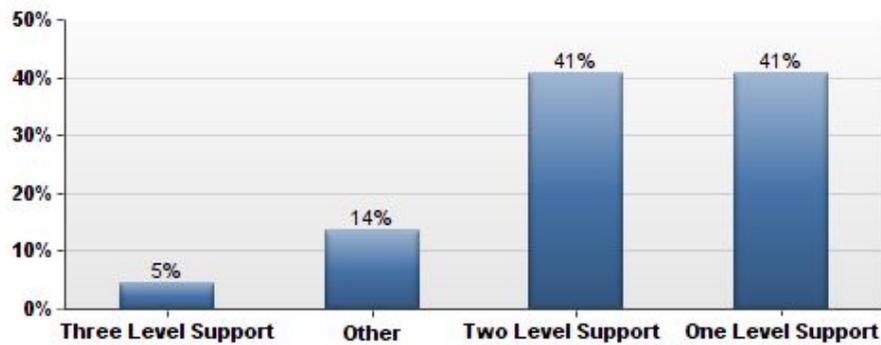


Figure 6.20 The percentages of responses belonging to the structure of support in climate e-Science infrastructure support.

Under the “other” category the comments of respondents recorded about the structure of support in climate e-Science infrastructure were as follows:

- Depending on the area (second level is getting the questions directly too)
- One person at the moment at Irish Centre for High-End Computing (ICHEC) i.e. Irish High Performance Computing Centre
- Up to three levels

In an interview with a technical expert, he claimed that there is no difference between the roles of FLS and SLS in ESGF user support process:

“Most of the people are acting both as First level support (FLS) and Second level support (SLS) both...” Interviewee C, ESGF Developer.

So, the tiers are not defined in the user support process of ESGF.

6.7.3 User support levels in ESGF

In the survey-questionnaire, a question was asked that whether the employees of ESGF do a distinction between the FLS and SLS activities, the responses to this question captured were interesting as shown in Figure 6.21. Almost 48% of the employees of ESGF that support users believe that they are carrying out Second Level Support (SLS) activities. Nearly, 44% of the respondents claim that they are carrying out First Level Support (FLS) activities on the average. 8.2 % of the responses of the employees do not consider them either FLS or SLS category. This is because either they are the only point of contact or they consider that First Level Support (FLS) and Second Level Support (SLS) roles are not applicable to them because they are “not that organized.”

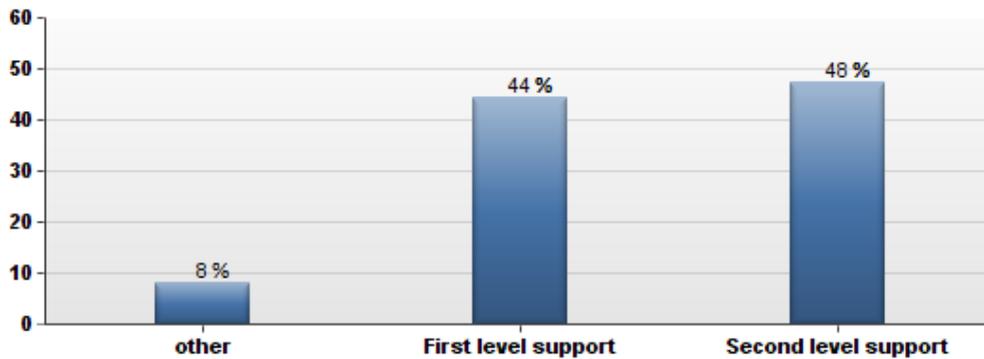


Figure 6.21 Distribution of support activities represented in the form of FLS, SLS and others, expressed in percentage.

Responses recorded under “Other” in Figure 6.21 were: “we’re not that organized” and “I’m often the only point of contact”.

6.7.4 ESGF user support is on-demand

92% of the employees of ESGF and ESGF-like e-Science infrastructures handle user requests mainly on “on-demand” basis i.e. as the user need arises, see Figure 6.12. Users normally send the queries as e-mails. The e-mails are received either in a mailing-list or a particular request tracking system at the concerned administrative body or institute. Any employee or even user (in case of mailing list) can look or jump in to answer the query right on the spot on volunteer basis.

“...People jump in at any time and answer question” Interviewee C, ESGF Developer.

Currently, there is no dedicated time slot in a time zone within which the user requests are serviced.

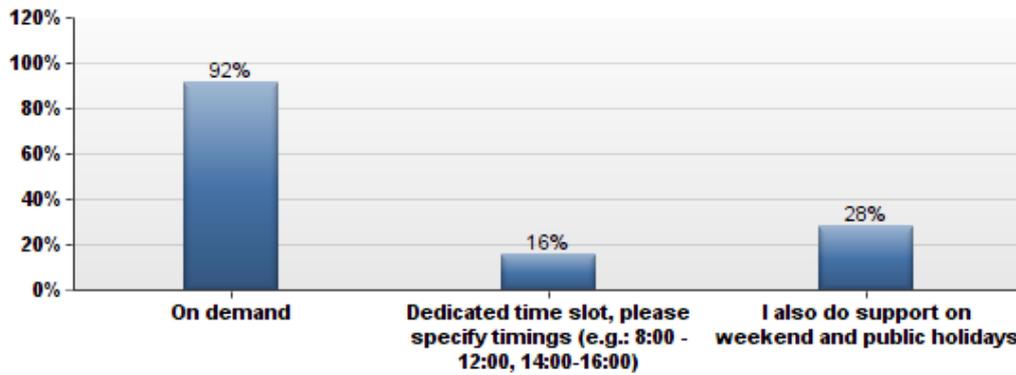


Figure 6.22 The user support in ESGF e-Science infrastructure is mainly “On-demand”.

6.8 Current practices of user support

The current practices i.e. the snap shot of providing user support can be examined by looking at the different aspects of the current user support process and by exploring the attributes of incoming user requests. For instance: Total number of incoming user requests, trend of user requests, understandability of user requests and other aspects are the indicators of the current practices of user support. They were partly already explored in Section 6.3.2 (page 136). The details about communication media used by the users to seek help from the ESGF staff are explored in Section 6.8.2. Similarly, the details about time factors involved in servicing user requests and responding to the users is provided in Section 6.8.3. Finally, the details about user request documentation is given in Section 6.8.4.

There are two modes for the users to get help on the troubles that they may face. Firstly, the users are supported mainly via help desk of ESGF. Secondly, the users also have a possibility to look at the relevant web-pages to get self-help. For self-help, currently relevant information on the ESGF websites and wiki is provided.

“Website and a wiki are in place to support the users. There are two schools of thought about the Website and the Wiki. One school of thought says Wiki is enough. The other wants both” *Interviewee C, ESGF Developer.*

“Normally, technical people including developers of ESGF and tech savvy users prefer wiki. However, other users and policy makers like information on the webpage” *Interviewee H, Climate Scientist and User.*

6.8.1 Examining the attributes of user requests

In this sub-section, the total number of incoming user requests, the trends of incoming user requests, understandability of user requests, average percentage of unresolved requests and other aspects are discussed.

6.8.1.1 Number of user-requests per working day

According to responses recorded in the survey questionnaire by the respondents, the average number of incoming requests from end-users per working day in a year, range from 1 to 5, as shown in the Table 6.4 below. However, the majority of employees are of the opinion that the incoming user requests per day are from 1 to 2 on the average. From this data one can infer that the workload in climate e-Science support service desks is lesser than in comparison to the commercial help desks in commercial organisations.

Table 6.4 The average number of user-requests received by the support staff per working day.

Total number of user-requests received per day	Responses
Less than 1	5
1	7
2	7
3	2
5	3
	Total responses: 24

The data captured from the interviews coincide with the data captured via survey-questionnaire. Once asked from the ESGF staffs in interviews that “How many requests you answer on the avg. per day?” The following replies were recorded:

“Four to five user requests per day on the average” Interviewee D, System Administrator and ESGF Developer.

“I am not sure, but may be two user requests per day” Interviewee C, ESGF Developer.

6.8.1.2 Trend of user requests

As the efforts to provide better user support information on the ESGF web-pages and wikis were initiated, the quantity of the incoming user requests have been decreasing since then.

Once asked in an interview “Do user requests increase or decrease with the passage of time?” The replies were:

“Recently the user requests have been decreasing which is good news” Interviewee C, ESGF Developer.

“They have been decreasing as a matter of fact. This is good news. There are reasons for that first of all, we had a big spike before the IPCC’s AR5 report, and now that’s done. I think some of the users find the solutions on the webpages. Some people who had trouble in the past their questions are cleared. But there are still some percentage of users who cannot use the system without support” Interviewee D, ESGF Developer and System Administrator.

Actually, once the climate researchers need to write research reports, the usage of the ESGF system is high and therefore the requests increase abruptly. For example, in case of IPCCs AR5 report, when the submission deadline is getting nearer there are many logins to the ESGF system and lot of pulling of climate data from the ESGF data archive. During the interaction with the ESGF system, the users face troubles and that’s why they send many queries to the ESGF user support. As a consequence, there is an increase in the trend of user requests and after publication of relevant research reports or publications the trend lowers again. Consequently, there is bust and a boom at different intervals of time depending on the activity of researchers. Similarly, the PhD students who pull the data at a single point in time may not access ESGF data at all, once they are done with their PhD or another research project.

6.8.1.3 Understandability of user requests

If the incoming user requests are not understandable to the ESGF staff then there is more consumption of time by the ESGF staff to treat the user requests. The responses in the survey questionnaire reveal the percentage of incoming understandable requests. According to the Figure 6.23, nearly 61% of the employees (support staffs) believe that a lot of incoming user requests are clear and understandable. 17% of the employees, who are involved in supporting users of ESGF, are of the opinion that only some of the incoming user requests are understandable. 9% are of the opinion that only little portion of the received requests are understandable. So, majority of the incoming user requests are seen as understandable by majority of the ESGF staffs. 13% of the employees mentioned “others” category in the survey questionnaire and mentioned that up to 50 to 70 percent of the total incoming requests were understandable.

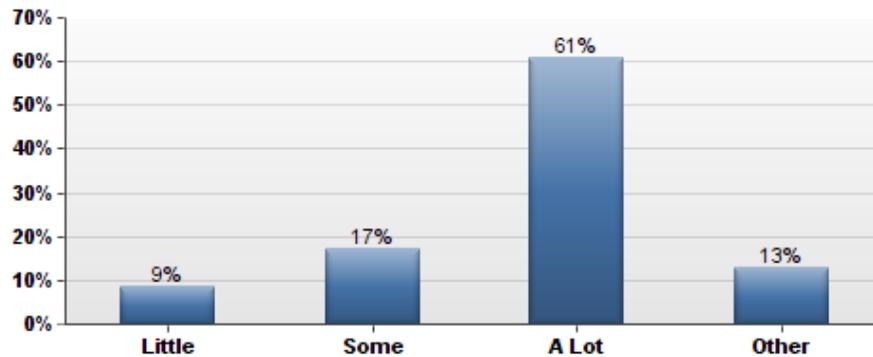


Figure 6.23 The proportion of incoming user-requests that is understandable.

6.8.1.4 Number of employee responses to attain clarity of user requests

If the incoming use requests are not understandable then the ESGF employees do need to write back to attain clarity about user`s request. Almost 91% of the respondents who are involved in supporting users in ESGF are of the opinion that they normally need to write back, at least one time or twice to attain clarity about the original request (see Figure 6.24). Out of 91%, nearly 45% of the respondents claim that they need to write back at least twice to understand a user request. Similarly, the remaining 45% of the respondents are of the opinion that they need to write back at least once to understand the request. No one said that they need to write back thrice which means that the incoming user requests get comprehensible after one or two times of further questioning by the support staffs of ESGF. Under the others category, the respondents responded that they need not to reply at all to get an understanding of the user requests or may be need to reply once, depending on the situation and not every time. The others category make 9% of all the responses of the ESGF staffs to the survey-questionnaire.

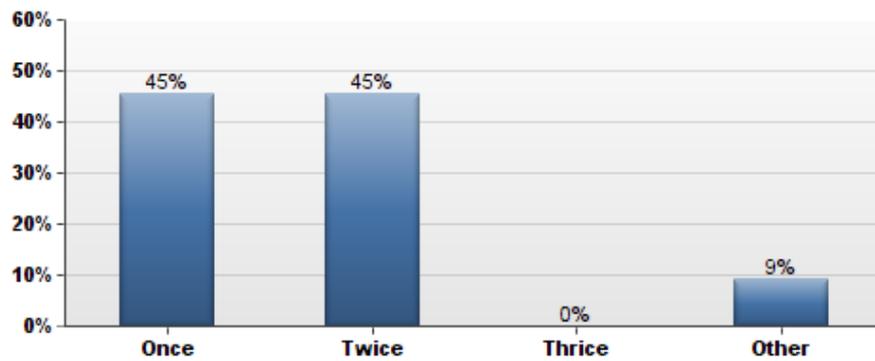


Figure 6.24 Number of times support staffs have to write back to a user to fully understand the user request.

The responses recorded under “other” were: *Mostly not at all* and *from zero to one (0...1) times*. So, only two respondents of the survey questionnaire were of the opinion that they most need not to write at all or once in a while that may vary from 0 to 1 e-mail to understand the original user request from time to time.

6.8.1.5 User request pattern

The user request pattern is not in place completely in the current user support process of ESGF. As according to the observation of respondents, only some of the users follow a particular request pattern, see Figure 6.25. Consequently, the user request pattern is not being utilized in the current user support process of ESGF e-Science infrastructure to confirm the uniformity of the incoming user requests.

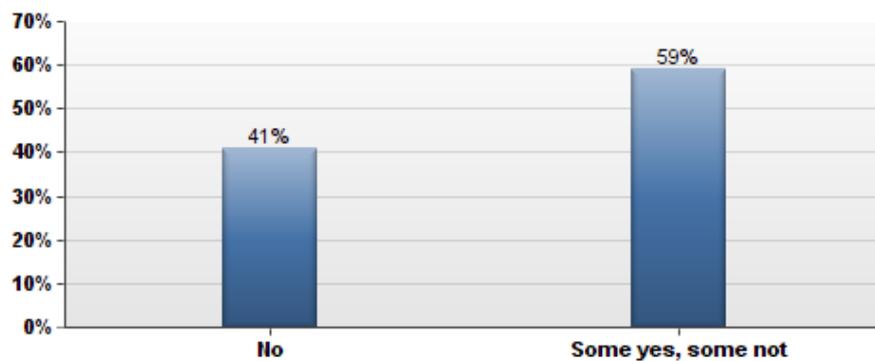


Figure 6.25 The utilization of user-request patterns by the end users.

6.8.1.6 Unresolved user requests

Unresolved user requests are the requests that are never resolved and probably ignored after some time from the point in time when they arrive to the ESGF staffs. Once asked in the survey questionnaire about the percentage of the unresolved requests, the majority of respondents claimed that only a little or less than 5% of the incoming user requests remain unresolved in a year, see Figure 6.26. However, up to 22% of the respondents believe that from 20% onwards incoming requests of the users remain unresolved in a year.

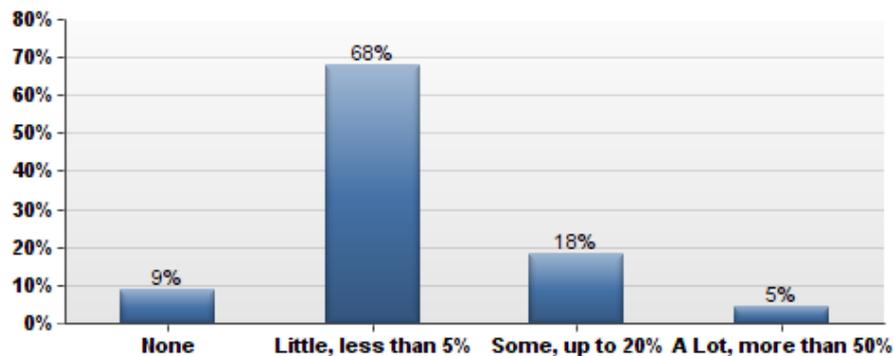


Figure 6.26 Percentage of the user-requests on the average that remain to be unresolved in a year.

In the interviews, once asked the same question, the reply was:

“I am aware of the fact that all of the incoming user requests do not get answered” Interviewee D, ESGF Developer and System Administrator.

So, the technical staffs of ESGF are aware that not all incoming user requests get answered and resolved. In the interview it was then asked if there is a *mechanism to detect unanswered or unresolved requests*. The answer of the same interviewee was:

“The only mechanism would be to go and look for all the e-mails. I think GITHUB would work just fine. Someone is needed to cut and paste the e-mail and put it in GITHUB, creating a ticket and then you know one can directly answer in it. Actually we can do that to be honest. But we need to research exactly. You need special permissions to do that I think anybody can write. This will be a good step forward. I like GITHUB, it is the best. I don’t think that there is a need for a commercial solution for tracking requests but we can use GITHUB that would work” Interviewee D, ESGF Developer and System Administrator.

Another interviewee is of the opinion that, partly the ESGF staffs believe that users are able to get data without any help that makes them not to answer queries:

“Although help desk was important but it was not absolutely fundamentally essential because, some people are able to get the data without the help desk. This is the reason why some of the queries were not answered” Interviewee B, Climate Scientist.

Moreover, the same interviewee argued that if the user queries are always ignored then there can be serious consequences under which the climate research activities through ESGF may suffer in the sense to achieve both the long term and the short term goals of climate research.

“...if the user queries are ignored ...In the long term we suffer because we do not have this information and in the short term we suffer too” Interviewee B, Climate Scientist.

Furthermore, the same interviewee said that 10% of the requests which are not responded to need to be avoided. Moreover, he proposed that the long delays to respond users of ESGF might discourage them in their research activities. The response from the ESGF staffs to the users within two days is expected to be fine.

“...if we are responsive in a day or two that’s completely fine. We have to avoid very long delays and we have to avoid like 10% of the requests not being responded to” Interviewee B, Climate Scientist.

From the mailing-list analysis it was found that 13% of the queries do not get response.

6.8.1.7 Multiple delegation

A practice exists amongst the ESGF staffs to ask more than one support staffs at the same time, consequently, forwarding the user request to more than one ESGF staffs. The phenomenon to forward the request to another ESGF staff for further treatment is known as delegation of the user requests. If a user request is forwarded to more than one ESGF staffs for further treatment, this phenomenon is known as “Multiple Delegation”. According to the survey-questionnaire, up to 61% of the ESGF staffs involved in user support are of the opinion that multiple delegations are never (13%) or rarely made (48%), see Figure 6.27.

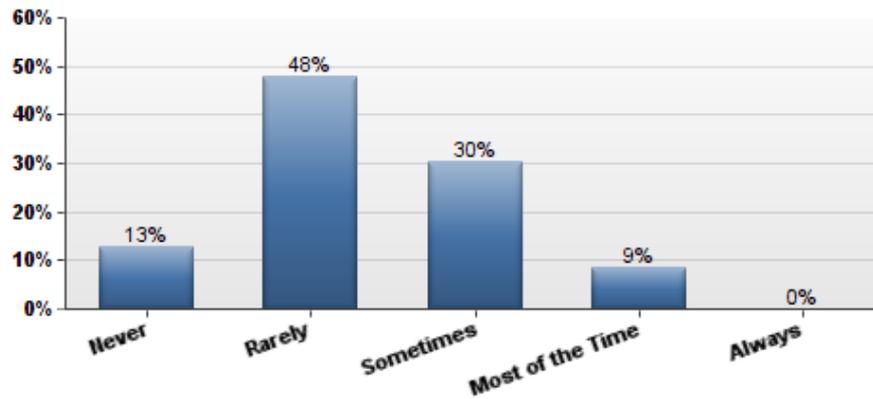


Figure 6.27 The average percentage of support staffs who delegate user-requests to multiple peers, simultaneously.

Once asked about multiple delegations in an interview, the interviewee replied:

“Multiple delegations may be avoided; most effective I think is the single delegation” Interviewee C, ESGF Developer.

“There may be confusion of responsibility in case of multiple delegation of user query to multiple staff” Interviewee C, ESGF Developer.

The feature of multiple delegations which is currently part of the user support process in ESGF is debatable. Consequently, in author’s opinion the practice of multiple delegations is discouraged unless the receiver of the user request acknowledges that they take responsibility to process the user request according to best of their knowledge and capability. However, the ESGF executive and the technical committees need to decide on it further.

6.8.1.8 Possible categories of user requests

Once the respondents of the survey questionnaire were asked about the categories of user requests that might be possible, they revealed the categories that they believed that the incoming user requests might fall into. The respondents clicked every category that they believed relevant to the user requests coming to the user support process of ESGF and its associated projects. This can be seen by the percentage and the responses in the Table 6.5. Uncovering the popular categories in the ESGF and its associated climate science e-Science infrastructures is vital to know about the possible classification of the incoming user requests in the user support process. Table 6.5 reveals the possible list of the categories to which the user-request might belong. Furthermore, the respondents provided the information about the categories in which the incoming user requests might fall in under the “others.”

The main categories were *general information, data access, workflow execution, visualisation, and others*. These categories are depicted in the table as items from 1 to 6, see Table 6.5. Other possible categories as suggested by the respondents under the category “6 i.e. others”, as depicted in the Table 6.5 are as follows:

- Projects: IPCC-AR5, CORDEX, CFMIP, LUCID;
- Models: ECHAM6, MPIOM, JSBACH, HAMOCC, CCLM4
- How to publish data i.e. data publication?
- Getting logged in
- Miscellaneous
- Post-processing of data
- Data citation
- Quality control

Table 6.5 Percentage of categories deemed relevant by the support staff, shown in percentage, in ESGF and associated projects support.

#	Answer	Bar Chart	Responses	%
1	General information about structure and functions of e-Science infrastructure		11	50%
2	Data access and download		18	82%
3	Workflow execution		7	32%
4	Scientific support		12	55%
5	Visualization		7	32%
6	Others		4	18%

The sub-categories of the main category: “1. General information ...” of Table 6.5 are depicted in the Table 6.6 from A to F, namely: General queries about e-Science infrastructure, data, format of data, workflows, registration information and others (see Table 6.6). Under the category “F” in Table 6.6, the further possible sub-categories are suggested as: Authentication (i.e. login), model compilation, debugging, optimization, feature development, software installation, software usage and finally queries about correctness of the data which are downloaded.

Table 6.6 Percentage of sub-categories recorded under “1. General information ...” first category of Table 6.5 considered relevant by the support staff members in ESGF and associated projects support, shown in percentage.

#	Answer	Bar Chart	Responses	%
A	General queries about anatomy of e-infrastructure		11	50%
B	Data queries, e.g. queries about the types of data sets available in the ESGF portal		16	73%
C	Format queries, i.e. queries about formats of the data which are available to be downloaded		8	36%
D	Workflow queries, i.e. queries about workflows which are available and their functionality		11	50%
E	Registration queries, i.e. queries about registration process e.g. Where do I need to register? What benefits can I have after registration?		16	73%
F	Others		3	14%

From the responses recorded in the survey-questionnaire, it is evident (as shown in Table 6.5 and Table 6.6) that most of the incoming user requests fall under data and registration inquiries.

6.8.1.9 Categories of user requests received

In the survey questionnaire the respondents provided information about the user requests belonging to particular categories that they normally receive. This information is depicted in percentage in the Figure 6.28. The respondents marked the percentage in front of each category under which they receive user requests. The more popular the category of user-requests is, the higher the percentage assigned to that category by the ESGF staffs involved in supporting users is.

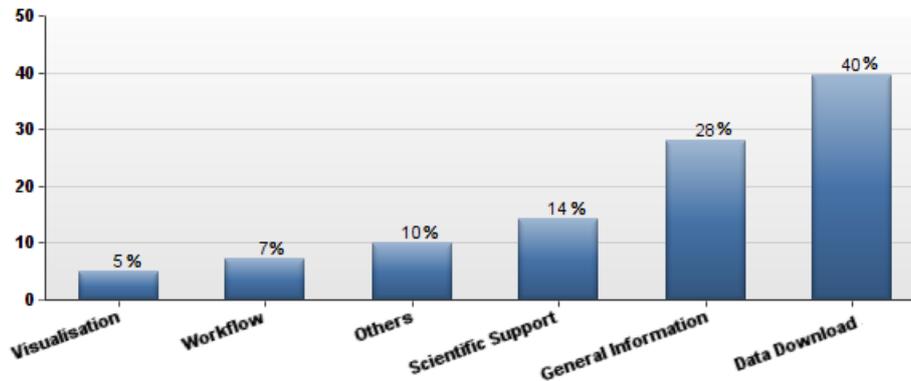


Figure 6.28 Category-mix of the user-requests that are received by support staff member in the current support process in climate e-Science, depicted in percentage.

Again it is evident from the Figure 6.28 that most of the incoming user requests fall under “data access and download” category, followed by general information.

6.8.2 Media of communication

The users of ESGF system send the requests to the staffs of ESGF to get help on the issues via a communication medium. The e-mails are sent by the users of ESGF if the users cannot get any information on their own which is usually the circumstance. In this sub-section, an overview of the characteristics of the communication media used by the users and the staffs of ESGF to get support is presented. In ESGF e-Science infrastructure, the two tracks of getting help via media are present in the current user support process as revealed by one of the interviewee.

“Actually user support comes in, how we assist the users. And the other one was how do we assist the developers. So we have two tracks. One is for the users and one for the developers” Interviewee L, ESGF Chair.

One track is about the internal communication between ESGF staff and the other is about the external communication between ESGF staffs and the users.

“For data users, external communication media are used. For internal users internal medium are used” Interviewee L, ESGF Chair.

6.8.2.1 Media of external communication with user

The external communication between ESGF staffs and the users is mostly based on asynchronous, e-mail based media, mainly text. Almost 78% of the external communication through e-mails is conceived either to Request Tracking Systems (RTS) or mailing lists (ML). The percentage of the e-mails received directly into Request Tracking Systems (RTS) is 43% and the one into the mailing lists is 35%. The

percentage of all the communication media used in the ESGF user support process is given in Figure 6.29.

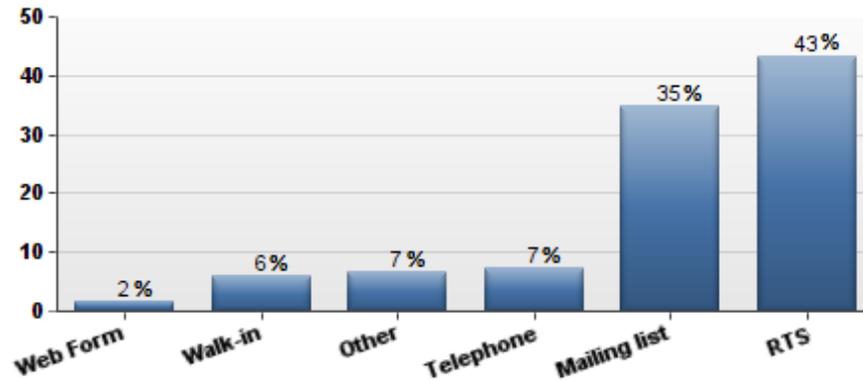


Figure 6.29 The popularity of medium used (in percentage) for external communication with users.

Under the other categories, the predominant categories for the external communication between users and the ESGF staffs mentioned by the users are; direct emails (2 responses), monitoring systems and skype. From the conducted interviews, the interviewees are also of the opinion that the major communication between ESGF staffs and the users is mostly based on asynchronous e-mail based media:

“We have an e-mail list; still the preferred way to do user support is e-mail. So that’s advertised on portals. Then anybody can answer” Interviewee D, ESGF Developer and System Administrator.

“Most of the users send their queries via mailing list” Interviewee C, ESGF Developer.

“It is very important, we have few forms of user support: First one we have, ESGF mailing list, we have websites...” Interviewee L, ESGF Chair.

6.8.2.2 Communication mode for internal communication (in between employees who support users)

The internal communication between ESGF staffs about their own issues as well as about the issues of users is mostly based on asynchronous, e-mail based media, mainly text. However, regarding the internal communication between ESGF staffs, the utilization of e-mails as a medium is a bit lesser i.e. 57% in total as compared to external communication between ESGF staffs and users. This includes 29% of e-mails leading to RT as well as 29% of e-mails leading to mailing lists, see Figure 6.30. However, respondents also practice walk-in; 20%, followed by telephone; 14%, if present at the same location, see Figure 6.30.

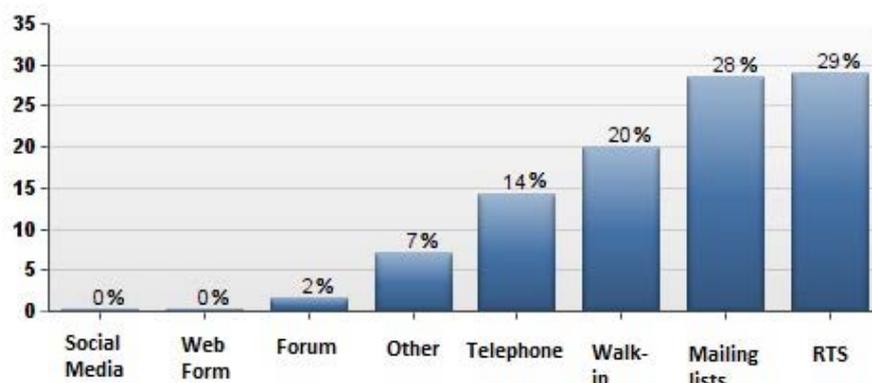


Figure 6.30 Modes of internal communication used (in percentage) within e-Science projects.

Therefore, for internal communication, the synchronous media is utilized thus increasing interactivity between the staffs. Under the “others” category, the users mentioned direct e-mail (2 responses) and sykpe. Under the “social media” category in Figure 6.30; facebook and twitter are meant.

6.8.2.3 Preferred mode of user communication

According to respondents of the survey-questionnaire users prefer *e-mails leading to RTS*, the most (see Figure 6.31). RTS provides tracking possibility, search capability, setting priority as well as maintaining statistics of user requests, therefore, it is preferred. After RTS, mailing-list(s) is the second mode of communication which is popular amongst users. Telephone is ranked as the third medium due to its interactive nature.

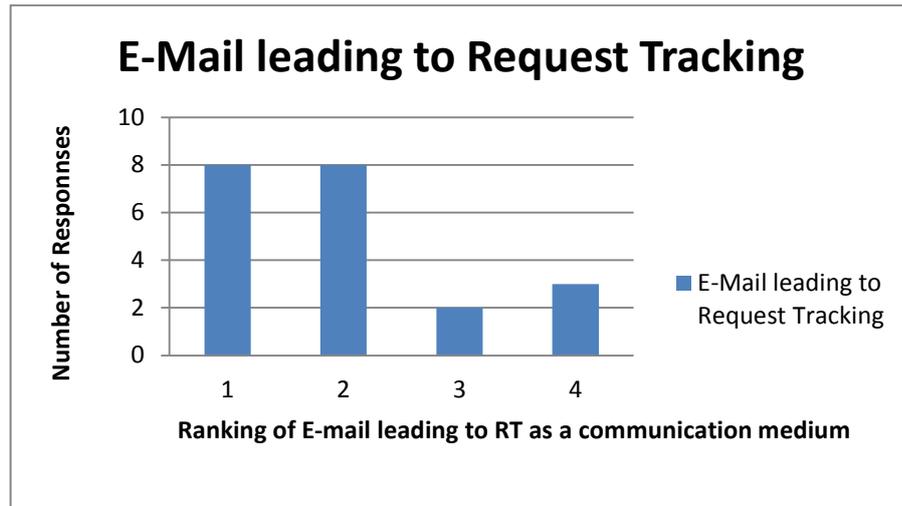


Figure 6.31 Ranking of e-mail leading to Request Tracking Software as a (1st popular) medium.

RTS is favourable amongst the ESGF staffs, a unanimous decision on using a particular RTS suite by the federating partner organisations of ESGF; however, did not take place. Once asked an interviewee that whether RTS is the best option, he replied:

“I think RTS is beneficial, we should put it in place. We made an attempt to do that and that was when BADC established the help desk. Somehow, it didn’t quite totally work well. I never got involved with that. I am not sure why it didn’t work well. We were using GIT HUB before, with that one can open tickets and provide an answer. I personally didn’t have time; I had so many other things to do” Interviewee D, ESGF Developer and System Administrator.

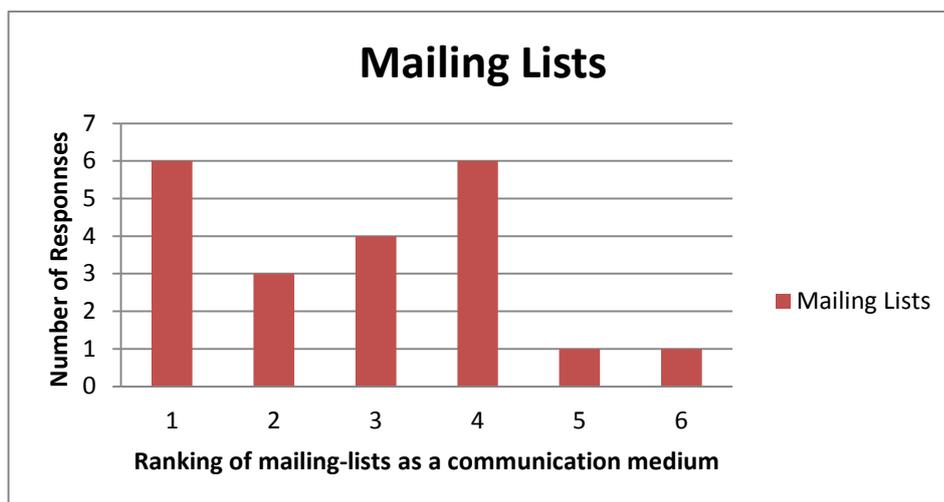


Figure 6.32 Ranking of mailing-lists as a (2nd preferred) medium.

Mailing-list is the second preferred medium by the ESGF staffs in the current user support process; see Figure 6.32 as according to the responses, collected from the survey-questionnaire. It is also evident from an interview protocol:

“We have an e-mail list, still the preferred way to do user support. So that’s advertised on portals. Then anybody can answer. Using e-mail list has been advertised on the webpages” Interviewee D, ESGF Developer and System Administrator.

6.8.2.4 Language of communication

Most of the incoming user-requests are in English language 72%, see Figure 6.23. Others use German language probably because of communication within national partners of the ESGF e-Science infrastructure and its associated projects 22%. Yet others use French and Italian, almost 3%. In global infrastructures like ESGF communication is mainly in English. Consequently, it is important to set a standard as English so that all nationalities may understand the language.

6.8.3 Time factors

Users expect handling of their requests in a timely manner, as they are working on their research questions and fetching the data from ESGF to do data analysis and submit the results of their research in the form of a publication. Hence, time is an important factor in the user support process. The researchers want their troubles to be addressed by the helpers without much delay. In the current user support process, the reply time factors and solution time factors have been captured via the survey questionnaire.

6.8.3.1 Reply time factors

The reply time factors include average reply time of the ESGF staffs involved in helping users as well as the reply time of staffs once the request has been forwarded to him or her by the staff who has received the user request initially. In addition, the placement of auto reply mechanism in the user support process of ESGF provides the acknowledgement that the user request has been received with the processing time estimation of the requests.

Automatic reply mechanism

The placement of auto reply mechanism in the user support process of ESGF informs the users that the request has been received. Most of the support staffs in the climate e-infrastructures user support process do not use automatic reply mechanism, 79%, see Figure 6.33.

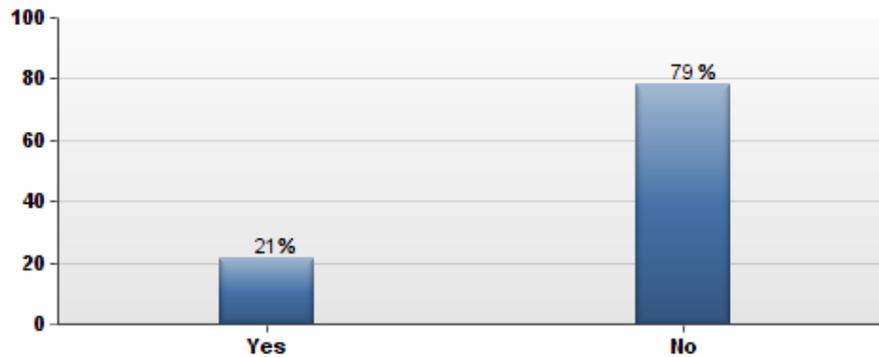


Figure 6.33 The percentage of (average) automatic reply mechanism present in the current support framework in e-Science.

Consequently, in the current user support process, the automatic reply mechanism has not yet been employed entirely by all partner institutions of ESGF.

Personal reply time

From the Figure 6.34, 60% of the respondents claim that they need more than 1 but less than 3 hours to provide personal response to particular user requests. Almost 30% of the support staffs need more than 3 hours to send a personal response which is also acceptable in case of distributed help desk.

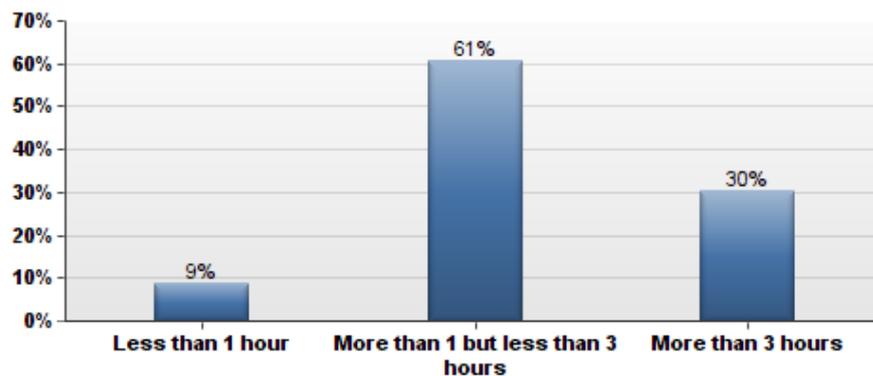


Figure 6.34 The average response time to a user-request by the support staff in climate e-infrastructures support frameworks.

It is important to ensure that the user receives the first reply preferably within 24 hours and the time limit of 48 hours shall not exceed. An interviewee told about the preferable time within which the user must get a reply, he said:

“I would say within 24 hours. Let us take an example of apple support, I do not know, how to run apple support? But I think I am guaranteed to get an answer with 24 hours” Interviewee D, ESGF Developer and System Administrator.

Similarly once asked: “How important do you see the time aspect in user support i.e. covering different time-zones?”

The same interviewee replied: *“If we have a user support we should have it for 24 hours, 7 days a week. If a user gets an answer within 3 to 4 hours, that would be better. That’s why covering different time-zones are important. May be one can solve a problem in a few hours’ time but at least one should be able to tell the users that we are working on the problem” Interviewee D, ESGF Developer and System Administrator.*

Reply time after delegation of a user request from a support person to another support specialist

From the Figure 6.35, 68% of the support staffs believe that it might take a day or up to two days to get a response after delegating it to a peer support staff. 31% of the support staffs believe that it might take up to 3 hours to get a reply from the support staffs after the user-request is delegated.

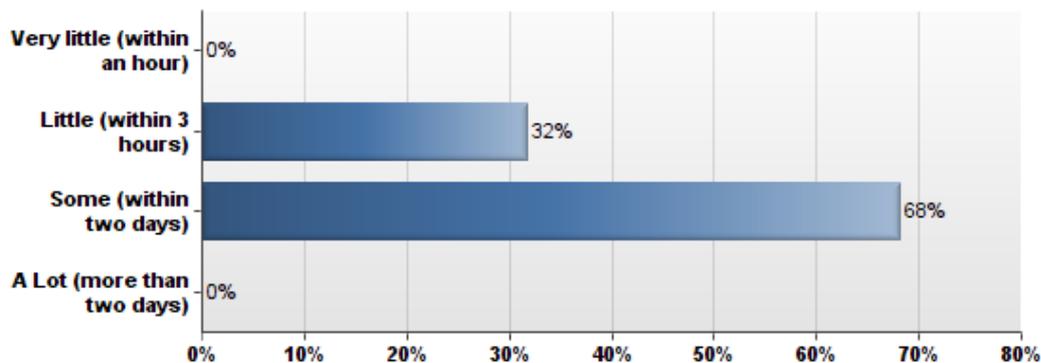


Figure 6.35 Support staff’s (average personal) reply time.

6.8.3.2 Request solution time

43% of the total incoming user requests (i.e. user problems) can be solved within an hour, as it is evident from the Figure 6.36. Now the user requests that can be solved within five minutes can be classified as simple problems (or petty requests with respect to time). The user requests that are solvable within 1 hour can be called normal requests with respect to time. Up to 28% of the user requests can be solved within 3 hours. These user requests can be classified as difficult problems (or time consuming problems). The

solution of up to 29% might take full day or may take longer than 1 working day. These can be classified as very difficult problems (or very time consuming problems).

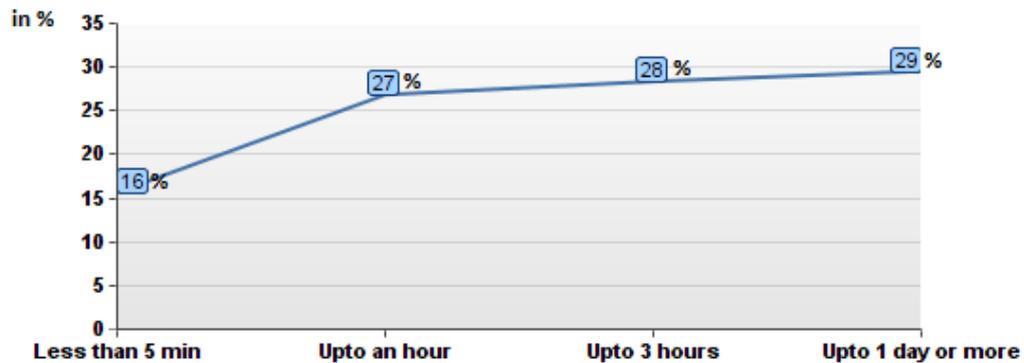


Figure 6.36 The percentage of request resolution time of incoming user-problems with respect to time frame.

6.8.4 Documentation of user requests

A very interesting question that can be asked is that how the incoming user requests are managed in the user support process in ESGF. The information about the current state of the documentation level of the incoming user requests was revealed by the survey-questionnaire. According to the results of the survey, mostly the incoming user requests remain undocumented. At the same time, the requests are neither centrally archived in the form of a central storage area e.g. a database or a central document, nor a statistics is maintained about them.

6.8.4.1 Handling of user-requests in a non-documented way

86% of the respondents, as according to responses collected from survey questionnaire do not document the requests, see Figure 6.37. Consequently, in the current user support process of ESGF e-Science infrastructure and its associated projects, there is no practice of documenting requests.

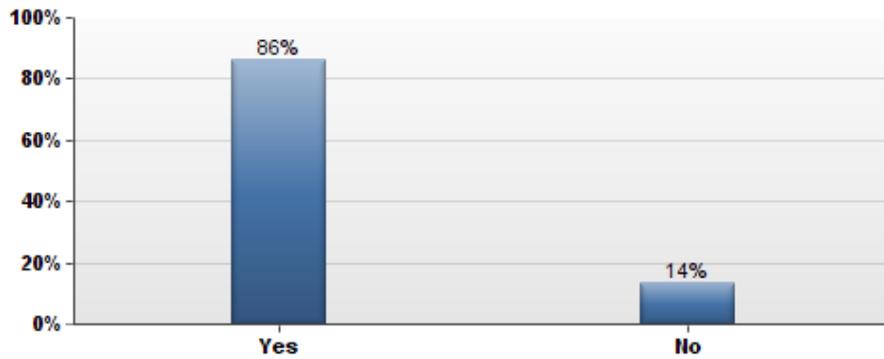


Figure 6.37 Support staffs (shown in percentage) who handle incoming user-requests in a non-documented way.

6.8.4.2 Frequency of non-documented requests

The responses of the survey-questionnaire reveal the frequency of user requests handled by the respondents in a non-documented way per-month in percentage. Up to 10% of all the incoming requests are undocumented as according to 40% of responses, see Figure 6.38. Furthermore, according to 30% of the responses the frequency of undocumented requests lies from 11% to 40%. However, according to 30% responses, the frequency of undocumented requests lies above 40%.

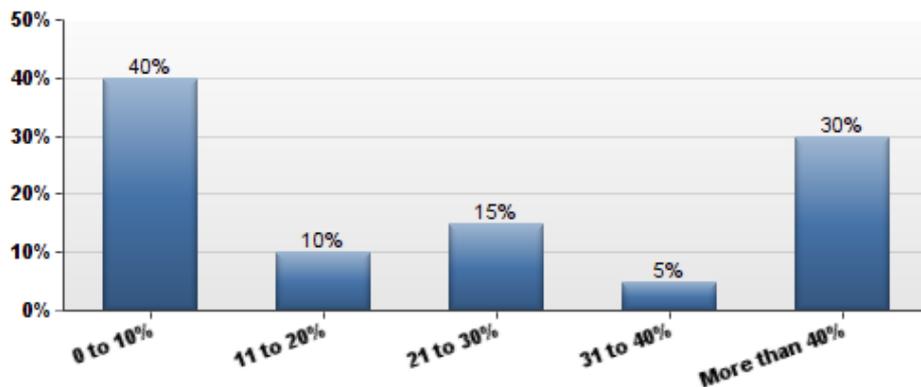


Figure 6.38 Frequency of user requests handled in a non-documented way in a month.

