eResearch2020

The Role of
e-Infrastructures
in the Creation of
Global Virtual Research
Communities

Final Report

A study by

empirica empirica Gesellschaft für Kommunikations- und
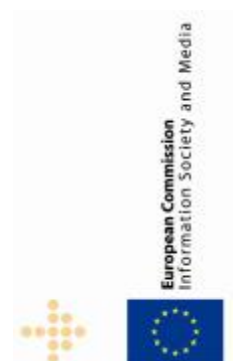Technologieforschung mbH

n|w Fachhochschule Nordwestschweiz
Hochschule für Wirtschaft

Oxford Internet Institute, University of Oxford

NORC National Opinion Research Center,
University of Chicago

European Commission
Information Society and Media

Disclaimer

The views expressed in this report are those of the authors and do not necessarily reflect those of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the information provided in this document.
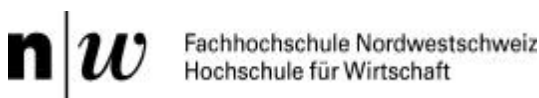
The study team

This study has been conducted by:

empirica Gesellschaft für Kommunikations- und Technologieforschung mbH

Tobias Hüsing, Simon Robinson

Fachhochschule Nordwestschweiz, Hochschule für Wirtschaft

Franz Barjak, Oliver Bendel and Gordon Wiegand

Oxford Internet Institute, University of Oxford

Kathryn Eccles, Eric Meyer and Ralph Schroeder

National Opinion Research Center, University of Chicago

Zack Kertcher and Erica Coslor

Contact

For further information about the study please contact:

empirica
Gesellschaft für Kommunikations- und Technologieforschung mbH
Oxfordstr. 2, 53111 Bonn, Germany
Fax: (49-228) 98530-12
info@empirica.com

Rights restrictions

Bonn and Brussels, February 2010

Executive Summary

>> Research Questions

e-Infrastructures radically change the way research is conducted, overcoming distance to support a growing multitude of virtual research communities across the globe. The eResearch2020 consortium has conducted research on a diverse sample of e-Infrastructures from around the world, talking to both developers and users. The aim is to improve policy, enhance technology adoption and facilitate the creation of global virtual research communities.

e-Infrastructures can be defined as networked tools, data and resources that support a community of researchers, broadly including all those who participate in and benefit from research. Following this definition, the term e-Infrastructure comprises very heterogeneous projects and institutions within the scientific community.

e-Infrastructures include services as diverse as the physical supply of backbone connectivity, single- or multi-purpose grids, supercomputer infrastructure, data grids and repositories, tools for visualization, simulation, data management, storage, analysis and collection, tools for support in relation to methods or analysis, as well as remote access to research instruments and very large research facilities.

The impact of e-Infrastructures on virtual research communities will especially be affected by:

- The regulation and governance of e-Infrastructures,
- the integration or separation of e-Infrastructures at national and disciplinary levels,
- different organizational and business models,
- considerations of research communities' needs and practices in the services provided by e-Infrastructures

The eResearch2020 case studies and survey have revealed a multitude of approaches to all these topics in today's e-Infrastructure development and operation.

Key questions addressed in the study included:

- To what extent do e-Infrastructures contribute to the establishment of global virtual research communities? Do they reduce disadvantages of researchers in peripheral regions and developing countries?
- What are the organisational structures and coordination mechanisms of e-Infrastructures, their key players in the interaction with the researcher communities, the relevant regulatory and policy aspects and the support they receive by funding and other external bodies?
- How well do e-Infrastructure providers define, consult, plan for, engage with and overcome bottlenecks in scaling up to match growth in their user community?
- How do e-Infrastructures ensure that they make an essential contribution to their community of beneficiaries?
- How do researchers use e-Infrastructures? What are the main benefits and costs for global virtual research communities, and to what extent do they influence adoption and use?
- Given current trends, what e-Infrastructure and virtual communities can we expect in the future?
- What policy action can enhance the impact of e-Infrastructures on virtual research communities and how can a Roadmap for European e-Infrastructures be devised?

e-Infrastructures represent a very heterogeneous subject of investigation, they span continents, scientific and professional practices, functions and technologies. eResearch2020 examined how both providers and the respective virtual research communities are using, shaping and steering e-Infrastructure services. The approach included a survey of users and interviews with e-Infrastructure officials.

Survey of e-Infrastructures

In a qualitative cross-case comparison, eResearch2020 selected a sample of e-Infrastructures to cover a wide range in terms of the existing development, geographic spread, project maturity, and size, for a total of 18 cases. In-depth interviews with key informants and archival analysis enabled the identification of common themes across the cases.

e-Infrastructure sample		
	e-Infrastructure	ESFRI category
Providers	DEISA	e-Infrastructure
	EELA-2	e-Infrastructure
	EGEE	e-Infrastructure
	GÉANT	e-Infrastructure
	OSG	e-Infrastructure
	Teragrid	e-Infrastructure
	Swedish National Data Service	Social Sciences and Humanities (Biological and Medical Sciences too)
User communities	C3-Grid	Environmental Sciences
	CineGrid	e-Infrastructure
	CLARIN	Social Sciences and Humanities
	D4science	Environmental Sciences
	DARIAH	Social Sciences and Humanities
	DRIVER	e-Infrastructure
	ETSF	Materials and Analytical Facilities
	MediGrid	Biological and Medical Sciences
	NVO	Physical Sciences and Engineering
	Swiss BioGrid	Biological and Medical Sciences
Standards	OGF – Open Grid Forum	e-Infrastructure

Survey of Research Communities

On the basis of the qualitative cross-case comparison, a survey was designed and administered. The invitation to take part was sent through e-Infrastructure contact persons and distributed widely to others participating in e-Infrastructure-related activities.

More than 400 individuals filled in the online questionnaire. Responses were obtained from a broad set of countries - more than 50% from respondents in the EU27 and a small share from other European countries. North America (exclusively the US) yielded 10% of responses and Latin America – above all, Brazil, Colombia, Argentina, Venezuela and Ecuador – another 21%.

The survey also achieved a good coverage of academic functions, including scholars, researchers, other professionals and administrators, and of fields of research and development.

Respondents by research domains, fields of work, or area of development activities		
	Frequency	in % of total
a) Research domains		
Astronomy or Astrophysics	24	6.2
Biological Sciences and Medicine	32	8.2
Chemical and Material Sciences	18	4.6
Computer and Information Sciences	36	9.3
Engineering and Technology	20	5.2
Earth and Other Natural Sciences	18	4.6
Physical Sciences	21	5.4
Social Sciences and Humanities	13	3.4
b) Fields of work		
Academic support services	12	3.1
Non-academic support services	17	4.4
c) Area of development activities		
Academic and IT support services	37	9.5
Supercomputing and distributed computing	66	17.0
Networking	16	4.1
Application Development	35	9.0
Other	23	5.9
Total	388	100

## >> The Empirical Picture: The User Perspective

Typically, virtual research communities are medium-sized, truly global or spanning several countries with grid computing being used as the most popular service.

Features of Existing Global Virtual Research Communities

Most of the virtual research communities that our respondents reported on are medium sized, with 21-100 researchers working on the same problems on a particular e-Infrastructure." to „The survey respondents mostly reported on small or medium sized virtual research communities: 28% work on related problems on a particular e-Infrastructure in communities of 21-100 researchers. 15% of the respondents reported a very small research community of no more than 5 researchers, and another 18% reported no more than 20 peers.
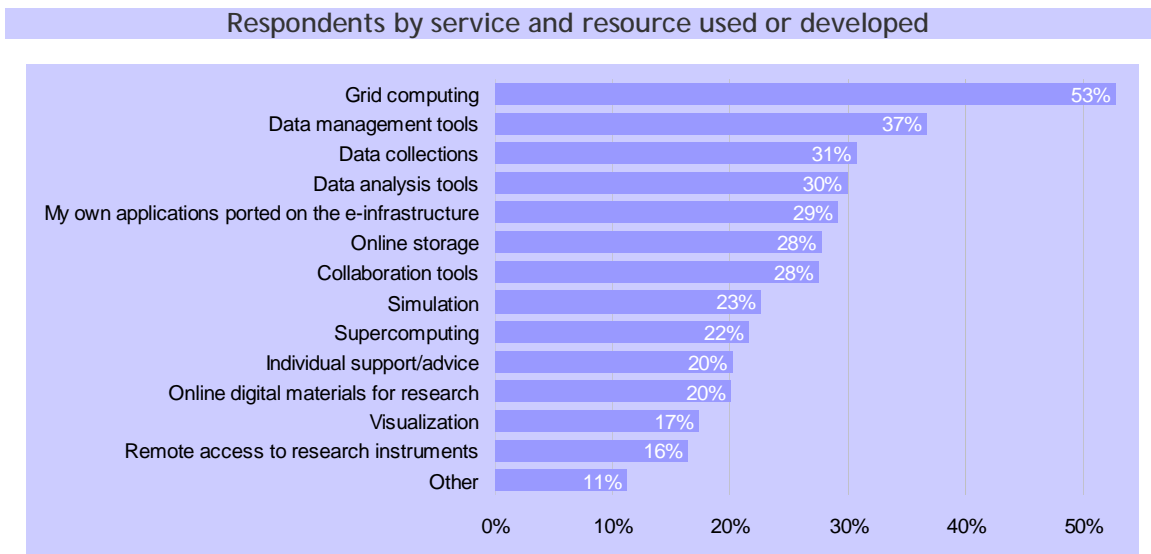
Survey respondents - Users of e-infrastructures
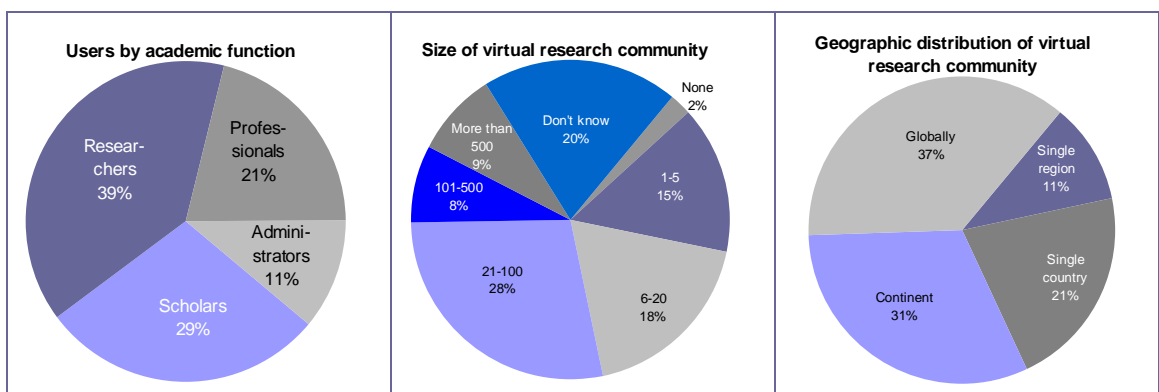
In geographic terms, most (37%) virtual research communities turned out to be truly global, spanning more than one continent, while 31% are continent wide and 32% national.

Grid computing is the service used by the vast majority of respondents. Communities using data management tools and data collections are also very prevalent.

Researchers who port their own applications on to the e-Infrastructure make up a sizable group, which cautions against assuming a clear distinction between „users" and „developers" in interpreting developments in e-Infrastructures.
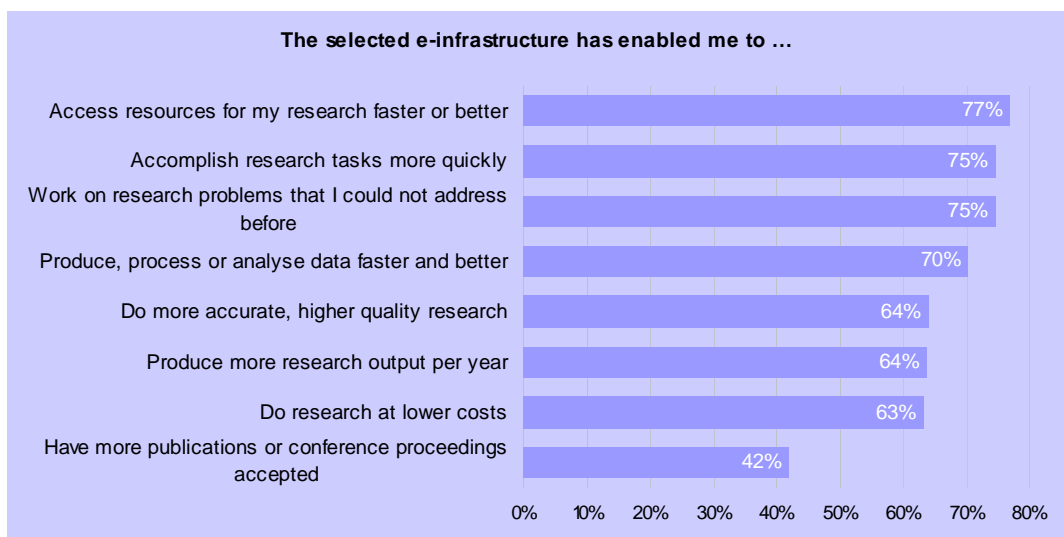
Respondents by service and resource used or developed

Service/Resource	Percentage
Grid computing	53%
Data management tools	37%
Data collections	31%
Data analysis tools	30%
My own applications ported on the e-infrastructure	29%
Online storage	28%
Collaboration tools	28%
Simulation	23%
Supercomputing	22%
Individual support/advice	20%
Online digital materials for research	20%
Visualization	17%
Remote access to research instruments	16%
Other	11%

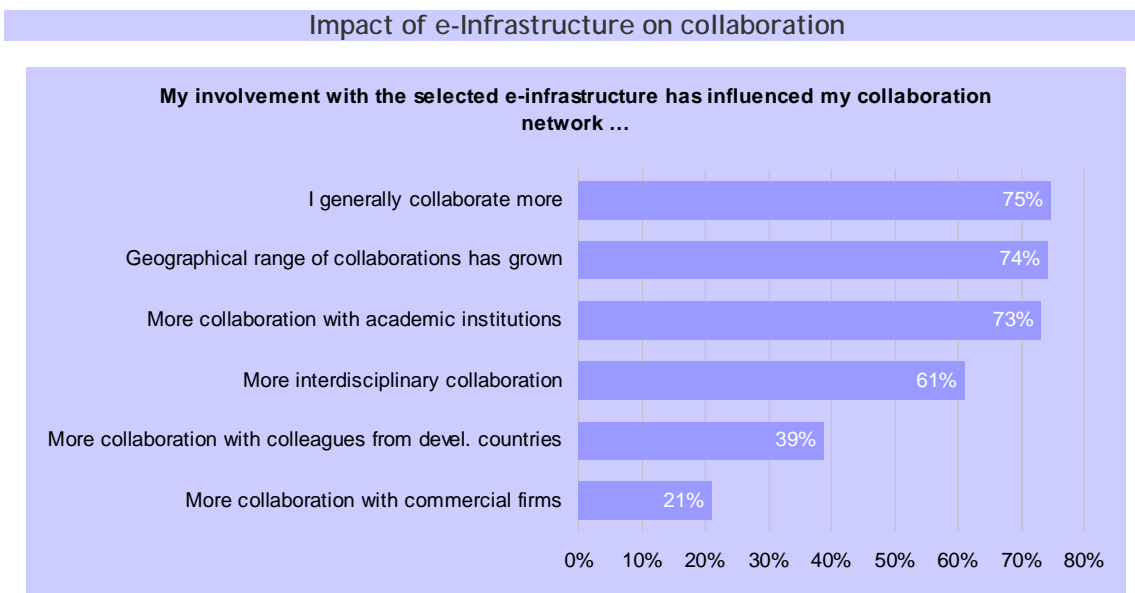## >> Impact of e-Infrastructure

Mostly Positive Impact

More than 85% of e-Infrastructure users classify e-Infrastructure as important or very important for their work. Most would also see their research or work programmes impaired if the e-Infrastructure did not exist. Early adopters more often report relying on the availability of the e-Infrastructure than those who became involved later. It apparently takes some time for benefits of e-Infrastructure to materialize, and benefits are often over-shadowed by costs at the outset.

Impact of e-Infrastructure on research practice and output

The selected e-infrastructure has enabled me to …

	Percentage
Access resources for my research faster or better	77%
Accomplish research tasks more quickly	75%
Work on research problems that I could not address before	75%
Produce, process or analyse data faster and better	70%
Do more accurate, higher quality research	64%
Produce more research output per year	64%
Do research at lower costs	63%
Have more publications or conference proceedings accepted	42%

Benefits that were most valued were having the possibility to experiment with new technology, obtaining access to high-end distributed computing, obtaining access to large-scale distributed storage or databases and training and learning effects. Obtaining access to other resources (new software/applications, standards, advanced visualization or remote

instruments) received fewer mentioning . The responses in this case are biased to respondents involved in computing infrastructures.

Impact of e-Infrastructure on collaboration

My involvement with the selected e-infrastructure has influenced my collaboration network …

I generally collaborate more	75%
Geographical range of collaborations has grown	74%
More collaboration with academic institutions	73%
More interdisciplinary collaboration	61%
More collaboration with colleagues from devel. countries	39%
More collaboration with commercial firms	21%

0% 10% 20% 30% 40% 50% 60% 70% 80%

There is widespread agreement about the positive impact of e-Infrastructures. For seven out of eight questions, more than 60% of the respondents agree that there is a positive impact. The main benefits relate to the speed of doing research or work: accomplish tasks more quickly, access resources faster or better, produce processes or analyse data faster or better. Equally important is the ability to work on new problems which could not be addressed with previously available technology. Slightly less frequently respondents agreed to positive effects on productivity ("Produce more output per year"), costs, and quality ("Do more accurate, higher quality research work"). The lowest number of positive responses was on the acceptance of publications, perhaps due to the particular difficulties of assessing this impact.

Catalysts and barriers in the adoption of e-Infrastructure - quotes from users

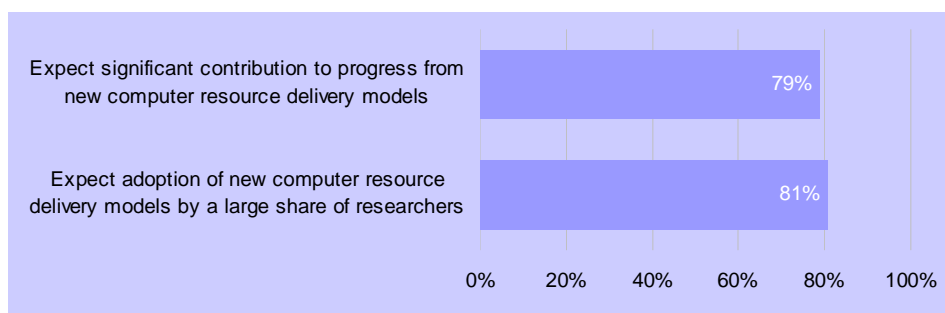|	Catalysts	Barriers
Access to resources	- Access to a larger distributed network than available locally - Sharing of data across multiple institutions - Additional resources available - Computer resources assigned to DEISA	- Reasonable existing local resources - Already have access to other resources elsewhere
Organizational	- Enthusiasm of most stakeholders • - Collaboration among scientists • - Job requirement - Developing high level analysis services for research that requires industrial-strength organization of computation flows - Good infrastructure and organization - Support from colleagues	- No support for radio astronomical data - Grid infrastructure changed often, changes to my application were needed as a result - EU legal constraints not compliant with my institution's requirements - Lack of support from my institution - Low administrative pressure to stimulate the use of these tools - Bureaucracy
Technical capabilities	- Need to bridge interoperability gaps among communities of practices - Reporting tool. - Computing Power and Fault Tolerance capability - Possibility to use state of the art technology	- It is not easy, in basic research, to make detailed statements on how much CPU time will be needed to complete a project - Time required to adapt usual workflows - Lack of structure to support anonymous access

	- Research interest on grid technology and remote instrumentation	- Download and Installation of applications
Ease of use	- User-friendliness - Easy application process - Availability & reliability - Easy writing and uploading project	- Interface - Slow to get to compared to other resources - Difficult to use in the beginning
Funding related	- Funding - Continuous funds to guarantee continuous research - Outsourcing infrastructure management and maintenance costs - The grant of the financing institution	- Developing fundraising and governance structure - Securing national (matching) funding - Cost of network infrastructure - Insufficient funds
Training related	- Technical support and training - Need of HEP communities in Latin America to create support infrastructure	- Time spent to get the application compiled and running - Learning curve - Lack of background in grid computing - Not known by individual researchers - Learning material is good, but sparsely distributed through the web

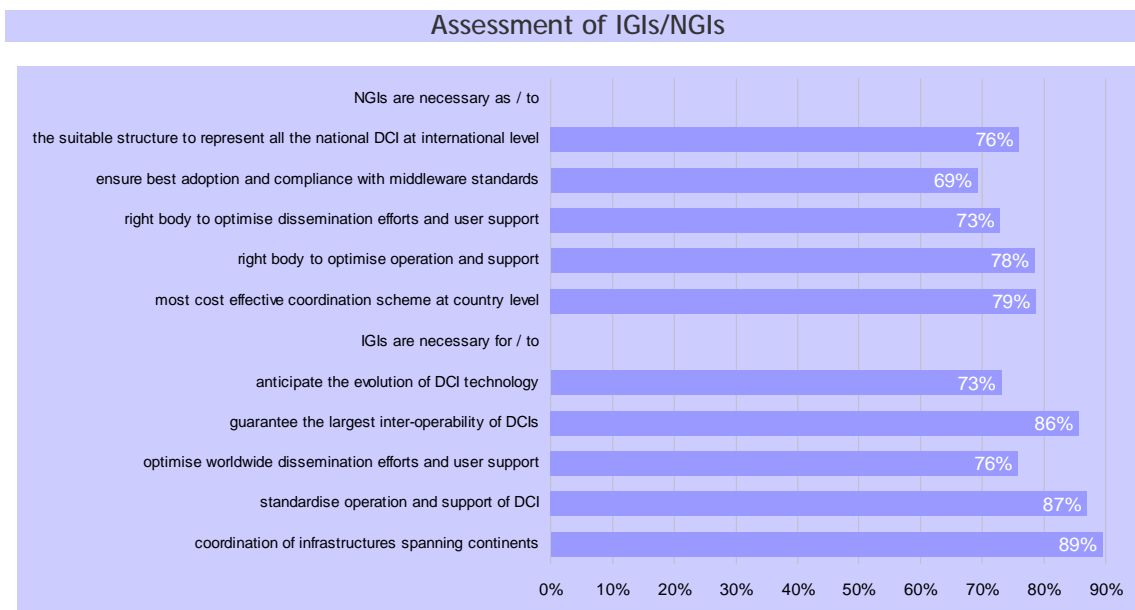## >> Perceived Trends and Policy Requests

A large majority, 80% of those responding, find it likely or very likely that new resource delivery models such as Software as a Service, Cloud Computing or Utility Computing will spread and have a significant impact in science in the next five years.

We see also wide agreement from the respondents to statements about the necessity and benefits of National and international Grid Initiatives. In particular, statements on the necessity for coordination bodies and for optimising operation and support of distributed computing services are acknowledged by at least four out of five respondents.

Expectations about cloud computing and other new resource delivery models

Expect significant contribution to progress from new computer resource delivery models — 79%

Expect adoption of new computer resource delivery models by a large share of researchers — 81%

0%   20%   40%   60%   80%   100%

Roughly 30% of the respondents also made policy recommendations. Most important among these are those addressing organizational or funding issues, which were suggested by more than 10% of the respondents. Examples are included in the table below.

Assessment of IGIs/NGIs

NGIs are necessary as / to	
the suitable structure to represent all the national DCI at international level	76%
ensure best adoption and compliance with middleware standards	69%
right body to optimise dissemination efforts and user support	73%
right body to optimise operation and support	78%
most cost effective coordination scheme at country level	79%
IGIs are necessary for / to	
anticipate the evolution of DCI technology	73%
guarantee the largest inter-operability of DCIs	86%
optimise worldwide dissemination efforts and user support	76%
standardise operation and support of DCI	87%
coordination of infrastructures spanning continents	89%

0% 10% 20% 30% 40% 50% 60% 70% 80% 90%

Policy requests - quotes from users

Category	Response examples	Category	Response examples
Access to resources	- Make it institutionally and ubiquitously available as if it were the telephone, mobile phone, electricity, or air we breathe. - Policy maker should push for a flexible and open GRID access to a variety of computational resources, both HPC and High Throughput oriented. - by providing tools allowing reallocation of resources for a given group of scientists on demand	Funding related	- 1) by rewarding and funding the development and evaluation of production-ready technology; 2) by providing stable funding for user support and training - By making clear decisions on sustained funding, not just funding projects. Basic for advancing einfrastructures is the long-term maintenance.
Organizational	- Support software applications design and provide career and career plans for whole generations of developers rather than living from hand to mouth on short term contracts well into their forties and fifties. - Provide clear national strategy around e-Infrastructure, outlining drivers and strongly connected research communities, and lead agencies and organisations; Facilitate the aggregation of research agendas towards developing and sustaining e-Infrastructure developments - A grid services brokerage company is required. Infrastructure use grants could be given.	Training-related	- Making the e-infrastructure familiar for more people, with workshops for the older and introducing or building e-infrastructure in public schools, for the children. Also teachers should enhance their knowledge to keep on with new technologies and teaching strategies. - In countries where the technology is not widespread, most of the effort should be placed in training people to use new scientific methodologies that can profit from the massive amounts of computing and storage available and that can be put together thanks to these e-Infrastructures.
Technical capabilities	- Focus on alternatives to „Grid", especially on web service standards. These have proved far more effective in promoting interoperability and integration of data-dependent services. - Creating standards and study previous cases such as the Internet evolution	Awareness raising	- There must be applications that create impact in the country's economic value, to make policy makers at the national level support and sustain the investment. In developing countries, immediate problems have priority. - Promoting through events and tutorials the use of grid, at least

| Ease of use | - By paying more attention to the needs of end users and less to the claims of those promoting technologies
- Improve the simplicity and accessibility of the user interface layer
- participation should be easier and encouraging | | once a year in all the involved countries.
- by showing good examples (pilot projects); by making it easy and relatively cheap to access the e-Infrastructure; by taking away the (emotional and political) barriers
- funding and articulation of a global vision explaining goals, plans and motivations |

>> The Empirical Picture: The Provider Perspective

Providers of e-Infrastructures also reported a number of inhibitors of effective use of e-Infrastructures. At an early development stage, cultural differences between developers and lead users occur. These are exacerbated when developers have little understanding of specialized user practices, when there are communication problems among e-Infrastructure collaborators or when divergent objectives are pursued. For example, developers may aim to work on cutting-edge technologies, in contrast to the basic and robust services that users seek. Other barriers were noted in reaching out to new users. There was also a negative attitude among some users toward computer-enhanced research environments, with a reluctance to spend the time and resources currently required to learn to use the new technology. However, it was apparent that our informants often lack detailed information about their users. Some infrastructures do not distinguish between individuals and organizations; many can only monitor access to their website, wiki or portal rather than actual use.

Strategies that Work

e-Infrastructures projects commonly accommodate cultural differences between developers and users by improving communication channels, such as through conducting routine meetings and telephone conferences. This helps in establishing a common ground. To enhance user adoption, a variety of strategies were employed - both passive strategies with limited user engagement and active ones that focus on ongoing interaction with users. All providers studied pursued user recruitment through direct dissemination of information and by giving presentations at conferences. Several projects have also ventured into more active recruitment, utilizing "engagement teams" to work with leading users in diverse communities, or "brokering" – the use of key individuals and relevant organizations. For example, the US TeraGrid has launched a program in which "campus champions" serve as institutional mediators for recruiting users and as local technical experts. The European D4Science and the US-based Open Science Grid utilize third-party organizations that offer e-Infrastructure technology to user communities. The advantage of the more active user recruitment strategy is that they build a communication channel between e-Infrastructure stakeholders, sensitizing developers to users' needs and helping adopters derive more benefit for their research.

Another class of strategies for enhancing adoption is the reduction in the cost of learning the new technology. Refining documentation, utilizing wikis and additional Web 2.0 mechanisms, and running training workshops is a passive approach common among the e-Infrastructures studied. Here, too, active cost reduction strategies appear more advantageous. Relying on brokers, some of the projects achieved good results by designing virtual environments that simulate the typical computational environment of users - for example through domain-specific portals. Another approach involves masking e-Infrastructure complexity from users, using specialized virtual technologies. This type of brokerage may offer considerable benefits in the long run.

>> Scenarios and Roadmap

Based on the empirical findings, the roadmap aims to inform research policymakers and e-Infrastructure developers about critical issues in e-Infrastructures for research in the European Research Area and beyond that must be addressed in the coming decade. The Roadmap proceeds as follows: It reviews how e-Infrastructures fit into recent changes in the relationship between research and society, and especially the changing scale and complexity of scientific and other research efforts. It highlights how it is important not to have a fixed or narrow conception of 'infrastructures', but nevertheless to identify them precisely and notice their protean nature. Next, the Roadmap reviews the relevant policymaking initiatives and the various reports and groups which have aimed to support the research policymaking process. Here it is noted that a number of groups have made contributions (for example, ESFRI, e-IRG), but there still exist important gaps and far more could be done.

The Roadmap then reviews some of the main elements of the eResearch2020 report on which it is based, including the case studies of e-Infrastructure providers and the survey of virtual research communities. From this report, a number of patterns can be elicited, including understanding that e-Infrastructures should not be regarded as uniformly 'top-down' efforts but also 'bottom-up' efforts, both of which may emerge within but also across disciplines and fields of research. This heterogeneity, and a balance of leading-edge and more well-established efforts, are highlighted at a number of points throughout this document as requiring a balanced approach in terms of support and planning. Further findings from the report include a selection of technical but mainly social bottlenecks to e-Infrastructures development, of which a current critical one is the sharing and re-use of data.

It is then detailed how e-Infrastructures will play a key role in industry, government, health, education and cultural heritage, which leads to an analysis of priorities for e-Infrastructures developments. These include management and governance, the latter a particular priority given the difficult multi-institutional nature of infrastructures. Further priorities include data and the need to engage with new technologies such as 'clouds'.

On this basis, the roadmap describes four scenarios: research revolutions, winners and losers, a many headed beast, and overtaken in the fast lane. These identify different outcomes depending on the level of e-Infrastructures uptake and whether this is across-the-board affecting many institutions - or encounters mixed fortunes in these, and the relevant risks and opportunities are identified in each case. Finally, the Roadmap concludes with a series of recommendations for action and recommends 14 steps to research policymakers, which address (to highlight just three) the need to ensure long-term planning, requirements for more extensive training, and the need for indicators of success.

>> The Future: Scenarios and Risks

„Research Revolution" represents the leitmotif scenario but further ones have also been sketched to develop possible futures.

Scenario 1 - Research Revolution

A future scenario that incorporates a best case ideal type, a „Research Revolution" resulting from e-Infrastructures, would be characterised by the following

- Large-scale collaboration, data- and tool- intensive
- The nature of research is fundamentally transformed and carried out in distributed mode
- Change takes place across all disciplines and cross-disciplinary fertilization
- Change takes place on all levels of research (infrastructures, applications, daily practices) and at all levels, including in schools

- Industry joins up with the research community and there are links to e-Government, e-Health and the public
- Public funding is complemented by private funding, an 'open science' ethos prevails

From a policy point of view it is clear that this scenario is likely to require the largest amount of funding and researcher effort. The benefits, for the research community and for society-at-large, are potentially enormous, but as with many innovations, it is possible that these benefits will only become realized after a considerable time. This 'lag' is the main risk of this scenario.

Another could be that despite good efforts, critical grand challenges to society (climate, energy, disease) that need to be addressed urgently will not be addressed quickly enough by an e-Infrastructure research revolution.

While this is the best possible scenario, three other scenarios can be sketched that involve a failure to reach one or more goals of the research revolution (the four scenarios are likely to be mixed in practice, but the analytical separation provides a way to think about different developments towards 2020). The difference between

the four scenarios can be mapped onto two dimensions: the vertical dimension is whether there is large or small uptake by virtual research communities, and the horizontal whether the impacts of e-Infrastructures are spread across all areas of technology and its effects on communities, or whether the effects are felt only in certain areas and not in others (or quite differently in different areas).

Scenario 2 - „Winners and Losers

- Some disciplines have strong uptake, succeed in creating strong communities, and move to new research questions
- Other disciplines have weak uptake, fall behind in creating collaborative communities, and retreat into disciplinary silos
- Some disciplines and transdisciplinary communities mature rapidly, others do not get beyond planning
- Some fields gain via data- and resource-sharing, others are unable to benefit
- Winners move forward and e-Research supports collaborati on and healthy competition in the field, losers are left behind

This scenario represents risks for certain research communities rather than others. The benefits for some fields or disciplines will be balanced against the losses for others, so that researchers and society-at-large must for example bear the cost of lacking an e-Infrastructure that would provide cultural heritage while having one for particle physics, or vice versa – with all that this entails for the research community and the public.

Scenario 3 - „A Many-Headed Beast"

Only certain fields develop e-Infrastructures - others concentrate on large facilities, still others focus on Web 2.0, e-Research is ignored in some areas – a plethora of directions

- Some areas duplicate efforts, in others there are no e-Research efforts or different directions
- A mixture of private and public funding, neither is provided across the board, and funding is concentrated in pockets
- There are enormous disparities between sciences, social sciences, and humanities in funding (with little for humanities, even though there is much potential for cross-pollination with cultural heritage, educational outreach, and public access)
- A mixture of strong and weak research identities, large geographical variation, and efforts are separated by technologies and possibilities for collaboration

Scenario 3 is based on the backdrop of a growing bottom-up Web 2.0 (or 3.0) tools and datasets. A recent study surveying e-Social Scientists found that many social scientists build their own tools and datasets, often in idiosyncratic ways, to meet their particular needs and because no other tools and datasets are available to meet these needs. With the growing popularity of Web 2.0 or Wiki-style forms of collaboration, this type of tool and data development has become widely accessible. And social
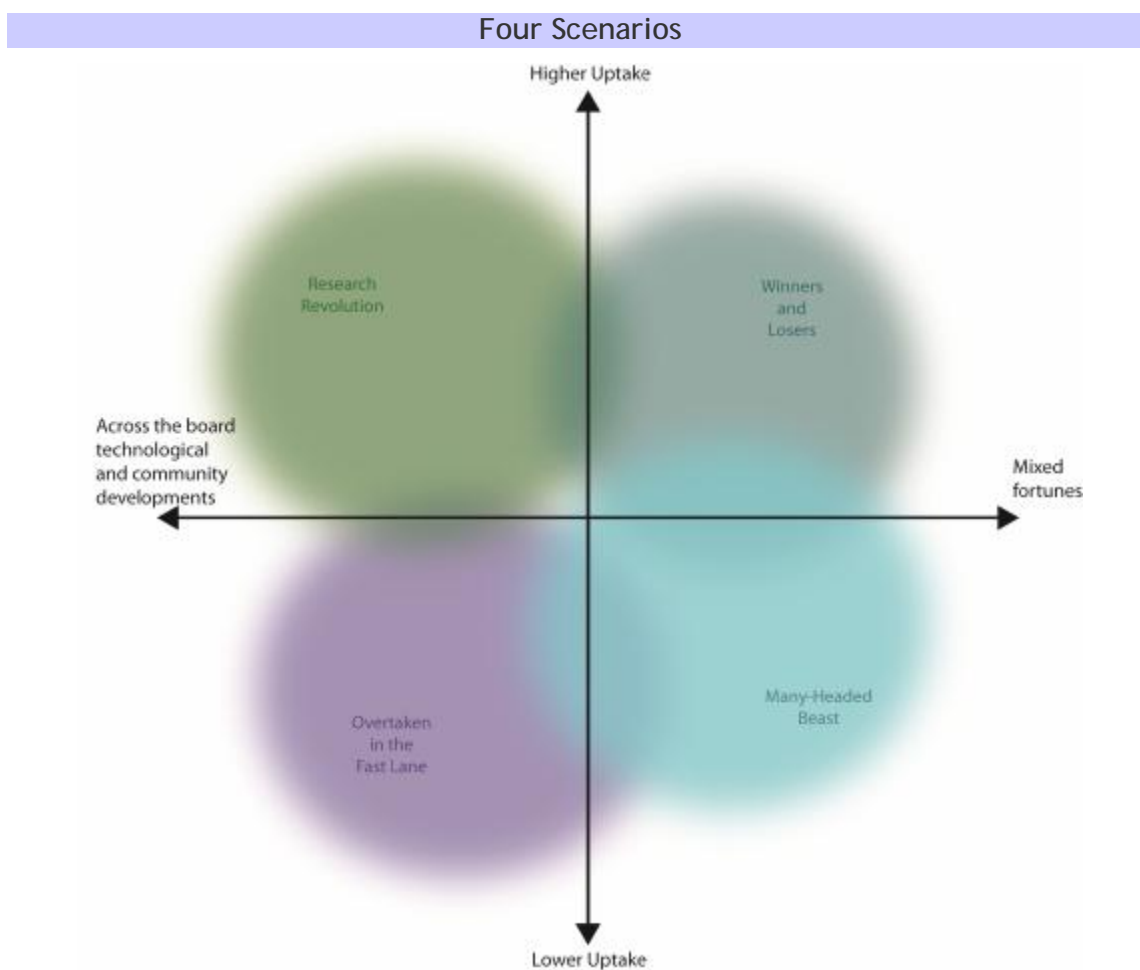
scientists are not the only ones engaging in this type of bottom-up activity. the bioinformatics and other communities are is also moving in this direction. Unless e-Infrastructures monitor, engage with, and either focus elsewhere or directly embrace these developments, this could lead to a scenario in which there is little uptake.

Scenario 3 suffers from a different main risk; namely, that the benefits of coordination and potential synergies between research communities are not realized. This could apply both to geographic spread and to spread within and between fields: some would be well-provided for (but without the possibility of linking to other e-Infrastructures since different technologies would not interoperate), others would be overprovided because of parallel efforts, and yet others would be left out altogether. One way to avoid this risk is to implement a policy whereby any funding allocated for infrastructure is granted on the condition that the e-Infrastructure must be open and must interoperate with other systems.

Scenario 4 - „European e-Infrastructures overtaken in the fast lane"

- EU e-Infrastructures are overtaken by developments in the US and Asia, where there is more uptake of newer technologies other than e-Infrastructures
- Technological and social developments (clouds become a commercial Google or Amazon service in the US, petabyte libraries on mobile phones become common in Asia) overtake Grids, supercomputing and other research infrastructures – enabling computing-based research to move onto different terrain
- Data storage and compute resources become a commodity outside of research, so that shared public e-Infrastructures have little uptake outside universities
- Within research, e-Infrastructure investment atrophies
- Research quality and competitiveness in the EU suffers decline compared to Asian and US research

Scenario 4 means that the research initiative passes to non-EU researchers and the private sector. The commercial sector, and especially software providers, play an important role in this case in future scientific developments. It is important to recognize that these are parallel efforts, and that these commercial efforts will both compete (for example, in developing software for the annotation of scientific texts) and collaborate. This scenario also involves the role of 'clouds' and data. The pay-offs from e-Infrastructure investment are not realized due to other provision channels, and the status of European research declines in relation to that of other parts of the world.

Four Scenarios

Higher Uptake

Research
Revolution

Winners
and
Losers

Across the board
technological
and community
developments

Mixed
fortunes

Overtaken
in the
Fast Lane

Many-Headed
Beast

Lower Uptake

>> A Roadmap to Research Revolution

A key challenge of e-Infrastructure policy is to recognize diversity and commonality in issues across disciplines. The various social, institutional and technical challenges to the formation of effective e-Infrastructure collaborations do not pose uniformly serious obstacles or impinge with equal severity upon all branches of scientific inquiry. Similarly, the potential transformative impacts of enhanced e-Infrastructures are not likely to be felt equally across all the domain sciences and emerging interdisciplinary fields.

Gaining a better sense of policy priorities will enhance the support of global research communities as e-Infrastructures become more complex and at the same time critical to the quality of research outputs as well as to productivity. The following recommendations for action by the European Commission and other research policy makers are to ensure arrival at Scenario 1 rather than Scenarios 2, 3, or 4.

eResearch2020 Recommendations to Clear the Way Towards the Research Revolution Scenario:

1. European and other researchers increasingly depend on the most technically and socially advanced e-Infrastructures to meet the world's most urgent research challenges. e-Infrastructures development underpins the future of meeting these challenges and should remain a key priority for policymakers.

2. Sustainability should be considered in a much longer-term perspective. Resources sustained at the European, national or other level must be committed for extended (10+ years) periods so this commitment provides a reliable and well-integrated platform for the research community and beyond.

3. The uncertainties around funding are the single-largest perceived barrier among providers, virtual research communities, and the yet-to-be-engaged. Clearer plans and funding agendas could overcome these uncertainties.

4. While data is not scarce any more, the key challenge has moved on to the coordination, proper safeguarding, sharing and re-use of data, also beyond its initial purposes. Mandating clear policies to share software and make data interoperable are essential.

5. There are currently few rewards for researchers both inside communities and among providers for their contributions to e-Infrastructures development, or for sharing data and tools. Reward mechanisms need to be promoted that recognize and reward researchers to do this.

6. „Openness" has been a much vaunted principle in e-Infrastructures development, but while open source software and open publishing can already show successes, much more by way of coordination is needed to apply openness to standards and interoperability in systems and collaboration platforms.

7. Governance and metagovernance (governance which coordinates the governance of individual efforts) strategies are still emerging in many ad hoc forms. Although ERICs are emerging as a possible single legal mechanism for the future, there is still uncertainty among the e-Infrastructures communities. Policy can be put in place to overcome this uncertainty.

8. Education and training efforts for e-Infrastructures lag behind e-Infrastructures development, but offer an excellent route for much more widespread engagement with the novel research possibilities and should thus be among the highest priorities in future planning and funding.

9. Many opportunities for shared best practices and for sharing resources between fields and sub-fields are currently unexploited and could be fostered by more funding that favours cross-disciplinary teams and efforts.

10. A fair share of future efforts is also to be dedicated to actions with a higher risk of failure (subject to constant monitoring and revision) which are hoped to generate completely novel applications to problems in which distributed computing and other e-Infrastructures have not yet been applied.

11. Mandating standards both in software and in the interlinking of metadata and data, although requiring a balance with flexibility, remain a high priority

12. Indicators of success and impact and quality are required in view of the need for coordination and resource planning. High priority should be given to providing resources for projects which undertake such measurement or to research from outside the e-Infrastructures to enable monitoring and comparison.

13. Existing barriers to participation by industrial research partner participants need to be removed so that potential benefits materialise more easily for both larger firms with sizeable R&D organizations and SMEs.

14. Research into the bottlenecks, effectiveness, and future potential of e-Infrastructures will be imperative. e-Infrastructures - as a relatively novel, still protean, and absolutely vital platform for research in the ERA and beyond - are a still largely unexplored territory in terms of their impact dynamic. Especially in relation to Recommendation 12, such research will have enormous pay-offs.

Contents

Tables and figures

PART 1 – THE EMPIRICAL PICTURE

1   Introduction and objectives

e-Infrastructure plays a pivotal role in the research landscape of today and fosters the creation of global virtual research communities. This study researches a variety of types of e-infrastructure, their state of development and their role in supporting productive research in Europe and beyond.

The field of e-Infrastructure is very heterogeneous. It ranges from the physical supply of research networks (a prominent example being the operating and development of the backbone as supplied by GEANT) to providing access to data for virtual research communities in single fields. e-Infrastructures include organisations and services as diverse as national and international multi-purpose grids, supercomputer infrastructure, data grids and repositories, tools for visualization, simulation, data management, storage, analysis and collection, tools for support with regard to methods or analysis as well as remote access to research instruments and very large research facilities.

The diversity of this field makes it difficult to compare any two e-Infrastructures in an evaluative sense. It should therefore in the first instance be our aim to understand each infrastructure: what they do, for whom, and in what environment. Once this goal has been achieved, it is our aim to point to similarities and dissimilarities in their approach and outcome.

As our research aims to address both e-Infrastructure providers and their respective virtual research communities, questions addressed in this project include:

- What kinds of e-Infrastructures are successful and less successful in anticipating and catering to the needs of virtual research communities?
- How well do e-Infrastructure providers define, consult, plan for, engage with and overcome bottlenecks in scaling up to match growth in their user community?
- How do e-Infrastructures coordinate with other complementary tools and resources to maintain a unique profile while also integrating with other synergetic efforts?
- How do e-Infrastructures implement strategies to ensure that they make an essential contribution to their community of beneficiaries?
- What kinds of instruments do e-Infrastructures need in order to gauge and adjust their provisions on an ongoing basis to cater to their communities?
- How do e-infrastructures link researchers globally and reduce the effect of geographical distance on research collaboration and other cooperation in academia, i.e. to what extent they contribute to the establishment of global virtual research communities?
- What organizational structures and coordination mechanisms of e-infrastructures exist? Who are the key players in their interaction with researcher communities, relevant regulatory and policy bodies and support from funding and other external bodies?
- How do researchers use e-infrastructures? What are the benefits and costs for global virtual research communities, where do they accrue and to what extent do they influence effective adoption and productive use?

The study's analytical objectives are reached through extensive new empirical work. The analytical approach followed in this enterprise is two-fold. Firstly, an analysis of each infrastructure was carried out via explorative semi-structured interviews with representatives of each provider. This was then followed by a standardized survey of users and developers of each of the e-Infrastructures.

This research and the resulting understanding of the e-Infrastructures investigated is presented in this document. As the study progresses, it will be used to formulate recommendations to the Commission - the roadmap for action – on how e-infrastructure development can best be promoted through EU policy. Furthermore, the study will provide a first contribution to policy implementation by raising awareness of recommended actions, through the construction and validation of the roadmap in collaboration with stakeholders.

The roadmap will highlight successful patterns of e-Infrastructures in productive research communities, together with scenarios of future development to help coordinate future policy and public sector action.

2 Literature review: e-Infrastructure and global virtual research communities[1]

2.1 Introduction: Scope of Relevant Literature

Past research and policy recommendations on e-Infrastructures can be subdivided under a number of categories:

- Economics, particularly economics of innovation
- Social analysis, including legal and ethical issues
- Historical, focusing on analogies with historical infrastructures
- Transforming scholarship, especially practices of data sharing and the importance of research instruments
- Measurement of scholarly production, including how it is affected by open access
- Policy recommendations and strategic visions

Since these are often intermingled in practice, it will only be possible to summarize some of the main findings and areas that have been singled out as highly significant. It can be pointed out immediately, by way of anticipating a conclusion, that the various categories and approaches do not overlap very much. In other words, previous research does not offer a synthesis. This will be a major task of this project's roadmap.

This review will consist of two parts: the first provides an overview of the relevant literature. In the second part, the review will focus on a few key topics that have been identified and addressed in the literature:

1. Openness
2. The Analogy with Historical 'Infrastructures'
3. The Heterogeneity of e-Infrastructures
4. e-Infrastructures and Public Perceptions of Research
5. The UK Experience: lessons from a mature e-Research programme
6. Cloud Computing

2.2 Part 1: Overview of Literature

An overview of past research that is relevant to e-Infrastructures includes:

1. The changing relationship between science and society. This has been addressed in the literature under the labels of "Post-Normal Science" (Funtowicz & Ravetz, 1993), "Mode 2" (Gibbons et al., 1994), or the "Triple Helix" (Etzkowitz & Leydesdorff, 2000; Leydesdorff & Etzkowitz, 1997). Key issues here include that researchers may have to have a more responsive way to address social problems which, moreover, are increasingly complex and face the constraint of diminishing research funding.
2. The shifting landscape of research towards new regions of the world (such as China) and the globalization of research (Leydesdorff & Wagner, 2007; Leydesdorff & Zhou, 2005).

---

[1] The principal authors of this section are: Kathryn Eccles, Eric Meyer and Ralph Schroeder

3. The increasing need for multi-institutional, large teams, and online collaboration (Wuchty et al.2007).
4. Policy documents outlining visions and needs related to e-Infrastructures. We will not try to summarize or reproduce key points here (these can be found in the documents listed, and they will be discussed in the roadmap), but to present only salient highlights.

This review will not go into detail of the background of changes in the research system in depth, as it can be assumed that it will be familiar - except insofar as it bears on the question of e-Infrastructures.

The first trend to note is simply the rising importance of team-based research and collaboration. Research is increasingly taking place in larger teams, and team efforts have a greater impact compared with individual efforts, or those of smaller groups, when measured by citations. This applies not just to the natural sciences, but also to social sciences and only to a somewhat lesser extent, to the humanities (Wuchty, Jones and Uzzi, 2007). Several reports have shown that the importance of scientific collaboration has grown in the last 25 years (European Commission, 2003; Narin, Stevens, & Whitlow, 1991; National Science Board, 2004). Growth rates vary by academic domain, but the overall trend is ubiquitous and visible for local as well as international collaborations, disciplinary as well as interdisciplinary collaborations, and those in the public research sector as well as university-industry collaboration.

These trends have created pressure for significant investment in technologies that support distributed research and collaboration. e-Infrastructures are being rapidly developed and deployed worldwide and across the European Research Area (ERA) to support team-based research. The European Commission is the main driver of these infrastructures in Europe through the Framework Programmes. In FP6 and FP7, this included supporting large networking and Grid infrastructures such as GEANT and EGEE, as well as domain specific ones such as BioinfoGRID, and projects developing high-performance computing such as DEISA. ESFRI (The European Strategy Forum on Research Infrastructures), EGEE II, and many others are carrying this work forwards.

There are now a number of bodies that address policy-related issues relevant to e-Infrastructures, such as the e-Infrastructures Reflection Group (e-IRG). There are also some initiatives that are just emerging or being planned, such as the design study for European Grids (EGI). Finally, there are a number of initiatives funded by national research funding bodies and initiatives beyond the ERA such as the UK £250 million (over 5 years) e-Science programme that started in 2000 (Hey & Trefethen, 2003), and the German D-Grid initiative starting in 2005 (see Schroeder, den Besten, & Fry, 2007, for a comparison of the initiatives).

Outside of Europe the initiatives of the US National Science Foundation through its Office of Cyberinfrastructure should be mentioned; these gained considerable momentum after the 'Atkins Report' was published in 2003 (Atkins et al., 2003). Significant funding in the US is directed to cyberinfrastructure development and deployment. Outside of the US and Europe, smaller, but by no means negligible efforts are being undertaken in countries such as China, Japan, Australia, New Zealand, and Canada. Some European efforts are already being coordinated with these (for example, EUChinaGRID), and such links will need to be increasingly coordinated in the future.

These investments in e-Infrastructures have produced without doubt some important advances towards realising Vannevar Bush's 1945 vision of a "memex": a "device in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory." (Bush, 1996, p. 43 [reprint]). However, there are still many issues that need to be resolved to make globally distributed, collaborative and multidisciplinary science (science will be used in what follows to encompass humanities) a reality and to ensure that the consequences are contributing to scientific progress. The key difficulties here, as argued in a number of seminal publications, are social as much as they are technical: for example, the

collaboration that is typical of e-Infrastructures or e-Research, especially across institutional boundaries, remains fraught with difficulties, perhaps even more so than collaboration across disciplinary boundaries (Cummings & Kiesler, 2005; David & Spence, 2003). Increasingly global collaboration requires further institutional underpinnings (Drori, Meyer, Ramirez and Shofer, 2003).

Moreover, high expectations on tools supporting distributed collaboration have often been disappointed and as a result, domain scientists have become critical of the promises of computer scientists and developers. For instance, case studies and descriptions of different tools for on-line meetings and remote collaboration suggest that there are still several technical shortcomings relating to hardware and software, and that the technical staff and users' proficiency with information and communication technologies is somewhat limited (Finholt, Rocco, Bree, Jain, & Herbsleb, 1998; Mark, Grudin, & Poltrock, 1999; Olson & Olson, 2002; Sanderson, 1996). The challenge is therefore not only to remove the technical deficiencies and constraints, but also to convince and prove to scientists that adopting them, investing the time to learn how to use them, and even changing established work routines to integrate them, is beneficial and leads to better science in the end.

e-Infrastructure has thus far emphasized technology – shared and distributed computing tools and resources – yet it is clear that the social pressures on science for more distributed collaboration and access to shared and distributed tools resources will increasingly put organizational, rather than technical issues at the top of the agenda. As e-Infrastructures become more widespread, it will be important to assess critically to what extent they are living up to the challenge of supporting collaboration and distributed research, and if they have been able to address the problems of distributed collaboration not just technically - but also socially. Social science can help address challenges common to other disciplines, contributing an understanding of how humans adopt and use e-Infrastructure, identifying the organizational and market mechanisms that create the best incentive structure for the development of e-Infrastructure, and the impact of e-Infrastructure on the economic, social and behavioural framework of society.

In different disciplines, e-Infrastructures are at different stages of maturity in terms of technologies and organization-building. The humanities have adopted e-research in different ways depending on the discipline (Nentwich, 2003). In the US, for example, it was found that 6% of humanities research was based on more complex forms of networked research and digital tools (Frischer et al. 2006, p. 4). However, the policy group which undertook this study recommended that construction of e-Infrastructures should be based on current practices that operate within existing research traditions, rather than training researchers to become 'new users' of tools developed through means and expertise that is external to existing practices.

Finally, the design and use of advanced Internet and Grid technologies in the social, natural and computer sciences as well as the humanities are reconfiguring not only how researchers obtain and provide data resources and other forms of evidence, but also what they and the public can access and know; not only how they collaborate, but with whom they collaborate; not only what computer-based services they use, but also from whom they obtain services (Gläser, 2003; Hilgartner, 1995). Scientific networks and communities need time to reshape and adjust to these broader social changes.

