

Evolution of e-Research: From Infrastructure Development to Service Orientation

Hashim Iqbal Chunpir^{1,2(✉)}, Thomas Ludwig^{1,2},
and Dean N. Williams³

¹ German Climate Computing Center (DKRZ), Bundesstr. 45a,
20146 Hamburg, Germany

{chunpir, ludwig}@dkrz.de

² Department of Informatics, University of Hamburg, Hamburg, Germany

³ Lawrence Livermore National Laboratory (LLNL), Livermore, CA, USA
williams13@llnl.gov

Abstract. E-Research has reframed the process of research. Researchers can now access distributed data around the globe with the help of e-Research infrastructure. This paper presents an overview of the developmental process and evolution of an e-Research platform: Earth System Grid Federation (ESGF) that evolved from a research infrastructure test-bed to a services oriented platform, in subsequent phases. ESGF is a leading distributed peer to peer data grid system in Earth System Modelling having around 27000 users distributed all over the world. Currently, it is a challenge faced by most of the e-Research facilities to provide user oriented services. Moreover, there is a strong need to conduct user experience and usability studies of e-Research facilities like ESGF, which is in demand. However, very few steps have been taken in practice to create a better user experience (UX), so that users' interest can be generated to interact with e-infrastructures, on an intuitive basis. Hence, thriving the practice of e-Research and making it more interesting overtime. Finally, this paper indicates at the service oriented and usability aspects of e-Science infrastructures.

Keywords: Service orientation of e-Science systems · E-Research · User support · Help desk · User experience · User-centered design · Federated e-Research facilities

1 Introduction

Usability and user service orientation of big data has always been an interesting topic in Human Computer Interaction (HCI), Computer Supported Cooperative Work (CSCW), e-Research, big data and other fields as it refers to use of technology via which a user can achieve big discoveries. The recent discovery of Higgs Boson Particle was also accomplished via an e-Research facility at CERN (reported by BBC on 1st August 2012). Users of e-Research technology are chiefly researchers and they realize multifarious tasks of their research that they need to complete within a specific time-frame, via e-Research infrastructures. In the past, in e-Research facilities, a massive development effort and money has been invested in making infrastructure

stable and reliable [1, 2]. Currently, in most of e-Research facilities, a relatively stable and reliable state has been achieved such that they can serve researchers by fulfilling their research needs [3]. Moreover, research and development in improving the user oriented services in e-Research facilities has been done to some extent [2, 3]. However, the usability aspects of e-Research infrastructure have hardly been addressed [4].

The organization of this paper is as follows: Sect. 2 provides a brief background on this study related to the field of e-Research. The research methods used in this study are explained in Sect. 3. An overview on the historical development of ESGF is given in Sect. 4. Discussion and future work on ESGF is given in Sect. 5. Finally, concluding remarks are provided in Sect. 6.

2 Background

This paper analyzes the historical development of a well-known e-Research facility Earth System Grid Federation (ESGF) that is serving data projects in climate science domain. ESGF facilitates to study climate change and impact of climate change on human society and Earth's eco system [5]. Moreover, this paper provides an overview of the structural, organizational, functional development achieved in ESGF, including the expansion in volume as well as number of data projects that is served with respect to user oriented services. With the passage of time the users of ESGF have also increased in number and currently there are 27,000 users using this e-Research facility for research purposes.

The structural, organizational and functional development of ESGF includes, for instance; programming, architectural design, connectivity of distributed components, strategic planning including governance with organization, node administration and others. The process of operating, maintaining and further expanding the e-Research facility including its data is iterative in nature [1, 6]. Apart from the structural, organizational and functional development in ESGF, progress in servicing end-user requests has also been done. The progress includes offering user support in the form of self-help via support websites, online tutorials, wikis or contacting an expert in the form of traditional help-desk [7, 8] and service-desk [9–11]. In this paper, the environmental complexity and the contemporary practices of user support services is presented. The paper then emphasizes on the need to enhance the usability and user experience in e-Research.