6.8.4.3 Communication channel of undocumented user requests

Almost 86% of the support staff as according to responses collected, do not document the requests. This evidence is collected from Q. 26 in the survey questionnaire; see Section 6.8.4.1 (page 177). It is important to know the different forms of communication via which non-documented user requests are received by the support staff. There are four main channels that contribute to unrecorded user requests, as according to Figure 6.39:

1. 29% unrecorded e-mails
2. 18% mailing list which is undocumented
3. 29% telephone (unrecorded) and
4. 22% walk-in (unrecorded)

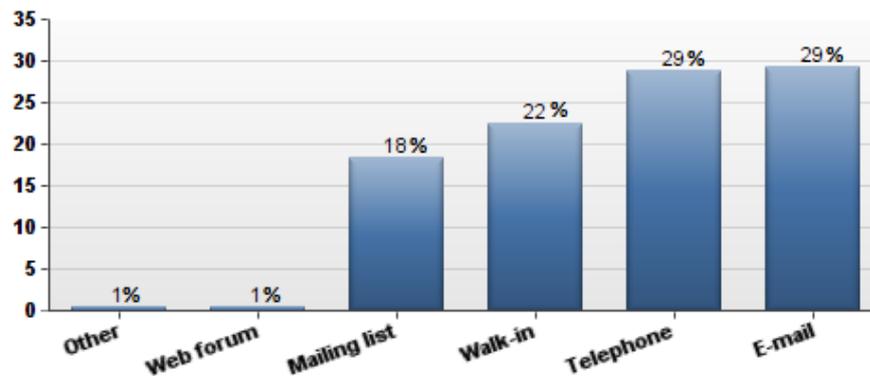


Figure 6.39 Communication media used (given in percentage in this figure) for undocumented user requests.

Under the “other” category, skype was specified by a respondent.

6.9 Satisfaction level

In this section the satisfaction of the ESGF staffs in relation to their work of supporting users with respect to user support process is highlighted. The survey-questionnaire was used to gauge the satisfaction level of the ESGF staffs involved in user support as well as what they think about the satisfaction level of the users in the form of seven-scale. This phenomenon describes whether the people are fine with the user support process functionality and usage of the system.

6.9.1 Satisfaction level of employees

The ESGF staffs who are involved in supporting users were asked in the survey questionnaire about their satisfaction level of the user support process in ESGF and its associated projects as shown in Figure 6.40. The respondents of the questionnaire were

mostly “to some extent satisfied” as they picked the choice of response *somewhat satisfied*, 36%. The response “somewhat satisfied” was followed by the response “satisfied” as the second highest percentage, 23%. No one selected “very satisfied”. However, 9% of the responses of the ESGF support staffs selected “somewhat dissatisfied”. 5% of the support staffs are “extremely dissatisfied” and “dissatisfied” respectively. This means at least that on the average the ESGF staffs involved in supporting users are not “*completely satisfied*” with the current user support process.

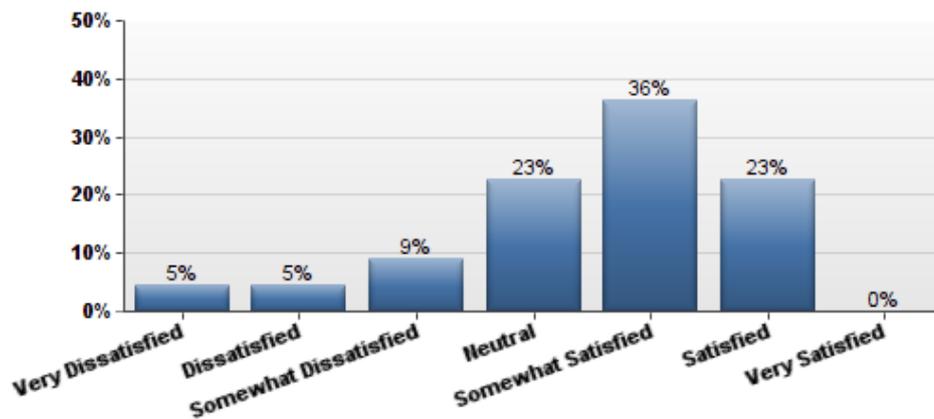


Figure 6.40 The responses of the support staff as far as satisfaction factor of the support staff is concerned.

6.9.2 Satisfaction level of the users as according to the employees about the user support process in ESGF

The satisfaction level of the users in the eyes of user support staffs has been measured in the questionnaire using a seven Likert scale. The responses are shown in Figure 6.31. From the Figure 6.31, it is obvious that most of the users of ESGF and ESGF-like infrastructures as according to the observation of the respondents (ESGF employees who support users) most of the users are satisfied with the user support process, 59%. 23% of the users are “some-what satisfied”. 14% are neutral. Only 5% are “some-what dissatisfied”. It is important to note that none of the end-users are “very much satisfied” with the current user support process. This implies that there is a room of improvement in the current user support process.

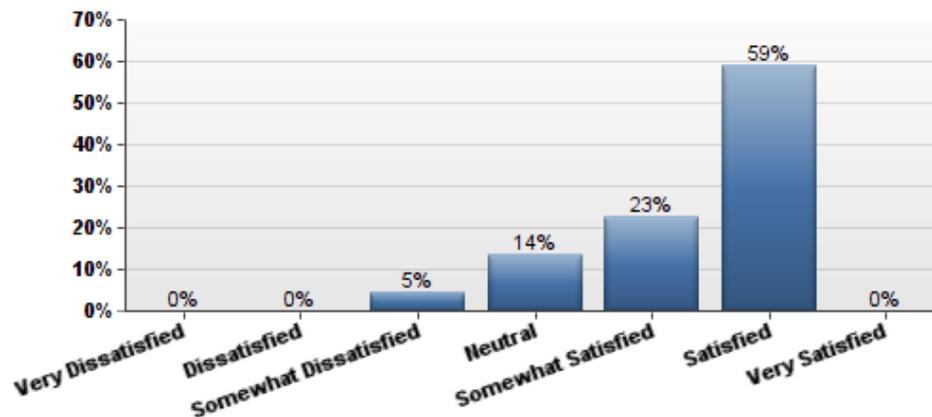


Figure 6.41 The satisfaction level of end users as estimated by the support staff.

6.10 Service level agreements

Technical people are against placement of SLAs unless there are enough resources SLA's might not be implemented.

“I’m against SLA’s; one has to put money behind SLA. SLA was mentioned from BADC. If you have operational resources that allow you to do SLAs, then yes. At the moment best effort from employees is in place of SLA. It is hard to measure best effort” Interviewee D, ESGF Developer and System Administrator.

However, users and other stakeholders of ESGF are for placement of SLAs.

“We need to implement SLAs to make services of ESGF reliable” Interviewee E, Principal Investigator.

“We want SLAs” Interviewee F, Master User.

“SLA can improve user support service of ESGF” Interviewee J, ESGF User

“Implementing SLA is an important step towards making user support better in ESGF” Interviewee K, ESGF User

6.11 Criticism of the current state of user support

In the current user support process in climate e-Science infrastructures, there is no assignment of user support roles. Consequently, top computer scientists end-up performing the duties of the front desk, in parallel to their main (core) duties. E-Science infrastructure employees already have a daily plan of action “p” in their memory i.e. doing activities such as software development, data curation, writing documents and others. So, the planned tasks are known beforehand. The incoming events “e” or event tasks, which are user’s or other employee’s requests that can be triggered at any point in

time, are actually interruptions to their daily work plan of action. Therefore, the incoming event tasks may affect the routine of employees, affecting the planned tasks. As a consequence, it requires e-Science infrastructure employees to allocate more time to process user queries or alternatively dedicating less time to their core tasks. This is evident from Section 6.5.2.1 (page 149). Employee's contribution to process user queries is not formally recognized by the e-Science infrastructure management in such a way that incentives are provided to the employees who process user requests. Additionally, there is no culture of users helping other users.

There can be two types of event tasks: First, explicit event tasks, meaning that task performed on the request of a user or an employee i.e. responding to users' or other employee's e-mail received. The second type is an implicit event task which is performed by employees based on their own cognizance (awareness); for instance; updating the web page by inserting relevant information for users or inserting an alternate solution in the solution repository after its discovery. Currently a central solution repository is not existent in the user support process of ESGF.

There is no priority rule existing in the current user support practices in ESGF. So, all of the planned tasks and event tasks are needed to be prioritized and there is no priority rule existing in the current user support practices. The current priority is based on employee's own judgment, which is noteworthy to be investigated further by asking employees. After prioritization of tasks, tasks are then actually performed within a specific time using certain means. The timings and the means to service user requests are generally not surveyed or recorded in the user support process of climate cyber-infrastructures. The current practice of time and means to process user requests varies from location to location of the federating partners that make up the ESGF consortium.

Since there is no central storage area in the form repository or database to record user requests and the solutions, consequently, the solution to the redundant enquires or enquires of similar nature cannot be known or matched via a central information retrieval system. Though, help webpages are available at the ESGF and its associated web links, however, a central information retrieval system is not available to the users or to the employees where relevant information can be sought by browsing the categories of redundant queries and their solutions. It is in the interest of users as well as employees of ESGF and ESGF-like infrastructures that a knowledgebase and information retrieval system may be designed, developed and made available online. At first, it seems a lot of effort and it may require resources but in the long run it will benefit the users and employees both.

From the survey results, it is an important observation that not all user requests in the form of questions are answered by the employees. There is no ticket assignment mechanism in the user support process, partly because of the multiple channels of

communications and partly because of lack of central ticketing system. Moreover, some of the user requests might be over-seen due to lack of central ticketing system. At times, the user support staffs that are scattered world-wide may rely on other colleagues and may not find responsible to answer or even to take all the user queries. Consequently, keeping in view the current arrangement of user support process, no one feels answerable, which is not a positive aspect of the current user support process in climate e-Science infrastructures.

In a nutshell, there is a process of user support which exists in the climate cyber-infrastructure but it is not formally defined in the e-Science infrastructures in such a way that it is measured in terms of quality, cost and standardization. Its resources are not explicitly allocated or it is organized as per wishes of e-Science infrastructure users. So far, there was no survey conducted to capture and analyse users' wishes or to measure usability and accessibility of UI of ESGF and ESGF-like e-Science infrastructures and users' help portals. There is no process owner(s) of the user support process and in the ESGF governance scheme there is no explicit regulated process about user support process in ESGF.

Unfortunately, there is no categorization of the incoming user queries and no information is available about the categories. Therefore, all the incoming user requests are answered based on the experience and knowledge of e-Science infrastructure employees. There is no document, repository (database) or an information retrieval system available where the user request categories can be stored for future reference in order to avoid redundant effort of working similar user requests. Similarly, there is no user self-help culture where the users may look at the self-help retrieval system to get the initial information about the problems they are facing. After the analysis of the survey-questionnaire, some of the categories of user requests were proposed. These categories are given in Section 6.8.1.9 (at page 169).

In an interview, it was asked about the deficiencies in the current user support process in ESGF, the interviewee replied:

*“I think **user support is very critical to the success of the project**. Unfortunately, the current user support system is pretty deficient. I think we need to do much better. My personal role is that I have been trying to help users as much as I can. For the longer period of time I have been answering 90% of the questions. I personally think that **we need to have a much better model for answering user questions**. Right now it is up to the goodwill of the individuals to answer (user) questions. I think we need to have more formal model for supporting users” Interviewee D, System Administrator and ESGF Developer.*

In the current user support process, users do not provide feedback and suggestions to improve the whole process. Moreover, the users do not provide requests for change in

the infrastructure and its associated processes. However, users do complain and these complains infer to a change in the ESGF e-Science infrastructure. These complains normally lead to development of new features in ESGF.

“We didn’t have users asking for a change in the infrastructure but we have users complaining about the infrastructure. People have said; like, can I get an FTP” Interviewee D, System Administrator and ESGF Developer.

Currently, changes mainly come from internal and external discussions of ESGF staffs participating in committees and meetings. Users are not the major contributors of the changes in the user support process as well as in the ESGF e-Science infrastructure (and its associated projects) as evident from the following excerpt:

“Users are not initiators of changes. Changes typically come either from internal discussions, or from the developers or from other projects like heterogeneous data project who want ESGF to go in one direction. I do not think that users have enough knowledge of the infrastructure to suggest changes” Interviewee E, Principal Investigator.

The main observation here is that since the e-Science infrastructures are created for the users of the infrastructures, how the participation and users of these infrastructures can be totally overlooked from the development of the e-Science infrastructures. The users must be included in the developmental process of ESGF that may lead to changes in the e-infrastructure. Users can be motivated to provide suggestions for changes that can further be discussed by the ESGF staffs in the committees present in the governance system of ESGF.

Similarly, there is no mechanism of rating solutions for the users or the staffs because the communication media used in the current user support process do not allow such features. Very few users are currently commenting on the problems of others and get involved with the correspondence with other users. The solution’s time of a user request, employee’s response time to a user request and response time between user support staffs, in case the request is delegated, is not uniformly defined and agreed upon by all the stakeholders. Consequently, suitable service level agreements (SLA) can be framed and applied to regulate the current user support so that the employees know that they have to deliver some quality of service to the users within a specified time frame. Plausible SLA can be formulated by forming a consensus between executive committee (EC), the technical committee (TC) and other stakeholders of ESGF e-Science infrastructure. The initial SLA recommendation is detailed by the author in the forthcoming chapter.

This research project was seen by most of the staffs (and other stakeholders) of ESGF, as a welcome gesture, about the steps taken by author towards performing research to progress user support process. Furthermore, it was expected that this research leads to a

better user support process and perhaps better user support system design for the staffs as well as the users as seen in the following interview excerpts:

“We are going to build a subsystem for looking at support which means that technical people have to speak to you what they want in a survey. You let us know what things are important to capture from survey point. Keep thinking about these kind of questions because that we can use (your research recommendations) for the subsystem and then we can see if we may get resources to build it. Think about, what do you ideally want the support system to tell you; what other features are important and needed? How far can we get modernizing? We create new software to build it. Support stuff depends on the evolution of infrastructure. Base lining us time series of support, the way it looks like, no. of types of users it supports” Interviewee G, Web Developer.

The missing gap in the ESGF is to strengthen the user support process, as pointed out by the chairman of ESGF that is needed to be covered:

“So, the users came in later after we had a product. Again we had converted the system from the research environment to production environment. Then we started putting together FAQs. One of the thing that we didn’t do, that might be useful, actually having someone active sitting together with the users, beginners, intermediate users and advanced users and looking over their shoulders and getting to know how do they operate ESG. And analyze those gap analyses that are missing. If your study can do that that would be awesome” Interviewee L, ESGF Chair.

Previously there hasn’t been any survey or study conducted in ESGF or any other climate e-Science infrastructure that would have involved users:

“We haven’t assessed how we are doing to know exactly what is needed to be done (in user support of ESGF) and how formal it has to be, whether it will take lot of resources or just nominal amount to improve things” Interviewee B, Climate Scientist.

Similarly, no survey has been conducted previously to assess user support in ESGF and other e-Science infrastructures of climate science domain. The survey and the interviews conducted in this research are of significant nature because these data collecting instruments shows the current state of the user support process and at the same time show the areas that are needed to be improved.

Once asked to an interviewee about whether someone knows if a user survey was collected previously it was stated:

“The CMIP panel put together a survey about CMIP5 (data-sets) and send it to modelling centers, more the WG1 type scientists. It wasn’t directed at the broad user base” Interviewee B, Climate Scientist.

The monetary aspect of financing user support activities is needed to be framed in the form of a working user support process that has the minimum financial resources

available. Currently, it is very hard to know the financial expenditures on the current user support activities as no estimate has been done. The cost model is hard to find out, as the author found it hard for him to get access to the documents relevant to funding of user support activities. The cost model of the current user support was missing in the current scenario:

“What missing in here is the cost model, one thing happening right now is that there are some of the projects that are not giving us any money but their user community is asking this questions and they are getting answers to it but they are not contributing any funding. The folks who are funding projects like; the CMIP5, is being used by others. That’s not fair! But how do you know how much to charge...” Interviewee L, ESGF Chair.

In an interview; the author asked about the funding, from where the money pumps in and how much is reserved for user support activities or shall be sufficient for it.

These projects, these folks are paying money for the ESGF. Those are the people who are primarily incharge of ESGF because they are giving the money. Who has more say ... more rights. If DOE is putting in 2 million dollars and others are putting in 200,000 \$ then who has more say. They are the project, stakeholders and shareholders” Interviewee L, ESGF Chair.

6.12 Chapter summary

Chapter 6 is about the outcomes of the empirical study, it shall serve to communicate them to stakeholders and interested parties with the help of graphics, tables, charts and text. Problems within the current user support are identified and in the end, a criticism of the current user support process is given. The data sources from which the findings were derived in this chapter are mainly: Survey questionnaire, interviews with the ESGF stakeholders and participatory observations. In addition, analysis of mailing list and a field study was conducted as well; however, they are not part of this dissertation.

In this chapter, the empirical results communicated to the reader cover in detail: Attributes of the user requests, media of communication between users and support staffs, time factors, satisfaction level of employees as well as users, service level agreements and documentational aspects of user requests. The empirical findings about the current user support system are represented by using soft systems methodology (SSM). The reason for using soft systems methodology is because all the stakeholders can understand, what is going on as far as user support service is concerned.

Soft systems methodology provides a holistic perspective of the whole system including the soft elements such as humans involved in the system. A conceptual model of the current user support is represented in the form of a rich picture and it explains the

structure and function of ESGF user support. A rich picture mind map is helpful in understanding the complex situation of a user support in a federated environment.

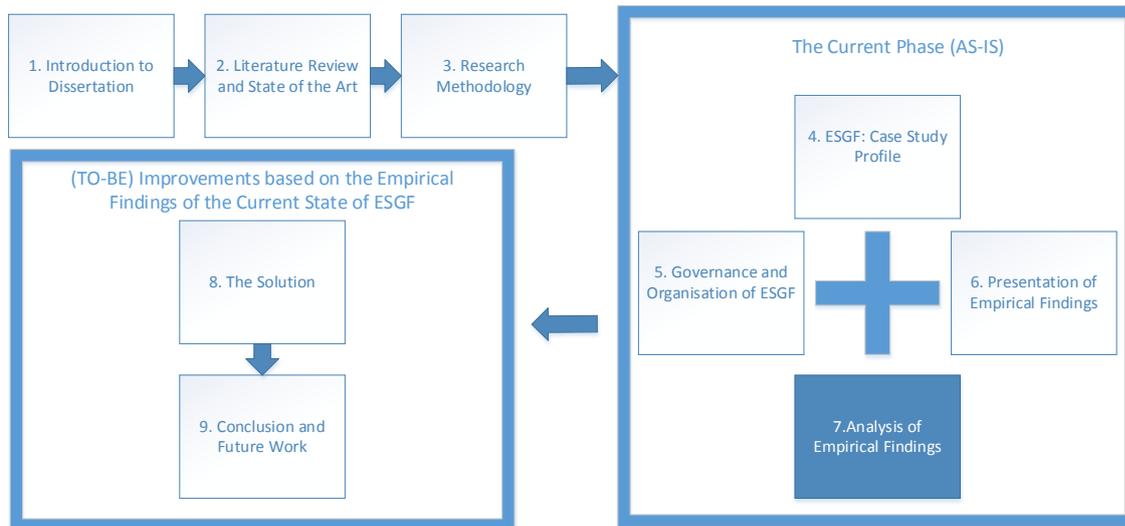
The key channel of user and employee communication is e-mail, as according to the survey-questionnaire; 78% of the incoming user requests are received as e-mails. The e-mail communication can be further separated into e-mail delivered via Request Tracking System (RTS) and via mailing lists (ML).

According to the survey questionnaire, the most common problems encountered in the current user support of ESGF were: Data access and data download problems, user authorization and authentication as well as registration problems. Neither a central repository for stored requests nor an information retrieval system exists in the current user support process of ESGF. So currently, there is no documentation about user problems.

The percentage of solved requests is around 77% as according to the survey-questionnaire and the mailing-list analysis. Thus, 13% of the incoming requests are abandoned or not pursued by the ESGF staffs, which is an alarming number. Online support material is available, but its usability and understandability are not determined yet.

Amongst the support staffs, no formal role of user support manager in e-Science infrastructure organization of ESGF is assigned. Any climate science e-Science infrastructure employee from any part of the world can jump in and answer a user request and provide solution to user's problem. There are no positions in user support defined by the organization or supported by the sponsors. The next chapter presents an analysis on the empirical findings presented in this chapter.

7 Analysis of empirical findings



From this chapter the following peer-reviewed papers were accepted:

- Hashim Iqbal Chunpir, Thomas Ludwig, Amgad Badewi: *A Snap-Shot of User Support Services in ESGF: Advances in the Human Side of Service Engineering. AHFE, 2015*
- Hashim Iqbal Chunpir, Thomas Ludwig, and Amgad Badewi: *Using Soft Systems Methodology (SSM) in Understanding Current User Support Scenario in the Climate Science Domain of Cyber-Infrastructures. HCI (10) 2014: 495-506*

7.1 Introduction

The empirical results presented in the previous chapter are further discussed and interpreted in this chapter. Based on these results, this chapter lays the foundation of “to-be” state of the user support process in the ESGF e-Science infrastructure and its associated projects, after analysing the “as-is” state. The to-be state is suggested in the next chapter that provides indications and recommendations to enhance the current user support process; however, this chapter provides the basis of to-be by analysing as-is, i.e. building upon the results of the empirical investigation mentioned in the last chapter.

This chapter is composed of the following main parts;

- the analysis and interpretation of the current empirical results i.e. the *as-is* state in the ESGF e-Science infrastructure, given in Section 7.2,
- the evolution of the user support process especially the communication media in the last few years; i.e. since this study started, provided in Section 7.3 and
- the major problems faced in the user support process, mentioned in given in Section 7.4.

7.2 Analysis of the current empirical results

In this section the analyses and interpretation of the survey results presented in the last chapter is given. The attributes of people involved in supporting users is given in Section 7.2.1. The expectations of users from the perspective of staffs of ESGF are given in Section 7.2.2. The satisfaction level of users and staffs is provided in Section 7.2.3. Similarly the structure and model of user support in ESGF is described in Section 7.2.4, the attributes of user requests are covered in Section 7.2.5, documentation aspects of user request is treated in Section 7.2.6. Communication media in user-support is presented in Section 7.2.7. Time related factors, strengths and weakness of the current user support process and finally discussion is covered in Sections 7.2.8, 7.2.9 and 7.2.10 respectively.

7.2.1 Attributes of people supporting users in ESGF

In this sub-section: The main characteristics of people i.e. the staff involved in the ESGF and the associated projects are discussed.

7.2.1.1 Sample description and introduction of basic social biographic data with self-rating about their expertise

In this study, 25 respondents participated, which form a well-defined population of the staffs who are supporting users. The total 36 of respondents started the survey-questionnaire out of which only 25 responses were useful. The audience of the survey was support staffs that are located at different institutions, either FLS or SLS who support multiple e-Science projects in parallel. In HCI studies with as few as 12 participants or users are not uncommon, results with 20 or more users are however, more convincing (pg. 373 Lazar, Hocheiser).

The major proportions of the respondents are engaged with the ESGF project, (see Figure 6.16). The role of FLS and SLS is not hardcoded in the user support process of climate e-Science projects (Figure 6.21). The respondents are qualified and sufficiently experienced with the support activities.

From the Figure 6.15, it is evident that the current user support staffs in e-Science projects is dominated by the technical experts followed by some staffs who also possess domain knowledge. Currently, there is no involvement at all of pure domain scientists. The questions asked by the users are mainly of technical nature; however, there are questions of scientific nature being asked as well. Therefore, it is worthy to have a collaboration and exchange as well as knowledge sharing between the technical experts and the pure scientific experts. Scientific experts are also known as domain experts. Consequently, in the current user support process of ESGF, there is a need to incorporate scientific experts.

It is worthy to note that the respondents of the survey-questionnaire, who are staffs involved in user support are involved in different activities which are not directly related to user support, see Figure 6.14. The activities popular amongst the support staff are: Software development (68%), writing documents i.e. project deliverables (68%), which are the major activities i.e. most popular activities amongst staffs that occupy their time. The other activities include system administration (60%), research work (44%), administrative activities (40%) and data related activities (40%). This implies that respondents mainly dedicate their time to these core activities apart from the user-support activity. If the user support process is streamlined to become more efficient and smoother then the support staff will have more time for these activities, hence their productivity might increase.

The issues mentioned in the user requests can contribute to the other domains. For instance; a programmer knows about particular software and can support in this regard. The user feedback helps to improve the software.

7.2.1.2 Time dedicated for support activity (Q. 10)

The average dedicated time for the support activity ranges from one hour to 3 hours on the average out of 8 hours of working day, which is almost up to 35% of the total time invested in support activities, see Figure 6.17. If the practice of self-help is introduced by creating accessible support websites and Wiki's for the end users then the average time for the support activities can be reduced. This observation is similar to a study conducted by (K. Y. Leung & Lau, 2005). This would mean that the time can be dedicated for the other activities other than support activities. Since the support activity is on-demand, it is important that the support staff sets his or her particular routine for servicing the user-requests. Since the support staff is not a dedicated support staff, it is important for support staff to manage support activities along with other activities that he or she is taking care of.

Suggestion: Time dedicated to the user support activities can be made shorter by doing the following:

- A norm can be created where a standard could be set (in future) where a support staff has to plan or reserve part of his or her time for support activity.
 - This activity can be formally duly rewarded and recognized in the ESGF. Some priority, recognition or value should be assigned to the user-support time in a working day where one can give more credit to the support activity performed by the user-support staff, as other activities that the support staff may be performed in parallel.
 - It is important to keep support activity productive and alive. For this reason, Users Support Workers Activity Model (USWAM) has been suggested by the author. USWAM aims to improve the task processing of the people who are
-

supporting users with their questions to interact with e-Science infrastructure such as ESGF. Further details on USWAM are not part of this thesis. It is, however, interesting to note that USWAM can be integrated into the FeUSE framework.

- Furthermore, it is also suggested to involve users in the user support process and motivate them in helping other users.
- Self-help mechanisms such as Wiki, FAQ, videos.
- Assigning roles such as FLS, user support process owner (USPO), user support coordinator (USC) to the staffs dealing with user support of e-Science infrastructures like ESGF as suggested in the FeUSE framework.

7.2.2 User expectations from the respondents` perspective

The attributes of users from the respondents` perspective (i.e. staffs of ESGF as well as its associated projects) are as follows:

7.2.2.1 Users base (Q. 20)

In ESGF and associated climate e-Science projects the number of end-users supported by the (distributed) support team varies. Most of the respondents believe that the users base (total number of end users) range from 50 to several hundred. This means in this case personal relationship can be made with the users.

The responses differ tremendously. This is partly because there are no user-statistics maintained (Q.39) and it could be based on a vague estimation of the respondents. However, it is important to consider that the responses given by the support staff vary because some support staffs act as front-line staff (FLS), others act exclusively as SLS and yet others act as both FLS and SLS. Moreover, there are some SLS respondents who are not aware of all users-base because they do not interact with all users. Finally, different e-Science projects are associated to ESGF and each project has a particular users` base. The statistics about users of the project have not been maintained therefore the accurate user-base of the users of ESGF is not known.

From the literature, there are almost 27,000 registered users in ESGF (Dean N. Williams et al., 2013). However, it is important to get a more insight and investigation into the users` base e.g. by formally collecting user statistics.

7.2.2.2 User recommendations (Q. 30)

It is essential to note that the users normally do not provide recommendations. Only very few users provide suggestions such as proposals, use cases etc.

Suggestion: It would be helpful to get the suggestions from users and encourage their participation for a number of reasons: Firstly, their feedback can improve the support process. Secondly, one can get an insight into their level of satisfaction. Thirdly, they

may rate the solution suggested by the support staff and finally since the number of users are not so high, a personal interaction may create mutual trust, empathy and better relationship amongst end-users and the support staff.

It is vital for the continual enhancement of user services to have such a possibility included in the user support process where end-users can submit their recommendations or feedback and rate the solutions to their problems or requests at least. Ideally the users might be willing to help other users. This will enhance the interaction with the support staffs as well. In ESGF, the users generally want to give feedback, as according to interviews with the stakeholders, especially users. Consequently, *Askbot* platform was deployed in ESGF to capture users' feedback. The feedback from the users was made possible via the *Askbot* platform that included comments like "correct solution" and others.

7.2.2.3 Collection of user statistics (Q.39)

Most of the support staffs (80%) do not collect statistics, as shown in the Figure 6.37. The user-statistics are important, in order to know:

- For maintaining a user-base to track record of users coming from different continents
- To be able to measure the overall user satisfaction
- It is important to have statistics in order to measure total number of incoming requests versus resolved requests
- It is important to try different suggestions or workarounds to solve similar problems to find out a better solution etc. And to archive these solutions

Appropriate technologies or tools such as request tracking system (RTS) and other platforms can be applied to collect user statistics.

7.2.3 Satisfaction level

In the survey-questionnaire, employee satisfaction level was gauged as well as user satisfaction level from the perspective of employees was measured with the help of Likert scale.

7.2.3.1 Employee satisfaction level

From the Figure 6.40, it is clear that there is a substantial tendency to increase satisfaction of users. For the ESGF e-Science infrastructure to urge staffs to support users, it is significant that the staffs that support users are also satisfied with the support system.

It is interesting and worthy to undertake a separate research to find out the causes of dissatisfaction of users in ESGF: There are following open questions: What could be the reasons of support staffs dissatisfaction? What specific situations, behaviors, tools and

events lead to respondent's satisfaction and dissatisfaction with-in the user support process? What attributes of these encounters cause them to be remembered favourably or unfavourably?

There could be many reasons that could lead to support staffs dissatisfaction e.g. as encountered previously in the relevant literature (Heckman & Guksey, 1998):

- Lack of user statistics
- Lack of cooperation from a peer support worker
- Lack of training and knowledge
- Lack of user cooperation
- Bad support tools and support procedures constituting the user support process
- Unaccredited work (low paid remuneration with no promotion chances)

One apparent reason is the uncredited work of the people involved in supporting users of ESGF. According to observation of the author, sometimes lack of cooperation from the peer support workers and users might cause dissatisfaction that makes the user support an unattractive task.

7.2.3.2 User satisfaction level (Q. 37)

According to the Figure 6.41, it is evident that most of the end-users i.e. 59% as according to the observation of support staff's perspective are satisfied with the support process. However, there is a room for further improvement in terms of satisfaction because no one is very much satisfied with the user support process, which is similar to the satisfaction-level of support staff. It is important to note that not even a single person responded that end-users are very satisfied. This further strengthens the belief that more can be done.

Suggestion: An attempt should be made to improve further the satisfaction level of end-users so that they are very satisfied with the support process.

End-user satisfaction can be improved by considering following improvement possibilities or factors (Heckman & Guksey, 1998):

- Reliability
 - Responsiveness e.g. too much time consumption or delay till the actual solution or support provided might lead to dissatisfaction.
 - Empathy and other psychological factors
 - Assurance: the knowledge and courtesy of support people and their ability to convey trust and confidence.
 - Interpretation of other communication materials (tangibles). Are they effective, interpretable and well-written?
 - User-support websites and other resources should be regularly updated.
-

7.2.4 Structure and model of user support in ESGF

The structure and model of the user support is given as follows:

7.2.4.1 *User-support model*

At the moment, the current user-support model used by the respondents is quite a mix, as shown in Figure 6.19. It is interesting to note that due to distributed nature of e-Science, the user-support model is also distributed. Very few users i.e. only 18% rely on e-support, see Figure 6.19. Consequently, e-support model is not much in practice in ESGF and associated Climate e-Science projects. The reason for less reliance on e-support can be partly the distributed and scattered user support information on the websites as well as on the associated Wikis of ESGF. On top of this, the information is not updated regularly by the staff members of ESGF.

It is interesting to note that ESGF has a centralized as well as decentralized support model i.e. two support models exist in parallel. The user-support model in e-Science is currently distributed at different locations. It needs Single point of Contact (SPOC), as much as possible. At the same time decentralized nature of handling user-requests is required because the incoming user requests are about certain components or concepts at a certain location in the federation of ESGF. Consequently, user-requests that need a special attention could be handled at a particular location by a concerned user-support team. There is no outsourcing model involved in support process of e-Science infrastructure, so, no outsourcing of incoming requests is done to any third party.

Since the current support model and organization of the current support process in climate e-Science is not uniform, it is vital to decide between centralized and decentralized model. This is an open question and can be tested in future. There are advantages and disadvantages associated to both of the models; however in user-support, in general, centralized model is generally preferred because of SPOC (I. Middleton, 1999).

The reason why respondents (i.e. staffs involved in supporting users) are using different support structures is that people are different and have different working styles, habits with different background and different experiences with support structure, therefore they have a particular style or way that they prefer. Consequently, in a federation like ESGF it very hard to stick to a single user support model; however efforts can be made in this direction.

Suggestion: E-support should be promoted in climate e-Science projects. A practice should be followed that the users try to get e-support by looking at the associated web sites and Wikis at first, to find the relevant information needed for self-help, as also suggested by (K. Y. Leung & Lau, 2005). Currently, there is a considerable potential to consolidate the user-help material and to make it more understandable for end-users.

Cook books sort of small recipe schemes can benefit users so that users can be able to find out the material that they require to interact with the ESGF infrastructure. In the meanwhile, e-learning materials such as learning games, avatars and other gamification techniques can be helpful to train users as well as staffs.

7.2.4.2 *User support structure and levels*

In the existing ESGF e-Science infrastructure user support process, there is no definite classification of the user support structure into particular N-tiers. There is a practice of two-tier system (41%) as well as one-tier (41%) support structure in parallel, as shown in Figure 6.20. However, sometimes SLS is getting questions directly too. Hence, there is no complete distinction between one and two tier user support structures in ESGF. Nonetheless, there may be a distinction in other support infrastructures in commercial or other domains.

Suggestion: A norm or a standard shall be developed so that the support units stick to a particular practice, as far as a user support model and a user support structure is concerned. From the observations and the collected data, it is recommended to have two-tier based user support model in ESGF.

7.2.4.3 *On-demand nature of user support*

The user support type in ESGF climate e-Science infrastructures is mainly on-demand, see Figure 6.22. So, for the users who are in Europe, for them the support people within Europe are in the relevant time-zone and the support people in USA are in a different time-zone. It is important to take care of the time-zones. Different time-zones do lead to problems such as delays in responses to the user requests as well as collaboration issues amongst the stakeholders of ESGF.

7.2.5 Factors related to user requests in the user-support process

The characteristics of the user queries coming to the ESGF support centre are as follows:

7.2.5.1 *Number of user-requests per working day (Q.21)*

According to recorded survey responses, the number of incoming requests from end-users per working day range from 2 to 5, in 2013 (see Table 6.4). Lower workload differentiates the help desks in e-Science to traditional help desks. In e-Science, user-support help desk is more like an ad-hoc support system where users call or contact the experts in case of technical support (Mckoen, 2000). This implies that at the moment, no big independent help desk or support unit is required, where the support staffs are only doing support activities mainly.

Suggestion: This means it is easier for a support staff to document the incoming user-requests properly for future reference as well as to put the incoming enquiries in the archive and maintain the solution archive for each important user-problem in a knowledge base. Currently, it is a good time for the ESGF partners to set milestones to establish knowledge base.

7.2.5.2 Staffs involved in handling user requests (Q.17)

In the e-Science support process, most of the incoming requests are handled by single support person (normally the first person who handles requests) which is a positive sign (see Figure 6.18). However, if the users are trained and have the knowledge of the sources such as FAQs and How To`s then incoming user-requests in general can be reduced and in particular, first level enquiries can be reduced too, thus disburdening the first level support staff. This observation is similar to Q.14 (see Section 6.8.3.2 page 176) as 16% of the incoming user requests are solvable within five minutes that means they are simple and routine queries. Similarly, in Q.18, delegations to multiple staffs are rarely done (see Section 6.8.1.7, page 166).

This section shows the distribution of user-requests in the form of a problem solved by number of people. The treatment of user requests by many support people shows the total number of human resources allocated to solve a user-request, therefore the treatment of user-requests might get expensive. If there are many people involved in handling a user-request it may lead to frustration because of possible time delay factor (Mckoen, 2000). In the user-support literature, it is also pointed out that handling of a user enquiry by too many support staffs (helpers) is a cause of dissatisfaction for the end users, if the problem is not solved in the end (page 78. Heckman, 1998).

7.2.5.3 Number of times employees' ask to attain clarity of user requests (Q.32)

The staffs of ESGF, who support users were asked in the survey questionnaire (Q.32) that how often do they write back to attain clarity of user requests. None of the support staffs said that they need to write more than two times (see Figure 6.24), which means that the incoming user requests are generally understandable. In case some user-requests are not understandable, after one or two times of further questioning by the support staff, they become clearer, which is a positive sign in the current e-Science support process. However, there is a potential to make the user requests even more understandable for instance by using a request pattern.

Suggestion: It is significant to make incoming requests clearer. As it is evident from Section 6.8.1.4 (page 163), only two respondents are of the opinion that they need to write back to the users to understand the user-request more often. However, by training and instructing end-users, clarity in writing user requests can be attained. A particular user-request pattern might be followed to make sure that all necessary information is

provided and the request is formulated by using clear words. The vague statements shall be avoided. If the incoming request is clear then a user support staff does not need to write back at all or only in seldom cases. The support staffs may also be trained to interpret the incoming user-requests efficiently. It is also recommended that the end-users be instructed to write a user-request clearly, this will result in saving time of staffs or other users who support.

7.2.5.4 Understandability of user-requests (Q.29)

There could be many reasons behind unclear user-requests, such as no instruction or training of users, problems with language expression, the way the users formulate request may not be comprehensible and some details may be missing or left out etc. One might define some parameters or patterns how to formulate a request.

Suggestion: A standard pattern of asking requests can be made and some specific elements about a respective problem must be provided in the request. In this way more percentage of incoming requests could be made understandable. Training users, a mechanism to rate incoming requests might help too. It is interesting to note that results of Q.29 and Q.32 (Section 6.8.1.3, page 162) are similar.

7.2.5.5 User-request pattern (Q.31)

Around 59% of users use a particular pattern of a request. But this practice is not followed all the times (see Figure 6.25).

Suggestion: A standard pattern of asking user-queries is recommended in the current user support process of ESGF. Some elements of the user-request must be specified in the corpus of the request about a respective problem. In this way more percentage of incoming requests could be made understandable. If a standard request pattern is used then unclear user-requests can diminish in number. Secondly, the incoming user-request can be made interpretable by automated tools such as CBR, CBA and others in a long term. It is strongly recommended to use user-request pattern in the user support of ESGF by the author in the FeUSE framework (see Chapter 8), however, this decision is needed to be taken by the policy makers of ESGF.

7.2.5.6 Possible categories of user requests (Q.33)

The possible classification of the incoming user-requests is a guiding instrument to know popular and vital categories of user requests and their respective solutions in the user support process of ESGF and ESGF-like e-Science infrastructures in climate science. Q.33 of the survey-questionnaire reveals the possible list of the categories to which a user-request might belong.

The first category is “*general information category*” that contains further sub-categories (see Table 6.5). One of the most popular sub-categories is “*Data: e.g. query*”

about the types of data sets available in the portal". "*Data access and download*" is the most popular category with-in the main categories. It is evident from the responses that the popular categories arise because of the activities: Login, search, find and fetch climate data performed by the scientists in these e-Science infrastructure projects.

Other possible sub-categories of the first proposed main category i.e. "general information" are suggested as follows:

- Authentication, i.e. login
- Model compilation, debugging, optimization, feature development, software installation, software usage
- Queries about correctness of the data which have been downloaded

Other possible main-categories as suggested by the respondents are as follows:

- Projects: IPCC AR5, CORDEX, CFMIP, LUCID
- Models: ECHAM6, MPIOM, JSBACH, HAMOCC, CCLM4
- How to publish data i.e. data publication?
- Getting logged in
- Miscellaneous
- Post-processing of data
- Data citation
- Quality control

Suggestion: The all suggested main categories and sub-categories can be included in the e-Support webpages as well as in a user support platform e.g. request tracking system and ask-bot forum (ask-bot is similar to stack overflow system), where user-requests can be accumulated and tagged with a particular category or a keyword.

Data access and download is the most popular category in climate e-Science infrastructures because the users are most concerned about "data access and download" queries. Similarly, related to the category "data access and download", users query about the "types of data sets available in the portal or a gateway" which is also part of general information. This is because the scientists search, find and fetch climate data in these e-Science infrastructure projects, which is the main purpose of the researchers to get data for their research projects.

Other important categories such as "*Registration*" i.e. queries about Login, password, about registration process e.g. where do I need to register? What benefits can I have after registration? All these details can easily be put in frequently asked questions (FAQs). Furthermore, it is important to note that category "*Scientific (Community) Support and Infrastructure*" is also an important one.

7.2.5.7 *Category mix of user-requests (Q.34)*

Figure 6.28 in Section 6.8.1.9, (page 169) shows the percentage of categories that users send requests about. Most of major queries are about data access (almost 40%), about general information (28%) followed by scientific support (almost 15%), which cannot be ignored. This category mix which is received by the support staff at the moment is very much similar to the results in the previous Section 6.8.1.8, page 167.

Suggestion: It is imperative to introduce sub-categorization in the following main categories:

- queries about general information
- queries about data access

The sub-categories of *data access* and *general information* should be clearly highlighted and made available in the e-support website, so that if end-users, especially novices lack the basic knowledge they can get information from an information portal. The novices may include laymen, policy makers or other people not from the scientific domain. The users shall have the possibility to be able to get faster introduction about information they require under a particular category so that they can solve their problems themselves via e-support. Moreover, it is pertinent to establish or introduce *scientific community support* in support webpages (e-support) in the form of FAQs or How To's as a separate section.

All people who are involved in supporting users shall be encouraged to suggest important categories of questions that they receive on a regular basis. The category list can be updated, revised and revisited as different types of user requests are encountered. FeUSE framework recommends having an agreement on the categories list.

7.2.5.8 *Percentage of unresolved user-requests (Q.40)*

An attempt should be made that minimum or none of the incoming user-requests remain unresolved. Most of the unresolved user requests are mostly features for further development in the infrastructure or errors that are needed to be discovered in an e-Science infrastructure. Currently, more than 14% of the incoming user requests are not pursued by the staffs as according to the mailing-list analysis conducted recently (H. I. Chunpir & Rathmann, 2015).

7.2.5.9 *Multiple delegation (Q.18)*

From the Figure 6.27, it is clear that multiple delegations (MD) are not always made in user-support process in ESGF and its associated climate data projects. However, there is a practice of multiple delegations in the user support process only to a limited extent. It is interesting to note that none of the respondents say that multiple delegations are “always” made.

Suggestion: Multiple delegations might be reduced if a *directory of user support specialists*, also known as user-support directory (USD) is maintained or a system automatically suggests to which staff the request might go. Moreover, if an incoming user query is represented visually in the form of a graph or a chart in such a way that from which source as well as location it was originated, who has treated it as far as FLS or SLS is concerned and to what people it was delegated to, then it can be a great feature.

Multiple delegations are useful if a single user-request contains multiple questions pertaining to different domains that can only be answered by multiple specialists. Multiple delegations might also be useful in reducing reply-time after delegation of user-request to SLS staff or a peer support staff because one of many staffs might reply quicker than the rest of delegated staffs.

Multiple delegations may turn to be harmful if it is delegated to more than one support staff. In this case confusion might exist that to which staff member the incoming user query is addressed and in the worst case none of the staffs might end up answering the query. Consequently, there is a risk that the user query might not be answered at all and for that reason the use of multiple delegations is discouraged.

7.2.6 Documentation of user requests

The attributes about the documentation of user requests are as follows:

7.2.6.1 *Handling user-requests in a non-documented way (Q.26)*

Most support people handle user-requests in a non-documented way, see Figure 6.37. As a result, the federated partners of ESGF do not maintain the records of user-requests.

Suggestion: It is helpful to document the important incoming user-requests properly by using a suitable request tracking system such as Git Hub or any other mechanism unanimously agreed by the stakeholders of ESGF. It is, however, not necessary to document *all* incoming user-requests as observed and witnessed by the support staffs in ESGF and ESGF-like projects. On the other hand, all user requests are needed to be logged and recorded, preferably in a central database or a central storage area.

7.2.6.2 *Frequency of non-documented requests (Q.27)*

The documentation of the incoming user requests is not maintained. Moreover, there is no consensus on the frequency of non-documented requests (see Figure 6.38). However, majority of respondents believe that frequency of non-documented user requests, i.e. unrecorded requests, is high.

Suggestion: It is important to reduce important user-requests which remain undocumented and unrecorded by using a federation-wide ticketing or request tracking

system. By using a tracking system, automated documentation can be generated about the recorded user requests.

7.2.6.3 Communication channel of undocumented user-requests (Q.28)

It is evident from responses to Q.26 in the survey-questionnaire; see Section 7.2.6.1 (page 200) that most of the support staffs (90%) do not document the incoming user-requests. Therefore, it is necessary to know the different forms of communication via which non-documented user-requests are received by the support staff. From the Figure 6.39, it is clear that unrecorded e-mails, telephone calls followed by walk-in and unrecorded mailing lists are the media which are used for undocumented user-requests.

Suggestion: A recorded mechanism should be followed, e.g. RT system or any other platform such as Git Hub, where all e-mails could be recorded for future references.

A practice should be adapted in the user support process, by following which, important telephone and walk-in requests could be documented for future purposes. Since the support staffs do not get more than five requests per day, as according to the responses from the survey-questionnaire, it is possible for support staffs to record the requests if needed.

7.2.7 Communication media in user-support

The attributes of user support in ESGF are as follows:

7.2.7.1 Updating user-support webpages (Q.19)

The update of user-support information on the user support portal i.e. e-support is not regular (see Table 6.3).

Suggestion: The update of user-support information on the user support portal(s) should be made regular. A Service Level Agreement (SLA) can be suggested, where the concerned support staff updates the information regularly. The information can be checked by other support staff or a usability expert to make sure that these communication materials are accessible, easy to grasp and are understandable.