The challenges to developing and making use of e-Research Infrastructures are thus as much social, legal, political and economic as they are technical. Some of these non-technical issues have already been partially addressed for some aspects of e-Infrastructures and particular e-Research projects. For example, some of the European activities worth mentioning are:

- Intellectual property issues: ESFRI has addressed this with workshops and including experts within each project, and Burk (2007) has clarified some of the issues
- Data sharing: Axelsson and Schroeder (2009) have identified the issues for Sweden, a country with uniquely favourable conditions for data sharing

- Uptake among user communities: the AVROSS project has surveyed the catalysts and barriers for social scientists and humanities researchers to use e-infrastructures (AVROSS 2007a, 2007b).

- Business models, including the sustainability of e-Infrastructures, are addressed, for example, in BEINGRID (http://www.beingrid.com/).

- Another issue that needs to be addressed are the effects of requirements for depositing and sharing data on research. This is also a longer-term issue which involves the IP of databases (Wouters and Schröder, 2002, 2003; Borgman 2007). Data and resource sharing are linked to wider issues. There is a link between trust and public understanding (or perception of) research and how researchers 'trust' in their data and make it more widely accessible.

Activities outside of Europe and in particular in the US also constitute part of the knowledge base on e-infrastructures and need to be monitored and evaluated in regard to their European applicability. The Atkins Report (Atkins et al., 2003) is a milestone in the advancement of Cyberinfrastructure in the US and the Berman & Brady Report provides an extension to the social sciences (Berman & Brady, 2005). The Science of Collaboratories (SOC) project at the University of Michigan has investigated the technical and behavioural principles of distributed collaboration (http://www.scienceofcollaboratories.org/). Nevertheless, since e-Infrastructures span many academic disciplines and institutions across different national research policy settings, it has been difficult to share information across initiatives and thus develop best practice.

2.3 Part 2: Key Topics

2.3.1 Openness

Openness can mean, among other things, open access, open source, open science and open data (David, den Besten, Schroeder 2006). Here, for the sake of brevity, these will be bundled together. In all cases, there have been initiatives by research funding bodies, government ICT policymakers, NGOs and others which favour (and in some cases mandate) 'open science' policies. At the policy level, while it is too simple to say that this will continue to be a struggle between the public and NGO sectors (funding bodies mandating open access publication, or the recommendations of the 'Science Commons' NGO, which is part of 'Creative Commons') versus the private sector (commercial academic publishing, pharmaceutical companies), it is clear that 'opening' research will continue to be an uphill battle (Schroeder 2007).

What is interesting to note is that 'on the ground', researchers' practices are still quite diverse: while they uniformly endorse the principles of open science, their practices only partly reflect this (Fry, den Besten and Schroeder 2009). For example, the way they publish their papers online or use common project repositories or make data available within and beyond their projects is very inconsistent and varied. This is partly due to researchers' lack of awareness of the rules and laws that apply, partly the effort that is required, partly that established routines are difficult to change, and in some cases that trust is felt to be a precondition for openness and this is perceived to exist only among project members. To summarize briefly: despite a continuing push against 'closedness' at the policy and macro-levels, there will continue to be a push 'downwards' and into practices, and the result of these tensions will be mixed.

2.3.2 The Analogy with Historical 'Infrastructures'

The notion of 'infrastructures' comes from the support structures that were created for societies mainly in the late 19th and early 20th centuries, such as transport, communications, and power. These 'systems' were created to support society-at-large with essential services. Since they required extensive networks of technologies and organizations, they have also been analysed under the label of 'large technological systems' (Hughes 1987). Edwards et al. (2007, see also Jackson et al. 2007) have provided an extensive analysis of the lessons that historical 'infrastructures' provide for contemporary 'e-Infrastructures' (or 'cyberinfrastructures', since they are writing in the American context), the main one being that infrastructures require planning for very long periods (decades and even centuries) and that early 'lock-ins' to particular choices can have profound and lasting effects on the nature of infrastructures.

These lessons are important, but they also overlook a number of points:

- That the very label of infrastructures often does not fit well with 'e-Infrastructures' (Eccles et al. 2009). For example, many e-Infrastructures are in reality projects and other types of more and less temporary and differently structured organizations. Furthermore, they often provide services not for the whole of society or the whole of the scholarly community (as infrastructures do), but rather for specialized niches.

- That infrastructures and large technological systems have a 'momentum' (Hughes 1994) of their own, which leads them to expand become solidified – except insofar as their monopoly is curtailed by competing systems.

- Edwards et al. (2007) stress the political nature of infrastructures, which is appropriate in so far as these infrastructures seek allies, consolidate their power, and institutionalize their own preferred design choices. However, this view presents an incomplete picture: for the most part, the main concern of infrastructure builders is simply to optimize the system, maximize its usefulness, and extend it from its potential to the largest possible constituency. This latter feature is also common in historical infrastructures.

- Following on from the previous point, a unique feature of infrastructures is that trade-offs must be made: for example, between scaling up to the maximum community of users and providing them all with the features that they want; or between providing a simple tools with few system requirements as against being able to interoperate with other systems (for example, importing data in certain formats) which involves add-ons and additional programming requirements.

2.3.3 The Heterogeneity of e-Infrastructures

It is worth reiterating that e-Infrastructures and their communities are not simply a single phenomenon. There are various types (Eccles et al.2009), for example:

- Some are aimed at the academic community at large (DRIVER); others are aimed at specialized sub-communities within disciplines (SwissBioGrid)

- Some provide services (DRIVER) or resources (SND), while others consist of research technologies (ATLAS)

- In terms of models of sustainability, some are stably embedded within larger established institutions, some are projects without a future beyond the end of project funding, others are international networks that federate the contributions of members, and there are also purely volunteer efforts (even if they may be linked to 'umbrella' organizations).

The user communities can also be differentiated. Although it is envisioned, as per the definition of an infrastructure, that all members of a particular community will need to make use of an infrastructure, in practice, this is not so. It may be argued that this limitation is

provisional – that ultimately all scholars will come to rely on these infrastructural tools – but this argument only needs to be stated clearly to see that it is misleading. Only in certain areas of research will it be universally necessary to use certain infrastructural tools. We can therefore divide the communities into those in which there are only early adopters, as opposed to those which are close to maximizing the relation between potential and actual adopters.

2.3.4 e-Infrastructures and Public Perceptions of Research

e-Infrastructures can affect the public's perception of research. For example, pollution monitoring via remote sensors might engage the public's interest in e-Research, but it might also make the public more sceptical towards researchers. Similarly, census data might be linked to patterns of consumer behaviour by means of shop loyalty cards, blurring the boundaries between 'official' and 'commercial' data, with social science researchers using e-Infrastructures increasingly wanting to take advantage of linking multiple sources of data. However, such a blurring of boundaries might make the public wary of social science researchers.

Three further examples can be given which highlight the very wide range of these impacts: the contribution of amateurs to astronomical data in the International Virtual Observatory Alliance, in which the EU e-Infrastructures play a part, will affect researchers and how they are perceived. Another example of engaging the public is www.climateprediction.net, which allows the public to contribute the computing power of personal PCs for climate modelling. These initiatives link e-Infrastructures to the public, making the public more aware of e-Infrastructures, and such projects are likely to become more common. In so doing, it will affect the public understanding of this part of science and research. A final example is medical data sharing. Here, there is a link between trust and the public understanding of science (or research), and how researchers 'trust' their data and yet wish to make it more widely accessible.

The public understanding of e-Infrastructures fits largely into two areas: the public awareness of the risks and benefits of the research (bio banks, pollution sensors, shared video recordings for social scientific analysis of sensitive settings), and the outreach efforts of e-Infrastructure projects (schools projects, museums, libraries). Both of these areas will, over time, change the image of research in society. The main reason for elaborating on this issue is that – apart from the voluminous literature on the public understanding of science (see the review in Bauer and Gaskell, 2002), there is as yet no literature which specifically addresses this topic for e-Infrastructures. It is therefore highlighted here as a major gap in the literature.

2.3.5 The UK Experience: lessons from a matured e-Research programme

The UK has come further in reviewing its e-Research programme than any other country or region. In a recent paper, the key players in this programme laid out a vision of the future based on the programme so far (e-Science Directors' Forum Strategy Working Group, 2009). This can be discussed in some depth since it is the first policy document with a strategy which: a) goes beyond the visionary statements that characterized the early e-Infrastructure documents, b) represents one of the most mature programmes (apart from the US, the most well-established e-Infrastructure programme), and c) develops lessons that are supposed to go beyond the end of a dedicated e-Infrastructure programme.

Several points are noteworthy. The UK, after investing more than £250 million during the period 2001-2006, is now looking to ensure that the gains that have been made with this investment are not squandered. To do this, the document makes a number of

recommendations. However, the main avenue by which this will happen in practice is a coordinated programme across UK research funding bodies called the 'Digital Economy', which is much wider than 'e-Science' and encompasses the myriad ways in which digital technologies can benefit the UK economy. Second, the document also points out the 'risk from inaction': namely, 'loss of competitive position', 'poor return on investment [from the e-Science programme] as opportunities for sharing are lost and as there is duplication and excessive fragmentation in communities, processes and provision', 'lack of dissemination about the approaches used by researchers as users would not reach a critical mass', and 'loss of international influence' (p.32).

This summary of the risks could equally apply to the EU e-Infrastructure programme in a few years' time. The document also notes that 'it is essential to appreciate that infrastructures, including e-Infrastructures for research, are not built "top-down" to the dictates of a master plan but grow from the "bottom-up" through the efforts of a wide range of players and stakeholders'(p.36). Finally, the document stresses that increasing 'ease of use' will be necessary in e-Research: 'as researchers gain experience of well-supported Web services, such as Google, Wikipedia, Flickr and Facebook, their expectations for ease of use and interfaces will rise'(p.40). In short, the document makes clear that e-Research and e-Infrastructures are moving into a new phase: from a phase of developing tools to a phase of consolidating gains, moving e-Infrastructures out into the wider world, and integrating them into the practices of future generations.

Again, this is worth highlighting because the EU's efforts will also be going in this direction. However, it is important to note the subtext of the UK strategy document. Funding for a separate programme is coming to an end, and the paper can be interpreted as a means of seeking further funding, with the proviso that funding should shift towards enabling bottom-up innovation and great user-friendliness. Again, reading between the lines, the message is that the programme is not as successful as it might have been because it did not concentrate enough at an early stage on users and on ensuring widespread adoption and integration of e-Infrastructures into the practices of researchers.

2.3.6 Cloud Computing

The e-IRG White Paper (2008) discusses cloud computing as part of the future. It argues that a 'mixture' of grids and clouds combining the 'best of these technologies' will be optimal (2008, p.10). Although the White Paper mentions a number of well-known limitations of cloud computing (sensitive data, transferability of data across clouds, and the like), it overlooks others. Cloud computing only provides solutions in certain cases (storing data and other materials in clouds). For many others, such as remote instrumentation, web-based research, sensors, data storage which requires bespoke solutions, and many others, clouds will not be an option. It is difficult to predict how large or small this subset is. The cloud option also presupposes that material stored or held in the cloud can be shifted from one cloud to another (without a lock-in to a particular provider, for example, or to certain formats), and that some cloud will be available indefinitely for this purpose. However, whether this is a reasonable assumption is open to question.

Put differently, the problem of 'where to park your data' is only a small part of what academic researchers have to address in relation to e-Infrastructures. In this respect, if researchers are able to make contracts to move their data and other materials into clouds, the main concern for them (as opposed to, say, banks who can move from one contract to store their customer databases to another) will not be the commercial storage capacity of clouds, but rather whether long-lasting shared resources which can be modified and shared across many sites will be developed.

3 Analytical and empirical approach[2]

3.1 The e-Infrastructures and virtual communities sample

For the different levels of involvement, services offered and developed, and organizational objectives, we distinguish between three interrelated levels of e-Infrastructure virtual communities.

- Providers – including distributed organizations that offer e-Infrastructure to virtual user communities. Among the services offered are dedicated high-bandwidth networks, supercomputing and Grid computing facilities, including data Grids, community portals, training and technical support.[3]

- User communities – virtual communities that utilize and further develop e-Infrastructure applications and instruments that are specific to their domain. We analyze communities from diverse disciplines and fields, including life sciences, hard sciences, social sciences and the humanities.

- Standards – this involves activities that may be carried out by dedicated organizations, or through the activities of e-Infrastructure providers and user communities. For example, e-Infrastructure providers may work to interoperate various middleware packages, and user communities often strive to integrate data and instruments. Standardization thus enables different e-Infrastructures to work with one another, thus enhancing its reach and adoption potential.

Figure 3-1: Layers of e-Infrastructure
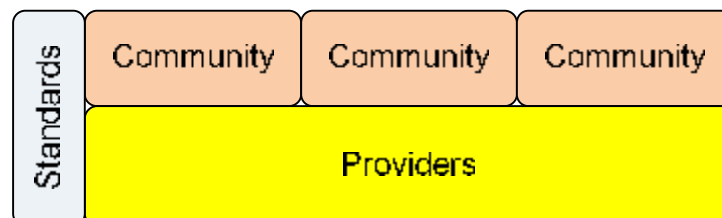
We investigate e-Infrastructure development, utilization and related activities at the three layers, as well as their interaction, by relying upon a comparative design. Three criteria guide our comparative case selection:

e-Infrastructure layer
As specified above we analyze cases from different layers of e-Infrastructure. For instance, we compare providers that specialize on Grid computing (OSG, EGEE) with those more focused on supercomputing services (TeraGrid, DEISA2), network providers (GEANT), as well as data providers (Swedish National Data Service, Driver). At the community layer, we compare communities across different fields of science. Accordingly, we study communities from the life sciences (MediGrid, Swiss BioGrid), the hard sciences (US-NVO), social sciences and humanities (DARIAH), as well as more eclectic communities that transcend traditional academic disciplines (Driver, CineGrid).. As standardization cuts across the provider and user community layers, we inquire about efforts to interoperate and rely on standards in each of

---

[2] The principal authors of this section are: Franz Barjak, Tobias Hüsing, Zack Kertcher, Simon Robinson and Ralph Schroeder
[3] While often operating as e-Infrastructure user communities, specializing in engineering and computer science, as other analysts have (e.g. Avery 2007), we distinguish e-Infrastructure providers from user communities, since their goal is the provision of services to these communities.

the studied cases.   We also include in our study a standardization organization: the Open Grid Forum (OGF).

Geographical range
e-Infrastructure has the potential to scale globally. However, there is considerable variation in the current geographical range of providers and user communities, some range across multiple continents, some cater to regional populations, while others concentrate their activities on a specific country.

Maturity
For our research objectives, as specified in the tender, the study focuses on cases beyond their initial phases, those that have reached production and offer substantial services to virtual research communities. Less mature projects do not enable a rigorous comparative analysis, as they are still in formation, some without actual users or developed application, still deliberating design approaches. While sufficiently mature for inclusion, the cases we chose to investigate are at different stages of "maturity." Some cases have already developed most of their technologies and have likely reached their peak(OGF) , other projects have offered tools and engaged users, but expect to considerably expand their repertoire and user community (US-NVO, DEISA2, OSG, TeraGrid), while others are at a much more formative stage (DARIAH, CLARIN).

	e-Infrastructure	ESFRI category
Providers	DEISA	e-Infrastructure
	EELA-2	e-Infrastructure
	EGEE	e-Infrastructure
	GÉANT	e-Infrastructure
	OSG	e-Infrastructure
	Teragrid	e-Infrastructure
	Swedish National Data Service	Social Sciences and Humanities (Biological and Medical Sciences too)
User communities	C3-Grid	Environmental Sciences
	CineGrid	e-Infrastructure
	CLARIN	Social Sciences and Humanities
	D4science	Environmental Sciences
	DARIAH	Social Sciences and Humanities
	DRIVER	e-Infrastructure
	ETSF	Materials and Analytical Facilities
	MediGrid	Biological and Medical Sciences
	NVO	Physical Sciences and Engineering
	Swiss BioGrid	Biological and Medical Sciences
Standards	OGF – Open Grid Forum	e-Infrastructure

Table 3-1: e-Infrastructure sample

3.2   Surveys of e-Infrastructures and research communities

e-Infrastructure providers have been addressed as those responsible for the characteristics of the technologies that undergird e-Infrastructure and the research communities using them. e-Infrastructure service providers are well positioned to help evaluate usage scenarios of various research communities, as well as to provide a coherent account of some of the challenges that have arisen over time.

However, service providers often do not have a detailed insight into the extent of collaborative research activity or into many aspects of research community behaviour relevant to this study. Also, it would be a mistake to neglect the possible conflict of interest there may be in some cases between an honest assessment of the history and current situation

of an e-Infrastructure, and the promotion of the economic success of the provider organisation.

Therefore, individual researchers and research communities have been addressed to provide essential insight into the research process using e-Infrastructure. The study therefore pursues two approaches: surveys of both providers and the research communities they serve.

3.2.1 e-Infrastructure Survey

This extended exercise in data collection provides quantitative and qualitative data on 18 e-infrastructures through a range of different empirical methods. The design is envisioned as rather strict in regard to the results that need to be produced, but open in regard to the methods employed in the field phase.

As a first step, common templates have been developed requiring compilation of detailed quantitative and qualitative information on the infrastructure at hand.

As preparation for interviews, an analysis of documents was conducted. Organizational documents and published materials are a chief source to examine the operation logic and structure of an institution. This step also served to identify interviewees, and guide the researchers on potentially unexplored aspects of e-infrastructure provision. Documents included designated web sites for the selected infrastructures, publications and presentations. From this initial work an Interview Guide was developed, which can found in the annex.

Using semi-structured interviews enabled us to gain much insight from a diverse set of providers, while maintaining analytic consistency across the cases. Since providers are often widely geographically distributed and due to time constraints, we relied primarily on telephone interviews, but in some cases face-to-face interviews were held. Interviewees were selected from the senior management and/or technical manager levels of e-infrastructures. Up to 7 persons per e-infrastructure were interviewed depending on the functional diversity within the organization.

3.2.2 Research Communities Survey

The Research Communities Survey adds data from the perspective of users who are engaged in specific communities and are using an e-Infrastructure in the sample. These data cover their use of e-infrastructures, their roles in the communities, the degree of collaboration in their research and their research output. It sheds light on how e-infrastructures contribute to the creation, growth and success of successful, coherent research communities.

Because of the diversity of e-Infrastructures covered and the different types of users involved, it was decided to distinguish three types of respondents:

- Research users, i.e. researchers utilizing the e-Infrastructure instrumental to their production of research outputs
- Other users, i.e. any other, non-research users utilizing the e-Infrastructure instrumental to their professional work
- Developers, i.e. those users closely involved with their e-Infrastructure, "working in the engine room", operating or optimizing the infrastructure.

Data gathered from each of these groups includes:

Background data

- Academic background, affiliation, working time spent on e-Infrastructure, level of experience

Motivation

- The catalysts and barriers in adopting e-infrastructures from technical, cultural, governance, and financial perspectives
- Motivation for the establishment of a research collaboration or for joining a research community, main users, geographic span, institutions, countries, growth, budget and funding sources

Tools

- The types of functionality of the e-Infrastructure used, applications (e.g. data stores, analysis and collaborative tools)
- Previous knowledge of using these tools and the time needed to make productive use of tools;
- Challenges and expectations

Impact

- Consequences for research approaches, work routines, time allocation, and other aspects of the research process
- Outcomes of the use of e-infrastructures at different levels in the research process, such as new data, new methods for analysing the data, new collaborations, publications

NB: The complete questionnaire can be found in the annex.

The Research Community Survey was designed as an online survey for individual e-Infrastructure users and research community members.

In most cases, the link (URL) to the survey was communicated by the e-Infrastructures to their user community using appropriate means, such as e-mail newsletters and newsgroups.

The survey produced the requested data for exploring the contribution of e-infrastructures to the creation of global research communities. The design can also be considered the first step towards the monitoring of e-infrastructures and research communities for which the necessary instruments have been designed and implemented for the first time.

4   Cases of e-infrastructures within virtual research communities[4]

4.1   C3-Grid

Case Overview

What does the project do mainly? C3-Grid links distributed data archives in several German institutions for earth system sciences. With the help of Grid technologies it creates an infrastructure which provides tools for effective data discovery, data transfer and processing for scientists in climate research. This can increase productivity in scientific work by climate scientists[5].

Motivations for setting it up: The original motivation for initiating the project was improving access to data needed for simulations. Until then no overview of the existing data archives of earth science existed; accessing proprietary data at other institutions was even less feasible. And if access was possible, researchers faced the problem that the format and structure of the data from other disciplines were completely different. It was quickly realized that Grid technology could solve the problem of connecting the distributed data repositories.

Main goals of the project: The mission for C3-Grid was to build a collaborative environment to facilitate data discovery, data access and data processing (Kindermann and Stockhause 2008). The C3-Grid user accesses data from simulations and observational data stored in institutionally and geographically distributed archives (e.g. WDCC/Hamburg, Pangaea/Bremen, DLR, DWD and others). Access to the data is provided via one portal and the data come in a standardized format. An integrated data management system supports typical workflows.

Project maturity: The project started in September 2005 and has been terminated officially in February 2009 after an extension in August 2008. The C3-Grid is a founding member of the German D-Grid initiative. A follow-up project is currently being evaluated for funding. To date a working prototype of the Grid has been implemented. To make the Grid ready for a reliable service production a further 3 years of development are necessary. The software is not fully stable yet; during access peaks the system frequently gets overloaded. A further problem is the improvement of the international operability, e.g. tools allowing a first data analysis will be implemented to get the users a first impression of the data.

Up to now there are around 50 users. Most of them are scientists working in Germany. In the future the Grid will be opened for scientists from other countries as well. A main target of the successor project is to expand the Grid architecture and functionality in a way that enables uncomplicated access for scientists from all over the world.

Project funding: C3-Grid is part of the German D-Grid initiative. The D-Grid Initiative (German Grid Initiative) builds a sustainable Grid infrastructure for education and research in Germany. The German Ministry for Education and Research funded D-Grid and C3-Grid accordingly. The funding was awarded for personnel costs only. Hardware and other infrastructure had to be provided by the participating organizations. Consequently a significant proportion of the costs were borne by the participating institutes.

---

[4] The authors of this section are: Franz Barjak, Oliver Bendel, Erica Coslor, Kathryn Eccles, Tobias Hüsing, Zack Kertcher, Eric Meyer, Simon Robinson, Ralph Schroeder and Gordon Wiegand
[5] This description is based on 90 minutes of face-to-face and telephone interview time with 3 informants as well as documents available on the C3-Grid website (http://www.c3grid.de) and some published material as cited.

Organizational Structure

Size and composition: C3-Grid consists of eight project partners, plus six associated partners. This is the core group of official project participants. The consortium consists of eight data-providing institutions. The task of this group is to provide and arrange the data, including e.g. descriptions of the data with metadata. Eight further members of the consortium, so called operators, represent the users. These are mainly universities and other scientific institutes. Two further members of the consortium are responsible for informatics. Additionally there are three associated academic partners (all universities) and three industrial partners.

Governance: The Alfred-Wegener-Institute for Polar- and Marine Research (AWI), Bremerhaven, coordinates the project. Each group of project partners has its specific task. The role of the domain scientists is to specify the requirements and to provide the domain-specific applications, like diagnostic scripts. The data providers contribute processed data. The computer science partners have to supply all the middleware.

Managing internal and external relations

Management of the project: If possible decisions are made consensually. Controversial issues are discussed with the AWI having the final decision-making power. To date all decisions have been taken jointly by all partners. The project partners meet roughly every six months.

Users: Up to now (May 2009) around 50 users have used the C3-Grid. In the terminology of the project users are always individual scientists. All users so far are based in Germany, but in the future the Grid will be opened to scientists from all over the world. Despite the users as persons are based in Germany the projects, in which they are involved, are typically international. Hence the impact is not limited to Germany. The formal extension of the user community to other countries is scheduled for the end of the project. Since the Grid has not yet reached its desired final production functionality the users still have to have a good knowledge of the technology. It is for many purposes not yet possible to use the Grid routinely. Hence not even in Germany itself all potential users are included.

User recruitment: After trying out different strategies with little success, it turned out, that the most effective way to raise interest among scientists for C3-Grid is to visit their institutions. So, C3-Grid project members travel and visit the institutions with potential interest making the user-recruitment strategy more flexible and thus user-orientated. Presentations at specialist conferences also turned out to be a powerful way to motivate scientists to become involved in C3-Grid.

Drivers and barriers to adoption: The main driver for a scientist to join C3-Grid is the easy access to huge data archives coming from both real measurement and simulations. Since the data are distributed over many institutions, C3-Grid is the only possible way to get the data. To find a demand for the Grid within the community is most difficult. But since the use of Grid technology is not as trivial as the use of other software, a certain amount of insight in the technology is required on the users' side. The C3-Grid is from the viewpoint of the user not a black box, but a "grey box" (C3-Grid interview). This involves the need to acquire specific knowledge and an inclination to computational research among the potential users.

Challenges in interdisciplinary collaboration: Different scientific cultures are a major problem in the project management. To solve the problem interdisciplinary task forces were installed which convened face-to-face meetings for discussing the appropriate way to proceed. A good deal of the workload of the project managers concerns the coordination of the different disciplines. The coordinator estimates that approximately 20% of the overall workload of each project member is needed to find a common basis with other project members. Additional 20% of the workload is required for amendments where the supposed mutual understanding only seemingly was existent.

Collaboration with other organizations: Many of the relevant German organizations which would qualify as partners are already included in the project consortium. This is especially true for all earth science institutions. The collaboration with partners from computer science is broad in scope as well. C3-Grid is embedded in the D-Grid initiative for example. Thus an exchange of experience concerning grids is guaranteed. Furthermore there is a dedication to the world wide Grid community, e.g. the project engages in collaboration with EGEE.

Technology

Main technologies, resources and services and the role of technology development: From a technical point of view the aim of the project was to "gridify" existing diagnostic workflows and to provide the Grid itself. C3-Grid did not extend the methods of earth system sciences; it was "only" focused on the technical, i.e. infrastructural aspects. Hence, many of the tools of the project are middleware. Existing tools were used as much as possible, but many had to be developed anew. The key challenges were to enable data discovery with automatic metadata generation, to ease data access by bridging heterogeneity, support data processing by workflow composition and organize the access to resources with a consistent security infrastructure.

Data sharing: Despite data sharing is the gist of the project one lesson learned was that a sophisticated access right management has to be implemented. In May 2009 a new internal project has started to implement a new access right management system.

Interoperability with similar or connecting infrastructures: As C3-Grid forms part of the D-Grid initiative, collaboration with other German grids is wide-ranging. C3-Grid is considered to be one of the most important Grid technology development projects in this initiative. The connection to EGEE is rather loose. Since both projects use different middleware, the main purpose of the collaboration is to ensure the re-usability of the tools by making them compatible. C3-Grid is not only an early but also a successful project within the Grid community. Since it is well documented and has published many of the preliminary tools, C3-Grid has become a model at least within the European Grid community.

Contribution

Main contributions of project: The impact of C3-Grid in the earth science community is substantial. It has enabled the analysis of data from different sources simultaneously which has led to new insights into the interaction of earth subsystems. Furthermore there is a strong impact on the methodology of earth science. It is common sense in the community that local data management in petabyte dimensions is not possible anymore. As pointed out, C3-Grid has become a model at least within the European Grid community.

Challenges The submission of an application for a follow-up project to the German Federal Ministry for Education and Research is planned for May 2009. The aim is to advance the Grid from the prototype to the production status. The software has to be stabilized and scalability needs to be reached. In regard to the content the work is done but it still needs testing at length. A further task is to improve the international interoperability. Pre-processing of the data has to be improved in order to reduce the size of data that is being transmitted in every data download. Users with limited internet download speed, e.g. from developing countries, can only handle customized data sets. Hence the functionality of C3-Grid will be broadened. A third task is to review and edit the access rights management in the Grid. The current version is not elaborate enough and needs to be refined. An additional task to improve the interoperability is the integration of C3-Grid into partner grids like Earth Systems Grid (ESG, http://www.earthsystemgrid.org/) and the Nerc Data-Grid (http://ndg.nerc.ac.uk/). From the viewpoint of the project management the most important task is to find better

communication solutions especially to improve communication between members of different disciplines

Informants' recommendations to policy makers

Not covered in C3-Grid interviews.

SWOT analysis

	Table 4-1: C³-Grid strengths and weaknesses	
	Strengths	**Weakness**
Long-term funding	The funding of the follow-up by the Ministry for Education and Research of the project is very likely but the final commitment remains to be made.	The funding of the project by the ministry concerns only the manpower costs. The costs for the hardware are contributed by the participating institutions. So, it is still somewhat unclear what happens, if one of the institutions should withdraw from the project.
Sustainability	Since the participating institutions switched their data storage step by step from local to Grid archives it is not easy to switch back. Once the commitment to participate is made, it is hard or even impossible to step back.	Even though it is an integral part of the project to open the Grid for an international community, it currently still is restricted to German scientists.
User recruitment	Users are recruited by visits and presentations at expert conferences. Different strategies of recruitment have been tried, so it is likely that the most effective way could be found.	The strategies might work well for the German community, but it will be expensive and demanding to recruit users internationally by personal visits. New strategies have to be established.
Involvement of current users	Current users are mostly highly committed. Many important projects within the earth sciences are not realizable without Grid technology anymore. There is a vital necessity to stick with this technology to acquire prestigious projects and to publish in high impact journals.	Since the Grid is not fully operable yet, it is still a problem to open up Grid technology to scientists who are not computer-savvy. It still needs specialist knowledge to use the Grid and this discourages potential users, as the workload is too high to have the Grid doing what it should do.
Organizational bedding	All involved institutions have a long tradition as research institutions. Many of them are flagships of the German research system.	
Institutionalised links	As figurehead of the German Grid community at least the bracing within German and European Grid projects is excellent. Furthermore there exist at least loose affiliations to most Grid projects all over the world.	
External use of software, tools	A significant part of the work of C³-Grid was to develop middleware and Grid standards. Many younger Grid projects in Germany have adopted the technology and tools.	The core of the Grid technology is the middleware. C³-Grid - like all D-Grid projects - use gLite as middleware. But from an international point of view much more research is done on Globus, an alternative to gLite. Globus is used

		in many other paradigmatic projects.

Table 4-2: C³-Grid opportunities and threats

	Opportunities	Threats
Funding of member organizations	All member organizations are major research institutes or universities. Their funding is guaranteed for the future.	
Technology monitoring	Within the earth science community C³-Grid is setting standards. The project is being presented and discussed at all major conferences in the affiliated fields. Furthermore, C³-Grid is an active member of the Grid community, so developments in this field won't be missed.	The purchase and maintenance of the hardware is a responsibility of the participating organizations. Hence, it is not guaranteed that all use the same high standards of hardware. There is no obligation to adapt the best technology. But up to now this is more a theoretical problem.
Competition with other infrastructures or technologies	C³-Grid is very well embedded in D-Grid. Hence not only C³-Grid but other projects as well help to improve the tools in use.	As mentioned the middleware gLite used is different from the middleware Globus of other major projects.
Security risks	Up to now no security risks are known.	A more sophisticated access right management system has to be developed. A separate project proposal has been developed and submitted for funding.
Change of user communities and fields	The current trend within the earth sciences is to develop models with huge data bases. These data bases can only be handled with Grid technology. It is conceivable that more and more scientists will use the Grid.	Since more scientists will use the Grid the Grid has to become more user-friendly. It has to work like usual software, which means that no highly specialized skills have to be necessary to use it. Furthermore the Grid will be opened to researchers from all over the world including countries without good internet access. Hence, access to data has to be simplified in a way that only the data really needed is downloaded. A more sophisticated pre-processing of the data has to be developed and implemented.

4.2   CineGrid

Case Overview

What does the project do mainly? CineGrid is a worldwide community or network of excellence in which organizations and individuals in the areas of electronic visualization, networking, media studies and engineering collaborate.[6] Its origins lie in the iGrid events and is closely related to the GLIF and scientific visualization communities. The fields of scientific networking and electronic visualization are subfields of computer science which has developed rapidly since the 1980s.

Motivations for setting it up: CineGrid was founded in order to apply e-science developments to anticipate and satisfy evolving needs in the global media industry in Hollywood and in digital cinema elsewhere. The CineGrid founders anticipated that there would be enough bandwidth available on high-speed networks to move for the first time very high-quality media images around the world in real time. Digital cinema and the community developing it could benefit from a closer interaction with the community developing e-science infrastructure on a global scale involving high-speed networks and high-resolution scientific visualization. Though there is a focus on digital cinema, any tools and applications developed within CineGrid are still equally usable in other collaborative environments (such as scientific visualization).

Main goals of the project: The worldwide CineGrid community joins forces in targeted R&D projects to play, prototype, experiment and do proof of concepts with new high-quality digital media and cinema on super fast optical photonic networks. One informant phrased this as follows: *"We are a kind of self-organizing user group riding on other people's infrastructure on a volunteer basis."* (CineGrid interview 4). CineGrid is a scientific and pre-commercial undertaking.

Project maturity: The CineGrid organization, cinegrid.org, was incorporated in 2006, four years after the idea was born and after the first proof of concept at the iGrid2005 conference. Though the CineGrid community has realized several demonstrations since then, it is still in an early phase: it lacks funding and does not produce its infrastructure services persistently (nor does anybody else do it at the CineGrid level).

Project funding: CineGrid is mainly funded through membership fees and eventual corporate partnerships, research sponsorships, grants and project fees.

Organizational Structure

Size and composition: According to its website CineGrid currently has (April 2009) around 50 member organizations; 80% of these organizations are located in North-America (the US and Canada), and around 10% in each Asia (Japan and Korea) and Europe. CineGrid members include Networking organizations (National Research and Education Networks, institutions working with lambdas), media schools and university institutes in the areas of computer science and media, non-profit and other public organizations that currently use high-performance digital media as a means of discovery, education and collaboration, IT and telecom corporations, film & media companies.

Governance: CineGrid has a Board of Directors and an Executive Committee as main governing bodies. The secretariat is run by Pacific Interface Inc., a California-based consultancy. The

---

[6]   This description is based on 380 minutes of face-to-face and telephone interview time with 6 informants as well as documents available on the CineGrid website (http://www.cinegrid.org) and several other websites (as indicated in the text) and some published material as cited.

community collaborates mainly through dedicated CineGrid projects, which are pre-commercial field trials, proof-of-concept demonstrations, technology test-bed experiments or first-of-a-kind remote collaborations with high-quality digital media and cinema on super-fast optical photonic networks. It convenes at the annual CineGrid workshop to present and demonstrate CineGrid projects and technological advances and discuss future technological challenges and progress.

Managing internal and external relations

Management of the project: Involvement in CineGrid is mainly obtained through these joint projects and the annual workshops. There are no centrally administered work-programs and work-packages producing predefined outputs. The motivations for participating in and contributing to CineGrid vary between the members and important catalysts include: the necessity to have counterparts for doing experiments on high-speed networks, identifying new markets for NREN's, exploring the use of new media technology in the cultural sector (and spreading what is useful) and the community's forward-looking developments in the area of distributed content management and retrieval.

Users: There is currently no constituency or group of end users of what CineGrid offers that goes beyond the CineGrid community. This mainly consists of technology developers and innovators and it is growing, as the attendance to the annual workshop shows.

User recruitment: CineGrid members reach and attract with their demonstrations and presentations large and diverse communities.

Drivers and barriers to adoption: The main drivers for new organizations to join CineGrid are similar to the motivations of the current members (see above). A general trend towards enhancing certain products and services with high-quality multimedia content and visualizations is also supportive; for instance, universities are more and more interested in using it for creating new teaching experiences, supporting scientific collaboration and communicating their research. Barriers against working with CineGrid originate in the necessity to having very fast (usually 10 Gbps.) optical fibre-connections, scepticism towards new technologies (e.g. in the motion picture industry), lack of expert knowledge needed to participate in such a community of excellence and funding.

Challenges in interdisciplinary collaboration: Interdisciplinary collaboration is built into the CineGrid community and one of the aspects that make joint projects technologically interesting and fruitful. Participants have been described as willing and able to compromise and find solutions:

"One of the reasons for CineGrid members to join projects is to throw their people into this challenging environment where they must learn by doing and must mature as people. One of our explicit goals of CineGrid is to help grow the next generation of media professionals who must understand the new tools and capabilities brought on in our case by high-speed networking." (CineGrid interview 4)

Interviewees mentioned problems of collaboration, for instance between technologically oriented people and content producers, stemming from different attitudes to and experiences with networks and digital technologies; problems also appear between practitioners and scientists. The different groups in CineGrid use many different languages, have different mindsets and paradigms and find it occasionally difficult to understand each other. As one informant stressed, when constituencies like this start a project, there needs to be a learning phase, in which the collaborators learn enough about each other to be able to work together. Hence, the projects tend to incorporate such learning phases. The main formal activity established to supporting such learning and creating a common technological knowledge base among the community members are tutorial sessions offered before the annual CineGrid workshop.

Collaboration with other organizations: N/A

Technology

Main technologies, resources and services: overview of available resources, technologies and services: CineGrid projects generally require a systems integration of media devices, computers, storage and networks (on the infrastructure level) and network engineers, computer scientists, audio/video technicians, performers and directors/producers (on the human level). The core technologies required for realizing CineGrid projects are high-resolution digital cameras and display technologies and high quality sound technologies as well as very high-performance fibre-optical network connections and computing resources. CineGrid does not own any of the hard- or software that is used in its projects, but its members contribute these on a voluntary basis.

Role of technology development: Technology development is mostly an outcome of the uncoordinated research and development activities of the groups involved in the community and the results are then shared. "Learning-by-doing" and solving problems through trial-and-error are common approaches during the realization of projects. The CineGrid exchange, however, is a coordinated development that was started in 2007, as CineGrid faced the growing need to improve storage and management of its collection of digital media assets.

"The CineGrid Exchange is a distributed digital media repository designed to support CineGrid Member-driven test-beds for digital media asset management, distribution and preservation applications. The Exchange consists of digital media (visual and aural) of varying resolution, subject matter and format made accessible to CineGrid Members via secure high-speed networks." (http://www.cinegrid.org/index.php?option=com_content& view=article&id=73&Itemid=32).

Data sharing: The main challenges in producing the CineGrid Exchange were related to funding the production of different types of content (educational, artistic, scientific) in experimental formats and Digital Rights Management (DRM). While the content contributors maintain copyright to their work in the CineGrid Exchange, CineGrid negotiates certain content usage rights for its members. Under the widest use agreement (CineGrid Gold and Silver Members) the material can be edited, transcoded, and/or used for experimentation and research without restriction. Commercial use is excluded and CineGrid Members must endeavour to prevent distribution and disclosure of CineGrid content, particularly to broadcast television and the Internet.

Interoperability with similar or connecting infrastructures: N/A

Contribution

Main contributions of project: Interviewees in all CineGrid interviews pointed to projects realized by usually several CineGrid members as the best way of making the intentions and achievements of CineGrid clear to external observers. These projects were described as experimental and pre-commercial; they have the intention of developing tools which enable others to effectively use photonic networks' bandwidth. Examples of these use cases as they were described by CineGrid informants and in publications are included in the longer case report (see also Shimizu et al., 2006, Shirai et al., 2009, Smarr et al., 2007). Other contributions stressed by the interviewees are: i) CineGrid has brought together people from different communities and supported their collaboration in joint projects; ii) an environment of trust and mutual understanding has been created in which collaboration and the sharing and joint use of resources are common practice; iii) CineGrid has raised awareness about new audio-/video-, production and post-production, transmission and display technologies and work-flows among artists, filmmakers and other media professionals; along the same lines, it has raised awareness of using visualization technologies among scientists; iv) it has

demonstrated the feasibility of new modes of high-quality video transmission and work-flows in production and post-production.

Challenges: The main challenges for the future are to secure the funding of the community, to develop the management and coordination activities further so they keep up with the growth of the community, and last but not least to continue providing value for money to its members, in particular *"move from ad-hoc one-time demonstrations to more persistent efforts"* (CineGrid interview 4).

Informants' recommendations to policy makers

Not covered in CineGrid interviews.

SWOT analysis

Table 4-3: CineGrid strengths and weaknesses

	Strengths	Weakness
Long-term funding	The long-term funding of the CineGrid organization is secured through its membership fees.	Funding is low level. The community depends on additional funding and contributions in kind from its members for realizing the CineGrid projects. In the past it has been possible to mobilize the necessary funds, but it cannot be said to what extent this will be achieved in the future.
Sustainability	There is no pre-defined project ending. The community is embedded in other, larger communities of networking and electronic visualization research with which it interacts to mutual benefit.	
User recruitment	CineGrid members attract with their presentations large and diverse communities. Individually CineGrid members engage in outreach activities, present projects at workshops and events, make demonstrations and performances. Thus, they raise interest, widen their individual networks and acquire contacts for future projects. The community is growing in numbers.	CineGrid does not have any users as such. Its members are researchers, developers, and professional practitioners interested in combining the technologies of research networking and electronic visualization with digital cinema technologies. The activities are experimental and pre-commercial. CineGrid does not have any dedicated activities or campaigns for increasing the community or involving potential end users beyond the CineGrid network.
Involvement of current users	Some CineGrid members have large intrinsic motivations and drive the community as it contributes to their home organization's core activities and mission.	The community does not have any strategy or guideline to have its members involved. Except for the CineGrid exchange, there are no coordinated R&D activities and much is done on an ad hoc basis as project opportunities appear. Some members, in particular from private companies, have been described as "developing members" who do not contribute much to the research, but show interest in the community's developments.

		The community is rather organizationally detached and not integrated into any organization such as an academic society or research institution.
Organizational bedding		
Institutionalised links	CineGrid is informed by the work of other e-infrastructure projects and communities, in particular the OptIPuter project and the GLIF community. This is mainly because CineGrid members contribute to or even drive these projects and communities.	Institutionalised forms of cooperation were not mentioned by any of the sources on CineGrid.
External use of software, tools	The major coordinated development in CineGrid is related to a distributed system for storing and retrieving at high-speed large high-quality audio and high-resolution video material, the CineGrid Exchange. The developments in this area are supported and closely monitored by those CineGrid members, who have similar needs in their home institutions. Besides, CineGrid members share their developments and achievements in projects with the community.	No examples of wider sharing or use of CineGrid results were mentioned by the informants.

Table 4-4: CineGrid opportunities and threats

	Opportunities	Threats
Funding of member organizations		An overall assessment of the funding of CineGrid organizations is very difficult due to their number and diversity. One of the main drivers of the community who also acts as secretariat, Pacific Interface Inc., is a small consultancy firm in the areas of business consulting and business development. An evaluation of its funding situation is not possible, but the funding is probably less dependable than that of a major university or other publicly funded organization.
Technology monitoring	The community receives first-hand information on new developments mainly through some of its members, who are at the forefront of their fields and involved in standardization and governance activities in academia as well as business.	
Competition with other infrastructures or technologies	Neither in technological nor commercial sense there is any strong competition for CineGrid as no similar initiatives exist. The Enhanced Digital Cinema (EDcine, http://www.edcine.org/) project was funded within the 6th	

	Framework Programme and focused on the advancement of digital cinema in Europe. However, the CineGrid community sticks out because of the excellence of its members and it can be considered as unique. CineGrid is doing experimental and pre-commercial work, hence commercial competition is not relevant.	
Security risks		Security problems could affect the CineGrid community negatively: Firstly, the production, archiving and subsequent use of the CineGrid Exchange audio/video material require a system of digital rights management that protects the material from commercial and other misuse. Second, the scepticism towards using networks for transmitting movie content or making content for network experiments and demonstrations available has been stated as one of the barriers to entering the community that applies in particular to the motion picture industry; the security of the content is also the major issue in this case.
Change of user communities and fields	The current trends and expected changes in the networking and electronic visualization fields and the media education/science sector have not been assessed in the interviews in detail. However, there are at least two different trends which are supportive to the community's work: 1) a general trend in higher education and research towards including high-quality multimedia content and visualizations; 2) the rising importance of digital cinema technologies in the motion picture industry.	

4.3   CLARIN

Case Overview

What does the project do mainly? CLARIN is a large-scale pan-European collaborative effort to create, coordinate and make language resources and technology available and readily usable.[7]

Motivations for setting it up: The project builds upon a strong existing network of researchers, digital tools and technologies, and data archives. The project was submitted under the ESFRI scheme in accordance with its criteria. The aim of CLARIN is to enable the latest developments and initiatives to support and develop language resources, and to broaden both the remit and the impact of current and proposed initiatives among network partners.

Main goals of the project: CLARIN will offer scholars the tools to allow computer-aided language processing, addressing one or more of the multiple roles language plays (i.e. carrier of cultural content and knowledge, instrument of communication, component of identity and object of study) in the Humanities and Social Sciences.

Project maturity: CLARIN is a nascent project, currently in the preparatory phase. This will be followed by a construction period of 5 years.

Project funding: Secured from the EC in the first instance (€4.1m), subsequent phases dependent on national investment from partner countries.

Organizational Structure

Size and composition: CLARIN is a true pan-European project, with partners in almost every European country. Partners include language resource archives, academic departments within universities, and academies of science. Total number of members: 156, Number of countries involved: 32.

Governance: The project is managed by a multi-tiered structure, comprising a Scientific Board consisting of high-level scientists, a Strategic Coordination Board consisting of representatives appointed by the funding agencies, an Executive Board consisting of 8 experts covering the required expertise and each leading a Work Package, an International Advisory Board to give advice to the Executive, Scientific and Strategic Coordination Boards on issues of common interest, and National groups to define an appropriate national coordination structure.

Managing internal and external relations

Management of the project: Management of the project is distributed between a Scientific Board, a Strategic Coordination Board, an Executive Board (composed of senior work package leaders and liaison staff), an International Advisory Board and CLARIN members. Work flows between these management structures are well developed and clear.

Users: Ensuring the continued development of the user community is a high priority for the CLARIN team. The large majority of the current CLARIN community consists of providers rather than users, CLARIN will have to work hard to ensure firm and structural liaisons with potential users in order to make sure that the resource will actually be used.

User recruitment: See above. The problem of user recruitment is one that the CLARIN team are aware of, but until the project is more mature, little action can be taken. CLARIN staff

---

[7] Number of informants: 2, totaling 125 minutes.

are actively seeking collaborative opportunities in order to build upon existing expertise (in recruiting users) and active communities.

Drivers and barriers to adoption: The main driver to adoption will be CLARIN's provision of persistent services that are secure and provide easy access to language processing resources. At present, in order to perform simple language processing tasks, one needs to find an appropriate program (to do translation, summarization, or extraction of information, etc.), download the program, make sure it is compatible with the computer that will execute the program, understand the form of input it takes, download the data (e.g. novels, newspapers, corpus, videos), and convert them to the correct format for the programs, and all this before one can get started. For most researchers outside computer science, at least one of these tasks will currently be an insurmountable barrier. CLARIN will provide resources for processing language, the data to be processed, as well as appropriate guidance, advice and training, and will be accessible over a distributed network from the user's desktop.

One potential barrier to adoption is that European countries have to opt in to the organisational structure and provide funding for the construction and operational phases of CLARIN. For researchers in countries that do not join the consortium, there may not be full access to all services..

Challenges in interdisciplinary collaboration: One of the major challenges is to co-ordinate efforts spread across such a wide and varied group of partners. The preparatory phase is to address the legal, organizational, financial and technical challenges to building infrastructures in this field.

Collaboration with other organizations: CLARIN has partners responsible not only for collaborating with other ESFRI preparatory phase projects such as DARIAH, but also with existing and recent language projects, and other European infrastructure projects such as Europeana. CLARIN has dedicated work packages looking at collaboration with other organizations, including researching similar efforts in North America and the potential for linking up with such efforts. CLARIN has recently initiated CHAIN, the Coalition of Humanities and Arts Initiatives and Networks, to work together with CenterNet, DARIAH, Project Bamboo and the Association of Digital Humanities Organisations (ADHO) to ensure interoperability of the shared services that we are developing.

Technology

Main technologies, resources and services: overview of available resources, technologies and services:

Processing: Incorporates advanced multi-lingual language processing technology that supports cultural and linguistic integration. Incorporates, and contributes to, Semantic Web technology to overcome the structural and semantic encoding problems

Network: Includes Data Grid technology to connect the repositories

Data/storage: Builds on ideas launched by the Digital Library community to create Live Archives, and will further such initiatives

Role of technology development: Ensuring interoperability of existing language processing technologies and data sets is an important part of CLARIN's work.

Data sharing: Existing digital resources which are being made available via repositories in the network will be made available through a new framework which will allow common resource discovery procedures, common metadata formats and procedures. It will also provide existing tools as web services, in a Grid environment, where currently disparate resources will be able to be used together. No privacy/security issues are foreseen. There are legal and ethical

considerations regarding language resources that contain copyrighted material and potentially sensitive material relating to people and communities.

Interoperability with similar or connecting infrastructures: CLARIN will rely on existing and emerging technologies to guarantee interoperability of language resources. They are confident that these resources will supply these needs. The main challenge here is to make sure at this early stage that the project avoids conflicting standards and competing services.

Contribution

Main contributions of project: The CLARIN project seeks to build upon and reinforce a network of researchers, language tools and technologies and digital archives to widen both the usage and impact of these resources within and outside the field. It will marry currently disparate tools and datasets, creating interoperability that will significantly enhance the types of research undertaken and widen access to language resources generally. In creating networks with other e-Humanities infrastructures such as the DARIAH project, CLARIN is committed to maximizing the reach of its resources.

Challenges: The main challenge is the timely development of the resource, currently on target, and the continued national and international support for this development. A further challenge pointed to by project personnel is the development of a robust user community. As the project is at an early stage of development, this second factor is likely to be addressed more directly as the resource develops.

Informants' recommendations to policy makers

CLARIN has recently issued a statement regarding copyright which proposes a research exception to European copyright law to allow researcher to make use of digital materials covered by copyright for educational and academic research purposes within a secure research infrastructures.

SWOT analysis

Table 4-5: CLARIN strengths and weaknesses		
	Strengths	Weakness
Long-term funding	Project has already secured some national funding for the next phase of the project, so potential further investment is likely.	Long-term funding has not been secured.
Sustainability	Project is currently in a preparatory phase, with clear objectives for next phase (building) and beyond.	
User recruitment	Project is well integrated in target user communities and has a well researched user engagement plan.	Immaturity of project means that no measures have yet been tested.
Involvement of current users	N/A	N/A
Organizational bedding	The project is well established within multiple institutions and a number of overlapping academic communities. This project is very much a 'bottom-up' effort, signifying strong commitment from the institutions involved.	
Institutionalized	Yes. CLARIN have worked hard to	

links	research and integrate themselves within similar projects and infrastructures. Co-operation is further secured through dedicated liaison personnel.	
External use of software, tools	N/A	N/A

Table 4-6: CLARIN opportunities and threats

	Opportunities	Threats
Funding of member organizations	Multiple participating organizations so difficult to say, although this could be seen as an advantage – being anchored to so many organizations reduces the threat to the project by unstable funding in one or more partner institutions.	
Technology monitoring	Yes. The project has considerable accumulated experience in technology development, and knowledge of potential and actual technologies currently outside the project which may be useful.	
Competition with other infrastructures or technologies	Competition is not strong – collaboration is extremely strong. CLARIN seeks to integrate itself within and consolidate existing efforts rather than compete with them.	
Security risks	N/A	N/A
Change of user communities and fields	Not known.	Not known.

4.4   D4SCIENCE

Case Overview

What does the project do mainly? D4Science, the successor project of DILIGENT, is one of the main European e-Infrastructure projects. DILIGENT produced a testbed e-Infrastructure and its enabling system, gCube. The developed e-Infrastructure provided basic functionality for: (1) controlled sharing and access to distributed heterogeneous content, services and computational resources; (2) on-demand creation of Virtual Research Environments (VREs) providing access to subsets of the shared resources. The VREs can be used for different requests, e.g. the monitoring of processes, the analysis of data and the collaboration of users. D4Science is currently an offer for the Environmental Monitoring (EM) and Fisheries and Aquaculture Resources Management (FARM) communities.[8]

Motivations for setting it up: The D4Science project aims to continue the path that the GÉANT (a multi-gigabit pan-European data communications network, see the separate case report), EGEE (Enabling Grids for E-science in Europe, see the separate case report), and DILIGENT (A Digital Library Infrastructure on Grid Enabled Technology) projects have initiated towards establishing networked, grid-based, and data-centric e-Infrastructures (Castelli and Michel 2008). These e-Infrastructures are expected to accelerate multidisciplinary research by overcoming several crucial barriers that stand in the way, primarily those related to heterogeneity, sustainability and scalability. When DILIGENT was designed, the core partners had quite a long background in the digital library field. The trend was to go to federated, distributed systems which allow the integration and sharing of digital content coming from different places. As informants indicated, DILIGENT realized a movement from the traditional digital library technologies towards e-Infrastructures which offer a radically less expensive organizational and development approach for supporting access and exploitation of shared knowledge and the construction of Virtual Digital Libraries (ancestors of the Virtual Research Environments).

Main goals of the project: The D4Science project aims at deploying the e-Infrastructures built so far by the EGEE and DILIGENT projects so that they address the needs of scientific communities affiliated with the broad disciplines of EM and FARM. The e-Infrastructure will provide facilities for creating VREs based on shared computational, data and service resources offered by EGEE and DILIGENT at a European level, as well as on data and domain-specific service resources offered by large international organizations. In particular, the DILIGENT testbed infrastructure will be brought into production by preserving its usage dependencies with the corresponding EGEE production infrastructure.

Project maturity: The project has recently started its second year. As the ground has been prepared by the predecessor project DILIGENT, D4sience was able to start with a quite sophisticated infrastructure in place. The general challenge is homogenizing the access to this infrastructure for several groups. At the moment, the e-Infrastructure is established and productive. gCube maintains the infrastructure, and based on gCube, domain specific VREs have been created for the two communities (see below).