3 Research Methodology

In this research, case study research method is used and an important practical use-case in e-Research, in the field of climate science is ESGF project. This study is based on a single, in-depth, synchronic case in the e-Research infrastructure user services sector. A case study is “a research strategy which focuses on understanding the dynamics present with single settings” [12]. It can also be defined as “an empirical inquiry that investigates a contemporary phenomenon within its real-life context” [13].

The methods of data collection used in this study are: Participatory observations, informal meetings and archival analysis of documents relevant to ESGF.

A reason for the choice of the case study approach as a research method is that it suits well for studying service processes that are linked to a complex organizational context, as these cases offer in-depth view of development in organisations [14]. Moreover, the use of the case study is appropriate since e-Research infrastructure support involves a large number of actors in the form of distributed users and support teams and practices where the boundaries between these constituents are not easily distinguishable.

4 Evolution of ESGF

In this section the history of development of the ESGF along with the significant changes in the infrastructure and organization 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. The summary of salient features of the historical development of ESGF in these phases is discussed in the forthcoming sections. Phase 1 from 1999 to 2001 when it was called ESG-I is presented in Sect. 4.1, phase 2 from 2001 to 2006 when it was called ESG-II is given in Sect. 4.2, phase 3 from 2006 to 2011 when it was called ESG-CET is described in Sect. 4.3 and finally the current phase of ESGF which is from 2011 onwards is given in Sect. 4.4.

4.1 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 analyze 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 program (NGI) to cater the needs of climate scientists and to fulfil the emerging challenge of climate data [15]. 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 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) program. 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 Inter-comparison 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” [16].

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-g¹ protocol. As a result, the system started supporting 10,000 registered international users and managed some 200 terra-bytes (TB) of data [16]. 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 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

¹ 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/).

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.

4.3 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 [16]. 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 first author, who was working on C3Grid, a collaborative project of the ESG; to investigate the user support process, usability as well as user experience aspects in e-Research infrastructure and ESG-CET (now known as ESGF) was chosen as a case study.

4.4 ESGF P2P: The Current Phase (from 2011 Onwards)

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 [17]. 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, 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 evolved in the form, what now 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 [18]. It now reflects a broad array of contributions from the collaborating partners.

It is interesting to note that with the enormous growth in the organizational structure of ESGF, an upward trend of registered users was recorded [19]. The registered users reached almost 27,000 in number, from different parts of the world, with almost 700 to 800 active users per month. With this rapid development in the ESGF organization, there is an ever-increasing need to meet long-term user-support requirements, as the number of data-holdings and number of users rise. 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.

5 Discussion and Future Work

Looking at the history of ESGF, one can conclude that the architecture and organization 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 [20]. 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.

ESGF is stated as: “The Earth System Grid Federation (ESGF) is a multi-agency, international collaboration of people and institutions working together to build an open source software infrastructure for the management and analysis of Earth Science data on a global scale” [21]. However, in this ESGF definition user support services are not explicitly made part of the definition of ESGF. Keeping the anatomy of ESGF P2P data archive system in view, the user support system of ESGF has its sub-units. Though these sub-units are not formally designated as support units within administrative