The ESGF policy makers shall be able to set agreements for updating the user-support web-pages. The responsible staff in ESGF like: FLS, SLS, User Support Process Owner (USPO) can send notices to update the web-pages. In addition, USPO, who is incharge of the overall user support process, can set deadlines. Moreover, understandability of the material can be checked as well.

7.2.7.2 Language of communication (Q.22)

Most of the incoming user-requests are in English language (up to 72%).

Suggestion: Sticking to one language standard is recommended, especially for international projects like ESGF.

7.2.7.3 Medium of external communication with user (Q.23)

From the results, (see Figure 6.29 for more details) one can deduce that in e-Science, “e-mail” leading to RT and mailing lists is the mode of communication between users and the support staff. Therefore, it is clear that e-mail is the main mode of communication as opposed to telephone in some other organizations using IT help desk support (N. Leung & Lau, 2007, p. 77) (Major sources of user contact table 21 Lau and Leung 2007 Journal of CIS).

At the moment in e-Science projects as far as the author has experienced from his participatory observation and meetings with other support staff, in the support process both “Mailing Lists” and “Request Tracking Tools” are used in parallel. According to ESGF support staff they need mix of mailing list’s features and request tracking tool’s features, mostly advantages of both. Request tracking software is even preferred to the mailing lists for external communication with the end-users, see Figure 6.29.

It is important to know and weigh the pros and cons of both of the media i.e. mailing lists and request tracking system.

Advantages of mailing lists in relation to e-Science Support process:

- all stake-holders (including end-users, developers, policy makers, scientists, data publishers, data managers) can participate
- least expensive method
- feasible for distributed support units
- no licensing issues

Disadvantages of mailing lists in relation to e-Science Support process:

- no history track of user requests
- no assignment method to further support personnel
- no search mechanism for user requests
- no priority mechanism for user requests
- may lead to ignoring some of user requests
- no solution archive connection
- no knowledge base
- no privacy

Advantages of request tracking software in relation to e-Science Support process:

- history of user requests is tracked
 - user requests are assigned to a concerned support personnel
 - search mechanism available for user requests
-

- priority mechanism available of user requests
- user requests have a particular number and they cannot be ignored
- solution archive connection can be made
- knowledge base can be available
- privacy settings can be made
- FAQs and How To's can be updated automatically³⁷

Disadvantages of request tracking software in relation to e-Science support process:

- Limited participation of all stakeholders. One needs to create logins and accounts for each person who want to login into a request tracking software. It makes even difficult to assigns hundreds of logins to support staffs that are present at distributed locations.
- System administration skills are required to configure, manage RT systems.
- It could be expensive in terms of license costs and human resource cost
- Security problems may arise too if access at distributed location is allocated.

Suggestion: A suggestion would be to come up with some common platform or a model that would combine the advantages of both mailing lists and request tracking software.

In RTS, it is required that all users have access to RT system and can be able to categorize incoming requests and sort the incoming requests in different categories. The users should be able to do configurations or customizations in RTS according to his or her preferences.

7.2.7.4 Preferred mode of user communication (Q.24)

According to respondents, users prefer e-mails leading to RTS the most, as it was ranked the highest (see Figure 6.31 and Figure 6.32). It is noteworthy to observe that this evidence is similar to the evidence from Q.23, see Figure 6.29. The users just like support staffs prefer e-mails leading to Request Tracking Software (RTS) the most. Mailing lists are ranked second. Therefore, from these results one can deduce that in e-Science users do not prefer other media as a priority, only "E-mails leading to RT" is the popular mode of communication between users and the support staff.

The reason for using e-mail as a major medium of communication is the distributed nature of e-Science and secondly e-mail is cost effective and decoupled in time. All advantages of RTS are better suited for the support staffs as wells as the end-users that are the reasons why they prefer it to the mailing list. For more information, see advantages of request tracking in Section 6.8.2.1, page 170. However, for the stakeholders and e-Science infrastructure developers, a mailing-list is a good mechanism to observe and participate in a user and support staff communication.

³⁷This feature might not be available in all request tracking tools.

7.2.7.5 *Communication mode for internal communication (between support people) (Q.25)*

From the results, one can deduce that even in the internal communication in e-Science between user-support staffs, “E-mail leading to RT followed by Mailing List” is a popular media of internal communications between the support staffs as well, see Figure 6.30. This result is similar to the result of popular medium of external communication. It is important to note that even for internal communication e-mail leading to RTS and mailing lists are used in parallel within user-support staff teams based at different locations.

7.2.8 Time related factors in the user support process

7.2.8.1 *Automatic reply mechanism (Q.15)*

Majority of user support staffs, almost 80%, do not use automatic reply mechanism, see Figure 6.33.

Suggestion: It is essential to consider using automatic reply mechanism in case of a RTS. The automatic reply should preferably look like a personalized message and not a machine response.

In case of a mailing list, an automatic reply is not necessary since each mail reply to the list is distributed to every list member. In a web forum such as *askbot* the automatic reply must not be published.

7.2.8.2 *Reply time after delegation of a user-request to a support specialist or Second Level Support (SLS) (Q.16)*

68% of the staffs who are involved in supporting users believe that it might take a day or up to 2 days to get a reply after delegating it to a peer support staff, see Figure 6.35. It is vital in the support process to reduce the reply time after delegation to make support process faster and efficient.

It is interesting to note that none of the respondents are of the opinion that the response can be within an hour. This observation further strengthens the distributed nature of the user support process in climate e-Science infrastructure, ESGF.

The FLS person solves only the routine or simple enquiries and passes the complex questions to SLS. It might be hard to get the personal response from the SLS within a certain short time period.

Suggestion: The support process will improve if the reply time can be reduced by an agreement or introducing a Service Level Agreement (SLA) for delegated response. A standard is needed to be defined and the support staffs who receive the delegated user-request might follow the SLA or an agreement for a delegated response time.

In case a support specialist is not available then either multiple delegations may be made. In case of non-availability of a support specialist, a peer support specialist might be consulted who could respond to a delegated user request in absence of the targeted user-support specialist.

In case there is no other support specialist available other than the targeted support specialist, then there should be a mechanism in place to deal with this situation. One solution could be that the end-user is intimated by FLS that the targeted support specialist is not available at the moment and as soon as he or she is available again, the request will be addressed. These circumstances can be termed as “*indispensable support specialist circumstance*” that may prolong the response time from the concerned support specialist. In this case, the reply needs time and no alternate person can be assigned. An example of indispensable support specialist circumstance is queries related to data access via OPeNDAP¹⁵. There is only one specialist available that can help in this regard.

7.2.8.3 Personal reply time (Q.13)

70% of the respondents of the survey questionnaire claim that the personal reply by staffs to user-requests is sent after passage of around three hours on the average from the point when the user-request is received see Figure 6.34. An effort can be made to reduce the time of personal response. It also depends on the time zones for example if the user has forwarded a request in Europe then it is vital to reply him or her from a support staff working within European time-zones. The greater the time of personal response, the slower user support process would be and as a result lesser would be the satisfaction of users.

Suggestion: An SLA or a flexible quasi-SLA can be introduced in this case where up to one working day at maximum could be set as a standard i.e. the maximum time for personal reply. In case a concerned support staff is not available (e.g. on holidays etc.) then an ersatz or a substitute support staff shall be made available to service the user-requests in each support unit. This is very much important in terms of showing empathy (i.e. building user’s confidence in the support process) to the end-user because end-user might get frustrated if he or she does not get a personal reply.

7.2.8.4 Request solution time (Q.14)

Almost 43% of the incoming user-problems or user-requests can be solved within an hour (see Figure 6.36, page 177). These incoming user-requests can be classified as simple problems (with respect to solution time). Previous research by Knapp and Woch, Dawson and Lewis has confirmed that a majority of the incoming queries are simple and routine (Knapp & Woch, 2002; N. Leung & Lau, 2007). “Simple and routine” technical queries are technical queries that can be solved by a user if relevant

information is provided to the user; e.g. password queries, account queries, hardware and software enquiries.

An attempt can be made to reduce these incoming user-requests because if an adequate documentation or information is provided to the users then these requests can be reduced. Up to 28% of incoming user-problems or user-requests can be solved within a day time i.e. 24 hours. These can be classified as normal problems (with respect to solution time). The solution of up to 29% might take longer than one day. These can be classified as big problems (with respect to solution time). It is significant to note here that the classification of user-problems depends not on the difficulty to solve them directly but the time-frame required for solution of the incoming user-requests.

Suggestion: It is important to note that more than 40% of the incoming user-requests or problems can be solved within an hour. From the request solution time, it is evident that 40% of the incoming requests can be solved by an FLS. Moreover, it also implies that incoming user-requests that are solvable within an hour might require an initiative of users for self-help (i.e. to help them). A mechanism should be devised to encourage self-help, for example via support websites with sections in the form of Frequently Asked Questions (FAQs), How To's or other relevant techniques. The incoming user-requests (problems) of simple nature, that constitute the major chunk, can be made solvable by the users themselves, if this is not possible then by the FLS staff preferably without involving the specialists. This will reduce the human resource cost in the support process of e-Science infrastructures, as it is evident from the respondents (in Q.12) that support staffs (respondents) are also engaged in other activities in e-Science infrastructures.

7.2.9 Strengths and weakness of the current user support process in ESGF

The strengths and weaknesses of the current support process from the stakeholders of ESGF and associated projects as responses recorded by the survey questionnaire are as follows:

7.2.9.1 Weaknesses

After examining the responses of respondents (as recorded weaknesses of the current support process in ESGF), the weaknesses of the current support process can be described in terms of the following list:

1. Insufficient training of the support staff and users (1 response)
 2. Weak link to the scientific experts (1 response). This observation is similar to Section 6.5.3 (page 152), see Figure 6.15.
 3. Support activities in e-Science are normally unfunded and left at the good will or mercy of people. This makes it imperative to improve the support process for the support staff to make it time and effort saving.
-

4. There is no standardized way for future-referencing and documenting the incoming user-requests. This observation is similar to Section 6.8.4.1, Figure 6.37 and Section 6.8.4.2 (page 178), Figure 6.38 (from the responses taken from support staffs).
5. No Single Point Of Contact (SPOC): This observation is also evident from the Section 6.7.1, page 156, Figure 6.19.
6. Understaffing and lack of time with the support staff (5 responses). This observation is similar to Section 6.5.4, page 154, Figure 6.17 and Section 6.5.5, page 154, Figure 6.14, where support staffs are also engaged in the tasks other than support activities.
7. Two parallel support mechanisms (RTS and Mailing Lists) in place. Both do not work due to their disadvantages. This observation is similar to Section 6.8.2.1, page 170, Figure 6.29 and Section 6.8.2.2, page 171, Figure 6.30.
8. Support activities cannot be measured.
9. Response time is not optimal or longer especially when a particular support person (especially a specialist) is unavailable to provide a solution or to answer a user-query. This observation is similar to Section 6.8.3.1 (i.e. reply time after delegation), page 176, Figure 6.35.
10. Missing support information on the support webpage or support webpages itself are missing. Too few updates of support webpages. This observation is similar to Section 6.5.2.2, page 150, Table 6.3, page 151.
11. No ticketing system. This observation is also evident from the Section 6.8.2.1, page 170, Figure 6.29, where one can observe that institutions that are carrying out e-Science projects are using ticketing system only to a limited extent (up to 42%). In Section 6.7.1, page 156, Figure 6.19, one can see that the whole support of e-Science do not use centralized Ticketing system as a SPOC (Single Point Of Contact).

7.2.9.2 Strengths

After examining the responses of respondents (for recorded strengths of the current support process in the survey-questionnaire) the strengths of the current support process can be described in terms of the following list:

1. All stakeholders can participate in the user support process. This feature is evident from Section 6.8.2.1, page 170, Figure 6.29, because of usage of mailing lists where everyone can participate in the support process.
 2. Required information can be successfully communicated to the users. Therefore, most of the questions do get answered and hence the support process works for now (5 responses). This is true as also evident from Section 6.8.1.6, page 165, Figure 6.26.
-

3. Developers can observe the user communication in mailing lists and this gives them a chance to see user-troubles. This feature is evident from Section 6.8.2.1, page 170, Figure 6.29, because of usage of mailing lists where everyone can participate in the support process.
4. Distributed nature of support. Various locations different time zones. (Section 6.7.1, page 156, Figure 6.19)
5. More requests are resolved by a single person. (Section 6.5.5, page 154, Figure 6.18).
6. Flexibility and dynamism in the user-support.
7. Personal contact is guaranteed.
8. Competence and knowledge of individuals (see Section 7.2.1.1 and Section 6.5.3).
9. The advantage of RT system.
10. No involvement of automatic answering machines. This is evident from Section 6.8.3.1 (see automatic reply mechanism), Figure 6.33 and page 175.

7.2.10 Discussion and conclusion

This empirical qualitative cum quantitative investigation has revealed number of issues where attention should be paid to and are needed to be taken care of in the current user-support process in climate e-Science infrastructure projects. These issues include allocation of time, allocation of human resources, time to solve the user-problems, support tools, support structure and many others. There are some important suggestions that are needed to be considered to be implemented in the end-user support processes in climate e-Science infrastructures. The suggestions include:

- Defining roles of FLS and SLS explicitly
 - SPOC model should be followed in true letter and spirit
 - 2-tiers can be defined
 - Since the user support in ESGF is *on-demand* different time-zones can be covered by allocating FLS at least 3 at different locations that cover different time-zones
 - Connecting scientific experts for scientific support
 - Acknowledging the user support activities and rewarding the performers of the user support activities
 - Regularly updating the user support information on Wiki, FAQs and making it more understandable to the users. It can be presented in the form of recipes or cook books. With this effort e-support can be made popular amongst the stakeholders especially the end users.
 - Engaging users to help others
-

- Understandability of the user requests can be enhanced by using request patterns and in the long run it might be a milestone towards machine learning.
- All different possible categories of incoming user requests must be decided and documented. Similarly, solution to the known problems should be made browseable.
- The practice of multiple delegations is currently discouraged and at a single time one staff should be responsible.
- User-requests must be recorded and stored for future reference or statistics purposes.
- Agreements such as SLA and OLA are suggested to be made. So that the response time from staffs can be standardized.
- Notification in an absence of peer support specialist to another specialist via USD.

From this empirical study it is evident that many different communication media are in use simultaneously which lead to confusion amongst support staffs and end-users. The mailing lists are not good at resolving complex issues requiring follow-up and several support staff. On the mailing list some queries get missed as there is no way to assign a query to support staff. The request tracking system is too cumbersome for small queries and does not work well for a globally distributed support staff. We have trouble keeping everyone engaged in the request tracking system. Moreover, users should be educated in a better way about how to use the system. It is important to form a consensus on the communication medium or platform between the users and the support staff as well as within support staff in the support process. With the efforts of this study ESGF consortium has realized this suggestion and an initial consensus on using Git Hub has been done.

The whole user support process and user support system cannot be completely standardized quickly in one go but an attempt can be made to standardize it in such a way that it is acceptable to the whole stakeholders, especially staffs of the partner institutions of the federated e-Science infrastructure, ESGF. The rationale to consider standardization in the user support process is the value that the user support process and user support system can provide to the stakeholders of the e-Science infrastructure, especially user communities for their research areas and eventually science. Like in other systems, processes in industry and academia the standardization can be made via following good practices like ITIL. Similarly, it is expected that standardization will create effectiveness and efficiency in the user support process.

The categories of user queries are mainly the problems related to data access and download, around 80% and 50% of the problems are about the general information. Implementation of FeUSE framework proposed in the next chapter aims to improve the

coordination and in the long run, the quality of the user support system in federated e-Science infrastructures can be enhanced.

From the survey questionnaire and interviews, the conceptual model of a support system in e-Science infrastructures is thought to be comprising of the following four elements (see Figure 7.1):

- **Communication system** (with standardization, unified pattern of communication): The better the quality of communication the better the quality of user support system. The employees cannot solve all of the problems unless they have a better communication system with each other as well as the end-users of the e-Science infrastructures.
- **Infrastructure problems:** Most of the problems occur as the user try to access data and download it. If the system is good enough in a way that there are fewer failures of nodes or components, people will have reduction of the total number of requests. Reliability of the infrastructure is the main factor. The infrastructure has to be robust. SLAs can be formulated for *data access and download* and other facilities. If OLAs and SLAs are maintained by the federating partners of ESGF, we can minimize the e-infrastructure errors. If infrastructure is stable then the quality of the user support will increase. Quality control is needed to be done i.e. assurance that services of the e-Science infrastructure are delivered with Quality of Service (QoS).

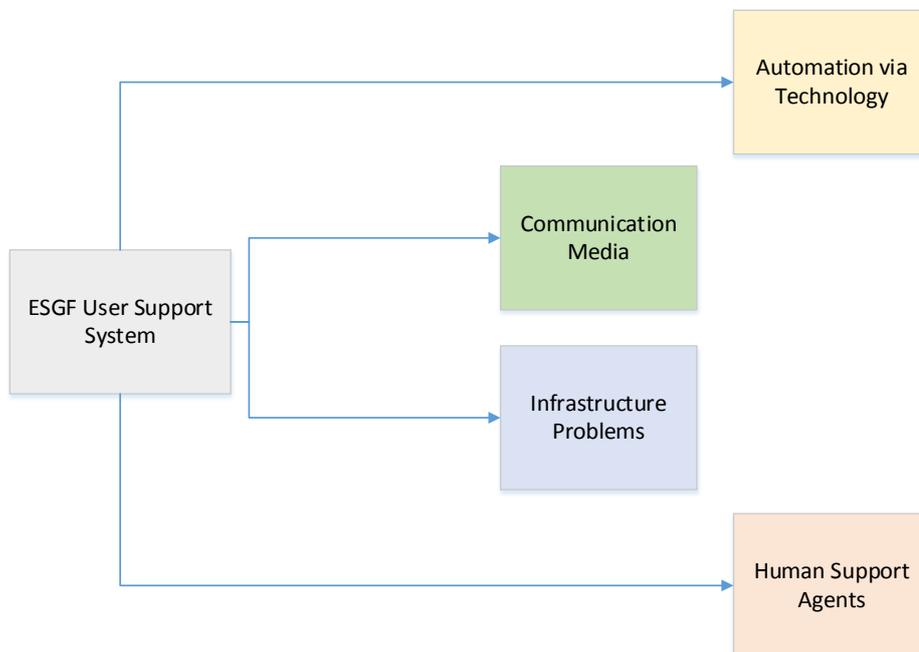


Figure 7.1 The conceptual dependency of ESGF User Support System.

- **Use of technology enabling techniques:** Currently, there is a minimal use of technology to enable automation within the user support process of e-Science

infrastructure in order to ultimately enhance the quality of user support. Currently, there is a simple system to track requests in place and partly a mailing list. However, if a more advanced or automatic troubleshooting techniques are used it could ultimately reduce the number of requests. For instance; if there is a problem experienced by the users, the users send it to a system and system answers the question automatically using techniques like CBR, CBA in association with central knowledge-base.

Even the ESGF system and its components can generate events and send the code name to the support system and the support system can interpret it. Issue log, enables to trace the problem. For instance; in windows there is a problem code, via which one can contact Microsoft to report a problem *xyz*. As a result, they understand the number and send you the solution. This is because there is a standardized way of solving a problem. Problem description is sometimes very much time consuming and staffs looking at lot of e-mails to understand the problem makes the whole user support process expensive.

If a problem (whether trivial or very simple) in the system is encountered; ESGF e-Science infrastructure may generate a number and from this number the staff or the ESGF user support system can detect what kind of problem the user is facing. In case it is a known error the problem can be solved in a matter of few minutes or few seconds.

Consequently, the type of problems can be filtered or routed to different parts of the federation i.e. partner institutions. There are some other issues that are needed to be considered such as problem in using technology, developing technology so this factor was evident mainly because of no resources. Resources are needed to be set after doing the cost and benefit analysis. Cost should be rationalized with the benefit.

- **Human agents of support service:** The skill and experience of user support agents who handle user queries is very important. The quality of user support process and user support system depends on the way the user requests are handled and processed. The total number of human support agents available to support users, the time zones covered, analytical skills of human support agents, their background knowledge, problem solving skills and empathy all matter to the success of user support system.

In the current user support process, most of the ESGF staffs are performing user support as a part time activity. There are no standard norms that have been set so far. As a result, there is no responsibility and accountability. With the initiation of this study, a full-time user support employee was employed to treat user queries. The full-time user support employee is funded by IS-ENES project, which is a European collaborator of ESGF consortium.

So these four factors affect each other.

7.3 Analysis of historical aspects of the user support process

The user support process in ESGF has been evolving in the last decade, as the nature of ESGF changed from the research test bed of grid technology to the production environment serving climate data projects. With the provision of interface to the users to facilitate them to get data, compute it and visualize it via variety of tools and methods, the user support process started its evolution. Up to now several evolutionary steps have been taken place in the ESGF user support process. Learning from these steps that occurred in the past, the author brings a new approach to consolidate them. In future, energy and biological data projects will be served via ESGF thus enabling further changes to take place. In the following sub-sections; the different approaches to support users are elaborated:

7.3.1 Evolution, advantages and disadvantages of the help desk media

The data gathered from the survey-questionnaire, interviews, observation and other sources revealed the details about the trends of the help desk models in use in the current user support process in ESGF as well as its associated projects. The information about the tools used in supporting users and to communicate between the users and the e-Science infrastructure's staffs have revealed a number of user support mechanisms. Each user support mechanism inbuilt in the current user support process in ESGF has its strengths and weaknesses.

From the survey, more than 80% of the all communication between users and staffs comprised of e-mails, see Section 6.8.2. Therefore, the main mode of communication between the users and the staffs to seek help was found out to be mainly asynchronous textual in nature; e-mail based through mailing-list (ML) and via request tracking software (RTS). Part of the description of the current user support process in ESGF was already published in conference papers (H. I. Chunpir, Ludwig, & Badewi, 2014a, 2014b). From the data analysis of the survey questionnaire the pros and cons of both of these methods; ML and RTS can be summarized in the table as under:

Table 7.1 The comparison of advantages and disadvantages of the mailing-lists and the request tracking software.

<i>Advantages of mailing lists (ML) in relation to e-Science Support process</i>	<i>Advantages of request tracking software (RTS) in relation to e-Science Support process</i>
<ul style="list-style-type: none"> • all stake-holders (including end-users, developers, policy makers, data managers, scientists) can participate • least expensive method to 	<ul style="list-style-type: none"> • history of user requests is tracked • user requests can be assigned to a concerned support personnel • search mechanism available for user requests • priority mechanism available for user requests

<p>implement</p> <ul style="list-style-type: none"> • feasible for distributed support units • no licensing issues 	<ul style="list-style-type: none"> • user requests have a particular assigned tracking number and they cannot be ignored • solution archive connection can be made • knowledge base can be made available • privacy settings can be made • FAQs and How To's can be updated automatically • statistics about user-requests can be generated
<p><i>Disadvantages of mailing lists (ML) in relation to e-Science Support process</i></p>	<p><i>Disadvantages of request tracking software (RTS) in relation to e-Science support process</i></p>
<ul style="list-style-type: none"> • no history track of user requests • no assignment method to further support personnel • no search mechanism for user requests • no priority mechanisms for user requests • may lead to ignoring some of user requests • no solution archive connection • no knowledge base • no privacy • long term costs associated due to overhead involved in handling e-mails and redundancy involved in treating e-mails 	<ul style="list-style-type: none"> • limited participation of all stakeholders. One needs to create logins and accounts for each person who want to login into request tracking software. It makes even difficult to assigns hundreds of logins to support staffs that are present at distributed locations. • system administration skills are required to configure, manage RT systems. • it could be expensive in terms of license costs and human resource cost, training costs • security problems may arise too if access at distributed location is allocated.

The results of the survey questionnaire were communicated to the ESGF executive and the ESGF technical board last year (May, 2013) to take decision on the type of tools to be used to facilitate user and staff communication to support e-Research activities. There was no unanimous decision taken so far by the ESGF board about employment of particular user support tool, federation wide. However, combinations of various methods have been tried in the form of help desk in the ESGF user support process to make the whole process better. The details are provided in the forthcoming section.

7.3.2 Different models and tools used until now to support users in ESGF

ESGF user support process is based on communication with the users and the staffs. The different forms of communication and tools used in the user support process are as follows:

7.3.2.1 Mailing list

One of the main tools for this communication is via the mailing list; esgf-user@lists.llnl.gov, which started from the beginning and is in place until now. Users send an e-mail to that list (see step 1 in Figure 7.2), and every mailing list member gets a copy, as it is usual for mailing lists (see step 2 in Figure 7.2). Someone who feels authorized sends an answer to the mailing list voluntarily (step 3) and the answer is again distributed to all list members (step 4). Occasionally, important questions and matching answers are collected and put together in the form of How To's and FAQs (in step 5, see Figure 7.2).

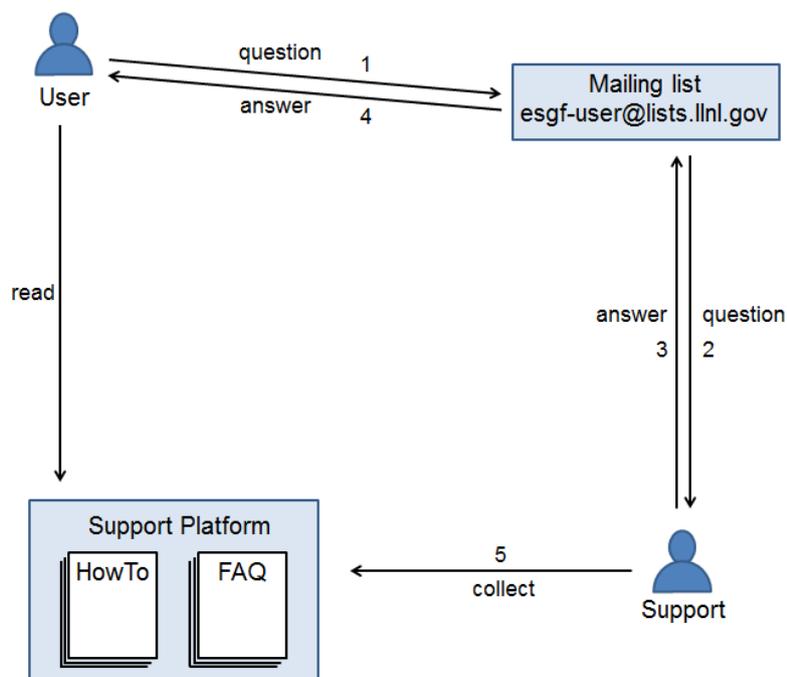


Figure 7.2 Current ESGF user support process with ML.

Nevertheless, a number of problems witnessed via mailing-list analysis remain with this mode of user-support staff communication:

1. **Lost issues:** Often e-mails from users are received and distributed via the mailing-list (ML), but from time to time; these received e-mails are not answered. Later these emails get forgotten in the heap of consecutive mails and are not answered.

2. **No assignment mechanism:** No formal assignment of user e-mails to a particular user support staff can be done. Consequently, it is often not known that who is responsible for the received e-mails to reply. Thus, no assignment mechanism is present.
3. **Incomplete history:** At times, users send feedback e-mails or a follow up e-mail and the preceding mails concerning the same issue are not included i.e. the history is not shown. Hence, it is troublesome for a support person to find the history and track it in the mailing-list. This leads to time loss, creating an inefficiency or dissatisfaction as well as irritation for the user support staff.
4. **Access to backups:** Every support co-worker should have an access to a stored copy of every e-mail without seeking further help. Until now mail list entries have been stored at LLNL, but one needs the help of an ESGF staff, e.g. Mr. X to get a copy and even Mr. X has archived only his part of history. Not the complete history is maintained.
5. **Difficulty in finding e-mails:** As a consequence, there is difficult to access stored copies of e-mails and the specific mail traffic on a special issue can hardly be found, thus making the whole process cumbersome and time consuming.
6. **Urgency:** The priority of an issue is not recognizable immediately. *No priority mechanism is present.*

7.3.2.2 Request tracking software

Request tracking software (RTS) was partly employed by some of the partner institutions in federated ESGF and is still being used by some partner institutions e.g. BADC. However, due to lack of consensus to use a common RTS federation-wide, RTS could not become the common platform to support users of the ESGF federation. Similarly, the underlying user support process (personnel) roles and features were not clear and defined unanimously in the ESGF, federation wide. The problem witnessed was not only lack of consensus but also lack of central RTS accessible by each partner institution in the federation. These problems included issues of accessibility and security.

Improvements in RTS based user-staff communication model are needed and are also possible. For instance; forming a RTS based help desk with federation wide unanimous policies about the user support process and assigning roles to the employees will improve the situation. The help desk consists of software that has synonyms; request tracker, best-practical tracking system, issue tracking system etc. In this model, the staff is recommended to be based on two-tiers; First Level Support (FLS) or Second Level Support (SLS). However, in ESGF currently there is no division of FLS and SLS staffs.

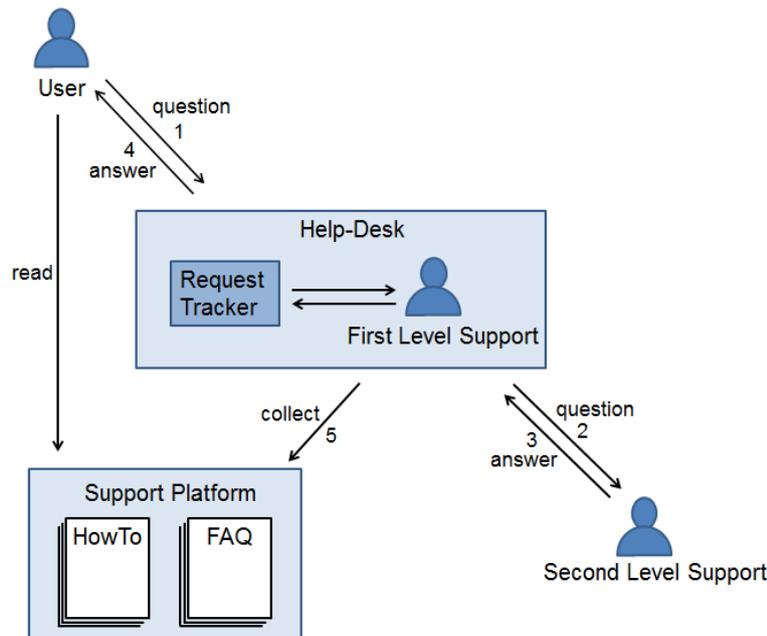


Figure 7.3 Integrating help desk in ESGF user support process.

The user sends an email to the help desk system (step 1 in the Figure 7.3). A ticket is automatically opened and an automatic response with the ticket number is sent back as a consequence of each arriving email (not indicated in the Figure 7.3). The FLS solves the problem or forwards the ticket to the contact address of the responsible ESGF partner, according to the content of the ticket (step 2). A *list or directory containing a contact address of each ESGF institution* is necessary to do this, which is currently not present and visible to support staffs. The support provided by an institution is the second level support (SLS) in view of ESGF as a whole.

SLS solves the issue and takes down a notice for the user in the help desk system, this way completing the ticket (step 3 in the Figure 7.3). Every support co-worker should be able to login on the help desk system. On the contrary, there is no necessity for users to login there. The solution is simply sent by email to the user (step 4), and the ticket is closed. If an issue occurs quite often, the FLS adds it to the FAQ or to How To's (step 5), maybe with the help of the SLS again. The main problem with this set-up in ESGF is that there has been no unanimous consensus on deciding which help desk software to use collectively. There have been suggestions about FootPrints, GitHub and RTS. Different request tracking softwares have been partially employed by different partner institutes in ESGF, at the moment.

7.3.2.3 Co-existence of help desk and mailing list

Figure 7.4 shows an intermediate phase that still exists in the ESGF user support process. In the intermediate phase, mailing list and help desk coexist. The intermediate phase started by adding the help desk system to esgf-user@lists.llnl.gov.

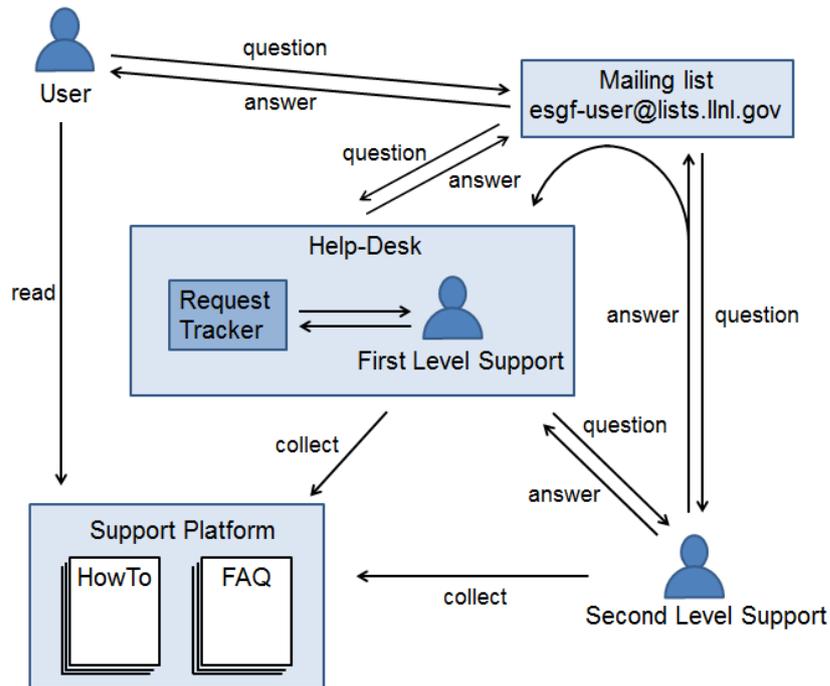


Figure 7.4 Intermediate phase with coexisting mailing list (ML) and help desk (HD) in ESGF user support process.

The intermediate phase is incumbent, since it was hardly possible to convince all ESGF partners to switch completely to the federation wide help desk. Currently, support staffs and users use the mailing list as the main medium and from mailing list; requests are then transferred to local help desk at institute level.

Help desk solves the following problems 1 to 5:

1. Issues could not get lost, since:
 - A ticket is opened for every issue which is numbered and assigned to a contact person.
 - If the contact person chosen by the FLS didn't match, the contact person would give back the ticket to the First Level Support.
 - Reliability of the user support service can be further increased by setting of response times, sending of reminders and an escalation mechanism. Of course, currently no one is forced to obey response times as there are no SLAs in place. However, methods like SLA have been seen to be effective in other grids without written contract about support activities.
2. History of the incoming user requests can be updated by the help desk system, if the correct ticket number is used. History would remain complete, locally though, since:
 - Support staff get access to the helpdesk system and could directly edit a ticket;

- Users may reply by providing feedback information using the ticket number. Then feedback could automatically be assigned to that ticket.
 - If the user's feedback mail didn't contain the ticket number, the FLS need to merge the original and the "new" ticket manually.
3. Since every support co-worker had access to the central help desk system, direct access to every ticket could be guaranteed.
 4. Support staffs shall be able to easily find tickets because of:
 - Direct access
 - Search function
 - Categorization of tickets, for example assigning the categories such as:
 - Data download
 - Login + User access
 - User registration
 5. A priority level could be assigned to the ticket. This might be done by a user or FLS.

It is further necessary to insert a list of contact addresses for each ESGF institution into local RTS. Despite all positive features provided by local RTS, a central RTS access is necessary. The security, role administration, moderation issues are hindrances in adopting a central RTS that is connected to all the partner institutions.

7.3.2.4 Co-existence of user forum and mailing list

Another important scenario was trial of user forum in ESGF. The central users' web forum coexisted along with the mailing-list from August until November, 2013 (see Figure 7.5). After a series of meetings, it was decided to employ a stack-overflow-type of communication forum known as *Askbot*. The web forum Askbot was installed as a pilot phase. However, after November, 2013 due to hard-disk failure at a partner institute of ESGF Askbot was disbanded. Recently in June 2014, it was made also active. Askbot is similar to Reddit or stack exchange system.

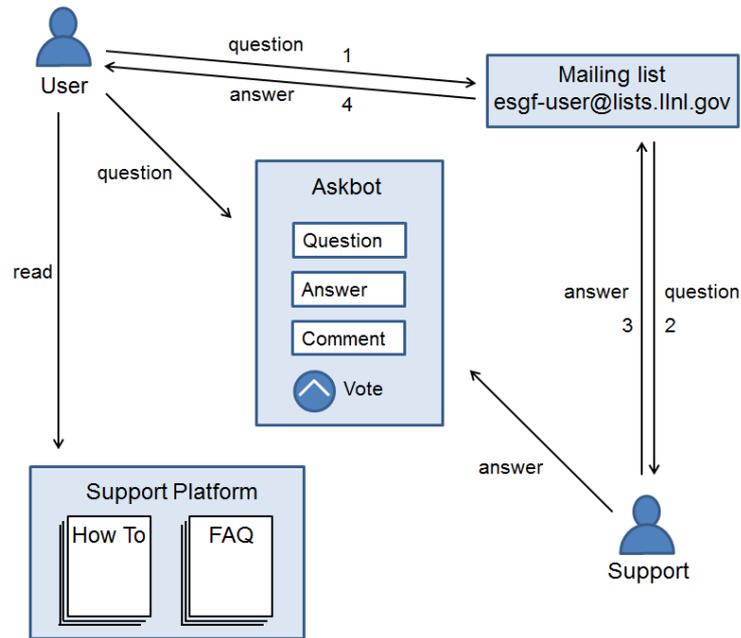


Figure 7.5 Intermediate phase with coexisting mailing list (ML) and Askbot user forum in ESGF user support process.

The positive points witnessed in this scenario were; every post remains visible to all. All registered members of the Askbot can post and rate. Good questions and responses can be cited in new posts. Many similar questions can be categorized together. Reward system can be integrated; however it was not integrated in the pilot-phase of Askbot employed at ESGF. The major problem with the web-forum is that it needs a constant moderation and over-look because users may set similar key-words to the posts. Some impolite posts may also appear from time to time and an active moderation is needed to sort these problems. Due to the pilot-phase of Askbot it was not clear to many users that a web-forum in the form of Askbot exists so they kept using the mailing-list (ML). For employing Askbot it is important to buy a license for server software.

7.3.2.5 Forum

The aim of employing Askbot was that in future the mailing-list (ML) will be closed and the Askbot type of web-forum will be used where the users as well as ESGF staff can post and rate the workarounds or solutions freely. However, this hypothetical state as shown in Figure 7.6 was not reached so far, though the ESGF staffs would like to reach it.

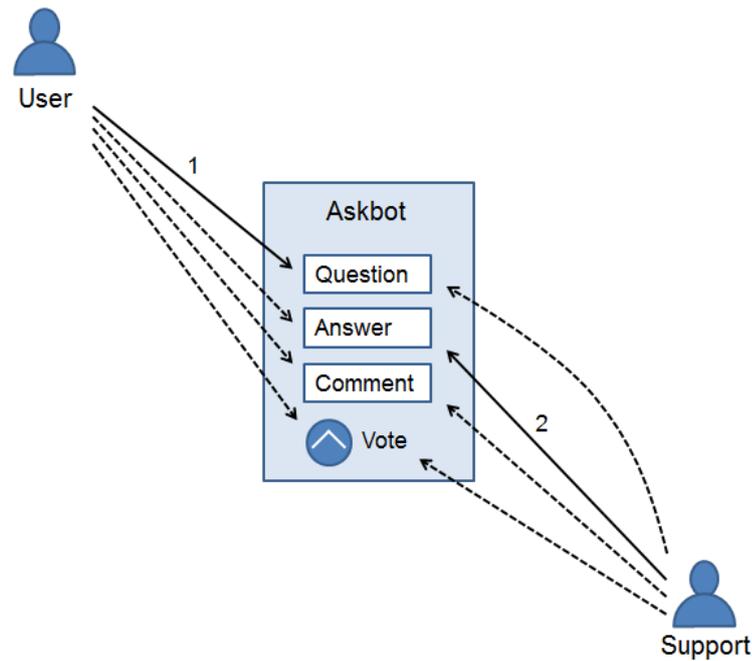


Figure 7.6 The next phase without coexisting mailing list (ML) that ESGF staffs want to achieve as an ideal ESGF user support process.

7.3.2.6 *Combination of RTS and users forum*

Another idea that was propounded by the author was the use of Askbot type of user forum along with the RTS at the same time in such a way that the users can post queries on the web-forum and concerned moderators can look at it and if it is needed to be treated by a particular partner institute of ESGF. Then, the incident can be shoved into the local RTS of institute and handled internally and finally if a solution or a workaround is found then the solution or workaround can be posted in the web-forum. Even in this case, there can be two variations of RTS; either single central RTS or a many local query tracking software. So this model is a mix of Figure 7.3 and Figure 7.6, as shown in Figure 7.7.

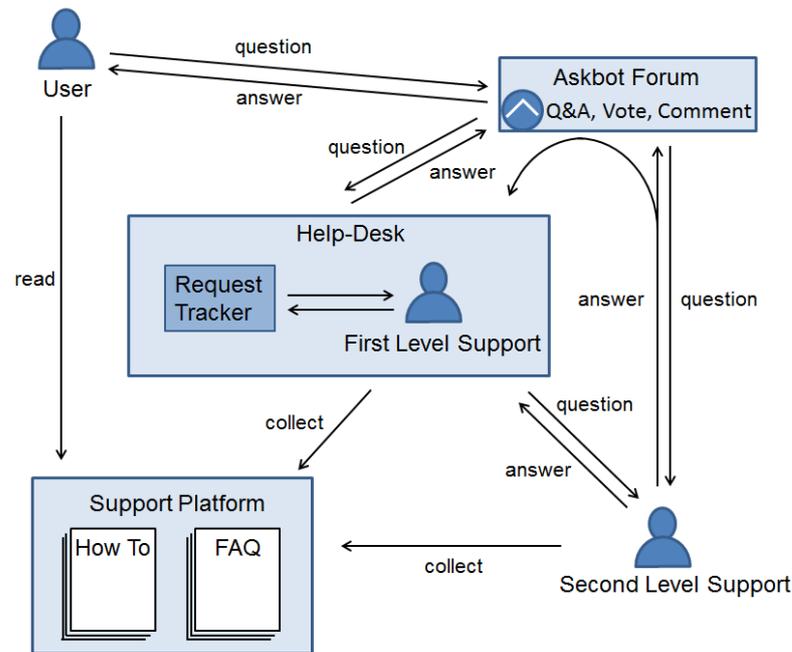


Figure 7.7 Using user's forum and RTS system in combination as a solution.

In summary, a central help desk system such as Github, federation wide in ESGF can offer and allow the following actions:

- Merging of tickets
- Access for all ESGF support co-workers
- Deletion of tickets (spam)
- Automatic reply
- Open a ticket via email
- Categorization of tickets
- Search
- Setting priority

A central RTS access along with the central user support forum is need of the hour in ESGF. The security, role administration, moderation issues may pose difficulties in adopting a central RTS that is connected to all the partner institutions.

7.3.2.7 Discussion

From this it was learnt that so far the changes that have been done directly at an operational level, these changes might not be effective until there is a complete user support process definition and a role allocation. It is mandatory to plan the process in a systematic way from top to bottom i.e. from strategy making for a service to the implementation of the strategy.

Like many e-Science projects and other scientific projects user support is not a primary activity but a “volunteer” activity. It is therefore important to provide framework for efficient user support. The current methods of help desk to support end users are

presented, however all of these methods have pros and cons. The framework that works independently of all help desk tools is made the part of the service strategy of ESGF and its associated scientific projects, the help desk alone cannot be sufficient.

The cost and benefits of all these methods are mandatory to be assessed. In this research the cost estimates of all these methods have not been analysed. However, two points are clear, mailing-list is a short-term solution, although it seems an inexpensive and flexible solution but in the long run it does not support cost effectiveness, sustainability and reinforcement of standards in the ESGF and its associated projects. This is due to cons stated at the start of this section. The RTS has its own pros and cons and the security and accessibility issues make it a non-suitable candidate to some extent. However, RTS-like system in combination with a user's form like *Askbot* seems to be a better option. Nonetheless its effectiveness is needed to be tested.

Human resources and financial resources are one of the main issues to consider in taking decision to apply an appropriate tool. It is recommended by the authors to invest time and money to introduce a platform that combines the benefits of ML, RTS and Askbot types of communication forums. This can be then tested and put in place as a pilot platform to see whether users and employees both have combined benefits i.e. pros of most of these methods above. A work has been underway to create a prototype that combines pros of all of these methods including some new features to facilitate the user support process in e-Science infrastructures, especially suitable for ESGF. This work has been elaborated in the next chapter (Chapter 8).

From all the current models of user support, the lessons are learnt that before implementing any model the pre-condition is needed to be considered. The pre-condition is definition and a unanimous agreement on a user support process. The technology oriented solutions in the form of help desk models can never be sustainable in the long run. The procedures behind them are needed to be elaborated, agreed and implemented first. The technology oriented solutions are not end in itself rather they are means to an end. Therefore, before implementing any solution or model, it is vital to carry out a comprehensive cost benefit analysis as well as the suitability of technology in the user support process and the acceptability of technology with the staffs that supports users.

Furthermore, it is important to investigate the interaction between users in this context as well. From all the lessons learnt, at first the technologies of RTS and Askbot can be taken because it combines the best features amongst other options and the pros of these technologies together offer a social network-like solution. However, in the longer run a social support network that suits and support the underlying framework must be developed by the ESGF team or others.

7.3.3 Self-help improvements

The self-help is for the users to help themselves online by using online resources. Over the last three years there has been substantial uplift of the webpages offering FAQs as well as Wikis. The content of Wikis and FAQs has been updated periodically. Similarly, the information architecture (IA) of the websites has been reformed. Video lectures have been made in the form of webcast and posted as well. Documentation has been made better now. The tables of contact persons who are in charge of different components of the ESGF federation at the different locations at the partner institutions have been updated. However, there are no specific duties allocated to people to update the information. The information updated on the websites has not been tested with the users as how effective and comprehensive it is.