Project funding: EU funding for D4Science amounts to 3.15 million EUR. The overall budget is 3.92 million EUR. D4Science is on the top of other projects and uses resources and technologies which were developed in EGEE and DILIGENT. Therefore it is very difficult to estimate direct and indirect costs.

---

[8]   This description is based on 170 minutes of telephone interview time with 3 informants as well as published materials, papers and documents available on the D4Science website, the project's public wiki pages and a set of presentations, photos and videos.

Organizational Structure

Size and composition: D4Science is one of the main European e-Infrastructure projects. Eleven partners (National research centres and several organizations) from seven countries participate (France, Italy, Greece, Switzerland, United Kingdom, Malaysia, Hungary) in the project.

Governance: The project's "governance" structure includes groups, functions and roles like External Advisory Board, Project Coordination, Members General Assembly, Project Management Board, Managers for each community, Project Executive Board, Technical Director and diverse managers. Futhermore, there is a taskforce quality insurance. The External Advisory Board (EAB) is a panel of external experts advising on project strategy and complex technical decisions. The D4Science EAB is comprised of four specialists, one from the digital library domain, one from the grid domain and two from the user communities. The Project Management Board (PMB) is the supervisory body of the project. It is designed to promote continuous sharing of project knowledge across all areas of activity.

Managing internal and external relations

Management of the project: Several tools facilitate the collaboration in the project team. Informants explained that a shared workspace is available which hosts resources and materials. A track system is used for the handling of bugs as well as for the monitoring of tasks and compliance with milestones. Mailing lists for every work package have been established. There are weekly telephone conferences and face-to-face plenary technical and managerial meetings every tree months. The website gives access to nearly all available resources. Furthermore there is a large number of monitoring tools which are used for maintaining the e-Infrastructure. Last but not least, an online event calendar lists activities in the project like events, related to the project activities.

Users: Several hundred users have participated in the project until now and further communities shall be addressed in the future. Five mediators act as connecting points to large communities of users which are not technology-oriented and not necessarily interested in the use of software.

User recruitment: The current users were recruited and trained by the technical team and the user community mediators. In workshops further users are attracted and trained.

Drivers and barriers to adoption: The EM Community consists of researchers and stakeholders operating over a widespread geographic scale to provide political and technological solutions to global environmental issues, like protection of the marine environment, preservation of forest ecosystems and studies of climate changes. Requirements converge on having secure collaborative computing environments where accessing huge amount of heterogeneous information and domain computing services be seamless tasks. The FARM Community consists of researchers and decision-makers from many disciplines spread worldwide and operating to facilitate and secure the long term sustainable development and utilization of the world's fisheries and aquaculture. Requirements are VREs, encompassing many resources on aquatic biodiversity and socio-economics, offering to the communities tools for collaboration on shared fishery assessments in a continual way. The participating user communities show many commonalities in their expressed requirements and are willing to share information and data whenever necessary; they also are in the process of investigating further usages of the infrastructure thanks to the availability of domain applications. Because of the generic gCube software, every community can be addressed in principle. From a technical standpoint there are no barriers to adoption.

Challenges in interdisciplinary collaboration: There are four main groups of members or participants: The first group is the management of the project, the second consists of computer scientists, e.g. developers and testers, the third are the domain specific mediators

which are responsible for the creation of the virtual research environments; and the fourth group are the users from the participating user communities EM and FARM. One member is basically doing the management, administration and coordination in the project. The managers control the compliance of milestones with deadlines and coordinate the computer scientists, technicians and work packages. Seven partners are from the computer science domain and responsible for technical tasks like developing and testing the e-Infrastructure and the specific applications. They have diverse expertise, e.g. in software testing, grid computing, and library systems. Three partners act as the gateways to the user community, actively promoting technological achievements and informing the technical team about relevant feedback from the end-users. It is also part of the activity of these partners to provide technical support and knowledge to share existing IT resources in their work domain.

Collaboration with other organizations: D4Science is collaborating with other FP6 & FP7 projects and R&D programmes. These collaborations are of different nature, as they range from technical exchanges involving mutual exploitation of technologies to the sharing of e-Infrastructure resources and joint organization of networking and dissemination events.

Technology

Main technologies, resources and services: D4Science is consolidating and enhancing the technology which underpins the D4Science e-Infrastructure operation, namely the gCube framework. Cube, successfully deployed within the testbed developed by the DILIGENT project, reflects within its name a three-sided interpretation of the grid vision of resource sharing: sharing of computational resources, structured data and application services. As such, gCube embodies the defining characteristics of computational grids, data grids and virtual data grids.

Role of technology development: The gCube it builds on the gLite middleware (developed by the EGEE project) for managing distributed computations and unstructured data, includes dedicated services for managing data and metadata, and offers a novel approach for managing application services. Rather than interfacing with the infrastructure, the software that implements the services is literally handed over to it, so as to be transparently deployed across its constituent nodes according to functional constraints and quality-of-service requirements. This is genuinely ambitious and entirely novel: like computational resources and data before, application logic in gCube becomes a pervasive commodity within an infrastructure that abstracts from the physical location of its resources at any point in time. D4Science now consolidates and enhances the gCube services to reflect the shift in functional and QoS requirements, which marks the passage from a testbed to a production-level infrastructure.

Data sharing: D4Science mainly serves the needs of the two communities EM and FARM by providing them with a portal through which practitioners from these fields can define and access various VREs giving organized and seamless access to the resources they use in their daily activity. Two kinds of resources are particularly relevant with respect to the community operation, data sources and tools. The data sources are for example repositories of various types of data and information ranging from digital versions of documents, to temporal series, data stored in databases, data gathered from satellites or sensors, and, in general, any other source of information the communities need to have access to for accomplishing their tasks.

Interoperability with similar or connecting infrastructures: N/A

Contribution

Main contributions of project: With regard to the main contributions of the project so far, the following benefits can be listed: An infrastructure was built which enables the sharing of resources; several search tools are available; domain-specific tools have been developed; a

dialogue between different communities has been established. According to an informant, the main contribution is the automatic generation of Virtual Research Organizations and Environments. The success should be measured in a few years from now on the basis of the number of the participating users and communities. Another informant agreed with the mentioned appraisal and underlined the importance of an infrastructure where end users can create and use Virtual Research Environments. It is a collaborative workspace integrated with existing grid technologies; this virtual environment for collaboration exploits the advantages of grid computing technologies and therefore supports very complex and high-demanding requests. In contrast to "normal" grid computing projects which only provide the capability to process and store large amounts of data, in D4Science there is the possibility for an efficient exchange of data.

Challenges: Not all problems are solved yet, not so much on a technical level but with regard to the support of the user communities and the attraction of new users. Further challenges are the interoperability with other infrastructures which have already aggregated content from different sources (to be realized in the planned successor project D4Science-II) and the simplification of access to the D4Science services for other infrastructures. The realization of the on-demand generation of reports will be a very positive impact on enhancing scientific collaboration. Last but not least, the infrastructure should have many facilities to automatically manage itself in order to reduce the cost of its maintenance.

Informants' recommendations to policy makers

As one informant pointed out in an article, the next step will be to move from the current model, which is characterised by the existence of autonomous, independently-operated heterogeneous e-Infrastructures to e-Infrastructure ecosystems, where e-Infrastructures are interoperable and can collaborate by sharing resources and capabilities (Castelli 2008). The author arrived at the conclusion that Ecosystems will serve a significantly expanded set of communities dealing with multidisciplinary challenges, the solution of which is beyond the reach of existing resources.

SWOT analysis

Table 4-7: D4Science strengths and weaknesses		
	Strengths	Weakness
Long-term funding	The funding of the current project is secured through the support by the EU.	It is not clear if there will be a successor project.
Sustainability	At the moment, the e-Infrastructure is established and productive. The user communities benefit from the e-Infrastructure. The chances are good that other communities can be attracted, provided that there will be a successor project.	The project will end in 2009.
User recruitment	The current users were recruited and trained by the project team and the mediators. In workshops further users are attracted and trained.	At the moment, there is a clear focus on two user communities. The communities have differences but also commonalities; they already attempt to share data and tools. It is unclear of totally different communities can be integrated by implication.
Involvement of current users	Some users have large intrinsic motivations and are very interested in a	

	further use of the e-Infrastructure; this is described in some research papers.	
Organizational bedding	D4Science seems to be embedded well, as it contributes to the core mission of its participating organizations, namely to deliver Grid computing services.	
Institutionalised links	D4Science is well informed of the work of other e-infrastructure projects and communities and is collaborating with other FP6 & FP7 projects and R&D programmes; furthermore, the user communities participate in several projects respectively programmes.	The collaborations are of different nature, as they range from technical exchanges involving mutual exploitation of technologies to the sharing of e-Infrastructure resources and joint organization of networking and dissemination events.
External use of software, tools	Within the VREs the users have the possibility of selecting a number of technologies and services and creating a bundle of them, for domain specific investigations and analysis. For example the users can share an archive or a database. Other communities can also benefit from the automated processes.	No examples of wider sharing or use of D4Science results were mentioned by the informants or in the available documents.

Table 4-8: D4Science opportunities and threats

	Opportunities	Threats
Funding of member organizations	The Environmental Monitoring (EM) and Fisheries and Aquaculture Resources Management (FARM) communities are big and strong communities and linked with the Food and Agriculture Organization of the United Nations (FAO) and the International Center for Living Aquatic Resources Management (WorldFish Center). Therefore a funding of member organizations could be possible.	
Technology monitoring	The project receives first-hand information on new developments mainly through some of its members, who are - like the participating Universities and the CERN - at the forefront of their fields and involved in standardization and governance activities in academia as well as business.	
Competition with other infrastructures or technologies	Developments like gCube and the automated generation of VRE result in a unique selling proposition.	There is a competition for D4Science because the work of EGEE is continued as in similar initiatives.
Security risks		Security problems could affect the D4Science community negatively. In the fields of Environmental Monitoring and Fisheries and Aquaculture Resources are strong political and commercial interests.
Change of user communities and	There are several trends which are supportive to the community's work, e.g.	

fields	the increasing need of climate data because of the global warming and the increased awareness of an ecological balance.	

4.5   DARIAH

Case Overview

What does the project do mainly? The Digital Research Infrastructure for the Arts and Humanities – DARIAH is a nascent project funded through the European Commission's Seventh Framework Programme (FP7).[9] DARIAH is the collaborative effort of several data centres across Europe to plan and support a digital infrastructure to underpin research in the arts and humanities. The project is currently in development.

Motivations for setting it up: The DARIAH project is ambitious. It aims to provide an infrastructure 'for the entire field of arts and humanities and access to [the] cultural heritage of Europe'.[10] It plans to create 'a common understanding of the cultural diversity and its history in Europe'. The planned impacts are stated to be: the facilitation of comparative research over time periods, cultures, languages, or regions, and the triggering of novel research questions, that with traditional access to cultural heritage sources dispersed over a multitude of different sites and institutions could up to now not be approached. It also proposes to help enhance national infrastructures.'

Main goals of the project: To create an international digital infrastructure for the Arts and Humanities.

Researchers will use DARIAH to:

- Find and use digital content from Europe and acquire tools to use and interpret it,
- Ensure the long-term preservation of data,
- Ensure that they work to accepted standards and follow best practice,
- Exchange ideas and knowledge of digital scholarship and seek advice, and
- Use DARIAH as a site of experimentation and innovation in collaboration with other scholars.

Archives, libraries, museums and other 'repository agencies' will use DARIAH to:

- Make their digital information known to a wider pan-European public,
- Ensure the long-term preservation of data,
- Get help with and advice on digitisation, curation and preservation of data,
- Use DARIAH as a site of data exploration and innovation in collaboration with other institutions.

Project maturity: Currently in a preparatory phase, called 'Preparing DARIAH', which started in September 2008. The aim of this stage of the project is to produce a blueprint for construction of the DARIAH infrastructure.

Project funding: 'Preparing DARIAH' is funded by the EC. The preparatory phase is estimated to cost €6 million, with construction costing another €10 million.   In order to secure the DARIAH project, annual funding of an estimated €6 million is required from national governments and funding organisations. The aim is to create an infrastructure of at least 25 partners, requiring a funding commitment of €250,000 per partner, although it is envisaged that large countries will pay more and small ones less, depending on national priorities.

---

[9] Number of informants: 1 totaling 120 minutes, several additional conversations.

[10] ESFRI Roadmap Report (2006), p. 33.

Organizational Structure

Size and composition: The organisational model for DARIAH is based on a three-tiered structure. Firstly, at the local or **thematic (domain) level**, research and digitisation projects, resource centres, **communities** of practice and other subject coalitions will form the basis for DARIAH. Secondly, at the **national level,** DARIAH partners will provide services ensuring permanent access to digital resources. They will also contribute to stimulating best practices and standards. Finally, at the **European level,** DARIAH will have several key functions: Enabling, coordinating and funding; Setting best practice and standards; Harvesting, harmonisation and combination of digital resources.

Governance: During the preparatory phase, DARIAH's workflow will be divided between and led by six institutions across Europe, project managed by *Data Archiving and Networked Services* in the Netherlands.

Managing internal and external relations

Management of the project: The project currently has a workflow divided between six lead institutions, supported by a number of project partners. No management structure for the building phase of the project has been released at this stage.

Users: The main users will be on the one hand researchers, and on the other, archives, libraries, museums and other 'repository agencies'. Researchers will find and use digital content from across Europe, deposit their own data (and work to common standards in so doing), use DARIAH to exchange knowledge and expertise and collaborate with other scholars. Archives, libraries, museums and other 'repository agencies' will use DARIAH to widen access to their resources, ensure long-term data preservation, ensure best practice in the digitisation, curation and preservation of data, and collaborate with other institutions.

User recruitment: The DARIAH project began with only 4 partners and has since increased to 14, representing 10 European countries and three types of partners: data centres, technical institutes, and humanities research partners. The project seems therefore to have united providers and users, with the latter group expressing a great deal of interest in digital applications. This growth was achieved 'without much effort at all': through speaking about DARIAH at conferences, sent around a letter to research funding organizations and that was it, more or less.

Drivers and barriers to adoption: A major early change to the project was the withdrawal of funding from the UK partner, the Arts and Humanities Data Service (AHDS) in April 2008. The AHDS had been funded by the UK's Joint Information Systems Committee (JISC) and the Arts and Humanities Research Council (AHRC). Many institutions have committed to carrying on the work of the AHDS, but the withdrawal of this organisation from the UK's national funding map was an unanticipated development for the DARIAH project, whose success depends on recruiting national support.

Challenges in interdisciplinary collaboration: The challenges to collaboration derive largely from disciplinary traditions: some areas of the Arts and Humanities remain the territory of the lone scholar, writing single-authored papers and attending occasional conferences. However, these scholars are increasingly turning to digital libraries and archives for research and teaching. Some areas of the Arts and Humanities are highly developed in terms of e-Research, with fields such as Archaeology and Linguistics at the forefront of digital technologies and collaborative working methods.

Collaboration with other organizations: DARIAH is currently collaborating with fellow ESFRI project CLARIN (http://www.clarin.eu/), and with another large European infrastructure project, Europeana (http://www.europeana.eu/portal/), in order to ensure interoperability of their services.

Technology

Main technologies, resources and services: overview of available resources, technologies and services: The general objective is to draft the technical reference architecture of DARIAH. This architecture will consist of drafting engineering plans for the construction as well as small proof-of-concept prototypes for key enabling technologies. There are four major activities:

- A scoping study to identify already existing technologies that enable research infrastructures, survey the technological infrastructure at partner centres, and recommending technologies and standards for the primary building blocks of DARIAH.
- Building the technical reference architecture, demonstrating the validity of an approach that focuses on the development of an integrated middleware for supporting digital access to arts and humanities data and services in Europe through proof-of-concept studies
- Proposing a technical and functional roadmap which establishes a model and a methodology for a future DARIAH network, and defining tasks and roles of the partnering data centres and overarching DARIAH services.
- Construction of the proof-of-concept demonstrators, design studies that focus on specific problems of arts and humanities data, which need to be overcome to implement DARIAH as a research infrastructure. The developed system will integrate the access, archiving, and organization of electronic resources and will permit the harvesting of metadata.

Role of technology development: Technology development will occur in Work package 7 of the 'Preparing DARIAH' project. It is a crucial phase of the initial project to show that the technology can support the envisaged architecture.

Data sharing: DARIAH is unrestricted in terms of the types of resources it seeks to make part of the infrastructure. It is equally interested in texts, images, video, map data etc. It builds upon the work and resources of the original partners, all national data archiving centres.

Interoperability with similar or connecting infrastructures: The aim is to build an infrastructure to ensure the interoperability of every collection within DARIAH, and to develop interoperability at an early stage with other EC projects such as CLARIN and Europeana (see above).

Contribution

Main contributions of project: The main contribution of DARIAH will be to facilitate the use of digital humanities and cultural heritage (DH&CH) information. Sharing of expertise, tools, and ICT methods for creation, curation, preservation, access and dissemination are key elements in the infrastructure.

Challenges: As with many of our case studies, one of the main challenges to the future of this project is the securing of sufficient future funding. In order to develop the main infrastructure, the Preparing DARIAH project needs to secure national support from each of its partner institutions. As the project develops, further challenges, particularly relating to user recruitment, may well arise but are not foreseen at this stage.

Informants' recommendations to policy makers

None stated.

SWOT analysis

Table 4-9: DARIAH strengths and weaknesses		
	Strengths	Weakness
Long-term funding		No. Initial phase is funded by the EC; subsequent development is dependent on national funders becoming involved.
Sustainability	The project is currently in a preparatory phase, with building due to commence in September 2010	No long term plans are available regarding sustainability.
User recruitment	Strong interest in the project from prospective users.	Immaturity of the project means no results of user engagement are currently available
Involvement of current users	Prospective users are being integrated into the project from the outset.	As above, immaturity of project means no results available.
Organizational bedding	Yes, the project is well embedded, but...	The variety of institutions involved in the project may be uncertain in future.
Institutionalized links	Some links being built with CLARIN and Europeana to ensure interoperability.	
External use of software, tools	N/A	N/A

Table 4-10: DARIAH opportunities and threats		
	Opportunities	Threats
Funding of member organizations		
Technology monitoring	Yes, and through collaboration with other infrastructures is likely to stay ahead of these developments.	
Competition with other infrastructures or technologies	Scarcity of similar Humanities projects means that there is no fierce competition, and more of a collaborative, integrated effort.	Currently working with other infrastructures but difficulties may arise if funding becomes scarcer
Security risks	N/A	N/A
Change of user communities and fields	Not known.	Not known.

4.6   DEISA

Case Overview

What does the project do mainly? DEISA, the Distributed European Infrastructure for Supercomputing Applications, connects the leading supercomputing centres in Europe—extending from Italy and Spain through Central and Western Europe to the UK and Finland—through high speed network empowered by GEANT2 and the national research networks (DEISA 2008). The fundamental objective of DEISA is to provide services which enable the high performance operation of remote computing platforms on remote data sets.[11]

Motivations for setting it up: The integrated supercomputing power of DEISA is intended to bring a boost in competitiveness for Europe in scientific areas where extreme performance is needed. The provision of high performance computing resources to researchers has traditionally been the objective and mission of the national HPC centres in Europe. However, the increasing global competition between Europe, USA, and Japan is inducing growing demands for computational resources at the highest performance levels, as well as a need of fast innovation. To stay competitive, major investments in replacing supercomputers are needed every five years—an innovation cycle that is difficult to follow even for the most prosperous countries because of the enormous costs. The limitations of national approaches became increasingly clear to policy makers and practitioners. In 2002, the idea of combining resources across countries emerged in order to overcome the fragmentation of supercomputing resources in Europe. According to this logic, aggregated supercomputers would offer both better system availability and the necessary skills for more efficient supercomputing support (Lederer 2008). In 2004, DEISA1 started as an Integrated Infrastructure Initiative of eight leading European supercomputing centres. The project substituted the available resources for networks and computers with a higher performance, installed HPC architecture and gave the users and communities access to combined resources. DEISA2 started in 2008 and continues to develop and support the computing infrastructure.

Main goals of the project: The purpose of DEISA is to enable scientific discovery across a broad spectrum of science and technology, by enhancing and reinforcing European capabilities in the area of high performance computing (DEISA 2008). Grid technologies are used to integrate national supercomputing platforms, and to provide to scientific users transparent access to a European pool of computing and data resources. The joint and coordinated operation of this environment is tailored to provide enhanced computing power and resources to end users, and to enable new, ground breaking research activities in science and technology. DEISA operates as a virtual European supercomputing centre.

Project maturity: All important elements for a distributed system exist and are functioning. However, several processes are still being optimized and automated. Further associated partners are still expected to be integrated, a process which is likely to be challenging. Three types of European-wide expert teams have been established.

Project funding: The project cost for DEISA1 was 24,351,100 EUR, the EU funding 13,976,000 EUR. The project cost of eDEISA (an additional interim project) was 13,145,700 EUR, the EU funding 7,000,000 EUR. The project cost of DEISA2 is 18,733,200 EUR, the EU funding 10,237,000 EUR. The difference is financed by the national supercomputing centres themselves.

---

[11]   Published materials and papers, documents and videos available on the DEISA website and the project's newsletters were used in preparing this case report in addition to interviews with 3 informants totalling 120 minutes of interview time.

Organizational Structure

Size and composition: Eleven principal partners, all leading national supercomputing centres, are involved in DEISA2. Each site is strongly integrated to form the DEISA HPC infrastructure. In addition, four associate partners act as coordinators of HPC activities in each participating country, as well as Russia.

Governance: The project's "governance" structure includes groups, functions and roles like Principal Partners, Associate Partners, DEISA Executive Committee, Extended DEISA Executive Committee, Technical Board, Advisory Scientific Committee, Project Management Team, Project Coordinator, Technical Coordinator, Quality Assurance Coordinator and Presentation Team. The management of the DEISA project is the responsibility of the DEISA Executive Committee and the Project Coordinator who, as the chairman of the Executive Committee, represents the DEISA project. The Project Coordinator is responsible for the day-to-day management tasks and assisted by the Project Management Team. He receives support from the Technical Coordinator.

Managing internal and external relations

Management of the project: On the operational level, 1—2 hour video conferences were held every two weeks, and 1—2 face-to-face meetings were arranged every year. Furthermore face-to-face meetings of the executive board and other strategic and management committees and teams take place regularly. An internal wiki is used to monitor frequently changing issues. In addition, a document management system provides more persistent information, e.g. on proposals and deliverables. The newsletter, founded in 2005, is aimed mainly at the scientific community and HPC end users at large, but it also spreads useful information for the Grid technology development community in Europe and the rest of the World.

Users: 90 extreme computing projects were conducted, run by more than 160 European universities and research institutions, including partners from outside of Europe. About 600 users in several countries for example have access to the computer in Munich. An example of a science community with DEISA HPC support is the European fusion community. The fusion community brings together some of the largest European research laboratories working in the areas of nuclear fusion research.

User recruitment: There are Europe-wide calls, supported by a press release to attract users and user communities. Most of the users were recruited and trained by the project team. The idea is to make the project known to other communities so they will be able to take the developed technologies and apply them in their daily work.

Drivers and barriers to adoption: The participating communities have a strong need for supercomputing. Many projects are not imaginable without support from an e-Infrastructure like DEISA, especially in the fields of natural science and in projects with a global reach and data input. With increasing numbers of project proposals submitted, it has become necessary to review and select submissions. DEISA is supporting users through various media such as user documentation, training programme, operating a centralized user help desk and monitoring the availability status of the DEISA software stack at the partner sites.

Challenges in interdisciplinary collaboration: On the project's side, the Operations Team, the Development and Technology Team and the Applications Support Team play an important part. An example of a science community with DEISA HPC support is the European fusion community. Further targets for DEISA support are EU FP7-supported computational projects. In addition to fusion energy research, the initiative is supporting, e.g., climate/earth system research, astrophysics/cosmology, life sciences and materials sciences. The initiative is also open for communities from other disciplines. Software engineers have deep knowledge on supercomputing hardware and software but not in detail of the natural sciences and other

user domains. They can help to make programs faster and solve technical problems. Users have direct contact with some supercomputing experts and communicate with them by phone or e-mail, especially in cases of troubleshooting. Documentations and FAQs are used to help solving recurring problems. It was suggested that there are hardly communication problems between the providers and user communities. The communities are mostly from the natural sciences and the users have a good understanding of possibilities and barriers of supercomputing.

Collaboration with other organizations: DEISA is cooperating with a long list of partners in different academic fields and domains, e.g. CLARIN, COSMOS, DANTE, EFDA/ITER, ENES and the European Psi-k Network in Material Science.

Technology

Main technologies, resources and services: The DEISA research infrastructure is constituted by leading national supercomputers in Europe interconnected with a high bandwidth point-to-point network provided by GEANT2 and the National Research Networks (NRENs). High bandwidth network connectivity is required to guarantee the high performance of the distributed services, and to avoid performance bottlenecks. DEISA incorporates several different platforms and operating systems (IBM AIX on Power5-6, IBM Linux on PowerPC, SGI Linux on Itanium, Cray XT, and NEC vector systems), and the consortium has deployed middleware that enables the transparent access to distributed resources, high performance data sharing at European scale, and transparent job migration across similar platforms.

Role of technology development: The DEISA system will be enriched by community interfaces to make it more user-friendly and to allow a more comfortable usage from external computers. A new research paper declared more precisely: "In DEISA2, the single-project-oriented activities (DECI) will be qualitatively extended for persistent support of virtual science communities. DEISA2 will provide a computational platform for them, offering integration via distributed services and web applications, as well as managing data repositories." (Lederer 2008)

Data sharing: In the fields of meteorological or cosmological research large amounts of data are produced, e.g. by satellites or other measurement instruments with gigabyte rates per application. In materials science simulations are also important and often connected with a lot of data. A big challenge is the transfer of huge data sets from computer to computer. For example, in meteorological research terabytes of data are produced on one computer. A growing problem is the long-term storage of raw data. A solution being discussed is to delete the raw data after a first analysis and re-generate it later if necessary. In the future, JAVA-based web interfaces will be developed for facilitating data access and transfer.

Interoperability with similar or connecting infrastructures: N/A

Contribution

Main contributions of project: An important contribution of DEISA is the collaboration among European supercomputing centres. Previously, there were only national initiatives and no interaction at a European level. The HPC sites in Europe had different strengths and special competences in different parts of the HPC field. The centres had to develop common platforms and distributed services. The possibility to process wherever it is beneficial for a project, respectively where resources are most appropriate and available, is revolutionary in every respect. One point is that countries which cannot provide specific systems for themselves are still able to do simulations and analyses on a high level. Another advantage for European science is the improved access to supercomputing resources for scientists from countries which cannot provide specific systems themselves. With DEISA, it's easier to apply for projects at an international level. The increasing transparency of the system and its

processes make the users aware of the hardware and resources that are available. A global central file system that is available for all DEISA users also increases transparency.

Challenges: In the future, the focus will be more on a complete European HPC ecosystem, with a heterogeneous set of machines and Grids at different levels of the infrastructure. One of our informants noted that these services need to be accessible to any scientist through a workspace that hides as much as possible the complexity of the DEISA systems, but facilitates the creation of workflows using differing system architectures or Grids as needed.

Informants' recommendations to policy makers

An informant indicated that there is an increasing demand for computing time. In the last call the demand exceeded the submission with a ratio of 3:1. Therefore more and more proposals must be declined. Another problem is that the development of hardware is faster than the development of software. It is necessary to support also the software development to overcome the gap. Otherwise in the future only a few users will be able to use the provided services and systems. Another informant pointed out that in addition to the technical challenges, human collaboration is one of the most important factors. It is necessary to consolidate the team and build a community to support a kind of collaboration which is based on trust and cooperation.

SWOT analysis

	Strengths	Weakness
Long-term funding	The long-term funding of the project is secured through the support by the EU.	It is not clear if there will be a successor project.
Sustainability	At the moment, the e-Infrastructure is established and productive. The user communities benefit from the e-Infrastructure. The chances are good that other communities can be attracted. Especially the calls seem to be effective instruments.	The project will end in 2011. It is well documented; but not all web pages are available.
User recruitment	There are Europe-wide calls, supported by a press release to attract users and user communities. Most of the users were recruited and trained by the project team. The idea is to make the project known to other communities so they will be able to take the developed technologies and apply them in their daily work.	At the moment, there is a strong focus on one user community, the fusion community. Further targets for DEISA support are EU FP7-supported computational projects. Ten European computational science grand-challenge projects from the DEISA Extreme Computing Initiative were presented at a recent conference, covering the fields of Weather and Climate Research, Engineering, Materials Science, Astrophysics, Computational Neurosciences, Plasma Physics and Computational Bio Sciences.
Involvement of current users	Some users have large intrinsic motivations and are very interested in a further use of the e-Infrastructure.	Because not all projects can be accepted, not all potential users can benefit from the e-Infrastructure and there is a preference for high-level projects with a strong need for supercomputing.

Table 4-11: DEISA strengths and weaknesses

Organizational bedding	DEISA seems to be embedded well, as it contributes to the core mission of its participating organizations, namely to deliver super computing services.	
Institutionalised links	DEISA is well informed of the work of other e-infrastructure projects and communities and is collaborating with other initiatives like CLARIN, COSMOS, DANTE, EFDA/ITER, ENES and the European Psi-k Network in Material Science.	
External use of software, tools	DEISA is one of the biggest projects in the field of grid and supercomputing and provides several e-Infrastructures with computing power and resources.	

Table 4-12: DEISA opportunities and threats

	Opportunities	Threats
Funding of member organizations	DEISA is linked with other strong organizations, i.e. supercomputing centres. Therefore a funding of member organizations could be possible.	
Technology monitoring	The project receives first-hand information on new developments mainly through some of its members, who are—like the participating supercomputing and research centres—at the forefront of their fields and involved in standardization and governance activities in academia as well as business.	
Competition with other infrastructures or technologies	Neither in technological nor commercial sense there is any strong competition for DEISA as no similar initiatives exist. The focus in EGEE is more on grid computing with the need of data exchange.	Cloud computing may be a threat; but it has unsolved problems in the field of security and can hardly reach the same computing power.
Security risks		Security problems could affect the DEISA community negatively. In the field of fusion and nuclear power are strong political and commercial interests.
Change of user communities and fields	There are several trends which are supportive to the community's work, e.g. the increasing need of electricity (which can be a product of nuclear power stations) and mobility (which can be supported by results of fusion research).	

4.7   Digital Repository Infrastructure Vision for European Research (DRIVER)

Case Overview

What does the project do mainly? DRIVER develops standards and infrastructure for sharing content (especially metadata) and functionality among digital repositories. DRIVER aims to integrate the metadata and data (that means publications and publication related information) that exist in repositories across *Europe*. In trying to harmonize this highly varied data across Europe, it is hoped that the community of repository managers will be brought together.

Motivations for setting it up: The main motivation behind the project is to bring together scattered scientific information in one (virtual) place and make it accessible to a wider audience. There are various scientific repositories storing research material such as texts, data and other material in Europe in universities, research institutions, and national organisations. At the time DRIVER started, there were scattered similar initiatives on the national level in several countries (e.g. Netherlands, Germany, UK) and DRIVER set out to bring all this information together, with repository managers and librarians targeting end users.

Main goals of the project: DRIVER is set out to develop and provide the e-Infrastructure and the interface capable to support the integration of various digital repository sources from diverse collections. A core building block to reach this aim is the software D-Net which can be used in all kinds of repositories and in digital libraries. In addition to this main goal, several other goals exist, such as fostering a European community of repositories, expanding the geographical reach of digital repositories, establishing a European Confederation of digital repositories, promoting the idea (and availability) of enhanced publications and advocating Open Access.

Project maturity: Having started in 2006, DRIVER is a relatively young e-Infrastructure. Nevertheless, good progress seems to have been made, with the DRIVER platform D-Net having already being adopted by four National Repositories.

Project funding: The funding is 2.7 M Euro for DRIVER2 and was 1.8 M Euro for DRIVER1. With cost of operation being at about 1 million EUR per year, 40% are indirect and 60% direct costs.

Organizational Structure

Size and composition: DRIVER involves participants from 13 institutions from eleven European countries. These institutions include universities and National repositories.

Governance: Project tasks are divided among several organizations: NKUA (University of Athens) is responsible for scientific, technological and management support. NKUA is the project coordinator, maintains the services provided by DRIVER-II, and provides support for enhanced publications and support and training to users. ISTI-CNR is the scientific and technological coordinator of the project. Technical aspects of the project are handled by inter-organizational collaboration among partners

Managing internal and external relations

Management of the project: The DRIVER Confederation of European Digital Repositories constitutes a network of content providers and involves academic institutions that host scientific digital repositories, universities as well as research centres, and other national, regional or subject-based federations.

Users: There are three types of users:

- Repository managers who provide DRIVER with their content or take up DRIVER infrastructure.
- National organisations (or other types of organisations) who are willing to take up the DRIVER infrastructure and build their own national digital repository systems
- End users using the portals

244 repository managers have submitted their data. Three European countries (Belgium, Portugal and Spain), together with China and India, are using or considering deploying the D-NET framework for their national repository.

User recruitment: DRIVER seeks to raise awareness among potential repository management level users in international workshops and conferences. Tutorials and demos are also presented in various European conferences (e.g. OAI6) and at a national level (Belgium and UK local conferences for repository managers). The work (papers) produced by the DRIVER partners is presented in European conferences related to the digital/information library fields.

Challenges in interdisciplinary collaboration: Co-operation between the different roles works smoothly according to the interviewed DRIVER representative. Challenges in interdisciplinary collaboration are not really encountered here.

Collaboration with other organizations: The DRIVER Confederation of European Digital Repositories involves academic institutions that host scientific digital repositories, universities as well as research centres, and other national, regional or subject-based federations. It is set up as a European federation of federations – a Confederation. It constitutes a network of content providers. It is intended to extend the DRIVER Confederation to assist those countries without developed national structures of repositories.

Confederation partners represent European and international repository communities, subject based communities, repository system providers, service providers, as well as political, research, and funding organisations Its members are organizations representing the key stake holders in the international repository landscape. Some of the DRIVER partners are, or will be, part of the confederation. Institutions and initiatives come from the majority of European countries, the U.S., Canada, Latin America, China, Japan, India and Africa. DRIVER has signed a Memorandum of Understanding or Letter of Intent with the following partner organisations: SPARC Europe, LIBER, eIFL.net, RECOLECTA Spain, DINI OA-Netzwerk Germany and DRF Japan.

The Confederation aims to advance DRIVER from test-bed status to a fully functional e-Infrastructure, including a sustainable organizational model, a geographical and thematic extension of the repository platform, the uptake of DRIVER technology and the close correspondence between the DRIVER infrastructure and communities of practice. It thus aims to provide an integrated concept for organisation, technology and content for the European Open Access repository landscape, in a virtual structure that is independent from the DRIVER project activities.

Technology

Main technologies, resources and services: The DRIVER system is implemented on an open service-oriented software architecture (SOA) logically organized into areas shown in the figure below. This architecture guarantees service extensibility and interoperability, system expandability and local repository autonomy. The DRIVER deployment may include multiple instances of the services identified, in order to guarantee better quality of service (e.g., availability, performance) or to support diverse functionality (e.g., different query languages). These instances, which may have either the same or a different configuration, are distributed on different sites /physical locations managed by the coordinating bodies. Choice

of sites and allocation and distribution of the service instances are driven by both organizational and technical quality parameters, e.g., independence, security, availability, performance, etc.
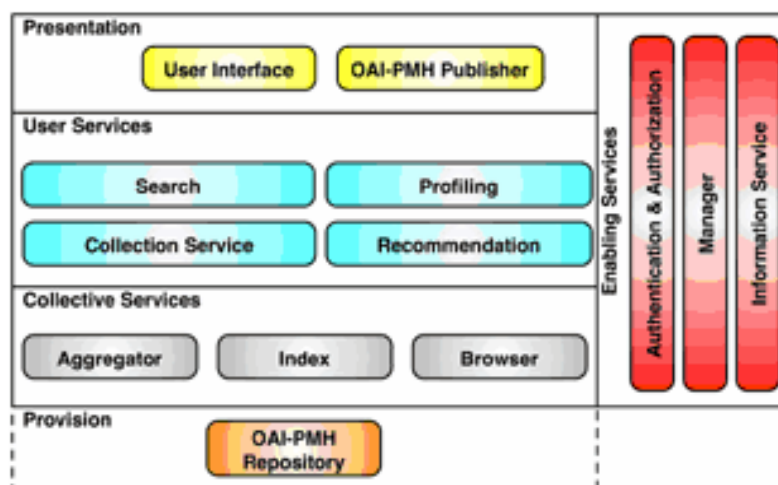
Role of technology development: DRIVER has developed D-Net. This open source software offers a tool-box for deploying a customizable distributed system featuring tools for harvesting and aggregating heterogeneous data sources. A variety of end-user functionalities are applied over this integration, ranging from search, recommendation, collections, profiling to innovative tools for repository manager users. A running instance of the software, namely the "European Information Space", maintained by the DRIVER Consortium to aggregate Open Access publications from European Institutional Repositories, can be accessed online at: www.driver-community.eu (Search the Repositories Portal).

Data sharing: N/A

Interoperability with similar or connecting infrastructures: N/A

Contribution

Main contributions of project: DRIVER has built the "European Information Space", the DRIVER search portal based on a robust network of content providers. The DRIVER software is running and can be used to set up similar portals by all kinds of institutions, also to develop new applications on top of the basic services. A support network for repository managers is up and running as well as services for the end-user. DRIVER is further advocating Open Access and promoting the idea of Enhanced Publications.
The webtool is now used in several European countries and many universities have registered. DRIVER's Open Access policy is especially valuable for smaller universities, giving them the opportunity to increase their visibility. About 245 institutions all over Europe participate in the DRIVER Information Space. There is strong community uptake and commitment. Sustainability is likely given the establishment of the confederation. Libraries generally have an interest in taking the DRIVER service and running it and repositories are willing to conform to the framework.

Challenges: The DRIVER representative interviewed states that the main objectives have been reached and the task is now to keep it working. The two main challenges for the future are studying and organizing the European repository landscape, trying to bring together a range of technologies and passing on standards to the various repository stakeholder groups, and secondly, making a production quality system with 24/7 operation, which requires planning and resources that go much beyond the scope of a project. The Enhanced Publication also

needs further attention and promotion. It is in the prototype phase and further research on "non-plain" publications and access to them is desperately needed.

Informants' recommendations to policy makers

Open access needs further advocacy. Although linking things has been a strong driving force in development in recent years, data and publications are still separated more often than not. Mandates are essential, if research funding hinges on making results available open access, this will push the idea forward.

SWOT analysis

Table 4-13: DRIVER strengths and weaknesses

	Strengths	Weakness
Long-term funding		Long-term funding of the DRIVER depends on funding subsequent projects
Sustainability		Accordingly, there is a defined project ending by the end of 2009.
User recruitment	245 repository managers have submitted their data. Three countries (Belgium, Portugal and Spain) are using or considering to deploy the D-NET framework for their national repository.	
Involvement of current users	Involvement of current users seems to be positive	
Organizational bedding	N/A	N/A
Institutionalised links	Links to large parts of the repository landscape exist	
External use of software, tools	Core business of DRIVER is external use of their software platform which has been quite successful	

Table 4-14: DRIVER opportunities and threats

	Opportunities	Threats
Funding of member organizations	Participating organizations are universities and national repositories that are not dependent on volatile funding.	
Technology monitoring		
Competition with other infrastructures or technologies		Google book search
Security risks	None disclosed	None disclosed
Change of user communities and fields		

4.8   EELA-2

Case Overview

What does the project do mainly? EELA is a four year project (in two separate phases) that was established across 14 European and Latin American countries to set up a high capacity, production quality, and scalable Grid facility and ensure the long-term sustainability of the e-Infrastructure by advancing the creation of National Grid Initiatives (NGIs) federated in a Latin American Grid Initiative (LGI).[12]

The EELA project is a multidisciplinary project that involves several academic fields in different functions. Probably the largest number of institutions and individuals contributing to EELA work in the field of Grid computing and the broader discipline of computer science; they are in charge of the provision of the computing and network services, supporting the applications and developing new services for applications and the infrastructure; major EELA user communities exist in high-energy physics (HEP), biomedicine and bioinformatics, and earth sciences; fewer users work with applications in fields like artificial intelligence and optimization, chemistry, civil protection, engineering, environmental science.

Motivations for setting it up: The establishment of EELA was largely technology-driven and advanced by scientists involved in Grid computing and the funding bodies behind them, in particular the European Commission. After the establishment of EGEE and development of the middleware, it became clear that other countries and regions worldwide could benefit from this investment. "Sister projects" were established with partners from the EU and partners in non-EU countries worldwide. The EELA initiative was welcomed by the HEP communities, computer scientists interested in Grid computing and other scientists and universities in Latin America which saw a chance to overcome their scarcity of computing resources and obtain access to more powerful computers.

Main goals of the project: The first EELA project mainly focused on the set up of the infrastructure and the human network, whereas the second EELA project looked to its extension and sustainability (Marechal, 2008, 3; Marechal, Gavillet and Barbera 2009). For the latter purpose it also engages in promoting and supporting the creation of National Grid Initiatives (NGI) and a continent-wide federation in Latin America.

Project maturity: Preparations on EELA started in mid 2004 and the first two-year funding period started on 01.01.2006, the second on 01.04.2008. The infrastructure is considered as quite mature by our informants, as it has successfully made the transition from a test bed to a production quality infrastructure with many contributing sites and applications running on it.

Project funding: The funding of 5.1 Mio. € in EELA-2 (3 Mio. € in EELA) is used for networking activities (dissemination, training, supporting applications), service activities (network and Grid computing services) and R&D (on middleware and applications) (see table 1).

---

[12]   This description is based on 340 minutes of telephone interview time with 7 informants as well as documents available on the EELA-2 website (http://www.eu-eela.eu/) and several other websites (as indicated in the text) and published material as cited.

Activity	EU countries	Latin American countries	Total	
			in %	In Euro
RTD	54%	46%	100%	507.0
Coordination	66%	34%	100%	2338.9
Management	100%	0%	100%	318.6
Other	42%	58%	100%	1942.2
Total Budget	57%	43%	100%	5106.7
of which EC funded	65%	35%	100%	2093.0

Table 4-15: EELA-2 budget and funding by continents

Source: EELA-2.

Organizational Structure

Size and composition: EELA-2 has 16 participants from 14 different countries, out of which five are European (Spain, France, Italy, Portugal and Ireland) and nine Latin American (Brazil, Argentina, Chile, Colombia, Cuba, Ecuador, Mexico, Peru and Venezuela) plus the multinational Cooperación Latino-Americana de Redes Avanzadas (CLARA, http://www.redclara.net/), the organization that connects the Latin American NRENs via its network RedCLARA.

Governance: The 16 participants in the project act as coordinators of so-called Joint Research Units (JRU) in their countries, if there are further EELA-2 partners in the respective country. A JRU is described as a partnership between organizations of the same nationality without any formal legal status (Marechal & Gavillet, 2008, 33). The JRUs are a new construct and were created to strengthen the process of establishing NGIs in the EELA-2 countries. However, they also have management functions in the project – e.g. all payments from the EC are channelled through the respective JRU coordinator. The set-up of the JRUs required much more time than was expected which created some tension in the governance of EELA-2. In the meantime, further partners – in total more than 50 – have become involved in EELA-2 (see full case report on the full list of partners as of May 2009). The JRU with most EELA-2 members is the Brazilian JRU with 15 partners. In Spain there are 8, in Chile 7, in Peru 4, in Portugal and Argentina 3, in Colombia, Venezuela and France 2 partners. Cuba, Ecuador, Ireland, Italy and Mexico have only single EELA-2 members (as of May 2009).

The project has three different boards with overlapping membership: the Management Board makes the day-to-day decisions required for running the project and the infrastructure and cares for the strategies and roadmaps for long-term sustainability; the Technical Board takes care of all technical issues securing technical coherence and progress of the project; the Consortium Board is described as the "parliament" of the project, with a rather symbolic role in practice. Six activities are differentiated in the EELA-2 work plan: three networking activities covering overall management, dissemination and training and support to applications (NA1, NA2, NA3), network and Grid computing services (SA1, SA2) and a research and technical development activity developing services for applications and infrastructure (JRA1).

Managing internal and external relations

Management of the project: Involvement of the existing EELA-2 members in the project is secured through regular telephone and Skype conferences, occasional face-to-face (f2f) meetings, mailing lists, a Wiki (http://Grid.ct.infn.it/twiki/bin/view/EELA2/WebHome) and Blogs (e.g. http://twitter.com/eela_na3), national and international meetings and workshops plus the annual EELA-2 conference. Furthermore, current members need to contribute to the

outreach and training events of the project and thus interact with each other as well as with (potential) users.

Users: The infrastructure currently (May 2009) supports 56 applications out of which 29 (52%) are biomedical, 8 (14%) from the earth sciences and 5 (9%) from high-energy physics. Application leadership is to more than 25% each in Brazil and Spain, followed by Mexico (10%) and the other involved European and Latin American countries. 32 or 57% of the applications have status 4 (testing) or 5 (deployed) and are considered as applications currently running on the EELA-2 infrastructure. The other applications have preparatory statuses. All applications are included in one single EELA virtual organization.

User recruitment: This is an important issue for the management as well as the involved partners and several activities are implemented, such as tutorials, Grid schools, workshops or "Gridification weeks". In addition, both the management of EELA-2 as well as local partners and their teams engage in further dissemination activities in their environment.

Drivers and barriers to adoption: The scarcity of computational resources in Latin America creates a convincing argument for universities to consider the EELA-2 Grid services (which are either based on gLite or the OurGrid middleware). Important barriers are the scarcity of funds and the still low maturity of Grid technology, making it difficult for early adopters to entirely avoid problems when installing Grids, porting applications to the Grid and running them later on.

Challenges in interdisciplinary collaboration: It is estimated that 70% of the people involved in the project are computer scientists and 30% are domain scientists. According to some informants interdisciplinary collaboration is working out smoothly in EELA-2. However, according to the problems mentioned by some informants, it cannot be ignored that EELA-2 is also affected by the type of problems that are frequently mentioned in e-science projects (Barjak et al., 2009): large costs of communication between computer scientists and domain scientists, no standardized research work flows, scepticism towards new computation models etc.

Collaboration with other organizations: EELA-2 is involved in a continent-spanning network of e-infrastructure stakeholders at network and Grid computing levels including NRENs, RedCLARA, EGEE, OurGrid and others.

Technology

Main technologies, resources and services: overview of available resources, technologies and services: As of May 2009 EELA-2 had 22 computing sites in production with a total of 5800 cores, of which 3800 computing cores can be provided to EELA-2 users; a 20 % growth in computing is planned over the project duration. EELA-2 uses two middlewares (see on this Brasileiro et al., 2008): 1) the gLite middleware for a service Grid developed in the EGEE project; 2) the OurGrid middleware, a free, open-source middleware that enables the creation of opportunistic peer-to-peer Grids. The OurGrid has been used to speed up the execution of "Bag-of-Tasks" applications, parallel applications whose tasks – the parts that run on a single machine – do not communicate with each other during execution. Currently, the OurGrid is mainly used by an active community of developers and users in Brazil. EELA connects Latin American sites through RedCLARA and with Europe through Géant.

Role of technology development: Research in EELA-2 is particularly addressing the co-existence of the OurGrid and the gLite middleware on the same infrastructure (Brasileiro et al., 2008).

Data sharing has so far not presented itself as a critical issue. Data security and privacy are maintained, though all users are integrated into one single EELA-2 VO. Access rights are allocated to individual applications and application owners decide what is shared.

Interoperability with similar or connecting infrastructures: The interoperability with EGEE was decided right at the beginning of EELA and the EGGE middleware and registration and certification procedures are used in EELA-2. Interoperability between the two middlewares gLite and OurGrid is a major issue of research in EELA-2 in order to integrate a service Grid and an opportunistic desktop Grid improving the performance of both. Interoperability with the Open Science Grid (OSG) in the US is another issue, as Latin American universities traditionally have strong collaborations with US universities and this also applies to e-science communities. This issue has not yet been resolved, but EELA-2 has a technical solution for Grid sites submitting jobs to both infrastructures, EELA-2 and OSG, which could be installed if a site desires this.

Contribution

Main contributions of project: EELA-2's continent-spanning network of e-infrastructure stakeholders at network and Grid computing levels and the general boost that it gave to the idea of Grid computing in Latin America are among its most important achievements. They build upon the fundament laid in the first funding period in which the "production quality" e-Infrastructure was established. EELA-2 today facilitates the better and faster production, mining, processing and analysis of data and helps to produce more accurate results in a shorter lapse of time. Examples for such contributions to scientific research are described in the literature (see for instance Dutra et al., 2007).

Challenges: The future challenge will be to make the infrastructure permanent and convince Latin American governments to build and dedicate resources to NGIs. As to be expected, the first and main challenge is to secure the funding for the NGIs. The most likely scenario is that the Latin American Grid Initiative (LGI) starts with a few NGIs and later on further NGIs join the federation. The same happened during the creation of NRENs and RedCLARA. It is currently being evaluated and negotiated whether the LGI can become a part of RedCLARA and this can be mirrored in similar national pairs of NGI/NREN.

Informants' recommendations to policy makers

Not mentioned in the interviews.

SWOT analysis

	Strengths	Weakness
Long-term funding		Long-term funding for NGIs and the LGI is still being negotiated with the Latin American governments and research and education networks.
Sustainability		The project ending is scheduled for March 2010. Then, NGIs should take over and provide the infrastructure services to their scientific communities.
User recruitment	User recruitment is an important part of the EELA-2 activities. There are several coordinated measures, like user tutorials, Grid schools, workshops in a community or country, and customized "Gridification weeks". In addition, both the management of EELA as well as local partners engage in further dissemination activities in their environment. These activities are to some extent successful, as the number of applications supported by the infrastructure is rising.	
Involvement of current users	Involvement of the existing EELA-2 members in the project is secured through regular telephone and Skype conferences, occasional f2f meetings, mailing lists, a Wiki, Blogs, national and international meetings and workshops plus the annual EELA-2 conference. Furthermore, current members need to contribute to the outreach and training events of the project and thus interact with each other as well as with (potential) users.	
Organizational bedding	It is currently being evaluated and negotiated whether the LGI can become a part of RedCLARA and this can be mirrored in similar national pairs of NGI/NREN.	EELA-2 is not (yet) embedded in any organization.
Institutionalised links	Institutionalised co-operations exist with different other infrastructures: The co-operation with RedCLARA, Latin American NRENs and Géant are necessary to provide transmission capacities. Co-operation with EGEE has been established right from the beginning and EELA-2 uses gLite, the EGEE middleware, for its service Grid. In addition to gLite, another middleware, OurGrid, is used to provide opportunistic Grid services for certain applications. OurGrid developers also participate in	

Table 4-16: EELA-2 strengths and weaknesses

		EELA-2.	
External use of software, tools		EELA-2 sites are also providing computer cores to EGEE.	EELA-2 has only few research & development activities. These mainly address infrastructure and application services for the project.

Table 4-17: EELA-2 opportunities and threats

	Opportunities	Threats
Funding of member organizations	The EELA-2 member organizations are mainly higher education and research organizations in Europe and Latin America. It is not possible to assess their funding situation.	
Technology monitoring	The project members are aware of the technological developments in the area of Grid computing as this is their core area of expertise.	However, they are only partly familiar with the computing models and possible alternatives in their application domains, such as biomedicine, HEP, earth sciences and the like.
Competition with other infrastructures or technologies		Technological competition between Grid computing and other computing models, e.g. local clusters or cloud computing, may already constitute or develop in the future as an alternative for many scientists.
Security risks	The combination of gLite and OurGrid middleware lowers security risks according to our informants' opinion. In the OurGrid all remote tasks are executed within a virtual machine that does not have access to the network and harm could only be done to the virtual machine.	
Change of user communities and fields	Diverse fields, changes cannot be projected.	

4.9   EGEE

Case Overview

What does the project do mainly? The project provides researchers in academia and business with access to a production level Grid infrastructure, independent of their geographic location. It develops the middleware gLite.[13]

Motivations for setting it up: EGEE was established in 2004 as follow-up to the EU DataGrid project (EDG) that produced a testbed of distributed computing and storage resources. The DataGrid software was the basis of the CERN Large Hadron Collider Grid Project first production infrastructure, the facility that has been set up for the analysis of data that is being produced by the CERN accelerator.

Main goals of the project: The main goal of EGEE is to refine the LHC computing Grid infrastructure of CERN (see below) to enable and encourage scientists from different fields to use it. The advancement of the Grid means to build a production-quality, i.e. secure, reliable, sustainable and robust, Grid infrastructure for scientific researchers to share computing resources across collaborative projects. Furthermore to re-engineer a light-weight middleware solution, gLite, specifically intended to be used by many different scientific disciplines. And to attract, engage and support a wide range of users from science and industry, and provide them with a production service supported by extensive technical and training support.

Project maturity: EGEE-III is the third project stage of EGEE. EGEE is closely linked to the LHC Computing Grid (LCG) of the Large Hadron Collider (LHC) of the CERN. LCG was designed to handle the massive amounts of data produced by the Large Hadron Collider. EGEE-I provided researchers with access to major computing resources. It has built a consistent, robust and secure Grid network and thus attracted additional computing resources. The second major task was to improve and maintain the middleware. The third core area was to attract new users. EGEE-II made a continuous operation of the infrastructure available. It introduced support for more user communities and added further computational and data resources. EGEE-III is focused on transitioning to a sustainable operational model, while maintaining services for its users. It is planned that the European Grid Initiative (EGI) will take over the e-infrastructure from EGEE after the end of the EGEE-III project.