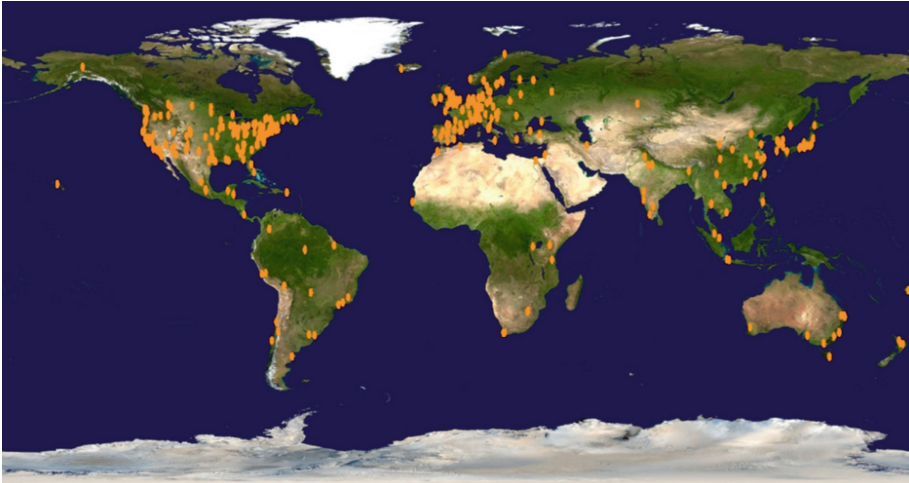


Fig. 1. Distribution of administrative units of ESGF

bodies, the support units are implicitly part of administrative bodies. Consequently, from the geographically distributed organization of ESGF P2P network, it is understandable that each and every administrative domain have their own practices of handling user-requests. This observation is also evident from the qualitative cum quantitative inquiry into the support practices of ESGF undertaken by the authors. The diversity of practices in handling user queries by the user support staff, who themselves are developers of the ESGF system, form different support structures and models followed in each of the administrative domains that tend to make a heterogeneous user support process. Subsequently, the user support system in ESGF does not comply with any set standards of processing user support requests.

From the Fig. 1 one can anticipate that the numbers of administrative bodies (principal investigation institutions plus developing teams of ESGF), data holdings, ESGF users, and ESGF staffs participating in ESGF P2P system are subject to increase. Additionally, the role of an administrative body and its attached components such as nodes are subject to change, therefore this whole ESGF set-up is a complex, dynamic and evolving in nature. In ESGF system there is a continuous architecture re-designs activities, software development, hardware changes, data publishing, data curation, data quality check and other activities. Attached to these core operational activities is the necessity of the user support activities that cannot be ignored. A dynamic and an ever-evolving infrastructure need a dynamic user support “service desk.” Therefore, e-Science is likely to be confronted with demanding issues of long-term and continuity of service, particularly related to user support services which is quite similar to data curation and software development.

In the future, ESGF will cover other scientific domains such as health sciences, biology, chemistry, energy as well radio-astronomy. With these additions of domains

more users will be using the ESGF e-Research facility, consequently making usability and user experience very important. Therefore, there is a need to measure the user experience and usability of the interfaces provided by ESGF to the actors who interact with the ESGF system. Moreover, the current user interfaces of ESGF are needed to be evaluated and based on the findings of the investigation, recommendations can be made in future to enhance the usability and user experience.

6 Conclusion

In this paper a historical overview of the organizational and structural advances achieved in ESGF e-Research facility of climate science domain is presented. Furthermore, in this study, the evolution of ESGF e-Research infrastructure is observed with the help of research techniques such as participant observation, archival analysis and informal meetings. ESGF has evolved from a non-user oriented research experimental testbed towards a user-oriented environment. Currently, it is evident from the observations that the aspects of usability and user experience within e-Research facilities especially in climate science need improvement. Attention of sponsors, operational and executive staff members of e-Research facilities in climate science such as ESGF is required to improve the usability needs of users. Further studies are needed to be conducted with the users of ESGF to capture the user experience and restructure the complex and dynamic interfaces provided to different types of users of the e-Research facility. Currently, the authors are working on a conceptual model to improve the user experience and usability standards in e-Research interfaces within the climate science and other domain. In the future, the authors will observe the effectiveness of transferring front line user support units to various institutes in developing countries.

Acknowledgement. We appreciate the sincere participation and support of our all colleagues and the users of ESGF.

References

1. Chunpir, H.I., Badewi, A.A., Ludwig, T.: User support system in the complex environment. In: Marcus, A. (ed.) *Design, User Experience, and Usability. User Experience Design Practice. Lecture Notes in Computer Science*, vol. 8520, pp. 392–402. Springer International Publishing, Cham (2014)
2. Chunpir, H.I., Ludwig, T.: Reviewing the governance structure of end-user support in e-science infrastructures. In: *Informatik 2014 Proceedings of Mastering Big Data Complexity. Lecture Notes in Informatics (LNI)*, vol. 232 (2014)
3. Chunpir, H.I., Ludwig, T., Badewi, A.: A snap-shot of user support services in earth system grid federation (ESGF): a use case of climate cyber-infrastructure. In: *Proceedings of the 5th Applied Human Factors and Ergonomics (AHFE) Conference*, July 2014