Figure 7.8 The screen shot of the tutorial in the form of video webcast on how to log on in the ESGF federation gateway to access climate data.

7.3.4 Participation of users in the ESGF

Users are encouraged to participate in the ESGF user support process in the last few years. The users are not only encouraged to participate in the discussions not only in the mailing-list and user-forum such as *Askbot* but also for the first time the users are invited to participate in the face to face (F2F) meeting. The participation of the end-users has been encouraged by the chair and other members of the ESGF on the suggestion of the author. In the mailing-lists the user participation is still very less. In the users forum *Askbot* it is a bit higher.

7.3.5 Appointment of user support staff

Before a year, staffs of ESGF and its supporting process were completely on volunteering basis. However, now one position of support has been created which is being funded by IS-ENES project from the EU. However, the problem with the position

is that a highly-qualified scientist has been allocated to perform user support. He is answering most of the e-mails. Most of these e-mails can be treated by FLS user support technician. Although for now there seems to be utility of the position, however in the longer run the money being allocated to pay the salary of a well-qualified user support person is waste of resources because a person with high capabilities and skills is doing a simple and routine job. The scientist who is doing this job is paid 4500€ per month. Either the person may be assigned the job as an in charge of the user support process for its enhancement or alternatively two to three FLS technicians can be engaged to do the same job with half of the budget what the highly skilled scientist is doing.

7.4 The core problems identified in the current ESGF user support process

The core problems identified via the empirical investigation are mainly related to the management and organisation of the user support process in ESGF. Moreover, establishing a shared understanding of user support in e-Science particularly in the federated ESGF is seen as a pre-requisite to the successful management of the user support process and implementation of the service desk function. From the results in this study, the current situation of the user support is rather heterogeneous, short-termed and non-goal oriented. Consequently, there are hurdles in establishing a well-organised user support process (H. I. Chunpir et al., 2014a). These hurdles are needed to be overcome. In the *as-is* situation though there are isolated efforts by organisations; which are partners in the ESGF, nonetheless there is no combined effort at a higher level to re-structure the user support process.

The role of managing user support activities is not defined exclusively and these activities are being performed by the employees of ESGF in addition to their core activities of their job description (H. I. Chunpir et al., 2014a, 2014b). For instance, there is no designation of user support manager who is exclusively responsible and accountable for performing user support activities. As a result, the user support activities are performed by the top computer scientists and climate scientists, in particular servicing simple and routine user queries. Henceforth, this is mainly misappropriation of their creative skills. Moreover, the additional responsibility on the shoulders of the technical staff does not reward them in monetary terms, promote them or bring them some form of recognition. Thus, the user support requests are not serviced by the employees on regular basis resulting in overlooking of some user requests (H. I. Chunpir et al., 2014a). All interviewees are of the opinion to assign user support responsibility to someone who is a structurally anchored person with the ESGF dealing with the user support process.

In addition to the misunderstanding of the term “user support”, some partner organisations of an e-Science infrastructure face three challenges: they need to structure

the user support process and thus define clear process steps. Secondly, clear definition of responsibilities and allocation of budgets, especially user support budgets are needed overall the structure of ESGF. Thirdly, some flexibility and informal work to support creativity in these processes must be allowed. Some interviews especially the technical employees oppose formal structure of the user support. In addition they do not agree to structure the user support and document its process. Therefore, the user support process shall be partially formalized and it should be allowed to some extent to work outside of its pre-defined structure.

All interviewees i.e. the employees are in favour of allocating user support budget exclusively for the user support services and resources from the overall budget of ESGF. Though there are some techniques and methods that are tested by some organisations participating in the ESGF in the past to support user support process but none has been adopted as a standard. This is because each method having its own pros and cons do not fit in the dynamic structure and culture of ESGF. Moreover, these methods were technologically oriented without having a rationale to use them from a holistic user support perspective of ESGF (H. I. Chunpir et al., 2014b).

Consequently, the overall strategy as well policy to govern ESGF's user support process is missing due to lack of the overview and the holistic perspective of the ESGF user support process. In addition, there is no consensus on the methods used to service the user requests. Furthermore, the methods; one or the other are not completely standardized, communicated, documented and known to the entire ESGF partners. Moreover, these methods are needed to be connected and interlinked to other processes and tasks carried out with in ESGF as a whole.

All interviewees wish to professionalize user-support process in ESGF e-Science infrastructure. Therefore, they recommend integrating user support process for the whole infrastructure for all partners participating in it with the general management. Some interviewees especially developers and members of the technical committee do not agree to structure the user support and document its process. The problems deduced from the interviews are briefly summarized in the Table 7.2.

Table 7.2 The table shows the number of problems and their description deduced from interviews.

Problem #	<i>Problem Description</i>
P1	different <i>understanding of the term user support</i> amongst stakeholders
P2	no clarity who <i>the end-users</i> of the ESGF are, amongst the stakeholders
P3	<i>unclear role definitions, compensation and accountability</i>

Problem #	<i>Problem Description</i>
	for the staffs of federated partner institutions performing user support
P4	<i>no central documentation recorded or data base</i>
P5	<i>no optimal use of human resources</i>
P6	<i>No allocation of budget for end user support: Responsibility vs. allocated amount</i>
P7	no defined methods of user support management and they are not used e.g. SLM
P8	<i>heterogeneous situation regarding user support practise and culture at every participating partner organisation in ESGF</i>
P9	missing combined overall <i>user support strategy</i> for the federated ESGF
P10	need to professionalize IT user support process (<i>missing structures in user support</i>)
P11	<i>lack of integration and connectivity of the user support process with other processes in the ESGF e-Science infrastructure</i>

7.4.1 Reasons to apply ITIL guidelines in e-Science

The growth and maturity of e-Science infrastructures after achieving technical excellence in the underlying technologies such as Grid and Cloud technology has enabled wide usage of user communities to interact with e-Science infrastructures. The user communities of all data projects hosted by ESGF use the e-Science infrastructure. Unfortunately, the overall management and coordination currently hardly address the two main challenges that the e-Science projects and e-infrastructures face, first the challenge to provide *user satisfaction* and secondly making the e-Science infrastructure *sustainable* (Appleton et al., 2014). If the user experience and usability of a system is better, the better the user satisfaction is. The soft factors are important to success of any system. A system is used by people thus there are social aspects that are essential to consider. One of these aspects is providing support to users especially when the users are researchers and do not want to wait for a longer period of time to fetch data from data centres and perform computations using HPC.

Researchers have pointed out relationship between ITIL and the service desk in the non-commercial environment (Arora & Bandara, 2006; Jäntti, 2009, 2012a, 2012b). ITIL is suitable in any organisation whose goal is to deliver IT services that meet the need and requirements in user or consumer oriented set-ups (Potgieter, Botha, & Lew, 2005).

ITSM manages the IT function as a service function and ITIL is the most popular ITSM framework amongst IT managers in organisations. Other frameworks to manage ITSM are Microsoft Operational Framework (MOF), HP ITSM and IBM ITPM (Iden & Eikebrokk, 2013).

Another reason to concentrate on ITIL is that ITIL can be used in harmony with other Best Management Practice (BMP) products or international organisation standards (Office of Government Commerce, 2011). Other BMP publications include: Management of Portfolios (MoP), Management of Risk (M_o_R), Management of Value (MOV), Managing Successful Programmes (MSP), managing projects with PRojects IN Controlled Environments, V2 (PRINCE2) and Portfolio, Programme & Project Offices (P3O). P3O is to develop, establish and maintain support structures (Iden & Eikebrokk, 2013).

ITIL is flexible and embraces a practical approach to service management. One can apply whatever works and leave whatever does not work within an organisation. Moreover, ITIL is vendor neutral i.e. any organisation irrespective of the platform it is using can utilize ITIL practices to add value to its business. ITIL is generic i.e. it offers robust, mature and time-tested practices applicable to all sorts of organisations. Furthermore, ITIL recommendations are based on best and successful service providers around the world (Office of Government Commerce, 2011).

ITIL is diverse because it incorporates cost benefit analysis, manages risk, measures as well as monitors IT services, manages knowledge, enables adoption of standards and improves interaction with customers. Moreover, ITIL encourages changes in the organisational culture to support goal oriented success and coordinates the delivery of services to deliver value to customer through services. Likewise, ITIL reduces costs for the organisation providing the service to customers, thus optimizing costs (Office of Government Commerce, 2011; Potgieter et al., 2005). Consequently, for the enhancement of the user support situation in ESGF e-Science infrastructure of climate domain, ITIL guidelines can be applied. In the following section, the problems identified earlier are handled in the light of ITIL V3 from 2011 recommendations. However, these recommendations are adapted to suite the dynamic and ever changing environment of e-Science infrastructure such as ESGF.

In e-Science, the underlying IT infrastructure is mostly federated. Therefore, instead of IT infrastructure, *federated e-Science infrastructure* is needed to be considered. The recommendations made in the light of ITIL guidelines combine the best practices from the industry, public sector and adaptations of ITIL in the academic sector. These guidelines are then applied to e-Science infrastructure to augment the IT management and administration that are deficient to bring quality of service (QoS) to large scale, heterogeneous and geographically distributed resources. ITIL has five major phases

starting from service strategy planning, design, transition, the implementation and continual service improvement (CSI). These five phases to manage IT services is also known as ITIL lifecycle. The lifecycle addresses the services provided by the IT infrastructure to the customer in the form of delivering value to the customer.

7.4.2 The need for a framework

The technical architecture and technical capabilities to deliver services are mature in the ESGF i.e. the e-Science infrastructure has achieved its production level. Therefore, in this research study, the main focus is on the *user support service* in the federated environment as one of the main services amongst other sets of services offered by the ESGF e-Science infrastructure. The federated environments such as federated IT infrastructures, e-Science infrastructures often lack the hierarchy and level of control seen in other situations, therefore; a federated user support framework is needed to provide user support interoperability across the federated e-Science infrastructure like ESGF. This can be achieved by defining best practices, documenting responsibilities that are well understood, defining clear inputs and their transition into useful outputs without consuming many resources and doing unmanageable changes. Consequently, Federated User Support Enhancement framework (FeUSE) is recommended to tackle the problems mentioned earlier in Section 7.4 (page 224). FeUSE does not recommend a heavily metricized solution that is tightly managed as in other frameworks like ITIL, COBIT and others.

The traditional ITSM frameworks normally tend to address traditional non-federated IT environments where there is a single point of control and there is no lack of hierarchy. However, the e-Science infrastructure federations tend to lack hierarchy and a single point of control that generally traditional ITSM assume to have in an organisation (Szepieniec, 2013, p. 4). Consequently, there is a knowledge gap in ITSM frameworks to cater user support of the federated e-Science environment. The traditional non-federated ITSM tools and frameworks do not take the distributed federated resources and norms into consideration.

FeUSE is a first step towards streamlining user-support and implementing FitSM. FitSM provides guidance in implementing different aspects of ITSM in the federated environment and targets a lightweight implementation of ITIL in the federated environments. FitSM is compatible with ISO/IEC 20000-1 (requirements for service management system). The FeUSE framework is seen as part of ITSM, as it harmonized user support across federated e-Science infrastructures. E-Science services are provided in the form of a loose collaboration of organisations acting as a federation. FeUSE defines minimal set of requirements for the user support process which are achievable and useful for federated environment as traditional ITSM practices are difficult to apply and their application in the federated environments is overambitious.

In the federated e-Science environment, FeUSE helps to create a user support management plan, implements it, maintains it, monitors and further improves it. The plan must include at minimum the objectives of user support, roles and responsibilities involved in it, the interface to other processes with the e-Science infrastructure, the tools, technologies, and techniques to operate the user support service. In addition, FeUSE helps to monitor user support process using suitable methods including reviews and internal audits and identifies the nonconformities to the defined norms, standards. Finally, corrective methods and actions are taken to prevent nonconformities.

7.5 Chapter summary

Chapter 7 analyses the empirical findings presented in the last chapter. The analysis of attributes of people involved in supporting users of ESGF and its associated projects, the user expectations from the user support service, their current satisfaction level from the service, the structure of user support and the factors related to user requests are presented in this chapter. Furthermore, analysis of communication aspects, documentation aspects and time related factors are also given in this chapter. Moreover, after the analysis of each attribute of the elements of current user support process in ESGF the suggestions to improve the current state of user support process are specified. The weaknesses and strengths captured from the survey-questionnaire are listed at the end of the analysis of the current user support process in ESGF.

From the analysis of all the attributes of the current user support system in ESGF, a conceptual model of a support system in federated e-Science infrastructures is derived. This conceptual model of a support system in e-Science infrastructures is thought to be dependent on four chief elements, which are: Communication media, stability of an e-Science infrastructure, skills of human agents involved in seeking or providing support, the technology as well as degree of automation of the user support system. Any disturbance or a problem in these four elements mentioned above can disrupt the function of user support system in an e-Science infrastructure.

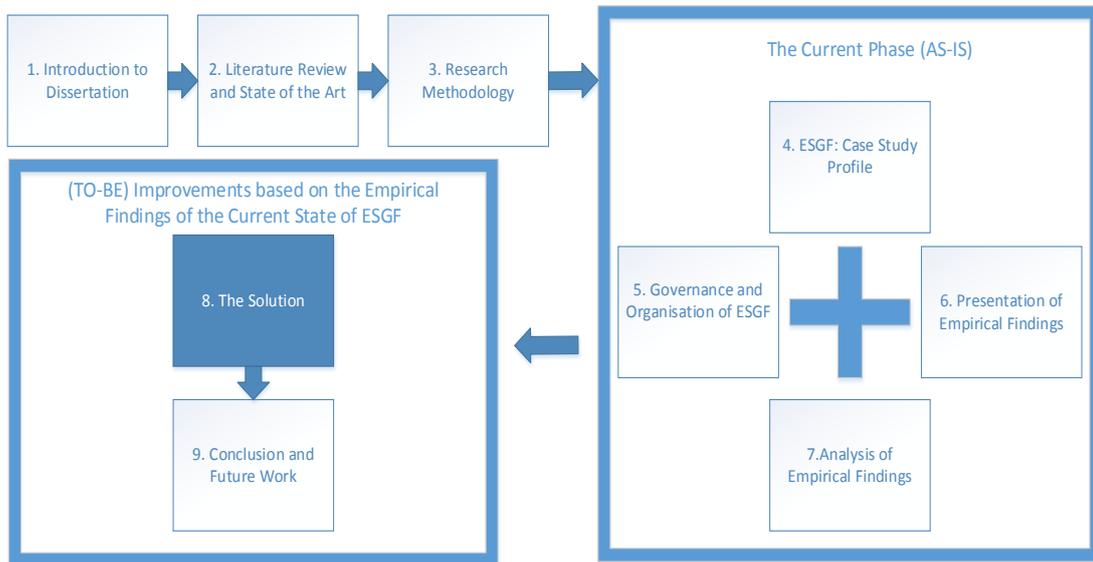
Additionally, in this chapter, an evolution of different models and tools of communication media, used until now to support users is provided. One of the major factors in not sticking to one communication media in ESGF is lack of consensus on using a sole communication model and tool for the whole federation. Moreover, suggestions of developing a self-help culture and a culture to persuade users to help others as well as participate in user support process in e-Science infrastructure are also given.

The core problems identified from the empirical findings in the current user support process are highlighted in this chapter. These problems include: Different understanding of the term “user support” amongst stakeholders, no clarity of the term “users” of ESGF amongst stakeholders, unclear roles of personnel, their accountability and monetary as

well as non-monetary compensation for supporting users in the user support process of ESGF and no optimal use of human resources. Furthermore, missing combined user support strategy, no central documentation, no explicit allocation of budget to the user support service, no service level agreements and lack of integration of user support process with other processes in an e-Science infrastructure were some of the other problems listed in this chapter.

The term “end-users” is preferred and instigated by the author to be used for the users of the services provided by ESGF. Moreover, to counter the problems listed above, the need of a light-weight framework called Federated e-Science infrastructure User Support Enhancement (FeUSE) was proposed by the author. It is also known as Federated e-Research User Support Enhancement (FeUSE). The reasons to use FeUSE were also given in this chapter. FeUSE is stimulated from frameworks like ITIL and COBIT which are found in different areas and different walks of life. The complete description and application of FeUSE on the ESGF user support process is given in the next chapter.

8 The solution (To Be) – FeUSE framework



8.1 Introduction

This chapter continues the discussion from the point raised in the last chapter about the need of the framework i.e. the guideline pattern for the federated e-Science infrastructures. This chapter describes the recommendations to enhance the user support process in the form of Federated e-Science User Support Enhancement (FeUSE) framework and its possible implementation in ESGF. The introduction to FeUSE framework is given in Section 8.2. The aspects of FeUSE framework (also known as phases or steps) are illustrated in Sections 8.3, 8.4, 8.5, 8.6, 8.7, 8.8 and 8.9, respectively. The readiness of implementation of FeUSE and other views about FeUSE framework by the ESGF stakeholders; such as its acceptability in ESGF are evaluated in Section 8.10. The future vision of this research such its extension in the form of a social network and social support forum is given in Section 8.11. Finally, the summary of the chapter is given in Section 8.12.

8.2 Federated e-Science infrastructure user support enhancement (FeUSE) framework

The Federated e-infrastructure user support enhancement (FeUSE) framework is based on the results of the empirical investigation that were carried out on ESGF as well as considering suggestions from popular ITSM best practice framework such as ITIL. The framework aims to contribute to sustainability and usability of the e-infrastructure's user support and hence the sustainability and usability of the e-infrastructure itself. This framework defines basic guidelines that any federated e-Science infrastructure can follow or adapt to stream line the user support activities, subsequently, turning these activities into knowledge that can be shared amongst all the stakeholders.

Correspondingly, the aim is that the innovation in the e-Science infrastructure can advance from the users' interaction with the people who support them. These framework recommendations are based on the ESGF and its associated projects. The validity and evaluation of this framework has been done with the focus groups as well. However, the complete evaluation of this framework has not been included in this thesis.

Likewise, user-centred approach has been exploited, so that users can be encouraged to provide suggestions to improve not only the user support system in place but also its UI, communication and collaboration. Moreover, users can be able contribute to even innovate in the e-Science infrastructure. Equally, a possibility has been offered to users to be part of the support team and contribute in helping co-workers to mitigate their technical and scientific problems. Consequently, the author proposes the seven major steps that must be considered by the federated e-Science infrastructure management and governance team to enhance user support process in any federated e-Science infrastructure. However, the applicability of each of the steps is left at the will of an implementer i.e. the policymakers of an e-Science infrastructure. These seven steps of the FeUSE framework are listed as follows:

1. Vision
2. Strategy
3. Communication
4. Agreement and Assignment
5. Support operation
6. Monitoring
7. Institutionalization

An overview of the FeUSE framework in the seven important steps is mentioned in the forthcoming sections. From all the seven steps; the fifth step is explained in further detail, keeping in view the T-model, as this step is concerned with the implementation of the user support operations.

8.3 Vision

In the FeUSE framework it is essential to have an explicit (long-term and short-term) vision of the user support process in the federated e-Science infrastructures.

The vision must have the following four objectives namely (see Figure 8.1):

- The user support system and process in the federated e-Science infrastructures like ESGF must aim to provide: *User satisfaction* to maximum
 - At the same time keeping the costs and resources as lowest as possible thus enabling: *Cost efficiency*
-

- The user support system and process has a key role in making the data projects and the over-all infrastructure stable, *reliable* and *sustainable*. This is because the support for these projects and the over-all infrastructure as a service. At the same time, FeUSE enables to document all details, capture user experience and provide improvements that enables data to be converted into knowledge and eventually wisdom
- FeUSE enables to provide transparency in upbringing *innovative ideas*, sharing them with the stakeholders, thus contributing the support activities towards *development* of the e-infrastructure

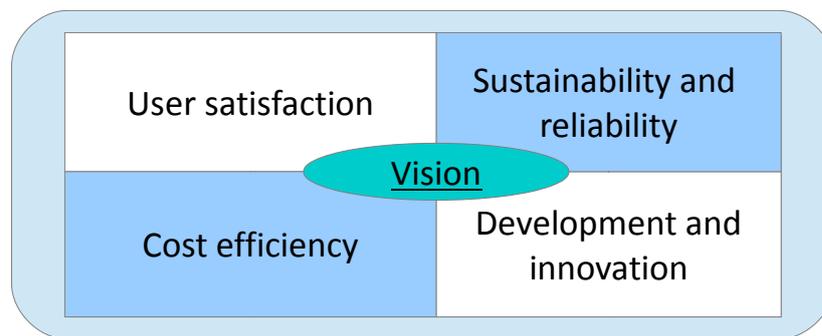


Figure 8.1 The first phase of FeUSE Framework.

Currently in ESGF user support process and practices are not explicitly following this vision. Yet, it is time to make clear to the stakeholders of ESGF e-infrastructure to organise support process and its activities in such a way that the vision of the user support process can be achieved. This vision is in accordance with the over-all vision of the ESGF-like e-Science infrastructure.

With the passage of time the vision of ESGF changed from the purely research oriented grid-technology based test-bed to a fully functional commodity serving climate research. ESGF has now a vision to serve researchers of climate science domain who study climate change to generate report. In order to support this vision ESGF provides services to its users. According to FitSM, the service is defined as: “A way to *provide value to a user or a customer through bringing about results that they want to achieve.*” The services offered by the e-Science infrastructure provide value to its consumers. Each service provided by the ESGF supports the vision of the e-Science infrastructure. Thus, the services contribute to the vision. Currently, there is a variety of services offered by ESGF to the users as well as other stakeholders such as data download, data post-processing, visualization and others. FeUSE recommends listing all these services being provided to the stakeholders especially the ones to the end-users.

FeUSE aims to enhance the user support process in federated e-Science infrastructures; therefore, in this framework, user support service is identified as one of the vital services provided by the federated e-Science infrastructure to the users. Without this service, users can feel difficulties to utilize other capabilities and services offered by an e-Science infrastructure. All other services and their nature may change, but user support service has to be available in every case for the users, to clarify their queries in order to retain them. Thus, it is essential to have a vision or a goal to be achieved in the longer run for the user support activities that constitute a user support process in the ESGF. This vision needs to be in compliance with the overall vision of the e-Science infrastructure.

However, currently in ESGF the vision of user support process is not explicitly defined. Therefore it is necessary to set goals and a vision to enhance user support for instance in ESGF. FeUSE proposed federated infrastructures to have four main goals (as already mentioned above): *User satisfaction, cost efficiency, sustainability* as well as *reliability, development* of e-Science infrastructure including *innovation*. The current version of FeUSE intends to improve all the four main goals in ESGF. Without the commitment of the top management of ESGF, the vision of user support service cannot be implemented.

In a nutshell, the services provided by the ESGF to the users must have a long term goal and a vision that fulfils the vision of the e-Science infrastructure. Supporting users may require some specific services from all partner institutions because of the distributed nature of e-Science infrastructure. Furthermore, it is important to ensure that all services from the supplying partner institutes are in alignment with the norms defined by the ESGF and the service vision.

8.4 Strategy

The FeUSE framework recommends developing a user support strategy based on the federated environment. A broad, combined, federated user support strategy is mandatory to the success of an e-Science infrastructure initiative. An e-Science infrastructure aims at fulfilling the research needs of research communities, therefore to meet the needs of e-Science infrastructure users. E-Science infrastructures like ESGF are providing user centric services and the aim is to retain user communities as well as attracting them to use the tools and capabilities of e-Science infrastructure to facilitate research.

8.4.1 Importance

An effective user support plays a pivotal role in attracting the user communities as well as retaining them. Presently, most of the users in ESGF with less technical affinity find difficulty in using e-Science infrastructures as evident from the interviews conducted in

this study. Consequently, the strategy to fulfil user needs i.e. providing them support shall be part of and well integrated into the overall strategy of an e-Science infrastructure. This creates trust amongst the users and provides service warranty of using the product to the users. Hence, adding to the overall reliability of the e-infrastructure in the eyes of users that can attract more and more users to use the system.

Although the ultimate goal of each e-Science infrastructure is to serve users, however, few e-infrastructures have structured their user support processes. In creating the strategy the following points are worth considering:

- How can we achieve the vision as far as user support is concerned in the form of long-term and short-term goals?
- Who exactly are the users i.e. the audience: Research domains and sub-research domains? What are their skill sets? At what times are they active?
- How many resources do we have? How much budget one can dedicate to user support? How much budget has been dedicated to the e-Science infrastructure as a whole?
- What is the value that we can achieve at the end of the investment? Why is it beneficial for the users?

All the stakeholders i.e. the policy makers of the federated institutions (of ESGF and ESGF-like infrastructures) must define a clear strategy unanimously in the light of the vision and the ultimate mission which is then translated into a plan. The strategy must ensure to provide user satisfaction and sustainability of the user support as already defined in the first step i.e. vision. It is important to analyse the internal and external environment of the e-infrastructure before defining the strategy. In order to formulate any strategy, Mintzberg's four Ps are worthy to be considered (Office of Government Commerce, 2011):

- *Perspective:* Have a clear focus and vision
- *Position:* Take a clear defined stance
- *Plan:* Form a precise notion of how an organisation should develop itself
- *Pattern:* Maintain consistency and actions

In ESGF, the user support policy was not defined explicitly in the form of a clear vision and mission, in the past. The detailed user support service strategy described in FeUSE framework is defined based on the other frameworks such as ITIL and COBIT, their applications in academia, industry and governmental sector. At the same time, the specific strategy for ESGF user support is formed on the basis of the empirical investigation collected via the survey questionnaire, interviews with the ESGF stakeholders, participant observation and other methods from the business point of view. It includes the specification of the type of service to be delivered to achieve the vision of user support.

Before developing a user support service strategy, the perception about the audience of the service i.e. the service beneficiaries are needed to be clear. Likewise, all the resources including the human resources available amongst the partner institutions must be known, documented, communicated and made aware to all stakeholders. Thus, financial planning and budgeting must be done. The staffs carrying out the operations and services in an e-Science infrastructure along with the relationships among them must be visualised and documented in the form of charts and tables. In case of ESGF the overall chart of all the institutions and the data projects was visualized (H Chunpir et al., 2013).

Similarly, all types of services ideally offered by the e-Science infrastructure must be identified and approved by the e-Science infrastructure's top management from all the partner institutions in terms of a business value. Likewise, connections of these services to each other as well as the user support must be defined in a central document or a database known as *service portfolio*. Defining all the services offered by an IT infrastructure is known as *service portfolio management* in ITIL. It is a dynamic method to govern investments in service management across the enterprise (Office of Government Commerce, 2011).

Consequently, a clear service strategy for all services shall be defined in service portfolio of each service including the new coming service, the current services and the former services. This includes the *services to internal people* of ESGF organisation and to the *external people i.e. users or customers*. Furthermore, the status of *service demand* can also be set, for instance: *High*; the time the service consumption is at its peak, *Medium* and *Low*. (This phase is similar to the first phase of the ITIL lifecycle that suggests setting up a *service strategy* for IT infrastructure.) However, in FeUSE strategy management is tailored for the federated user support service for the e-Science infrastructures.

8.4.2 Proposed strategy of FeUSE framework

The five main aspects to consider forming a proper user support strategy to enhance user support, as recommended by the second step of FeUSE framework are as follows: (Note: The sub-steps of the second step of the FeUSE framework are labelled as numbers from 2.1 to 2.5.)

- 2.1 Know the stakeholders
 - 2.2 Form a governance scheme
 - 2.3 Organise
 - 2.4 Plan resources
 - 2.5 Stakeholder relationship management (SRM)
-

In ESGF until yet, no formal strategy exists on which basis the user support process is governed, organised, resources are allocated, relationship with the users and other all stakeholders are made. Similarly, all user communities, staffs, specialist and other stakeholders are also not known. However, there are some steps which have started in the recent past. Due to inception of the IS-ENES2 project, one staff (a highly qualified scientist) has been hired to answer the user queries and he has started analysing the incoming e-mails. But still, the phases defined in the second step (i.e. Strategy) of the FeUSE framework are missing.



Figure 8.2 Main points of the strategy of the FeUSE framework to enhance the user support services of ESGF.

8.4.3 Problems solved

The following identified problems from the empirical study of the current user support process “*as-is*” in ESGF can be solved by this phase:

- *P2: Who are the stakeholders?* This problem can be solved from 2.1.
- *P3: Who does what?* Responsibility, Accountability and Role definition can be solved from 2.2 by form a governance scheme and assigning roles
- *P10: Missing organisational structures in user support.* Organisation chart of the federated e-Science infrastructure and the user support directory (USD) in sub-phase 2.3 helps to organise and visualize the organisational structures in user support and other services. An example of organisational chart is given in below.
- *P6: No allocation of resources.* Resources must be planned and allocated for the user support in the sub-phase 2.4 *Plan Resources*.

- *P5: No central recorded documentation and relationship management between users, staffs and other stakeholders.* In the sub-phase 2.5 Stakeholder Relationship Management (SRM) the statistics of users and staffs is maintained about what users (work groups or community groups) ask what type of questions and these questions are tackled by what staff. At the same time what staffs does to build a positive relationship with the users and other stakeholders.

The five points highlighted in the “strategy” step are highlighted in the forthcoming sections, from Section 8.4.4 till Section 8.4.8.

8.4.4 Know the stakeholders

As discussed earlier, the basic information about a service must be specified including the *name of service, service description and user of a service* in a service portfolio (FitSM, 2014). Therefore, it is important know the list of stakeholders in the services of e-Science infrastructures. Since in this case the user support service is concerned, the stakeholders can be divided into end users and others. Figure 8.3 shows an overview of this sub-step. A user is a major stakeholder in the user support services and is considered extremely important. The following sections provide description about users and other stakeholders that are needed to be known as guided by the FeUSE framework.

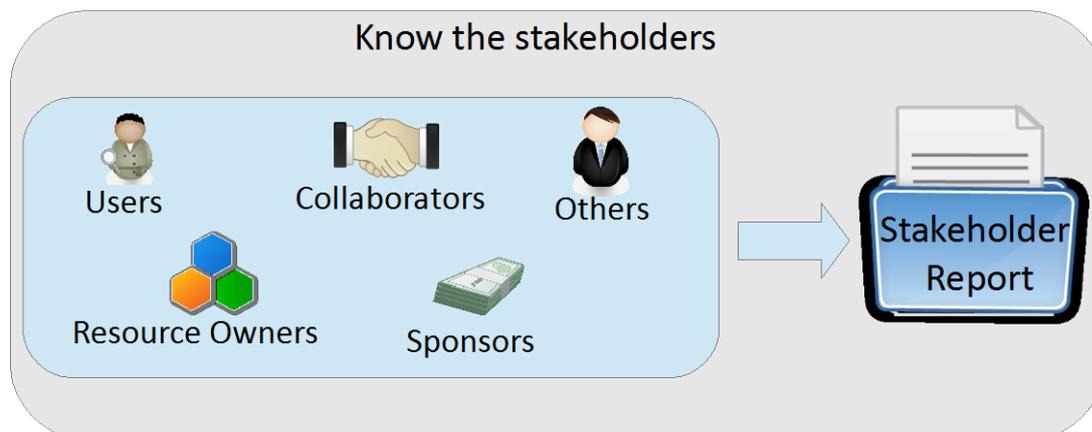


Figure 8.3 The sub-phase of the second step of the FeUSE framework.

8.4.4.1 Knowing the users (audience of the e-Science infrastructure)

The end user communities are the seekers of the data. One of the major end user communities in ESGF is CMIP5. Currently, ESGF is serving user communities of climate science. In future the new domains such as biology domain and energy domain data have been planned to be served by the ESGF.

The first point is to get a perception of the end users of the e-Science infrastructure for whom the infrastructure is made. In the case of ESGF, end users are the research communities related to different climate models and data projects served by the ESGF

e-Science infrastructure. All of the users of ESGF irrespective of their role and skills need to access data from the ESGF data-archive system. The users can be at any location in the world. The users of ESGF are not only limited to researchers but also commercial users and some other categories that include policy makers and journalists.

“Globally we have tens of thousands of users may be 20,000 to 30,000. People who are really engaged in this community are may be few thousand who are active in the past let’s say 12 months. We do not have metrics but every day there are almost up to 100 users who are downloading data throughout the world. Some are active and inactive. E.g. researchers who are doing PhD they get the data and then they become inactive”
Interviewee F, ESGF Master User.

1) Classification of ESGF users with respect to purpose of data access

The researchers who are using the ESGF data-archive system can be categorized into categories as shown in Figure 8.4, as according to the empirical investigation done in the ESGF. The two main categories as far as commercial-orientation of the audience of the e-Science infrastructure are concerned; *commercial* and *non-commercial* users.

Commercial Users

ESGF is mainly developed for research activities for climate science, but there is a potential of commercial usage as commercial users from insurance companies are already using the ESGF system. However, all climate data sets are not made available to them by the climate data modelling centres. It may be helpful to make policies for a business model and payment scheme for the commercial users. The commercial users shall be able to become paying members of ESGF with a fee structure such as subscription fees set by the financial policies. The fee from the commercial users can be used as revenue for the user support activities or other ESGF activities.

Non-Commercial Users

The non-commercial users can be further categorized into climate data users and other future domains (such as energy, biology, and chemistry etc.) data users. Climate data users are currently the main users of the ESGF. They can further be categorized into pure climate scientists, who do climate modelling and use climate data for their research, impact researchers; who research about impact of climate change e.g. on economy and integrated assessment modelling (IAM) researchers. In IPCC terminology, the climate scientists are known as working group 1 (WG1: The Physical Science Basis), working group 2 (WG2: Impacts, Adaptation and Vulnerability researchers) and working group 3 (WG3: Mitigation of Climate Change researchers).

Other domain data, such as energy, biology and chemistry are planned to be introduced in future (Dean N. Williams et al., 2013). All these user communities have their own issues, problems and demands and these may be incorporated into the policy of an e-

Science infrastructure. ESGF and other similar e-Science infrastructures shall create user-centric support services for the users that aim to retain the user communities for longer term and ease their research needs.

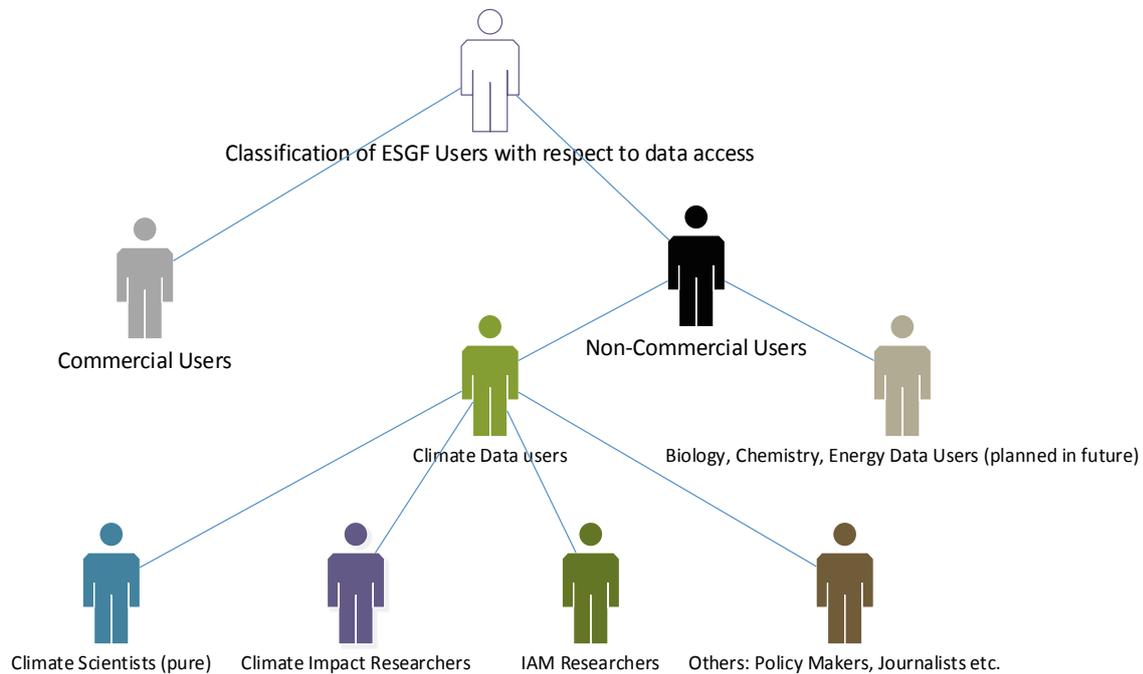


Figure 8.4 Classification of ESGF with respect to data download and access via ESGF system.

2) Classification of users of ESGF with respect to skill-level

Another classification of users of ESGF with respect to skill-level can be categorized into following categories as shown in the Figure 8.5.

- a. *Technical savvy users* also known as power users, who have enough technical knowledge, can write scripts etc. It is expected that they do not need much user support to access and download data from ESGF system.
- b. The ones who are *not technical savvy*. It is expected that they need more user support to access and download data from ESGF system.

“We have people that simply cannot use the system for number of reasons” Interviewee H, Climate Scientist.

*“Looking at e-mails that we get, there are **big power users** in the business since last 10 years. They know how to write script they need massive amount of data for climate analyses etc. Then we have simple graduate students who need particular data for their*

research. Then there are people from the private sector, who need to forecast climate for insurance companies” Interviewee D, System Administrator and ESGF Developer.

“We have students, researchers, scientists and then people in the private sector. Then we have developers and data managers” Interviewee D, System Administrator and ESGF Developer.

“There is one side which is climate scientists and there is the other side that is the help desk that would allow answering questions. Some of them can be done with better documentation” Interviewee B, Climate Scientist.

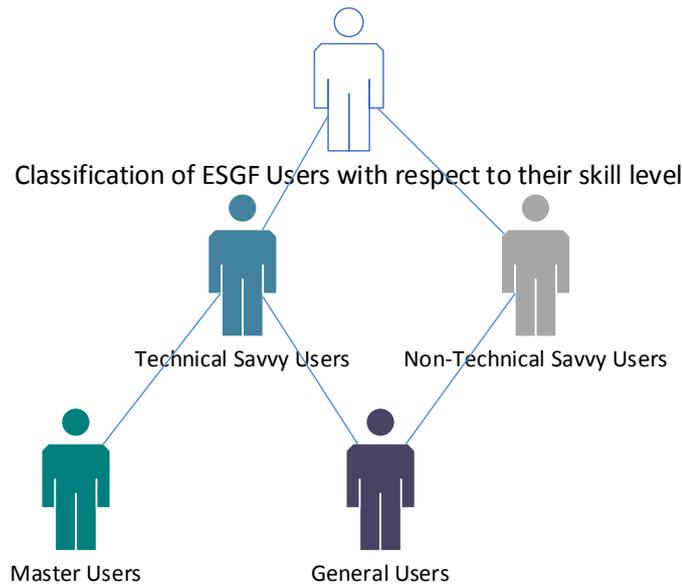


Figure 8.5 Classification of ESGF users with respect to technical affinity and skill sets.

The support services are also provided to the users, other than the end-user communities. These users fall under the category of collaborators of ESGF i.e. other projects connected to the ESGF who extend the services of the ESGF. For instance, the developers of ESGF at the partner institutions have many issues to be solved. Similarly, the data projects who offer their data connectivity to the ESGF need also support. The user communities and other data owners as well as ESGF extenders are all considered to be users. However, it is important to make a distinction between end users and others. The “others” category include mainly the staffs of organisations connected to ESGF for instance they are using ESGF software stack, extending it or connecting data to the ESGF. The fact is that the non-end user communities also need support but the separation is made. For instance the excerpts from the interviews explain the different nature and role of non-end user communities:

“There are users of the ESGF services but you can call them users too. So we can distinguish between data users and software users. E.g. in the EU there is climate4impact portal (ENES), this is a project where they search ESGF data holdings to process data, plot maps etc. These are high level software people who want to

connect with ESGF and write their own GUIs. The people who want to connect to ESGF and extend its services whether through a web-portal, desktop application including UI or whether tool like UV-CDAT in the US” Interviewee D, System Administrator and ESGF Developer.

“Errata page which are online and there are many different things that we do for support e.g. user support is not just accessing the data, we have support for the data center that how do they get the data in the right format. We had CMOR (Climate Metadata Output Re-writer). So we had to write the support of data center, like how to get data Net CDF semantics and supporting the CF (Climate Forecast) convention and supporting handling, publication of data, then, installation of ESG itself. Because you know ESG is not that easy to install” Interviewee L, ESGF Chair.

From the interviews, it is clear that the staffs of ESGF consider the software staffs from associated projects such as IS-ENES also as users. The data providers, who want to provide their data in the ESGF, so that it can be hosted to the researchers, are also considered users. However, making a distinction between all the end-users and others is important from end user i.e. audience point of view. The others who want to connect to the ESGF to use its software stack to host data can be called *service extenders or service collaborators*.

A user centric approach to design the services of the e-Science infrastructure can only be achieved once the attributes of the user communities of the e-infrastructure and other communities who contribute to the extension of ESGF are known. These attributes of the users of an e-Science infrastructure can be collected by a number of methods. First method is to capture important user attributes when a user registers for a user account to get access to the data offered by the e-Science infrastructure. These attributes include: *users affiliation, commercial and non-commercial usage, users can choose the data-holdings they are interested in e.g. CMIP5, PMIP, EUCLIFE and all others. What exactly group do they belong to e.g. Working Group classification, what instruments would they like to use? E.g. UV-CDAT for data visualization, scientific and technical skill sets and other special requirements to use the ESGF or its user support service.* Once this information is captured via the registration form, a basic guide on the ESGF system will be sent to them via e-mail. Once all the communities interacting with the ESGF are known, then user support policy can be altered according to the needs of the users. The needs of the users can be captured once they are encouraged to provide feedback. It is a key step that before making a user support policy, the policy makers are well aware of the capabilities and research needs of the user community. This step can further be strengthened by the F2F meeting where the user representatives are encouraged to participate, so that ESGF can attract the users and facilitate them in their problems.

Initially, at the start of an e-Science infrastructure when the information via registration about the users has not been collected; surveys, interviews, field studies and other methods can be used to get adequate information. In this way the perception about the users can be made vivid.

8.4.4.2 *Knowing the other stakeholders*

Getting the background information on all the stakeholders and their relationship with each other is instrumental input to the policy makers for setting a strategy of the ESGF services, especially in the user support service. Background information ascertains who does what and why as well as the scope and nature of the stakeholders. The responsibilities and accountability in the processes, especially in the user support, process can only be set after understanding the nature of activities that each type of stakeholder is meant to do. The main stakeholders of federated e-Science infrastructures including the ESGF community can be sub-divided into:

1) *Sponsors or funding agencies*

The sponsors or funding agencies basically pump in the money to programmes that run for several years with decisive goals. Their representatives form guiding principles on which the federated institutions orientate. They do influence major policies of the federated e-Science infrastructure. Partly they form policies as well. Examples of funding agencies are DOE, NASA and others.

2) *Data owners and data projects*

The data owners and data projects get the funding to prepare their data sets in an appropriate format, so that it can be delivered to the data users mainly for research purposes. The contact points (person) of each data project collaborate and cooperate with the e-Science infrastructure, so that their data can be made available via software to deliver it to the relevant audience. The data owners make data related decisions and set policies for data access and data replication. Examples of data owner projects are CMIP5, CORDEX and others.

3) *E-Science infrastructure owners*

The e-Science Infrastructure owners are responsible for the architecture, software and hardware functionality of the e-infrastructure. They do major changes in the architecture and software to enable federated infrastructure to function. The chairperson and the principal investigators are the policymakers for the services that the e-Science infrastructure must provide in the light of the guidelines provided by the funders. The e-Science Infrastructure has main sites and the federated sites. These include regional computing and data centres. In case of ESGF, the main site is LLNL, DKRZ, BADC, ANU are the federated sites who own resources, mainly data storage and compute

facilities. The policy makers for the services are normally principal investigators (PI's) followed by the technical operational staffs.

Data owners and e-infrastructure owners are termed as *resource owners* in the gSLM project (Appleton et al., 2014).

4) Service extenders (collaborators)

The service extenders are the projects that collaborate with the e-Science Infrastructure owners and use their resources as well as connect to them. An example in case of ESGF is IS-ENES project. Another example is the PRACE project. Both of the projects seek to federate large number of resources though they do not own them. The service extenders are known as *federated e- infrastructure operators* (Appleton et al., 2014).

In a nutshell, information about all the stakeholders is vital in user support strategy making. The representatives of each stakeholder must be involved in set up the strategy of user support service.

8.4.5 Governance

ESGF consortium is an international body of national collaborating institutions connected to each other in a federation. Every participating institute has its own norms, work culture, community, specialization, components (hardware/software), human resources, standards, standing (reputation), goals and expectations. Although, ESGF currently has its own governance policy based on the principles of “governance” developed and amended over years. However, yet the governance scheme needs a revamp because it is not clear to some of the stakeholders.

8.4.5.1 Criticism on the current governance scheme of ESGF

Additionally, the current governance scheme does not have specific user support governance structure that can be integrated in the over-all governance of ESGF. In ESGF federation, every participating organisation influences the governance of the ESGF. The user support in ESGF needs an explicit governance strategy that must be visible and understandable to all the participant institutions of the federation. This governance structure is comparable to general principles of governance or the “constitution” found in the arena of Political Science, adhering to the principles of: Critical ideas of fairness, transparency, measurability, responsibility, accountability and performance.