Project funding: EGEE-III: The European Commission (through the Directorate-General for Information Society and Media) contributes 32m €. The funding of the EC is always matched by investments of the project partners and users. Total budget is 47.15m € (with a further estimated 50m € worth of computing resources contributed by the partners). The long-term objective of EGEE is to make the infrastructure self-sufficient.

Organizational Structure

Size and composition: EGEE connects more than 140 institutions in 33 European countries. It employs around 1'000 persons with a full time equivalent of about 380 persons.

Governance: EGEE-III features many horizontal groups that cover various aspects of the project's operation. Among them are the Administrative Federation Committee, the Activity Management Board, the Collaboration Board, the External Advisory Committee, the Project Management Board and the Technical Management Board.

Figure 4-1: Governance of EGEE

---

[13]   This description is based on 110 minutes of interview time with 3 informants as well as documents available on the EGEE website and published material as cited.
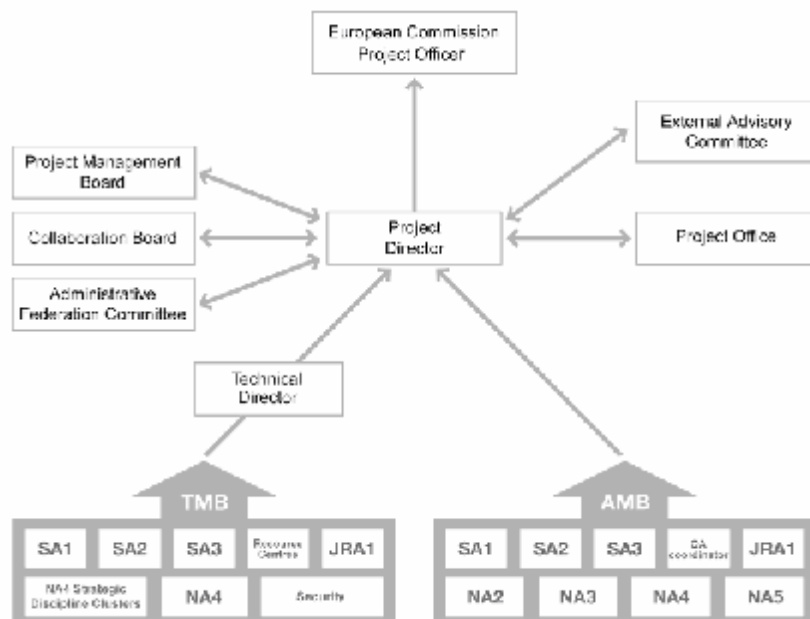
Managing internal and external relations

Management of the project: The objective of the management activity is to provide an overall project management and reporting to the European Commission. The scope of duties includes the daily management of the project activities, resource allocation, conflict resolutions and corrective actions, overall quality assurance for the project, establishing and maintaining relations with key external bodies and projects, collaboration with EGI-DS to ensure long-term sustainability plans are successful. All management tasks are accomplished by the coordinating partner, CERN, with the exception of the Quality Assurance which is the responsibility of the partner BT Infrastructures Critiques.

Users: More than 200 virtual organizations use the EGEE-Infrastructure, 152 of them are registered. The total number of registered users exceeds 16000. 15 application domains use EGEE. Though EGEE aims at both, scientific and business users, the vast majority of the users come from science.

User recruitment: It is not particularly hard to recruit new users. EGEE is very well known at least in Europe and recognized as "leading brand" in Grid technology (Interviewee 1). Most of the scientists interested in using a Grid, start by looking at EGEE.

The high profile of EGEE is a result of its leading position within the Grid community. Not only is EGEE one of the most mature Grid projects, it is also very well connected to other organizations and scientific communities. EGEE is driven by the needs of its user communities. Hence it has become an integral part of the workflow of not only single scientist but of certain scientific fields in the whole. In some fields, such as many research areas within particle physics, there is no question as to whether or not to adopt EGEE; it is simply standard (Interviewee 2).

EGEE aims to reach all scientific communities with the need for Grid technology. EGEE is still often considered as a particle physics project. EGEE III has taken much effort to correct this perception. It went into all scientific communities coming into question to use Grids. For instance EGEE offered training for interested parties. Often the activities to reach new user groups were accomplished by other organizations like DANTE and the National Research and Education Networks (NRENs). Additionally EGEE is very much embedded in activities of the EU

commission. Its achievements and benefits are communicated very well by the EU commission, the most important financer of research projects in Europe.

Drivers and barriers to adoption: EGEE is the biggest Grid project in Europe. The e-Infrastructure has reached production level and the middleware is mature. For many projects with a very high amount of data and need for large computational power only EGEE can satisfy the needs. Compared to other Grid projects (including the set up of an own, local Grid by for instance a university), there are actually only minor barriers to adopt EGEE for scientists.

Many potential users from industry are nevertheless deterred because their security needs can not be fulfilled properly. Data privacy may not be a big problem within many scientific communities but for private companies their compliance is vital.

Challenges in interdisciplinary collaboration: None mentioned by the informants or in the evaluated documents.

Collaboration with other organizations: As one of the most important and influential projects, the cooperation network of EGEE is wide. All significant Grid related projects worldwide are linked to EGEE somehow. The core tasks of EGEE are to develop the middleware gLite and to set up a community. Many of the sub-tasks needed are accomplished collaboratively. Among the most important collaboration partners are Géant2, DANTE and the European NRENs.

Technology

Main technologies, resources and services: EGEE provides all a Grid needs: It maintains and coordinates the e-Infrastructure and develops gLite. By now EGEE runs around 12m jobs a month on its computing cores. The overall number of CPUs available to users 24/7 is approximately 114000. Not considered tape MSS the total storage capacity available is 25 PB. These numbers are according to GStat and published on the EGEE web site.

gLite is the middleware stack for grid computing used by a very large variety of scientific domains. It provides a complete set of services for building a production Grid infrastructure. gLite provides a framework for building grid applications tapping into the power of distributed computing and storage resources across computer networks. The gLite services are currently adopted by more than 250 Computing Centres and used by more than 15000 researchers in Europe and around the world.

Role of technology development:

Data sharing: The amount of data produced by the LHC alone will amount up to 40 PB a year. A multiple of that amount will result from simulations run with the original data. Hence, the distribution of the data to scientists will be a continuous task of the e-Infrastructure.

The nodes are organized in a hierarchical structure. At the top, Tier0 is responsible for the storage of all the raw data, the first reconstruction of the data and the data distribution to the Tier1 level. Tier0 has a special role as probably the only repository where all the raw data can be found. However, data (both raw data and reconstructed data) will be copied to Tier1 centres to provide better access to data and to provide a degree of redundancy in case of failure. The twelve Tier1 centres are large computer facilities (regional or national) acting as repositories of the entire reconstructed sample. Their computing power is used to complement the Tier0 capacity, in particular, providing capabilities for successive reprocessing of the data and other very data-intensive analyses. The Tier1s act as data distribution centres and provide support to the lower levels. The Tier2 level is expected to provide a large fraction of the total available CPU for analysis purposes. Tier2 facilities (and possible lower Tiers down to level 4) should be able to effectively contribute computing power without having to locally deploy significant tape storage facilities, which are provided in a centralized way by the Tier1 (and Tier 0) levels (Lamanna 2004). There are approximately 150 Tier 2 nodes.

Interoperability with similar or connecting infrastructures: EGEE collaborates with all major Grid projects in the world at least informally. The EGEE Collaborating Projects Liaison Office is a point of contact for projects which are collaborating with EGEE, and facilitates the relationships between those projects and the EGEE activities. For many projects, the first step to collaboration is receiving a Letter of Support from EGEE to accompany their proposal. Other projects have drawn up a Memorandum of Understanding, stating explicitly what they will need from EGEE and what they will offer in return. Collaborative activities range from technical work on interoperability to community activities such as organizing joint training events and dissemination material. Depending on the issue at hand, different forms of co-operation might be appropriate and EGEE is open to suggestions and initiatives from any project that wants to help advance Grid computing.

Contribution

Main contributions of project EGEE contributed to establishing a Europe-wide and globally well connected Grid user community. Many of the contributions of Grid technology as a whole are also due to the efforts of EGEE. EGEE was and is one of the projects that bring e-Science to the specialist scientific communities. It has enabled many scientists to use the advantages of Grid technology without having to become experts in e-Science. A second important contribution is the development of gLite. This is a mature software with high usability. No special e-Infrastructure knowledge is necessary to adopt it. EGEE provides training courses to learn using gLite. There is also ample support for users. gLite is freely available and continuously updated and therefore attractive.

Challenges The follow-up project EGI has existed only as so called EGI Design Study since 2007. The tasks of the Design Study are to evaluate the requirements and use cases for the EGI, to identify processes and mechanisms for establishing an EGI, to define the structure of the EGI and to initiate the construction of the EGI organization. The actual project EGI has not yet been approved by the EU, and project funding is currently unresolved. Despite the fact that none of the interviewees anticipates that funding will not be granted, it is unclear what amount it will be. There remain, therefore, some uncertainties among the users of the Grid. These fears we be allayed when the EU commission has approved the funding which will be in November 2009 at the latest.

Informants' recommendations to policy makers

Suggestions:

- Continue to fund scientific communities' direct use of e-infrastructure to encourage uptake in the communities (e.g. INFRA-2010-1.2.3: Virtual Research Communities);
- Do not give funds to individual researchers or communities for computing equipment unless they accept to connect it to a shared infrastructure;
- Only fund software development by research communities if they agree to distribute it under an open-source license;
- Encourage convergence of e-Infrastructures by insisting on interoperation as a key objective for funding.

SWOT analysis

Table 4-18: EGEE strengths and weaknesses

	Strengths	Weakness
Long-term funding	Guaranteed for the next years. A	EGEE aims to get self-financed in the

	follow-up project is in the pipeline. Many institutions are involved in the project. If one drops out – except for CERN – this won't be a vital problem. Total investments so far have simply been too high to abandon the project.	long run. Even though the investments of the participating institutions are very high yet, financial support by the EC is still essential.
Sustainability	The use of the e-Infrastructure coordinated by EGEE has become an integral part of the workflow of literally thousands of users and hundreds of institutions. Even if the EC would drop EGEE there would probably emerge a surrogate very soon.	
User recruitment	EGEE is still often considered as particle physics project. EGEE III has taken much effort to correct this perception. It went into all scientific communities coming into question to use Grids. These efforts were successful and now there are user communities from 15 different scientific fields. Since EGEE has strong ties with specialist communities the communication of the benefits is very easy and institutionalized.	Up to now it was not possible to recruit a substantial amount of users from the industry. EGEE provides differing levels of secure data storage including an encrypted service (called Hydra) used by the life science communities. Business are reluctant to put valuable data outside their own enterprise resources and this is an issue just not for EGEE and grids but all external services (grids or clouds).
Involvement of current users		Since there are so many users, EGEE does not put much effort in "customer retention".
Organizational bedding	The organizational bedding of EGEE can't be any better. It is connected to virtually every Grid project of at least national importance. Many tasks of EGEE are done in cooperation with other institutions. An example is the communication with potential user groups via NRENs.	
Institutionalized links	One important link to other big projects is the fact that EGEE is funded by the EC like most of the leading potential user projects. EGEE has institutionalized links to most Grid projects and the NRENs.	
External use of software, tools	The gLite services are currently adopted by more than 250 Computing Centers and used by more than 15'000 researchers in Europe and around the world.	

Source: eResearch2020

Table 4-19: EGEE opportunities and threats

	Opportunities	Threats
Funding of member organizations	All member organizations are research institutions like universities. They may not lavish money but at least the long-term funding is guaranteed.	

Technology monitoring	EGEE is not only a cutting-edge project with a leading position within the community, it has also established institutions to enhance the cooperation with "competing" projects like the Globus Toolkit to set conjoint standards to harmonize and ease the interoperability.	
Competition with other infrastructures or technologies	A competitive technology might be cloud computing. EGEE has commissioned a study to compare the advantages and disadvantages. The study (Bégin 2008) comes to the conclusion that cloud and Grid computing may be integrated and are not competitors regarding implementation.	
Security risks	Within the scientific communities no problems are reported	Users from the industry can't adopt EGEE because of a lack of data privacy.
Change of user communities and fields	Diverse fields, changes cannot be projected.	

4.10 European Theoretical Spectroscopy Facility (ETSF)

Case Overview

What does the project do mainly? ETSF is delivering services by theoretical physicists to experimental physicist users. These services consist of theoretical support and code to analyse experiment results. In this regard, ETSF is a rather unusual variety of e-Infrastructure, its main asset not being a network or supercomputer but the brains of the participating scientists and the codes they develop and have developed in the past.

Motivations for setting it up: Briefly, the motivation is to reach out in a novel way to experimental physicist users, providing theoretical services to them by theorists. The problem motivating the creation of ETSF has been the recurrent inability of experimentalists to find theorists with state-of-the-art theoretical tools to do the necessary calculations for them.

The idea of the ETSF structure is to have a more reliable mechanism for the experimentalists to keep in touch with theorists who are able to work together on the same problems.

Main goals of the project: The main goal is to broaden access to the knowledge and expertise which have been built up in the field of theoretical spectroscopy across the public or private sector, bridging the gap between theoretical methods and real applicants. In particular, ETSF lists the following goals:

- Developing theory and methods: The ETSF groups, together composed of more than 150 researchers, extend the potential of theoretical spectroscopy by developing more efficient and more accurate methods and techniques.

- Developing scientific software: The ETSF offers several scientific codes that translate state-of-the-art methods into tools for studying the properties of real materials. Scientific programmers and software engineers support ETSF researchers in developing and providing efficient, user-friendly, and well-documented codes.

- Providing training in theoretical and computational techniques: The ETSF regularly organizes training events targeted at young researchers pursuing, or wishing to pursue, a career in the area of theoretical spectroscopy. This service can be extended upon request to other users, e.g. experimentalists, scientists working in

industry, or researchers working in a similar field. ETSF users can apply for specifically targeted training, for small groups or for a single person.

- Undertaking scientific projects on demand: In analogy to large experimental research infrastructures, such as synchrotron facilities, the ETSF users can propose projects for which scientific and technical support is provided by ETSF researchers.

Project maturity: ETSF has been working for 3 years. The first call for user proposals was in spring 2007 and subsequently twice per year. Scientifically it is well developed. It already has a wide range of capabilities serving a wide range of users. In terms of reaching out to potential users, ETSF is an early stage, however, but it is becoming increasingly well known in the community.

Project funding: ETST is funded through EC funding plus both national and local funding. EU funding is 3.8 million euro. By funder this is approx. €10M per annum of which about 60% is from ETSF institutions, 15% national and regional funding (mostly scientific research councils), 5% private sector, and 20% EU.

Organizational Structure

Size and composition: The ETSF Core includes ten nodes from seven countries. There are also six associate nodes.

Governance The ETSF has an administrative structure formed by a Steering Committee, a Governing Board and an Advisory Board. All ETSF activities are controlled by the Steering Committee which consists of representatives from the 10 core groups. The steering committee consists of one member from each research group and e-infrastructure has 11 partners. The decision making body is the steering committee.

The ETSF also has a set of working teams, who work more regularly together on specific aspects of the project. These teams are not localised geographically, but each team comes across different nodes of the project.

Managing internal and external relations

Management of the project: The ETSF is divided into 7 beamlines which are concerned with specific topics. A beamline coordinator is responsible for the contact with the users of each line. He/She also serves as the contact person for users who want to submit a proposal to the ETSF.

Users: Users are mainly experimental physicists, but also from neighbouring disciplines such as chemistry, material and earth sciences.

User recruitment: Generally, users can approach ETSF with a proposal to solve a particular problem. Experimentalists are invited to submit these proposals to the ETSF website. The proposals are evaluated twice a year by an external evaluation panel. Finally, nodes of ETSF are allocated to the project and work in communication with the users.

Drivers and barriers to adoption: N/A.

Challenges in interdisciplinary collaboration: Most of the scientists of the ETSF itself are physicists and a few of them are actually located in chemistry departments. The application of quantum mechanics to matter is on the border line between Physics and Chemistry. There are links with material science and device engineering and some of their users come from these communities. Further communities, to which the ETSF has yet to make better contact, would include molecular astronomers and geological scientists.

Collaboration with other organizations: N/A

Technology

Main technologies, resources and services: overview of available resources, technologies and services: High performance computing is much used in ETSF. There is specific use of supercomputing, particularly newly developed in Barcelona. Massive use is made of parallel computing, but these are located in single sites.

Role of technology development: The main asset of ETSF, apart from the underlying theories developed, is the set of computer codes which are developed. Most of these are able to run on very large parallel computers. This code simulates what is going on spectroscopic processes.

Data sharing: There is no massive use of external data. The services rather refer to ab-initio calculations, basic equations of quantum mechanics, which are purely arithmetical and theoretical. Data is brought in by the experimentalists.

Interoperability with similar or connecting infrastructures: N/A

Contribution

Main contributions of project: ETSF is structurally a new way to enable the collaboration of theorists and experimentalists. ETSF is producing around 160 publications per year, specifically high quality journals, and 150 to 200 scientists are invited to speak at the annual conference. The main expected contribution of ETSF for the future is to continue working successfully with users.

Challenges: The main issue is the difficulty of funding on the European scale. At the moment, ETSF has no funding in place after December 2010, when the e-infrastructure grant ends. European funding is the most restricting factor to the future.

Informants' recommendations to policy makers

While the e-infrastructure funding provides welcome funding for user projects and for developing the facilities for user projects, it is proving worryingly difficult to secure EU funding for European collaboration on the other essential strands of the ETSF: training young scientists (who will be the future ETSF scientists to work with users) and collaboration on fundamental science among the ETSF nodes, both of which were built up - to warm praise from the EC - under the Nanoquanta Network of Excellence project. This lack of funding jeopardises the future of the ETSF.

SWOT analysis

Table 4-20: ETSF strengths and weaknesses

	Strengths	Weakness
Long-term funding		Long term funding of ETSF is not secured.
Sustainability		Depending on funding
User recruitment	Institutionalised way of user recruitment	
Involvement of current users	Good rate of repeat-users	
Organizational bedding	N/A	N/A
Institutionalised links	Links to large parts of the repository landscape exist	
External use of software, tools	N/A	

Table 4-21: ETSF opportunities and threats

	Opportunities	Threats
Funding of member organizations		Rather than on engagement by member organizations, ETSF hinges on top-down funding
Technology monitoring	Technology here means codes, with a very probable longevity	
Competition with other infrastructures or technologies	ETSF is a very unique infrastructure that caters for a previously untapped research need.	
Security risks	None disclosed	None disclosed
Change of user communities and fields	Communities not yet served provide opportunities to expand the services of ETSF	

4.11 GEANT

Case Overview

What does the project do mainly? GEANT provides the European Internet Network for Research and Education. It connects 34 National Research and Education Networks and coordinates their inter-operability. It also links to a number of other world regions and so is at the heart of global research networking. GEANT operates the backbone, but also does research about networking technologies (examining the future of research networking and further developing its services), and support of users.

Motivations for setting it up: Each country had its own NREN in the early 1990s. From about 1990 on, they came together and started successive projects to build a pan-European capability to complement the national capabilities. Cooperation had also previously existed; there had been bilateral national cooperation in the 1980s. Precursors were EuropaNET (Trans-European Network), TEN-34, and TEN-155. For the Commission a major motivation is the building of the European Research Area (ERA). GEANT is however not part of the ERA (which is DG Research) but DG InfSo.

Main goals of the project: The main goals include the operation of the backbone, research about networking technologies (also examining the future of research networking), and support of users. Thereby gaps in networking provision should be closed, thus closing digital divides.

Project maturity: GEANT1 started in November 2000 and ran until August 2004; GEANT2 started subsequently and was supposed to end in 2008 but was extended to June 2009 when GEANT3 will continue the work. There are relatively smooth transitions from project to project.

There is usually a 5-6 year lifetime until there is a quantum change in the performance or approach of the technology, which the project then reflects and examines how to implement. At the moment the technology is working and is mature, but the next generation of networks will be doing new things. There was a steady growth in the performance of IT networks until 2004. In 2004 GEANT began to use overlay networks to multiply capacity by exploiting the optical characteristics of fibre. This lead to an explosion in capacities, with the possibility of having multiple networks on top of one another.

Project funding: The total budget amounts to about 40m Euro per year. Of this,

- ~ 50% are long term commitments, fibre, depreciation of equipment, maintenance,
- ~ 40% buying short term capacity (12 moth basis)
- ~ 10% project management (Dante)

Organizational Structure

Size and composition: GEANT2 consists of 30 national educational networks voting, Dante as coordinator, plus Terena as a European networking lobby organization, Terena and Dante are non-voting partners. A governance body called the NREN Policy Committee exists which holds 3-5 meetings per year, members from all partners are included. This committee is responsible for the day-to-day management, while the project is responsible for the technical coordination.

Governance: The project is split into work packages according to service activities: procurement, operations, development of services, global connectivity and spread out.

DANTE's purpose is to plan, build and operate pan-European research networks. It was set up, and is owned, by a group of NRENs. It was established in 1993, and has since played a pivotal role in five consecutive generations of pan-European research network: EuropaNET, TEN-34, TEN-155, GÉANT and now GÉANT2.

TERENA, the European association of research and education networking organisations, also has significant responsibilities within the project. It handles a number of the outreach activities, and supports the co-ordination of the research and development effort among project partners. In particular, it encourages the common exploration of new technologies between project partners and other groups that are active in technical development of particular relevance to research and education networking, through the continued operation of TERENA task forces.

The Policy Committee (often referred to as the NREN PC) consists of appointed representatives from each partner in the project. It meets at least three times a year, and is responsible for setting and overseeing overall policy. The Policy Committee Chairman is elected by the committee members for a two-year term.

The Policy Committee has appointed a second body to contribute to the management of the GÉANT2 project - the Executive Committee, which consists of a small group elected by the Policy Committee. It is primarily responsible for preparing the yearly work programme for the project, and for quality assurance and supervision relating to its implementation.

The Technical Management (direct management of technical activities within the project) is carried out by technical activity leaders. Technical activities are overseen by the Technical Committee

Managing internal and external relations

Management of the project:. See governance.

Users: Users are connected to the national NRENS and those using the network structure, which means that there are millions of users, many of whom are unaware that they are using GEANT.

As to specialized services: there have always been pan-European co-operations, and the aim of GEANT was to provide technical and operational support. Today they have resources such as virtual networks, e.g. the LHC has been given a 10 gigabit path to communicate with each other across national borders. From this point of view, there are unseen possibilities today that were not there before.

Another example is the very long baseline interferometry in radio astronomy: the larger the diameter of a dish the better resolution, with a maximum size limited to about 100 meters. As an alternative, smaller devices which are connected to each other, anywhere on the globe, aligned in same direction, bring huge amounts of data together. Today, with the help of GEANT networks, the advantage is immediate feedback of data which makes it possible to decide more quickly to deepen research on certain area, loop backwards to decision in days instead of months. The role of GEANT here is to provide the bandwidth needed, and it was included in the set-up phase. There is no interaction on a day-to-day basis.

User recruitment: N/A.

Drivers and barriers to adoption: N/A.

Challenges in interdisciplinary collaboration: N/A.

Collaboration with other organizations: N/A.

Technology

Main technologies, resources and services: The main technologies relate to the operation of networks and are packet rooting, circuit switching, transmission technology, DWDM, next generation SDH.

No supercomputing is provided, however it provides interfaces to the Grid layer (below Grid layer), generic monitoring measurement applications and security applications are provided that will often work in non-Grid environment as well. Not all can be categorised as Grid based.

Role of technology development: The own developments are limited to making technology work on the operational side.

Data sharing: N/A

Interoperability with similar or connecting infrastructures: N/A

Contribution

Main contributions of project: Without GEANT a significant bulk of pan-European research cooperation would simply not function. GEANT is a critical infrastructure for many projects.

GEANTs reach continues to expand, having begun in the early 90s with Western European countries, later adding the Czech Republic, Poland, Hungary, the Baltic states and Turkey, and Russia being the latest addition. Internationally, connections to Japan, China, Latin Africa have also increased.

The link to the US increased in capacity recently, with bandwidth quadrupling in effect. Also, the LHC has had a strong effect.

With regard to expected major future contributions, with a view to the next four or five years, the main objective for GEANT is to make services more available and accessible at high performance levels on a pan-European basis. Each of the national networks has a track record of developed national services. These services are to match one another, and should be recognizable as pan-European services. Research institutions today are keen to receive Europe-wide services. The aim is to make these services usable seamlessly across multiple domains, provided on a backbone by GEANT, while the services are distributed nationally. The provider characteristics have to match: GEANT's aim is to find a way to match these services each behind a services interface of the different national service providers.

Going back to the late 1980 or 1990s, just getting a connection at all was innovative. Today, it is the performance of these connections and the ability for groups of users to have virtual networks of their own that make the difference. To make that an efficient process, where people who are not primarily concerned with technology can easily utilize it, is a primary future goal of GEANT.

The webtool is now used in several European countries and many universities register. DRIVER's Open Access policy is especially valuable for smaller universities, giving them the opportunity to increase their visibility. About 225 institutions all over Europe use DRIVER.

There is strong community uptake and commitment. Sustainability is likely due to the establishment of a confederation. Libraries generally have an interest in taking and running the DRIVER service, and repositories are willing to conform to the framework.

Challenges: The major challenge can be seen in translating national domains into a common service portfolio. Instead of having just one technology provider on a single operation environment in a single office, GEANT works in a heterogeneous technological environment distributed across multiple domains and is aiming to make it appear as if it is just one big domain.

This is a major challenge for the future on the technology level, but also a challenge with regard to the service model, and the business case based on that service model. Networks are rather different in organizing and funding structures, some are more top-down, some involve government funding, others not. Therefore, there is a high degree of variability, with different demands for planning and different views upon services.

Also, national budgets run on a 12 months periods basis while GEANT sets up a 5 year budget plan which often makes modifications in the later years necessary for GEANT.

Another threat relates to the question in how far members are prepared to pay for solidarity of services. This is a question of commitment of different member states. Of course, every participating nation is trying to get a fair share, which is often difficult to decide upon. The problem is to some extent related to geography. Turkey is for example most expensive to connect (due to geography and market size), so success depends on solidarity.

Solidarity today is growing and threatened at the same time. It is growing in terms of size and reach, but endangered because with new countries joining the network, additional costs have to be borne by all members, including those who are not doing joint research projects with the new countries.

Trust is not reported to be a major concern at the GEANT level, the issue is related to the level of national commitment. Commitments on the European level, of course, to some extent restrict freedom of choice about how things should be done at a national level. The benefits of being part of the network are counterweighed by obligations.

Informants' recommendations to policy makers

Not covered

SWOT analysis

Table 4-22: Géant strengths and weaknesses

	Strengths	Weakness
Long-term funding	Being an indispensible and mission critical factor in the European research landscape, funding can be expected to be secured	Co-ordination among an ever growing number of participants.
Sustainability	Ditto	
User recruitment	Support of very large co-operative research endeavours will sustain the user base	
Involvement of current users	Integration with very large research projects	Invisible to end users
Organizational bedding	Integration of NRENs through DANTE and Terena	Some challenges as to solidarity
Institutionalised links	Strong (as core business)	
External use of software, tools	N/A	N/A

Table 4-23: Géant opportunities and threats

	Opportunities	Threats

Funding of member organizations	Being an indispensable and mission critical factor in the national research landscapes, funding of NRENs can be expected to be secured.	No objective measure of fair shares in funding contribution of partners, solidarity is crucial.
Technology monitoring	Driver of networking technologies	
Competition with other infrastructures or technologies	None	
Security risks	Undisclosed	Undisclosed
Change of user communities and fields	Due to little competition not to be expected	

4.12 MediGrid

Case Overview

What does the project do mainly? The MediGrid project was started in 2005 as part of the D-Grid initiative financed by the German Federal Ministry of Education and Research (BMBF). MediGRID consists of thee modules in representative arenas of biomedical research: image processing, biomedical computer science, clinical research. In these areas, each four methodical modules (middleware, ontology tools, resource fusion and eScience) are applied. Exemplarily applications from the three research areas were transmitted to the prototypical Grid structures. The aim of MediGRID was to develop Grid technologies for the testbed use in medical and biomedical research.

Motivations for setting it up: The project's primary motivation was developing Grid technologies. Medical and biomedical research was considered to be a field of application that could benefit to a great extent from the Grid computing technology. In a joint initiative with German research and industry, the Federal Ministry of Education and Research (BMBF) is funding the development of D-Grid. The first D-Grid projects started in September 2005 with the goal of developing a distributed, integrated resource platform for high-performance computing and related services to enable the processing of large amounts of scientific data and information. MediGrid is one of the so-called Community Grids which were developed for different research disciplines.

Main goals of the project: The main goal of the MediGrid project was to develop a Grid infrastructure and test it through several test applications. The main motivation was to develop a technical platform in which then different applications could be integrated with comparatively little effort.

Project maturity: At the end of the project period, the infrastructure has been established successfully and the first test applications have been run. The original project aim has therefore been achieved. However, during the lifetime of the project the main funding agency, the BMBF has been raising issues of sustainability. This objective could not be achieved, since the Grid infrastructure and the services based on it are in a very early stage of the market cycle. They lack yet the maturity to become financially sustainable. Nevertheless, several spin-off projects from MediGrid have been designed – some already funded, some awaiting funding – that are hoped to advance the services provided based on the Grid infrastructure to marketability.

Project funding: MediGrid project received around 5 million Euro from the BMBF. An estimated 80% of this funding is needed for personnel and the remainder (around 20%) are special investments funds for infrastructure costs.

Organizational Structure

Size and composition: There are eight partners to the MediGrid project, all of them are German public or semi-public institutions.

Governance The project coordinator is TMF (Telematikplattform für medizinische Forschungsnetze). The TMF office has been responsible for the organization of so-called „speaker-meetings", and for the exchange of documents between the project partners. The project is organized in 8 different modules (coordination, resource fusion, ontology tools, middleware, image processing, biomedical informatics, clinical research, e-Science).

Managing internal and external relations

Management of the project: The module leaders plus the consortium spokesman constitute the "speakers board" that meets quarterly to take decisions concerning the MediGrid project. The decisions concern mostly implementations issues. Since all project partners have submitted and been granted individual project proposals, they all have their specific project aims and implementation infrastructure. The "speakers board" then is responsible for the overall cooperation and presentation of the results

Since each project partners is fully responsible for their particular project, the motivation for cooperation has to be mostly intrinsic since there are no formal provisions to ensure cooperation and the decision of the "speakers board" are not fully binding. Consequently, the participation of the partners has been variably distributed due to different degrees of intrinsic motivation. As in many other projects with several project partners involved, the efforts undertaken by the partners depended very much on the persons representing the partner organizations.

Users: The aim of the project was to provide a testbed for services and rather than to provide fully-fledged services during the project time. Consequently, there have been rather low numbers of users that cannot be pinned down to any concrete number. Especially, there are no users completely external to the project. The subproject "Augustus", dealing with a genetic sequence analysis, had a slightly bigger (potential) user community since the tool existed and had been in use before, and the project only advanced it through the use of Grid technology. All in all, more users are expected for the spin-off projects being implemented after the end of the MediGrid project.

User recruitment: N/A, no external users have been recruited.

Drivers and barriers to adoption: N/A, no external users have been recruited.

Challenges in interdisciplinary collaboration: The project staff includes around 40% computer scientists and other engineers, 40% computational biologists, and additional each 10% of purely medical staff, and economists and other. All take on different functions according to their field of expertise. Difficulties concerning the cooperation resulted not primarily from the interdisciplinary compositions of the project teams, but stemmed mostly from the cooperation between different partner organizations. Since no enforcing measures existed, some partners were more active than others. All in all, the communication is estimated to have worked well, for some applications better than for others

Collaboration with other organizations: N/A

Technology

Main technologies, resources and services: overview of available resources, technologies and services: Software: MediGrid has not been devised as a Data Grid but as a Computing Grid. For this purpose, the project team has tried to use and adapt existing open source software, such as Globus and Unicore. A more complex middleware used was "Workflow Manager", and "Gwess", an in-house development of Fraunhofer Institute that was further developed and adapted for project purposes.

The website portal has been devised as the unique entry point for all applications and has actually been successfully used by all MediGrid applications. The software used was the Open Source "Gridsphere" by Windows. OGSA-DEI has been used as the main software for data integration, the standard for image data was Dicom. During the course of the MediGrid project this standard has been updated with additional security features and developed to a new standard "Grid Dicom".

Data/storage: The "Biomed" application of MediGrid has used internet-based genetic databases, while the image data used for other applications has been data accumulated for and by the project partners. All data used in MediGrid has been stored locally, which means that it has been stored at the sites of the computing centres. However the local data storage did not mean that all project partners were allowed to access the data.

Role of technology development: See above for Gwess.

Data sharing: See above.

Interoperability with similar or connecting infrastructures: N/A

Contribution

Main contributions of project: The main contribution of the project is the software developed that can be used for future projects. One of the technical contributions of MediGrid was the development of the enhanced "Grid Dicom" image data standard.

The experience gained as to how such a cooperation based on Grid technology can work, can also be seen as a major outcome of the project that should be useful for future co-operations, especially with a view to policy mechanisms and regulations.

Concerning the geographical scope of the project, one of the benefits of the project was the dissemination about the MediGrid activities on a global level. Especially the awareness level in the United States has been raised. This enhanced knowledge about the project has been achieved mainly through the membership of TMF in the healthgrid.org (Europe) and through a comparatively close and successful cooperation with caBIG, a big health Grid program based in the United States. In addition, TMF has initiated a "Forum Grid", as an exchange platform for all actors involved or interested in health grids. The MediGrid project has contributed significantly to this higher level of interconnectivity and networking that can be perceived today.

Challenges: A principal challenge seems to be the project structure that results from funding requirements. Since each partner is only responsible for a specific subproject, it becomes difficult to motivate the partners to cooperate in view of the overall project objectives and requirements.

Informants' recommendations to policy makers

Our informant recommends that cost structures of public research institutions in Germany become more transparent, so as to facilitate the budgeting and billing of services between project partners. This higher transparency is seen as an absolute necessity in order to envisage and achieve financial sustainability of similar research projects. The lack of transparency concerns, in particular, indirect costs, but staff costs are also not always calculated realistically.

In addition, the legal regulations concerning the use of patient data should be revised. Up to now it is near to impossible to transfer or share non-anonymous patient data via a network or Grid due to legal uncertainties. The data transfer is possible inside one institution or between two institutions, but difficult when more actors are involved. The biggest problem in this case is the need for the patients' consent when his/her data is transferred to a different place. Even if the data concerned is made anonymous, the federal government's ethics board has to be consulted first.

Another difficulty is related to funding: in Germany in the IT sector data processing systems (hardware) are financed comparatively often and easily. However, the operating costs and overhead expenditures are financed far less often. This selective funding encourages organizations to buy their own hardware instead of using shared resources that they cannot

bill on project budgets or receive funding for. This means that at the time being "normal" projects cannot receive funding for the use of Grid Infrastructure.

Another problem encountered by MediGrid were differing legal environments – in Germany due to the federal structure, the exchange of computing capacity between federal states proved difficult. On an international level, differing legislation concerning data protection might also hamper (medical) Grid projects.

SWOT analysis

Table 4-24: MediGrid strengths and weaknesses

	Strengths	Weakness
Long-term funding		Not secured, rather follow up projects
Sustainability	Knowledge spills over to follow-up projects	No user base beyond the project
User recruitment		Ditto
Involvement of current users		
Organizational bedding	Project well anchored in participating organisations	Little integration between partners
Institutionalised links	Good links to international Grid projects	
External use of software, tools	Use in follow up projects	

Table 4-25: MediGrid opportunities and threats

	Opportunities	Threats
Funding of member organizations	N/A since project has ended	N/A since project has ended
Technology monitoring	N/A since project has ended	N/A since project has ended
Competition with other infrastructures or technologies	N/A since project has ended	N/A since project has ended
Security risks	None disclosed	None disclosed
Change of user communities and fields	N/A since project has ended	N/A since project has ended

4.13 National Virtual Observatory (NVO)

Case Overview

What does the project do mainly? NVO develops standards and protocols to support astronomical analysis of multiple types of large-volumes of celestial data from disconnected astronomic ground and sky instruments.

Motivations for setting it up: The concept for NVO arose at a meeting of the Decadal Survey Panel on Theory, Computation, and Data Discovery in 1999. In the following two years, a series of workshops and conferences were held to flesh out this concept. These developments received considerable encouragement from the National Academy of Sciences Decadal Survey, which positioned the idea of the VO as a top priority for small astronomy projects—i.e., those funded up to $100 million over ten years. Both the panel and NVO pioneers envisioned a new approach to astronomy, shifting from observations of small samples of objects limited to one or a few wavelength bands, to studies based on multi-wavelength avalanche of data that consists of billions of celestial objects. As a part of other research in data Grid technologies, computer scientists were interested in NVO as an opportunity to further develop data representation and interoperability schemes.

Main goals of the project: NVO is mandated to develop and provide the e-Infrastructure and the interface capable to support the integration, or federation, of various astronomical digital data sources from diverse instruments. A major component of this virtual telescope is to enable efficient processing and visualization of these massive amounts of data. More broadly, the project is spearheading a fundamental change in astronomy—from the study of a single to multiple wavelengths.

Project maturity: Starting in the 2001—and later as part of a network of peer projects around the world, the International Virtual Observatory Alliance (IVOA)—the NVO is one of the most mature e-Infrastructures. At the same time, since the objective of the project has centred on developing the foundational technologies for this infrastructure, not on attracting or serving users, it may be considered immature in these respects.

Project funding: In September 2001 NSF's Information Technology Research program awarded $10M for a 5-year period. Along with funding from NSF's astronomy division, this award has been extended until September 2009 with additional funds, totalling approximately $14 million in overall allocation from the start of the project. Currently, most of the funding is for software developers. In its next stage of funding, NVO is promised to receive $36 million for the next five years, which will enable meeting this more comprehensive target.

Organizational Structure

Size and composition: NVO involves participants from 17 US-based institutions. These institutions include astronomy data centres, national observatories, supercomputer centres, university departments and computer science specialists. The project's staff consists of 13.5 full-time equivalent positions, which are shared by 51 people.

Governance: An Executive Committee sets project priorities and allocates resources. It is comprised of the PI, a co-PI, NVO's project manager, project scientist, and chief architect, as well as other senior personnel. The Executive Committee works closely with Technical and Science working groups on prioritizing developments.

Managing internal and external relations

Management of the project: Work is divided across project teams of astronomers and astronomy data experts, computer scientists and information technology experts, as well as

individuals who focus on training and outreach. Computer scientists and information technology experts develop the software that is the basis of the NVO's e-infrastructure. Astronomers guide these developments from the scientific perspective and considering the needs of future constituents. And outreach specialists are responsible for information dissemination.

Users: Informants consider the entire astronomical population as potential users. However, in its current formant, the NVO is not aiming to engage a large user community; it is more of a research and development facility. Beyond polling number of hits on the project's website there are no data on actual number of users. Informants' rough guesses indicate over a hundred users, but this figure is not based on a concrete measure. Individual data centres are able to tell who is accessing data through their own websites, as opposed to virtual organization data requests. One measure of usage is that at the Space Telescope Science Institute, they see about 20-30% of online requests for data arriving via NVO requests.

User recruitment: Some of the NVO's core users have been recruited through their role as scientific advisors on the project, which likely gave them a high degree of exposure to various technical details and has likely made adoption easier for those individuals. NVO also holds summer training workshops and recruits users in this way; these users may then go back to their home institution and train other staff and researchers on the interface as an in-house expert. Other people become users as they start using NVO to find data for class projects as graduate or even undergraduate astronomy students. Aside from those who find the NVO through a web search, knowledge about the NVO has circulated through "word of mouth"—i.e. social networks—primarily within a home institution.

Drivers and barriers to adoption: An important driver to adoption is that there are no other options aside from the NVO to analyze multiple wavelength data across a large-number of diverse datasets. Should investigators wish to access even one more format or one more dataset, it would require them considerable time investment and having deep knowledge in not only the astronomical principles but also the computational technologies. At the same time the present format of the NVO is obscure to most astronomers. Perhaps more challenging, the dominant paradigm in astronomy involves a study of a single wavelength. NVO has initiated several user recruitment strategies, including the lowering of its technological complexity and increasing training.

Challenges in interdisciplinary collaboration: There are profound cultural differences between astronomers and their computer science collaborators. Computer scientists aspire to develop software that is "cool and cutting edge." As a result, developments suffer from what is colloquially known as "feature creep," in which designers continuously add software features to make it universally applicable. More concerned with diffusing an integrative wavelength paradigm in their scientific community, astronomers, on the other hand, require simple and efficient designs that would enable scientists to address particular research problems and facilitate the use of the NVO e-Infrastructure. Among the strategies currently pursued is to work more closely with users so that development would be more related to their research needs, and aiming to simplify interfaces rather than adding unnecessary features.

Collaboration with other organizations: NVO assumes a central role in a densely connected network of international e-Infrastructure astronomy projects called the International Virtual Observatory Alliance, or IVOA. Members of the NVO regularly attend IVOA meetings, chair working groups and exchange information with their international peers. NVO also collaborates with commercial organizations, which provide useful in-kind sponsorship. They are currently working with Microsoft Research on the Worldwide Telescope program, as well as Google Sky, where NVO participating organizations are providing datasets.

Technology

Main technologies, resources and services: The NVO includes the following several core services that are being continuously updated. Presently the list includes DataScope, Open SkyQuery, and the Registry (renamed Directory), plus an Inventory tool, a data mining tool called VIM, and a command-line interface for scripting called VO-CLI. Grid middleware provides NVO users a distributed high performance processing facility of federated astronomical data.

Role of technology development: In its current phase NVO is mostly developing technologies and standards that enable the integration of diverse large-scale astronomical datasets. While data do not reside in the NVO, middleware layer (e.g. Globus and web services)—or the "collective layer"—supports authentication and access to data, as well as distributed computing, data, and visualization capabilities. A user layer enables users to query registries in search for data. Further tools the NVO develops provide workbenches, portals and procedures that present NVO capabilities to the end user. Among these tools is the OpenSkyQuery tool, which provides data from ten astronomical surveys. Other layers relate to technologies, standards and protocols for data discovery, delivery, and management, which would make it easier for tool and search developers beyond the NVO.

Data sharing: As other e-Infrastructures data is epicentre of the project, the actual data always reside in the home institution. The NVO does not curate, or store its own data—but it does provide several important features. One is to provide a capability to access a logical federation of astronomical data, as well as to analyze it, and while they do not try to fill the role of the "data police," they do try to make sure that the data descriptions are accurate, an important concern in making this data useful for researchers. Much of the other NVO activities involve developing protocols, specifications and standards concerning data format that would enable its federation.

Interoperability with similar or connecting infrastructures: Both the NVO and the IVOA have worked for about eight years to develop common metadata standards and models that would make integration more transparent to the users. Prominent among these developments is the joint specification among the NVO and AstroGrid of a VOTable, which is an XML formatting standard for astronomical tables. Another notable development includes an Astronomical Data Query Language that was completed through IVOA, and is a syntax of the query language that follows SQL standard with extensions for astronomical coordinates and regions. The newer tools for translating data into the same coordinate system or other frame of reference have been a welcome timesaver for research astronomers.

Contribution

Main contributions of project: On the scientific front, the NVO advances a new paradigm in astronomy that integrates high-quality, homogenous, multi-wavelength data on millions of objects from different observational sources. Supported by the NVO, this new paradigm has already led to several discoveries. A second, more practical contribution is to enable astronomers from institutions not directly connected to observatories, or from teaching-oriented institutions, to still work on research. These individuals are able to download NVO data and could even work on research using a laptop at home. This aspect of the NVO is likely to benefit astronomers in other, less developed countries as they would easily gain access to research resources. A related efficiency is that NVO allows for easier collaboration across co-PIs on a project, who in the past may have had to set up a website or mail data to share with the other researchers. The common standards set by NVO have also created greater efficiency for astronomers, particularly through the sharing of tools and techniques used to translate data coordinates between different coordinate systems. Other benefits include the engagement of commercial partners that include Microsoft and Google.

Challenges: Engaging users with still relatively complex tools is likely to continue to be challenging. However, funding is likely to cease unless large numbers of astronomers will find the NVO of immediate relevance to their work. In addition, although software development and specifications are robust and could be used by other entities who want to pick it up there are no plans to continue developments or provide support should funding stop.

Informants' recommendations to policy makers

Informants offer several recommendations to policy makers. They suggest that policy makers should be aware that tool development is directed by user needs, not just computer science priorities; they propose that stable funding stream would support more focused staff; find ways to acknowledge and reward the work on an e-Infrastructure that increasingly turns to be, or "under the hood" for user scientists.

SWOT analysis

Table 4-26: NVO strengths and weaknesses

	Strengths	Weakness
Long-term funding	Long-term funding is secured	
Sustainability	Funding for the next round of the NVO, called Virtual Astronomical Observatory (VAO) is secured. VAO will focus more on engaging users and providing to them production facility.	
User recruitment	There is a strategy for recruiting new users.	The strategy is not very helpful when the design is cumbersome, or when astronomers are unwilling/unable to analyze multi-wavelengths across datasets.
Involvement of current users	The NVO does not target users at the moment. A small core of users who are directly engaged with the project play more of an advisory role.	
Organizational bedding	The NVO is important to participating institutions, especially to astronomical ones (as opposed to computer science). This group is likely to pick up some of the development made in the project, should funding stop. They will also continue to take part in the project in its next round.	
Institutionalised links	There are good, ongoing relationships with IVOA, and some degree of collaboration with e-Infrastructure providers, such as TeraGrid.	
External use of software, tools	NVO's collaboration with IVOA on standards as well as technologies ensures that developments made in the US will be used elsewhere. Commercial partners have also used this work.	

Table 4-27: NVO opportunities and threats

	Opportunities	Threats
Funding of member organizations	Participating organizations are universities and national labs that are not dependent on transient funding.	
Technology monitoring	NVO involves renowned experts on distributed academic computing as well as astronomy experts. However, as noted above, there is a clash between computer scientists and astronomers resulting from over exploring technological opportunities, and not being focused on delivering a simple to use production facility.	
Competition with other infrastructures or technologies	There is no discernable competition.	
Security risks	While implementing commonly used authentication and other security mechanisms, the type of data used does not require extensive security measures.	
Change of user communities and fields	NVO is promoting a new research paradigm in astronomy. Informants note that this new paradigm is well regarded, but it needs to be backed by a robust and easy to use e-Infrastructure.	

4.14 Open Grid Forum (OGF)

Case Overview

What does the project do mainly? OGF serves as an open forum for participants to develop standards, specifications and recommendations for distributed e-Infrastructure computer technologies.

Motivations for setting it up: OGF started in 1999 as a grass-roots effort of computer scientists in US national labs and research universities to serve as an open forum for setting standards concerning developments in Grid computing. The OGF is a result of previous mergers with e-Infrastructure regional standardization bodies in 2000 and then with the Enterprise Grid Alliance, a commercial spin-off, in 2006.

Main goals of the project: The OGF is the main standardization body of e-Infrastructure. In addition to writing technical specifications and recommendations, the OGF also advocates the uptake and diffusion of "applied distributed computing" in various fields in science and industry. This is a shift from the previous narrower focus on of the OGF on Grid computing.

Project maturity: Operating for almost a decade, the OGF is not only one of the most mature e-Infrastructure actors, it is also one of the most influential. During this period thousands of individuals from hundreds of organizations participated in 31 meetings around the world, and over a hundred documents were published through the work of dozens teams. While the OGF has passed its peak, it dynamically updates its mission and scope to accommodate rival alternative technologies, most recently cloud computing.

Project funding: Undisclosed amounts have been received from individual and organizational membership, as well as from sponsors. The annual budget is estimated to be somewhat less than $1 million. Informants note that the overall budget has substantially declined in the past years, primarily as a result in diminishing interest in Grid computing. The budget is used to operate three meetings per year, and to provide administrative and organizational support to the ongoing work of dozens of groups and committees.

Organizational Structure

Size and composition: OGF is an open forum that joins various academic, commercial and government/non-profit organizations, all being key players in the diverse fields of e-Infrastructure. Assessing the "size" of the OGF is a difficult task. One of the complicating factors is that the OGF staff is very small—currently a full-time Director and an assistant to the Director—while there are several other position holders, the main organizational strength comes from those that can be more appropriately described as stakeholders, or constituents of the OGF community. These individuals contribute their time to standardization activities, study and share community practices, advocate the OGF and its aims, and to attend meetings. All in all, the OGF has engaged thousands of people and hundreds of organizations from around the world to take part in these efforts. One of the cited measures of the OGF's size is conference attendance. However, this figure can be misleading because the number of attendees considerably fluctuates—from around 100 to 800—based on the location of the meeting, whether the meeting is held jointly with another organization, or due to a declined interest in Grid computing. Informants estimate the core, stable group of its participants to be 150-200.

Governance: OGF's decisions are based on a system the IETF has pioneered, called "rough consensus." In this decision making model individuals in a group are encouraged to reach consensus, and when a certain opinion appears to prevail, roughly reflecting most

participants, it is selected as the decision. Many of the decisions in the OGF take place in the three main "functions" of eResearch, enterprise, and standards. Each of these functions are divided into "Areas" and then further broken down into more specialized "Groups," each consisting of a handful up to dozens of participants.

Managing internal and external relations

Management of the project. Work in OGF groups as well in most of the organization's governing bodies is voluntary, but it often requires committing substantial efforts and time, including attendance at multi-day face-to-face meetings, participation in weekly telephone meetings and contributing in other forms. At the same time, several motivating factors encourage participation and contributions to the OGF. These factors include organizational advantages. Learning is one of the most salient advantages. Firms and e-Infrastructure providers gain access to cutting edge knowledge about technologies and developments in peer organizations—thus influencing internal research and development activities. Equally important, organizational representatives benefit by influencing the overall direction the field of Grid computing is taking. Marketing is a third benefit. OGF provides participating organizations exposure to hundreds of other organizations, which may lead to collaborations or the engagement of new users. Individuals benefit from the recognition of their peers to their work. The ability to interact with many of their peers—some of them luminaries in IT and e-Infrastructure more specifically—assist in recruitment and career development. It is therefore not uncommon to see individuals moving to work from one e-Infrastructure to another or to a firm. Nevertheless, interest in Grid computing has substantially declined in recent years, particularly in the commercial arena. While several years ago there were slightly more academic participants than commercial ones, interviewees estimate the current commercial portion at about 30 per cent.

Users: Users are individuals and organizations who adopt a specification that the OGF published. Over the years groups in the OGF generated dozens of specifications that vary considerably in their scope—some being much more particular than others while others may apply across niches in computer science/information technology. Partly because of this technological breadth the OGF does not measure the adoption of its specifications; instead, groups may survey possible users, or contribute their knowledge about the uptake of a certain standard.

User recruitment: The open structure of the OGF, which includes diverse organizational participants, enables an expansive reach that includes constituents from the finance industry to telecom, to commercial and academic life sciences, high-energy physics to the social sciences and humanities. OGF's management considers overall strategies to engage users in new communities or in communities that have decreased their interest in grids—by encouraging the establishment of new groups for example. Each of the various OGF groups also works to actively solicit adoption in a respective community they cater to.

Drivers and barriers to adoption: A somewhat cynical way to consider the motivation for the adoption of OGF standards is their close to monopoly on the field of e-Infrastructure. Another adoption driver is the availably of existing distributed infrastructures that are based on OGF standards. If newer e-Infrastructures wish to interoperate with more established ones, they need to also adopt a set of OGF specifications. Barriers to adoption include complexity and the availability of alternatives. The emergence of virtualization and cloud computing have shifted the interest of large IT vendors away from Grid systems to these alternatives, as they provide more immediate commercial benefit.

Challenges in interdisciplinary collaboration: N/A

Collaboration with other organizations: OGF is based on collaboration with and among organizations. In addition to bringing together diverse stakeholders in applied distributed

(Grid) computing, the OGF also has relationships with comparable organizations. Appointed liaisons maintain formal relations with to all related major standard bodies in the computing industry. These relationships are also manifested informally through individuals who participate in the OGF as well as other standardization organizations.

Technology

Main technologies, resources and services: N/A

Role of technology development: N/A

Data sharing: data used in the project and the practices and challenges in sharing data: N/A

Interoperability with similar or connecting infrastructures: Since interoperability is the essence of standards, many of the specifications OGF groups generate are focused on interoperability. One notable effort is the work of the Grid Interoperation Now community group (GIN)—and coming out of it, the Production Infrastructure working group—that aims to identity and then develop a set of standards to enable major e-Infrastructures to work with one another. To date, these efforts have achieved partial interoperation.

Contribution

Main contributions of project: An important contribution of the OGF has been the crucial role it has played in both demonstrating the effectiveness of the open model of standardization, with more direct contributions to participants from the distributed computing community. Fostering an open community, enabled commercial and academic organizations to obtain the most recent information about Grid computing and to receive feedback on their developments. Compared to other fields in information technology, the organizational inclusiveness that the OGF has fostered has kept the field truly open, making single vendor dominance less feasible. The work in OGF groups has also offered e-Infrastructure providers with necessary standards that undergird their operation. Some consider the main contribution of OGF specifications to go beyond the e-Infrastructure community into other related areas, such computer clusters. Thinking forward, interviewees have pointed that the emergent cloud technology suffers from relatively weak management capabilities across clouds—a problem the OGF has arguably already solved. The most recently established OGF group, Open Cloud Computing Interface Working Group, will seek to connect this past work with current developments in cloud computing.

Challenges: It is increasingly apparent that Grid computing and thus the OGF are highly sensitive to the development of alternatives, most recently cloud computing.