4. Hey, T., Trefethen, A.E.: Cyberinfrastructure for e-science. *Science* **308**(5723), 817–821 (2005)
5. Cinquini, L., Crichton, D., Mattmann, C., Bell, G.M., Drach, B., Williams, D., Harney, J., Denvil, S., Schweitzer, R.: 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 (2012)
6. Chunpir, H.I., Ludwig, T., Badewi, A.A.: Using soft systems methodology (SSM) in understanding current user-support scenario in the climate science domain of cyber-infrastructures. In: Marcus, A. (ed.) DUXU 2014, Part III. LNCS, vol. 8519, pp. 495–506. Springer, Heidelberg (2014)
7. Leung, N., Lau, S.: Information technology help desk survey: to identify the classification of simple and routine enquiries. *J. Comput. Inf. Syst.* **47**(1), 70–81 (2007)
8. Leung, N.K.Y.: University of Wollongong Thesis Collection Turning user into first level support in help desk : development of web-based user self-help knowledge management system (2006)
9. Jäntti, M.: Improving IT service desk and service management processes in finnish tax administration: a case study on service engineering. In: Dieste, O., Jedlitschka, A., Juristo, N. (eds.) PROFES 2012. LNCS, vol. 7343, pp. 218–232. Springer, Heidelberg (2012)
10. Jäntti, M.: Lessons learnt from the improvement of customer support processes: a case study on incident management. In: Bomarius, F., Oivo, M., Jaring, P., Abrahamsson, P. (eds.) PROFES 2009. LNBIP, vol. 32, pp. 317–331. Springer, Heidelberg (2009)
11. Jäntti, M.: Examining challenges in IT service desk system and processes: a case study. In: Seventh International Conference on Systems, ICONS 2012, pp. 105–108 (2012)
12. Yin, R.: *Case Study Research Design and Methods*, 5th edn, p. 240. Sage Publishing, Thousand Oaks (2013)
13. Eisenhardt, K.: Building theories from case study research. *Acad. Manag. Rev.* **14**(4), 532–550 (1989)
14. Buchanan, D.A.: Case studies in organisational research. In: Symon, G., Cassel, C. (eds.) *The Practice of Qualitative Organisational Research: Core Methods and Current Challenges*, pp. 373–392. Sage Publishing, London (2012)
15. Bernholdt, D., Bharathi, S., Brown, D., Chanchio, K., Chen, M., Chervenak, A., Cinquini, L., Drach, B., Foster, I., Fox, P., Garcia, J., Kesselman, C., Markel, R., Middleton, D., Nefedova, V., Pouchard, L., Shoshani, A., Sim, A., Strand, G., Williams, D.: The earth system grid: supporting the next generation of climate modeling research. *Proc. IEEE* **93**(3), 485–495 (2005)
16. Williams, D.N., Drach, R., Ananthkrishnan, R., Foster, I.T., Fraser, D., Siebenlist, F., Bernholdt, D.E., Chen, M., Schwidder, J., Bharathi, S., Chervenak, A.L., Schuler, R., Su, M., Brown, D., Cinquini, L., Fox, P., Garcia, J., Middleton, D.E., Strand, W.G., Wilhelmi, N., Hankin, S., Schweitzer, R., Jones, P., Shoshani, A., Sim, A.: The earth system grid: enabling access to multimodel climate simulation data. *Bull. Am. Meteorol. Soc.* **90**(2), 195–205 (2009)
17. ENES: The ENES portal: European network for earth system modelling (2013). <https://verc.enes.org/>. Accessed 02 December 2014
18. ESGF: ESG: Earth system grid. ESGF (2010). <https://www.earthsystemgrid.org/about/overview.htm>. Accessed 13 May 2014

19. ESGF: ESGF use statistics for CMIP5 data sets (2015)
20. Jackson, S.J., Edwards, P.N., Bowker, G.C., Knobel, C.P.: Understanding infrastructure: history, heuristics and cyber infrastructure policy. *First Monday* **12**(6) (2007). <http://pearl.acc.uic.edu/ojs/index.php/fm/article/view/1904/1786>
21. Williams, D.N.: Earth system grid federation (ESGF): future and governance, Livermore, CA, USA (2012)