At the moment, ESGF governance structure is composed of the following constituents; see Figure 5.1 (page. 107) in the Chapter 5. ESGF has a review board known as “the ESGF Review Board (ERB)” which closely coordinates and monitors the three committees: Executive Committee (EC), Steering Committee (SC) and the Technical Committee (TC), see Figure 5.1. The main functions of all governance units namely

ERB, EC, SC and TC are depicted in the ovals near each unit in the same colour. The ESGF organization and the governance in place are primarily focussed on the ESGF data archive software development function and it does not include an explicit user support process and servicing strategy. The ESGF infrastructure is mature enough to leverage its services to users, therefore, it is important to make a policy that satisfies user concerns and makes ESGF service delivery efficient and ultimately lowers the costs of service.

Consequently, it is recommended to create ESGF Support Service Committee (ESSC) to safeguard the support services provided to the concerned groups of stakeholders. The ESSC comprises of three task-forces namely: ESGF Developers Task Force (DTF) to cater the needs of developers of ESGF, User-support Services Task Force (USTF) to cater the needs of users of ESGF and finally ESGF Data Management and Publishing Task Force (DMPTF) to cater the requirements of ESGF data managers, as shown in Figure 8.6.

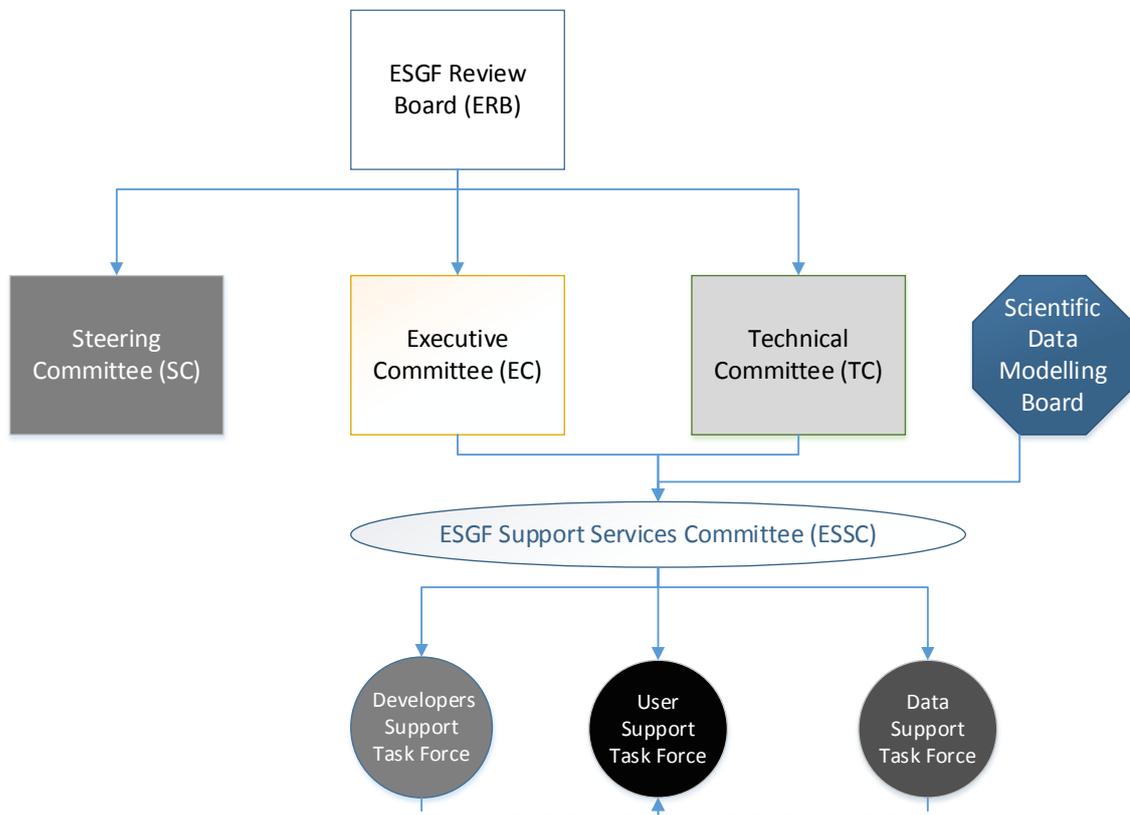


Figure 8.6 The new "to be" governance structure of ESGF.

In order to address user support concerns in the current ESGF governance scheme, user support committee or task force (USTF) is needed to be created. The USTF can comprise of staffs with concrete roles namely; the user support process owner (USPO), user support coordinators (USC) and members of executive committee (EC) as well as technical committee (TC). In addition, USC for scientific data modelling projects can also take part in USTF.

In Figure 8.7 the actors that are part of support services committee are shown.

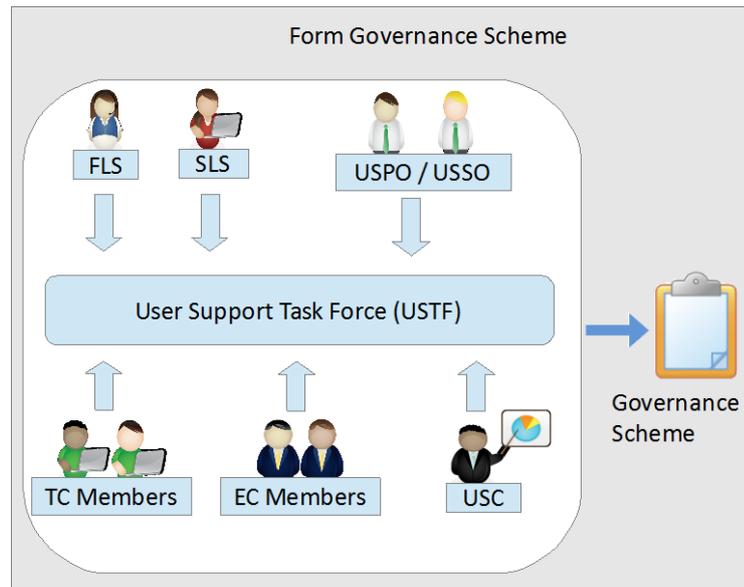


Figure 8.7 The sub-phase “form governance scheme” of the second step of FeUSE framework

The detailed responsibilities about these roles are defined in the forthcoming subsection.

8.4.5.2 Proposed roles in the user support process

All the roles connected to the user support process must be defined. E.g. User support process owner (USPO), User support coordinator (USC) (see Figure 8.8). The boundary of the assigned roles must be clear. The people assuming the role must be held responsible and accountable for their responsibility. Appropriate compensation and remuneration must be offered to the responsible personnel.

In ESGF, the user support process and activities is a responsibility and is taken by the following roles.

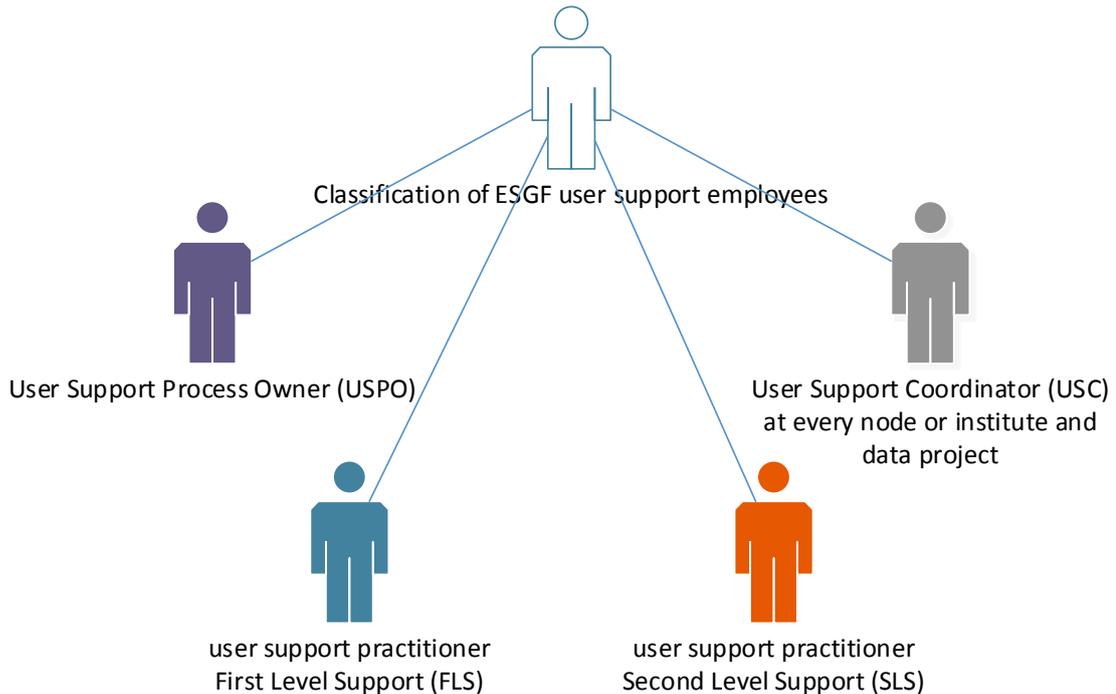


Figure 8.8 User support employee roles in ESGF. USPO and FLS are the full-time user support employees. SLS and USC (represented in red) do user support tasks on demand as the need arises. However, USC monitors the user support process.

1) User support process owner and user support service owner

FeUSE framework recommends preferably one person to be accountable for the user support process. Any suitable person, who is paid full-time for user support, can do this task. User support process owner (USPO) is in charge of the user support process policy making and is responsible for the development, innovation, extension and research of the user support process in ESGF. USPO is a dynamic role and interacts with the executive committee (EC), technical committee (TC) and users of ESGF to formulate and implement user support process policies. USPO allocates and monitors funding allocated for user support development and its innovation, called user support development fund (USDF). It is proposed that 6 to 9 % of the overall funding or resources dedicated to the development of e-Science infrastructure must go to user support. On top of it, the USPO handles public relation affairs (user relations) at the global level.

USPO develops strategies to attract new users and create awareness about the usage. He decides with other technicians from the TC to automate the user support process and monitors the automation procedure where-ever possible within the budgeting constraints. In case another employee, *User Support Service Owner (USSO)* can be employed additionally if the resources of the e-Science infrastructure allow. In this case USSO is the one in charge of the automation of the user support process. USPO is

accountable for the user support process. It is also recommended to employ at least a USPO.

ITIL suggested separate designations: USPO and USSO. However, in FeUSE framework the designation of USPO and USSO is merged together. Once the organisation grows or if need is felt, then USSO can be employed as well to improve the user support service, mainly for improving the tools and technologies used in the user support process. The decision to employ USSO is taken by the policymakers of the ESGF depending on the funding available to ESGF. Additionally some technical person may be assigned in the charge of USSO, if the policymakers decide to do so. For each service an owner must be assigned. This person must have accountability for the whole service, from a management point of view. USPO and USSO must be documented in the service portfolio of a service, in particular user support service.

Similarly, there must be logical and generic contact information for the designations in the user support process as also suggested by the FedSM project i.e.

servicename.enquiries@companyname.com . In the case of the user support service in the ESGF it is recommended to have usersupport.enquiries@esgf.org . In order to contact USPO and USSO, the following e-mail contacts are suggested;

processowner.usersupport@esgf.org and serviceowner.usersupport@esgf.org,

respectively. In case only USPO is present in the organisation, the contact information can be owner.usersupport@esgf.org .

In ESGF, currently there is one position of user support which is fully funded by the IS-ENES2 project. This position is suggested to be given the responsibility of USPO. At the moment, this position only expects answering of user queries.

2) *User support coordinators*

User support coordinator (USC) is responsible for the user support issues at a particular institute or data node. USC deals with all issues related to user support at the local level. Furthermore, USC monitors the implemented policies, standards and progress of user support process at local level. USC updates and maintains the directory of the specialists (SLS) or FLS present at the local organisation. USC may provide recommendation of future enhancement in the user support process. In FeUSE framework, USC responsibility is recommended to be assigned to the administrative heads at each partner institutions participating in ESGF. USC can be a *technical-USC* in case that the institute participates at the ESGF software stack development and data node maintenance. USC may be a *scientific-USC* in case of data modelling centre at an institute.

Technical user support coordinators (*technical-USC*) are present at the partner institutions having data nodes, HPC facilities as well as other relevant instruments at

institutes where ESGF software stack, ESGF associated tools are developed, network and hardware is maintained projects. Technical user support coordinators keep track of the specialists who may do the SLS (technical hardware and software support); in case scientific question has arisen from the users of the ESGF system.

An excerpt about the need of the scientific collaboration in ESGF:

“Then there are people who can use the system and they are looking for data so they are asking like what does this variable mean” Interviewee D, ESGF Developer and System Administrator.

The contact information for the user support coordinator is recommended in the FeUSE frameworks as:

parnterinstitutename.typeofcoordinator.servicename@organisationname.org e.g. for general coordinators the e-mail addresses can be suggested as follows:

dkrz.coordinator.usersupport@esgf.org, llnl.coordinator.usersupport@esgf.org and mpi.coordinator.usersupport@esgf.org.

In the case that both scientific and technical coordinators who handle the user support queries are present in an organisation, then the contact information can be as follows:

dkrz.technical.coordinator.usersupport@esgf.org and

dkrz.scientific.coordinator.usersupport@esgf.org

3) *Second level support employees*

SLS employees are the staffs of the partner institutions that are mainly involved in the development of the ESGF e-Science infrastructure. SLS are generally not directly involved in user support activities until a query is forwarded to them. The major activities of SLS may include programming, data curation, developing plans, policy making, architectural design, research work, documentation and any e-infrastructure component specialization. However, if the query arrives and their specialized knowledge is needed, then they assist to solve the problem. USC has a list of all SLS in a particular organisation that can assist FLS and users to fix the problems, perform workarounds or answer specific questions.

SLS can help in technical queries as well as scientific queries. Currently in ESGF, there is not a direct link to scientific staffs, such as data modelling scientists, to help the users in scientific aspects. Therefore, it is suggested to establish a link with scientific data modelling centres and put scientists from these centres also on board to answer scientific user queries. A distinction has been made in the FeUSE framework between the technical SLS which is depicted by “SLS-t” notation and the scientific SLS which is depicted by “SLS-s” notation. The contact of the SLS is suggested to be parnterinstitutename.SLS.componentname@organaisationname.org. An example may

be nasa.SLS.search@esgf.org, dkrz.SLS.access@esgf.org and badc.SLS.security@esgf.org.

4) First level support employees

In order to provide user support, a set of dedicated full-time or part-time user support employees are strongly recommended to be employed based on the demand of users to get support. User support employees must be responsible for FLS i.e. looking at the incidents, categorizing them, replying incidents, providing solution to their best of knowledge, forwarding requests that they cannot handle to the concerned specialists (SLS or their substitutes in case they are not available). The FLS staff must make sure that the user concerns are met. In addition, FLS must be able to write and update user support documentation, controlling their versions, updating the information in user support web-pages such as FAQs, Wikis, following the documentation standards as set by the User Support Task Force (USTF), the body taking decisions on user support activities.

“In general in ESGF the documentation is not complete as it should be. And it is a matter of resources and they spend all of the resources trying to get things to work and basically didn’t have enough to cover other things. This is because we needed all money to make the system robust and make sure to simplify the things at the node. The people who are managing the nodes, they were successful in doing that although there was a bit of delay” Interviewee B, Climate Scientist.

FLS must update the user support material (text, figures, audio, video files) on the concerned web-forums and user information portals. Depending on the qualification of FLS, FLS might be involved in the automation procedure of the user support process. FLS may also communicate the new user-support, ESGF UI or any other features deemed necessary to be introduced in the ESGF data-archive system or ESGF user support system. Moreover, FLS should follow standardized procedures to update the contact directory of SLS called USD (i.e. scientific and technical staff from the collaborating institutes) under the supervision of USC.

As far as the qualification of FLS is concerned, FLS can be IT-related graduate students or IT graduates with good communication, technical knowhow and language skills. In addition, they may possess some programming skills to automate part of user support process. An online-internship for up to six months can be offered on part-time basis to support users.

From the interview which was conducted in this research, a system architect, a developer and a principal investigator of ESGF, who suggested the number of FLS employees, are quoted as follows:

“Two persons in the US, two persons in the EU, one in Asia and one in Australia, totally responsible for answering questions, covering five time zones” Interviewees C, D and E.

The number of FLS people may depend on the total load of incidents and may vary from time to time. However, at the present situation it is suggested to have either two fulltime FLS or three part-time FLS, qualified students, employed to handle the user requests, one each from the US and the EU, respectively. The contact of the FLS is suggested to be partnerinstitutename.FLS@organaisationname.org

e.g. dkrz.FLS@esgf.org,

lnl.FLS@esgf.org.

Additionally, there are certain other elements that were highlighted by the interviewees during the interviews conducted with the stakeholders as follows:

- Each user query is a case.

“Each request should correspond to a ticket and the ticket remains open until it is solved. Then that ticket may be routed to the most appropriate scientist, technician or a data manager or an expert” Interviewee C, ESGF Developer.

- Every user has a right to get response.

“So we have to have that structure where every user gets an answer” Interviewee C, ESGF Developer.

- Replacement of people shall be documented.

“People get sick, for example for search I can be the main contact responsible for that, but if I am not available some other person should be able to answer the question” Interviewee C, ESGF Developer.

5) User Support Task Force (USTF)

USTF is a special committee which is formed to serve the interests of users, attract new users for more usage and improve the usability (i.e. ease of use) of the ESGF. USTF interacts closely with the data publishing and management support (DPMS) and the ESGF software team. In the long run, USTF aims to have a comprehensive user support platform that covers the complete user support process, automatizing it wherever useful for the ESGF user community.

In this chapter, only USTF is discussed in detail as this study is focused on user support process. USTF is composed of USPOs, USCs, EC members as well as TC members. All members of USTF, especially USPO, USC and FLS get together at-least once a month and discuss the short-term and long-term activities for future and the progress achieved

so far. USPO may liaise with other user support strategy developers in other e-Science infrastructures of other domains as well as industry.

The USTF is a directing, monitoring and evaluating body that directs monitors and evaluates the strategy of user support services as well as the responsible staffs allocated by the USTF. Moreover, it implements the policy at the local administrative bodies or the institutes participating in the ESGF consortium. Furthermore, USTF adapts the user support services to the changing needs of the e-Science infrastructure. The manifesto of USTF includes defining norms and standards in the form of SLAs³⁸ and OLAs³⁹ to make user support process reliable and monitoring these agreements at regular intervals.

Categories of user issues must be sorted out by unanimously agreeing on the categorization standards. All known and unknown problems must be documented. If possible, a database can be created. These problems and their corresponding categories can then be shared with other e-Science infrastructures as well. It is useful to present a list of all problems that users encounter and the list can be discussed in the meetings. Finally, a strategy can be made to reduce the user issues by solving the underlying root causes. The global user support directory (USD) must be maintained and updated at regular intervals. The USD will include all the FLS staffs, SLS staffs along with their components for which they are responsible and accountable.

6) Remuneration and reward to the support staff

“Of course there are people who spend significant amount of time doing that and they are needed to be paid. So we need to arrange money from the organization for this”
Interviewee C, ESGF Developer.

The staffs as well as users who spend their time helping the other users of ESGF in assisting to solve their problems are recommended to get monetary and non-monetary reward by the FeUSE framework. Different sorts of reward systems are proposed:

The three different types of ESGF certificates of recognition (both for users and staffs) that can be awarded are:

- *Certificate of active participation in the user support process:* This award can be conferred to users as well as staffs of the ESGF who actively support other users over a particular time frame e.g. over a year.
- *Certificate of appreciation for willingness to support users:* Awarded to people whose workarounds and solutions are liked and appreciated by others.

³⁸ Service Level Agreement (SLA) is an agreement between users who use the service and the service providers.

³⁹ Operation Level Agreement (OLA) is an agreement between different service providers within an organisation who provide operational services under the same organisation.

- *Certificate of appreciation for the most effective solutions*: Awarded to a person who presents the most effective solutions to the problems faced by the user community.

ESGF award of recognition (only for users and staffs):

- This award is conferred to a user or a staff who helps other users to solve the problem of user communities and helps innovation and further development of the ESGF e-Science infrastructure. The award may include a return flight to the ESGF conference and a souvenir.

ESGF certificates of didactics:

- *Certified User Support Trainer*: Awarded to a person who is able to train SLS and FLS
- *Certified User Support Professional- FLS*: Awarded to a person who is able to perform FLS activities
- *Certified User Support Professional- SLS*: Awarded to a person who is able to perform SLS activities
- *Certified User Support Social Media Moderator*: Awarded to a person who is able to moderate the conversation between users and staffs at user support platforms like Askbot
- *Certified User Support Process Planner*: Awarded to a person who is able to plan and improve the user support activities
- *Certified User Support Process Automation Engineer*: Awarded to a person who is able to automate key user support activities
- *Certified User Support Service Social Media Expert*: Awarded to a person who is able to develop social media channels for support activities

It is important to make the funding agencies understand that these certifications and awards must be sponsored too. The precondition of each award must be discussed by the user support task force (USTF) decision makers concretely, before introducing the awards.

8.4.6 Organization

An organisation chart is recommended to be prepared where the sponsors, data projects, the partner institutes as administrative bodies, the node types and the number of nodes and the resources that the partner institutes control are represented. Moreover, the number of human resources related to the user support such as USC, FLS, SLS and their relationships shall be mapped as well. In this way, the organisational chart and the user support directory (USD) can be prepared. Figure 8.9 shows the major outcomes of this sub-step.

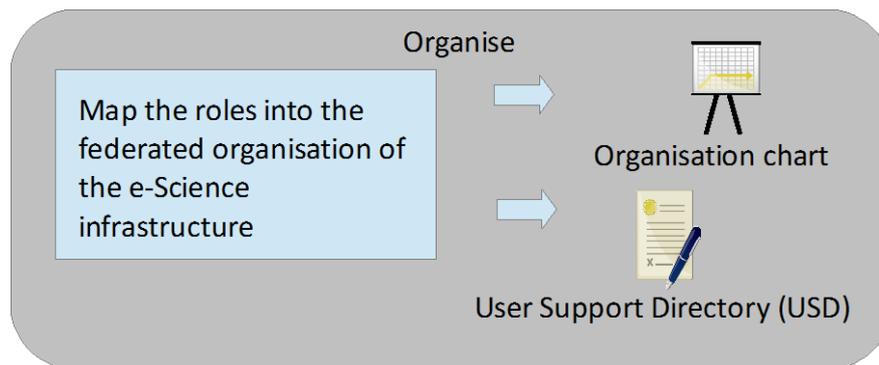


Figure 8.9 The sub-phase of the step two of the FeUSE Framework.

The organisation chart for user support within ESGF can be formed across the federated e-Science infrastructure, as already an attempt is made by the author, as shown in Figure 5.2 (at page 115 in Chapter 5). The four Ps; people, process, products and partners provide the main “machinery” of any organisation. People, products and partners are recommended to be represented in the organisation chart. Communication is the main element in the federated infrastructure that serves as a lubricant for the machinery (FedSM) and the flow of the communication may also be represented in Figure 5.2 in a way that one can map the user queries.

In ESGF organisation, user support process is needed to be recognized and introduced formally by all the administrative bodies participating in the ESGF consortium and its associated projects. Staffs of the ESGF shall participate fully in the user support process as per their capacities. User support process shall be monitored constantly and quality can be controlled as well as assured by the user support coordinator (USC) at the regional level at every participating organisation. A role of USC shall be introduced and assigned at every administrative body of ESGF comprising of a node or more nodes, as shown in Figure 5.2 (page 115). A user support process owner (USPO) is the main person(s) responsible for the user support service. From the interviews it is evident that there is a lack of user support organisation in the ESGF, see the following interview excerpts:

“We need to have a formal plan in answering user questions; this could involve some people in the world to do this. The US, EU, Australasia should try to do the same. We need to have a way to track the questions which has not been answered.

We need to have different group of people being in charge of different questions. We have people that simply cannot use the system for number of reasons...” Interviewee D, ESGF Developer and System Administrator.

The staffs who receive user queries must know to whom the queries shall be routed and re-routed to. This is possible via the organisation chart. The user support structure in

ESGF is proposed to have two-tier i.e. first level and second level user support. At the moment, there is no distribution of tiers in ESGF. The assignment of staffs especially FLS staffs depend on the demand of the users, if user queries increase in number then for a time frame more staffs that may provide support to the users can be engaged to supply the user support demand. Therefore, user demand shall be taken into consideration and a policy must be in place to manage it. According to the number of users and the number of queries, the number of FLS positions can be assigned. The FLS and the SLS may follow a user's support worker activity model (USWAM) to be efficient in their performance of servicing user requests along with other necessary tasks that are needed to be done. USWAM is proposed by the author and is not part of this thesis.

8.4.7 Resource planning

The resources available to the ESGF e-infrastructure must be clear and transparent at least to the top-management. Once all the resources are known to all the stakeholders, especially the policy makers, then the resource planning can be done in an efficient manner. Currently, the finances and other resources offered by the organisations in the ESGF are not clear and known to all the stakeholders. The resources available to an e-Science infrastructure can be divided into *soft resources* that include; software, people i.e. human resources, metrics, budget and *hard resources* that include data storage centers, HPC facilities and other hardware which is available to be utilized by the users and the staffs of an e-Science infrastructure.

All hard and soft resources available to the e-Science infrastructure must be recorded, documented and depicted in the form of charts and tables visible to authorized people, particularly people involved in planning and policy making of services including user support service. From this list of resources, the resources that are dedicated for the user support process in e-Science infrastructures shall be defined in the form of *service building blocks* and made known to the concerned stakeholders. The building blocks of the services are all the activities and components required for the service to function e.g. help desk.

The core resources are the key building blocks that must be available to the users of the e-Science infrastructure for the support service to be provided. The additional resources can be added as additional building blocks that can enrich the user support service further. Amongst the all services offered by an e-Science infrastructure, (also specified in the service portfolio) the core resources allocated, as well as additional resources to each service shall be listed. FeUSE framework strongly recommends it to be specified in detail for the user support service. One can also call core resources as *core service building blocks* and additional resources as *additional service building blocks*.

Moreover, the relationship amongst these building blocks along with the dependencies must be specified too.

A decision must be made that whether the financial resources in the form of revenue can be generated from the user support process. It is vital to decide whether to charge for providing support service to commercial users of the ESGF infrastructure as well as for the access to data hosted by ESGF. In this case, additionally it is important to determine that how much it costs to provide a user support service and whether the e-infrastructure would like to charge profit for it or not in the form of pricing or subscription fees.

Currently, in ESGF no pricing strategy and business model has been indicated so far. Financial policy is needed to be set in this regard, specifically for the commercial users or customers. It might be possible to split the costs of providing user support across the services offered by the ESGF. Cost can be calculated per unit e.g. €per CPU hour for processing, €per terabyte of storage, for subscription cost per month for access. The cost of service at first might be a rough estimate but later it can be refined over time.

Financial planning ensures that the funding for user support is transparent and that the authorized stakeholders of the ESGF understand it. It is important to determine from where (which sponsor) the funding comes in, how much it is and for what activity or component the money is allocated in the user support process. Funding can be taken from the sponsors and might be generated internally via selling a service. In ESGF, almost all of the funding is done by the sponsors. Some sponsors dedicate more funding for user support whereas others dedicate less. Since the user support of ESGF is for all users and differentiation in servicing their user requests is not made, thus, user support is done irrespective of the funding weightage.

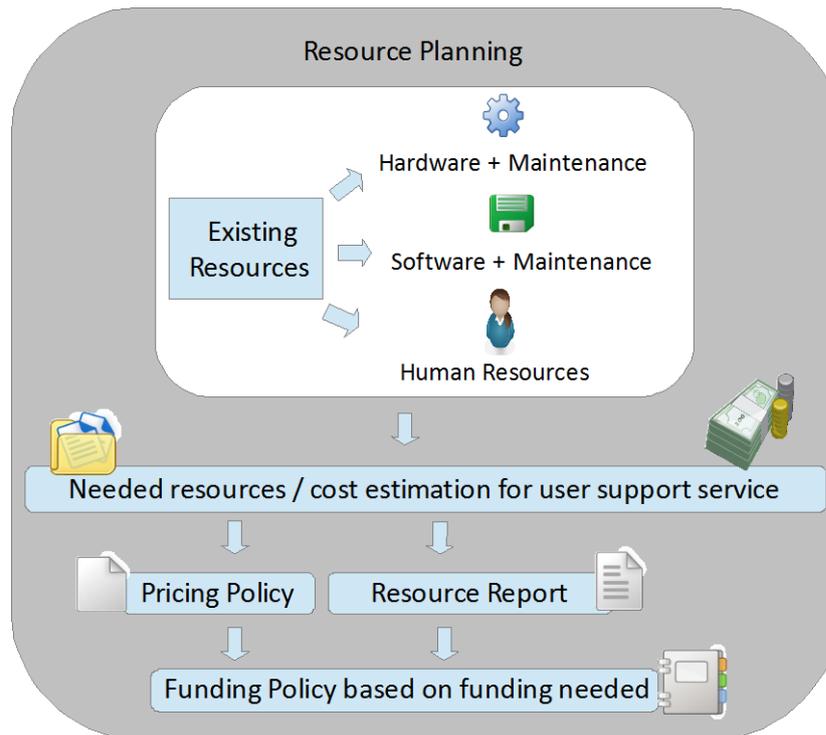


Figure 8.10 The major steps in the resource planning sub-phase of the second step of the FeUSE framework.

Figure 8.10 shows the major steps involved in this sub-phase that include cost estimation of the user support service based on the resources available and resources needed. The outputs are resource report that documents the resources available and the resources needed. The pricing policy and the business model is also an outcome of this sub-phase and these outputs serve as an input to the overall funding policy allocated for user support service.

User support task force (USTF) must also look for and apply for an adequate funding from the sponsors and set budget for user support services. The funding must be then invested in an efficient manner. In addition, USTF must investigate the need of providing automation in the user support process necessary after the cost benefit analysis. For instance; automatic answering of request via case based reasoning (CBR) and information retrieval.

8.4.8 User-staffs relationship management

User relationship management policy manages the links between the services provided by the ESGF or any other e-Science infrastructure and the users especially in the user support process. This may include staffs present at the different locations around the world. It sets the strategy and provides tactical guidance to involve users in the user support process and understanding as well as prioritizing their requirements to meet their needs. The policy of user relationship management must address to the changing

needs of the users. The primary measure that the user relationship is being achieved is level of user satisfaction. It includes mediation in the case of conflicts.

It is vital to invite and encourage users as well as staffs of ESGF to provide feedback on the ESGF operations and services. These suggestions may lead to inclusion of new features i.e. user driven development and enhancement of ESGF as well as its user support. In addition, short surveys to capture user and support staffs satisfaction level must be conducted at regular intervals. The users can be involved in training sessions, workshops, meetings, conferences and teleconferences at local as well as global levels. If researchers and open-source contributors are attracted to provide feedback the investigation and development of user support process of ESGF, usability and scalability of ESGF can be improved drastically. Furthermore, data modelling centers can be contacted to provide more data as well as data support in the ESGF data archive system. The point is to attract researchers and open-source contributors to enhance ESGF and ESGF user support. Figure 8.11 shows the major outputs of the user-stakeholder relationship management sub-phase in the form of user statistics and user and staff relationship report that are helpful in orienting the strategy of the user support in e-Science infrastructures.

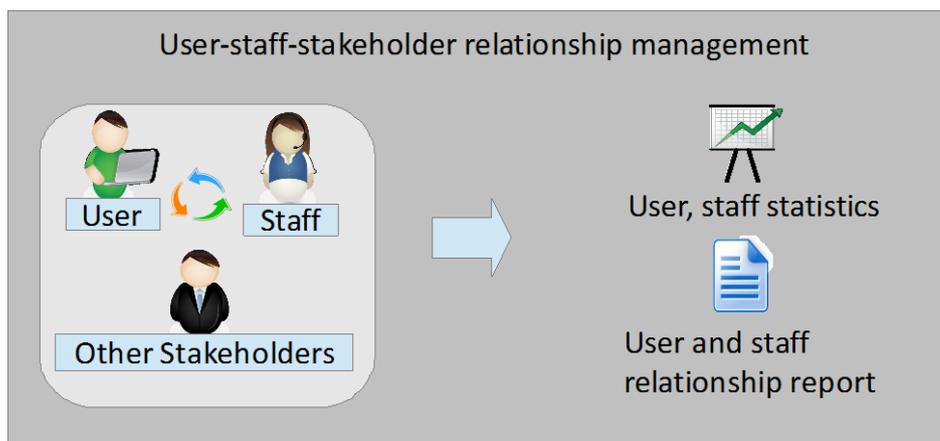


Figure 8.11 The main output of the user-stakeholder relationship management sub-phase.

Policies on possible incentives to the users and the support staffs may be introduced in order to improve the user support process. Responsibility and accountability must be allocated to staffs. To make the user support process interesting, gamification techniques can be applied. Similarly, self-help can be promoted by disseminating understandable task-oriented support information to the users via ESGF web-pages, wikis, in the form of documents and online tutorials. Moreover, user statistics must be collected and communicated to interested stakeholders of ESGF. Eye-tracking technologies can be applied to evaluate and enhance UI components of ESGF. Commercial access may be granted to the commercial users with some fees. For

instance insurance companies can use climate data. Finally, appropriate metrics such as usability surveys to measure the quality of user support must be introduced.

8.5 Communicate the strategy

The vision and the strategy regarding the user support must be communicated to all the stakeholders especially the policy makers from all the federated partner institutions, to achieve the goals set by the user support task force (USTF). The communication is done to create awareness via meetings, proposals, white papers, web-pages and board meetings to all the stakeholders, see Figure 8.12, and the feedback is collected.

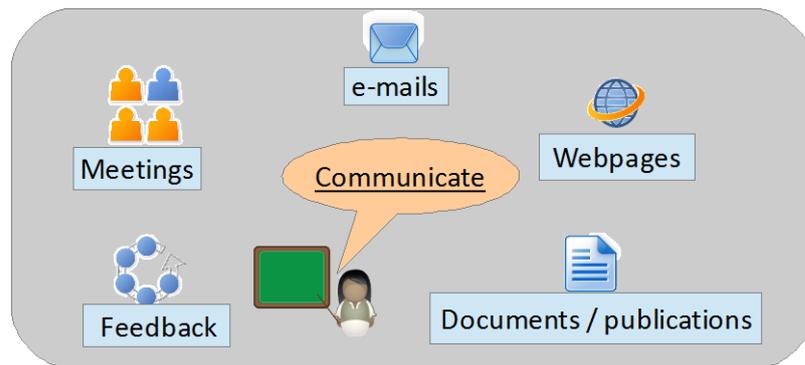


Figure 8.12 Step three of the FeUSE framework; communicating the strategy to all the stakeholders and collecting feedback.

On the user support strategy, the feedback from the stakeholders can be collected, considered and if necessary some changes can be made in the policy, only if it is feasible. It must be ensured that all the stakeholders are on the same page about the user support policy. In this phase, it must be ensured that the stakeholders must understand the manifesto of the user support process i.e. the set policies, the need to allocate resources for the user support activities and finally they are in line with the implementation of the user support strategy federation wide.

8.6 Agreement and assignment

Forming agreements is the most critical aspect in federated e-Science infrastructure. The agreements can be reached using a voting process where every partner institute of the federation has a right to vote on critical matters. It is recommended that one partner institute can cast one vote which is agreed within the institute. The right of vote and number of votes per partner institute can further be discussed within the federation by the policy makers. Once all the stakeholders know the vision and goals; agreements on how to implement the strategy in the form of policies can be achieved in its true spirit. An agreement must be reached in this step by all the policy makers for the user support service on the delivery of support service and its associated services. The aim is to ensure that a particular agreed level of service provision is attained and maintained.

In this step, agreements shall be made via consensus on the areas such as establishing priority of incoming user issues, service level agreements (SLA), operational level agreements (OLA) and service level delivery (SLD). The demand and availability of service must be monitored regularly. As a result, SLA and OLA charts must be maintained to gauge performance against the set agreements at the end of each year.

Agreements on the *Information security and access policy* are also done in this phase, see Figure 8.13. *Information dissemination policy* and efforts to attract potential users is needed to be agreed as well. If agreed, the tools to support users must be developed or a license shall be acquired federation-wide, collectively. If the tools to support users are developed in house then these tools can be licensed to other e-Science infrastructures or firms to generate revenue.

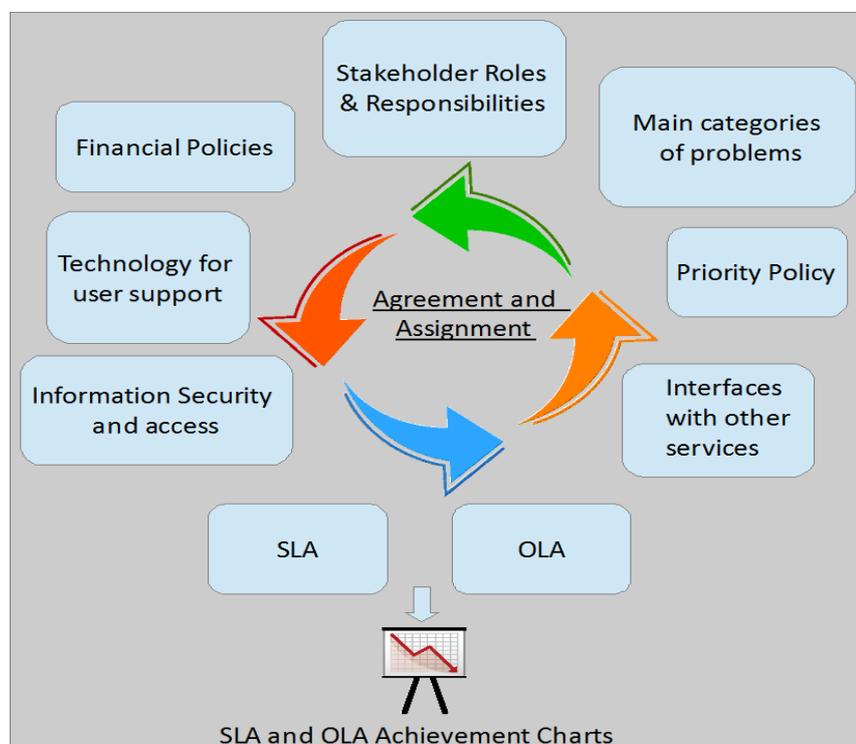


Figure 8.13 Agreement of the stakeholders and assignment of responsibilities to all concerned stakeholders for the user support process in federated infrastructures.

8.6.1 Priority policy

Priority defines the severity of problems and has an impact on the support centre's resource engagement time line. According to FeUSE framework, urgency of all the incoming user requests shall be determined. *Urgency* is defined by the impact of the issue, from the user's or customer's perspective, on the individual, overall business, or organization. The support center uses urgency level, along with other factors or pieces

of information, to determine the priority level. If a user makes a support request by telephone, e-mail or web submission form, the support centre frontline representatives (FLS) will help a user setting an urgency level using the same criteria that defines the priority. The FeUSE framework recommends the following levels priority:

A. *Urgent* (Emergency e.g. e-infrastructure down): Defects or major problems that interrupt or halt operations, for instance; service of e-Research or e-infrastructure in climate science. Some data or e-Research results might be date/time (calendar) driven or dependent upon hard deadlines; these issues may also come under the urgent category. Urgent priority support requests take precedence over all other support requests.

B. *High* (Making e-Research difficult): Defects or major problems that might slow down e-Science infrastructure. However, with technical intervention or a functional work-around, the user can complete the current e-Research cycle or access the hosted service, e-Science application or web-based information. Examples include, but are not limited to WGET scripts.

Some issues might require the following fixes via:

- Scripts
- Patches
- Revisions to web-based information
- News releases
- Other announcements

High priority support requests take precedence over medium and low priority support requests in a specific area of support.

C. *Medium* (Standard problems): Problems related to development, system administration, and management issues of e-infrastructure updates not defined as urgent or high.

This is the most common priority level and is generally processed using a “bank queue” FIFO (First in, First out). Medium priority support request takes precedence over low priority support requests.

D. *Low* (Petty issues): This includes problems that are less critical e.g. general questions on use, informational items, testing, or minor defects that are reported. However, without solving these issues normal functionality of e-Science infrastructure or implementation can continue. Low priority support requests are processed when there are no urgent, high and medium priority support requests in a specific area of support.

8.6.2 Service level agreements

Service level management (SLM) ensures that the agreed levels of services are met by the service providers. SLM is a process that includes planning, coordinating, agreeing,

supplying, monitoring on *service level agreements*. SLM ensures a quality of service and provides warranty for it. An SLA is a written agreement between a user or a customer and the organisation providing the services. It binds mutual goals and responsibilities on both of the parties. All the details in the SLA's must be clearly defined otherwise disputes may arise amongst the federated partners.

Before defining any particular SLA for any service provided by the e-Science infrastructure, it is vital to know requirements for the services for the users; service level requirements (SLR). SLR for each services offered by the e-Science infrastructure can be determined from statistics and experience of stakeholders. In FeUSE framework, however, the focus is on user support service.

An implementation of SLA makes it binding on the staffs to respond or react within a specified and agreed time frame e.g. agreed response time and other agreed timings. *Response time* is the period between the users contacting the support centre with a problem, question, incident or a support request and the time when the FLS or SLS representative acknowledges a receipt of the request, begins diagnosing or exploring the issue. Response time also refers to the initial response to the request, and to a response to an additional request on an existing support request.

In case of user support service of ESGF and other similar infrastructures, the following timings are worth considering to be agreed upon for answering the user queries, as suggested by the FeUSE framework:

- *Mean personal response time (MPRT)*: The time when a user receives a personal response from the human support agent.
 - *Mean time to provide a workaround (MTPW)*: The mean time to provide a user a workaround from the human support agent(s).
 - *Mean time to provide a solution (MTPS)*: The mean time to provide a user a workaround from the human support agent(s).
 - *Mean time to provide user feedback (MTPUF)*: Once the workaround or a solution is sent to the users. It is the responsibility of the user to provide feedback. If no feedback is provided then it is assumed that everything is alright and the problem is solved.
-

The following Table 8.1 shows the ideal MPRT and MTPW as well as MTPS, as recommended by the FeUSE framework. Similarly,

Table 8.2 shows the target MTPUF, as recommended by the FeUSE framework. The priority of the user queries are divided into urgent, high, medium and low as shown in these tables.

Table 8.1 Table shows the ideal target MPRT and MTPW as well as MTPS as recommended by the FeUSE framework.

	Urgent	High	Medium	Low
Respond to user within (MPRT)	1 hour	2 hours	4 hours	4 days
Resolve or provide a workaround within (MTPS / MTPW)	12 hours	2 days	4 days	3 weeks

Table 8.2 Table shows the target MTPUF, recommended by FeUSE.

	Urgent	High	Medium	Low
Provide feedback on a solution or a workaround within (MTPUF)	48 hours	5 days	10 days	3 weeks

8.6.3 Operational level agreements

Similarly, an operational level agreement (OLA) is an agreement between the service provider body and other partner(s) of the same organisation. In case of ESGF, an OLA can be between partner institutes that is providing some services to another partner institute. For instance; a particular component and data set is needed to be available to another institute which is part of the ESGF. Similarly, an FLS may refer the user's request to a concerned SLS by referring to a central user support directory (USD) maintained by the ESGF and expect a response from an SLS within a specific time frame.

- *Response time between staffs (RTBS)*: The FLS may refer the user's request to a concerned SLS of a federating partner by referring to a central user support directory (USD) maintained by the ESGF. For this, an agreement about the response time between staffs (RTBS), whether FLS and SLS or between SLS, is important. Response time between staffs (RTBS) is defined as the time when a staffs receives a query from the human support agent and forwards it further to another staff of the federation.

Table 8.3 Table shows the ideal target RTBS as recommended by the FeUSE framework.

	Urgent	High	Medium	Low
Respond to staff within (RTBS)	1 hour	2 hours	4 hours	4 days

Since all services delivered by the e-Science infrastructure depend on the federated components, therefore the availability of components do affect the availability of services, including user support service. The service availability includes all aspects of components availability as well as unavailability and the potential impact of components availability and unavailability on the service. The support service also depends on the potential availability and unavailability of the human resources available to service user requests. The reliability indicates how long a component can perform its agreed function without interruption. The maintainability of service or components indicates how fast it can be restored after failure. In agreed level of support service supply, the availability, reliability, capability and maintainability of components as well as human resources is needed to be considered and agreed upon.

The services provided to the users in the e-Science infrastructure such as ESGF depend on the availability and capacity of the federated components. Therefore, an agreement is recommended to be done within the federation on the availability and capacity levels of the components. The following KPI's can be agreed upon and measures to improve the reliability of components involved in providing services relevant to support service can be set:

- *Mean time to restore a service (MTRS)*: If a particular service is interrupted, suspended or not available, then the mean time to restore a service (MTRS) must be set. A particular service must be agreed upon by the policy makers and other stakeholders of ESGF.
- *Mean time between failures (MTBF)*: The average time that a service or a component can perform its agreed function without interruption.
- *Mean time between service incidents (MTBSI)*: Mean time when a system fails until the next time it fails. E.g. the mean time once the node is up and servicing; if the node is down, then the node is up and when it goes down again.
- *Mean time to repair (MTTR)*: The average time taken to repair a system or a component which is a part of an ESGF.

8.6.4 Service achievement charts

Both OLA and SLA result in standardization of a process and contribute to ensure customer satisfaction. Any breach of the agreements is recorded i.e. logged and complaints or compliments can be logged too in the form of *service achievement log* and finally *service achievement charts* can be generated.

The services achievement charts are recommended to be made to monitor performance of the partner institutions in delivering user support service and other services by FeUSE. Services achievements charts depicted in the form of Red, Amber and Green (RAG), also recommended by ITIL. The FeUSE framework recommends the following colour depiction scheme:

- If the services agreements are not reached, the achievement charts are depicted *Red*; in this case there is much deviance from the agreed terms.
- If the services agreements are reached to some extent (more than 50% but less than 85%) the achievement charts are depicted *Amber*; i.e. there is a deviance amongst the agreed terms but not much.
- If the services agreements are reached (more than 85%) the achievement charts are depicted *Green*; i.e. there is a little or no deviance amongst the agreed terms.