Informants' recommendations to policy makers

Informants advocated two positions to policy makers. The first suggested implementing mechanisms that would ensure the continual involvement of participants in OGF activities—for example by funding agencies' encouragement to its grant recipients to participate in and contribute to the OGF activities. Without these extrinsic motivations, suggested informants, e-Infrastructures will have a hard time to find a place that would allow them to learn about successes and challenges from their peers around the world. A second, somewhat contrasting view, suggested that policy makers should start considering models that would move existing e-Infrastructures to more recent technologies, such as clouds.

SWOT analysis

Table 4-28: OGF strengths and weaknesses

	Strengths	Weakness
Long-term funding		Funding has gradually declined and if the OGF will not manage to connect its work to such technologies as Cloud computing, it is not likely that the organization will be viable in the long-run.
Sustainability	N/A	
User recruitment	The open, network structure of the OGF engages diverse communities through face-to-face meeting, marketing and in other means. However, competition from alternative technologies weakens these innate advantages.	
Involvement of current users		OGF manages to engage its current users—that is, the adopters of its standards mainly by the fact that they have little other choices if they are e-Infrastructure providers. Commercial users are much more challenging, since they often follow technological fashions that diverge from grids. Preliminary efforts are taking place to connect grids to cloud computing, but it is too soon to assess their potential.
Organizational bedding	Grid computing, and the work of the OGF is well recognized and facilitated by the vast majority of e-Infrastructure institutions.	
Institutionalised links	Cultivated for a decade, OGF is a quintessential cooperation network among diverse organizations in the e-Infrastructure ecosystem—including all major e-Infrastructure providers.	
External use of software, tools	Specifications and recommendations OGF developed are being widely used, particularly in academic e-Infrastructure providers, but also by commercial providers of applied distributed (Grid) computing.	

	Opportunities	Threats
Funding of member organizations		Funding to Grid activities within commercial participating organizations has reduced considerably and a similar trend is found in the US academic field.
Technology monitoring	Through the work of dozens of groups and committees, new technologies are identified and considered to be incorporated or addressed by the OGF.	
Competition with other infrastructures or technologies		Each e-Infrastructure has its own middleware system. Being a highly complex technology, even the devoted work of groups such as GIN (see above), can only support partial interoperation.
Security risks	N/A	
Change of user communities and fields		Again, this largely depends on the degree to which new technological paradigms—such as cloud computing— will be adopted, both in commercial, as well as in academic e-Infrastructures.

Table 4-29: OGF opportunities and threats

4.15 Open Science Grid (OSG)

Case Overview

What does the project do mainly? OSG provides, supports and enable scientists more effective sharing and utilization of available compute cycles in participating organizations, and to more easily use distributed storage and software through its "opportunistic computing" model.

Motivations for setting it up: Two related developments promoted the establishment of the OSG: (a) the need for a worldwide distributed Grid computing infrastructure for the high-energy physics (HEP) experiment, the Large Hadron Collider (LHC); (2) the development of data Grid projects that focused on the providing distributed data solutions for HEP.

Main goals of the project: OSG has aimed to make collaborative scientific research more effective and widespread, stimulate new and transformational approaches to computationally based scientific discovery, and build intellectual capital for future scientific research relying on distributed cyber-infrastructures. Informants also mentioned their goal as promoting and enabling—by partnerships and interoperation—a truly global Grid, a "Grid of grids." Facilitating the vision of e-Infrastructure, members of the OSG envision such a comprehensive e-Infrastructure fabric to transform the practice of collaborative science, making it more effective and widespread.

 Project maturity: While the project officially launched in 2006, and it released its first major middleware version last year this is a mature project. OSG staff and related personnel have engaged in ongoing e-Infrastructure developments for almost a decade. To date, over 2,500 unique users have shifted or processed data using the OSG infrastructure. Undergirding the OSG is mature technology: Condor cycle scavenging solution developed since the late 1980s and is now a part of the Linux Red Hat operating system; the Globus toolkit has been under development since the mid 1990s and is the most prominent middleware technology not only in the scientific community but is also used by various IT computer vendors; and the Virtual Data Toolkit (VDT) has been developed since early data Grid projects in 2000.

Project funding: OSG is supported by $30 million of joint funding from the Department of Energy (Scientific Discovery through Advanced Computing program) and the National Science Foundation, for an initial five-year term.

Organizational Structure

Size and composition: OSG is organized as a consortium that presently consists of 53 academic and research institutions, mostly from the US. The main stakeholders are HEP experiments, major Grid technology development projects and national laboratories. OSG is also a collection of dozens of virtual organizations—research and development groups in various fields of science. The organization includes 34 full time-equivalent staff positions, which are handled by about 50 people in 16 institutions throughout the US. OSG staff support virtual organizations (VOs), and site administrators and ensure robust secure operation, engage new users, and conduct outreach activities that include training sessions, workshops and regular newsletters.

Governance: Representatives from central OSG stakeholders take part in the Council. The Council is responsible to govern the consortium and to ensure the provision of services in accordance with OSG's scientific mission. The Council holds monthly teleconference and face-to-face meetings. In addition, an Executive Board directs the OSG's program of work, write policy and represent the OSG Consortium in relations with other organizations and committees. Informants also noted that trust, which lubricates any long-term economic and

social interaction, is an essential ingredient in facilitating decision making and participation. While for short-term projects trusting of external members is a challenge, the longer duration of the OSG project and it gradual evolution has enabled sustained interpersonal and organizational relations. In addition to time, several mechanisms help promote trust, such as the structure of the consortium, which is based on consensual decision making.

Managing internal and external relations

Management of the project: An Executive Director oversees the daily operation of OSG. OSG staff designs, operates OSG core middleware, supports users and work with software providers to solve problems users and administrators report. In addition, staff conduct outreach and training activities to engage potential users from new research communities in the OSG.

Users: Participation in the OSG consortium is based on an acknowledgement that relaxing the grip on internal organizational resources will offer more value and benefit over the long run. To date, the consortium has served about 2,500 unique users who have moved data or have run computations on the OSG infrastructure. Users are a part of diverse fields of science, such as theoretical physics, industrial engineering, computer science and natural language processing, chemistry, biochemistry, computational biology, genetics, structural biology and economics. However, throughput cycles from these diverse communities amounts to only about 10% of the overall usage; HEP is the "heavy" user, consuming all other resources. OSG delivers capacity to the LHC, as well as the Tevatron CDF and D0 HEP experiments.

User recruitment: OSG has established a dedicated function called the "engagement team" that operates in a unique three step procedure: user identification through conferences, talks, newsletters and public announcements; engagement fosters close relations with a carefully study of each new user's research environment and adapting the OSG infrastructure to it; contagion is the last stage in which many others within a community follow the examples set by scientist working with the engagement team. There have been several successful examples, such as structural biology.

Drivers and barriers to adoption: Often described as the hallmark of the OSG, "opportunistic computing" is one of the main contributions the Consortium provides to its members. Resources made available to OSG by contributing members, which enables the automatic facilitation of compute cycles and storage in times when there is a need for resources beyond those locally available, in an automatic fashion and with a low overhead. Since the opportunistic model is based on facilitation of resources contributing members do not use, the cost for each contributor is negligible while the aggregate benefit substantial. Barriers include the lack of understanding or knowledge of the complex technologies that undergird the OSG infrastructure. However, the work of the engagement team serves to ameliorate this challenge.

Challenges in interdisciplinary collaboration: One of the difficulties has been the apparent cultural difference between computer scientists and computational physicists. Computer scientists are more concerned with performing—and publishing—cutting edge research on computer technology, but care much less about service provision. At the same time, IT physicists who serve big-scientific experiments are more used to building, managing and sustaining large-scale infrastructures.

Collaboration with other organizations: OSG is a central hub in a global network of e-Infrastructure organizations. As such, it fosters various relationships with three communities: e-Infrastructure providers, Grid middleware development groups and HEP experiments. Providers: OSG has strong relationships with its EU counterpart EGEE. Liaisons represent EGEE and TeraGrid—the other US large-scale e-Infrastructure—in the OSG council. Developers: Two of the main e-Infrastructure middleware developers, Globus and Condor occupy primary positions within the OSG, including in the Council and Executive Board. HEP experiments:

Much of the OSG is geared to the provision of services to the US-based HEP community. Among these constituents are CMS and ATLAS, the two largest experiments in the Large Hadron Collider (LHC).

Technology

Main technologies, resources and services: OSG provides access to and sharing of the set of autonomous processing and storage resources through operations of a coherent facility. The OSG provides common, shared services including monitoring, accounting, security, problem reporting and tracking, towards the goal of operating a robust, effective system. Additionally the OSG provides a common, shared integration and validation facility, and provides functional, performance and full-system testing of new releases of software, services and applications. Software, includes, but is not limited to, the Virtual Data Toolkit (VDT), provides the technologies used by OSG as well as other equivalent infrastructures, such as the TeraGrid and the EGEE. Each project, including the OSG, augments the VDT with specific configuration scripts and utilities for its own environment and users. The OSG provides software repositories from which the packages can be downloaded, installed and configured on processing, storage, VO management or user client computers.

Role of technology development: While the OSG does not develop software, the facility is used as a platform for developing and testing distributed system technologies. The software developers collaborate as members of (and through the extensions activity sometimes receive contributions from) the OSG. OSG staff: release, deploy and support, integrate and test new software at the system level; supports operations of Grid-wise services; provide security operations and policy; troubleshoot end to end user and system problems.

Data sharing: N/A. Shared data is provided through the Storage Resource Manager interface and GridFtp. Allocation of shared storage through agreements b/w sites and VOs facilitated by OSG; also dCache for large scale, high I/O disk caching system for large sites; DRM — NSF based disk management system for small sites. Virtual Data Toolkit to enable access to and use of the ensemble of processors and storage of 20-30 petabytes tertiary automated tape storage at 12 centres world-wide.

Interoperability with similar or connecting infrastructures: OSG works to bridge its infrastructure and services with other grids – from campus, state and national grids, to international and worldwide community infrastructures. Such bridges enable the submission of OSG jobs to other grids, give the ability for OSG sites to accept jobs from other grids, and also allow for the transport and management of data across Grid boundaries. OSG, then, ensures interoperation with such e-Infrastructure providers as EGEE and the Nordic National Grid Infrastructure (NorduGrid) in the face of independently evolving software and processes and while supporting a different broader set of user communities.

Contribution

Main contributions of project: OSG considers its contribution holistically: more than examining indicators of the amounts of data transfer and storage, it seeks to enhance scientific productivity through e-Infrastructure of computation, experimentation and simulation research. Acknowledging that this objective is not easily quantifiable, OSG's leadership is working on developing concrete measures (Pordes, Altunay, and Bockelman 2008: 18). However, reports and informants highlight two central contributions to distributed research across different fields of science: technological provision of robust e-Infrastructure to main stakeholders and sharing and dissemination of the distributed models undergirding the practice of distributed research.

Challenges: OSG does not anticipate or actively encourage the commercialization of its core technology. At the same time, most commercial vendors have recently moved from OSG's

core technology—Grid computing—to the much different cloud computing. Even within the realm of Grid computing there are problems with user engagement. Although the computational portion of the e-infrastructure is mature, security mechanisms are not yet at a stage that can support analysis of sensitive data. At the organization level, competition over funding resources is largely responsible for lack of trust among OSG and some e-Infrastructure providers. While competition invigorates innovation, often leading to economic benefits, the outcome is likely more negative when establishing a global production e-Infrastructure. Different Funding mechanisms may be applied to ameliorate some of these tensions.

Informants' recommendations to policy makers

Informants suggested that a grassroots approach is insufficient in the creation of global research communities; it should be supplemented with top-down requirements from funding agencies for collaboration among providers, as well as among research communities.

SWOT analysis

Table 4-30: OSG strengths and weaknesses

	Strengths	Weakness
Long-term funding	Funding until 2011 is secured.	The project is in the middle phase of its funding cycle. No specific plans for continued renewal were discussed. However, based on opinions on informants who are not directly associated with OSG, their model is sufficiently robust to support additional funding after the end of the current project.
Sustainability	N/A	
User recruitment	The project has strategy for recruiting new users. As detailed above, the OSG follows a three pronged approach that utilizes an engagement team that works closely with users to enter into new communities. Informants have noted that the engagement has helped them join the OSG e-Infrastructure and offer these resources to users in over 100 labs, mostly in the US.	
Involvement of current users	Users are exposed to and often take part in decisions. The engagement team keeps newer users involved.	
Organizational bedding		Fermilab is the main player in the OSG, and appears to be the most committed than others. Nonetheless, the opportunistic computing model that offers economies of scale and is based on trust does promote higher institutional commitment.
Institutionalised links	There are good relationships, including interoperation with multiple international e-Infrastructures, most notably the EGEE. OSG also has relations with TeraGrid and Japan's NAREGI.	

External use of software, tools	OSG uses common, well established middleware tools such as Condor, Globus and VDT. Many other e-Infrastructures rely upon these tools.	

Table 4-31: OSG opportunities and threats

	Opportunities	Threats
Funding of member organizations	Participating organizations are universities and national labs that are not dependent on transient funding.	
Technology monitoring	OSG involves some of the world's most renowned experts on distributed academic computing. However, there is no indication that efforts are being made to consider alternative technologies, such as clouds.	
Competition with other infrastructures or technologies	The open all-inclusive model of the OSG offers a low barrier entry and economic benefits to participants. Still, cloud computing offers a significant threat, should it be publicly available and be able to serve the specialized needs of high-end computation.	
Security risks		While implementing commonly used authentication and other security mechanisms informants acknowledge that this infrastructure is not ready for handling of sensitive data
Change of user communities and fields	OSG has implemented various mechanisms that accommodate specialized environments—thus not requiring users to change. Among these are community tailored engagements, as well as partnerships with mediators who provide OSG resource to a particular community, without them knowing that they are using these resources.	

4.16 Swedish National Data Service (SND)

Case Overview

What does the project do mainly? The Swedish National Data Service (SND) is the national academic data service for Social Sciences, humanities and parts of medicine.[14] They are a service organization for all Swedish universities and colleges, whose purpose is to collect, document and disseminate data within the designated areas. An important purpose of this network is to further international collaboration and the exchange of data by enhancing the research infrastructure. SND also offers professional advice on matters of documenting and archiving data based materials.

Motivations for setting it up: The purposes of national data archives are to preserve research data and to make these data available for further research. SND preserves and provides access to data, both nationally and internationally.

Main goals of the project: To provide ease of access to high quality datasets; to provide a resource to science; to supply professional help and support on how to manage project data for re-use; and to preserve and maintain data from different sources.

Project maturity: The SND grew out of the Swedish Social Science Data Service (SSD), which was launched in 1981, and was incorporated into the Social Science faculty of the University of Gothenburg in the 1990s during the Swedish financial crisis. An evaluation of the organisation and a review of its situation were undertaken in 2002. The Swedish Research Council launched a major infrastructure initiative, the Database Infrastructure Committee (DISC) in 2006, the mission of which is to promote the development of an effective infrastructure for sharing research data resources in Sweden. One of DISCs first jobs was to transform the existing SSD into the Swedish National Data Service (SND), which then took a broader scope incorporating the Humanities and part of medicine (mainly epidemiology). The SND signed an agreement with the University of Gothenburg in 2007, establishing the university as its host for the next five years.

Project funding: Funding is currently secured until 2013 from the Swedish Research Council and the University of Gothenburg. The SND has been provided with more generous funding than the previous SSD, coinciding with the Swedish Research Council's recent initiative to support national data infrastructures.

Organizational Structure

Size and composition: Approximately 8 full-time staff working at the SND office. There is a larger network of persons at the research funding councils, Gothenburg University, and associated researchers working on the databases in which SND, which can be seen a data service, is embedded.

Governance: SND has an advisory committee, and its own director and support staff.

---

[14] Number of informants: Two, totaling 80 minutes.

This case study also builds on previous research, which involved interviews conducted with a number of stakeholders in the Swedish e-Science data sharing enterprise. These included 15 in-depth, face-to-face interviews and one phone interview with database managers, researchers and funders, and authorities responsible for protecting database privacy.

Managing internal and external relations

Management of the project: The project is managed by a team of 8 full time staff working at the SND office.

Users: The users are both existing users of the services of SND's forerunners, current ones, and anticipated future ones which are expected to comprise of a larger number of users due to the enhanced accessed brought about by SND.

User recruitment: SND is actively pursuing outreach to attract more users, and there is also a survey being currently undertaken (this will become available during the 2020 project).

Drivers and barriers to adoption: Drivers are the uniquely good databases and to secure access to sensitive micro-data.

Challenges in interdisciplinary collaboration: None.

Collaboration with other organizations: With CESSDA, funding agencies, and with universities and researchers who are involved in databases.

Technology

Main technologies, resources and services:

- Software: Some software options for secure access to data are being considered, though it is unlikely that a bespoke system will be developed just for this project.
- Network: Part of the Swedish university network, SUNET.

Data/storage: Being provided by Gothenburg University.

Data sharing: SND aims to provide its data as accessibly as possible, while maintaining security and anonymity of sensitive data.

Interoperability with similar or connecting infrastructures: This is not an issue, though the secure sign-in may be an issue in the future. Data formats are always an issue for data service providers, but this is not unique to Sweden or to SND.

Contribution

Main contributions of project: Providing access to the uniquely good Swedish datasets.

Challenges: Maintaining the unusually high level of trust that has so far been maintained about the use of sensitive data in Sweden, despite a number of incidents that have challenged this trust and provoked public debate.

Informants' recommendations to policy makers

Sweden is in a unique position to capitalize on its system of personal identifiers, long-established and comprehensive databases, and trust between researchers and the population – for the coming generation of shared databases in medical, social science and other forms of research. Sweden may provide a model of how this kind of data sharing, especially of sensitive micro-data, can take place, although its unique conditions are unlikely to be found elsewhere.

SWOT analysis

Table 4-32: SND strengths and weaknesses

	Strengths	Weakness
Long-term funding	Secured. The project is fully supported until 2013.	
Sustainability	Although funding has been invested until 2013, it is expected that the infrastructure will remain after this date.	
User recruitment	SND is currently undertaking a user survey and adopting new strategies for outreach.	These efforts are still in progress, so no results can be reported here.
Involvement of current users	See above.	
Organizational bedding	Tightly embedded within governmental agencies and university institutions. Well established in Swedish e-Science initiative.	
Institutionalized links	Yes, with CESSDA and with other researchers and institutions using data.	
External use of software, tools	The project uses the Swedish University network SUNET.	

Table 4-33: SND opportunities and threats

	Opportunities	Threats
Funding of member organizations	Solid funding for both agencies and institutions.	
Technology monitoring	Not known.	Not known.
Competition with other infrastructures or technologies	No competition exists due to the unique nature of this resource.	
Security risks	Currently this is not a problem but….	Secure sign-in may be more of a problem in future. This is still being evaluated.
Change of user communities and fields	Current investment by Swedish government in the exploitation of this data will ensure continued growth of user community and research potential.	

4.17 SWISS BIOGRID

Case Overview

What does the project do mainly? The Swiss Bio Grid project ran from 2004-2008, supporting large-scale computational applications in bioinformatics, biosimulation, chemoinformatics and bio-medical sciences by utilizing distributed high-performance computing, high speed networks, massive data collections and archives, as well as the necessary software tools and data integration capabilities. [15]

Motivations for setting it up: The motivation for the project was to assess whether Grid computing technologies could be successfully deployed within the life science research community in Switzerland.

Main goals of the project: The Swiss Bio Grid goal was to establish a value added collaborative platform focused on solving key scientific challenges of the life sciences. The project successfully completed two proof-of-concept studies, one in proteomics and the other investigating high throughput docking for dengue virus research.

Project maturity: This project ended in 2008, although lessons learned during the project continue to feed into a national initiative, the Swiss National Grid project (SwiNG).

Project funding: No external funding was committed to this project. Likely funding agencies, such as the Swiss National Science Foundation, were reluctant to put money into building infrastructures that weren't directly related to the achievement of a specific scientific goal.

Organizational Structure

Size and composition: Six different academic groups across Switzerland, one being the research lab of pharmaceutical company Novartis. There were three infrastructure personnel involved (up to 70% FTE in total), and two distinct scientific groups developing the separate projects.

Governance: A project coordinator led the project, with a steering committee supporting him and monthly teleconference to raise any developing problems.

Managing internal and external relations

Management of the project: Swiss Bio Grid developed a simple management structure, assisted in some ways by absence of any funding bodies with which to negotiate. However, the IP issues arising from the involvement of commercial partners took considerable time and effort to resolve (although this was achieved in the end).

Users: The users were mainly those involved in the project itself, both academic researchers and those in the pharmaceutical industry. However, the project was aimed at being extensible in the future, and may become so.

User recruitment: The users of this Grid were exclusively those researchers who were involved in the project from its development. No extra users were recruited to the team.

---

[15] Number of informants: 1 (in this round*), totaling 90 mins.
*The research for this project built on previous research conducted by Ralph Schroeder and Matthijs den Besten into Swiss BioGrid. Schroeder and den Besten had interviewed 6 informants, totaling an estimated 6 hours.

Drivers and barriers to adoption: The drivers for adoption of these tools and techniques were the academic users who sought to apply them in specific projects related to their work. The organic nature of the development of the project largely eliminated barriers to adoption.

Challenges in interdisciplinary collaboration: None – the two scientific projects were entirely separate. The only shared resource was the computing Grid. The scientists were also highly competent in technology, and were familiar with writing code and developing solutions to technological problems. There were therefore no major clashes between the biological scientists and computer scientists, as each understood enough about the challenges involved that they were able to work together sympathetically to resolve any problems.

Collaboration with other organizations: In this project the research results belonged to the institutions and groups leading the investigation, they simply used a computing Grid across different institutions to distribute computing power to enable the data processing. The potential for expanding this Grid exists, but the nature of the field makes collaborative research less likely.

Technology

Main technologies, resources and services:

<u>Processing</u>:   Spare computing cycles on PC clusters were used.

<u>Software</u>: A new piece of software was developed during the project to form a bridge between the Unix machines and PCs, to allow a job to be distributed between these two types of computer. This was known as a meta-scheduler. No such software existed at the time the project was developed, although similar software has since been developed.

<u>Network</u>:   Largely within the institutions.

<u>Data/storage</u>:   Hosted by and within the partner institutions.

Role of technology development: The only development is that mentioned under 'software' above. The goal of the project was to share largely redundant computing resources rather than to create new technology resources.

Data sharing: The sharing of research data among academic groups was not the focus of this project, rather the intention was to share computing resources to enable large processing capabilities for each of the proof-of-concept projects described above.

Interoperability with similar or connecting infrastructures: Swiss Bio Grid decided against using EGEE, as the academic communities were resistant to installing this kind of software. They felt EGEE was too intrusive, too time consuming to install, and assumed homogeneity of infrastructure that was not realistic within this community. Swiss Bio Grid therefore installed a 'less ambitious' middleware, NorduGrid (ARC) which was much less intrusive and heavyweight. The installation time was minimal, a major consideration in a project operating through goodwill.

Contribution

Main contributions of project:

<u>Political:</u> A major contribution was that it was possible to build grids organically. You could build something that didn't require top-down governance or dictation of what technology was going to be used. Swiss Bio Grid showed that it was possible to do this by consensus.

Swiss Bio Grid also had a profound impact on national engagement with Grid technology. It was one of the factors which led, in 2008, to a new Grid initiative was set up to represent

Switzerland in international Grid efforts. Several of the partners in SBG are now active members of this new structure, the Swiss National Grid (SwiNG).

Technical: The software developed to build a bridge between computers running different operating systems was a major contribution at the time.

Scientific: The virtual screening project has identified c.100 potential drug candidates, a number of which will be put into experimental validation by Novartis. A new drug for Dengue would have a dramatic effect.

The proteomics project showed that this infrastructure worked for the specific project that was developed, revealing the potential for similar projects to attempt to employ this kind of solution. It is unlikely that this exact infrastructure will be widely used by other groups, as each lab tends to develop its own approach and technologies for doing analysis. A general purpose solution is unlikely to be realised at this time.

Challenges: One of the key challenges for Swiss Bio Grid was to gainfully employ an immature technology in a heterogeneous environment without prior funding. Thus, as a proof-of-concept project, Swiss Bio Grid bore a heavy burden as it was set to shape the future of e-Research in the life sciences in Switzerland. With the recent establishment of the Swiss National Grid, the challenge seems to have been met.

Swiss BioGrid also illustrates a wider problem in the life sciences, which is that rather than becoming integrated around a shared computational infrastructure, e-Science initiatives have resulted in the promulgation of countless heterogeneous resources and efforts. While much of the development of the Grid has been geared towards applications in particle physics, which tend to be centralised and fairly homogeneous, the more heterogeneous requirements of computational biology have been poorly supported by existing Grid solutions, and it is unlikely that this picture will change unless there is a concerted effort in adapting Grid tools and putting them on a permanent footing.

Informants' recommendations to policy makers

Bottom-up development is highly desirable in order to sure that real scientific needs are addressed. On the other hand, the longer-term usefulness of the system that has been developed can only be sustained if the context of a longer-term structure in which it is embedded has been ensured. Put differently, an infrastructure that smaller projects can be part of is essential to ensure that gains are not lost after a project finishes.

SWOT analysis

Table 4-34: Swiss BioGrid strengths and weaknesses

	Strengths	Weakness
Long-term funding	N/A Project has finished.	N/A Project has finished.
Sustainability	The project has now ended, but the infrastructure is still in use by some of the scientists involved.	The project has now ended, and some project staff have transferred to a new initiative, SwiNG.
User recruitment		None was developed during the project as it was an organic, bottom-up project devised by the scientists.
Involvement of current users	N/A	N/A
Organizational bedding	The project was extremely well embedded within the participating institutions.	
Institutionalised links	N/A	N/A
External use of software, tools	Lessons learned from SBG have been transferred to the SwiNG project.	

Table 4-35: Swiss Biogrid opportunities and threats

	Opportunities	Threats
Funding of member organizations	Stable, particularly the commercial partner, which continues to invest in exploiting the results.	
Technology monitoring	N/A	N/A
Competition with other infrastructures or technologies	There was none within Switzerland at the time.	
Security risks	N/A	N/A
Change of user communities and fields	N/A	N/A

4.18 TeraGrid

Case Overview

What does the project do mainly? TeraGrid is a national e-Infrastructure that provides a distributed set of high-capability computational, data-management and visualization resources to academic users.

Motivations for setting it up: A 1999 report by the President's Information Technology Advisory Committee (PITAC) suggested that for the US to retain its leading role in basic research, scientists and engineers needed to gain appropriate access to the most powerful computers, which at this time were at the teraflop level (10^{12} operations per second). Along with the growing interest in Grid systems and a more specific focus on data grids, these discussions have also raised attention around the problem of managing, interoperating, analyzing and visualizing an exponentially growing amount of data from scientific instruments. Funders identified Grid computing as a technological infrastructure that could meet needs that go beyond the individual technical elements of computing, data and storage technologies, moving toward a more holistic facility of seamless, balanced, integrated computational and collaborative environment that supports scientific research.

Main goals of the project: Three objectives guide the project: (1) "petascale science"—the use of intensely high-end computational capabilities to advance computational science in multiple fields; (2) empowering science leaders through "science gateways" methodologies (see "user recruitment"); and (3) providing a reliable, general purpose set of e-Infrastructure services and resources. More recently TeraGrid stated that its aim is to enable science that could not be done without TeraGrid; to broaden the user base, simplify users' lives, improve operations, and enable connections to external resources.

Project maturity: With continuous streams of funding since 2001, a partnership of organizations that include some of the world's most experienced institutions in supercomputing provision, and five years of production, TeraGrid is a very mature e-Infrastructure provider.

Project funding: Over the past eight years the National Science Foundation has directly and indirectly awarded approximately $250 million to the TeraGrid. Of the $12.1 million allotment to the Grid Infrastructure Group (GIG), the largest expenditure went to outreach and user support (44%). It was followed by allocation of basic infrastructure, resources and services (24%); Management, finance, and administration (14%), Science gateways (11%), and CI development (7%). Having a different mandate, most of the $31.1 million to the Resource Providers (RPs) supported basic infrastructure, resources and services (56%). Other efforts included user support and science outreach (26%), management, finance and administration (12%), science gateways (4%) and CI development (2%).

Organizational Structure

Size and composition: TeraGrid began operation in 2001 as a partnership among four RPs: University of Chicago/ANL, California Institute of Technology (Caltech), National Center for Supercomputing Applications (NCSA), and San Diego Supercomputer Center (SDSC). As displayed in Figure 1, organizational membership has grown over the years to include eleven resource providers (RPs), the Grid Infrastructure Group (based at the University of Chicago/ANL), as well as four Software Integration partners. As noted elsewhere, this expansion was not planned at the outset and was mostly a result of subsequent NSF awards

made to additional sites. The combined resources support a staff of approximately 130 full-time equivalent positions.

Governance: Work in TeraGrid is distributed following a matrix approach to the distribution of work, so that individuals responsible for particular areas or tasks are not necessarily the direct supervisors of those who work on those tasks, and team members are often located across several sites. Two main entities in the TeraGrid lead the project: the TeraGrid Forum and the Grid Infrastructure Group. A recently established body, the Science Advisory Board, provides external evaluation and consulting role.

Managing internal and external relations

Management of the project: Resource Provider (RP) Forum is responsible for setting policy and governance for the project, the Forum consists of principal investigators from the RPs and the GIG. An elected Chairperson, who is funded through the GIG, leads the Forum. Working closely with RPs that implement and support resources and services, the GIG is charged with providing coordination, operation, software integration, management and planning for the TeraGrid. Work is divided across various subject areas, each with its own Area Director. Area Directors manage, oversee, coordinate and maintain TeraGrid activities within their area. Working groups consist of teams of experts available in partner sites. The GIG management team provides general oversight and management to working groups. Different organizational cultures, goals and competition among TeraGrid organizations make collaboration challenging. To address these challenges TeraGrid has recently implemented project management processes that support clearer division of labour, and bolstered communication and coordination mechanisms to help the synchronizing of its inter-organizational activities.

Users: According to the NSF Cyberinfrastructure Allocation Policy, individuals eligible for resource allocations are those who are a "researcher or educator at a U.S. academic or non-profit research institution." In recent years, with the expansion of the scope of PIs, many more people are eligible to use—and do use—TeraGrid services, growing from fewer than 1,000 users in October 2005 to over 4,000 users at the end of 2008 (TG annual report 2009). The number of active PIs in 2008 was about 1,500. A breakdown of all active users shows that most are graduate students (36%), followed by faculty (22.3%) and post-doctorates (12.7%). While the stated number of industrial users is negligible, individual computing centres may have separate undisclosed provision contracts with commercial clients.

User recruitment: TeraGrid uses traditional publicity mechanisms to attract users: a project website, press releases, and public news announcements directed at the served scientific community (TeraGrid Science Highlights and International Science Grid this Week), as well as dedicated, often large-scale training events, where participation is partially supported. In addition, to broaden its direct reach to users, TeraGrid has implemented in the past years two novel mechanisms: Science Gateways and Campus Champions. Science Gateways enable users to maintain their familiar work environment, while porting their applications to the Grid. Campus Champions involves technology leaders in a campus that advocate the use of TeraGrid in their local community.

Drivers and barriers to adoption: Access to the unique resources TeraGrid offers is the main driver to adoption for those researchers who need high-end computational data or data visualization resources. No less important, according to some informants, NSF channels funded research to facilitate this e-Infrastructure. However, even with these carrots and sticks both our informants and past analyses have indicated several challenges relating to barriers to users. These barriers can be categorized based on two distinct user populations: 1) the highly computer savvy and 2) those less familiar with the operation of supercomputers or e-Infrastructure computer resources. The first group has repeatedly complained about the functioning of TeraGrid, claiming that the system is unreliable at times and that they often need to wait a long time to have their job reach the top of the processing queue. Technical

design constitutes perhaps a more considerable barrier to a second, larger group of scientists. Since these people are less computer savvy, they have little tolerance to accommodate cumbersome interfaces, software that requires them to spend much time to obtain new knowledge.

Challenges in interdisciplinary collaboration: N/A

Collaboration with other organizations: TeraGrid has limited external relationships with other e-Infrastructure providers. These partners include the US Open Science Grid (OSG), and international collaborations mainly at the level of sharing knowledge and experiences with Enabling Grids for E-sciencE (EGEE), National Research Grid Initiative (NAREGI), Distributed European Infrastructure for Supercomputing Applications (DEISA) and others on occasions that bring together e-Infrastructure providers, such as the annual Supercomputing conference or the more specialized meetings that providers tend to organize.

Technology

Main technologies, resources and services: As of 2009, TeraGrid hardware capacity include 161,000 processor cores across 22 systems, offering more than a petaflop of computing capability and more than 30 petabytes of online and archival data storage, with rapid access and retrieval over high-performance networks. Another major service TeraGrid provides is Science Gateways. In addition to supporting individual gateway projects, TeraGrid personnel provide and develop general services for all projects. Among these efforts are: help desk support, documentation, SimpleGrid for basic gateway development and teaching, gateway hosting services, a gateway software registry, and security tools including the Community Shell, credential management strategies, and attribute-based authentication.

Role of technology development: See main technologies and interoperability.

Data sharing: N/A

Interoperability with similar or connecting infrastructures: TeraGrid development efforts aim to provide transparent use of the project's distributed resources among the heterogeneous set of computers and devices found in participating sites. Toward that end, sites have developed a Coordinated TeraGrid Software and Services (CTSS) Capability Kits, which are defined as "collections of software related by users-oriented HPC [high-performance computing] tasks." Examples of CTSS Kits include: Remote Login, Remote Compute, Data Movement and Science Workflow Support. TeraGrid representatives have worked with the Open Science Grid on interoperability across the two e-Infrastructures, specifically MPI parallel job submission through Globus. In addition, senior TeraGrid members have participated in the Grid Interoperation Now group, which, under the auspices of the Open Grid Forum, aims to develop and demonstrate interoperation among the major e-Infrastructure providers.

Contribution

Main contributions of project: Advancing the set of technologies required to integrate distributed heterogeneous supercomputers and other high end performing computers into a cohesive and persistent fabric is one of the most direct and important outcomes of the project. Another less direct, but nonetheless important contribution of this work, is that collaboration across sites that did not traditionally work together has created the social and organizational fabric that has enabled important technology advancements. These relationships are likely to sustain additional collaborative research partnerships, particularly in 2011, when the next funding program will be implemented. TeraGrid also offers significant improvement in resources available to scientists in fields that have traditionally relied upon advanced computational infrastructure to advance their research, including high energy physics and climatology.

Challenges: There are currently no sustainability mechanisms being implemented that would enable TeraGrid development to continue should funding cease. In fact, informants note that without continued, persistent streams of funding, many of TeraGrid's efforts will be terminated. Perhaps more challenging is the development of commodity commercial alternatives, which are based on cloud computing. Should comparable resources be offered through these vendors at a lower operational cost, the prospects of continued investments in TeraGrid are limited.

Informants' recommendations to policy makers

Informants suggested operating longer funding cycles. While more dynamic temporality is suitable for scientific research, it is less efficient for infrastructure construction—especially across multiple organizations—because it is an activity that requires a much longer time horizon. In addition, they also recommended more direct involvement of program officers that would allow funders to gain a clearer understanding of the complexities involved in their funded project, and would also enable them to recognize individual contributions each partner makes.

SWOT analysis

Table 4-36: TeraGrid strengths and weaknesses		
	Strengths	**Weakness**
Long-term funding	Long-term funding is secured.	
Sustainability	In 2010, TeraGrid will continue as TeraGrid Extreme Digital Resources for Science and Engineering, likely though a different mixture of participating organizations.	
User recruitment	The infrastructure has a strategy for recruiting new users.	
Involvement of current users	After recognizing that a "build it and they will come" approach is untenable, TeraGrid has moved to an innovative three pronged strategy that includes marketing and information dissemination, a novel Science Gateways program that minimizes the need for users to change in adopting the TeraGrid, and Campus Champions that leverages local presence of technology advocates in university campuses. These programs managed to attract users that were not traditionally associated with supercomputing, but require high-end computation and data resources	
Organizational bedding		While strongly embedded in participating institutions, continual competition for grants—especially the upcoming TeraGrid Extreme Digital Resources—weakens the overall commitment to the project.
Institutionalised		Aside from efforts to collaboration with

links		the Open Science Grid, there are no established interoperation mechanisms, only exchanges of knowledge and practices.
External use of software, tools		TeraGrid has a very large number of users. At the same time most of its developments serve various participating sites in TeraGrid, but there is no evidence to suggest that it is used elsewhere.

<p style="text-align:center">Table 4-37: TeraGrid opportunities and threats</p>

	Opportunities	Threats
Funding of member organizations	Although the organizational composition would likely change, the allocation of next round public funding has been ensured. Nevertheless, until the winner of the bid is announced, there is fierce competition among current collaborators, which clouds day-to-day operation.	
Technology monitoring	TeraGrid involves some of the world's most renowned experts on distributed academic computing. However, there is no indication that efforts are being made to consider alternative technologies, such as clouds.	
Competition with other infrastructures or technologies	Being a highly complex and specialized operation, there are no alternatives e-Infrastructure technologies that can be implemented across participating sites to support the provision level of TeraGrid. Still, cloud computing poses a significant threat, should it be publicly available and be able to serve the specialized needs of supercomputing/high-end computation and data scientific users.	
Security risks	There is a stream of research and development on security, including identity management and advanced authentication mechanisms.	
Change of user communities and fields		It does not seem likely that the need for high-end distributed resources TeraGrid provides will quickly expand beyond communities that are currently served.

5   Multi-case comparison[16]

5.1 Size and composition

This section describes part of the structural background of the e-Infrastructures in the sample. In terms of size, we cover the whole spectrum of e-Infrastructures, starting from single institutions such as the Swedish National Data Service and single country Grid initiatives such as Swiss BioGrid and MediGrid with fewer than ten partners, to very complex organisations such as GEANT with its 30 NRENs to be co-ordinated, CineGrid with 50 partners from industry and academia, to OSG and EGEE, arguably the largest infrastructures in the sample with more than 50 partner organizations. A number of 11 to 20 partners appears to be the size of choice for most infrastructures (although it has to be noted that we do not have a random sample), with nine cases being within these size class brackets.

Size is not depending obviously on the endeavour being a purely academic one or an academic and industry partnership.

Table 5-1: Size in terms of participating organisations

Size	e-Infrastructures
Single	SND
5-10	SWISS BIOGRID*; MediGrid
10-20	D4SCIENCE; ETSF; DRIVER; DARIAH; DEISA ; EELA-2; TeraGrid; NVO; C3-Grid*
>20	CLARIN, EGEE*; GEANT; CineGrid*; OSG, OGF*

* Case has industry participants

With regard to international scope, several projects are single country endeavours, C^3-Grid and MediGrid (DE), SND (SE), SWISS BIOGRID (CH), NVO, OSG and TeraGrid (US). Really global projects, spanning two or more continents are CineGrid, D4SCIENCE, EELA-2 and OGF. Others are purely European (DARIAH, DEISA, DRIVER, ETSF and GEANT) at least in regard to their partner organizations.

Table 5-2: Scope of participants

Case	Participating organizations	Scope of participants and project's staff	Academia /industry
C3 Grid	18	8 data-providing institutions, 8 "operators" (users), 2 IT service institutions. Earth science and IT, all partners from Germany	Academia /industry
CineGrid	50	80% North-America, 10% in Asia (JP, KR), 10% in Europe. Networking organizations (NRENs, institutions working with lambdas), media schools and university institutes (computer science and media), non-profit and other public organizations , IT and telecom corporations, film & media companies.	Academia /industry
CLARIN	156	Partners in 32 European countries, including universities, national language councils, institutes, and libraries.	Academia
D4SCIENCE	11	University institutes and public research organisations. 10 Europe, 1 Asia	Academia

---

[16] The principal authors of this section are: Franz Barjak, Kathryn Eccles, Tobias Hüsing, Zack Kertcher, Eric Meyer, Simon Robinson and Ralph Schroeder

Case	Participating organizations	Scope of participants and project's staff	Academia /industry
DARIAH	14	National data archiving centres, university institutes and public research centres	Academia
DEISA	15	11 principal partners and 4 associate partners (coordinators) National supercomputing centres (principal partners), plus coordinators	Academia
DRIVER	13	Universities or university libraries and National Repositories Representatives	Academia
EELA-2	16	14 different countries:Coordinators of Joint Research Units (JRU) 5 European (Spain, France, Italy, Portugal and Ireland) and 9 Latin American (Brazil, Argentina, Chile, Colombia, Cuba, Ecuador, Mexico, Peru and Venezuela)	Academia
EGEE	More than 140	1'000 persons (380 full-time equivalents) from27 European countries	Academia /industry
ETSF	11	Mainly solid matter physics departments, one material sciences	Academia
GEANT	32	30 NRENs (voting), two administrative service partners (non-voting)	Academia
MediGrid	8	German public or semi-public institutions. Staff 8-12 FTE	Academia
NVO	17	All US-based: astronomy data centers, national observatories, supercomputer centers, university departments and computer science specialists. Staff 13.5 FTE positions, which are shared by 51 people	Academia
OGF	Assessing "size" difficult	various academic, commercial and government/non-profit organizations. Staff currently a full-time Director and an assistant to the Director	Academia /industry
OSG	53	Academic and research institutions, mostly from the US, HEP experimenters, major Grid technology development projects and national laboratories, plus dozens of virtual organizations—research and development groups in various fields of science. Staff: 34 FTE staff positions, which are handled by about 50 people in 16 institutions throughout the US	Academia
SND	1	N/A. Staff: 8 FTE at the SND office, further persons at associated institutions	Academia
Swiss BioGrid	6	University institutes of medical and life sciences, industry, bioinformatics. Staff: 2.1 FTE purely infrastructure plus research groups staff working on projects	Academia /industry
TeraGrid	16	11 resource providers (RPs), the Grid Infrastructure Group 4 Software Integration partners. Staff approximately 130 FTE positions	Academia

5.2   Background of the e-infrastructure (problem setting, motivations, goals)

The problem settings of the analyzed infrastructures are very heterogeneous. Some are feasibility studies with test bed character (MediGrid, Swiss BioGrid) others serve the complete science community of Europe or large scale big science endeavours (Geant, EGEE, TeraGrid, OSG). A detailed overview is given in the annex table 1-1.

We can distinguish between the motivations for setting up the projects in regard to community-driven and bottom-up (e.g. C3-Grid, Cinegrid) versus developer-driven and top-down (e.g. D4Science, EELA-2). The following table gives a tentative overview:

Table 5-3: User community driven vs. developer driven e-Infrastructures

Community driven		Developer driven	
C3GRID	CineGrid	D4Science	DARIAH
EGEE	NVO	DEISA	DRIVER
OGF	OSG	EELA-2	ETSF
CLARIN	SWISS BIOGRID	GEANT	MediGrid
		SND	TeraGrid

The e-Infrastructures main goals and motivations as they were stated by the interview partners could also be categorized and distinguished between technical, scientific and socio-cultural (see table 1-2 in the annex on the goals in detail). While most e-Infrastructures pursue technical goals, such as to build a grid infrastructure and to agree on technical standards, conspicuously few e-Infrastructures define scientific goals. ETSF is one exception in that it is mainly concerned with scientific application rather than the technical conditions of doing science. NVO is another example.

Socio-cultural and political goals are for instance to fight the digital divide by expanding networks and access to infrastructures to remote regions (GEANT, DRIVER, EELA-2) or to promote open access (DRIVER).

Table 5-4: Types of main e-Infrastructure goals

Technical	Socio-cultural	Scientific
C3GRID	CineGrid	C3GRID
CineGrid	D4Science	EGEE
D4Science	DRIVER	ETSF
DEISA	GEANT	MediGrid
DRIVER	OSG	NVO
EELA-2	TeraGrid	SWISS BIOGRID
EGEE	CLARIN	TeraGrid
GEANT	SND	OSG
MediGrid	DARIAH	SND
NVO	SWISS BIOGRID	
OGF	DEISA	
OSG	EELA-2	
SND	DARIAH	
SWISS BIOGRID		
TeraGrid		
DARIAH		

5.3   Funding arrangements: current and future

The assessment of project funding turned out to be a rather complex and difficult undertaking: e-Infrastructures received direct funding through EU contracts, national, or other sources and in most cases project participants cofunded the work indirectly with their own organizational budgets, unpaid labour and/or significant contributions in kind. This indirect funding could not be assessed reliably and the following Table 5-5 and Table 5-6 can therefore only be considered as very rough approximations to the real budgets, probably underestimating total budgets in most projects.

The heavy weights in terms of funding among our sample clearly are GEANT (40m € annually), EGEE (23.6m € annually, 48.6 m € including estimated partner contributions) and TeraGrid (31m US-$ annually). Still quite large are DEISA (6.24m € per year) and OSG (6m US-$ per year). Most projects range between 1 and 3m € or US-$ annually, namely C3-Grid, CLARIN, D4Science, DRIVER, EELA-2, ETSF, MediGrid, and NVO. The exclusively self-funded projects CineGrid, OGF and SwissBioGrid are probably the smallest e-Infrastructures included in this study.

Table 5-6 may give the impression that US-based e-infrastructures (OSG, Teragrid, NVO) have durations 5 years or more and longer perspectives than the European cases. However, for projects funded by the EU this is an artefact of the funding constructions: Projects were at the time of the case studies in their second (e.g. DEISA, EELA-2, Driver) or third (EGEE, Géant) funding rounds or had developed on the bases of predecessors (D4Science), so that the overall durations of the larger projects are also in the ranges of 5-10 years. It seems, however, that in Europe the projects at national levels (Swiss Biogrid, MediGrid, C3Grid, and SND) lack a long-term perspective and struggle with securing the funding for more than 5 years.

Table 5-5: Funding arrangements: current and future

	Funding	Future funding
C3-Grid	Total funding: 3m € personnel costs. Additional hardware had to be provided by the participating institutes, additional 3m €. Total budget 6m €.	If the follow-up project will be approved, it will be funded by the department for Cultural, Earth System and Environmental Research of the German Ministry for Education and Research, Department for Information and Communication.
CineGrid	Funding through membership fees and considerable contributions in kind from the members for realizing the CineGrid projects.	The long-term funding of the CineGrid organization is secured through its membership fees. Funding is low level. The community depends on additional funding and contributions in kind from its members. In the past it has been possible to mobilize the necessary funds, but it cannot be said to what extent this will be achieved in the future.
CLARIN	EU funding of €4.1m for initial stage	Future funding depends on national investment from the countries supporting CLARIN.
D4Science	The EU funding for D4Science amounts to 3.15 million EUR. The overall budget is 3.92 million EUR. The cost of the predecessor project DILIGENT was about 8.9 million EUR; the European	In the future, the costs of administration should be reduced and it is planned to reach a more or less autonomous e-Infrastructure. The

	Funding	Future funding
	Community contribution was 6.3 million EUR. Informants indicated that D4Science is on the top of other projects and uses resources and technologies which were developed in EGEE and DILIGENT. Therefore it is very difficult to estimate direct and indirect costs.	users could get the software, learn to use it and set up the VREs without causing any costs.
DARIAH	Preparing DARIAH' is funded by the EC. The preparatory phase is estimated to cost €6 million, with construction costing another €10 million. In order to secure the DARIAH project, annual funding of an estimated €6 million is required from national governments and funding organisations. The aim is to create an infrastructure of at least 25 partners, requiring a funding commitment of €250,000 per partner. Peter Doorn, director of the Dutch national archiving organization DANS and one of the founders of the DARIAH project, has stated that 'large countries will pay more and small ones less, depending on national priorities, and it is likely that the country hosting the central office will pay slightly more.	N/A (in preparation)
DEISA	The project cost for DEISA1 was 24,351,100 EUR, the EU funding 13,976,000 EUR. The project cost of EDEISA (an interim project) was 13,145,700 EUR, the EU funding 7,000,000 EUR. The project cost: of DEISA2 is 18,733,200 EUR, the EU funding 10,237,000 EUR	
DRIVER	The budget of DRIVER II is 3.06m €. Funding from the EC is 2.7m € Euro for DRIVER2 and was 1.8m € for DRIVER1. With cost of operation being at about 1m € per year, 40% are indirect and 60% direct costs. 432 person months are funded under DRIVER II	No information yet
EELA-2	5.1m € in EELA-2 (3m € in EELA)	The future challenge will be to make the infrastructure permanent and convince Latin American governments to build and dedicate resources to NGIs. As to be expected, the first and main challenge is to secure the funding for the NGIs.
EGEE-III	Total budget is 47.15m € (with a further estimated 50m € worth of computing resources contributed by the partners) 32m € EC funding	
ETSF	EC funding: 3.8m € for 36 months (+ ext. 6.2m € provided by partners)	At the moment, ETSF does not get any funding after December 2010.
GEANT	Total budget 40m € / year	Very probable
MediGrid	MediGrid project received around 5m Euro from the BMBF	Ended
NVO	In September 2001 NSF's Information Technology Research program awarded $10M for a 5-year period. Along with funding from NSF's astronomy division, this award has been extended until September 2009 with additional funds, totaling	In its next stage of funding, NVO is promised to receive $36 million for the next five years.

	Funding	Future funding
	approximately $14 million in overall allocation from the start of the project. Currently, most of the funding is for software developers.	
OGF	Undisclosed amounts from individual and organizational membership, as well as from sponsors. The annual budget is estimated to be somewhat less than $1 million. Informants note that the overall budget has substantially declined in the past years, primarily as a result in diminishing interest in Grid computing.	No information yet
OSG	$30 million of joint funding from the Department of Energy and the National Science Foundation, for an initial five-year term.	Funding until 2011 is secured
SND	Strong, ongoing commitment of funding until 2013, history of government support for this project	Short term funding is secure (until 2013), but long-term future may be in some doubt
SWISS BIOGRID	No external funding was committed to this project. The project largely evolved using machines already available within the institutions involved, and Swiss BioGrid's mission was to free up these resources for scientific purposes.	Ended
TeraGrid	Over the past eight years the National Science Foundation has directly and indirectly awarded approximately $250 million to the TeraGrid. Of the $12.1 million allotment to the Grid Infrastructure Group (GIG), the largest expenditure went to outreach and user support (44%). It was followed by allocation of basic infrastructure, resources and services (24%); Management, finance, and administration (14%), Science gateways (11%), and CI development (7%). Having a different mandate, most of the $31.1 million to the Resource Providers (RPs) supported basic infrastructure, resources and services (56%). Other efforts included user support and science outreach (26%), management, finance and administration (12%), science gateways (4%) and CI development (2%).	Very probable

Table 5-6: Annual funding and structure of funding by sponsors[a]

	Total funding/duration	Funding per year	Structure of funding		
			Government/EC	Project participants	Others
C3-Grid	6m €/42 months	1.7m €	50%	50%	0%
CineGrid	N/A	N/A	0%	100%	0%
CLARIN	4.1m €/36 months	1.37m €	100%	0%	0%
D4Science	3.92m €/24 months	1.96m €	80.4%	N/A	N/A
DARIAH	Projected: 16m € /unknown duration	N/A	N/A	N/A	N/A
DEISA	18.7m €/36 months	6.24m €	55%	45%	0%
DRIVER	3.06m €/24 months	1.53m €	N/A	N/A	N/A

	Total funding/duration	Funding per year	Structure of funding		
			Government/EC	Project participants	Others
EELA-2	5.1m €/24 months	2.55m €	42%	58%	0%
EGEE-III	47.15m €/24 months + 50m € (est.) contributions in kind	23.575m € (48.575 m €)	32.9%	67.1%	0%
ETSF	3.8m €/36 months + 6.2m € (est.) contributions from partners	1.27m € (3.33m €)	35%	60%	5%
GEANT	N/A	40m €	48%	52%	0%
MediGrid	5m €/48 months	1.25m €	80%	20%	0%
NVO	14m US-$/8 years	1.75m US-$	100%	0%	0%
OGF	N/A	N/A (< 1m US-$/year est.)	0%	100%	0%
OSG	30m US-$/5 years	6m US-$	100%	0%	0%
SND	N/A	N/A	100%	0%	0%
SWISS BIOGRID	N/A	N/A	N/A	N/A	N/A
TeraGrid	250m US-$/8 years	31.25m US-$	100%	0%	0%

a Participants are frequently funded from public sources (EU, national level). Partly considerable contributions in kind and labour provided for free by project participants could not be estimated consistently. The figures therefore represent only the lower bounds of total resources available to a project.

5.4   Context of academic domains and fields

This section takes a closer look at the academic fields and non-academic communities which are involved in the cases with a focus on both, the developer fields as well as the user fields. It describes different characteristics of these fields with a perspective on their influence on the uptake of e-infrastructure.

As our sample of e-infrastructure projects is purposive and by no means representative of any larger population of projects, it cannot give a general overview of the fields involved in e-infrastructure activities. It comes as no surprise that Grid computing and supercomputing predominate among the fields from which the developers come, with some contributions by high energy physicists and other fields of computer science (networking, scientific visualization) and neighbouring fields (bioinformatics, computational linguistics). Among the user fields are biosciences, HEP and other fields of physics, earth and environmental sciences, computer science, astronomy and astrophysics the most prominent fields in our sample. Social sciences, arts & humanities, materials science, chemistry and medicine are also involved in some of the projects.

Table 5-7: Developer and user fields

	ESFRI category	Developer fields	User fields
C3-Grid	Environmental Sciences	Grid computing	– Climatology – Geophysics – Biogeography – Hydrology – Oceanography – Other earth system sciences

	ESFRI category	Developer fields	User fields
CineGrid	e-Infrastructure	– Computer networking – Scientific visualization – Media science	
CLARIN	Social Sciences and Humanities	– Computer science – Computational Linguistics	– Linguistics – Languages – Computational Linguistics – Literature
D4science	Environmental Sciences	Grid computing	– Environmental Monitoring – Fisheries and Aquaculture Resources Management
DARIAH	Social Sciences and Humanities	– Library science – Computer science	– Arts & Humanities – Social Sciences
DEISA	e-Infrastructure	Supercomputing	– Nuclear fusion – Climate/earth system research – Astrophysics/cosmology – Computational Neuro Sciences – Plasma Physics – Computational Bio Sciences – Materials sciences
DRIVER	e-Infrastructure	– Library science – Computer science	N/A (any)
EELA-2	e-Infrastructure	Grid computing	– High-energy physics (HEP) – Biomedicine and bioinformatics – Earth sciences – Artificial intelligence and optimization – Chemistry – Civil protection – Engineering – Environmental science
EGEE	e-Infrastructure	– Computer Science/Tools – High-Energy Physics	– Archaeology – Astronomy & Astrophysics – Civil Protection – Computational Chemistry – Computational Fluid Dynamics – Computer Science/Tools – Condensed Matter Physics – Earth Sciences – Finance (through the Industry Task Force) – Fusion – Geophysics – High-Energy Physics – Life Sciences – Multimedia – Material Sciences
ETSF	Materials and Analytical Facilities	Theoretical physics	– Condensed matter physics – Chemistry – Biology – Material science – Nanotechnology
GÉANT	e-Infrastructure	Computer networking	N/A (any)
MediGrid	Biological and	Grid computing	– (Clinical) Medicine

	ESFRI category	Developer fields	User fields
	Medical Sciences		– Biomedicine – Biomedical informatics
NVO	Physical Sciences and Engineering	Grid computing	Astronomy
OGF	e-Infrastructure	Grid computing	Grid computing
OSG	e-Infrastructure	– Grid computing – HEP	– HEP (~90%) – Others (10%), such as theoretical physics, astrophysics, industrial engineering, computer science and natural language processing, chemistry, biochemistry, computational biology, genetics, structural biology and economics
Swedish Nat. Data Service	Social Sciences and Humanities	– Grid computing	– Humanities – Social Sciences – Medicine
Swiss BioGrid	Biological and Medical Sciences	– Grid computing	– Biological Sciences – Pharmaceutical research
TeraGrid	e-Infrastructure	– Supercomputing – Grid computing	– Molecular Biosciences – Physics – Chemistry – Astronomical Sciences – Materials Research – Earth Sciences – Advanced scientific computing – Chemical, thermal systems – Atmospheric Sciences – 19 other fields (<3% used NUs)

Next we assessed several characteristics of the case studies' user fields. We can distinguish between cases that were developed for and often also in close interaction with a rather narrow community of users and those that were developed as general purpose infrastructures for any interested community. Only for those of the former is an assessment of the field characteristics possible.

Collaboration is an important element in all user fields involved in the e-infrastructure cases. However, there is usually an intricate mix of collaboration and competition; OSG may serve as an example: the HEP community collaborates in developing the technology for running its competitive experiments. Also there are strong incentives to using e-infrastructure services in all of the cases: the fields are confronted with an increasing necessity of using large amounts of heterogeneous data from different sources and they require fast network connections and high-performance computing power to transmit and process it. The dynamics could only be assessed for half of the included cases; however we see that the need for e-infrastructure does not necessarily go in parallel with a fast pace of change in regard to problems, paradigms and approaches. Rather to the opposite, some projects - C^3-Grid, OSG and this certainly also applies to DEISA, EGEE, TeraGrid for which this question was not answered at general level due to the many user fields to which they cater - serve fields needing the infrastructure to move forward on big challenges which they have been addressing for some time already, e.g. the search of the Higgs boson (HEP), better climate modelling and identification of human influences on climatic change (environmental sciences), computer-based or *in silico* screening of compounds for drug discovery (biomedicine/-informatics). And even if there is a strong need, for example, in joining heterogeneous datasets in health, biological and social science research, it is not clear if the demand is starting to be met, or if

there is a large demand which is going unmet (only interviews with domain scientists could answer this).

Table 5-8: Structure of the user fields[a]				
	User disciplines	Collaboration & competition	Infrastructure/ facilities	Dynamics
C3-Grid	Few	Frequent & large scale collaboration among several fields	Increasing production of data and demand for data management	Rather low dynamics, persistent work on big challenges
CineGrid	Few	Partly frequent and large scale collaborat-ion, partly secretive developments (film industry)	Varying: Networking, visualization infrastru-cture are research ob-jects, but not yet common in commercial film productions	Varying: partly highly dynamic, partly rather conservative
CLARIN	Few	Increasing prevalence of collaboration	Successful infrastructure projects exist, CLARIN builds on these	Slow but steady dynamic
D4science	Few	N/A	Demand for integra-ting, customized pro-cessing and rearrangement of heterogeneous data from multiple sources	N/A
DARIAH	Many	Likely in some fields, very unlikely in others	Drive for collation of fragmented data sets to improve access for researchers	Slow but steady dynamic
DEISA	Many	N/A	N/A	N/A
DRIVER	Many	N/A	N/A	N/A
EELA-2	Many	N/A	N/A	N/A
EGEE	Many	N/A	N/A	N/A
ETSF	Few	Collaboration in small teams	Demand for computing power to analyze complex systems of atoms	Low dynamics, theoretical basis is dating back to the 1920s
GÉANT	Many	N/A	N/A	N/A
MediGrid	Few	N/A	N/A	N/A
NVO	Few	Frequent & large scale collaboration in astronomy sub-fields	1. Large importance of telescopes, observatories, data processing and management services 2. Growing practice of using data from other sub-fields (wave-lengths)	N/A
OGF	Few	Varying degrees of collaboration and competition between different Grid projects and players	Development and promotion of Grid computing is the main objective	Highly dynamic, project tries to bring more coherence into Grid development

	User disciplines	Collaboration & competition	Infrastructure/ facilities	Dynamics
		from academia, business, government and NPOs		
OSG	Many	HEP: Collaboration in technology R&D but competition in physical experiments	Access to distributed computing infrastructure is essential for HEP research	Rather low dynamics, persistent work on big challenges
Swedish Nat. Data Service	Many	Collaboration with other infrastructures growing	Access to data is important	Slow but steady dynamic
Swiss BioGrid	Few	Grid computing was the primary focus, no research collaboration	Access to Grid computing acknowledged to advance research in some areas	N/A
TeraGrid	Many	N/A	N/A	N/A

a User disciplines: Number of user fields to which the e-infrastructure caters; Collaboration and competition: Between fields, roles of theoreticians, empiricists, method/tool developers; Infrastructure/facilities: Importance of infrastructure/facilities, computing, data; Dynamics: Pace of change in regard to problems, paradigms and approaches in the fields.

5.5   Use and user communities

The chances of the cases for achieving long-term sustainability depend to a large extent on the e-infrastructure's ability to mobilize user communities, such as scientists, researchers from outside academia or other professionals who draw benefits from using the infrastructure. There is no standardized way of measuring use and therefore we let our case informants decide what the appropriate unit of measurement is.

We see first of all that international projects also cross continental boundaries to become truly global undertakings (see Table 5-9). Among the investigated cases there are ten international and eight national cases. The number of user organizations is often not measured in the projects but the most common available data relates to the individual user. But even taking these figures, for a variety of reasons, we experienced some difficulties with measuring usage:

- Users connect through gateways or portals which then do not appear as distinguishable organizations or individuals to the e-infrastructure providers;
- Registration and authentication are handled at a higher level (organization) and the individual user's identity is not revealed at log-in;
- Users register with the e-infrastructure and then there is little monitoring of what tools and applications they actually use.
- It is impossible to distinguish between a former user who stopped use, e.g. because of a more suitable alternative, and someone who just interrupted use.

For these reasons, the numbers of users stated by the informants vary sometimes by several orders of magnitude. Some infrastructures reach already very large and multidisciplinary user communities, above all EGEE, TeraGrid, OSG, DEISA and DRIVER, whereas most others still deal with a rather narrow set of 50 to up to 200 people.

Table 5-9: Extension of user communities

	Continents	Countries	Organizations	Individuals	Fields	Other

	Continents	Countries	Organizations	Individuals	Fields	Other
C3-Grid	1	1	N/A	around 50	5-10	
CineGrid	3	around 10	around 50	> 200	3-5	
CLARIN	1	32	156	N/A	3-5	
D4science	> 1	N/A	N/A	around 80	2	
DARIAH	1	10	14	N/A	5-10	
DEISA	> 1	N/A	>160	Several 100	5-10	
DRIVER	1	21	245	around 10000 hits per month	N/A	245 repository managers
EELA-2	2	14	around 50	50-100	5-10	56 registered, 32 deployed app-lications
EGEE	> 1	N/A	N/A	> 16,000	15	
ETSF	3	N/A	N/A	100-150	3-5	
GÉANT	1	34	N/A	N/A	N/A	30 NRENs
MediGrid	1	1	N/A	few	3	
NVO	1	1	N/A	> 100	< 5	
OGF	Several	Several	N/A	N/A	1	
OSG	1	1	N/A	around 2500	> 10	
Swedish Nat. Data Service	1	1	1	Estimate hundreds	<5	
Swiss BioGrid	1	1	6	<50	1	
TeraGrid	1	1	N/A	around 4000	> 20	around 1500 PIs

In the cases with small user communities, these are mostly restricted to pilot users, i.e. users from organizations participating in the project and a few scientists from other organizations who have learned of the project and its services and became involved because of congruencies to their work and needs (see Table 5-10). Out of the 19 cases shown in Table 5-10, 7 therefore are considered to be of low maturity, 5 of medium maturity and 7 of high maturity in regard to their user communities.

Table 5-10: Description of user communities

	Type of users	Status of use	Maturity[a]
C3-Grid	Pilot users, mainly from the project members	Ongoing development, infrastructure not fully operable	Low
CineGrid	Community members, technology developers and innovators	Eventual demonstrations at conferences and other events	Low
CLARIN	Pilot users from the project members	Nascent project, still in development	Low
D4science	Pilot users from the project members, external users from the EM and FARM communities	First version of a production infrastructure; still heavy involvement of mediators or community managers who realize the demands of end users	Low
DARIAH	Pilot users from the project members	Nascent project, still in development	Low
DEISA	Scientists and other users of	Production quality infrastructure	High

	Type of users	Status of use	Maturity[a]
	supercomputing services		
DRIVER	– Repository managers provide content – Organisations building their repository systems with DRIVER – End users using the portals	Production quality infrastructure	High
EELA-2	Pilot users from the project members, external users of Grid computing services	Production quality infrastructure	Medium
ETSF	Scientists and other users	Production quality infrastructure	Medium
GÉANT	Anybody in Europe transmitting data through an NREN internationally	Production quality infrastructure	High
MediGrid	Users from project members	Testbed	Low
NVO	Community members, research astronomers	Ongoing development, infrastructure not fully operable	Medium
OGF	Community members, organizations and individuals involved in Grid computing	Continuous publication of specifications	High
OSG	Scientists from many different organizations and fields	Production quality infrastructure	High
Swedish Nat. Data Service	Researchers from diverse fields	Operational	Medium
Swiss BioGrid	Scientists from a sub-field, pharmaceutical researchers	Project is now over	High when complete, now low if unrevived
TeraGrid	Graduate students (36%), faculty (22.3%) and post-doctorates (12.7%) from the user fields	Production quality infrastructure	High

a High: established large user population; medium: established small user population; low: no established user population

5.6   Interdisciplinary collaboration

In previous studies on e-infrastructure and e-science, collaboration between developers (computer scientists) and users (scientists from different domains) has been found to be problematic (Barjak et al., 2009). It therefore deserves particular attention in the present case comparison. It is noticeable that in addition to the developers and users there may be several further groups which need to collaborate and communicate and may encounter problems in the process (see Table 5-11), such as infrastructure and application developers (e.g. EELA-2, D4Science, TeraGrid), users from different user fields (e.g. NVO, OSG), or scientists and practitioners (e.g. CineGrid). This may create multiple layers of interest, work modes and communication practices which may be difficult or even impossible to reconcile (see Figure 5-1: Different involved stakeholders in e-infrastructure projects).

Figure 5-1: Different involved stakeholders in e-infrastructure projects
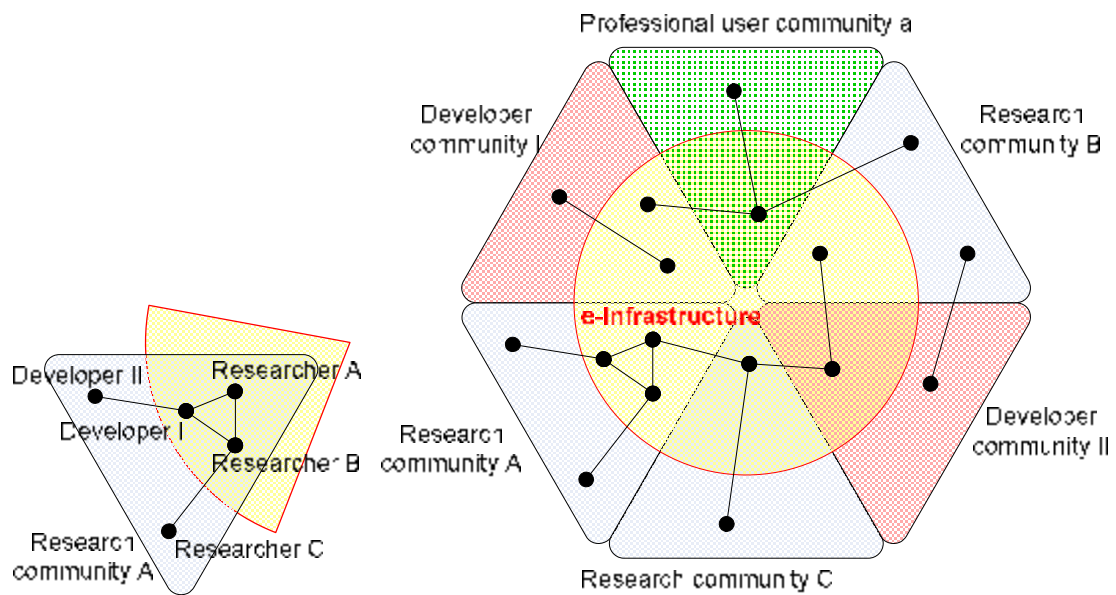
Typically our informants listed a negative attitude towards technology and computer-enhanced research, little understanding of domain-specific practices, general problems of field jargon and communication, and divergent objectives (cutting-edge research versus service provision) as the strongest challenges (see Table 5-11).

Table 5-11: Challenges of interdisciplinary collaboration

	Involved fields and groups	Type of challenges	Scope of challenges
C3-Grid	Computer science, earth systems sciences	Different scientific cultures, languages and jargons	Large; estimated at least 20% of the effort of each project member
CineGrid	1. Fields of computer science and electronics, media science 2. Practitioners and scientists 3. NRENs and artists, researchers and movie professionals	– Attitude to technology, – Defining and solving problems, – Differing time horizons	Medium; large commitment to making collaboration work
CLARIN	1. Project management 2. Developers 3. Practitioners/Researchers	– Interoperability of resource – Establishing user community – Securing funding	Large, but strong commitment from community
D4science	1. Project management 2. Developers and testers 3. Domain-specific mediators 4. Users from EM and FARM communities	None mentioned.	N/A
DARIAH	1. Development Team 2. Librarians and Collections managers 3. User communities	– Interoperability with community resources and other infrastructures – Establishing user community – Securing funding	Large
DEISA	1. Operations Team	Developers have little	Small; natural scientists

	Involved fields and groups	Type of challenges	Scope of challenges
	2. Development and Technology Team 3. Applications Support Team 4. User communities	knowledge of user needs and practices	have a long tradition with using supercomputers and good understanding of possibilities and barriers
DRIVER	1. Librarians 2. Computer scientists	N/A	N/A
EELA-2	70% computer scientists 30% domain scientists	− Lacking domain-specific expertise − Differing research/ work practices − Scepticism towards new technology and computation	Medium; strong commitment to establishing user communities on the infrastructure
EGEE	None mentioned.	None mentioned.	N/A
ETSF	1. Physicists 2. Chemists 3. Other user fields 4. Theorists 5. Experimentalists	None mentioned.	N/A
GÉANT	N/A	N/A	N/A
MediGrid	40% computer scientists and engineers 40% computational biologists 10% medical staff 10% economists and others	None mentioned.	Low
NVO	1. Computer science, astronomy, physics 2. Developers and users 3. Astronomy sub-fields (wavelengths)	− Differing epistemic cultures − Divergent objectives: "Feature creep" (fast changing software features) versus simple and stable designs − Interpretation of integrated data	N/A
OGF	Computer science	N/A	N/A
OSG	1. Computer science 2. HEP and physics 3. IT professionals 4. Users from several fields	− Differing epistemic cultures − Different languages and jargons − Divergent objectives: cutting-edge research versus service provision	N/A
Swedish Nat. Data Service	1. Infrastructure developers 2. Data managers 3. Users from several fields	− Security of data − Public support for service	Low
Swiss BioGrid	1. Grid developers 2. Users	− Lack of funding − Sustainability − No collaboration outside one country	N/A
TeraGrid	Developers from TeraGrid and users	N/A	N/A

In order to deal with these challenges many projects engage in activities aimed at generating a common base of understanding among the different groups (see Table 5-12). Among these activities we find tutorials and training (e.g. CineGrid, EELA-2, D4Science), involvement of mediators or translators of user demands (e.g. D4Science, DEISA) or different forms of web-based support such as Wikis, FAQ pages, mailing lists (e.g. D4Science, DEISA).

Though it is impossible to measure the effects – as in most other cases when such activities are employed – the solution taken up by NVO seems to stick out, as it does not only intend to raise users' computing knowledge but also sensitizes developers much more for users' needs and possibilities: it follows a strategy that has focused on enhancing the flow of requirements from astronomers to computer science developers, and simplifying access to the VO through user-friendly portals for scientific research. A similar strategy seems to be implemented by D4Science that also tries to better capture user requirements through involving mediators.

	Measures	Outcome
Table 5-12: Measures to enhance interdisciplinary collaboration		
C3-Grid	Interdisciplinary task forces which convene face-to-face meetings	N/A
CineGrid	1. Learning phases at start of demonstration projects 2. Tutorials at annual workshop	N/A
CLARIN	N/A	N/A
D4science	1. Mediators translate end user requirements and reduce the complexity of communication 2. Workshops and conferences bring the different groups together 3. Website, Wiki and mailing list 4. Trainers from the two communities are trained by technical experts	N/A
DARIAH	Optimizing interoperability of resources	N/A
DEISA	1. Direct contact between domain scientists and supercomputing experts 2. Documentations and FAQs	N/A
DRIVER	N/A	N/A
EELA-2	Training activities like tutorials and Grid schools to increase users' Grid literacy	N/A
EGEE	N/A	N/A
ETSF	N/A	N/A
GÉANT	N/A	N/A
MediGrid	1. Website and mailing list 2. Video and telephone conferences	N/A
NVO	1. Iterative development process for portal design and tool creation 2. Direct involvement and feedback from users 3. Simplification of interfaces	Medium, more efforts are needed
OGF	1. Involve users in decisions, make them full partners 2. Simplify interfaces 3. Work with "brokers" who better understand user needs	Successful, managed to engage many more users in fields new to e-Infrastructure
OSG	N/A	N/A
Swedish Nat. Data Service	1. Open availability of data sets 2. Opportunity to deposit new data sets	Successful, demand for both services exists
Swiss BioGrid	Good computer science and life science collaboration	Project is finished,

	Measures	Outcome
		benefits may revive
TeraGrid	1. Simplify interfaces 2. Develop "gateways" that simulate user environment with an e-Infrastructure engine.	Successful, particularly science gateways.

5.7   Extending use

In seeking out new users, an essential part of becoming a sustainable production quality e-infrastructure, most projects employ several measures, though this does not have highest priority in all the investigated cases. Among these measures, the most common are (see Table 5-13):

- Tutorials and training
- Targeted communication to potentially interested organizations and individuals
- Presentations at conferences, workshops, events
- Word of mouth and social networking

Some projects invest considerable resources and try to increase their understanding of the user communities by cultivating relationships and developing solutions which particularly address their needs, like OSG and TeraGrid. Over the course of years these projects have realized that the recruitment of new users is something that necessarily happens once the technology is developed and made available to the communities, but that more efforts than merely raising awareness and training are needed.

At first sight it may seem astonishing that the recruitment of new users is not a top priority for all projects(see Table 5-13). There are different reasons for some cases not to invest too much effort into recruiting users:

- Projects may follow a sequential approach of technology development, innovation and diffusion. Then, being in an early phase of technology development with a primary focus on building, testing and improving the e-infrastructure, they postpone involvement of a broader set of users to later phases. NVO and D4Science seem to be cases which apply this strategy currently.

- Others, like CineGrid, do not want to serve a broader user community, but see their main purpose as technological innovators doing proof of concepts and demonstrations which may or may not be taken up at some later point in time by others.

- A third group of projects with few efforts in enlarging their user base are those which have low prospects of being continued in the future.

Table 5-13: User recruitment

	Measures	Importance	Results
C3-Grid	1. User-meetings with organizations from outside the consortium 2. Tutorials 3. Visits to potentially interested organizations 4. Conference presentations	High	Moderate for 1 and 2, as large effort for potential users
CineGrid	No dedicated activities, outreach activities, interest raised through presentations, demonstrations and performances at various events	Low	Few, community grows slowly, funding situation tight
CLARIN	No activities as yet, project is at an early	High	N/A

	Measures	Importance	Results
	stage, but user recruitment is recognized as a high priority		
D4science	1. Targeted recruitment and training by the project team and mediators 2. Workshops and conferences	Low	N/A
DARIAH	No activities as yet, project is at an early stage	N/A	N/A
DEISA	1. Europe-wide calls for proposals 2. Documentation 3. Training 4. Centralized help desk	N/A	N/A
DRIVER	1. Raising awareness at international events 2. Summer school for repository managers 3. DRIVER summits 4. User tutorials	N/A	N/A
EELA-2	1. User tutorials 2. Grid schools 3. Workshops, decision maker days; 4. Customized "Gridification weeks" 5. Conference presentations 6. Local promotion by members	High	Promising, growing number of users, progress with establishing NGIs
EGEE	1. New users appear as they are socialized into major user fields (as HEP) 2. Presentations of the infrastructure to potential major user communities 3. Collaborations with other organizations (e.g. DANTE, NRENs) to reach new users 4. Training sessions for potential users 5. Participates heavily in EC project dissemination and concertation activities	Medium	EGEE is "the" European Grid brand and most potential users turn to it; success of recruitment activities is hard to assess.
ETSF	1. Calls for proposals 2. Training events 3. Manuals and tutorials 4. Users' Newsletter	High	Good, satisfying percentage of repeated usage
GÉANT	N/A	N/A	N/A
MediGrid	None.	Low	N/A
NVO	1. Summer training workshops 2. Inclusion in teaching at member organizations 3. Social networks of members 4. Funding of small research projects	Low in the past, high in the future	N/A
OGF	1. Involved groups actively solicit adoption in their communities 2. Targeted surveys on user requirements 3. Inviting "lead users"	Low, but rising, engagement of new user communities planned	N/A
OSG	1. "Engagement team" to involve new users 2. User identification: workshops, "Grid schools", announcements at domain specific conferences, "cold" email correspondences or telephone calls 3. Cultivating relationships to understand the	High	Good, large and growing number of users

	Measures	Importance	Results
	particulars of users' research and technological environments and tailor suitable solutions 4. Regular meetings between OSG managers and major user communities 5. Social networks		
Swedish Nat. Data Service	Training and support offered (particularly for deposit of data)	Medium	Good, users are successfully recruited
Swiss BioGrid	N/A	Project benefits may be revived	Project benefits may be revived
TeraGrid	1. Traditional publicity mechanisms (website, press releases, announcements) 2. Dedicated training events 3. Science Gateways: community-specific portals 4. Campus Champions	High	Good, large and growing number of users

Most projects state specific catalysts and barriers which are influential – according to their knowledge and experience – in the adoption process (see Table 5-14). As to be expected, the strongest drivers towards adoption are access to data, computing power and other resources; nearly every project mentions one or more of these motivations. Becoming involved with other people with particular expertise and knowledge and the support to collaboration were also mentioned as influential in several cases.

Among the barriers we find a notable variety. A lack of knowledge about the technology – combined with insufficient time to benefit from training and support activities – and different facets of the immaturity of the technologies are named most often across the board. To our surprise, funding problems are mentioned only in relatively few cases (CineGrid, EELA-2, DARIAH), of which the majority address constituencies in the social sciences and arts & humanities. This may indicate that these domains still encounter problems in justifying their e-infrastructure involvement and setting up a sustainable funding of their efforts.

Table 5-14: Catalysts and barriers of adoption

	Catalysts	Barriers
C3-Grid	Access to data from different sources and of different types	1. Grid-specific knowledge missing 2. Inclination to computational research missing
CineGrid	1. Counterpart for networking experiments, 2. Exchange of know-how with a global community of excellence, 3. Reducing effort and costs of transmitting audio/video data 4. Forward-looking developments in the area of distributed content management and retrieval 5. General trends towards scientific visualization, digital cinema	1. Demanding high-speed fibre-network connections 2. Scepticism towards new technologies 3. Lack of expert knowledge 4. Funding
CLARIN	1. Access to data 2. Access to software to analyze data 3. Potential collaboration	1. (Potential) technological barriers 2. (Potential) Lack of funding
D4science	Obtaining access to heterogeneous data in a Virtual Research Environment	1. Complex process of use involving mediators to translate users' demands

	Catalysts	Barriers
		into infrastructure services 2. Production quality of infrastructure is very recent
DARIAH	Access to data	1. (Potential) Lack of funding 2. (Potential) Lack of computational, technological knowledge
DEISA	Proposed projects have a strong need for supercomputing resources	None mentioned.
DRIVER	N/A	N/A
EELA-2	1. Scarcity of computational resources 2. Strong interest in international scientific collaboration	1. Scarcity of funds 2. Low maturity of Grid technology 3. Limited time for learning Grid use 4. Application programming is not supported
EGEE	1. Well known European Grid brand 2. EGEE use is standard in some fields (like HEP)	1. Data protection needs of users from the industry can't be guaranteed 2. Set-up and use of middleware are complex and require considerable understanding
ETSF	Scientific interest	N/A
GÉANT	N/A	N/A
MediGrid	N/A	N/A
NVO	1. Growing importance of multi-wavelength astronomy 2. Useful tools, such as a "name resolver" for celestial objects	1. Time investment to learn and utilize e-Infrastructure designs & technologies 2. Not familiar with Grid use and limited time for learning
OGF	1. Monopoly status of specifications in the e-Infrastructure space 2. Availably of existing distributed infrastructures based on OGF standards	1. Complexity: some standards are lengthy and difficult to implement 2. Availability of alternatives, e.g. virtualization and cloud computing
OSG	1. Personal interest 2. Access to experts 3. Access to distributed high-end computational and data resources	1. Sensitive data 2. Lack of trust
Swedish Nat. Data Service	1. Access to data 2. Facilitation of collaborative research	Sensitive data
Swiss BioGrid	Small project easy to manage	Project finished, partly no follow-up because of lack of infrastructure
TeraGrid	Need for high-end computational, data or data visualization resources	1. System not delivery oriented and at times unreliable 2. Long waiting times and latency 3. Technical design problems (cumbersome interfaces, software, unstable system)

5.8   Governance structure

The governance of projects and infrastructures that are so varied in terms of scale and resources is bound to involve considerable differences. Indeed, we see in the table below that

there are a number of subdividing roles and bodies that are responsible for managing and guiding the projects. What we also see, however, is simply that larger and more complex infrastructure projects, as we might expect, also have larger and more complex governance arrangements. However, it is also likely that the greatest variety in governance arrangements is likely to be – not in the chart-like information about formal organization that we have captured here – but in the degree of hands-on as against more laissez-faire practical governance practices, which are difficult to capture and assess.

Table 5-15 Governance structure

	Size and composition	Division of labour
C3-Grid	Consortium of 8 data providers, 8 operators, 2 informatics partners, and 3 universities	1. Project lead coordinates 2. Scientists specify requirements 3. Data providers supply data 4. Informatics partners supply middleware
CineGrid	1. Board of directors 2. Executive committee 3. Advisory committee	Run by consultancy firm
CLARIN	1. Scientific board (scientists) 2. Strategic board (funding reps) 3. Executive board (8 experts) 4. International advisory board	Multi-tiered distributed structure
D4science	1. External advisory board 2. Members general assembly (one rep for each of the 11 partners) 3. Project management board 4. Community managers 5. Project executive board 6. QA task force	Project coordinated by central office Series of task specific managers to direct work (technical, outreach, service, research)
DARIAH	Six institutions in a three-tiered structure (research, services, standards)	Managed by consultancy
DEISA	1. Executive committee of 11 partners and 4 associate partners 2. Technical board 3. Advisory scientific committee 4. Presentation team	Project coordinator chairs executive committee and manages project
DRIVER	13 partners	Project coordinator maintains services Technical partners develop software
EELA-2	1. Management board 2. Technical board 3. Consortium board (16 national partners)	1. Management board runs operations 2. Technical board deals with technical details 3. Consortium board makes strategic decisions 4. 16 coordinators lead national networks of organizations
EGEE	1. Administrative Federation Committee (AFC) 2. Activity Management Board (AMB) 3. Collaboration Board (CB) 4. External Advisory Committee (EAC) 5. Project Management Board (PMB) 6. Technical Management Board (TMB)	Project coordinator chairs activity management board and manages project; Technical Director and TMB coordinate technical progress
ETSF	1. Steering committee (includes	1. Decisions made by steering committee

	Size and composition	Division of labour
	representative from 11 core groups) 2. Governing board 3. Advisory board	2. Working teams assembled on *ad hoc* basis for specific sub-projects
GÉANT	1. Policy committee with all 30 members 2. Executive committee with 5 elected members, plus 6 non-voting members	1. Activities divided by work packages 2. Policy committee meets 3-5 times per year 3. Project managed by DANTE
MediGrid	Speakers board, with all 8 partners	Each of the 8 projects is managed individually Speakers board manages cooperation
NVO	17 partner institutions; Executive committee of senior personnel	Executive committee sets priorities Technical and science working groups carry out projects
OGF	Open community in the thousands, with core group approximately 200 individuals representing dozens of institutions) Hierarchical structure of "functions", which are divided into "areas", which are further divided into "groups"	Decisions at all levels made by "rough consensus"
OSG	53 partner institutions; Council, with central stakeholders Executive board	Council has monthly teleconference to make strategic decisions Executive board directs work
Swedish Nat. Data Service	Advisory committee	Run by a director and small support staff
Swiss BioGrid	Steering committee of 6 partners	Project coordinator assigned tasks
TeraGrid	11 partners Forum Infrastructure group Science advisory board	Distributed via matrix structure which allows for non-direct supervision and geographically dispersed teams

In terms of the organization of governance, there is a scale from the small and informally organized (CineGrid is an example) to larger multi-tiered and more elaborate and complex structure (Geant). One feature that is common to all larger projects is an advisory or steering committee of some sort (in some cases both, such as for Clarin) – a group which oversees the project and guides the management level. These are sometimes internal, sometimes external. They are also sometimes constituted so as to provide guidance, sometimes more to ensure 'democratic' representation from among all project members or stakeholder groups. Further, it is noticeable that in some cases, both the advisory or steering committees and the management group seem to come from among the researchers and from within the disciplines themselves (NVO), whereas in other cases a broad constituency from across disciplines is represented (Swedish National Data Service). A further dimension is whether the governing bodies are permanently constituted and include core staff that is constantly occupied by governance tasks, or if there is only episodic governance by means of regular face-to-face or teleconferencing type meetings.

More versus less centralization is of course a key factor in governance, but it seems that, unless projects are so small as to be a 'one man show' (SwissBioGrid), there are either one or a few coordinators who delegate tasks, with only the larger projects in addition having a larger more representative body which coordinates and delegates tasks. Only in a few cases (OGF, TeraGrid, OSG) is there a move away from a centralized towards a more federated or 'flat' organization which has multiple coordinators for different tasks (though some projects have such a body underneath the centralized coordinator or coordinating body). Apart from

centralization, the main variety in governance comes from the high or low degree of division of labour.

Whether there is good match between governance structure and the project functioning is difficult to generalize about. What is clear is that a variety of governance styles is possible, and that oversight and strategy as against management are separated in all the cases of larger infrastructures.

5.9   Internal & external communication

Communication within projects and between projects and the outside world of user communities and other stakeholders is handled in quite different ways. While internal communication is mainly mixed (via a variety of channels) and mainly a question of effectiveness (which is hard to assess, but mainly a practical problem), external communication is much more uncertain and varied in terms of how much attention to this the different e-Infrastructure projects are paying. At the same time, external communication is much more than a practical problem since it is likely to be vital to the long-term success of the project. In the case of external communication, the different strategies adopted or envisaged seem to be much more open-ended.

Internal communication ranges from project-internal teleconferences and email circulation of information only (as in the case of SwissBioGrid) to more elaborate internal meetings and regular forms and channels of communication. Project websites, which are mainly for external communication, are also sometimes used for internal project communication (for example, about stages of completion of work packages, as in the International Virtual Observatory Alliance - IVOA - of which NVO is a part). Typically, internal communication is via a mix of channels, and a particularly interesting topic to pursue (though it goes beyond the scope of the current research) is determining the effectiveness of different internal communication tools (which, in some cases, are also used for external communication) – not only looking at websites, but also at project management software, Wikis, and collaborative platforms such as BSCW.

External communication, as in relation to other issues, cannot be divorced from whether the infrastructure projects are at a stage of engaging with users or stakeholders or if they are still at the development phase only. In the latter case, the question of external communication mainly relates to how they *plan* to do this in the future. Some such infrastructures at the development phase clearly have elaborate user engagement plans (Clarin is a good example) while in others these plans are as yet unspecified.

OSG has a dedicated 'engagement team', which provides a model at the successful end of the spectrum. This also applies to the extensive engagement that projects like EGEE foster through their user community conferences. Such proactive outreach seems to be vital for all projects except those where the user communities already exist in a pre-given form.

5.10 Main technologies, resources and services

5.10.1 Providers of computing and network services

Core technologies: supercomputing and grid computing. Supercomputing provides high-performance processing, with very low latency and high reliability. Additionally, it provides large, online data storage, as well as advanced computationally hungry visualization technology. Expensive, specially designed machines are used to enable high-end capabilities,

and the purchase of these machines accounts for the relatively higher funding supercomputing providers receive. In the two studied cases that focus on supercomputer provision, DEISA2 and TeraGrid, developers use Grid technology to connect specialized supercomputers in different locations.

Grid providers, in contrast, utilize grid technology to aggregate available computing cycles and data stores in participating, geographically distributed institutions. The operational cost of grid-based providers is considerably lower than supercomputing providers because no specialized expensive equipment is necessary. Rather, capacity is a result of engaging additional institutions. On the other hand, grid e-Infrastructures such as EGEE and OSG grid, cannot compete with the lower latency and high-reliability levels found in the supercomputing provision model.

Geographic scope: national and international. Some providers, such as EGEE and DEISA typify the distributed international model, with partners across most European countries In contrast, OSG and TeraGrid, the two main US infrastructures, are comprised of similar types of organizations within the country.

Table 5-16: Main Distinctions among e-Infrastructure Providers			
Core services			
National distribution	Supercomputing	Grid	Network
National	TeraGrid	OSG	NRENs, Lambdarail, Internet2 (not included in study)
International	DEISA2	EGEE, EELA-2	GEANT

Basic Grid services are mature and in common use. Our findings confirm that much of the development of basic distributed technologies providers use has been completed and that these are in wide use. e-Infrastructure providers including TeraGrid and OSG in the US and the European EGEE or DEISA2 are using well-established middleware solutions. Among these technologies are different versions of the Globus Toolkit, Unicore, the Virtual Data Kit and Condor. Informants involved in Asian e-Infrastructures interviewed by eResearch2020 partners in related studies further indicate that this is a global trend; NAREGI and others are also using these common technologies. In fact, having recognized that after over a decade of research and implementation core middleware technologies are sufficiently mature US funders have stopped supporting these developments.

Considerable differences among providers make interoperability difficult. The ability to have one infrastructure accept jobs from others, and the transport and management of data and instruments across infrastructures are key aspects of e-Infrastructure. They are also a goal of most of the providers included in our study. We find this to be more of a long-term vision than a goal that can be soon accomplished. Paradoxically, although e-Infrastructure providers are using the same basic technology, they also have developed their own middleware packages. Our informants indicated that this development is essential. They also noted that it is difficult and time consuming because each package needs to accommodate unique hardware, such as proprietary supercomputing machines, or specialized user environments that require the construction of portals to a particular community—for example, "Science Gateways" in TeraGrid.

Cloud computing poses a real threat. According to multiple informants, the salience of Grid computing in the commercial sector has considerably diminished in the past few years, moving instead to cloud computing (Miller 2008). One reason for the growing popularity of clouds is that as opposed to Grid computing and web services that interoperate across distributed infrastructures by exposing infrastructural detail, clouds, and more generally Web 2.0 expose

almost nothing. Hiding complexity enables a simple interface that requires very little knowledge from users, make clouds a strong commodity alternative to Grid technology. As a result, most large IT vendors have moved away from Grid systems to clouds, as they provide more immediate commercial benefit. To what extent this development in the commercial sector applies to the academic sector, however, remains to be seen (see section 2.3.6). In any event, funding to the OGF, the main e-Infrastructure standardization organization, has thus been considerably cut.

5.10.2   Providers of data and analysis tools

Metadata integration is mature and well established. Since scientists rely on different data standards based on their sub-field orientation and use of certain types of instruments, much of the work of domain e-Infrastructure projects relate to the development of data protocols and standards that enable the analysis of data from various types and across locations. A typical example is the National Virtual Observatory (NVO). As a part of a larger international effort to develop protocols and standards for the astronomical community, NVO has developed specifications and protocols that enable the aggregation of observational data from multiple observatories in the US and across multiple wavelengths, leading to a revolutionary potential in astronomy.

Figure 5-2: Common Layers of Technology Development in Domain e-Infrastructures
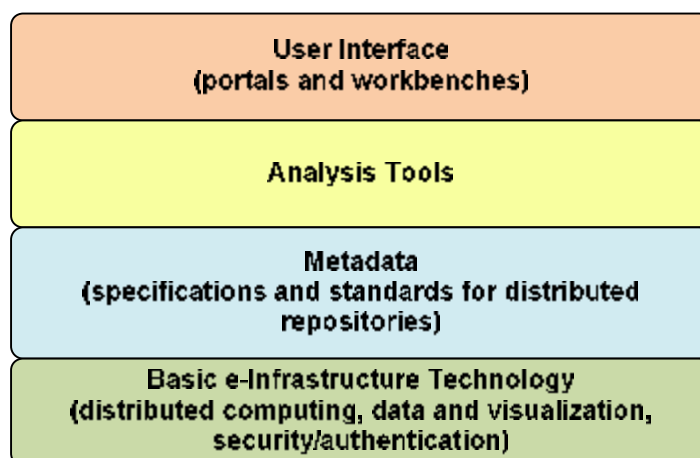
e-Infrastructures will advance to the development of specialized tools and interfaces. Against this background, we anticipate that the developments in the "upper" levels of technology development in domain e-Infrastructures—as illustrated in the above figure—to assume a more central role in the coming years. Although these developments require less integration among e-Infrastructure formats, they require more integration into user environments. This is a substantial challenge. For example, informants in the US-NVO indicated that their users desire to focus their resources on research, and thus have little tolerance to accommodate cumbersome designs or new technologies that considerably differ from their own. Thus, the NVO, DEISA2, TeraGrid and OSG envision directing considerable efforts into gaining better understanding of user environments and to develop adapters, or "gateways" to enable e-Infrastructure use. In other areas, like data for the life sciences, health and social sciences, there are still major issues to be tackled, both on the side of data integration and privacy and security technologies.

Table 5-17: e-Infrastructure Development Stage

	Metadata	Analysis Tools	Portal and Interface
C3-Grid	Mostly completed	Under development	Under development
CineGrid	Under development	Under development	Under development
CLARIN	Under development	Under development	Under development
D4science	Mostly completed	Mostly completed	Mostly completed
DARIAH	Under development	Under development	Under development
DRIVER	Mostly completed	Mostly completed	Mostly completed
ETSF	N/A	Under development	Mostly completed
MediGrid	Mostly completed	Mostly completed	Mostly completed
NVO	Mostly completed	Under development	Under development
OGF	N/A	N/A	N/A
Swedish Nat. Data Service	Mostly completed	N/A	Under development
Swiss BioGrid	N/A	Completed, may be revived	Completed, may be revived

5.10.3 Approaches to the development of specialized tools and interfaces

Three approaches to the development of specialized tools and interfaces. To accommodate these challenges several of the studied e-Infrastructures—including provider and domain e-Infrastructures—have shifted their efforts to work more closely with their users.

- *Working with "lead users" (provider and domain infrastructures).* We found in domain e-Infrastructures such as NVO involves actual work or plans to work more closely with users to gain direct input and simplify the interface of the portals and tools.

- *Generating field-specific environments (provider infrastructures).* The second approach involves the generation of an e-Infrastructure environment that is unique to a field, such as physics, astronomy or biology. Since 2004, TeraGrid has spent much effort in developing its "Science Gateways" technology. Using standard Web interfaces, gateways portals serve to connect users from a diverse number of scientific communities with supercomputing and Grid middleware, sometimes without users realizing that they are using these resources. The idea of Gateways is not just to serve a single laboratory; rather the aim is to open the Gateway to an entire community, which may consist of thousands of additional users, such as in the case of nanoHUB (http://nanohub.org/). Another example is D4Science, which addresses the development of field-specific virtual research environments (for the Environmental Monitoring EM and Fisheries and Aquaculture Resources Management FARM communities) exemplifying the further insertion of technologies developed in predecessor and parallel projects (Diligent, EGEE) into potential user communities.

- *"Brokerage" (provider and domain infrastructures).* A third approach that appears to be less common but offer substantial benefits is to "broker" the development of tools and interfaces to a partner that is more familiar with the requirements of a certain field, while affording users with basic e-Infrastructure services through a partner. A good example of this approach is the relationship between the Structural Biology Grid (SBG, http://www.sbgrid.org) and OSG. Charging a modest annual fee of $3,500 from its members, the SBG, which involves a small team of biologists and computer scientist at Harvard serves scientists from over a hundred structural biology labs throughout the US. Recently, SBG has expanded its operation to also serve several international institutions.

Table 5-18: Approaches to the development of user environments in the studied cases

	Lead users	Field specific	Brokerage
C3-Grid	+	+	-
CineGrid	(+)	-	-
CLARIN	+	+	-
D4science	+	+	+
DARIAH	-	+	+
DEISA	+	+	-
DRIVER	+	-	-
EELA-2	+	-	-
EGEE	+	+	+
ETSF	+	-	-
GÉANT	+	+	-
MediGrid	+	+	-
NVO	+	+	
OGF	NA	NA	NA
OSG	+	+	+
Swedish Nat. Data Service	+	-	-
Swiss BioGrid	+	+	+
TeraGrid	+	+	

Note: "+": implemented in the case; "-": not implemented in the case

5.11 Inter-organizational collaboration

The e-Infrastructure ecosystem is composed of a dense network of participating organizations, which may be divided in internal and external organizational collaborations. Domain infrastructures involve internal collaborations that aggregate core competencies to develop and offer advanced distributed services to constituents within a field. Each of these constituents is also a part of an organization. Provider infrastructures are based on partnerships or consortia of national or international organizations, and their users are spread across many institutions. Furthermore, providers collaborate with peer providers, regionally and internationally, as well as with domain infrastructures, standardization organizations and in some cases firms. For example, a large number of European e-Infrastructures we studied are associated with EGEE and D-Grid, both have ties with US providers and take part in the OGF.

Figure 5-3: Inter-organizational Collaboration Structures in e-Infrastructure

Interorganizational Relations
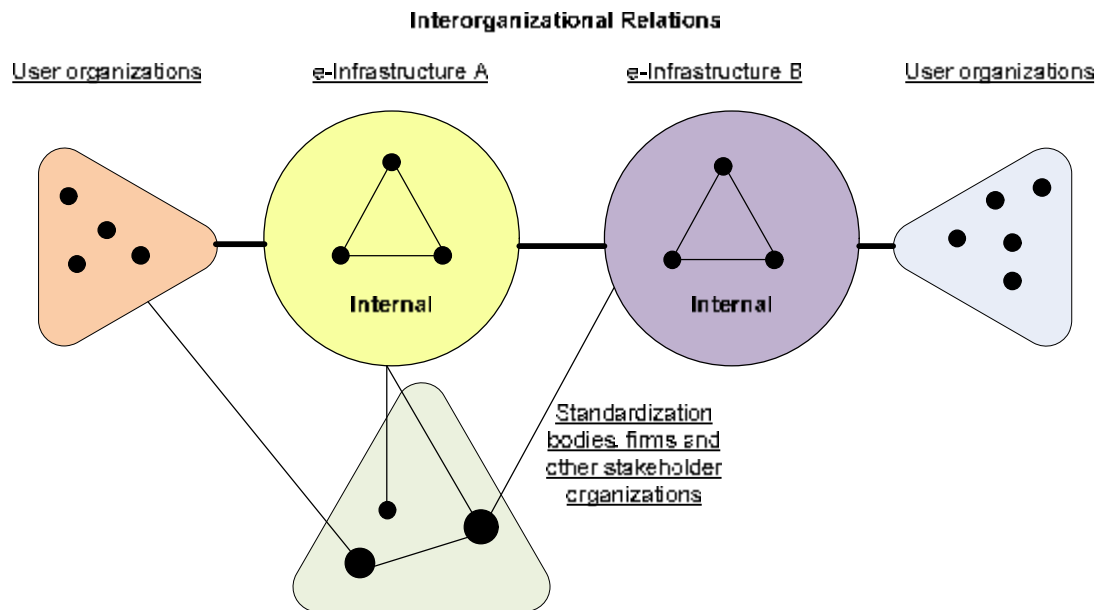
1. By sharing areas of specialization and exchanging knowledge, inter-organizational collaboration drives innovation development and the advancement of e-Infrastructure. However, cultural and technical differences among partner organizations in an e-Infrastructure often lead to collaboration barriers (Cummings and Kiesler 2008). Cultural differences can be broken down into two aspects: field affiliation and identity. Related technological distinctions further exacerbate these differences.

   - *Field affiliation.* Being affiliated with different fields exacerbates collaboration because each field has its own paradigm and utilizes different scripts of action. e-Infrastructure providers appear to suffer from these barriers partly because they join organizations that are associated with diverse fields (see also section 5.6 on interdisciplinary collaboration). For instance, TeraGrid informants have reported to experience difficulties while integrating the much different US Department of Defense culture found in national laboratories with that of supercomputing centres or academic research centres.

   - *Identity.* Organizational identity hinders collaborations when participants need to shift to work under a different, collective banner of the e-Infrastructure, such as DEISA2 or TeraGrid. When an e-Infrastructure requires lower levels of embedding—being a part of the collaboration is less involving for participating organizations, such as in EGEE or OSG, informants report less friction.

   - *Technology.* Organizations either employ different technological systems, or they use the same systems in different ways (Barley 1986; Orlikowski 1992). e-Infrastructure providers that integrate resources available in participating institutions, then, find it difficult to accommodate technological peculiarities. This is a particularly challenging problem for supercomputing providers, as integration among specialized supercomputers is more complicated than among commodity systems.

2. Competition among partners reduces trust and overall productivity. Competition that is an essential ingredient in innovation in the short-run may thwart e-Infrastructure collaborations. We found this barrier to relate to organizational identity. For example, informants observed competition among TeraGrid partners to partially be a result of inability to put forward the shared interest of the entire collaboration. They further noted that to remain competitive partners need to achieve external recognition—typically from funding agencies—to their own work or to the resources they invested. Under these conditions, organizations tend to avoid situations where credit goes to the entire collaboration.

3. Long term collaboration enhances trust and facilitates development. Informants indicated that there are several tension-mitigating factors to collaboration with e-Infrastructure partners. Ongoing relations among organizations are known to foster trust and facilitate stronger cooperation (Uzzi 1997). In the case of TeraGrid, even entering organizations that had a past relationship with one of the major providers—for example having staff previously working in that organization—found cooperation relatively pain free.

4. Coordination and communication are costly in the short-run but reduce overall collaboration barriers. Better communication can ameliorate the free-riding problem; a more transparent and updated sense of each others' activities leads to better scrutiny, self regulation of conduct (Olson 1965). However, distributed organizations, and virtual organizations in particular (Cummings and Kiesler 2008), require higher investments of time and resources to coordinate their efforts. The growing scale of the project, which includes diverse organizations working in a matrix fashion on a complex array of tasks leads to high coordination costs. This was found to be effective in TeraGrid and OSG.

5.12 External Organizational Relationships: Interoperability, dependencies and standards

Interoperability is inherently about joint operation of otherwise distinct infrastructures. This model is different from relationships among *internal* collaborators who work on a single infrastructure, and typically involves more formal organizational arrangements (e.g. partnership or consortium). Most of the external efforts take place among peer organizations that have a similar scope of operation, either for updating ongoing activities or to interoperate. For instance, standardization bodies, such as the OGF, aspire to specify compatible standards to those other standard organizations develop. Some e-Infrastructure providers coordinate their work to facilitate interoperability that would enable the establishment of a global computer infrastructure.

Synchronizing developments with peer institutions enables interoperation. In this model an infrastructure learns about the activities of its peers, and takes these into account when developing its core technologies. The main purpose of this process is to ensure that developments do not drift too far apart from other technologies that other organizations develop. This type of interoperation is common in cases where there is low dependency among e-Infrastructure technologies. It is also typical where there is competition among infrastructures. A good example of lower level interoperation is the OGF. OGF has instituted a formal liaison function that includes individuals from member organization whose role is to participate and monitor ongoing developments in other standardization organizations that compete as well as complement OGF's areas of coverage.

For example, the Storage Networking Industry Association focuses more closely on the development of offline and online, distributed storage systems. But the OGF may develop Grid related storage standards. OGF liaisons update peer standard bodies on activities that take place at the OGF. In turn representatives from these organizations may present key developments that take place in related areas. While many standardization activities occur behind closed doors, these external relationships aid standard organizations to channel developments in appropriate tracks. Similarly, e-Infrastructure providers we studied maintain relationships with peer infrastructures to gain a more detailed understanding of their expected development trajectory.

Synchronization of developments is also important for e-Infrastructure sustainability. As mentioned above, cloud computing poses one of the more severe threats to the sustainability of grid computing, and by extension to e-Infrastructure. At the same time, interviewees have pointed out that the emergent cloud technology suffers from relatively weak management

capabilities across clouds. Arguably, as suggested by some of our sources, the grid computing community has already solved this problem. The most recently established OGF group that was formed this year (2009), Open Cloud Computing Interface Working Group, thus seeks to facilitate interoperability and connect past work in grid computing with the emergent cloud computing.

Interoperation also occurs among peer e-Infrastructures that do not necessarily have ongoing, formal relationships. In these cases, projects collaborate with their peers, often from other countries, toward the development of a common or interoperable specification, to enable access to globally distributed resources, to transparently roam from one infrastructure to another, or to use data standards that all infrastructures accept. US NVO and its peer astronomical infrastructures offer a useful example of this model. NVO assumes a central role in a network of international e-Infrastructure astronomy projects called the International Virtual Observatory Alliance, or IVOA. Members of the NVO regularly attend IVOA meetings, chair working groups and exchange information with their international peers. Synchronizing these operations has enabled joint development of common astronomical data standards, which serve as grounds for a second stage production facilities that will enable astronomers transparent—or close to transparent—access to diverse data sets from around the world. However, astronomical data may be rather unique in being relatively easy to standardize and share.

When having a different scope of operation, interoperability among providers is more challenging and unlikely in the short-run. The ability to have one infrastructure accept compute jobs from others, as well as the transport and manage data and instruments across infrastructures, is a key aspect of e-Infrastructure and a goal of most of the infrastructures included in our study. Such upper level interoperability requires a higher degree of technological and organizational coordination. Although e-Infrastructure providers are using the same basic technology, they also have developed their own middleware packages. Our informants indicated that this process has proven to be difficult and time consuming because it needs to accommodate unique hardware, such as supercomputing machines, or specialized user environments that require the construction of portals to a particular community—for example, "Science Gateways" in TeraGrid—providing a common set of tools across the infrastructure to domain users, or requiring certain levels of service.

Developments that are specific to an e-Infrastructure lead to divergence among providers. This is a considerable challenge for infrastructures such as the OSG that must ensure interoperation with EGEE and the Nordic National Grid Infrastructure (NorduGrid) in the face of independently evolving software and processes. However, EGEE uses an older version of the Globus Toolkit for its gLite middleware. OSG, in turn, has developed a different package, on top of Condor, and using a newer version of Globus. While both infrastructures cater primarily to the high-energy physics community, the OSG and EGEE are working to interoperate their infrastructures, with some demonstrated success.

These focused efforts are supplemented by broader interoperation attempts. For the past few years, the 'Grid Interoperation Now' (GIN) community group in the OGF attempts to identify and then develop a set of standards to enable major academic e-Infrastructures to work with one another (Riedel et al. 2008). According to our informants, to date these efforts have only achieved partial interoperation; other, more pessimistic informants portrayed this work as a "heroic effort," speculating that upper level interoperation among the various e-Infrastructure providers is certainly not likely to happen in the short run, perhaps even not in the long run.

5.13 Recommendations to policy makers

Recommendations were not mandatory in our interviews. However, in future reports and in the Roadmap, a considerable number of recommendations will be gleaned or inferred not just from the interviews, but also from other materials collected. Hence the paucity of responses at this stage is not a major concern. Below we have divided the recommendations into those relating to 'funding and sustainability' and a catch-all category of 'other'. In both cases, as can be seen, the responses were, as might be expected, quite diverse.