8.6.5 Interface with service catalogue

Service catalogue comprises the details of all the services and service level requirements. The user support service has interface with other services as well. The outputs from other services can be an input to the user support service. Similarly, outputs from the user support service can be an input to other services. Therefore, FeUSE framework recommends maintaining an interface with the service catalogue.

Currently, there are no practices to document these services that an ESGF e-Science infrastructure offers, in the form of a catalogue, which may be visible to the stakeholders of the ESGF as well as the public. As different end-user communities and other stakeholders view ESGF differently, depending on how they would like to use the infrastructure and to get benefit of the services that the ESGF offers, it is imperative to document and maintain the services that the ESGF e-Science infrastructure offers. For instance: A computer scientist may be interested in a working component e.g. a particular software; whereas a climate scientist is interested in the data that she or he gets and yet, user support personnel are interested in helping users. So there is a diversity of use and utilization of services.

“At the moment, no practices exist in ESGF where all services that it offers are shown to the stakeholders. Different communities that use the infrastructure have different views how they want to use the infrastructure. Similarly, different contributors to the

infrastructure have their perspectives of the priority of contribution to the ESGF.” S. Kindermann, ESGF Architect at DKRZ.

FeUSE recommends to have a service catalogue where the list of all the services offered by ESGF are listed in such a way that the contributors of the ESGF can claim that this is our product that we have developed and these are the services that we offer to our diverse stakeholders, mainly end users.

The catalogue shall include *current services* that the e-Science infrastructure offers, the *retired services*; the services that were offered in the past and are not available any longer. Furthermore, the *future services* that the e-Science infrastructure is developing to offer must also be part of the catalogue. It is recommended by the FeUSE framework to include following fields in the service catalogue:

- List of all services (current, future and retired)
- Audience of services
- Benefit of the service and the reason why it is provided
- Budget and resources allocated for services
- Type of services (major and minor)
- Changes and updates done to the services in the last year
- Incharge of services (name, designation, institute, contact)
- List of contributor to the services and their e-mails
- Date the services were introduced
- Interface with other services
- Major components of a service and their locations

It is essential for the management to collect information about all services and the perspectives of stakeholders associated with this. User support process owner (USPO) must get the relevant information about the attributes mentioned above. It is interesting to note that the e-Science infrastructure environment is so agile and dynamic that the services that it offers are constantly changing.

Similarly, same can be applied to a system or a component of the federated e-Science infrastructure. In a nutshell, agreements must be made on

- Strategic plans including user inclusion in the user support process and capturing their requirements,
 - competencies of support personnel,
 - roles of support people,
 - available resources and their update,
 - MTRS of systems and components,
 - technologies employed used in the user support process, their configuration and development,
 - reactive and proactive procedures as far as user support is concerned,
-

- all procedures in the user support process,
- information security and access issues and
- dealing with redundancy.

8.6.6 Agreement on problem categories

The categories of problems, the possible solutions and the workarounds must also be agreed upon. Information security policy about the users and the staffs involved in user support is also needed to be defined eloquently. The information security policy must be in line with the overall information security policy of the e-Science infrastructure. The top executive and the policy planners of the federated ESGF must cover policies regarding asset control, information access, data access, password settings, registration issues, escalation issues, access to the media such as e-mail, RTS, websites and document control. The distribution of information must be sent to authorize users and workers of user support.

The following are examples of the main problem categories based on the empirical results given in Table 6.5 (page 168) and Section 6.8.1.8 (page 167):

- i. Data download problem (specify problem and the data set as well as data project which is problematic)
- ii. Workflow problem
- iii. Infrastructure problem
- iv. General query
- v. Management query
- vi. Download Format query
- vii. Scientific query

8.6.7 Stakeholder roles, risks and responsibilities

Once the policies have been designed, discussed, decided (DDD) then clear roles and responsibilities must be dispensed through authorized personnel (i.e. executive members of ESGF). Additionally, risks associated to each responsibility must also be defined, if the responsibility is not done. This is accomplished by assigning the mapped activities to the concerned staffs. At this stage, all stakeholders must know their roles and responsibilities and they are accountable for all the actions. In the current user support process, the incoming user issues are worked upon, mainly by the user support staffs on volunteer basis. Since the study about user support on ESGF started, there is one position which has been dedicated to a full-time user support staff at DKRZ recently.

RACI model is important in this regard to determine the roles of people involved in processes in a structure of an organisation (Office of Government Commerce, 2011). FeUSE framework recommends RACI chart to be made for people involved in servicing user requests. RACI is an acronym for *responsible, accountable, consulted*

and *informed*. According to the RACI principle; the individual activities must be mapped to the concerned people in the user support process. RACI model defines an authority matrix which is used to explain roles and responsibilities of workers involved in processes and activities across the organisation. RACI also presents OLA opportunities. It explains who, what, where, why and when the appropriate person needs to work together with others during a service or process execution.

The roles and responsibilities must be assigned to users, USPO, USSO, USC, FLS, SLS and any other roles defined in the user support process. In FeUSE framework, it is highly recommended to form RACI matrices in each partner institution of the ESGF e-Science infrastructure. Moreover, there is no agreement on classification of tier 1 (FLS) and tier 2 (SLS) in the current user support process in ESGF, which shall be done. FeUSE recommends a two-tier user support structure. The responsibilities and roles are attached to each and every actor who is interacting with the user support process.

8.6.7.1 User responsibility

A user has a right to seek support while interacting with an e-Science infrastructure like ESGF. However, a user of an e-Science infrastructure is also encouraged to help other users to provide support. In the first step, the user needs to determine whether he or she needs a support or a user can solve the problem him or herself by consulting online or offline material.

Ideally a user of e-Science infrastructure should be able to perform the following steps:

- 1- *A user must try to solve the issue* by him or herself via self-service support by consulting documentation and online support information available at ESGF information portal at first. A user may explore the self-help environment i.e.:
 - Review the technical information available online and offline (e.g. refer to component documentation for more information, review the FAQ'S or other related documentation). View video tutorials if available on a particular topic online.
 - Review the scientific information in case of a scientific query (e.g. refer to scientific documentation for more information).
 - Anticipate where and under what conditions, various errors might occur. A user may try to isolate the problem, e.g. check whether the problem is with the data download or data format, work flow or some other component.

 - 2- If a user cannot do self-service support, the user must seek help of a professional or a colleague or consult home institution's problem analyst i.e. technical staff, if available at the institution where an ESGF user is based at. Experienced
-

institution personnel are the best resources to provide primary support for applications, services and infrastructure.

- 3- If a user cannot do self-service support and seek help at his institution, he might submit a request to the First Level Support (FLS) via e-mail or a web forum as specified by the user support centre of an e-Science infrastructure. The FLS may supervise the users to get the problem solved by pointing at e-Documents or well documented FAQs. A user shall have a possibility to point out the category where the problem may belong to. However, a user might assign up to three categories and if a user cannot locate a particular category or area then he may not assign any category.
- 4- Any diagnostic information including error message number(s), if applicable, shall be provided by the user including diagnostic information in the form of text, attached screenshots, video etc. (if any possible) to the FLS. Additionally, contact information such as user name, telephone numbers, e-mail address and name of the institution can be provided to support staffs.
- 5- A user may specify the urgency level (i.e. urgent, high, medium, low) however, later the support staff determines the urgency based on the impact of the issue.

An e-mail based method is the primary channel of communication in ESGF, as according to the survey-questionnaire conducted in ESGF, as evident from Sections 6.8.2.1 (page. 170), 6.8.2.3 (page. 172) and 6.8.2.2 (page. 170). It is however recommended to employ user request patterns if found feasible by the executive staffs of ESGF. In ESGF, a user query sent via e-mail might lead to the mailing-list or to RTS and initially treated by FLS. If FLS cannot handle it, the query is sent to SLS for further treatment. Finally, the solution or a workaround is sent back to the user after some time frame as agreed by the policymakers in ESGF.

8.6.7.2 Support centre responsibility

According to the FeUSE framework, the support centre of the federated e-Science infrastructure such as ESGF shall be established as an entity and must take responsibility to provide support service to all possible stakeholders, chiefly end users. Apart from end users, software developers, data publishers, data curators and others may also take advantage of the support centre facility.

Step 1: FLS establishes the priority of the incoming issue. Refer to establishing priority definitions in the above Section 8.6.1.

Step 2: Engages appropriate support specialist (groups) i.e. SLS, if necessary.

Step 3: Responds and resolves support request based upon target response and resolution times. Refer to *target response and resolution times* as stated in the Sections 8.6.2 and 8.6.3 above.

Step 4: Communicates the solution to the customer, solicits customer feedback about the solution, and closes issue when appropriate.

The solution procedure and closure procedure of the issues shall be thoroughly documented and guidelines should be given to personnel who handle these issues and suggest solutions.

Recommended hours of operation:

The hours of operation of the user support centre can be decided by the ESGF federated partners keeping each time zone in consideration. FeUSE framework recommends to have operational hours of each institute where FLS and SLS are present from Monday to Friday in a week from 9:00 a.m. to 6:00 p.m. Exceptions are national holidays and announced scheduled holidays. The scheduled holidays are subject to each region where institute is located.

Factors affecting target resolution times:

While the support centre makes every effort to minimize conditions that may have an impact on resolution target times, the following factors may affect target times for *urgent, high, medium* and *low* priority support requests:

- Host announced holidays and other announced closings.
 - Service (or business) interruptions affecting many users and multiple federated sites.
 - Completion of checklists that the support center may require of an institution before going live with a new or updated service, such as local database or data-archive upgrades, middleware updates, workflow updates etc.
 - Lack of timely or accurate information provided to the support centre, such as information to begin incident or problem diagnosis or resolution, or confirmation of a resolution.
 - New developments that may require remediation and continued testing, such as scripts to be run through job submission
 - Dependency upon a third-party partner or vendor, such as manufacturer defect correction
 - Complexity of an incident or problem (e.g. performing a database intervention may take longer than resolving less complex issues)
 - Support resources may be redirected to day-one, go-live implementation issues that impact multiple sites and users and this may take longer resolution times for other priority support requests
 - Acts of nature (weather), power outages, or unexpected occurrences where demand exceeds service or support supply limits
 - FLS or SLS misinterprets user communicated symptoms or staffs misinterpret urgent, infrastructure down, or emergency guidelines (support centre must provide adequate training to handle issues)
-

- User does not reply to communications or discontinues engagement with the support centre staffs or other support resources
- Incident or problem meets guidelines to force support request closure

8.6.8 Agreement on technology for user support

A consensus on the best medium and technology to communicate to the user support must be determined by the policy makers of the ESGF. A review of the methods and technologies used in ESGF user support process are provided in Section 7.3.1 and Section 7.3.2 in Chapter 7.

8.6.9 Agreement on financial and funding policies

FeUSE recommends to have agreements on the financial resources that are dedicated for the user support service and the other services provided by an e-Science infrastructure. The funding for the user support must be transparent. The funding bodies (sponsors) must understand the value of money allocated for the user support. The rules shall be made to attract funders who invest more into the infrastructure and the user support. Similarly, a clear business model for the e-Science infrastructure and the services that it provides must be developed unanimously by the executive members. FeUSE recommends charging commercial users for data access, download, compute and support facilities provided to them. However, the exact arrangement and agreement on financial policies of e-Science infrastructures shall be made by the executive staff members.

8.7 Support operation

In ESGF and other federated infrastructures, it is recommended by the FeUSE framework to implement all the agreements reached in the last phase. Furthermore, FeUSE recommends making a distinction between the diverse types of queries that arrive from the users or other sources in the e-Science infrastructure user support centre. The incoming user requests initiated from the users (or any other entity) are not classified into different types with respect to nature of origin and their further categories based on nature of problems, in the current user support process in ESGF. FeUSE framework in this step splits the incoming issues based on the principle of *divide and conquer*, similar to divide and conquer algorithms found in the computer science domain.

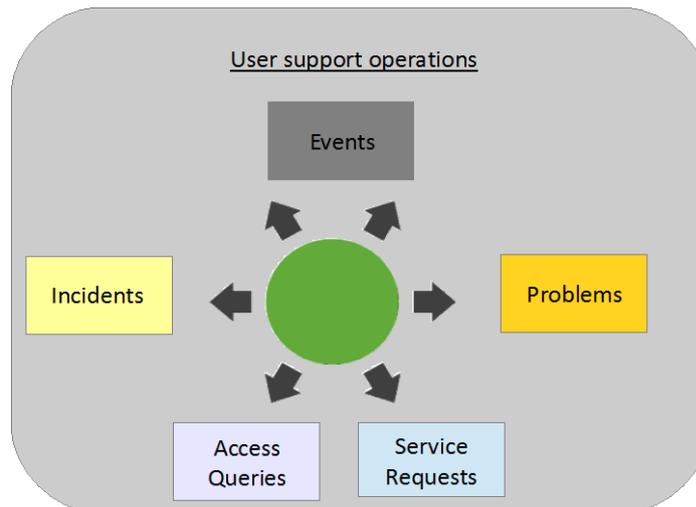


Figure 8.14 The fifth phase of the FeUSE framework that addresses the division of incoming issues into different types depending on the nature of origin of issues.

According to FeUSE framework, a classification of the incoming issues can be done based on their nature (as also recommended in ITIL Service Desk) and the incoming user requests are recommended to be categorized into:

- *Events or alerts* that initiate from hardware or software instruments that are part of the federated e-Science infrastructure and can be captured by dashboard in ESGF,
- *incidents* that may be initiated from end users or any other stakeholder of the e-Science infrastructure,
- *service requests* started from end users or any other stakeholder of the e-Science infrastructure e.g. a request of a new feature in the federated e-Science infrastructure,
- *problems* that may cause one or more incidents to occur and
- *access queries* e.g. a problem with password, user-name, webpage access, blog access etc.

8.7.1 Events and event management

An event is an *alert* or a *notification* created by any technical component, a monitoring tool or a service. In this connotation an event is the change of state that has significance for the management of an IT infrastructure service (Office of Government Commerce, 2011). In this study, it is an ESGF e-Science infrastructure service. An event with a significant impact on the service of the e-Science infrastructure needs an intervention of an e-Science infrastructure staff or an automated agent to take appropriate actions to restore the normal service of the e-Science infrastructure. Event often lead to incidents being logged.

Events occur all the time in different infrastructure e.g. traffic infrastructure and e-Science infrastructure is not an exception. It is imperative to catch events in the federated e-infrastructure to ensure proactive and reactive maintenance of service delivery to the end-users and other stakeholders. In ESGF, the events of the e-Science infrastructure are monitored by the monitoring tool known as “Dashboard”. Dashboard is a central management tool of the ESGF e-Science infrastructure that monitors “health” of the infrastructure based on intensity of event as logically programmed in the hardware or software components which are part of the federated peer-to-peer ESGF e-Science infrastructure. More details on dashboard are found in Section 0.

The types of events that may occur can be of the following type:

- a. *Events that indicate a normal operation* e.g. a user logging on to use a service.
- b. *Events that indicate abnormal operation* e.g. a user trying to access unauthorized datasets, or the gateway or a node is down.
- c. *Events that warn about an unusual situation* e.g. the capacity of disk storage in a server reach its highest acceptable level.

The process of event management is recommended in the FeUSE framework to monitor all events occurring in the e-Science infrastructure in a similar manner as suggested by ITIL V3 2011 Service Operations handbook (Office of Government Commerce, 2011).

The main activities of this process are depicted in the Figure 8.15. The steps are as follows:

- **An event occurs:** Not all types of events may be significant. However, it is important to agree upon what types of events must be registered in the components of the federated peer-to-peer ESGF e-Science infrastructure.
 - **Event notification:** The components of e-Science infrastructures, also called *configuration item* (CI) in ITIL terminology should be designed in such a way that they generate useful information. This information can be then communicated to a central tool such as dashboard that must collect all the significant information coming from the components of the federated ESGF.
 - **Event detection:** An agent or management tool may generate an event report after detecting it and interprets it directly from the hardware or software component.
 - **Event logging:** An event and subsequent actions must be logged in the device (hardware or software) i.e. an event management tool.
 - **Initial event filtering:** An event can be communicated from the hardware or software to the central monitoring tool such as dashboard. Dashboard filters information generated from an event.
 - **Gauging the significance of an event:** The significant events recorded in the central monitoring tool dashboard, can be categorized into for instance: An *informational* event, a *warning* and an *exception* event.
-

- **Further event analysis:** The event is interpreted according to the user support policy that incorporates event management policy according to the business rules set by the management of e-Science infrastructure.
- **Action required? :** Based on the event analysis, further action can be decided whether it is required or not.
- **Responding to the event:** At this stage, the action can be taken by either an automatic agent or by a human agent based on the information provided by the central management tool.
- **Review the actions:** It is essential to review the actions taken in order to ensure that the event has been treated properly and record the occurrence and re-occurrence of the event.
- **Closing the event:** Events can be closed. However, they may remain open until specific actions have been taken.

The major purpose of event management in e-Science infrastructures is to survey all the events happening amongst the whole federation in order to ensure the regular performance of the e-Science infrastructure and take appropriate action in case of service failure. Event management is the first step to resolve anomalies in the service and make the e-Science infrastructure reliable. Consequently, it leads to improvement of the overall user service and leading to a robust user support process. Events may lead to incidents.

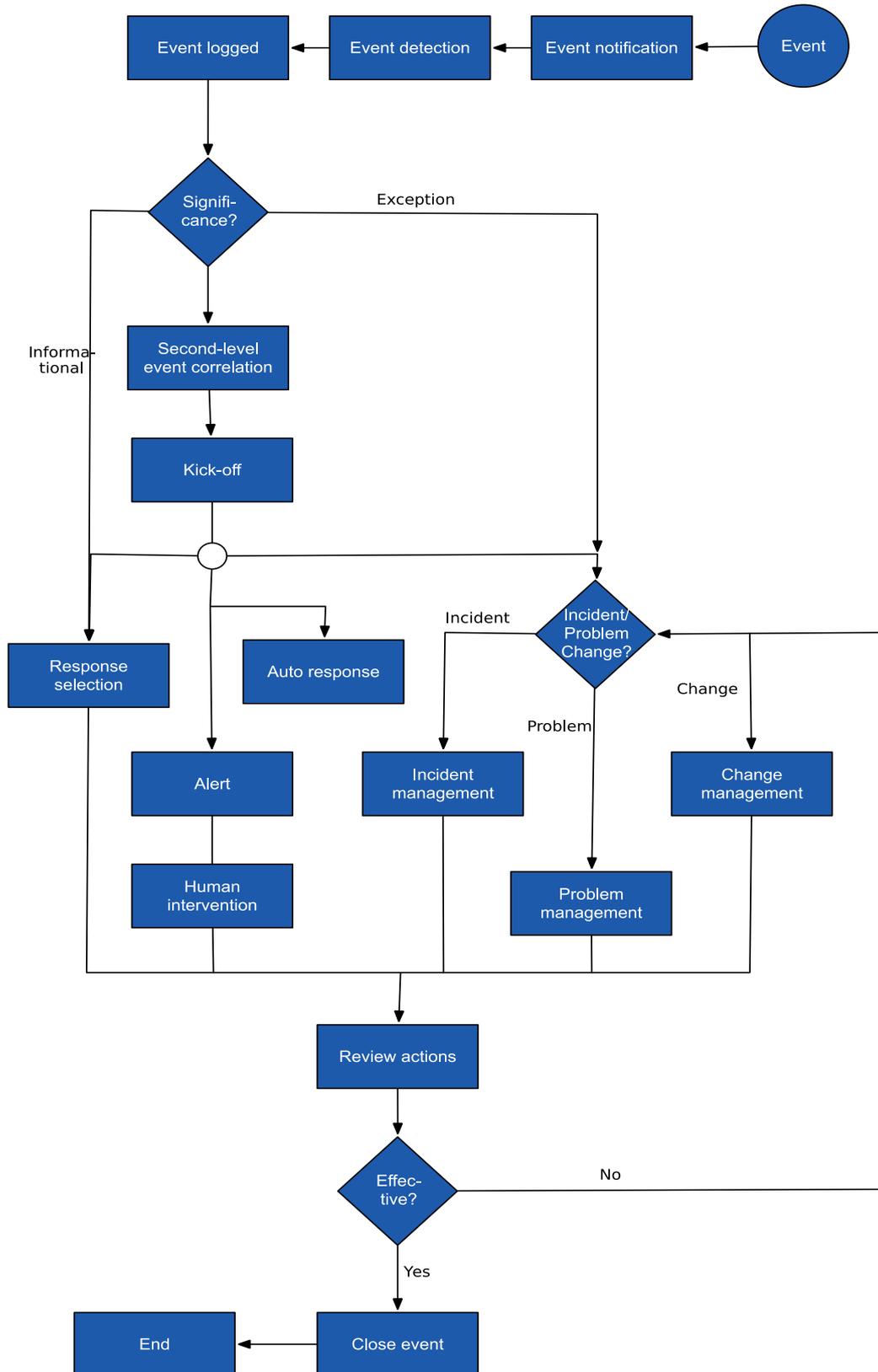


Figure 8.15 The workflow of events created by the federated components of the e-Science infrastructure, adapted from OGC (Office of Government Commerce, 2011).

8.7.2 Incidents and incident management

An *incident* is an unplanned interruption to the service delivery of an e-Science infrastructure or reduction in the quality of service that has an impact on the users (and customers). The impact on the users of an e-Science infrastructure could be hindrance in fetching data, problems with data analysis and computation trouble. Failure of a component, for instance; an outage of a data node that might have not affected the overall service, is also deemed as an incident. The aim of incident management is to restore the normal service, as soon as possible, so that the users are not affected. The normal service is the delivery of service to the users within the agreed service and operation levels.

The workflow of the activities of the incident management is provided as follows:

- **Identification:** The incident is reported or detected.
 - **Recording:** The incident is logged with the complete description into specific software such as RTS, Git Hub or other.
 - **Classification:** The incident is classified by type of incident, impact, status, urgency etc.
 - **Prioritization:** Every incident gets its priority depending on the two factors *urgency* and *impact*. If impact and urgency are high then priority can be high. The codes can be assigned as urgent, high, medium, low or alternative numeric codes can be assigned. For more details on prioritization see Section 8.6.1.
 - **First inspection:** The first inspection can be done by an automated agent or human agent to try to discover full symptoms of the incident. The incident data can be compared with other incident data of similar nature as well as known problems for a quicker resolution of an incident. This is performed by the first-level support staff (FLS) using different tools.
 - **Escalation:** The incidents that do not get resolved by the first level staffs can further be escalated for further support by the specialists. If the situation is very serious, then concerned managers might be informed to handle them.
 - **Diagnosis and investigation:** Further investigation is done to get to the possible solution. Specialists of a particular component or a domain might be helpful in this case.
 - **Recovery and resolution:** Once the solution has been found effective, the issue is resolved.
 - **Closing the incident:** At the end, if the user is satisfied with the solution, then the incident is closed. The incidents once closed can be reopened. However, the rules of reopening and the time-frame must be agreed upon by the policy makers of the e-Science infrastructure.
-

In the FeUSE framework it is recommended to document all the incidents, particularly the major incidents, at least. A major incident is an incident that has a higher degree of impact on the user community. It is noteworthy that with the passage of time the priority of an incident may change. A problem is the underlying cause of an incident and is seen as a separate entity that is handled separately.

8.7.3 Standard request and request management

The term user request is used and has been used in ESGF e-Science infrastructure as a general term for various requests that the users submit to the service desk (i.e. user support centre). A *service request* can be a request for the change of password or an additional installation of a software etc. A service request is a request for information, advice, a standard change or access to service. Since these requests have a lower risk as far as the negative affect on the e-Science infrastructure is concerned, it is recommended by the FeUSE framework to treat these requests in a separate process.

Many of these requests reappear in e-Science infrastructures as per observations of the author in two e-Science infrastructures, ESGF and C3Grid. That is why it is vital to have predefined *request models*. The requests can then be assigned different statuses such as in progress, rejected, cancelled, completed, waiting authorization as offered by different request tracking systems. Requests can also be prioritized in the same manner as incidents. Similarly, requests can be escalated from first level to the second and so forth in the same way as the incidents are escalated.

Some service requests may incur additional cost, adding to the running costs of servicing requests. In this case, the e-Science infrastructure may have an appropriate charging policy in accordance with the overall policy of the e-Science infrastructure. The request that might initiate a major change in the service of an e-Science infrastructure or may affect the e-Science infrastructure itself might be treated as a *request for change* (RFC) and not as a standard change. RFC can be handled by a separate process known as *change management*, see Section 8.8.1 and 8.8.1.1, page 283.

The activities of request management are as follows:

- **Receive request:** A formalized request is sent by the means specified by the e-Science infrastructure e.g. a web form, e-mail, phone call, social media or any other means.
 - **Request validation and logging:** The validation of the request is checked by automated means or by human agent. If it is a valid request, the request is recorded or logged into a particular system or documents with all details. The details include priority, sender information, description, category and other details.
-

- **Request categorization:** The request can be categorized according to the rules set by the e-Science infrastructure management or depending on the patterns found.
- **Request prioritization:** The request priority can be determined by the impact and urgency of a request as determined by the human agent of the e-Science infrastructure.
- **Request authorization:** All incoming requests need to be approved by an authorized person before they can be processed. At the same time, it must be determined whether the user is authorized to request requested services e.g. a commercial user might request permission to execute code on a particular HPC facility in the federation. The authorization levels and authorization staff to process each request might be set by the e-Science infrastructure management.
- **Request execution and review:** Once authorized the request can be executed by the authorized staff to give the desired function.
- **Request model execution:** Executing the service request of similar nature might enable the staffs of e-Science infrastructure to create a request model that may be followed to execute the requests of similar nature in a repeatable and a consistent way.
- **Request closure:** Once the staffs of ESGF have ensured that the request has been fulfilled then they can be closed. The rules of reopening user requests and the time-frame within it might be reopened and the time frame to accomplish the request must be agreed mutually by the management i.e. the policy makers of the infrastructure.

8.7.4 Problem and problem management

In FeUSE framework a *problem* is an unknown case that generates one or more incidents. *Problem management* records all such cases that might cause one or more incidents and controls the lifecycle of a problem. The aim is to eliminate or decrease the impact of incidents by identifying the root cause of these incidents. Consequently, the root causes or problems are administered separately in the FeUSE framework. In problem management, it is vital to know the two concepts: Work around and known error.

A *work around* is an attempt to eliminate or reduce an impact of an incident for which a complete solution may not be available.

A *known error* is a problem with a documented root cause. If a component or a system has a problem that cannot be resolved but, the *root cause(s)* is known, then they are called known errors. Creating a database known as known error data base (KEDB) is recommended by ITIL for faster diagnosis of problems (ITIL Handbook, 2011).

Like in the three processes mentioned earlier like event management, incident management, request fulfilment in the issue of problem management, there can be two aspects that can be taken into consideration at the same time. These are:

1. *Reactive problem management* is about responding to one or more incidents by analysing and resolving the problem that causes them.
2. *Proactive problem management* is about resolving the potential problem before it even occurs and can cause one or more incidents. *Proactive problem resolution* includes identification of potential weaknesses and analysis of trends.

The activities of performing reactive and proactive problem management are illustrated as follows (also visualized in Figure 8.16):

- **Problem detection:** The problems can be detected by studying and analysing the underlying causes of incidents either in a reactive or proactive manner.
 - **Problem logging:** All of the problems must be recorded in a central database or a document with appropriate and complete details.
 - **Problem categorization:** Recorded problems can then be split into categories with similar details as incidents that probably match the incident categories.
 - **Problem prioritization:** After categorizing problems can then be recorded with similar details as incidents in categories that probably match the incident categories.
 - **Problem diagnosis and investigation:** The investigation and diagnosis can be done using a variety of tools including “Ishikawa diagrams”, brainstorming, chronological analysis, hypothesis testing and so on.
 - **Workaround:** All workarounds must be noted. They are probably helpful to overcome difficulties. Until or unless the ultimate solution of a problem is found the problem record remains open.
 - **Known error record:** The root causes, details of a problem and the workarounds are needed to be put in a document or a data base known as *known error record*.
 - **Problem resolution:** If a solution of a problem has been found then the problem is resolved more likely by doing some changes in the e-Science infrastructure.
 - **Problem closure:** If an appropriate solution is found and the user is satisfied then the problem can be closed.
 - **Major Problem review:** The lessons learnt should also be documented and a review can be conducted to analyse problem and its solution.
-

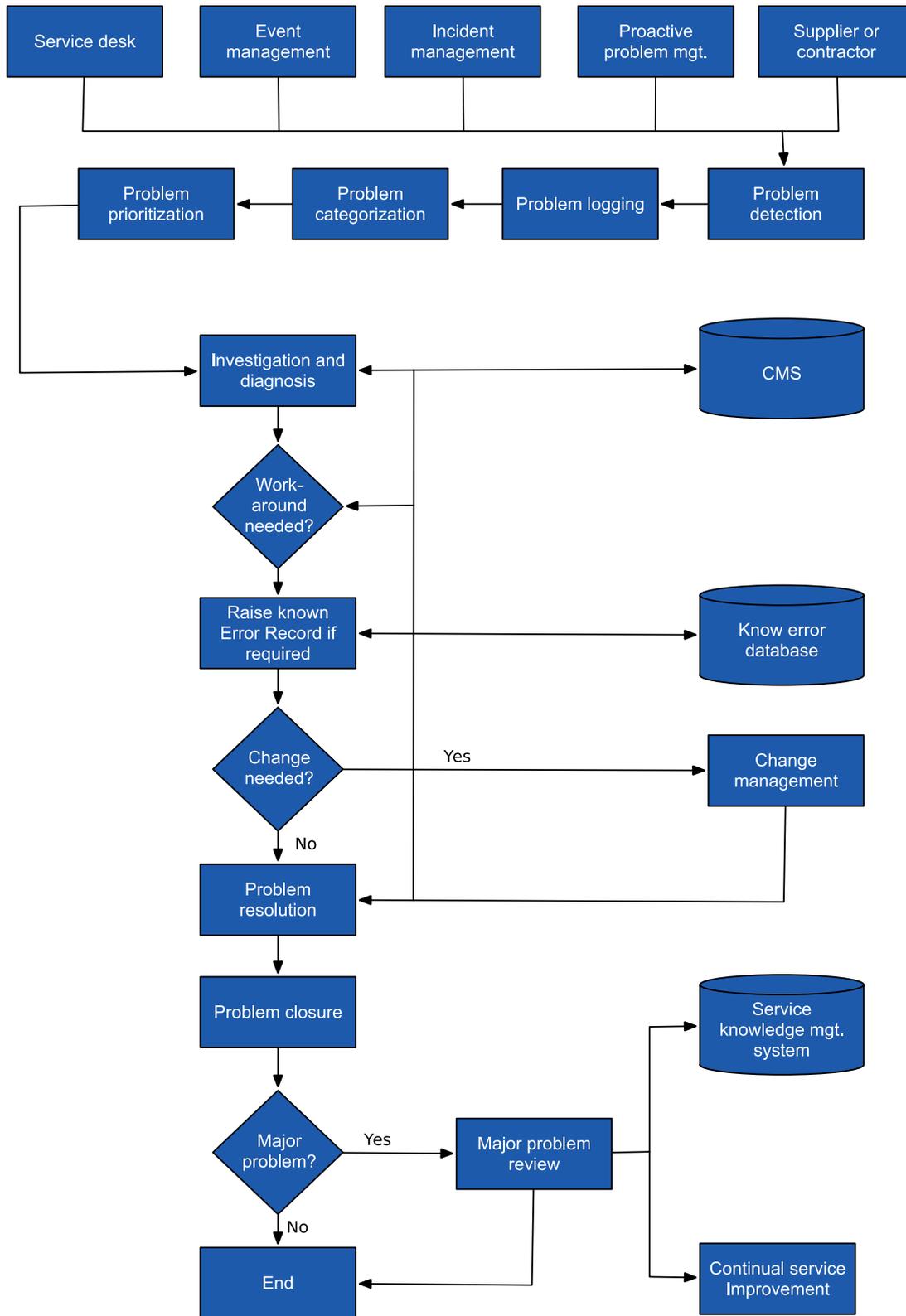


Figure 8.16 The workflow of the problem management in FeUSE framework (Office of Government Commerce, 2011).

8.7.5 Access management

Access is about allowing users to get the services, data and other assets of the e-Science infrastructure. ESGF has its own identity management policy that is implemented via identity provider (IdP). *Access management* ensures that the authorized personnel get access to assets offered by the e-Science infrastructure or modify them. Access management implements the security policy set by the ESGF e-Science infrastructure committees at a higher level. It is also referred to as identity management or rights management. The aim of the access management is to oversee that all resources of the e-Science infrastructure must not be improperly used.

Service or service groups: There are separate groups of users and internal staffs that must be authorized to access the resources according to rights set by the information security policy of the ESGF e-Science infrastructure committees. There are different tools that manage user access. They are referred to as *directory services* in ITIL terminology.

The description and workflow of activities of access management are given as follows (also represented in Figure 8.17):

- **Request an access:** The users of the e-Science infrastructure and the internal staffs might request certain privileges or access rights to access certain resources or services. This can be done via webpage, e-mail or other means, even RFC via request fulfilment process.
 - **Verify:** The request to access e-Science infrastructure is then checked based on whether the person's access request is authorized to access the services or resources and secondly whether are they really the person, who they claim to be.
 - **Provide rights:** Once the verification is done then the rights are provided by the authority as set by the policy set by the e-Science infrastructure for a time frame, the person is authorized to access the resources and the services.
 - **Check and monitor identity status:** Once the rights are provided to a person, the identity status is checked and monitored as the person's role may change such as being promoted, dismissed, retirement and expiring of access due to contract.
 - **Log and track access:** Access monitoring and control must be added to all activities, log must be maintained.
 - **Remove or restrict rights:** Access can be revoked at any time. The decision to do that must be taken by appropriate and authorized human agents in the e-Science infrastructure.
-

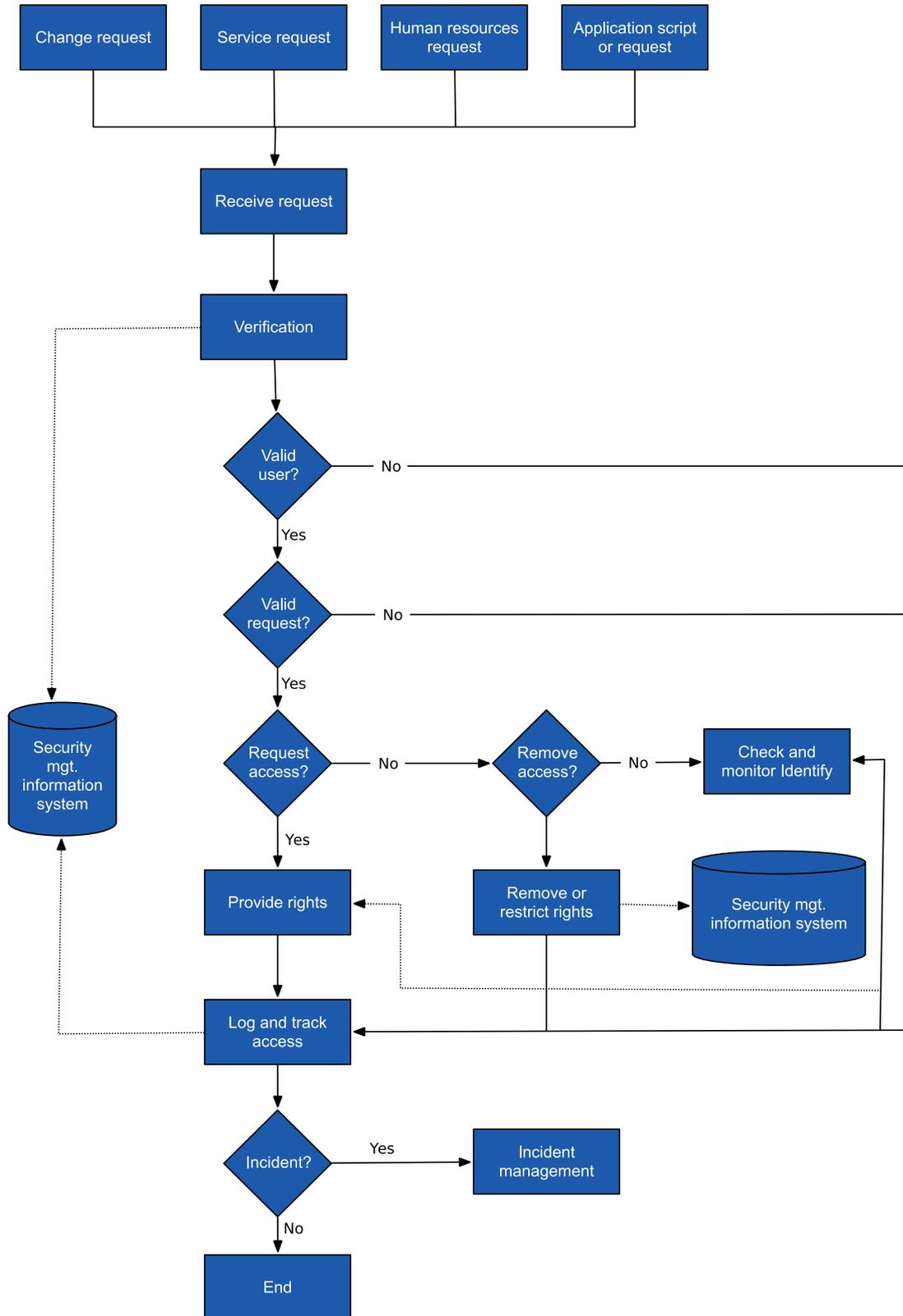


Figure 8.17 The access management workflow in the FeUSE framework adapted from OGC (Office of Government Commerce, 2011).

8.8 Monitoring

In this phase or aspect of the FeUSE framework, the changes are monitored constantly in the user support process of an e-Science infrastructure and the opportunities for improvements in the process are identified. This phase can also be identified as monitor, change and improve.

8.8.1 Change management

Federated e-Science infrastructure is ever-changing in nature as new data domains and data projects become part of the e-Science infrastructure and therefore it is significant to monitor the changes. *Change* is an addition, modification or removal of anything that could have an effect on the services provided by the e-Science infrastructure. The change can apply to architecture, processes, tools, metrics, components (or configuration items - CI), services as well as documentation. The changes can be at the strategic level or at an operational level. Changes can be major, significant and minor depending upon level of a cost or risk involved and on the scope of the change.

The stakeholders, especially users, may recommend changes any time in a year. User support process owner (USPO) and other support staff along with the executive committee (EC) and the technical committee (TC) form *change advisory board* (CAB) and take care of all changes of higher impact and magnitude. Each member institute of the ESGF federation from the technical committee as well as executive committee have a right to vote for a change or against a change. After every four months a CAB provides a decision about the changes i.e. in a year e.g. April, August and December. The changes can be approved, disapproved or deferred depending on the nature of changes and the resources needed. However, petty changes are taken care in the service requests, a sub-aspect of operations step of the FeUSE framework (for details see Section 8.7.3, page 277).

The process of controlling all changes in the form of a lifecycle is known as *change management*. The aim of change management is to enable beneficial changes to be part of the e-Science infrastructure with minimum disruption to the services that the infrastructure delivers. Any individual (from stakeholders) or a partner institution organisation may submit a *request for change* (RFC).

8.8.1.1 Request for Change (RFC)

Request for change is the formal change to the services or components of the e-Science infrastructure. The complex changes may be preceded by change proposals. There are three types of change requests: a *normal change*, a *standard change* and an *emergency change*. Normal change is a routine change done to the services and the federated e-Science infrastructure by adding, modifying or eliminating certain features to the architecture, processes, tools, metrics and documentation including changes done to

services, components and human agents involved in the federated e-Science infrastructure.

A standard change is a pre-approved change after the discussion and consultation with the technical, scientific and executive committees and other stakeholders. A standard change is a common routine change. It is a relatively low risk. All changes must be documented and recorded in a document or a database. The emergency change must be introduced and implemented as soon as possible. An example of the emergency change to be implemented is to repair the failure or resume the service as soon as possible. It can have a large negative impact on the e-Science infrastructure or its service itself.

The change advisory board (CAB) must take care of all the changes. It must prioritize and schedule changes. A post implementation review is significant to review the impact of changes. It is also important for the advisory board to analyse that in a year how many major changes were made to the e-Science infrastructure and from whom the change request was received, who contributed to implement it. The initiator of the change can be a partner or a user or any other entity. How effective the change was? How much time and resources (including cost) was taken? What advancement did it provide to the e-Science infrastructure? It is important to maintain the statistics of major changes as well? Once they have been documented, one can study the patterns of change in the e-Science infrastructure and their causes, retrospectively. These changes may lead to a particular innovation in an e-Science infrastructure. No change should be done without a backup (fall back) plan. Once the change has been implemented it must be reviewed to check whether the change was implemented successfully or if there are other changes required.

The workflow of activities of RFC is depicted in Figure 8.18. The description of these activities is highlighted as follows:

- **Create and record RFC:** Any individual (from stakeholders) or a partner institution organisation may submit an RFC. All RFC must be recorded in a document or a database. RFC may be about any radical change in the services of the e-Science infrastructure, including the user support service or any other change in the organisation, governance and architecture of ESGF.
 - **Review RFC:** Once the RFC has been submitted then the review board or the concerned committee may examine the RFC that whether it is logical, feasible, and necessary or if a similar RFC has already been submitted.
 - **Assess and evaluate the change:** A proposed change request is evaluated and assessed to see its impact on the e-Science infrastructure, potential risks, benefits, costs and the justification. Based on all these assessment and evaluations it is then decided by the CAB whether to take forward the change or to stop it.
-

- Authorize the change build and test:** The change is then authorized by an individual or a group of people before the change undergoes in the e-Science infrastructure.

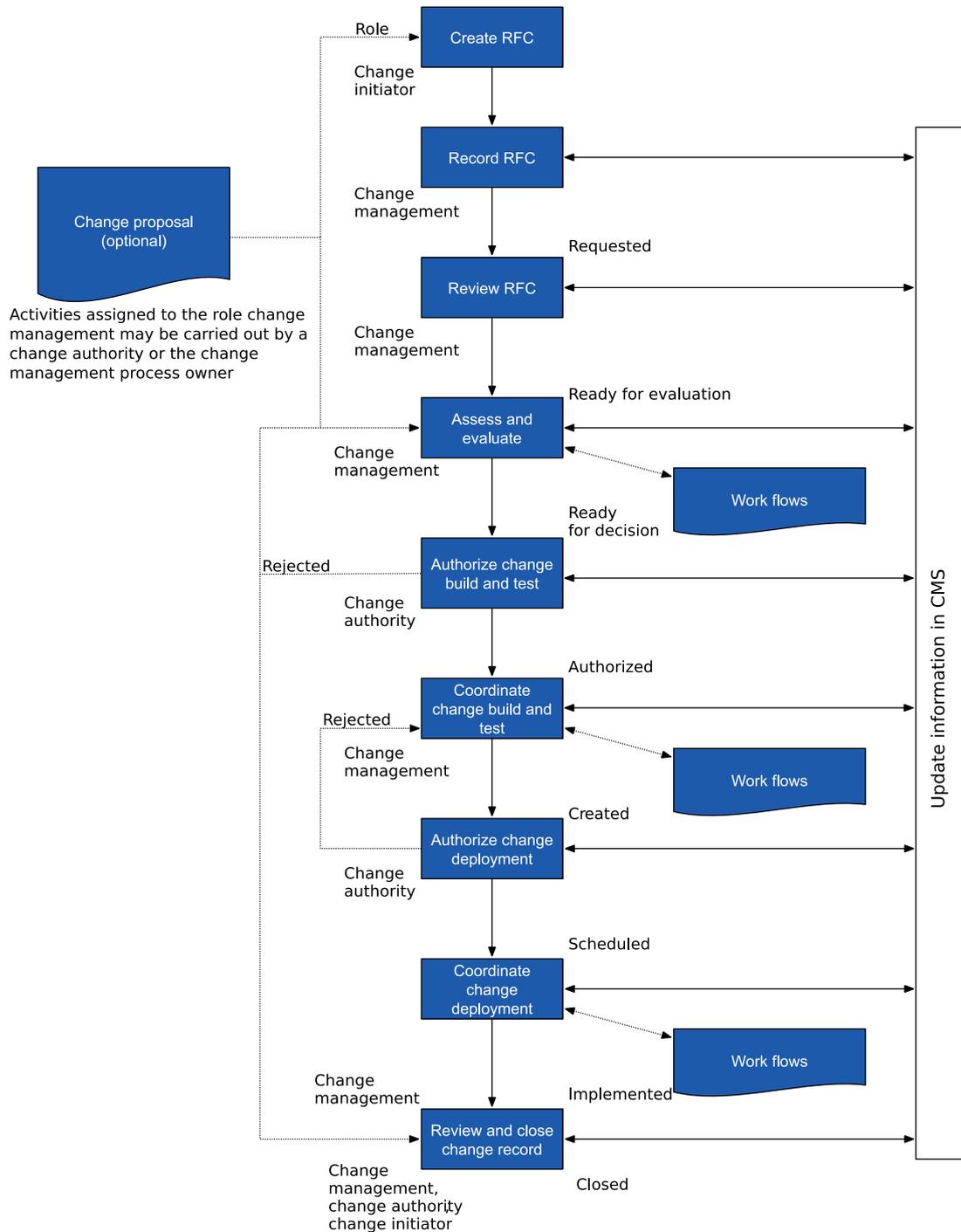


Figure 8.18 The workflow of activities in RFC adapted from OGC (Office of Government Commerce, 2011).

- **Coordinate change build and test:** Authorized changes are passed to the relevant groups to be implemented, tested and released.
- **Authorize change deployment:** The change authority requires a formal proof that the change has been fully implemented and properly developed, built, released in case of a component. This step may require formal report, a demonstration followed by a formal approval.
- **Coordinate change deployment:** All steps must be followed to deploy the intended change in all federated partners of the e-Science infrastructure. Remediation procedures should be prepared and documented in advance.
- **Review and close change record:** Deployed changes are reviewed after some time and post implementation review is done. If change is successful then the record after putting all the details can be closed.
- **Evaluate change:** The continuous changes done in the e-Science infrastructure user support process are needed to be evaluated. Change evaluation is all about determining the performance of a service change in the context of its likely impacts on the whole e-Science infrastructure, especially the services provided to the users. The analysis of intended and unintended effects of the change is presented in a change report.

8.8.1.2 Change evaluation

Once the change is approved and implemented in the user support process of the ESGF e-Science infrastructure as well as the e-Science infrastructure itself then a change report in the form of a summary of the implemented changes is made public for all the stakeholders. The change report also explains the rationale behind the implemented change i.e. why it was necessary to do the change and feasibility of the change. The report also includes the changes made to the UI of the e-Science infrastructure including the rationale behind them. The change report must also document the changes suggested but disapproved or deferred for future due to some reasons.

After the changes are implemented in the user support process of e-Science infrastructure, they are needed to be evaluated. *Change evaluation* is all about determining the performance of a service change in the context of its likely impacts on the whole e-Science infrastructure especially the services provided to the users. The performance of a change is measured against its predicted performance. Change evaluation helps in continually improving the service and changes done in it. Change evaluation report is the outcome of this phase.

8.8.1.3 Risk management

Risk management is part of each and every step in FeUSE framework. Risk management is “*a process of identifying, controlling, and eliminating or minimizing uncertain events that may affect system resources. It includes risk analysis, cost benefit*”

analysis, selection, implementation and test, security evaluation of safeguards, and overall security review” (OGC, 2011). For instance in ESGF, if the funding finishes for some service provided by the ESGF; it is essential to have backups plans and risk mitigation strategies. Similarly, in case a component in an infrastructure fails then there must be a risk plan that tells exactly what shall be done to continue a service or how to repair and replace the component.

Risks associated to any change happening in the e-Science infrastructure are needed to be evaluated. For example, in ESGF, such a case happened in the past: A forum for providing user-support, *Askbot* (the platform for user queries in the form of a Q&A blog) disappeared in November, 2013. As a result, all the queries asked by the users were lost and these queries were not recovered because there was no backup storage where all the queries were saved. Despite the efforts to recover data about the queries, it was unsuccessful. Consequently, no risk management was done in ESGF. Currently, in ESGF there is no explicit central risk management process present.

8.8.2 Knowledge management

Knowledge improves the quality of user support process. If the knowledge is created it enhances the decision-making of the users, the people who support users and other stakeholders in an e-Science infrastructure. Knowledge management is a process that transforms data into knowledge and eventually wisdom (Potgieter et al., 2005). The steps are data-information-knowledge and wisdom (DIKW). Quantitative data about all sorts of user queries are transformed into qualitative information. The data gathered in the user support process and other associated processes becomes information. Information with experience in a particular context, interpretation and reflection becomes knowledge.

Knowledge helps in right decision making which comes down to wisdom.

A central database must be made accessible to authorized personnel to access relevant data about the user requests, incidents, problems, request for changes, changes and other aspects such as log information of components (hardware and software) and other documents. This type of a repository is known as user support service knowledge management system (USSKMS). USSKMS can contain a number of databases such as service requests, incidents, problems, known errors, request for changes, changes made, releases, demand data, service strategy, financial data, service models, service reports, improvement plans. Moreover, the information is important for the entire e-Science infrastructure and all the stakeholders. The knowledge is an asset for the e-Science infrastructure and facilitates sustainability of the infrastructure. In addition, it saves time for the users and employees to acquire wisdom via knowledge.

The USSKMS may contain information about experience and skills of staff, information about behaviour of the users, performance of the user support service and other services of the e-Science infrastructure. It also includes the requirements, expectations of the federated partners and other peripheral issues such as problems and conflicts. The stored knowledge can then be transferred via knowledge transfer techniques such as e-learning and tools that enable knowledge visualization and reports, publications, seminars and workshops. These sessions and events can be conducted presenting the knowledge to the stakeholders of the e-Science infrastructure. Thus, the knowledge base or a central knowledge management system can serve as a binding tool and an integration mechanism for different stakeholder in e-Science infrastructures like ESGF.

8.8.3 Continual service improvement

Continual service improvement is a permanent part of the organisational culture.

CSI analyses the quality, process compliance and performance of the process. Moreover, it provides value to the stakeholders especially the users and the staffs of the partner institutions.

A register must be kept for all improvement initiatives in the user support process and service in an e-Science infrastructure. The register is part of user support service knowledge management system (USSKMS).

Improvements can be performed in the internal aspects within the organisation like organisational structures, culture, new knowledge, technologies and some external aspects such as customer requirements, reward system development, GUI improvement, and technology improvement. A baseline should be available as a starting point for later comparison of a service (ITIL-CSI). This baseline must be accepted, recognized, documented throughout the organisation.

Three types of baselines in the forms of metrics are recommended to be provided by the FeUSE framework to gauge whether a set target is achieved, these metrics are as follows:

- *Technology metrics*: These metrics gauge performance and availability of components and applications.
- *Process metrics*: These metrics measure performance of components that constitute a process including OLAs.
- *Service metrics*: These metrics test user satisfaction and end service results.

All these metrics help improving the service(s) provided by the e-Science infrastructures, chiefly the user support service. A user is interested in the services that he or she can consume and does not care much about the technical details of the infrastructure. “A customer is not interested in details about the functioning of the

technical infrastructure through which the services are provided, but only in the service itself.” (Foundations of ITIL, Volume 3, 2011)

The *critical success factors* (CSF) for technology, process and service are needed to be agreed upon by the top management of an e-Science infrastructure such as ESGF. CSF defines essential elements to achieve the main mission of the e-Science infrastructure in the form of *key performance indicators* (KPIs). The KPIs can be qualitative in the form of user satisfaction surveys or quantitative in the form of cost of solving a problem, cost of employing a technology to support users. The author recommends conducting user and staffing satisfaction surveys at least once in a year e.g. by end or beginning of a year.

8.9 Institutionalization

If the changes are not institutionalised in the partner institutions of the e-Science infrastructure like ESGF, then there may be potential risks, such as demotivation of staffs because the staffs are prone to use old procedure or there may exist a general resistance to change. If there is a situation that involves using new technology or an existing one to be used by the staffs or the users of the e-Science infrastructure then it is better to motivate the audience at their part to adopt it. There may be unforeseen circumstances and expenses due to the technology licenses, standards, staffs salaries, rewards and other issues in the user support process.

Despite the usage of a USSKMS, an SKMS and other techniques; there may still be a lack of knowledge sharing. Lack of maturity and integration of tools and processes with the user support process may hinder to achieve the results and value expected by the stakeholders of the ESGF. In addition, there can be cultural aspects such as organisational culture of the partner institutions that may hamper smooth transition of the changes done to the user support system and process in ESGF. Therefore, the socio-cultural aspects, legal aspects, conflict aspects amongst the users or user support workers, relation-development aspects, acknowledgement of the service are considered in the institutionalize changes phase.

8.10 Implementation of FeUSE in ESGF

So far, ESGF has tried to implement some of the suggestions of the FeUSE framework in the past informally. However, the full formal implementation of the FeUSE framework is needed which has not yet been done. It is significant for the applicability of FeUSE on any e-Science infrastructure like ESGF to observe the implementation of it and acquire a full feedback on the FeUSE framework's implementation.

In a survey conducted recently at the 4th Face-to-Face (F2F) ESGF meeting at LLNL in December 2014, policy makers from different partner institutions in ESGF were asked whether they think that the application of the FeUSE framework will improve the user

support process and the user support service of the ESGF. It is significant to note that in this survey, 24 policy makers from different partner institutions of ESGF participated. Most of the policy makers were of the opinion that the application of FeUSE framework in ESGF is needed to enhance the user support service in ESGF, as shown in the following Figure 8.19:

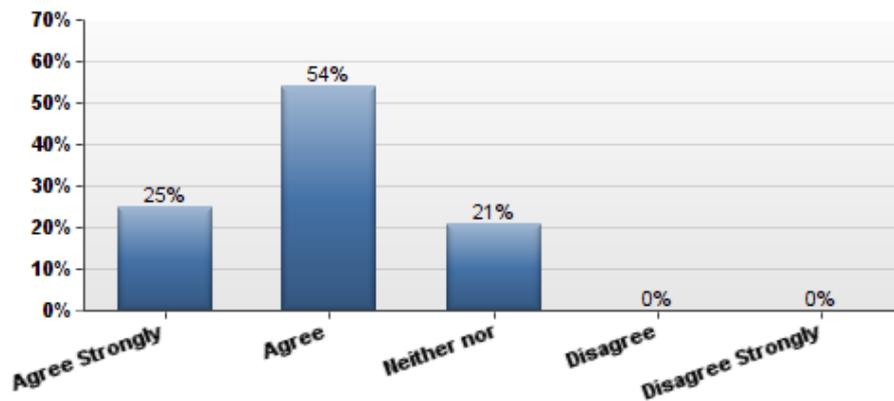


Figure 8.19 The graph shows the number of policy makers of ESGF and its associated projects that are in favour of introducing FeUSE framework to enhance user support service in ESGF.

It is evident from the figure that 25% of the respondents strongly agree to introduce FeUSE framework to enhance user support service in ESGF, 54% of the respondents agree and 21% of the respondents are neutral. It is vital to note that none of the respondents disagreed to introduce FeUSE framework to enhance user support service in ESGF and its associated projects. This shows a popularity of the view that FeUSE framework is beneficial to the ESGF e-Science infrastructure in enhancing its user support.

Similarly in the same survey, it was asked about the time needed to implement the FeUSE framework in ESGF, most of the policy makers are of the opinion that it is more likely to take upto two years to completely implement the FeUSE framework, see Figure 8.20. So, it is noteworthy to take this much of amount of time into consideration to implement the FeUSE framework, unless adequate resources are available to allow implementing the framework quickly.

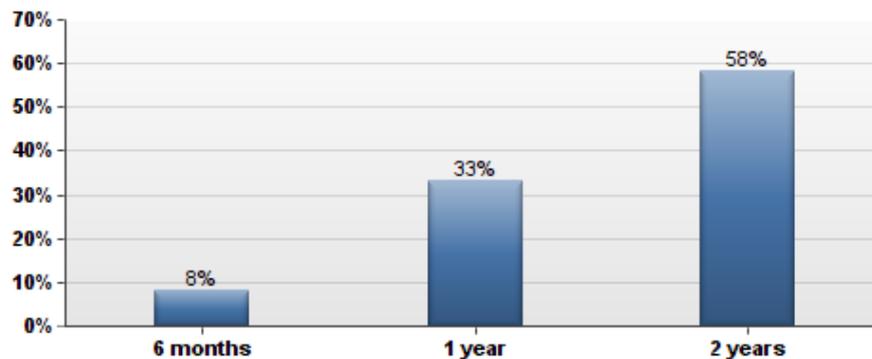


Figure 8.20 Time to implement the FeUSE framework in ESGF.

Until now, somewhat progress in ESGF has taken place as far as user support process is concerned that go in the direction of implementation of FeUSE. This includes a full user support position, which is responsible for answering user queries, funded by ENES (a partner of ESGF) project. Furthermore, a position at LLNL has been allocated where an employee is actively updating the user support material which is available online. Awards have been given for the first time to people who contributed to the research and development of ESGF including the user support. End-users were exclusively invited for the last Face-to-Face ESGF conference held at Livermore, California from 8th till 11th December, 2014 as according to suggestion of FeUSE framework. Moreover, a user support survey session was also organised by the author with the help of ESGF team during this conference where stakeholders were briefed about FeUSE and after that they took the survey about applicability of FeUSE in ESGF. For the upcoming conference which is planned in December, 2015, the author and end-users were also invited.

The organisation model and organisation chart suggested by the author were formally adapted by the ESGF administration to represent ESGF and UV-CDAT organisation structure. Support Working Team (SWT), which is equivalent to ESGF Support Services Committee (ESSC) already suggested by this research, has been formed recently in ESGF. The suggestion on categorization of incoming user problem along with the suggested categories based on the outcomes of this research were also put in use by ESGF. In a nutshell, one can say that with the help of this study, the significance of user support was made known to the stakeholders of ESGF. Moreover, this research has open doors to streamline the user support process and other associated processes in ESGF.

Besides, after the application of this framework, another similar framework on the same footings can be suggested to enhance other processes in e-Science infrastructure that are

similar to the user support process. These processes may include a process to support the developers and a process to support data managers of the e-Science infrastructure. In these processes, the focus is more on mitigating problems of developers and data managers of the e-Science infrastructure. These processes are very similar to the FeUSE framework in nature and application.

8.11 The future vision of this research

This study makes it pertinent to set a vision for the user support in e-Science infrastructures that shall lead to a boom in e-Science infrastructures, as far as communication and collaboration amongst the stakeholders is concerned. It is vital to recall the list of technologies currently being employed in the ESGF e-Science infrastructure and relate them with the future work that is needed to be accomplished. Just for a short recapitulation, these technologies are as follows:

Wikis: In ESGF, Wikis are being used especially for the advanced users, such as software developers, installers and users of the ESGF software stack. The information is detailed from topic to topic to guide the readers, especially who are interested in installing a software stack at a remote site (that may assume the role of data node, identity provider IdP, security or gateway and compute provider). In addition, it contains information about visualization softwares like Live Access Server (LAS), UV-CDAT and other third party software.

Webpages: The webpages cover information mainly related to frequently asked questions (FAQs), about other issues that include data formats, type of variables and list of available data sets. The list of servers that are up or may be down for maintenance. These webpages are hosted by ESGF, mainly related to technical FAQs and ESGF user-interface (UI) or scripts (via client software) related guidelines such as how to write Wget scripts to access and download chunks of data sets from particular data projects served by ESGF e-Science infrastructure e.g. CMIP5.

Forums (Askbot): Forums like Askbot were used to post user requests and get reply from the support people, whether employees of ESGF or users. Askbot was used to exchange views, to put comments, best solutions and to rate the solutions or comments. However, it was used only for several months as a pilot project and was not known to many users. The hard disk crashed at one of the collaborating institutes of ESGF and as a result the whole forum disappeared and the discussions in Askbot were lost.

E-Mail Lists: Mailing lists are the most popular communication media. As already described, users and developers are used to submit the user requests at a particular mailing list. At some partner institutes of ESGF, the e-mails from the users land in request tracking software (RTS).

The technologies currently used in the user support process of ESGF (as listed above) can further contribute in the following concepts that can serve as assets for the process: *Knowledge Management (KM)*, *Communication and Collaboration (C&C)*, *User-Relationship Management (URM)*, *innovation*, *development of the ESGF e-Science infrastructure and training* (learning), see Figure 8.21. At this stage, the relationship between the user support technologies and the assets that can emerge from these technologies has not been made explicit in the existing user support process. However, the use of FeUSE framework assists in achieving these aims in the long run by the use of appropriate technologies and by following the standards as well as best practices which are also proposed by the different steps of FeUSE.

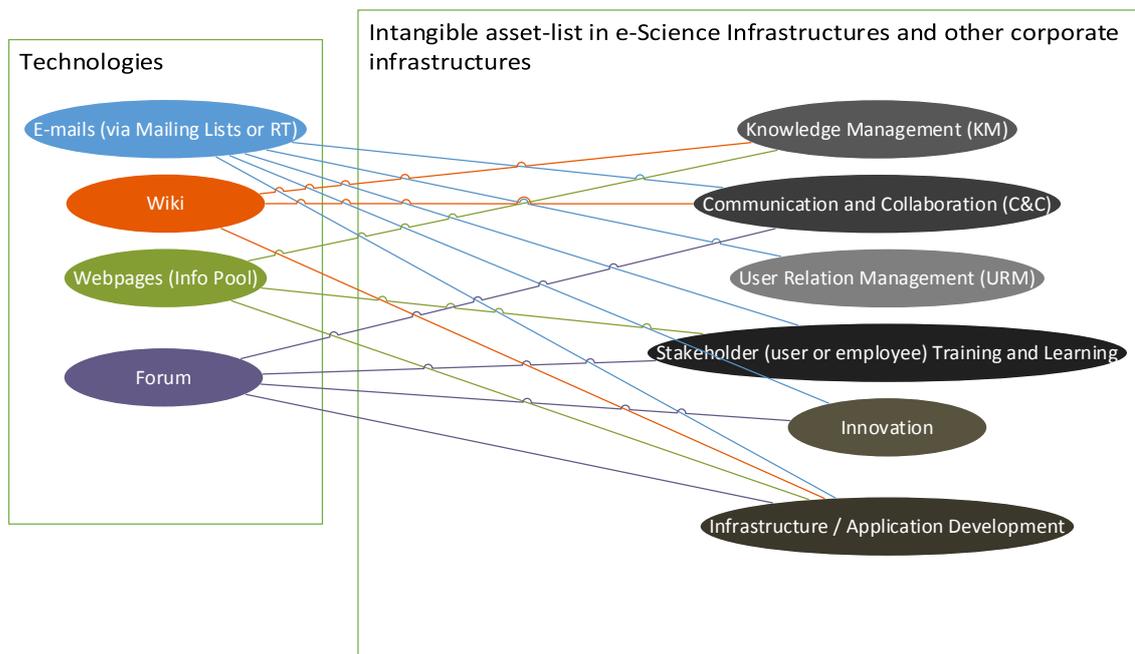


Figure 8.21 The technologies currently being deployed in the ESGF e-Science infrastructure, which provide intangible benefits to an organisation.

The technologies that can be applied in the user support process, along with the benefits that can be realized and driven out of the use of these technologies are listed in Figure 8.22.

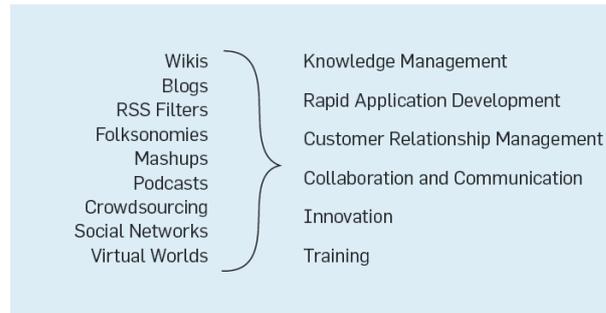


Figure 8.22 The web 2.0 tools that contribute to the various concepts that can bring value to the stakeholders of ESGF and other e-Science infrastructures taken from (Andriole, 2010).

All these concepts, presented in Figure 8.22, such as blogs, rich site summary (RSS) filters, folksonomies, mash-ups, crowd-sourcing and others can be applied in federated e-Science infrastructure support platform. RSS helps to change the web content regularly. It is helpful to the people who want to stay informed about the new content of the web. For instance, ESGF and other e-Science projects can offer RSS feeds so that the interested communities can be aware of the latest content offered, specifically about the e-infrastructure usage.

Mashups help to merge different features offered by open APIs and data sources to create enriched results (Andriole, 2010). Mashups might be applied to the proposed platform. A folksonomy which is a portmanteau of folk and taxonomy can be applied for categorising and tagging the content which is relevant in assisting the stakeholders of the e-Science infrastructure.

The use of these technologies and their combination can facilitate in the rapid application development, better collaboration and communication between stakeholders. Furthermore, the use of these technologies can be a driving force towards innovation in the infrastructure, training and e-learning of the stakeholders, customer (user) relationship management and knowledge management. The enormous possibilities that can be provided by the use of these concepts (as mentioned above) are provided in Figure 8.23. These possibilities can be utilized in the user support system for e-Science infrastructures to facilitate e-Research and the development of the e-Research environment.

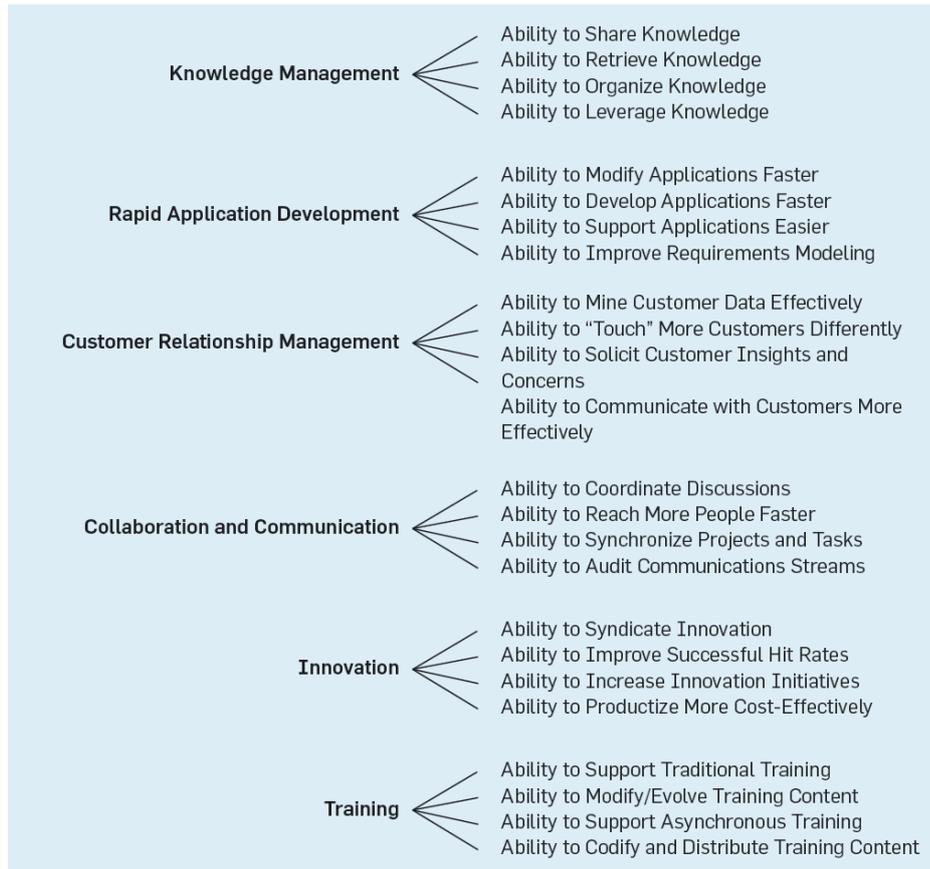


Figure 8.23 The abilities of the different concepts that the web 2.0 tools can support taken from (Andriole, 2010).

The author considers wiki's, blogs, request tracking systems (RTS), web pools and mailing-lists that help in collaboration in the e-Science activities as e-Research 1.0. However, an integration of all these tools, where the users can have access to a single platform that serves as a multi-purpose, multi-activity access point to the stakeholders of the e-Science infrastructures is known as *e-research 2.0 platform*. The combination of these technologies can be bundled together and this can lead to a combined platform that can be developed in the longer run by amalgamating these technologies. The sketch of this platform is provided in Figure 8.24.

When logged into this proposed platform, users have the ability to vote on submissions and comments to increase or decrease their visibility and submit links and comments just like reddit ("Reddit," 2014). Reddit is a bulletin board system. Users can also create their own subreddit based around any topic of their choice and interested users can add it to their frontpage by subscribing to it. Users earn points for submitting popular links and comments, which accumulate as point values on a user profile. According to the points collected by a user, a monetary or non-monetary award and recognition can be given to encourage users or staff members.

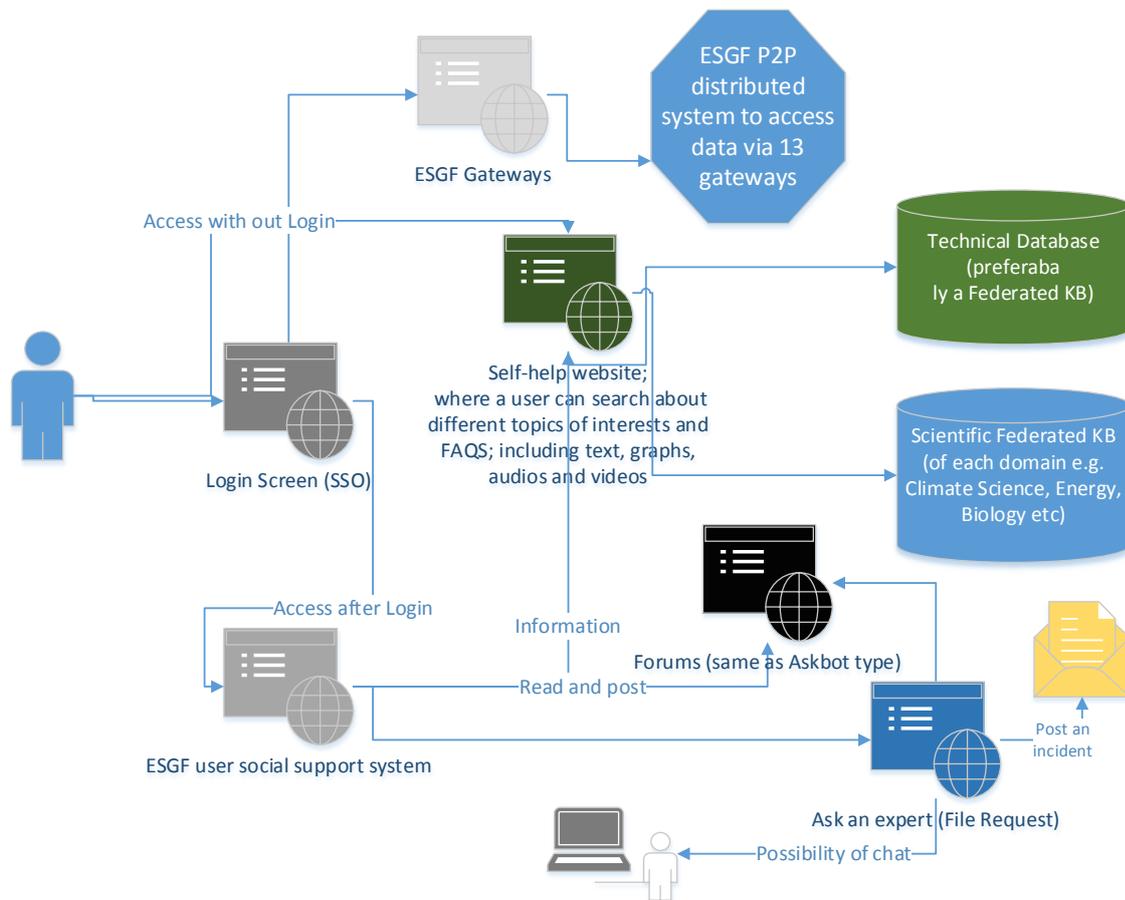


Figure 8.24 The sketch of e-Research 2.0 platform

Until now in ESGF and other e-Science infrastructures, the tools and web-techniques such as wikis, mailing-lists, forums and web-pages have been used to disseminate information and to process user requests. However, collective, collaborative social networking tools are not present to connect users as well as staffs and other stakeholders with each other. In e-Research 2.0 platform, the main commodity is data (like in e-Science infrastructure) and then computations on data are performed by some users as well. However, there is a need not only to get a user support but also to internetwork with similar people who are also looking for similar sets of data and performing related operations on the data. Therefore, the solution is not only to offer a solution that can help to support users but also offering social networking platform with the intention to promote e-Research activities supported by ESGF.

For instance, with-in the users of ESGF there are sub-user communities and sub-sub-user communities. For instance; the users who are interested in climate research and are looking for data sets from LUCID, CMIP5 data projects and then interested in particular climate models. This sub-category of users has not only the ability to support each other in terms of technical and scientific questions but also may collaborate with each other in scientific publications and in scientific projects. Thus, this community shares scientific aims and career paths. This is the reason why social e-Science network is significant as

far as the scientific collaboration in terms of data platform is concerned. This can undoubtedly increase the discovery rate via scientific collaboration through the proposed platform. The components of this platform, known as e-Research 2.0, social collaboration and support platform or a Science 2.0 platform are given in Table 8.4. An overview of the proposed functionality of the platform is given in Figure 8.24.

Table 8.4 The description of the components of e-Research 2.0 social collaboration and support platform.

Components of e-Research 2.0 social collaboration and support platform	Description and explanation
Single access point, networking, ITSM and collaboration tools available to all stakeholders	Users (researchers), staffs (developers, managers, executives, policy makers) and sponsors
Connectivity of users of various e-Science infrastructures world-wide	The proposed platform may connect e-Science infrastructures like ESGF, C3Grid and others.
Social networking capability	Like Facebook, Research Gate but the focus is on data communities and their usage as well as collaboration in discoveries, emanating from these data sets that the e-Science infrastructure offers as data projects
User support capability	The problems, incidents of the end-users can be handled via user support capability. It further includes following concepts: <ul style="list-style-type: none"> • Self-help capability for the users • Reward system • Training material • User request tracing capability
Developers support capability	The problems, incidents of the developers can be handled. Self-help capability for the developers as well as training material for e-learning
Data managing collaboration	The problems, incidents of the data managers and data curators can be handled. In addition, it may further include following capabilities: <ul style="list-style-type: none"> • Self-help capability for the data managers

	<ul style="list-style-type: none"> • Reward system • Training material • User request tracing capability
ITSM compatibility	The platform can offer tools and techniques to implement frameworks like FeUSE framework.

ESGF currently covers climate science domain but in the future it will cover biology as well as energy domain. The registered users of the ESGF infrastructures are 27,000. It is recommended to create a login by default to the platform for user support and networking. The users need other users’ support as well and can also use the networking capability.



Figure 8.25 Functionality aspects of the eResearch platform.

E-Research 2.0 is the future of e-Science. This concept of a social networking platform in the scientific communities of e-Science infrastructures can be instrumental in starting a new era of establishing collaboration and supporting the stakeholders of e-Science infrastructure. In addition, it reduces the wastage of resources and redundancy of efforts by the staffs of e-Science infrastructures. Satisfaction of users will increase as the user support process gets smooth and trouble-free. At the same time, the shared understanding of human support agents to mitigate root causes of user problems will enhance. This in turn can be a rich source for generating practical implications for system design, user modelling, and user instruction within e-Science infrastructure federations.

8.12 Chapter summary

In Chapter 8, the FeUSE framework is illustrated in detail along with examples of its application on ESGF, which is taken as a case study in this dissertation. FeUSE has seven steps namely: Vision, strategy, communication, agreement and assignment, operation, monitoring and institutionalization.

Vision is the first step in FeUSE and in this step the main goals of the user support process and service are set. Developing a combined federated strategy is the second step, in this step a strategy of the user support process of federated IT infrastructure is set based on analysis of: All types of stakeholders (primarily users), governance scheme, organisation of the infrastructure, resources available and managing stakeholders' relationships.

In the third step, the combined strategy is communicated and disseminated to all the stakeholders and feedback is collected. A consensus by the majority of the members of the federated partner institutions in the e-Science infrastructure is sought in the fourth step. An agreement is sought on issues like financial budget, business model, prioritization of incoming user requests, categorization of requests, service level as well as operational level agreements, key performance indicators, roles and other features. Similarly, once the agreements are made, the agreements are then implemented by assigning them to the concerned parties, which is also part of the fourth step of the FeUSE framework.

In the fifth step, once all agreements about the user support service operations are known and set, the events in the infrastructure and incoming user requests are classified into five different categories depending on the nature and origin of the issues. These include monitoring events in the federated infrastructure, splitting the incoming user requests into incidents, access requests, information service requests and finally the underlying problems are made known too. In the sixth step, the complete monitoring of the user support process and service is done. All types of changes and risks happening in the user support process are managed. The data and information gathered and stored in the user support process of e-Science infrastructure is managed following Data, Information, Knowledge, Wisdom (DIKW) model. This sub-phase is termed as

knowledge management. Moreover, opportunities for continuous service improvements are also highlighted in this step of the FeUSE framework. In the seventh step, institutionalization of the changes in the user support practices federation wide and conflict management is recommended in the FeUSE framework.

Apart from the description of the FeUSE framework, the changes done so far in the user support process and service of ESGF, motivated from this research are also highlighted. Moreover, part of the results of the evaluation of the FeUSE framework by the stakeholders of ESGF who attended the annual face-to-face ESGF conference in December 2014 is also presented. Finally, the future extensions of FeUSE and its integration with a proposed user support social network for the usage of variety of research communities is illustrated in this chapter. In the next chapter the final conclusions of this study are drawn.

9 Conclusion and future work

9.1 Introduction

This dissertation covers the relevant research work that has been done within the course of last years to uncover and understand the user support process in an e-Science infrastructure of a climate domain. The major research problem in this dissertation was about the user support practices that can be applied to improve the current state of user support process in e-Science infrastructures. The trends of an evolution of the user support process in the federated e-Science infrastructure, i.e. ESGF have indicated certain aspects of the process that are needed to be addressed in order to standardize the process federation wide. These aspects were determined after studying the current user support process in ESGF, empirically, which was part of the research objective of this research. These aspects include; finding out an optimum communication model for the user and support personnel communication, highlighting the characteristics of ESGF that can have an impact on the user support process and finally adapting best practices to enhance the user support service in e-Science infrastructures. This chapter brings the discussion to an end by concluding the major findings of the research.

In Section 9.2 of this chapter, the researcher summarizes the empirical findings of the study keeping the research objectives in view. In Section 9.3, the researcher discusses the proposed models i.e. the FeUSE framework that is the outcome of this study, chiefly based on the empirical analysis conducted by the author. In addition, the proposed FeUSE framework is partly grounded on the guidelines suggested in the relevant literature that illustrate the theoretical and practical aspects of models found in the government sector, academia and industry. Section 9.4 provides a brief overview of the limitations of this study as well as future works to enrich the foundations laid in this thesis.

9.2 Conclusion of the empirical findings

This research is focused on the problem of improving the user support process in e-Science infrastructures, which was identified over the years of observation of development and operation of e-Science infrastructures by the author. The scientific and practical problem faced by the ESGF consortium was the lack of formal user support process to assist users and stakeholders. The user support process within the ESGF e-Science infrastructure was analysed by using design science research, which is in some cases similar to natural science research, keeping with the guidelines of Hevner et al. and with March & Smith 2004. The design science research was chosen to study the existing informal user support process in ESGF in order to dig deeper into the process

and to establish evidence from multiple sources that the current view of the existing user support in ESGF is legitimate.

In order to achieve the eventual research goal, Peter Checkland's soft systems methodology (SSM) was used to model the prevailing user support process in ESGF, using rich pictures. Furthermore, the guidelines of Wilson were also considered to include various types of information about the user support process from different sources of evidence. The evidence was established via relevant literature, documents about ESGF, semi-structured interviews, field study and surveys. The salient empirical findings about the current user support process are as follows:

In the existing user support process, the employees of ESGF and users are not extremely satisfied. Moreover, the users tend not to provide feedback and recommendations to enhance the user support process in ESGF. In the user support process, there are more tendencies towards technical oriented staffs; however, the link to scientific staffs for answering scientific questions is almost missing. The tiers in the user support process are not defined. It is recommended to split user support into two tiers: FLS and SLS. The reason for this split is that currently people who handle the incoming user requests are mainly performing tasks other than user support. These tasks include mainly: Software development, documentation, research, administration and data curation.

Around 40% of all the user queries are of routine in nature and can be solved by non-specialists i.e. people having a general view of ESGF. Therefore, it is important to employ at least two to three FLS staff members covering different time zones. And only specialized queries that are not answerable by the FLS can be escalated to SLS. Employing FLS and training them adequately will allow the time dedicated by the SLS to support users to become shorter. Other mechanisms to save time and financial resources may include introducing a culture within users to help other users by using FAQs, Wikis and providing rewards to the users. Currently, e-support is not popular amongst the users of ESGF. User-assisting material such as FAQs and Wikis need improvement in terms of understanding, accessibility and up-to-date information. Cook books or recipes are needed to be introduced, so that the users are able to solve the problems themselves by looking at the relevant material available, whether on the web-pages or printed material published and issued by the ESGF consortium in the form of magazines or books.

The active phase of the usage of the ESGF e-Science infrastructure varies, at times hundreds of users of ESGF are active, whereas at another instance it may rise to thousands of users of ESGF to become active and pull the data from ESGF. The existing ESGF user support aims to be a single point of contact (SPOC) using a mailing-list but there are also other possible points of contact existent. The user support model is on-demand in nature and the user requests are processed in the first in first out (FIFO)

order. User requests are handled by a single person. Furthermore, user requests are generally understandable but there is a potential to make them even more understandable. For this, there is a need to introduce a user request pattern in the long run.

The categories of the incoming user requests are not formally suggested and the user requests are not categorized. After the collection of empirical results categorization of the incoming user requests were suggested. The chief categories included: Registration issues, access problems, data download, data format, post-processing and others. Up to 15% of the incoming user requests have remained unsolvable i.e. they have remained stuck in the user support centre and were ignored by the staffs. There has been a practice of multiple-delegation (MD) in the user support process within ESGF.

The incoming user requests are not recorded in a central location within the ESGF consortium. It is significant to have the user requests recorded at a central location at least for a particular time-frame agreed by the policy makers of the ESGF consortium, so that the user requests can be tracked by the users as well as the staffs. Furthermore, the incoming user requests can be sorted in different possible categories as preferred and decided by the user support staffs unanimously. The most important user requests that identify a particular problem should be documented. The relevant information of the user requests should be transferred to FAQs, Wikis and other relevant information portals for the users of the ESGF e-Science infrastructure in accessible and understandable format. In the existing user support process, there are no regular updates of the user support information websites including FAQs and Wikis as there are no OLAs or SLAs existing to accomplish it. In addition, web-page communication materials must be made easier to grasp by the users.

In the existing user support process, communication between users and the staffs is in English. There are different channels of communication between users and the staffs as well as within staffs that include walk-in, e-mail based forum and others. However, e-mail based textual communication is most common, up to 75% in the user support process of the ESGF e-Science infrastructure. The e-mail communication between users and the staffs as well as with-in staffs is further separated into user's mailing-list and RTS. However, there is no central RTS and each federating partner institute of ESGF has its own type of RTS. Most of the federating partner institutes of ESGF do not have RTS. Previously, neither the federating partner institutes of ESGF reached a consensus on an RTS technology that may be adapted by the whole federation, nor a unanimous decision on adaption of central RTS was reached.

Askbot, a user's forum was applied in the year 2013 for a short period from August till November 2013. *Askbot* user forum was a test version and a try to replace mailing-list completely because of the interactive nature offered by askbot. However, due to a hard-

disk drive failure at the partner institute of ESGF institute the data about the communication between users and the staffs were lost. This has not been recovered so far despite many trials to retrieve it back, which are still on-going in the ESGF federation.

There are no standardized procedures and time-frames set for replying to the incoming user queries as well as the escalated queries, referred by the staffs of ESGF to other staffs. This leads to longer time-frames of no response as well as treatment of the user queries by the user support staffs in ESGF. So in order to minimize the longer time-frames of no response, it is imperative to establish service level agreements (SLAs) and operational level agreements (OLAs) between the federating partner institutes of ESGF in the current user support process to speed up answering user queries for research purposes. The user support directory (USD) is currently not maintained in ESGF to know whom to redirect queries to, which cannot be answered by particular staff. Moreover, there is no automatic reply mechanism present in the existing user support process of ESGF. The personal response from the staffs of ESGF to the users of ESGF regarding their queries is helpful to the users, but efforts have to be made to improve it and make it more understandable.

In a nutshell, after examining the overall existing user support process from the empirical results, it becomes apparent that the current user support process is not explicitly defined and it needs an attention of the policymakers to re-engineer it and make it more transparent. Since the user support process is not explicitly identified, similarly there is lack of awareness about the user support service. The delineation of user support service and its constituents is needed to be developed. The user support service in ESGF is core service in mitigating the problems that the users of ESGF face. The ESGF and other infrastructures have been developed by the public funding agencies like European Commission, Department of Energy (DOE) of USA, NASA etc., i.e. they are funded by the public money. Thus, the user support process is enabling the research via e-Science infrastructure to take place. If the user support process was more effective then the research via e-Science infrastructures would thrive.

From the empirical study conducted in this research, it became evident that there are number of teething troubles and fallacies that are being faced by the ESGF consortium, as far as delivery of the user support service is concerned. These problems include; first and foremost there is a lack of an agreement on the term *users* also known as *end-users*. The meaning of *end-users* is not clear amongst the stakeholders including staffs, policy-makers, data-managers and others. Similarly, there is a different understanding of the term *user support* amongst stakeholders. For instance, some staffs use the term *user support* for the software installation activities performed by partner institutions such as new federating partners who want to set up data nodes in the ESGF federation. There is minimal central documentation recorded or a storage area for incoming user support

queries. The procedures of user support process where the operational staffs and even users can refer to are not clearly defined.

In ESGF, there is *no explicit allocation of budget for end user support*. The budget allocated for the user support is not visible to the stakeholders, since, at the moment concrete budget is not defined or calculated. At the same time, the policy makers fail to consider *optimal use of human resources*, which can be possible, if there is an adequate planning to allocate FLS and SLS to contribute towards the user support process. Similarly, there is neither responsibility expected from the staffs nor an amount in the form of remuneration allocated to the staffs for the user support service in ESGF. *Unclear role definitions of human resources*, lack of appropriate compensation and accountability of the staffs of federated partner institutions of ESGF (who are performing user support activities) leads to unreliability of the user support process in ESGF as well as ESGF-like e-Science infrastructures.

In essence, the *combined overall user support strategy* for the federated ESGF user support service is missing. Each partner institution has its own way of handling the user queries, updating the information on the webpages. This leads to *heterogeneous situation* regarding user support practice and culture at every participating partner organization in ESGF. Besides, there are *no defined methods of user support management* such as key performance indicators (KPIs), service level management (SLM); that include SLAs and OLAs, which are not set in ESGF.

Likewise, there is a *lack of integration* and connectivity of the user support process with other processes in the ESGF e-Science infrastructure. The outputs of the user support process can be inputs to other processes such as data management process, ESGF software development process, data curation process and others. Consequently, there is a need to *professionalize IT user support process*. The missing structures in the user support, especially the governance of user support in ESGF; its organisation and communication are needed to be present.

This study reveals that there is an obvious rebellion amongst the policy makers and other stakeholders of the federated partners of the e-Science infrastructures to adapt the heavy-weight IT service management frameworks. For instance; ITIL, MOF, COBIT, FedSM and other frameworks offer guidelines and best practices to create a service oriented atmosphere, to provide service management capabilities of an IT infrastructures in different domains and to fill the gaps of infrastructure governance, organisation and communication of services in IT infrastructures: However, the implementation and the acceptance of these advanced frameworks is questioned in e-Science infrastructures such as ESGF. This is mainly because these IT service management frameworks are heavy-weight.

A movement towards introduction of a very light-weight IT service management framework is needed to pave the way towards user's service orientation in e-Science infrastructure including ESGF. The movement towards infrastructure service management and towards users' orientation in the e-Science infrastructure has just begun with this research. The contemporary focus of this movement is supporting users and collecting their feedback to make the process of user support in e-Science infrastructures better, as well as to make the user support service worthy enough to mark the blossoming era of e-Research.

The next section illustrates the first version of FeUSE framework that came into being, after the rigorous effort performed to explore the facets of the user support process using the empirical study and by uncovering the relevant literature in this field of research.

9.3 Conclusion and summary of the FeUSE framework

In order to counter the problems that were the outcomes of the empirical analysis of the existing user support process, a Federated User Support Enhancement (FeUSE) framework was suggested by the author. The FeUSE framework is the contribution by the author towards structuring the user support process within the federated e-Science infrastructure consortium i.e. ESGF. Especially, this contribution is specifically directed towards service management within e-Science. The FeUSE framework is also based on guidelines mainly from ITIL and other frameworks including COBIT, FedSM and MOF.

The FeUSE framework comprises of seven main aspects that are needed to be taken into account before designing a process, especially a user support process in federated environments, such as; ESGF or other similar e-Science infrastructures. The beauty of FeUSE framework is that it identifies *what* aspects need attention. Moreover, FeUSE provides some general guidelines in the form of best practices that how these aspects can be reflected in the federated user support process. However, it does not state exactly *how* to address these aspects of the federated user support process in e-Science infrastructures and other federated IT infrastructures in a concrete manner. The *how-to* part is left to the decision of the policy makers in e-Science infrastructures.

The first aspect of the FeUSE framework is to *create a common vision* of the user support process and the user support service. The unanimous vision of the user support process should be agreed with all the partner institutions of the federation in an undisputed manner. The vision includes four major goals which are recommended in the FeUSE framework that include: User satisfaction, cost efficiency, reliability and sustainability of the user support service as well as the e-Science infrastructure and finally leveraging development as well as innovation of the e-Science infrastructure via the user support process. This facet helps in creating a common and combined policy by

the federated partners of the e-Science infrastructure to provide user support service. Additionally, it defines the boundary and an ambit of user support service as well as all the aspects that come underneath the user support service.

Once the goals of the user support service are set and agreed by the federated partners of the consortium, then a common strategy to deliver the user support service is developed. The second aspect in the FeUSE framework is termed as: *Developing a common user support strategy*. The FeUSE framework suggests all the steps necessary to develop a strategy for the user support service. This includes steps to gather relevant information about the stakeholders, their roles, their purposes and publishing a report called stakeholder report that can be issued by a consortium, such as ESGF, to know about the salient features of the stakeholders. The governance structure of the user support process in a federated environment is demarcated in this aspect of the FeUSE framework by defining the roles of people in the federated organisation like ESGF. The organisation chart of the user support team and the user support directory (USD) are the main outcomes of this sub-aspect.