Table 5-19 Projects' recommendations to policy makers

	Funding and sustainability	Other
C3-Grid	No data	No data
CineGrid	No data	No data
CLARIN	No data	No data
D4science	No data	Infrastructure 'ecosystems' may become too complex
DARIAH	No data	No data
DEISA	No data	No data
DRIVER	Funding should demand open access to results	Open access needs advocacy and mandating
EELA-2	No data	No data
EGEE	Continue to fund scientific communities' direct use of e-infrastructure to encourage uptake in the communities (e.g. INFRA-2010-1.2.3: Virtual Research Communities); Do not give funds to individual researchers or communities for computing equipment unless they accept to connect it to a shared infrastructure; Only fund software development by research communities if they agree to distribute it under an open-source license;	Encourage convergence of e-Infrastructures by insisting on interoperation as a key objective for funding.
ETSF	Secure funding.	Beyond funding user projects, thought should be given to funding also training of young scientists and the collaboration on fundamental scientif problems between the nodes.
GÉANT	No data	No data
MediGrid	Greater cost transparency is needed IT overhead funding is needed	Legal issues surrounding transfer of data need clearing up
NVO	Funders should support focused staff	Tool development should be driven by users, not computer scientists Reward work on infrastructure
OGF	Funders should encourage participation by funding mechanisms	Move towards cloud computing
OSG	Grass-roots approach is not enough, needs supplementing with top-down requirement from funders for collaboration	No data
Swedish Nat.	No data	Sweden is in a unique position to share

	Funding and sustainability	Other
Data Service		sensitive micro-data, but this remains to be fully implemented
Swiss BioGrid	Embedding in a longer-term infrastructure needs to sustain the gains of a small project	Bottom-up user-led development is good for science
TeraGrid	Longer funding cycles are needed for the time horizon of infrastructures	More involvement of funding program officers so they recognize complexities

What we can see is that quite diverse suggestions for ensuring the funding and sustainability of infrastructures are envisaged. In some cases, it is the mechanisms of funding that need to be improved (longer cycles, funding targeting certain aspects), while in other cases, funding should be used to promote certain project goals (open access, open source, more participation).

As for the miscellaneous other suggestions, we see that these vary by what the infrastructure is trying to accomplish: Teragrid, for example, needs to maintain a complex structure, whereas OGF needs to adapt to changing technological circumstances. Or again, the Swedish National Data Service and Medigrid both need better regulation for sensitive data, but this will not be an issue for most other infrastructure projects. EGEE points to stepping up funding regulations for leveraging interoperability.

The suggestions made by e-Infrastructure projects will provide a good indication of the scope of their concerns for the future and perhaps allow a ranking in terms of the priority of these concerns, but we need to be careful in drawing too many conclusions tied to the suggestions for individual projects which runs the risk of losing the 'bigger picture'.

5.14 Role of e-infrastructure in virtual research communities

We find two different views on conceptualizing and measuring the relationship between technology and society: one is usually called "technological determinism" receiving often a strong negative connotation among social scientists and scholars of science and technology studies. Technological determinism follows a top-down logic according to which external, largely independent and fixed technologies determine or force change in the social system that can be measured, modelled and predicted (for an extensive critique to technological determinism see for instance: McLoughlin, 1999). In opposition to technologically deterministic thinking a number of different constructionist theories have been developed which reject at large the expectation that technology is a driver of social change and instead argue for a more balanced analysis. These constructionist approaches partially differ widely but most of them agree that a better conceptualization of the interaction between technology and society is one of a mutual shaping of both: non-generic, configurable technologies are understood, adapted and used in different ways and according to the (dynamic) needs of specific communities (see Edge, 1995, McLoughlin, 1999, for an overview).

At the current point in time, we would categorize most e-Infrastructures still as non-generic (with the exception of research networks) and therefore rather the later type of perspective seems adequate, trying to understand how e-Infrastructures become e-Infrastructures and how they are adapted and appropriated in different research communities. The previous sections provided some evidence along these lines and we would now like to describe, how the informants on the different e-Infrastructure cases – in most cases the people who run and coordinate them and are responsible for providing the services – perceive the impact of their infrastructure in the wider community. Table 1-3 in the annex contains an overview of the

contributions that each infrastructure has achieved by the time the case studies were conducted in spring 2009. If we categorize these achievements into technical, scientific, socio-cultural (which includes political integration), other and none we get the following result as shown in figure 6-1. Most projects listed contributions in one or two areas, only few contributed in three areas and none in all four areas.

Figure 5-4: Main contributions of the cases by type
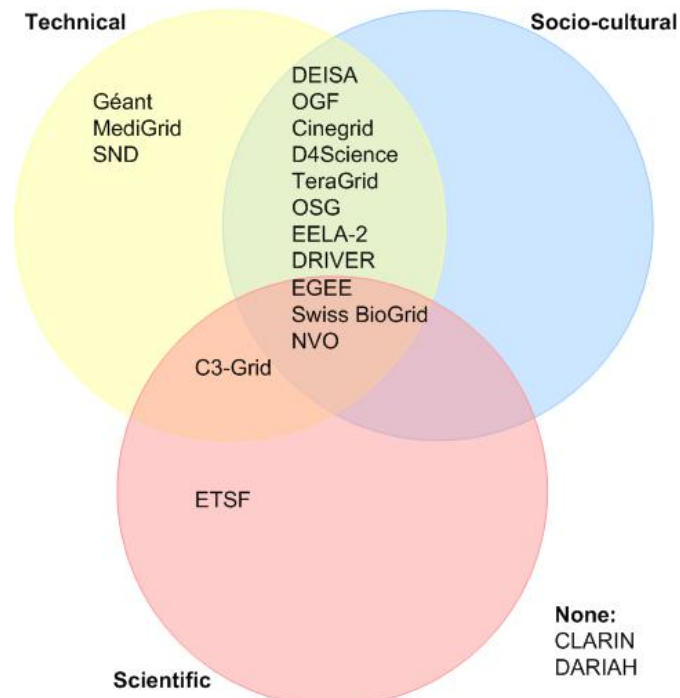
No contributions: CLARIN and DARIAH see overview of contributions Table 1-3 in the annex.

Scientific contributions were mentioned in 5 out of 18 projects, though we probably can take for granted that the other large infrastructures, DEISA, OSG, and TeraGrid, also contributed to some extent to the advancement of the fields which use them. We would expect that our informants, which were in most cases coordinators and managers of the projects or project parts, are only partially aware of the projects' contributions to science.

We note that out of 18 projects 15 projects or more than three quarters point to technical achievements in the widest sense as major outcomes. These achievements can be new software, tools, standards, proof of concepts of technical or organizational solutions, access to data, access to computing resources, increased data transmission capacities or the like – the broadness reflects the wide range of cases included in the study. However, a key point is that these contributions do not yet directly reflect scientific advancements from the research communities surrounding the infrastructure. This pronounced position of technical outcomes is probably a consequence of the early phase of e-infrastructure development for which the investigated projects mostly stand, in which technical progress was necessary and a main focus of nearly every project. Some technical progress can be taken for granted in virtually all projects, as out of the three projects not listed here two did not report any contributions at all (CLARIN and DARIAH are still in an early phase). Finally, it is worth pointing to the survey results, where access to resources, organizational (benefits) technical capacities, but also ease of use, funding and training are mentioned as catalysts (but also as barriers).

The socio-cultural (including political) contributions refer in particular to the establishment of new initiatives and organizations for the long-term provision of e-infrastructure services to science and technology, such as National Grid Initiatives, international organizations and the

like. Projects like EELA-2 had defined this as a major goal right from the start whereas for others – the Swiss BioGrid is here a case – this turned out to be the best path to reaching sustainability (even the Swiss only belatedly started a national grid infrastructure, after the project finished). Having said this, we would certainly hope for more projects going along these lines in the future as this is seen as a major catalyst by survey respondents.

Socio-cultural contributions or achievements refer to one of the core interests of this study, namely the impact of e-infrastructures on research communities. Notably in more than half of the cases our informants listed this type of impact. Taking a closer look at the communities which benefited, we find that in four cases – DRIVER, EELA-2, EGEE, and TeraGrid, all providers of e-infrastructure services, and the standardization effort OGF – it was mainly communities of e-infrastructure developers, such as supercomputing, Grid computing, or digital library communities, which received a boost from the involvement in the projects. In another set of projects – DEISA, ETSF, and NVO – the effects of community building were felt (according to our informants) more in the user domains. For ETSF and NVO this could be expected, as they are essentially community efforts of physicists, respectively astronomers. DEISA is somewhat an exception, but the classification is due to the fact that DEISA facilitates access to supercomputers for international projects and thus strengthens community networks. A third type of projects contributes to better interweaving e-infrastructure communities with user communities: CineGrid links networking, scientific visualization and digital cinema/media experts, D4Science brings Grid and digital library services to environmental monitoring and fisheries and aquaculture management communities, and last but not least OSG provides computing infrastructure to several fields. We do not wish to rank these socio-cultural contributions and projects in any way. They are all highly valuable results of an emerging e-infrastructure landscape that is still under development. However, the impact on domain-based virtual research communities outside of e-infrastructure development is probably higher in the second and third group of projects than in the first one. This stresses again the necessity for e-infrastructure developments to engage with user communities as early and comprehensively as possible.

It is notable that economic or commercial contributions were not mentioned by any of the cases, merely the involvement of partners from industry and private businesses was considered as a success in a few cases (NVO, OGF). Such commercial effects could be for instance spin-offs in the form of private service providers, partnerships with existing service providers, revenues through services to the private non-research sector or the like. For some cases such effects are largely prevented because of the contracts or the statutes of the funders which limit services to non-academic undertakings and purposes (e.g. EGEE, Géant and or CineGrid mentioned this explicitly). For others such commercial effects may exist but still at small levels and not worth mentioning, or the development stage may not yet be advanced enough for generating such effects. Last but not least, we also have to reflect that none of the projects included economic or commercial goals among its core goals (see section 5.2 above).

If we cross-tabulate this typology of contributions with another variable, the international versus national reach of the infrastructure, we obtain a further interesting result: Projects with socio-cultural impact are all multinational, except for three US-American projects. None of the four European national e-infrastructure cases that we investigated (C3-Grid, MediGrid, Swedish National Data Service, Swiss BioGrid) mentioned contributions in community building. This might be just a coincidence; however, we would rather make a different argument: scientific communities in European countries reserve national "playgrounds" for themselves and don't show a lot of interest in opening them up to colleagues from other countries. Individually very different reasons may explain such behavior, in particular technological and economic competition, but also data protection (and trust in its maintenance) or specific national research interests and paradigms. Research communities are then sealed off to outsiders and an extension or uncontrolled increase of dynamics is of little interest to the

insiders.[17] International projects and European projects in particular may dissolve these stalemates, whenever they manage to formulate goals of a higher priority and benefit.

	Scientific	Technical	Socio-cultural	Other	None
Table 5-20: Main contributions of the cases by type and geographical scope of the infrastructure					
International infrastructure	EGEE ETSF	CineGrid D4Science DEISA Driver EGEE EELA-2 Géant OGF	CineGrid D4Science DEISA Driver EELA-2 EGEE ETSF OGF	Géant	CLARIN DARIAH
National infrastructure	C3-Grid NVO Swiss BioGrid	C3-Grid MediGrid NVO OSG SND Swiss BioGrid TeraGrid	NVO OSG Swiss BioGrid TeraGrid	NVO	

See overview of contributions in Table 1-3 in the annex.

In addition, we also distinguished the contributions of the e-infrastructures between mono- and multidisciplinary infrastructures and between infrastructures producing computing services versus those that produce access to data or other resources (see annex tables 1-4 and 1-5). The differences do not reveal a particular pattern except for one issue: computing infrastructures are more active when it comes to socio-cultural (political) integration (see Annex table 1-5). Clearly this is a side effect of the currently ongoing institutionalization of Grid computing in national initiatives, a phenomenon that can not (yet) be found to the same extent for data infrastructures.

Another aspect covered in the case reports refers to the main challenges that the case informants perceive for reaching their future goals (see table 1-6 in the annex). As we would expect, funding is listed as the most important single main challenge perceived in more than half of the cases. Organizational challenges are also quite common. They refer in particular to issues such as maintaining attractiveness to users in the light of competitive offers (OGF, OSG, and TeraGrid), keeping current users involved or reaching out to new users (CineGrid, CLARIN, D4Science, DEISA, NVO, and OSG) and improving internal management (CineGrid, MediGrid, and SND). Technical challenges address primarily the service quality of interfaces and applications targeted to users (C3-Grid, D4Science, DEISA, OSG), interoperability issues (D4Science, EELA-2) and only in one (immature) case, CLARIN, the development of the core e-infrastructure. Further challenges are legal – in the case of MediGrid where data protection laws place a burden on the sharing of patient data – and relating to the change of scientific practice, as the establishment of "enhanced publications" aimed for by the Driver project.

Thus we see a wide range of challenges, but, as will come as little surprise to those familiar with e-infrastructure development, the sustainability of the funding and use of the infrastructures are the most frequent ones faced in one way or another by nearly all the cases that we looked at.

---

[17]   Because of the sheer size of the country and its science system the US is different in this regard – not that the same arguments don't apply there, too, but community-building and extension may still go on at national level.

6 Quantitative analysis of the survey among e-infrastructure communities[18]

6.1 Method

Quantitative data were collected using an online survey of e-infrastructure cases explored and presented in detail in the previous sections, and among a purposive sample of respondents in the area of e-infrastructure. We targeted a subset of our studied cases for design and response considerations. For example, TeraGrid was not included because they distribute surveys at a particular time that exceeded by several months the timeframe for our data collection. Or, as a purely standardization body, OGF did not fit the respondent profile in our survey. Our goal, then, was to select a purposive sample of cases on which we already had acquired deeper knowledge, to be able to explore hypotheses that pertain to other, similar e-Infrastructure efforts.

Given the variety of projects and types of involvement in these infrastructures it was impossible to define and delimit a clear research population for the survey, assess its size and employ techniques of representative sampling. Hence, the sample is non-representative and includes a wide set of respondents who are involved primarily as research users, other users, or developers in e-infrastructure projects. Respondents received a link to the questionnaire through our contact persons in the case projects included in the survey, our network of interested parties or through a mailing to the address database of BELIEF, a major European project in the area of Grid computing and e-infrastructure support.[19]

The questionnaire was structured in ten different modules (see Figure 6-1 and the full questionnaire in the Annex.

- Module A assessed the personal and professional background of the respondents (e.g. affiliation, time allocation, country of work, highest degree);

- Module B asked for their involvement with e-infrastructures, above all the main project to which they wanted to respond in more detail, when and how they became involved in it, the sponsors of their involvement, and – as a filter for the following sections – the type of involvement differentiating between research use, use for non research purposes, and development work.

- Modules C and D then asked further questions on use for and impact on research of the project that the respondent had chosen in the previous module. The modules included several questions on the community of which the respondent is a part, as well as on the personal use of the project services, and last but not least its impact on research and collaboration networks.

- Modules E and F asked very similar questions as C and D addressing professional practice or non-research use instead of research.

- In the same vein, Modules G and H asked again similar questions as C and D focusing on the development work and its impact on the respondents' development activities and collaboration networks.

- Module I consisted of four questions on the importance of national and international Grid initiatives for the respondents work.

---

[18] The principal authors of this section is Franz Barjak
[19]   We are thankful for support the BELIEF project team provided, in particular Tiziana Lombardo from Metaware, Pisa/IT, and the EELA-2 coordinators, Bernard Marechal from the Universidad Federal Rio de Janeiro and Philippe Gavillet, CERN, with the distribution of the questionnaire respectively the link to it and the development of the questionnaire.

- Module J ended the survey with a short, open-ended question on recommendations to e-infrastructure policy makers.

Figure 6-1: Structure of the eResearch2020 questionnaire
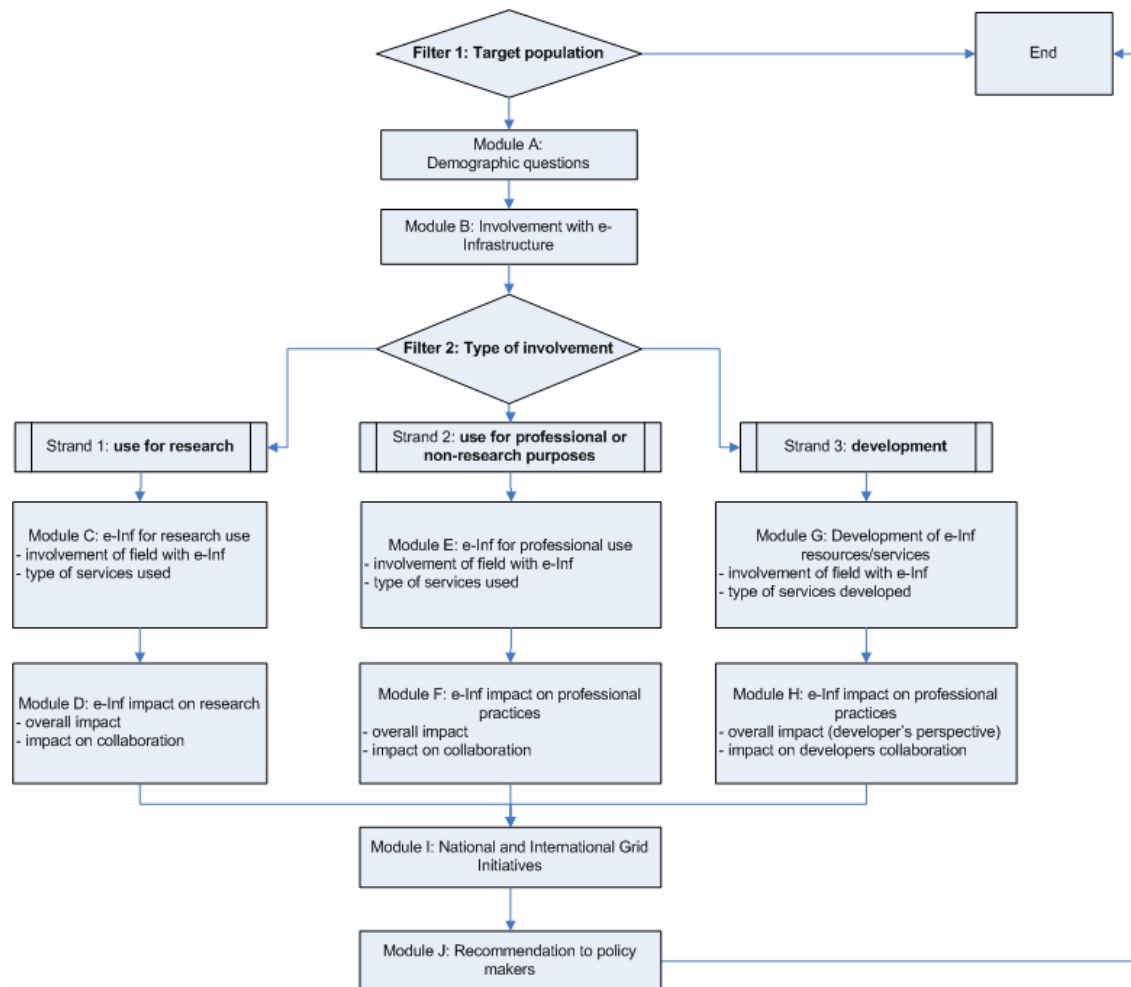
6.2   Overview of responses

The overall number of respondents ever starting the survey was 646. After eliminating mostly incomplete and non-usable responses, 407 responses were included in the following analyses.

6.2.1 Individual characteristics

We obtained responses from a broad set of countries, with the US, Italy, the UK and Germany contributing the largest shares (see Figure 6-1). In order to be able to compare response patterns by geographic location we defined country groups (coinciding partially with continents, see Table 3-1). More than 50% of the responses were contributed by respondents from the EU27 countries, an additional small share came from other European countries (Switzerland, Norway, Russian Federation, Turkey). From North America (exclusively the US) we gathered 42 responses and from Latin America – above all, Brazil, Colombia, Argentina, Venezuela and Ecuador – another 84.

Figure 6-2: Respondents by country

Country	Respondents
United States of America	42
Italy	41
United Kingdom	34
Germany	30
Brazil	20
Colombia	20
France	19
Spain	17
The Netherlands	14
India	13
Greece	11
Portugal	11
Argentina	10
Switzerland	10
Romania	9
Venezuela	8
Ecuador	6
Mexico	6
Finland	5
Belgium	4
Hungary	4
New Zealand	4
Russian Federation	4
Sweden	4
Australia	3
Austria	3
Chile	3
Cuba	3
Denmark	3
Ireland	3
Norway	3
Panama	3
Peru	3
Poland	3
Cyprus	2
Israel	2
Philippines	2
Serbia	2
Uruguay	2
Afghanistan	1
United Arab Emirates	1
Armenia	1
Bulgaria	1
Czech Republic	1
Estonia	1
Hongkong	1
Japan	1
Lithuania	1
Latvia	1
Moldova	1
Macedonia (former Yugoslav Republic)	1
Malaysia	1
Sudan	1
Singapore	1
Slovenia	1
Thailand	1
Turkey	1
Taiwan	1
Ukraine	1

		Table 6-1: Respondents by country group			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	EU	223	54.8	54.9	54.9
	Non-EU Europe	25	6.1	6.2	61.1
	North-America	42	10.3	10.3	71.4
	Latin America	84	20.6	20.7	92.1
	Asia	24	5.9	5.9	98.0
	Australia	7	1.7	1.7	99.8
	Africa	1	0.2	0.2	100.0
	Total	406	99.8	100.0	
Missing	System	1	0.2		
Total		407	100.0		

Note: Israel, Russian Federation, Turkey = Non-EU Europe.

If we distinguish responses between developed and high income countries and less developed, low and middle income countries, we count a 73% majority of responses from the former and approximately 27% from the latter type of countries (see Table 6-2).

		Table 6-2: Respondents by development status of countries of residence*			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Least Developed Countries	2	0.5	0.5	0.5
	Lower Middle Income Countries	49	12.0	12.1	12.6
	Upper Middle Income Countries	59	14.5	14.5	27.1
	Developed and high income countries	296	72.7	72.9	100.0
	Total	406	99.8	100.0	
Missing	-9	1	0.2		
Total		407	100.0		

* Development status was assessed according to the OECD DAC list (see http://www.oecd.org/dac/stats/daclist).

More than four out of five survey respondents work in academic institutions, either research universities, teaching universities or non-profit research organizations. Around 13% work for governments and international organizations and 6% for the private and commercial sector

Table 6-3: Affiliation of respondents				
		Primary institutional affiliation		
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Research university	198	48.6	51.8	51.8
	Teaching university or college	37	9.1	9.7	61.5
	Government agency	32	7.9	8.4	69.9
	Non-profit research organization	75	18.4	19.6	89.5
	International organization	18	4.4	4.7	94.2
	Commercial firm or service provider	22	5.4	5.8	100.0
	Total	382	93.9	100.0	
Missing	System	25	6.1		
Total		407	100.0		

The majority of respondents have a doctoral degree, while a smaller portion have a bachelor's or master's degree (see table 2-1 in the annex). We asked respondents what percentages of their annual working time they used for teaching, research, other professional work and administration. In previous work we have shown that this can then be aggregated to activity profiles which to some extent help to explain different uses of e-infrastructures (Barjak et al., 2008, 20). Thus, we conducted a cluster analysis of the responses on time use producing four different clusters of respondents (see Figure 6-3: Clusters of respondents according to time use pattern ("activity profiles")).[20]

- Cluster 1: 117 respondents grouped in cluster 1 are classified here as "Scholars". Their time use pattern reflects the typical pattern of scholars who have to reserve a considerable share of their time to research – the cluster average is 36.5% – and just about the same to teaching (37.5%). Administration takes up around 15.5% and professional work 10.5% of the working time in this group.

- Cluster 2: The cluster of "Researchers" (159 respondents), forms the largest cluster. Its members use more than two third of their working time for research, 11% for administration and the rest more or less equally for teaching and professional work.

- Cluster 3: The cluster "Professionals" with 86 persons consists of respondents who use more than 70% of their time for professional work, 11-13% for each research and administration and a little rest for teaching.

- Cluster 4 is again a rather small cluster of only 45 respondents, consisting of "Administrators", who use the biggest share of their time (close to 60%) for administration. Research and professional work take up less than 20% each and teaching is of little importance (6%) in this group.

It should be added that not only the cluster structure of the responses, but even the size of the clusters in the data set and the time use in each of the four clusters are very similar to the results obtained in our previous work with a different data set of early adopters of e-infrastructure in the social sciences and humanities (see Barjak et al., 2008, 20).

---

[20] The data of the 4 time use variables was processed in a Hierarchical Cluster Analysis using the squared Euclidic distance as the distance measure and the Ward algorithm to group the cases. The 4-case solution appeared to be the most appropriate solution. The initial clustering was revised in a cluster centre analysis with the cluster centres from the hierarchical analysis as the initial input values. 24 cases were re-grouped in this analysis.

Figure 6-3: Clusters of respondents according to time use pattern ("activity profiles")
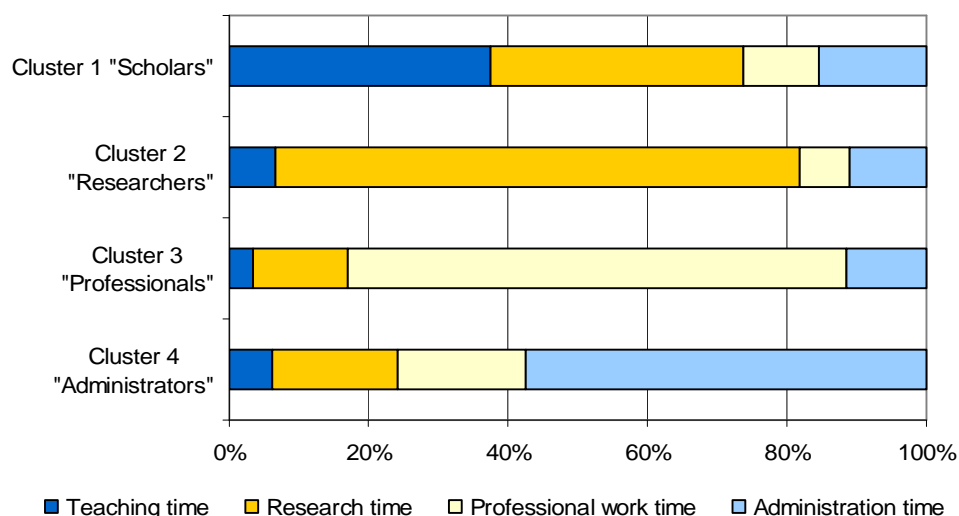
Data for this figure in table 2-2 in the annex.

Respondents show only little variation in regard to their attitude towards new technologies and the vast majority is technology-savvy: more than 75% agree to the statement "Among my peers, I am usually the first to try out new technologies" and disagree to "In general, I am hesitant to experiment with new technologies." (See table 2-3 and table 2-4 in the annex).

6.2.2 Project-level characteristics

A further set of variables includes answers to questions on a particular e-infrastructure project which the respondents could choose from a list of projects (the case projects from section 5) or enter through an open-ended question. The responses cover a broad set of projects as can be seen from the annex table 2-5. Out of these projects only four were selected by 20 or more respondents which we consider as a sufficient number to compare projects; the rest was grouped into the category "other projects" (see Table 6-4).

Table 6-4: Respondents by e-infrastructure project which they selected to report

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	DEISA	40	9.8	9.8	9.8
	EELA-2	73	17.9	17.9	27.8
	EGEE	55	13.5	13.5	41.3
	US NVO	25	6.1	6.1	47.4
	Other	214	52.6	52.6	100.0
	Total	407	100.0	100.0	

In order to include all the information provided by the respondents and avoid excluding the cases in which only few responses of a particular e-infrastructure were available, we then classified the e-infrastructures according to four criteria:

- National versus international,
- Disciplinary versus multidisciplinary,
- Computing versus data services,

- Community- versus developer-driven.

The distribution of responses on each category is shown in table 2-6 to table 2-8 in the annex.

About 70% of respondents became involved with the project on which they reported in the years 2005 or later and only a small percentage worked for/with it already before the year 2000 (seeTable 6-5). However, it has to be noted that the start dates of the projects vary considerably and therefore it may be more meaningful to see, how many years after the inception of the project the respondents joined it. Therefore, we added the start dates of the projects to the dataset which was possible for 295 out of 407 responses. The responses were then grouped into four categories reflecting the stage of the project at which the respondents became involved in it (see Table 6-6). Approximately half of the respondents were involved from the start of the project or a maximum of two years after it, whereas the other half became involved later on.

Table 6-5: Respondents by year of first involvement with the selected e-infrastructure project

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1990-1999	23	5.7	5.8	5.8
	2000-2004	90	22.1	22.6	28.3
	2005-2009	286	70.3	71.7	100.0
	Total	399	98.0	100.0	
Missing	System	8	2.0		
Total		407	100		

Table 6-6: Respondents by time of first involvement after the start of the selected e-infrastructure project

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Involvement from the start	80	19.7	27.1	27.1
	Involvement 1-2 years after project start	70	17.2	23.7	50.8
	Involvement 3-5 years after project start	104	25.6	35.3	86.1
	Involvement more than 5 years after project start	41	10.1	13.9	100.0
	Total	295	72.5	100.0	
Missing	System	112	27.5		
Total		407	100.0		

6.2.3 Field characteristics

Depending on the type of e-infrastructure involvement respondents to the survey classified themselves into a) research domains, b) fields of work, or c) area of development activities. The listing of research domains shows that all domains are represented with computer and information science being the most frequently listed domain and few responses from the social sciences and humanities (seeTable 6-7). Fields of work consist of academic and non-

academic support services. Developers mostly develop for supercomputing and distributed computing, academic and IT support and applications; other areas are less well represented.

	Frequency	in % of total
Table 6-7: Respondents by a) research domains, b) fields of work, or c) area of development activities		
a) Research domains		
Astronomy or Astrophysics	24	6.2
Biological Sciences and Medicine	32	8.2
Chemical and Material Sciences	18	4.6
Computer and Information Sciences	36	9.3
Engineering and Technology	20	5.2
Earth and Other Natural Sciences	18	4.6
Physical Sciences	21	5.4
Social Sciences and Humanities	13	3.4
b) Fields of work		
Academic support services	12	3.1
Non-academic support services	17	4.4
c) Area of development activities		
Academic and IT support services	37	9.5
Supercomputing and distributed computing	66	17.0
Networking	16	4.1
Application Development	35	9.0
Other	23	5.9
Total	388	100

In addition, the questionnaire contained a number of questions on characteristics of the field: the importance and types of collaborations, division of labour, intensity of competition, and maturity of and pace of change in the field. Field characteristics have been found to explain varying patterns of e-infrastructure adoption and to constitute major inhibitors to the penetration of science with new ICT (Fry, 2004, 2006; Kling & McKim, 2000; Talja, Vakkari, Fry, & Wouters, 2007; Wouters & Beaulieu, 2006; Wouters et al., 2008). (Whitley, 2000) suggests a framework for relevant characteristics of fields stressing the roles of mutual dependence and task uncertainty. We analysed the responses to seven statements on field characteristics[21] using cluster analysis and obtained three interpretable field clusters (see table 2-10 in the annex):

- Established low collaboration fields (104 cases): These are established fields in which collaboration is still the dominant mode of work, but less so than in other fields; respondents agreed more often than in the other clusters to the statement that work is typically done by individuals. Competition and the change of research problems, paradigms, approaches or methods are denoted as low.

- Novel dynamic collaborative fields (63 cases): These fields stick out by the fact that they are described as novel and with a comparatively fast pace of change of research problems, paradigms etc. In addition, collaboration is deemed as essential for achieving progress in these fields and work is more often done in large-scale collaborations of more than ten people.

- Dynamic competitive fields (73 cases): The third cluster of fields is characterized by the high intensity of competition in combination with a fast pace of change.

---

[21] See items 1, 2, 3, 4, 5, 7 and 10 in question 32 in the annexed questionnaire.

The importance of collaboration is average, but more in small than in large groups of collaborators.

Certain patterns apply to the description of fields (see Table 6-8): Biological sciences and medicine, chemical and material sciences, computer sciences, social sciences and humanities, non-academic support services and application development are common and overrepresented among the "established low collaboration fields". Earth and other natural sciences, academic support services, supercomputing and distributed computing, and networking were often classified as "novel dynamic collaborative". Among the "dynamic competitive fields" astronomy and astrophysics, engineering and technology, other development areas and academic and IT support services are particularly frequent. Physical sciences, engineering and technology, and academic and IT support developers are quite well distributed across all three clusters.

Table 6-8: Fields by field characteristics (frequency of a field in %)

	Field characteristics			
	Established low collaboration	Novel dynamic collaborative	Dynamic competitive	Total
a) Research domains				
Astronomy or Astrophysics	37.5%	18.8%	43.8%	100.0%
Biological Sciences and Medicine	55.6%	27.8%	16.7%	100.0%
Chemical and Material Sciences	54.5%	9.1%	36.4%	100.0%
Computer and Information Sciences	58.8%	17.6%	23.5%	100.0%
Engineering and Technology	33.3%	25.0%	41.7%	100.0%
Earth and Other Natural Sciences	50.0%	37.5%	12.5%	100.0%
Physical Sciences	38.9%	27.8%	33.3%	100.0%
Social Sciences and Humanities	55.6%	11.1%	33.3%	100.0%
b) Other user domains				
Academic support services	0.0%	66.7%	33.3%	100.0%
Non-academic support services	87.5%	0.0%	12.5%	100.0%
c) Development areas				
Academic and IT support services	39.1%	21.7%	39.1%	100.0%
Supercomputing and distributed computing	22.2%	44.4%	33.3%	100.0%
Networking	27.3%	45.5%	27.3%	100.0%
Application Development	68.0%	12.0%	20.0%	100.0%
Other development areas	35.7%	21.4%	42.9%	100.0%
Total	43.3%	26.1%	30.7%	100.0%

6.3   Characteristics of the virtual research community involved in an e-infrastructure

In a first set of variables the respondents were asked about the extent of involvement in their field in the same e-infrastructure. We see this group of peers from the same field who use the same e-infrastructure (or participate in the development of the same e-infrastructure in the case of developers) as a good approximation of a virtual community that has formed around an e-infrastructure. The questions assessed the number of other individuals involved, their geographic distribution and last but not least their affiliation by sector (academic versus non-academic).

6.3.1 Size of the virtual research community

Likely because estimating general involvement in the field requires comprehensive outlook, only about 75% of the respondents had an idea of how many other individuals from the same field participated in the selected e-infrastructure (see Table 6-9). 36% of those who answered the question (N=388) pointed to a small community and in another 28% to a medium-sized community of 21-100 people. Just about 16% work with larger e-infrastructure-based communities of more than a 100 participants.

Table 6-9: Number of other individuals working in the field that are using/participating in the e-Infrastructure

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	None	9	2.2	2.3	2.3
	1-5	58	14.3	14.9	17.3
	6-10	71	17.4	18.3	35.6
	21-100	109	26.8	28.1	63.7
	101-500	30	7.4	7.7	71.4
	More than 500	33	8.1	8.5	79.9
	Don't know	78	19.2	20.1	100.0
	Total	388	95.3	100.0	
Missing	System	19	4.7		
Total		407	100.0		

Next, the data permits us to assess the correlates of community size. We find that the e-infrastructure that respondents use is one of the most remarkable correlates (see Table 6-10): participants to DEISA and EELA-2 point most often to a small number of peers working with the e-infrastructure, whereas those participating in EGEE and US NVO point to mid-size and large communities. Classifying the e-infrastructures into computing respectively data infrastructures (see annex Table 2-13), we also obtain an interesting pattern. Respondents involved in computing infrastructures point out more often than those involved in data infrastructures, that only few peers from the same field work with the e-infrastructure (see

Figure 6-4: Size of the community from the same field using/participating in the e-Infrastructure by type of e-infrastructure (in %)

Figure 6-4). This suggests that data infrastructures tend to involve a larger number of people in the same manner and with similar needs, whereas computing infrastructures rather serve small groups in different ways. A similar pattern appears if we differentiate between community- and developer-driven e-infrastructures (see again

Figure 6-4). Those that are community-driven are considerably more often backed by larger communities.

Table 6-10: Number of other individuals from the same field using/participating in the e-Infrastructure by e-infrastructure (in %)

	Selected e-infrastructure					
	DEISA	EELA-2	EGEE	US NVO	Other	Total
None	10.8%	1.4%	0.0%	0.0%	2.0%	2.3%
1-5	18.9%	28.2%	11.8%	4.0%	11.8%	14.9%
6-10	32.4%	22.5%	9.8%	12.0%	17.2%	18.3%
21-100	13.5%	23.9%	25.5%	36.0%	31.9%	28.1%
101-500	0.0%	2.8%	11.8%	8.0%	9.8%	7.7%
More than 500	0.0%	.0%	13.7%	12.0%	11.3%	8.5%
Don't know	24.3%	21.1%	27.5%	28.0%	16.2%	20.1%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Figure 6-4: Size of the community from the same field using/participating in the e-Infrastructure by type of e-infrastructure (in %)

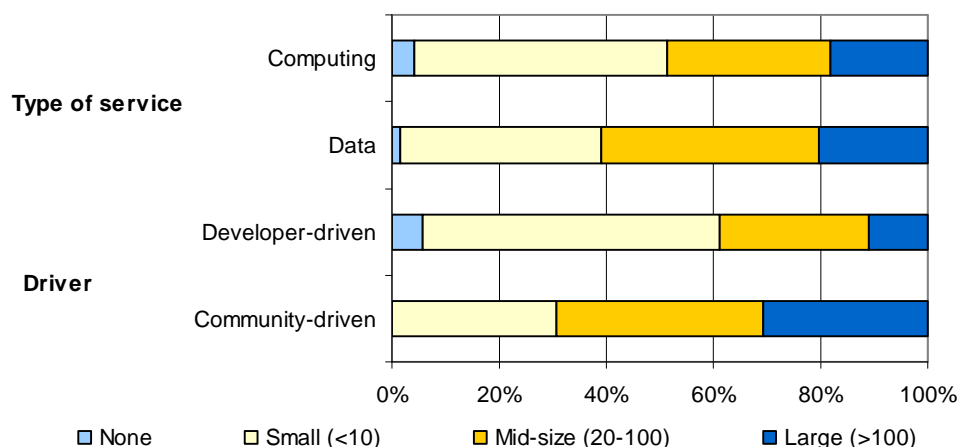Data for this figure in table 2-12 in the annex.

These results are partially due to differences between the fields involved in the projects. We see that in particular astronomers and social scientists point more often to large numbers of other e-infrastructure users, whereas the other fields indicate smaller numbers of users (see table 2-11 in the annex). The field characteristics also relate to the number of individuals involved in the e-infrastructure: in established low collaboration fields there are less other people from the same field involved, in novel dynamic collaborative and dynamic competitive fields there are more other people involved (see Table 6-11).

Table 6-11: Number of other individuals from the same field using/participating in the e-Infrastructure by fields of professional work and development areas (in %)

	Field characteristics			
	Established low collaboration	Novel dynamic collaborative	Dynamic competitive	Total
None	2.9%	1.6%	0.0%	1.7%
1-5	22.3%	9.7%	15.3%	16.9%
6-10	15.5%	16.1%	19.4%	16.9%
21-100	24.3%	37.1%	36.1%	31.2%
101-500	6.8%	12.9%	9.7%	9.3%
More than 500	6.8%	9.7%	6.9%	7.6%
Don't know.	21.4%	12.9%	12.5%	16.5%
Total	100.0%	100.0%	100.0%	100.0%

These assessments of the size of the communities of users and developers from the same field participating in an e-infrastructure are interesting, but they have to be interpreted with caution: there is obviously a learning effect taking place: The longer respondents have worked with an infrastructure, the more often they can answer the question and the higher the number of peers of which they have become aware (see table 2-12 in the annex).

6.3.2 Geographic distribution of the virtual research community

The next questions in the questionnaire modules for research users, other users and developers asked for the geographical distribution of the communities participating in the specified e-infrastructure. Local communities, i.e. those confined to a single region within a country, are not common. Approximately one third of the communities are national and two third are international (see).

Table 6-12: Geographic distribution of other individuals in the field that are using/participating in the e-Infrastructure

Geographic distribution of peers		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	In a single region	39	9.6	10.7	10.7
	In multiple regions within a country	78	19.2	21.3	32.0
	Across multiple countries within a continent	115	28.3	31.4	63.4
	Across continents	134	32.9	36.6	100.0
	Total	366	89.9	100.0	
Missing	System	41	10.1		
Total		407	100.0		

An important correlate of the geographic distribution of the peers from the same field is again the e-infrastructure in question (see Table 6-13): e-infrastructures either involve national communities with strong international links (NVO), international communities with a strong focus on Europe (DEISA), or extend on more than one continent (EELA-2, EGEE). It will come as little surprise that national infrastructures involve mostly peers from the same country, whereas international infrastructures cater to international communities (see table 2-14 in the annex).
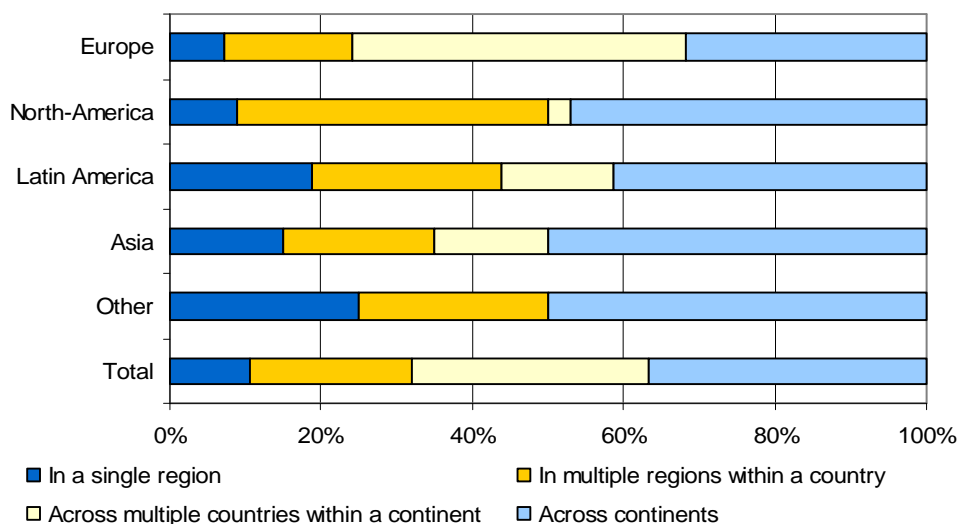
Table 6-13: Geographic distribution of other individuals in the field that are using/participating in the e-Infrastructure by e-infrastructure (in %)

Geographic distribution of peers	Selected e-Infrastructure					
	DEISA	EELA-2	EGEE	US NVO	Other	Total
In a single region	3.2%	15.9%	6.1%	4.5%	11.8%	10.7%
In multiple regions within a country	12.9%	14.5%	8.2%	40.9%	26.2%	21.3%
Across multiple countries within a continent	74.2%	13.0%	38.8%	4.5%	32.3%	31.4%
Across continents	9.7%	56.5%	46.9%	50.0%	29.7%	36.6%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

The geographic distribution again relates to the disciplines which e-infrastructures serve (see table 2-15 in the annex). Astronomers, physicists, chemists and material scientists and life scientists are more distributed at international level, whereas engineering communities (including computer science), earth scientists and social scientists and humanists are more bounded to national level (Note: Case numbers for most fields are rather small).

Another correlate of the geographic distribution of other community members involved in the selected e-infrastructure is the continent of the respondent (see Figure 6-5). European respondents have the smallest shares of national communities and the largest shares of communities limited to one continent (Europe, of course). North-American and other (mostly Australian) respondents point to either national or global peer communities. Collaboration with other countries on the American/Australian continent is in both cases negligible. Respondents from developing countries[22] state more often that their communities are national or even bounded to a single region than respondents from developed countries (see Figure 6-6).

Figure 6-5: Geographic distribution of other individuals in the field that are using/participating in the e-Infrastructure by continent of respondent (in %)

Note: Case numbers Europe: 224; North-America: 34; Latin-America: 80; Asia: 20; Other: 8; Total: 366

---

[22] It should be noted that whenever we mention "respondents from developing countries" this represents the sample of respondents. It includes a large share of Latin American respondents, not least because of the inclusion of EELA into our selection of projects covered. Several regions of the world are not or not adequately represented, including especially Africa.
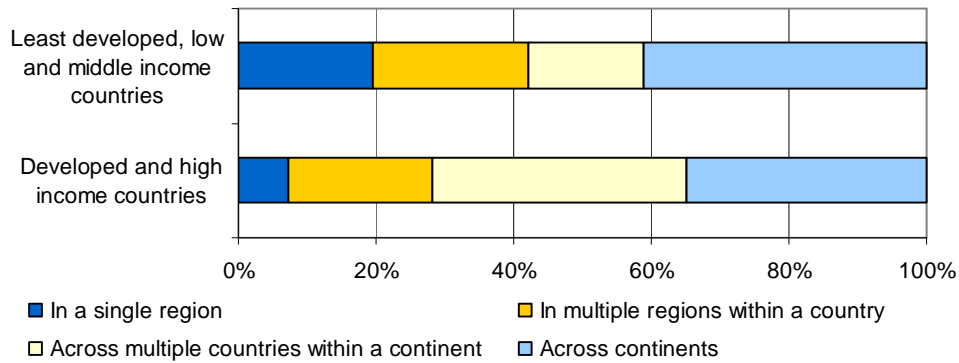
Figure 6-6: Geographic distribution of other individuals in the field that are using/participating in the e-Infrastructure by development status of the of respondent's country (in %)

Legend:
- ■ In a single region
- ■ In multiple regions within a country
- □ Across multiple countries within a continent
- ■ Across continents

6.3.3 Affiliation of the virtual community members

Last but not least respondents were asked about the institutional affiliations of their peers, differentiating between academic and non-academic organizations. Approximately 41% of the respondents stated that their community consists exclusively of academics and another 54% point out that their peers are both academics and non-academics (see Figure 6-7). Purely non-academic communities are infrequently reported. As we would expect, there is some bias resulting from the affiliation of the respondents themselves (see Figure 6-8): Those who are affiliated to an academic organization more often see their peer affiliations also as academic. Respondents involved in non-academic organizations, governments, international organizations or the private sector, point more often to peers with non-academic affiliations.

Figure 6-7: Affiliation of other individuals in the field that are using/participating in the e-Infrastructure (in %)
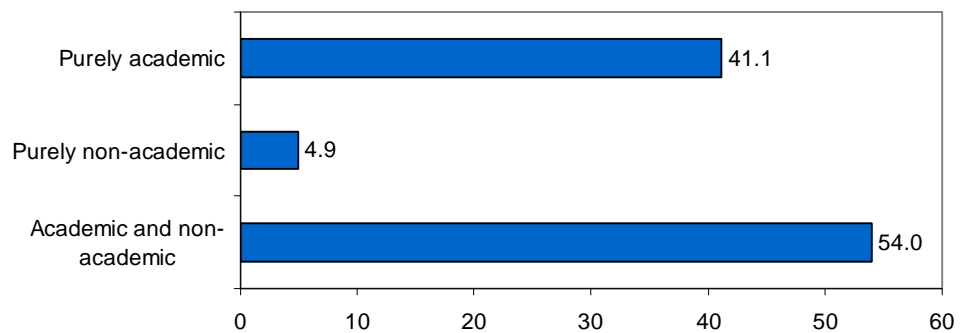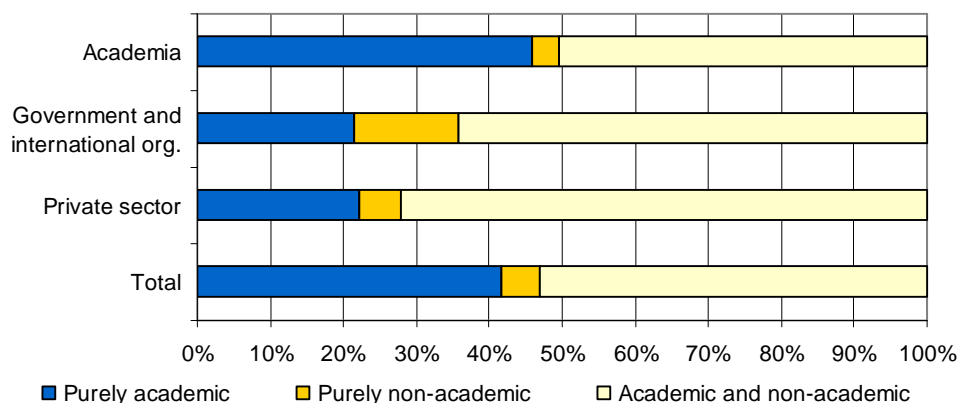
Figure 6-8: Affiliation of other individuals in the field that are using/participating in the e-Infrastructure by affiliation of the respondent (in %)

We find notable differences between the e-infrastructures also for this issue of institutional affiliation of community members (see Table 6-14). DEISA is strongly characterized by academia and US NVO by mixed communities; EGEE, too, but not to the same extent as US NVO. Non-academic communities were mostly mentioned by respondents involved in the other e-infrastructures. Distinguishing the affiliation of participants in an e-infrastructure by the type of the infrastructure we get a few striking differences (see Table 7-16). Disciplinary e-infrastructures serve more often mixed than purely academic communities whereas in multidisciplinary e-infrastructures both community types are equally important. Infrastructures offering computing services cater more to academic communities, non-academic communities are not that important. Data infrastructures, on the other hand, are more often dealing with non-academic communities.

Table 6-14: Affiliation of other individuals in the field that are using/participating in the e-Infrastructure by e-infrastructure (in %)

Affiliation of other individuals in the field	Selected e-Infrastructure					
	DEISA	EELA-2	EGEE	US NVO	Other	Total
Purely academic	79.4%	44.3%	31.3%	18.2%	38.2%	41.1%
Purely non-academic	0.0%	0.0%	2.1%	4.5%	8.4%	4.9%
Academic and non-academic	20.6%	55.7%	66.7%	77.3%	53.4%	54.0%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 6-15: Affiliation of other individuals in the field that are using/participating in the e-Infrastructure by type of e-infrastructure (in %)

Affiliation of other individuals in the field	Geographic scope		Disciplinary scope		Type of service		Driver	
	National	Inter-national	Disci-plinary	Multi-disciplinary	Computing	Data	Devel-oper	Com-munity
Purely academic	35.8%	40.9%	22.5%	47.2%	44.6%	28.6%	46.1%	30.0%
Purely non-academic	5.7%	2.2%	8.5%	1.0%	1.0%	7.1%	2.0%	5.0%
Academic and non-academic	58.5%	57.0%	69.0%	51.8%	54.4%	64.3%	52.0%	65.0%

Last but not least, responses on the affiliation of the other people involved in an e-infrastructure also correlate with respondents' continent (see Figure 6-9). European, Latin

American and other (mostly Australian) respondents are more often aware of colleagues with academic affiliation. North-American and Asian respondents perceive a larger importance of non-academics, from the governmental or private sector, in the communities participating in an e-infrastructure.
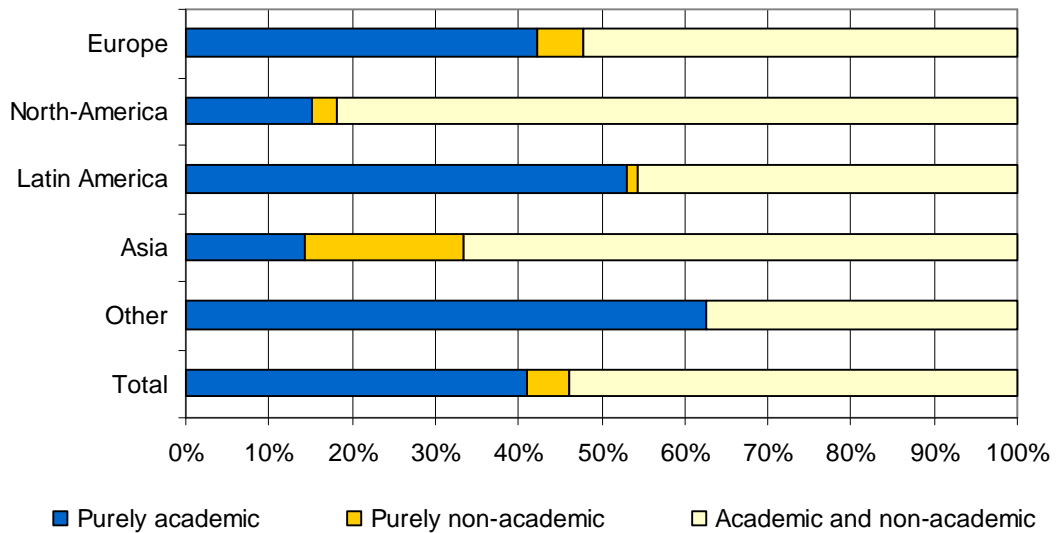
Figure 6-9: Affiliation of other individuals in the field that are using/participating in the e-Infrastructure by continent of the respondent (in %)

6.4   Involvement of respondents in e-infrastructures

6.4.1 Ways of involvement in e-infrastructures

Asked for their capacities or roles in the e-infrastructure most respondents reported to be users or researchers of it (see Figure 6-10). Equal shares classify themselves as PIs or Co-PIs and software developers; fewer are members of governing bodies or project managers, a proportion which is expected given that less people tend to serve in these functions. Respondents could then choose in the questionnaire whether they wanted to answer the remaining questions on the involvement with the selected e-infrastructure and the impact of this as a) research users, b) other users or c) developers. Large groups answered the questions as research users or developers and only a small share classified themselves as other users (seeTable 6-16).