The FeUSE framework advocates tracking an overview of all the resources that are allocated to the e-Science infrastructure, particularly to the user support service at a federation level. This is also part of the second aspect of the FeUSE framework. An estimate to calculate all the available resources including human resources is encouraged by the FeUSE framework. At first, this estimate can be a rough estimate. Future resource planning of the federated services is also an essential element proposed by the FeUSE framework. Taking a snapshot of all the resources needed for the user support service is the key aspect in strategy planning for the user support process. Managing the users and the staff relationship across the federation of e-Science infrastructure is also a part of the service strategy aspect of the FeUSE framework. The management of the relationship between users and staffs include; collecting user statistics and persuading users to take an active part in the user support process of e-Science infrastructures, such as motivating users to assist other users. A reward system based on the performance of users and staffs is also proposed, in this particular aspect of the FeUSE framework.

Once the strategy about the user support process has been made by the policy makers of the federation, the strategy must be communicated as well as circulated to all the stakeholders and the federating partners. *Communicating* the strategy to all the stakeholders and the federating partners constitutes the third aspect of the FeUSE framework. Communication is a lubricant in the federation like ESGF. The feedback about the user support service strategy from all the representative partner institutions of the federating organisation is needed to be collected after the communication of the strategy.

The *agreement and assignment* of the policy matters important to the user support process constitute the fourth aspect of the FeUSE framework. After the strategy has been communicated to all the partner institutions of the federation and the agreement has been made on the different characteristics about the user support service. Then, the strategy can be assigned to the federation i.e. the federation partners who are then responsible for the agreements made. The agreement is sought on the various facets of the user support service such as; financial policy, service level achievements, categorization of the problems, and the roles of the support staff. The staffs of the federated e-Science infrastructure need to implement the agreed strategy.

The implementation of the user support strategy includes carrying out the user support *operations* which is the fifth aspect of the FeUSE framework. The user support operations mainly include answering user queries and preparing user help material. The technique of divide and conquer is applied in the FeUSE framework borrowed from ITIL V3 2011. Following the divide and conquer technique, it is recommended to split the incoming user queries into events, incidents, problems, access queries and service requests. Events are machine generated notifications, e.g. notifications from federated components of the e-Science infrastructures such as ESGF. By following event notifications, one can perform precautionary measures to minimize the user requests.

FeUSE is dynamic in nature as it proposes to monitor the constant change occurring in the federated infrastructure to suggest the improvements in the user support process, which is a continuous phenomenon. This is done in the *monitor*, change and continuous improvement phase, which is the sixth aspect of FeUSE. FeUSE also suggests having interfaces with other processes such as data curation process, data delivery and download process through which services are offered by the e-Science infrastructure to users, in addition to the user support service. When changes, whether bigger or smaller, occur in the e-Science infrastructure consortium, they are needed to be deeply rooted in the federation. These changes then must be fully established in the federation as well as the risks, associated to these changes are needed to be assessed. This is done in the seventh aspect of the FeUSE framework entitled as: *Institutionalization* of changes.

The reasons that the FeUSE framework was proposed, is based on the bigger picture of the common problems (already discussed in the last section i.e. last paragraphs of Section 9.2) that were faced by the ESGF and other similar e-Science infrastructures. Additionally, the observation by the author and the empirical analysis suggest some more traits that have caused the FeUSE framework to be established. These include a commitment on a broad federal level towards structuring the user support process and providing a user support service strategy in ESGF.

The current lack of commitment triggered insufficient engagement of the stakeholders, especially the staffs and the users, to communicate and coordinate on a federal level to

work collectively on forming a process and delivering the service to the users. This also led to poorly defined and poorly communicated vision of the user support process in federated e-Science infrastructures. The inadequate consideration of recognition, rewards and allocation of resources to the staffs and users by the management added to the need to introduce a framework for user support. Furthermore, there was no focus on the benefits that can be provided by the user support process to other processes and services provided by the ESGF e-Science infrastructure.

The effect of this study and the FeUSE framework have identified the approach towards looking at e-Science infrastructures from the service point of view and not just from developing the anatomy of e-Science infrastructures and their components. The service point of view in e-Science infrastructures has not been highlighted in the past, due to the fact that the e-Science infrastructures like ESGF didn't reach its production level. Since recently e-Science infrastructures have attained the production level, it is essential now to look at them from the service delivery to the users' perspective. Thus, it is necessary to manage services that are provided to the stakeholders especially the users. The users need the support service to fuel their research activity that they are doing by interacting with the e-Science infrastructure.

With this situation the FeUSE framework, in essence, is a blueprint that suggests how maximum capability of a user support process can be achieved, keeping the resources available with the federated consortium of e-Science infrastructure such as ESGF. The intention of the FeUSE framework is to create a programme in the form of a user support service that adds value to the major activities performed by the federated e-Science infrastructure and providing help in using the data projects that are served.

9.4 Limitations of this study and future works

This research is diverse and in depth, nevertheless it suffers from the potential drawbacks of a single case study method. The information driven from the single case study may not be completely generalizable to other cases. However, single case-study method offers profound understanding to the characteristics to investigate a particular phenomenon, which in our case is the user support process in ESGF (Dyer & Wilkins, 1991; Langley & Royer, 2006). Consequently, the single case study method facilitates theory building. Additionally the potential findings that are outcome of the single case study can be easily replicated in other cases as well as in the findings from the cross-case studies.

Another limitation that can be considered is the selection of the climate science domain within the domains of e-Science infrastructures. One may argue that the outcomes from this domain might not be applicable to other domains. However, this proposition can be countered by the argument that the climate science domain is very similar to other domains such as biology, astronomy, medicine and others. Moreover, the selection of

climate science domain can be considered a strength of this research as none of such sort of research has been conducted so far in the climate domain of the e-Science infrastructures, as to the best of the knowledge of the author. Yet, the transfer of the empirical findings of this research and the transfer of the FeUSE framework to the other domains is needed to be done as a priority in order to let the theory and practice of the FeUSE framework to develop further.

ESGF has tried to implement some of the suggestions of the FeUSE framework in the past informally; however, the full implementation of the FeUSE framework has not yet been done. It is significant for the applicability of the FeUSE to observe the implementation of it and acquire a full feedback on the FeUSE framework's implementation. Moreover, after the application of this framework, another similar framework on the same footings can be suggested to enhance other processes in e-Science infrastructure that are similar to the user support process. These processes may include a process to support the developers and a process to support data managers of the e-Science infrastructure. In these processes, the focus is more on mitigating problems of developers and data managers of the e-Science infrastructure. These processes are very similar to the FeUSE framework in nature and application.

The conceptual model of the user support system in ESGF is considered to be comprising of the following four elements: The stability of the e-Science infrastructure, the behaviour, knowledge and training of the human resources, the capability of the communication media used in the e-Science infrastructure user support process, the technology as well as capability and degree of automation achieved via technology in the user support process. The dependency of user support system on the infrastructure in this model has not been tested yet and is considered to be a future work. Similarly, new communication technologies can be employed and tested in the user support process of ESGF to learn further lessons.

In ESGF, the input to the empirical results was taken from number of stakeholders, including users; however, even more users can be invited to participate in the study. Moreover, this study focussed more on the ITSM part and the processual nature behind the user support part of the e-Science infrastructure. However, the interaction, usability, user experience direction of the user support system and the interfaces of the ESGF gateways could not be deeply covered in this study.

Research on organisational culture (OC) of the federated partner institutions of the e-Science infrastructure can also be done in future. The significance of the research on OC is finding out, to what extent the federated partner institutions are ready to orient themselves towards implementing user-centred services, such as user support in a federated environment. This includes ITSM and changes in the overall governance as well as organisation of the e-Science infrastructures. Likewise, a deeper study can be

conducted to analyse and improve the cost model of the user support process which is suggested in the second aspect of the FeUSE framework.

10 Bibliography

- Aamodt, A., & Plaza, E. (1994). Case-based reasoning; foundational issues, methodological variations and system approaches. *AI Communications*, 7(1), 39–59.
- Aberg, J., & Shahmehri, N. (2001). An empirical study of human web assistants: Implications for user support in web information systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 404–411). Seattle, USA: ACM. doi:10.1.1.32.856
- Anderson, C. (2010). *What are the ten core business processes?. Bizmanualz*. Retrieved October 12, 2014, from <http://www.bizmanualz.com/blog/what-are-the-ten-core-business-processes.html>
- Andriole, S. J. (2010). Business impact of web 2.0 technologies. *Communications of the ACM*, 53(67-72).
- ANSI. (1998). ANSI/EIA-632 Processes for engineering a system. Electronic Industrial Alliance.
- Appleton, O., Schaaf, T., Belloum, A., Kocot, J., Heikkurinen, M., Metzker, M., ... Serrat, J. (2014). *A strategic roadmap for federated service management* (pp. 1–66). Retrieved from http://www.fedsm.eu/sites/default/files/FedSM-D4.2-Qualification_Scheme-v1.1.pdf
- Arora, A., & Bandara, W. (2006). IT Service Desk Process Improvement – A Narrative Style Case Study. In H. Lee & B. Tan (Eds.), *The Tenth Pacific Asia Conference on Information Systems (PACIS 2006)* (pp. 1274–1285). Kuala Lumpur, Malaysia. Retrieved from <http://www.pacis-net.org/file/2006/1178.pdf>
- AstroGrid, & Institute for Astronomy University of Edinburgh. (2012). *AstroGrid*. Retrieved June 21, 2013, from <http://www.astrogrid.org/>
- Babbie, E. (1990). *Survey research methods* (2nd ed.). Belmont, CA: Wadsworth Publishing.
- Beisse, F. (2010). *A guide to computer user support for help desk and support specialists* (4th ed., p. 28). Boston, MA: Course Technology, Cengage Learning.
- Bernholdt, D., Bharathi, S., Brown, D., Chanchio, K., Chen, M., Chervenak, A., ... Williams, D. (2005). The Earth System Grid: Supporting the next generation of climate modeling research. In *Proceedings of the IEEE* (Vol. 93, pp. 485–495). doi:10.1109/JPROC.2004.842745
- BiGgrid, Nikhef, & NBIC. (2010). *BiG grid: The Dutch e-Science grid*. Retrieved December 20, 2013, from <http://www.biggrid.nl/>
-

- BIRN. (2012). *Biomedical informatics research network*. Retrieved from <http://www.birncommunity.org>
- Boje, D. M. (2001). *Narrative methods for organization and communication research*. Thousand Oaks, CA: Sage Publishing.
- Bon, J. Van, Pondman, D., & Kemmerling, G. (2002). *IT service management: An introduction*. (J. van Bon, Ed.) (p. 217). Van Haren Publishing, ebrandedbooks.
- Brasileiro, F., Andrade, N., Lopes, R., & Sampaio, L. (2011). Democratizing Resource-Intensive e-Science Through Peer-to-Peer Grid Computing, 53–80. doi:10.1007/978-0-85729-439-5
- Broome, C. and Streitwieser, J. (2002). *What is e-Support? Service and support handbook* (p. 31 to 40). Colorado Springs, CO: Help Desk Institute.
- Buchanan, D. A. (2012). Case studies in organisational research. In G. Symon & C. Cassel (Eds.), *The Practice of qualitative organisational research: Core methods and current challenges*. (pp. 373–392). London: Sage Publishing.
- Buyya, R., Abramson, D., & Giddy, J. (2007). *Economy Grid*. Retrieved September 10, 2014, from <http://www.buyya.com/ecogrid>
- Buyya, R., & Murshed. (2002). GridSim: A toolkit for the modeling and simulation of distributed resource management and scheduling for grid computing concurrency and computation practice and experience. *Concurrency and Computation: Practice and Experience (CCPE)*, 14(13), 1175–1220. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.12.4262>
- Bv, G. B. (2010). *Enterthegrid: Directory on grid computing projects*. Retrieved June 10, 2013, from <http://www.enterthegrid.com>
- Campbell, A., & Tawadey, K. (1990). *Mission and business philosophy: Winning employee commitment*. Oxford, U.K.: Heinemann.
- Carney, T. F. (1972). *Content analysis : a technique for systematic inference from communications*. London: B.T. Batsford Ltd.
- Cempel, W. A. (2010). Logistic process reengineering: A case study. *Total Logistic Management*, 1(3), 5–20.
- Checkland, P. (1985). Achieving “desirable and feasible” change: An application of soft systems methodology. *The Journal of the Operational Research Society*, 36(9), 821–831.
- Checkland, P., & Winter, M. (2006). Process and content: Two ways of using SSM. *The Journal of the Operational Research Society*, 57(12), 1435–1441. doi:10.1057/palgrave.jors.2602118
-

- Chunpir, H. (2013). *C3Grid - INAD : Support centre with semi-automatic notification and coordination for experts work* (pp. 1–22). Hamburg.
- Chunpir, H. I., Ludwig, T., & Badewi, A. (2014a). A snap-shot of user support services in Earth System Grid Federation (ESGF): A use case of climate cyber-infrastructures. In L. Freund & W. Cellary (Eds.), *Proceedings of the 5th Applied Human Factors and Ergonomics (AHFE) Conference*. Krakow, Poland: AHFE Conference, CRC.
- Chunpir, H. I., Ludwig, T., & Badewi, A. A. (2014b). Using soft systems methodology (SSM) in understanding current user-support scenario in the climate science domain of cyber-infrastructures. In A. Marcus (Ed.), *Design, User Experience, and Usability. Lecture Notes in Computer Science, Volume 8519, Proceedings of HCI, Part III* (pp. 495–506). Heraklion, Crete, Greece: Springer International Publishing.
- Chunpir, H. I., & Rathmann, T. (2015). *E-mail analysis of a user- staff communication in e-Research* (pp. 1–12). Hamburg.
- Chunpir, H., Williams, D., Cinquini, L., & Kindermann, S. (2013). *Third annual ESGF and ultrascale visualization climate data analysis tools face-to-face meeting report. ESGF* (pp. 4–18). Livermore, CA. Retrieved from http://uvcdat.llnl.gov/media/pdf/ESGF_UV-CDAT_Meeting_Report_December2013.pdf
- Cinquini, L., Crichton, D., Mattmann, C., Bell, G. M., Drach, B., Williams, D., ... Schweitzer, R. (2012). The Earth System Grid Federation (ESGF): An open infrastructure for access to distributed geospatial data. In *8th IEEE International Conference on E-Science* (pp. 1–10). Chicago, Illinois, USA: IEEE. doi:10.1109/eScience.2012.6404471
- CMMI. (2009). *Improving processes for better services* (pp. 1–531). Hanscom Afb, MA. Retrieved from <http://www.sei.cmu.edu/reports/09tr001.pdf>
- Couper, M. (2000). Web surveys: A review of issues and approaches. *Public Opinion Quarterly*, 64(4), 464–94. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11171027>
- Czegel, B. (1998). *Help desk practitioner's handbook*. Hoboken, NJ, USA: John Wiley.
- Davenport, T. H. (1993). *Process innovation: Reengineering work through information technology*. Boston: Harvard Business School Press.
- David, M. (2005). *GGUS training for supporters: Duties and tasks* (pp. 1–31). Karlsruhe: EGEE.
- Denzin, N. K., & Lincoln, Y. S. (1994). *Handbook of qualitative research*. (N. K. Denzin & Y. S. Lincoln, Eds.) (1st ed., pp. 105–117). Thousand Oaks, California, U.S.A.: Sage Publishing.
-

- Dillmann, D. (2000). *Mail and Internet surveys: The tailored design method* (2nd ed.). New York, USA: John Wiley & Sons.
- Dyer, & Wilkins. (1991). Better stories not Better constructs, to generate better theory: A rejoinder on Eisenhardt. *Academy of Management Review*, 3(16), 613–619.
- Earth System Grid Federation. (2010). *ESGF research and development report. LLNL-ESGF* (pp. 1–135).
- EGEE. (2010). *Enabling grids for e-Science. EGEE and EGI*. Retrieved February 23, 2013, from <http://www.eu-egee.org>
- EGEE, & EMI. (2011). *gLite: Lightweight middleware for grid computing. CERN*. Retrieved April 20, 2014, from <http://glite.cern.ch/>,
- EGI. (2015). *EGI wiki: GGUS user guide and FAQ*. Retrieved February 10, 2015, from https://wiki.egi.eu/wiki/FAQ_GGUS-User_Guide
- EGI contributors. (2013). *EGI Wiki: GGUS helpdesk. EGI*. Retrieved March 16, 2014, from <https://wiki.egi.eu/wiki/GGUS>
- Eisenhardt, K. (1989). Building theories from case study research. *Academy of Management Review*, 14(4), 532–550. Retrieved from <http://www.wiggo.com/mgmt8510/Readings/Readings2/Eisenhardt1989AMR.pdf>
- ENES. (2013). *The ENES portal: European network for earth system modelling*. Retrieved December 02, 2014, from <https://verc.enes.org/>
- ESGF. (2010). *ESG: Earth system grid. ESGF*. Retrieved May 13, 2014, from <https://www.earthsystemgrid.org/about/overview.htm>
- ESGF. (2013). *ESGF. ESGF LLNL Web*. Retrieved September 15, 2014, from <http://esgf.org/mission.html>
- ESGF. (2015). *ESGF use statistics for CMIP5 data sets*.
- Evans, K., & Jones, W. T. (2005). Building an IT help desk: From zero to hero. In *Proceedings of the 33rd Annual ACM SIGUCCS Conference on User Services (SIGUCCS 05)* (pp. 68–74). Monterey, CA, USA: ACM Press. doi:10.1145/1099435.1099452
- FitSM. (2014). *FitSM Guide : Specifying services for portfolios and catalogues* (pp. 1–6).
- Flick, U. (2004). Triangulation in qualitative research. In U. Flick, E. von Kardof, & I. Steinke (Eds.), *A Companion to Qualitative Research* (pp. 178–183). London: Sage Publishing.
-

- Fulmer, J. (2009). What in the world is infrastructure? *PEI Infrastructure Investor*, 30–32. Retrieved from http://www.corridortrust.com/uploads/Infrastructure_Investor.pdf
- FusionGRID. (2007). *The national fusion collaboratory project*. Retrieved April 23, 2013, from <http://www.fusiongrid.org/projects>
- Gilb, T., & Brodie, L. (2005). *Competitive engineering: A handbook for systems engineering, requirement engineering and software engineering*. Burlington, MA, USA: Elsevier.
- Gomes, J., David, M., Martins, J., Bernardo, L., Marco, J., & Marco, R. (2004). First prototype of the CrossGrid testbed. In F. F. Rivera, M. Bubak, A. G. Tato, & R. Doallo (Eds.), *Proceedings of First European Across Grids Conference in LNCS Volume 2970* (pp. 67–77). Santiago de Compostela, Spain: Springer-Verlag.
- González, L. M., Giachetti, R. E., & Ramirez, G. (2005). Knowledge management-centric help desk: specification and performance evaluation. *Decision Support Systems*, 40(2), 389–405. doi:10.1016/j.dss.2004.04.013
- Greetz, C. (1973). *Interpretation of cultures* (First.). New York, New York, USA: Basic Books.
- Guilyardi, E., Balaji, V., Callaghan, S., DeLuca, C., Devine, G., Denvil, S., ... Valcke, S. (2011). The CMIP5 model and simulation documentation: A new standard for climate modelling meta-data. *CLIVAR Exchanges No. 56*, 16(2), 42–46. Retrieved from http://home.badc.rl.ac.uk/lawrence/static/2012/01/05/GuiEA11_CE.pdf
- Gupta, G., Mazumdar, C., & Rao, M. S. (2004). Digital forensic analysis of e-mails: A trusted e-mail protocol. *Journal of Digital Evidence*, 2(4), 1–11.
- Halverson, C. A., Erickson, T., & Ackerman, M. S. (2004). Behind the help desk: Evolution of a knowledge management system in a large organization. In *Proceedings of the 2004 ACM conference on Computer Supported Cooperative Work - CSCW '04* (pp. 304–313). New York, New York, USA: ACM Press. doi:10.1145/1031607.1031657
- Hammer, M., & Champy, J. (1993). *Reengineering the corporation: A manifesto for business revolution*. Harper Business.
- Heckman, R., & Guksey, A. (1998). Sources of customer satisfaction and dissatisfaction with information technology help desks. *Journal of Market Focused Management*, 3(1), 59–89. Retrieved from <http://link.springer.com/article/10.1023/A:1009794622860>
- Herrick, D. R., Metz, L., & Crane, A. (2012). Effective zero-cost help desk software. In *SIGUCCS '12 Proceedings of the 40th annual ACM SIGUCCS conference on User services* (pp. 157–160). Memphis, Tennessee, USA: ACM New York, NY, USA.
-

- Hess, J., Reuter, C., Pipek, V., & Wulf, V. (2012). Supporting end-user articulations in evolving business processes: A case study to explore intuitive notations and interaction designs. *International Journal of Cooperative Information Systems*, 21(4), 263–296. doi:10.1142/S0218843012500049
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105.
- Hey, T., & Trefethen, A. E. (2005). Cyberinfrastructure for e-Science. *Science (New York, N.Y.)*, 308(5723), 817–21. doi:10.1126/science.1110410
- Hiller, W., & Budich, R. (2013). ESM data archives in times of the grid. In W. Hiller, R. Budich, & R. Redler (Eds.), *Earth System Modelling* (Vol. 6, pp. 1–3). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-37244-5
- Holsti, Ole. (1969). *Content analysis for the social sciences and humanities*. London: Addison-Wesley.
- Iden, J., & Eikebrokk, T. R. (2013). Implementing IT service management: A systematic literature review. *International Journal of Information Management*, 33(3), 512–523. doi:http://dx.doi.org/10.1016/j.ijinfomgt.2013.01.004
- ISO/IEC. (2005a). *ISO/IEC 20000: IT service management, Part 1: Specification for service management*. Geneva, Switzerland.
- ISO/IEC. (2005b). *ISO/IEC 20000-1: IT service management, Part 2: Code of practice for service management*. Geneva, Switzerland.
- ISO/IEC. (2010a). *ISO/IEC FDIS 20000-1 IT service management, Part 1: Service management system requirements*. Geneva, Switzerland.
- ISO/IEC. (2010b). *ISO/IEC TR 20000-3: IT service management, Guidance on scope definition and applicability of ISO/IEC 20000-1*. Geneva, Switzerland.
- Jackson, S. J., Edwards, P. N., Bowker, G. C., & Knobel, C. P. (2007). Understanding infrastructure: History, heuristics and cyberinfrastructure policy. *First Monday*, 12(6). Retrieved from <http://firstmonday.org/ojs/index.php/fm/article/view/1904>
- Jäntti, M. (2009). Lessons learnt from the improvement of customer support processes: A case study on incident management. In *Proceedings of the 10th International Conference on Product-Focused Software Process Improvement, PROFES 2009* (pp. 317–331). Oulu, Finland: Springer. doi:10.1007/978-3-642-02152-7_24
- Jäntti, M. (2012a). Examining challenges in IT service desk system and processes: A case study. In *The Seventh International Conference on Systems, ICONS 2012* (pp. 105–108). Saint Gilles, Reunion Island: IARIA XPS Press.
- Jäntti, M. (2012b). Improving IT service desk and service management processes in Finnish tax administration: A case study on service engineering. In D. Oscar, J.
-

- Andreas, & J. Natalia (Eds.), *Proceedings of the 13th International Conference on Product-Focused Software Process Improvement, PROFES 2012* (pp. 218–232). Madrid, Spain: Springer Berlin Heidelberg.
- Jääntti, M., Cater-Steel, A., Shrestha, A., & Campus, N. (2012). Towards an Improved IT Service Desk System and Processes : A Case Study. *International Journal on Advances in Systems and Measurements*, 5(3), 203–215.
- Jääntti, M., & Kujala, T. (2011). Exploring a testing during maintenance process from IT service provider's perspective. In *5th International Conference on New Trends in Information Science and Service Science (NISS), 2011* (pp. 318–323). Macao: IEEE.
- Jirotko, M., Lee, C. P., & Olson, G. M. (2013). Supporting scientific collaboration: Methods, tools and concepts. *Computer Supported Cooperative Work (CSCW)*, 22(4), 667–715. doi:10.1007/s10606-012-9184-0
- Jirotko, M., Procter, R., Rodden, T., & Bowker, G. C. (2006). Special Issue: Collaboration in e-Research. *Computer Supported Cooperative Work (CSCW)*, 15(4), 251–255. doi:10.1007/s10606-006-9028-x
- Johansson, H. J., McHugh, P., Pendlebury, J., & Wheeler, W. (1994). *Business process reengineering: Breakpoint strategies for market dominance*. John Wiley & Sons.
- Juling, W. (2010). *NGI-DE: German national grid initiative*. Karlsruhe Institut für Technologie (KIT). Retrieved May 02, 2014, from www.ngi-de.eu
- Kajko-Mattsson, M. (2002). Problem management maturity within corrective maintenance. *Journal of Software Maintenance and Evolution: Research and Practice*, 14(3), 197–227. doi:10.1002/smr.252
- Kajko-Mattsson, M. (2004). Problems within front-end support. *Journal of Software Maintenance and Evolution: Research and Practice*, 16(45), 309–329. doi:10.1002/smr.298
- Kajko-Mattsson, M., Ahnlund, C., & Lundberg, E. (2004). CM3: Service level agreement. In *20th IEEE International Conference on Software Maintenance Proceedings, 2004*. (pp. 432–436). Chicago, Illinois, USA: IEEE. doi:10.1109/ICSM.2004.1357830
- Kapella, B. V. (2003). *A framework for incident and problem management* (pp. 1–20). Santa Clara, CA, USA.
- Kendall, H. (2002). "Prehistoric Help Desk!!." *Support World. Help Desk Institute*, (Oct-Nov), 6–8.
- Kettler, K., & Willems, J. R. (1999). A study of the outsourcing decision: Preliminary results. In J. Prasad (Ed.), *ACM SIGCPR Conference on Computer Personnel*
-

- Research (SIGCPR '99)* (pp. 182–189). New Orleans, LA.: ACM Press, New York, NY.
- Kirchmeyer, R. (2002). The consolidated help desk. In *Proceedings of the 30th Annual ACM SIGUCCS Conference on User services - SIGUCCS '02* (pp. 184–185). New York, New York, USA: ACM Press. doi:10.1145/588646.588688
- Knapp, M., & Woch, J. (2002). Towards a natural language driven automated help desk. In *Third International Conference on Computational Linguistics and Intelligent Text Processing, CICLing, 2002*.
- Krajewski, L. J., & Ritzman, L. P. (2005). *Operations management: Processes and value chains* (7th ed., p. 831). Pearson Education, Upper Saddle River.
- Krauter, K., Buyya, R., & Maheswaran, M. (2002). A taxonomy and survey of grid resource management systems for distributed computing. *Software: Practice and Experience*, 32(2), 135–164. doi:10.1002/spe.432
- Krippendorff, K. (1980). *Content analysis an introduction to its methodology*. Newbury Park, U.S.A.: Sage Publishing.
- Kunda, G. (1992). *Engineering Culture* (p. 351). Philadelphia, USA: Temple University Press.
- Kurzyniec, D., Sławinska, M., Sławinski, J., & Sunderam, V. (2007). Unibus : A contrarian approach to grid computing. *The Journal of Supercomputing*, 42(1), 125–144.
- Langley, A., & Royer, I. (2006). Perspectives on doing case study research in organisations. *M@n@gement Special Issue: Doing Case Study Research in Organisations*, 9(3), 81–94.
- Lazar, J., Feng, H. J., & Hochheiser, H. (2007). *Research Methods in Human-Computer Interaction*.
- Lazar, J., Feng, J., & Hochheiser, H. (2010). *Research Methods in Human-Computer Interaction*. Indianapolis, IN: Wiley.
- Leung, K. Y., & Lau, S. K. (2005). The development of a user self-help knowledge management system for help desk: Deployment of knowledge management approach and software agent technology. In D. Bunker & B. Campbel (Eds.), *16th Australasian Conference on Information Systems* (pp. 1–11). Sydney: Australasian Association for Information Systems.
- Leung, N., & Lau, S. (2006). Relieving the overloaded help desk: a knowledge management approach. *Communications of the International Information Management Association (CIIMA)*, 6(2), 87–98. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.98.1334&rep=rep1&type=pdf>
-

- Leung, N., & Lau, S. (2007). Information technology help desk survey: To identify the classification of simple and routine enquiries. *Journal of Computer Information Systems*, 47(1), 70–81. Retrieved from <http://ro.uow.edu.au/era/859/>
- Marcella, R., & Middleton, I. (1996). The role of the help desk in the strategic management of information systems. *OCLC Systems & Services*, 12(4), 4–19. doi:10.1108/10650759610153993
- Markopoulos, P. (2013). *Lecture notes on forming a qualitative research* (pp. 1–98). Eindhoven, The Netherlands.
- Marom, Y., & Zukerman, I. (2005). Analysis and synthesis of help desk responses. In R. Khosla (Ed.), *9th International Conference on Knowledge-Based Intelligent Information and Engineering Systems, KES 2005* (pp. 890–897). Melbourne, Australia: Springer Berlin Heidelberg.
- Mavrogeorgi, N., Gogouvitis, S., Voulodimos, A., Katsaros, G., Koutsoutos, S., Kiriazis, D., ... Kolodner, E. K. (2012). Content based SLAs in cloud computing environments. In *2012 IEEE Fifth International Conference on Cloud Computing* (pp. 977–978). IEEE. doi:10.1109/CLOUD.2012.106
- Mckoen, P. (2000). Creating a help center from scratch: A recipe for success. In *28th ACM Conference on Special Interest Group on University and College Computing Services (SIGUCCS) SIGUCCS '00* (pp. 186–189). Richmond, VA, USA: ACM New York, NY, USA.
- Microsoft. (2009). *Cross reference ITIL V3 and MOF 4* (pp. 1–33). Microsoft.
- Microsoft. (2015). *Microsoft operations framework (MOF)*. Retrieved January 13, 2015, from <http://technet.microsoft.com/en-us/solutionaccelerators/dd320379.aspx>
- Middleton, I. (1999). *The evolution of the IT help desk: From crisis centre to business manager in the public and private sectors. MSc Thesis*. The Robert Gordon University. Aberdeen, UK.
- Middleton, I. A., & Marcella, R. (1997). In need of support : the, (August), 120–127.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis* (2nd ed.). Thousand Oaks, California, USA: Sage Publishing.
- Morgan, D. L. (2008). *The SAGE encyclopedia of qualitative research methods*. (L. M. Given, Ed.) (1st ed., pp. 816–817). Thousand Oaks, CA, USA: Sage Publishing.
- Morse, J. M., Barrett, M., Mayan, M., Olson, K., Spiers, J., & Hon, D. (2002). Verification strategies for establishing reliability and validity in qualitative research. *International Journal of Qualitative Mehtods*, 1(2), 1–19.
- NEES, & NSF. (2012). *Network for earthquake engineering simulation*. Retrieved July 14, 2014, from <http://nees.org/>
-

- Negash, S., Ryan, T., & Igbaria, M. (2003). Quality and effectiveness in Web-based customer support systems. *Information & Management*, 40(8), 757–768. doi:10.1016/S0378-7206(02)00101-5
- neuGRID. (2011). *Grid-based e-infrastructure for data archiving in the medical sciences*. Retrieved October 10, 2014, from <http://www.neugrid.eu/pagine/home.php>
- Neuman, W. L. (2006). *Social research methods: Qualitative and quantitative approaches. Relevance of social research* (6th ed., p. 560). Pearson International.
- NGS. (2010). *The National Grid Service*. Retrieved June 17, 2014, from <http://www.ngs.ac.uk/>
- Niessinka, F., Clerca, V., Tjeldink, T., van Vliet, H. (2005). *The IT service capability maturity model: CMMI version 1.0*. Amsterdam.
- Nordugrid. (2012). *Advanced resource connector (ARC) middleware*. Retrieved June 12, 2014, from <http://www.knowarc.eu/>
- Office of Government Commerce. (2011). *ITIL V3 2011: Service Operation* (Version 3., p. 387). The Stationary Office.
- OGC. (2011). *ITIL Service Transition. Published by TSO (The Stationery Office) for the ...* Retrieved from <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:ITIL+Service+Transition#0>
- Open Science Grid*. (2013). Retrieved February 13, 2015, from <http://www.opensciencegrid.org/contact/>
- Oppenheim, A. (2000). *Questionnaire design, interviewing and attitude measurement*. New York, New York, USA: Continuum International Publishing Group-Academic and Professional.
- Ozok, A. (2007). *Survey design and implementation in HCI* (pp. 1151–1169). New York, New York, USA: Lawrence Erlbaum Associates.
- Pacholski, L., Cempel, W. A., & Pawlewski, P. (2009). *Reengineering: Reforming business and production processes in company*. Wydawnictwo Politechniki Poznańskiej, Poznan: Wydawnictwo Politechniki Poznańskiej.
- Patton, M. Q. (1988). *How to use qualitative methods in evaluation?*. Newbury Park, California, U.S.A.: Sage Publishing.
- PCMDI. (2013). *Overview of coupled model intercomparison project phase 5 - CMIP5. LLNL-WEB*. Retrieved May 24, 2014, from <http://cmip-pcmdi.llnl.gov/cmip5>
-

- Peters, S. L. (1993). Expanding the help desk to simplify customer access to CIT services. In *Proceedings of the 21st annual ACM SIGUCCS conference on User services - SIGUCCS '93* (pp. 330–334). New York, New York, USA: ACM Press. doi:10.1145/263814.263908
- Pettigrew, A. M. (1997). What is a processual analysis? *Scandinavian Journal of Management*, 13(4), 337–348. doi:10.1016/S0956-5221(97)00020-1
- Plakoyiannaki, M.-E. (2002). *Doctoral Dissertation on Customer Relationship Management (CRM): Effect of organisational culture - a longitudinal case study*. Strathclyde.
- Potgieter, B. C., Botha, J. H., & Lew, C. (2005). Evidence that use of the ITIL framework is effective. In *18th Annual Conference of the National Advisory Committee on Computing Qualifications* (pp. 160–167). Tauranga, Newzealand.
- QSR. (2013). NVivo [Computer software]. QSR International.
- Qualtrics. (2013). *Qualtrics survey suite online*.
- Rathmann, T., & Stöckle, G. (2012). *Blueprints for construction of the technical support infrastructure for WissGrid (in German)* (pp. 1–29). Karlsruhe, Germany.
- Reddit. (2014). Retrieved October 10, 2014, from www.reddit.com
- Reed, W. A. (2014). *User-centered design: A guide to ROI and ISO 9241-219*. Las Vegas, NV, USA.
- Rocco, T. S., Bliss, L. A., Pérez-Prado, A., & Gallagher, S. (2003). Taking the next step: Mixed methods research in organizational systems. *Information Technology, Learning and Performance*, 21(1), 19–29.
- Rolland, C. (1998). A comprehensive view of process engineering. In B. Pernici & C. Thanos (Eds.), *Proceedings of the 10th International Conference on Advanced Information Systems Engineering, CAiSE'98*. Pisa, Italy: Springer. doi:10.1007/BFb0054216
- Rosemann, M. (2001). *Business process lifecycle management*. [White Paper].
- Roth-Berghofer, T. (2004). Learning from HOMER, a case-based help desk support system. In G. Melnik & H. Holz (Eds.), *6th International Workshop on Advances in Learning Software Organizations* (pp. 88–97). Banff, Canada: Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-25983-1_9
- Rousseau, D. (1990). Quantitative assessment of organizational culture: The case for multiple measures. In B. Schenei (Ed.), *Frontiers in Industrial and Organizational Psychology* (pp. 153–190). San Francisco, U.S.A.: Jossey Bass.
-

- Rummler, G. A., & Brache, A. P. (1995). *Improving performance: How to manage the white space in the organization chart* (Second.). San Francisco: Jossey Bass.
- Scullen, J. (2001). Re-engineering desktop support at Griffith University. In *EDUCAUSE Conference* (pp. 33–39). Gold Coast, Australia: Council of Australian University Librarians (CAUL). Retrieved from <http://nla.gov.au/nla.arc-31940>
- Service Level Agreement Zone. (2013). *Outlining core elements of SLA. The SLA Toolkit*. Retrieved March 13, 2014, from <http://www.sla-zone.co.uk/>
- SETI@home. (2011). *SETI@home is a scientific experiment that uses internet-connected computers in the search for extraterrestrial intelligence (SETI)*. Retrieved from <http://setiathome.ssl.berkeley.edu>
- Sharifi, M., Ayat, M., Rahman, A. A., & Sahibudin, S. (2008). Lessons learned in ITIL implementation failure. In *International Symposium on Information Technology* (pp. 1–4). IEEE. doi:10.1109/ITSIM.2008.4631627
- Siehl, C. (1985). Organisational culture. In P. J. . Frost (Ed.), (pp. 125 – 140). Beverly Hills: Sage Publishing.
- Soehner, C., Steeves, C., & Ward, J. (2010). *E-Science and data support services* (p. 65). Washington, DC, USA.
- Sofaer, S. (1999). Qualitative methods: What are they and why use them? *HSR: Health Services Research*, 35(5), 1101–1118.
- Srivastava, R. K., Shrevani, T. A., & Fahey, L. (1999). Marketing, business processes, and shareholder value: An organizationally embedded view of marketing activities and the discipline of marketing. *Journal of Marketing*, 63(January), 168–179.
- Sutton, R. I. (1997). The virtues of closet qualitative research. *Organisation Science*, 8(1).
- Szepieniec, T. (2013). *D 6.1: Service management tools implementation and maturity baseline assessment framework*. Munich, Germany.
- Tashakkori, A., & Teddlie, C. (2000). *An excellent introduction to mixed methodology research*. Thousand Oaks, CA, USA: Sage Publishing.
- Taylor, K. E., Stouffer, R. J., & Meehl, G. a. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93(4), 485–498. doi:10.1175/BAMS-D-11-00094.1
- Team, xGUS help desk. (2013). *xGUS user`s manual* (pp. 1–33). Karlsruhe, Germany.
- TextGrid. (2013). *TextGrid*. Retrieved December 02, 2014, from <https://www.textgrid.de/>
-

- The Belgian Grid For Research. (2011). *BEgrid*. Retrieved September 23, 2013, from <http://www.begrid.be/index.php?module=webpage&id=5>
- The DataGrid Project*. (2004). Retrieved September 25, 2014, from <http://eu-datagrid.web.cern.ch/eu-datagrid/>
- The IT Governance Institute. (2007). *COBIT 4.1* (p. 167). Rolling Meadows, IL, USA.
- Underwood, J. A., Hegdahl, D., & Gimbel, J. (2003). A proper set of tools are needed to corral support. In *Proceedings of the 31st annual ACM SIGUCCS conference on User services - SIGUCCS '03* (pp. 23–26). New York, New York, USA: ACM Press. doi:10.1145/947469.947478
- Unicore Forum. (2013). *UNICORE*. Retrieved November 13, 2014, from <http://www.unicore.eu/UNICORE>
- Van Aken, J. E., & Romme, G. (2009). Reinventing the future: Adding design science to the repertoire of organization and management studies. *Organization Management Journal*, 6, 5–12. doi:10.1057/omj.2009.1
- Van de ven, A. H. (1992). Suggestions for studying strategy process: A research note. *Strategic Management Journal*, 13(S1), 169–188. doi:10.1002/smj.4250131013
- Venugopal, S., Buyya, R., & Ramamohanarao, K. (2006). A taxonomy of data grids for distributed data sharing, management, and processing. *ACM Computing Surveys*, 38(1), 1–60. doi:10.1145/1132952.1132955
- Vu, L. (2013). *Earth system grid federation: A modern day “silk road” for climate data*. *Energy Science Network*. Retrieved August 20, 2014, from <https://es.net/news-and-publications/esnet-news/2012/ESGF/>
- Ward, C., Aggarwal, V., Bucu, M., Olsson, E., & Weinberger, S. (2007). Integrated change and configuration management. *IBM Systems Journal*, 46(3), 459–478. doi:10.1147/sj.463.0459
- Watson, T. J. (1995). Rhetoric, discourse and argument in organizational sense making: A reflexive tale. *Organization Studies*, 16(5), 805–821.
- Williams, D. N. (2012). *Earth system grid federation (ESGF): Future and governance* (pp. 1–17). Livermore, CA, USA.
- Williams, D. N. (2015). *ESGF project overview* (p. 5). Livermore, CA, USA.
- Williams, D. N., Bell, G., Cinquini, L., Fox, P., Harney, J., & Goldstone, R. (2013). ESGF: Federated and integrated climate data from multiple sources. In W. Hiller, R. Budich, & R. Redler (Eds.), *Earth System Modelling* (Vol. 6, pp. 61–77). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-37244-5
-

- Williams, D. N., Drach, R., Ananthakrishnan, R., Foster, I. T., Fraser, D., Siebenlist, F., ... Sim, A. (2009). The earth system grid: Enabling access to multimodel climate simulation data. *Bulletin of the American Meteorological Society*, 90(2), 195–205. doi:10.1175/2008BAMS2459.1
- Williams, D. N., & ESGF. (2013). *ESGF governance model*. Livermore, CA, USA.
- World Meteorological Organization, & UNEP. (2015). *Inter-governmental panel on climate change. IPCC*. Retrieved January 10, 2015, from <http://www.ipcc.ch/organization/>
- XGUS website*. (2013). Retrieved from <http://xgus.scc.kit.edu/90.php>
- Yin, R. (2013). *Case study research: Design and methods* (5th ed., p. 240). Thousand Oaks, California, U.S.A.: Sage Publishing. Retrieved from <http://www.sagepub.com/repository/binaries/pdfs/HistoryofMethods.pdf>
- Zukerman, I., & Marom, Y. (2006). A comparative study of information gathering approaches for answering help-desk e-mail inquiries. In A. Sattar & B. H. Kang (Eds.), *Proceedings of the 19th Australian Joint Conference on Artificial Intelligence, AI 2006* (pp. 546–556). Hobart, Australia: Springer Berlin Heidelberg.
-

11 Appendix A

Table 11.1 The summary of the development of ESGF since its inception.

Version of ESGF	geographical range	Years Active	Data-holdings supported	Total registered users	Active users per month	Scalability	Main Features	Stake holders - Partners	Participating Institutes	User-support features
ESG	Cross country	1999-2001	Parallel Climate Model (PCM) 860 GB, 2000 downloadable files (df) available via FTP in 1999, 20,000 df, 3.7 TB in yr. 2000, 30,000 df, 6.8 TB in yr. 2001	Few hundreds	none	Very low	Client-server architecture (CSA), Prototype: climate “data browser”, Data-grid technologies for managing and replication of large data sets	DOE (main funder)	US: ANL, LANL, LBNL, LLNL, NCAR, USC/ISI	none
ESG II	Cross country	2001-2006	PCM, CCSM, CMIP3 (at later stage)	4,000 to 10,000 users	100	low	One gateway (web-based user portal), Web-based user registration and authentication, data	DOE (main funder)	ORNL, ANL, LANL, LBNL, LLNL,	informal communication initiated by CMIP3 data project

			53,000 df (13 TB) in 2002 100,000 df (16 TB) in 2003	In 2004 opened to friendly users			transfer to client via OPeNDAP-g distributed protocol, metadata, access to archival storage		NCAR, USC/ISI	users
ESG - CET	Global	2006-2011	CMIP3, NARCCAP, CSS EF, CCSM, POP, AIS, MLS 1 Petabyte served	22,000 to 25,000	500-800	fair	Four gateways, DataMover Tool, Storage Resource Managers (SRM)	DOE (main funder)	ORNL, ANL, LANL, LBNL, LLNL, NCAR, USC/ISI, PMEL	e-mail communication channels and mailing lists
ESGF	Global	2011-till present	CMIP3, CMIP5, GeoMIP, PMIP, TAMIP, obs4MIPs, ANA4MIPs, C-LAMP,	27000	1500	Very good	13 gateways Integration with many external projects: ESC, Metafor, GIP,	DOE (main funder), IS-ENES	US: NOAA's, GFDL ESRL, PMEL, DOE's	Request tracking systems: RT3, Github (US),

			CORDEX, NARCCAP, CSS EF, CCSM, EUCLIPSE, POP, AIS, MLS 8 Petabytes				C3Grid, IS-ENES etc., Distributed and federated architecture and others ⁴⁰ .		LANL, ORNL, ANL, LLNL, LBNL; NSF's NCAR, NASA's JPL. US Universitie s ⁴¹ , EU ⁴² , ASIA ⁴³ and Australia ⁴⁴	FootPrints (especially from UK), user-support mailing list, later Askbot, Wiki, FAQ website, limited web- casts
--	--	--	---	--	--	--	--	--	---	--

⁴⁰ Other features include: discipline specific portals, Support browser-based and direct client access, Single Sign-on (SSO), Automated script and GUI-based publication tools for data publishers, Full support for data aggregations: A collection of files, usually ordered by simulation time, that can be treated as a single file for purposes of data access, computation, and visualization, User notification service, Users can choose to be notified when a data set has been modified.

⁴¹ USC's Information Sciences Institute (ISI), Purdue University, University of Michigan, University of California, Davis

⁴² IS-ENES, BADC, UK-MetOffice, DKRZ, MPIM, IPSL, LSCE

⁴³ Univ. of Tokyo, Japanese Centre for Global Environmental Research, Jamstec, Korea Meteorological Administration

⁴⁴ ANU, Australian Research Collaboration Service, Government Department of Climate Change, Victorian Partnership for Advanced Computing, Australian Environment and Resource Management

Declaration on Oath

I hereby declare, on oath, that I have written the present dissertation by my own and have not used other than the acknowledged resources and aids.

Hamburg, 03.03.2015

A handwritten signature in blue ink, consisting of several loops and a horizontal stroke extending to the right.

Hashim Iqbal Chunpir