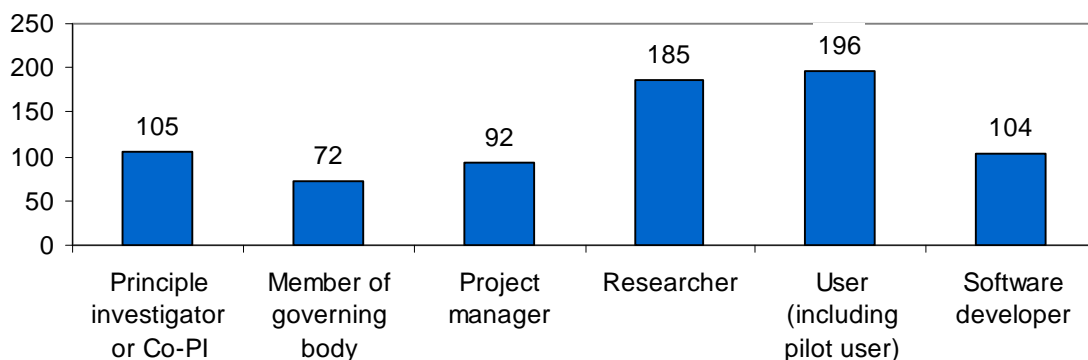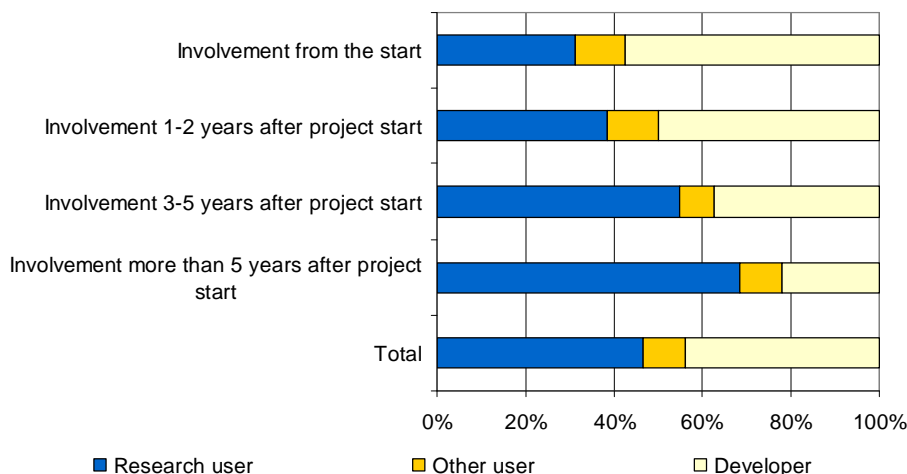**Figure 6-10: Respondents according to role in the selected e-infrastructure**

Table 6-16: Respondents by function of involvement in the selected e-infrastructure project

	Frequency	Percent	Valid Percent	Cumulative Percent
Research user	187	45.9	45.9	45.9
Other user	37	9.1	9.1	55.0
Developer	183	45.0	45.0	100.0
Total	407	100.0	100.0	

There are no clear patterns if we differentiate the functions of the respondents by countries (see table 2-16 in the annex) or e-infrastructure projects – except that the DEISA respondents to the survey are nearly exclusively researchers (see table 2-17 in the annex). Moreover, it is not surprising to see that research users are more often latecomers to the projects and their share is larger among those who became involved in a project at a rather late point in time (see Figure 6-11). In contrast, developers have been involved from early on and rather few became involved after the projects had been running for three or more years. This reinforces an impression that we already obtained in our case analyses: e-infrastructures follow traditional models of technology innovation and involve most research users at late stages of development. This finding may have some negative effects in regard to usability and to what extent users' problems and needs are actually addressed in the technology developed.

One surprising result is that, if other users are involved, they also tend to be involved quite early; however their numerical presence in e-infrastructures is rather small.

Figure 6-11: Respondents by function of involvement in the selected e-infrastructure project and years after project start at which this involvement began (in %)

Figure 6-11: Respondents by function of involvement in the selected e-infrastructure project and years after project start at which this involvement began (in %)

6.4.2 Funding of involvement in e-infrastructures

Looking at the primary sponsors of the involvement in the e-infrastructure projects, we find that for 160 out of 360 respondents (44%) their national (research) funding agencies were the primary source (see Table 6-17). More or less equal shares were funded by their own organizations and the EU or other international funding bodies. A negligible share of the respondents had their involvement funded from private funding sources.

Table 6-17: Respondents by primary sponsor of the activities with the selected e-infrastructure

Primary sponsor		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Governmental funding agency (national)	160	39.3	44.4	44.4
	International governmental funding agency (e.g. EU)	100	24.6	27.8	72.2
	Private funding agency	8	2.0	2.2	74.4
	Own institution	92	22.6	25.6	100.0
	Total	360	88.5	100.0	
Missing	System	47	11.5		
Total		407	100.0		

Depending on the continent of the respondent we see some variation to this funding structure (see Figure 6-12). There are also variations by the development level of the respondents' countries: while national and private funding are of about the same levels, funding from the own organization substitutes the lower level of international funding in less developed countries (see Figure 6-13).

Figure 6-12: Respondents by primary sponsor of the activities with the selected e-infrastructure project and continent (in %)

- ■ Governmental funding agency (national)
- □ Private funding agency
- ■ International governmental funding agency (e.g. EU)
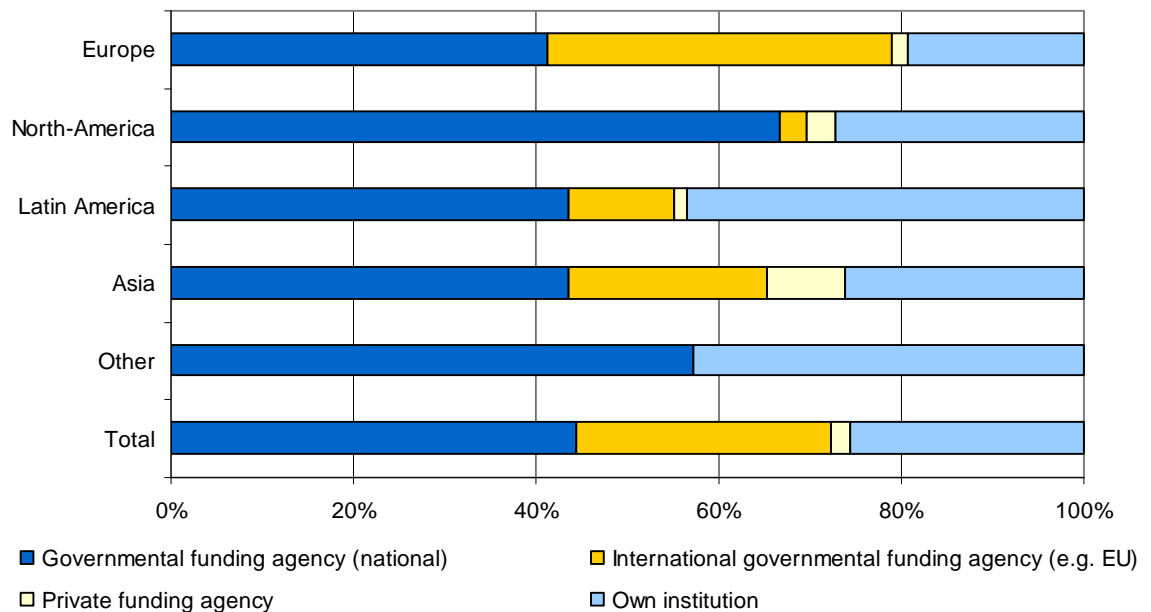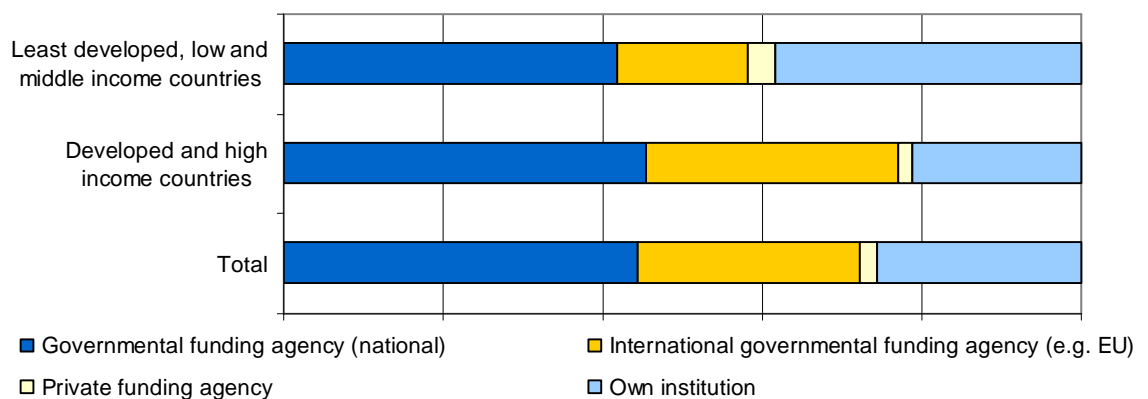- ■ Own institution

Figure 6-13: Respondents by primary sponsor of the activities with the selected e-infrastructure project and development level of their country (in %)

- ■ Governmental funding agency (national)
- □ Private funding agency
- ■ International governmental funding agency (e.g. EU)
- ■ Own institution

Notable variations appear also depending on the e-infrastructure the respondents had selected (Table 6-18): DEISA and NVO participants are most often funded by their national governmental funding agencies, EELA-2 participants by their own institutions, EGEE participants by international governmental funding agencies (e.g. EU). It is not surprising, that respondents involved in national e-infrastructures rely to 69% on national funding, whereas for those working with international infrastructures the share of national and international funders as primary sponsors is about the same with 34-37% (see table 2-18 in the annex). The funding structures of the users' involvement with the e-infrastructure – not the e-infrastructures themselves – also vary by service provided: Those involved in computing infrastructures depend to nearly half on national government funding. International funding is more or less the same for both types, and the users of data infrastructures rely also to about the same degree on institutional and national governmental funding (see Figure 6-14).

Table 6-18: Respondents by primary sponsor of the activities with the selected e-infrastructure project and project

Primary sponsor	Selected e-Infrastructure					
	DEISA	EELA-2	EGEE	US NVO	Other	Total
Governmental funding agency (national)	56.8%	25.4%	38.0%	63.6%	47.4%	44.4%
International governmental funding agency (e.g. EU)	13.5%	27.1%	54.0%	4.5%	26.6%	27.8%
Private funding agency	5.4%	1.7%	0.0%	4.5%	2.1%	2.2%
Own institution	24.3%	45.8%	8.0%	27.3%	24.0%	25.6%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Figure 6-14: Respondents by primary sponsor of the activities with the selected e-infrastructure and type of service of the selected e-infrastructure (in %)
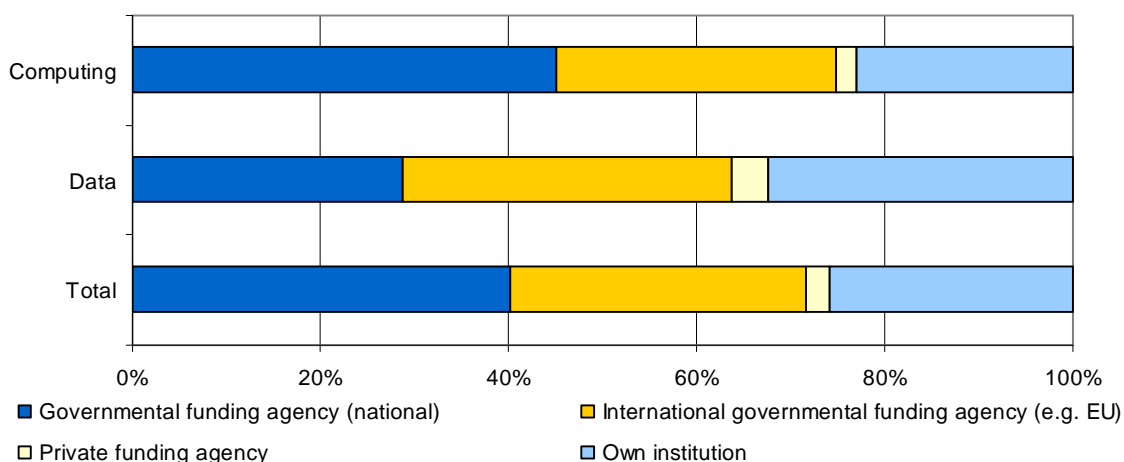
Another interesting result is obtained if the primary sponsor of the activities with the selected e-infrastructure project and the number of years after project start at which this involvement began are cross-tabulated (see Figure 6-15). Those who have been involved in an infrastructure early on, receive funding more often from the EU or other international funding agencies, whereas those who became involved later rely on funding more often from national agencies and their own institutions (see table 6 18 in the annex).

Figure 6-15: Respondents by primary sponsor of the activities with the selected e-infrastructure project and years after project start at which this involvement began (in %)
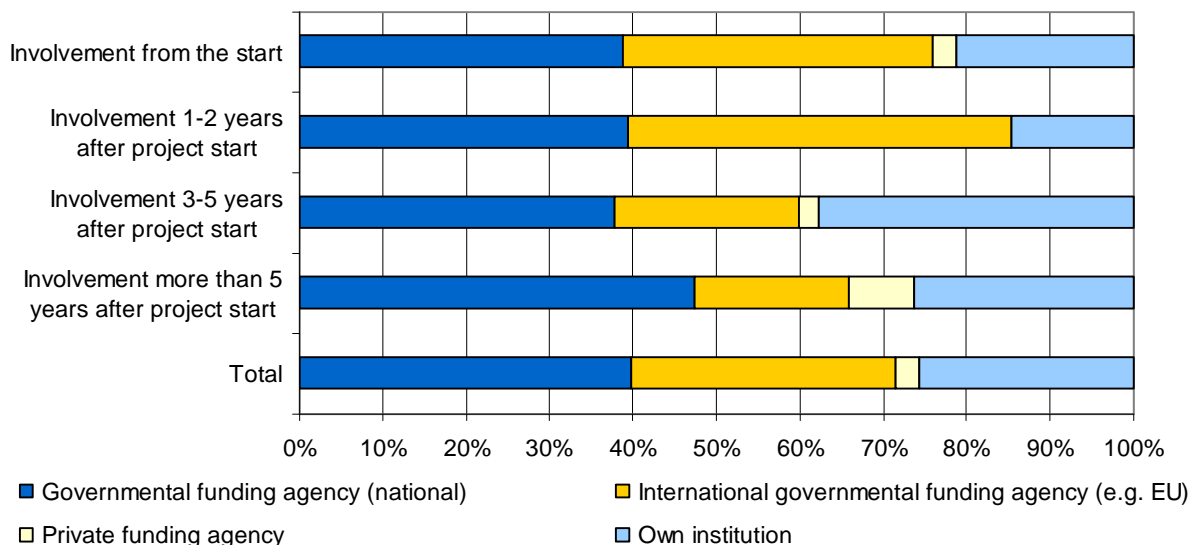
Between research fields also appear some differences in regard to the funding of e-infrastructure participation (see Table 6-19): National funding is the most important source in many fields, except for life sciences and social sciences and humanities, where respondents rely on funds from their own organizations and international organizations to about the same extent. Developers also answered this question on the primary sponsors of their development activities. We again find a few differences (see Table 6-20): Among our respondents development in the area of networking is very much funded by national sponsors. Application development, on the other hand, is more often funded from international and institutional budgets. Finally, differences also appear when we classify the fields according to their characteristics in regard to collaboration, competition, dynamics and the like: In particular users in fields classified as novel, dynamic and collaborative obtain to large shares international funding. Those from established fields rely more on their own organizations.

Table 6-19: Respondents by primary sponsor of the activities with the selected e-infrastructure project and research field (in %)

Primary sponsor	Astronomy or Astro-physics	Biological Sciences and Medicine	Chemical and Material Sciences	Computer and Information Sciences	Engineering and Technology	Earth and Other Natural Sciences	Physical Sciences	Social Sciences and Humanities	Total
Governmental funding agency (national)	61.9%	29.0%	66.7%	48.3%	40.0%	38.5%	68.4%	23.1%	47.2%
International governmental funding agency (e.g. EU)	0.0%	32.3%	11.1%	27.6%	13.3%	23.1%	21.1%	30.8%	20.8%
Private funding agency	4.8%	3.2%	0.0%	0.0%	6.7%	0.0%	0.0%	7.7%	2.5%
Own institution	33.3%	35.5%	22.2%	24.1%	40.0%	38.5%	10.5%	38.5%	29.6%
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%

Table 6-20: Respondents by primary sponsor of the activities with the selected e-infrastructure project and development area (in %)

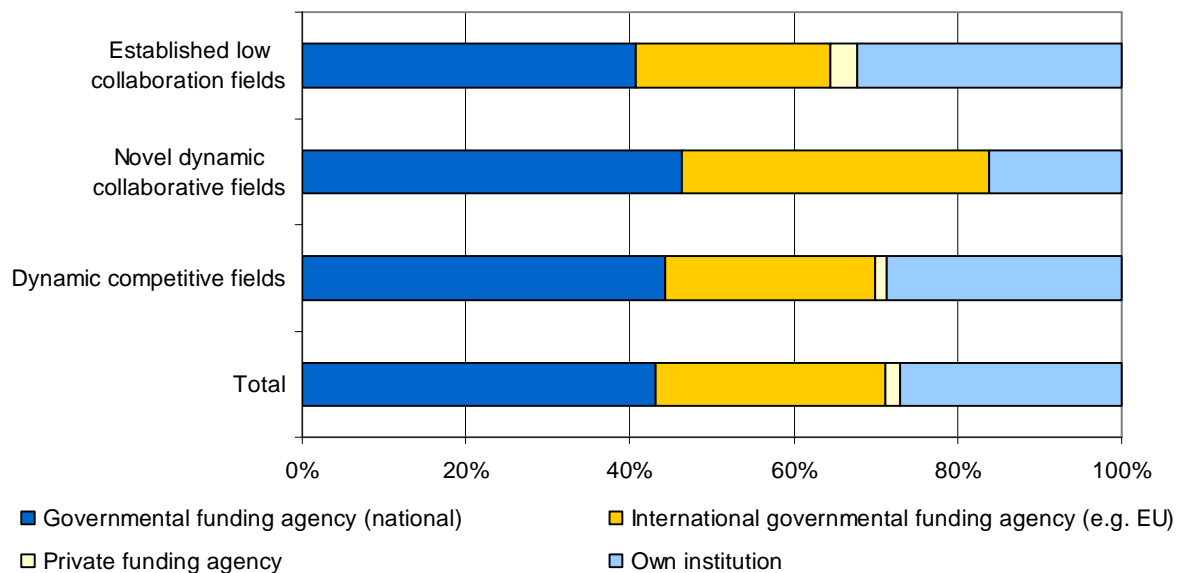| Primary sponsor	Academic and IT support services	Supercom-puting and distributed computing	Networking	Application Develop-ment	Other	Total
Governmental funding agency (national)	41.2%	47.6%	62.5%	30.3%	45.5%	44.0%
International governmental funding agency (e.g. EU)	32.4%	38.1%	25.0%	45.5%	22.7%	35.1%
Private funding agency	5.9%	0.0%	0.0%	0.0%	0.0%	1.2%
Own institution	20.6%	14.3%	12.5%	24.2%	31.8%	19.6%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Figure 6-16: Respondents by primary sponsor of the activities with the selected e-infrastructure project and type of field (in %)

■ Governmental funding agency (national)   ■ International governmental funding agency (e.g. EU)
□ Private funding agency   □ Own institution

6.4.3 Use or development of services and resources

Each e-infrastructure offers a specific set of services and resources that may also correlate to the impact it has on a served community. In the questionnaire we therefore asked the respondents, whether they used or developed 14 services or resources within the previous six months and – if yes – how often this was the case (see for instance questions 21-23 in the annexed questionnaire). More than half of our respondents were involved with grid computing resources (see Figure 6-17).

Figure 6-17: Respondents by services and resources used or developed (in %)

| Service/Resource	%
Grid computing	52.8%
Data management tools	36.7%
Data collections	30.7%
Data analysis tools	29.9%
My own applications ported on the infrastructure	29.1%
Online storage	27.8%
Collaboration tools	27.5%
Simulation	22.6%
Supercomputing	21.6%
Individual support/advice	20.2%
Online digital materials for research	20.1%
Visualization	17.3%
Remote access to research instruments	16.4%
Other	11.3%

At least 30% of the respondents dealt with data related services and resources (data management tools, data analysis tools, data collections). Several other tools and resources were called upon by 20-30% of the respondents: Above all their own application ported on the e-infrastructure, next online storage and collaboration tools, simulation applications and supercomputing resources.

We also assessed how frequently respondents worked with any of the services and resources in the previous six months; we distinguish between irregular use (just once, quarterly, or monthly), regular use (weekly) and intensive use (daily). However, the pattern that appears is not very clear (see Figure 6-18).

Figure 6-18: Services and resources used or developed by frequency of use (in %)
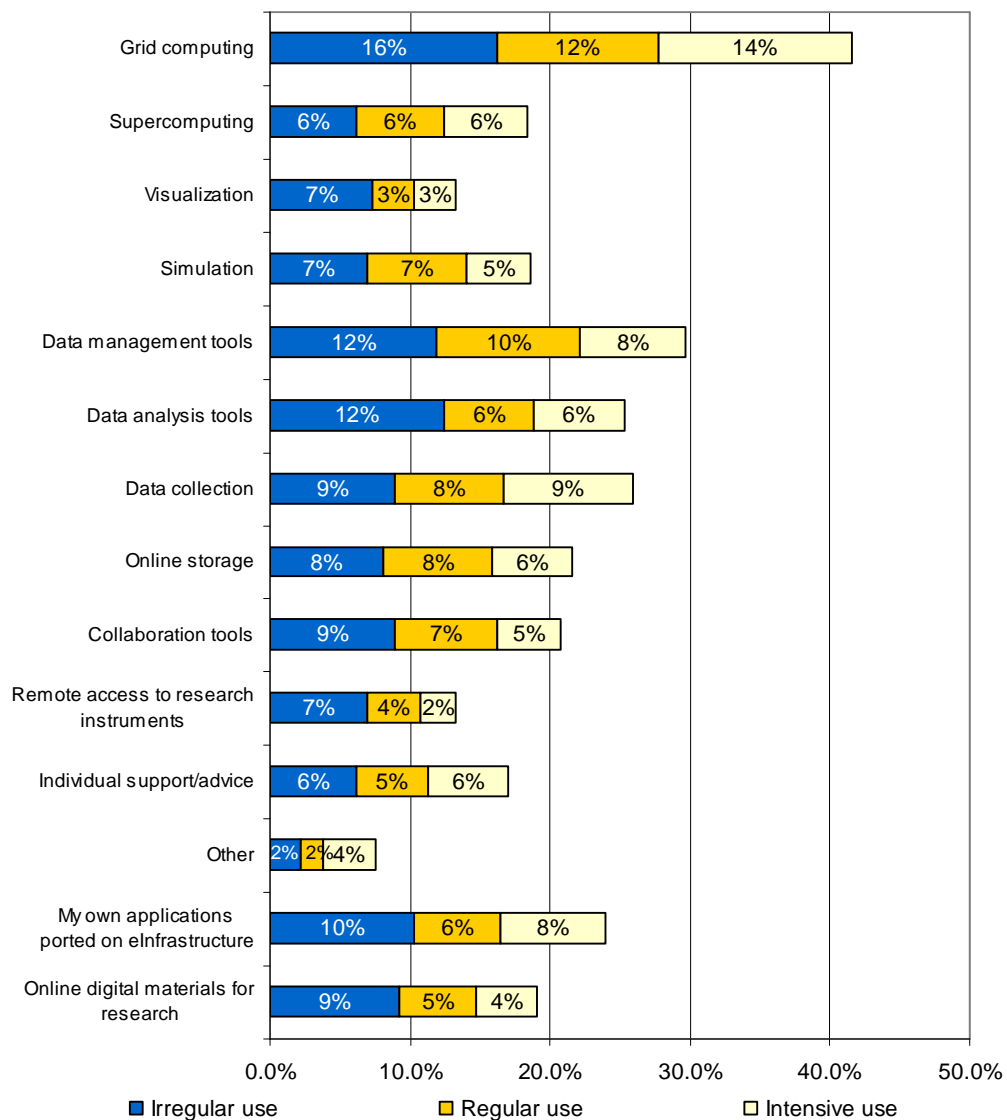
Respondents' affiliation correlates with the resources with which they are involved (see Table 6-21). Respondents from governments and international organizations less often rely on computing and simulation resources, and more often on data-related tools and online storage. Private sector responses were not analyzed due to the low case numbers. Activity profiles also correlate to some extent with the services and resources and confirm this finding (see table 2-19 in the annex). The most interesting group in this case are professionals, i.e. respondents who spend a large share of their time on professional (and not academic) work. This group is a lot less involved with computing resources – in particular supercomputing – and the services and resources which support analytical tasks (simulations, remote access to research instruments, or own applications ported on the e-infrastructure). However, they particularly often refer to data management tools, online digital material and data collections.

Table 6-21: Respondents by services and resources used or developed and affiliation (in %)

Services and resources used or developed	Academia	Government and international org.	Private sector	Total
Grid computing	50.0%	36.0%	50.0%	48.2%
Supercomputing	21.9%	16.0%	4.5%	20.2%
Visualization	15.8%	18.0%	13.6%	16.0%
Simulation	22.6%	14.0%	18.2%	21.2%
Data management tools	32.6%	44.0%	22.7%	33.5%
Data analysis tools	27.1%	38.0%	9.1%	27.5%
Data collections	26.8%	40.0%	13.6%	27.7%
Online storage	24.5%	38.0%	13.6%	25.7%
Collaboration tools	24.8%	30.0%	9.1%	24.6%
Remote access to research instruments	16.1%	16.0%	0.0%	15.2%
Individual support/advice	18.1%	26.0%	4.5%	18.3%
Other	8.1%	14.0%	22.7%	9.7%
My own applications ported on the e-infrastructure	28.4%	32.0%	13.6%	28.0%
Online digital materials for research	18.6%	33.3%	33.3%	20.3%

Note: The number of responses from the private sector is only 20.

As we would expect, we also find quite different profiles of service and resource usage, depending on the users' field of research (see table 2-20 in the annex, note however the small case numbers). For instance astronomers and social scientists use more than most others data-related tools, including visualization applications. Grid computing is more common among computer scientists and biologists and supercomputing among chemists and material scientists. Physicists and earth scientists use both to similar extent. Collaboration tools are particularly commonly used among social scientists and computer scientists. Categorizing the fields according to how respondents assess competition, collaboration, maturity and pace of change (see 0 above) we also obtain some interesting patterns (see Table 6-22). Respondents from fields characterized as "novel dynamic collaborative", i.e. fields which are not yet established, experience a fast pace of change of research problems, paradigms and approaches and rely to large extent on collaboration, use distributed computing and collaboration tools more than respondents from the other two categories of fields. Data-related tools and online storage are most often used in established low collaboration fields with low levels of competition and dynamics. In dynamic competitive fields with high intensity of competition, fast pace of change, and collaboration mostly in small groups of collaborators, several services are less common –exceptions are supercomputing, simulation applications, and the respondents' own applications ported on the e-infrastructure. As others before us have already put forth (Fry, 2004, 2006; Kling & McKim, 2000; Talja et al., 2007; Wouters & Beaulieu, 2006; Wouters et al., 2008), this finding suggests that field characteristics influence the types of services and resources needed and used. In other word, a one size fits all approach is destined to fail. In highly competitive fields in particular, Grid proponents are well advised to enable scientists to reliably use their own applications on the Grid.

Table 6-22: Respondents by services and resources used or developed and field characteristics (in %)

Services and resources used or developed	Field characteristics			
	Established low collaboration	Novel dynamic collaborative	Dynamic competitive	Total
Grid computing	49.0%	63.5%	47.9%	52.5%
Supercomputing	23.1%	23.8%	30.1%	25.4%
Visualization	21.2%	19.0%	19.2%	20.0%
Simulation	20.2%	23.8%	30.1%	24.2%
Data management tools	42.3%	34.9%	37.0%	38.8%
Data analysis tools	33.7%	31.7%	26.0%	30.8%
Data collections	31.7%	33.3%	26.0%	30.4%
Online storage	33.7%	27.0%	23.3%	28.8%
Collaboration tools	25.0%	44.4%	19.2%	28.3%
Remote access to research instruments	12.5%	25.4%	16.4%	17.1%
Individual support/advice	22.1%	25.4%	21.9%	22.9%
Other	4.8%	22.2%	5.5%	9.6%
My own applications ported on the e-infrastructure	35.6%	27.0%	38.4%	34.2%
Online digital materials for research	24.5%	20.0%	12.1%	20.4%

6.4.4 Intensity of involvement

Two variables permitted us to assess the degree of involvement of the respondents with the specified e-infrastructure:

- The share of work time invested in using the e-infrastructure or developing for the e-infrastructure (question 13 of the questionnaire, see annex),

- An index that was calculated by counting the number of services and resources used (developed) and the frequency of involvement distinguishing between i) high involvement: respondent involved with four or more different services and resources at least once a week; ii) medium involvement: respondent involved with two or more different services and resources at least once a week or with more than four different services and resources in total; iii) small involvement: respondent involved with one to three different services and resources but at maximum one out of these weekly or more often; iv) no involvement: no service or resource used or developed within the previous six months.

Involvement in the e-infrastructure looks rather similar, both measured in time or the number of services and resources in which the respondents are involved (see Table 6-23 and Table 6-24). Small involvement or less than 25% of the working time appears for roundabout 45% of the respondents; high involvement and more than 75% of the working time appear only for 16-17%. Both indicators do not correlate perfectly (see table 2-21 in the annex), and in combination they give a good picture of involvement in e-infrastructure.

| Table 6-23: Respondents by time of involvement in the selected e-infrastructure					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	less than 25%	177	43.5	47.7	47.7
	between 25 and 75%	131	32.2	35.3	83.0
	75% or more	63	15.5	17.0	100.0
	Total	371	91.2	100.0	
Missing	System	36	8.8		
Total		407	100.0		

| Table 6-24: Respondents by intensity of involvement in the selected e-infrastructure					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Small involvement	160	39.3	43.1	43.1
	Medium involvement	150	36.9	40.4	83.6
	High involvement	61	15.0	16.4	100.0
	Total	371	91.2	100.0	
Missing	System	36	8.8		
Total		407	100.0		

The continents of the respondents do not covary with the time indicator, but with the index on the intensity of services and resources used (see Figure 6-19). Respondents from Asia show highest involvement and Latin-Americans are least involved according to this measure. The differences are rather small, however.

Figure 6-19: Respondents by number of services and resources used or developed and continent (in %)
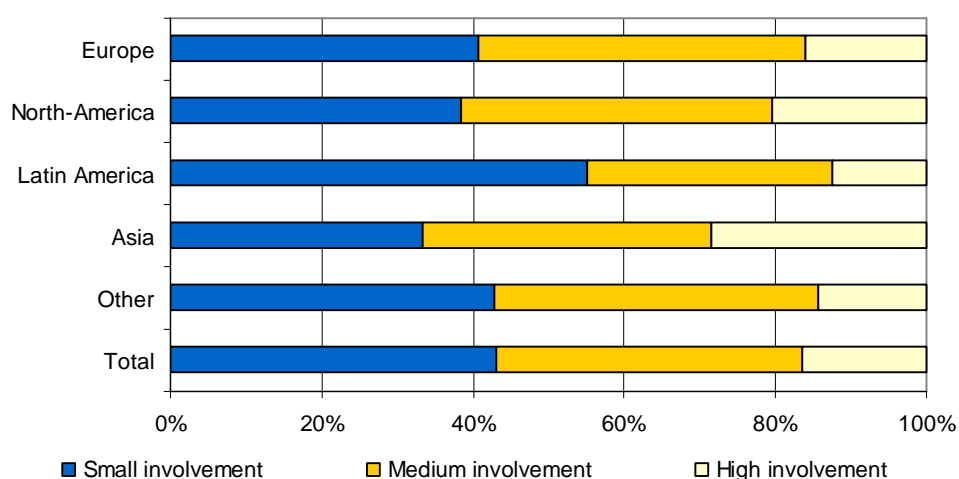
Both indicators convey the same message when we relate involvement with e-infrastructure to respondents' primary affiliation (see Figure 6-20 and table 2-22 in the annex). Respondents from governments and international organizations are the ones involved most intensively, whereas respondents from the private sector appear the least involved.

Figure 6-20: Respondents by time of involvement in e-infrastructure and primary affiliation (in %)
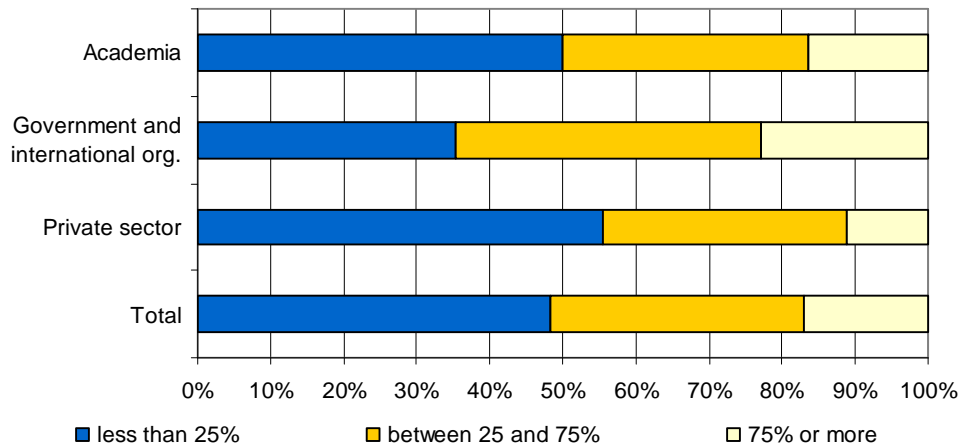
Interestingly, the indicators differ somewhat if we look at the function of involvement, i.e. whether a respondent started to work with an infrastructure primarily as research user, other user or developer. Whereas developers are the most involved with both indicators, research users are the least involved when it comes to the amount of working time they dedicate to working with the e-infrastructure (see Figure 6-21); other users are the least involved when it comes to the number of services and resources used and the frequency of use (see Figure 6-22). This result essentially points to different use patterns of both groups: research users require a broader set of applications without being able to spend too much time with each, whereas other users require only very few applications which they then use more frequently.

Figure 6-21: Respondents by time of involvement in e-infrastructure and main function of involvement (in %)
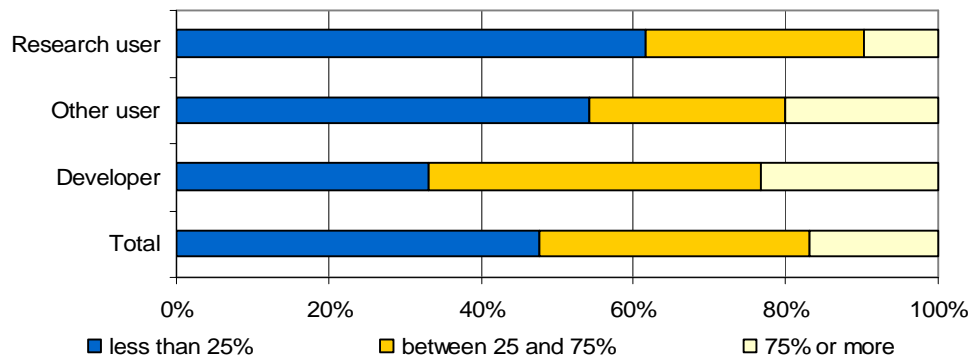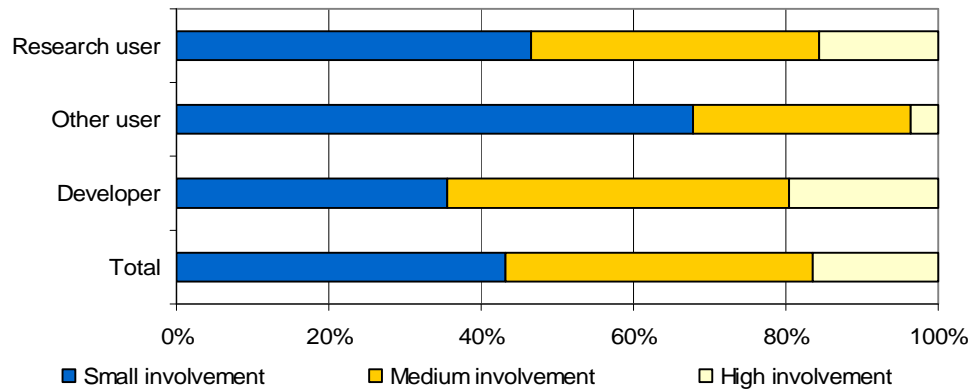
Figure 6-22: Respondents by number of services and resources used or developed and main function of involvement (in %)

As we may expect, the longer the respondents have worked with an e-infrastructure, the more they are involved today, both in regard to the time they work with the infrastructure and the services and resources with which they work (see Figure 6-23 and Figure 6-24). We get similar but slightly less clear results, if we relate the year of first involvement to the start year of the project and differentiate the indicators for involvement by how many years after the project start respondents became involved in it (see Table 2 23 and Table 2 24 in the annex).

Figure 6-23: Respondents by time of involvement in e-infrastructure and calendar year in which they became involved in it (in %)
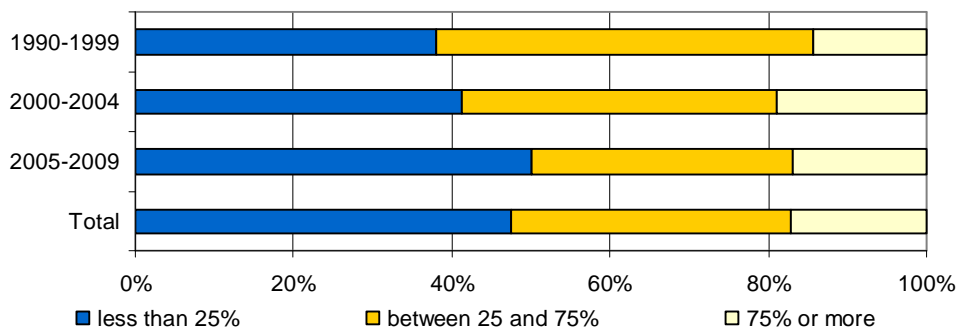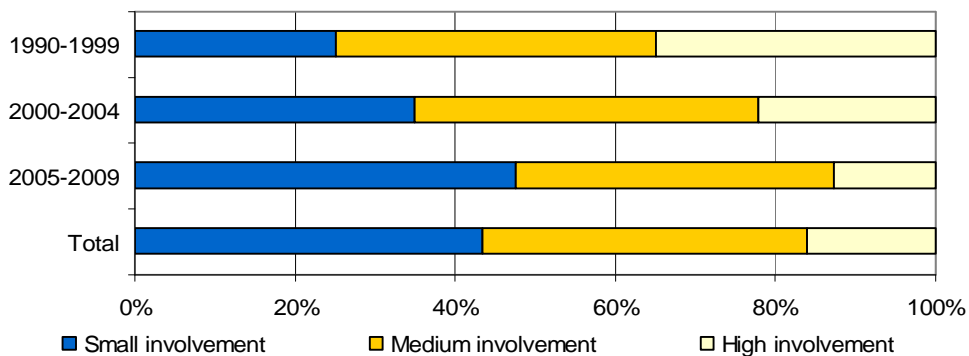

Figure 6-24: Respondents by number of services and resources used or developed and calendar year in which they became involved in the e-infrastructure (in %)

There are no clear patterns for the relationship between research fields and intensity of involvement (see Table 2 25 and Table 2 26 in the annex) and there are no differences between the different types of "established low collaboration", "novel dynamic collaborative" and "dynamic competitive" fields (see Table 2 27 and Table 2 28 in the annex).

6.4.5 Catalysts of and barriers to involvement

Catalysts of and barriers to involvement were asked in open-ended questions which permitted that the respondents themselves assessed and prioritized the most important influences on the adoption process. These responses were then coded independently by experts from the project team. Examples for the answers and the respective codes are shown in Table 6-25.

Table 6-25: Examples for answers on catalysts and barriers

	Catalysts	Barriers
Access to resources	– Access to a larger distributed network than available locally – Sharing of data across multiple institutions – Additional resources available – Computer resources assigned to DEISA	– Reasonable existing local resources – Already have access to other (much larger) resources in the US
Organizational	– Enthusiasm of most stakeholders – Collaboration among scientists – Job requirement – Developing high level analysis services for research that requires industrial-strength organization of computation flows – Good infrastructure and organization – Support from colleagues at UCSD, UvA, USC, Lucasfilm, Sony, Keio University, NTT Labs	– No support for radio astronomical data – Grid infrastructure changed often, and a few changes to my application were needed as a result sometimes – EU legal constraints not compliant with my institution's requirements – Lack of support from my institution – Low administrative pressure to stimulate the use of these tools – Bureaucracy
Technical capabilities	– Need to bridge interoperability gaps among communities of practices – Reporting tool – Computing Power and Fault Tolerance capability – Possibility to use state of the art technology – Research interest on grid technology and remote instrumentation – MT may only improve by having machines learning differently from humans	– It is not easy, in basic research, to make detailed statements on how much CPU time will be needed to complete a project. – Time required to adapt usual workflows to DEISA – Lack of structure to support anonymous access – Download and Installation of applications
Ease of use	– User-friendliness – Easy application process – Availability & reliability – Easy writing and uploading project	– Interface – Slow to get to compared to other resources – Difficult to use in the beginning
Funding-related	– Funding – Continuous funds to guarantee continuous research – Outsourcing infrastructure management and maintenance costs	– Developing fundraising and governance structure – Securing national (matching) funding – Cost of network infrastructure – Insufficient funds

	Catalysts	Barriers
	– The grant of the financing institution	
Training-related	– Technical support and training – Need of HEP communities in Latin America to create support infrastructure	– Time spent to get the application compiled and running – Learning curve – Lack of background in grid computing – Not known by individual researchers – Learning material is good, but sparsely distributed through the web
Other catalysts/ barriers	– Need for European A&H e-infrastructure – Personal interest – Desire to help the researchers – Part of my work	–

As our sample is not representative in any way, these results do not allow for any conclusions on the influences on adoption in general. We see that access to resources was the most important catalyst, mentioned by 28% of the respondents (see Figure 6-25); organizational factors and the technical capabilities of the e-infrastructure were also listed frequently among the catalysts. The latter were also the most important barriers that had to be mastered in the adoption process, together with organizational barriers and low usability of the e-infrastructure.

Figure 6-25: Respondents by catalysts and barriers (in %)
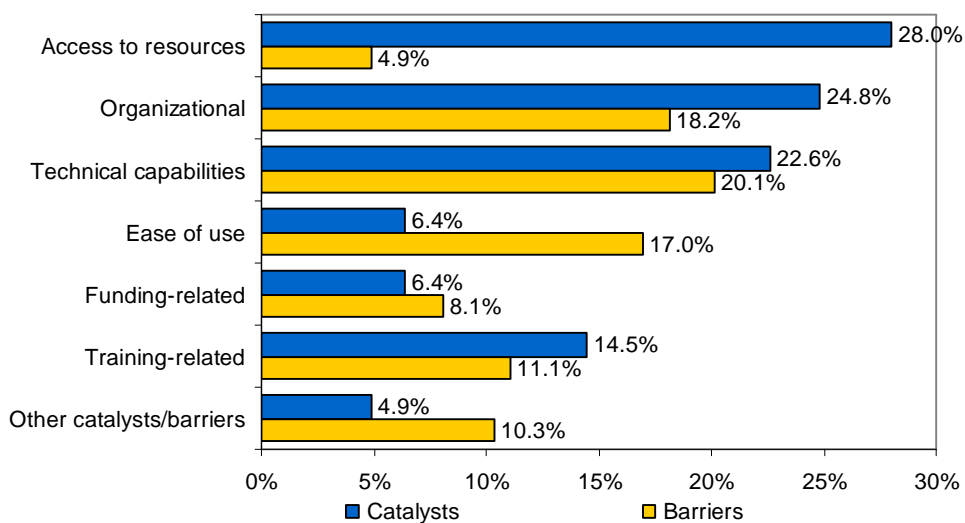
Distinguishing the catalysts and barriers by continents of the respondents we get some notable differences (see Figure 6-26 and Figure 6-27): North-Americans rated access to resources as the most important catalyst, Latin Americans organization-related issues and Asians technical capabilities. Organizational, training- and funding-related barriers were particularly important among Latin-American respondents. Among North-Americans technical capabilities, usability and organizational barriers were the most important. European and Asian respondents mentioned barriers less often than respondents from the American continent.

Figure 6-26: Catalysts by continent (in %)

Europe   North-America   Latin America   Asia

Figure 6-27: Barriers by continent (in %)

Europe   North-America   Latin America   Asia
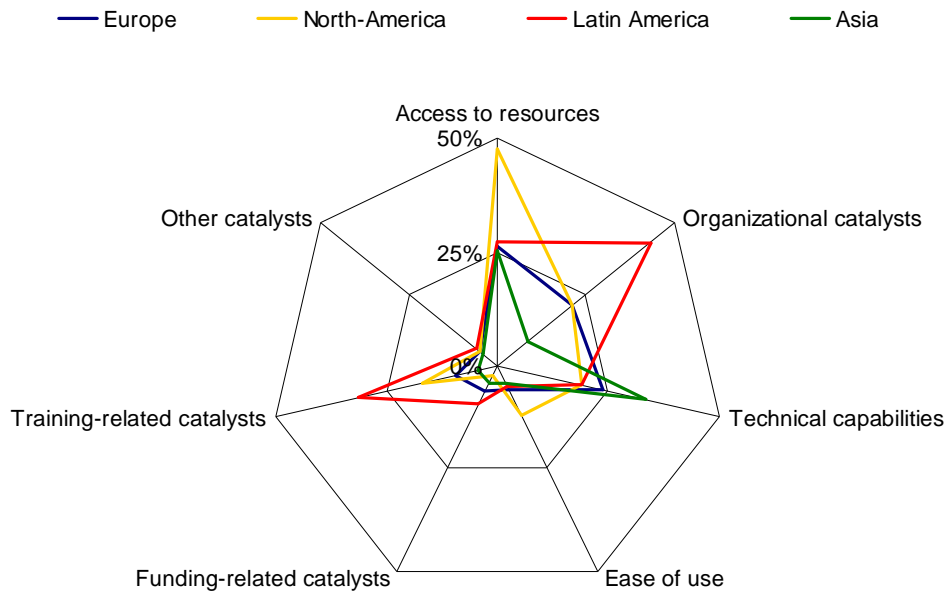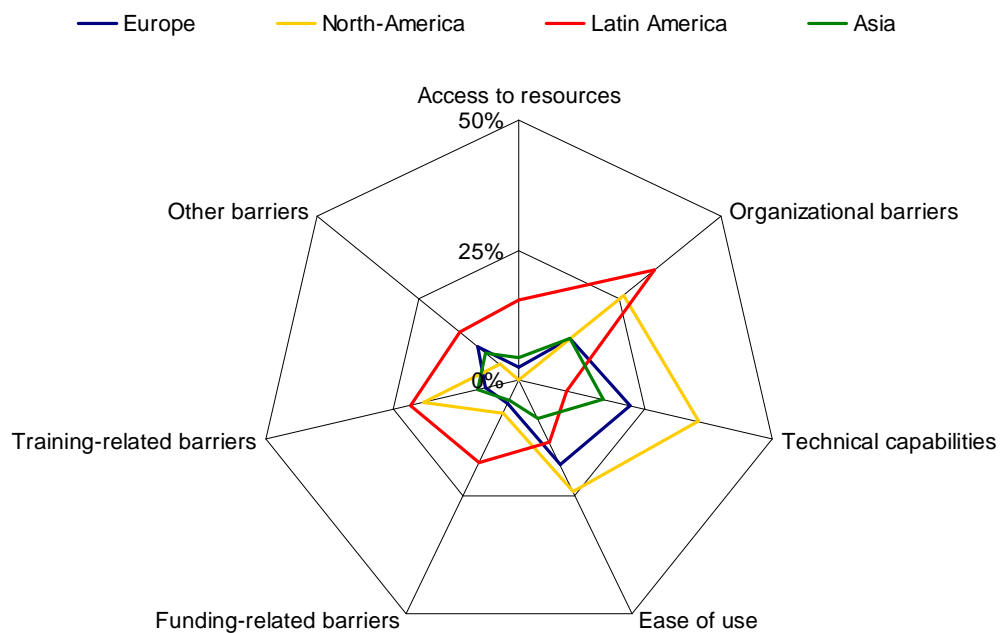

If we regroup respondents by the development level of their country we also get some interesting differences (see Table 6-26): Organizational, funding and training issues are mentioned both as catalysts and barriers more often among the developing countries. Technical capabilities and usability are more often barriers in developed countries than in the developing world.

Table 6-26: Catalysts and barriers by development level of the country (in %)				
	Catalysts		Barriers	
Catalysts and barriers	Least developed, low and middle income countries	Developed and high income countries	Least developed, low and middle income countries	Developed and high income countries
Access to resources	28.2%	28.0%	11.8%	2.4%
Organizational	33.6%	21.6%	28.2%	14.5%
Technical capabilities	20.9%	23.3%	11.8%	23.3%
Ease of use	4.5%	7.1%	11.8%	18.9%
Funding-related	8.2%	5.7%	14.5%	5.7%
Training-related	24.5%	10.8%	19.1%	8.1%
Other catalysts/barriers	6.4%	4.4%	12.7%	9.5%

Next we also find some differences between the responses on catalysts and barriers if we categorize the respondents according to their institutional affiliation (see Table 6-27). The differences are particularly strong for technical capabilities and usability. These appear – both as barriers and catalysts - as much more important for respondents from government and international organizations. Training-related catalysts and barriers were more often mentioned by respondents from academia.

Table 6-27: Catalysts and barriers by institutional affiliation (in %)					
	Catalysts			Barriers		
Catalysts and barriers	Academia	Government and international org.	Private sector	Academia	Government and international org.	Private sector
Access to resources	27.7%	36.0%	27.3%	5.5%	6.0%	0.0%
Organizational	25.8%	22.0%	22.7%	18.4%	20.0%	13.6%
Technical capabilities	21.0%	38.0%	18.2%	18.7%	36.0%	9.1%
Ease of use	5.8%	12.0%	4.5%	16.5%	26.0%	13.6%
Funding-related	6.5%	6.0%	9.1%	9.4%	6.0%	0.0%
Training-related	17.4%	8.0%	0.0%	12.3%	8.0%	0.0%
Other catalysts/barriers	5.5%	2.0%	4.5%	11.3%	8.0%	13.6%

Comparing the catalysts and barriers between the types of e-infrastructure which we distinguished – national versus international, disciplinary versus multidisciplinary, computing versus data e-infrastructures, and developer- versus community-driven – there are only few differences (see Table 6-28 in the annex). In particular for technical capabilities we get some notable differences: They were mentioned more often as catalysts in connection with national, multidisciplinary and computing infrastructures. As barriers they were mentioned more often in connection with national, disciplinary, data and community-driven infrastructures. Organizational catalysts were the dominant catalyst fro respondents involved in developer-driven e-infrastructures. Usability and training were more often mentioned as problems for becoming involved with a computing infrastructure.

Several catalysts have become more important over time and for those who joined an e-infrastructure rather late (see Figure 6-28): Access to resources, organizational catalysts, usability and funding issues were more often mentioned by those who became involved three or more years after the project had started. For barriers we would expect an opposite trend,

namely that they become less important over time and those who become involved in later phases encounter fewer barriers. However, this is not supported by the responses (see

Figure 6-29). It is not possible to identify any clear trends; only technical capabilities were clearly mentioned more often as barriers by newcomers to the infrastructure than by those involved from the start.

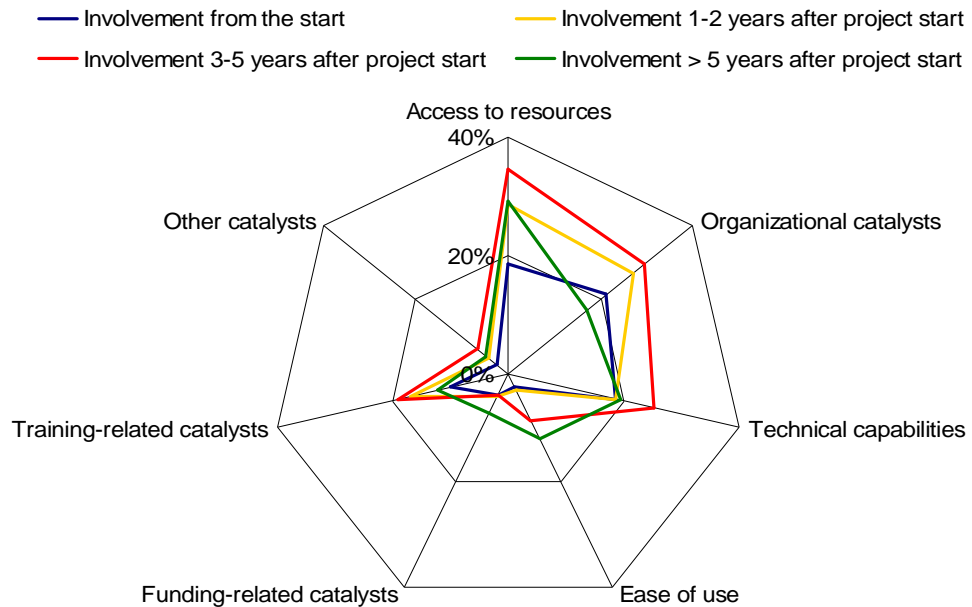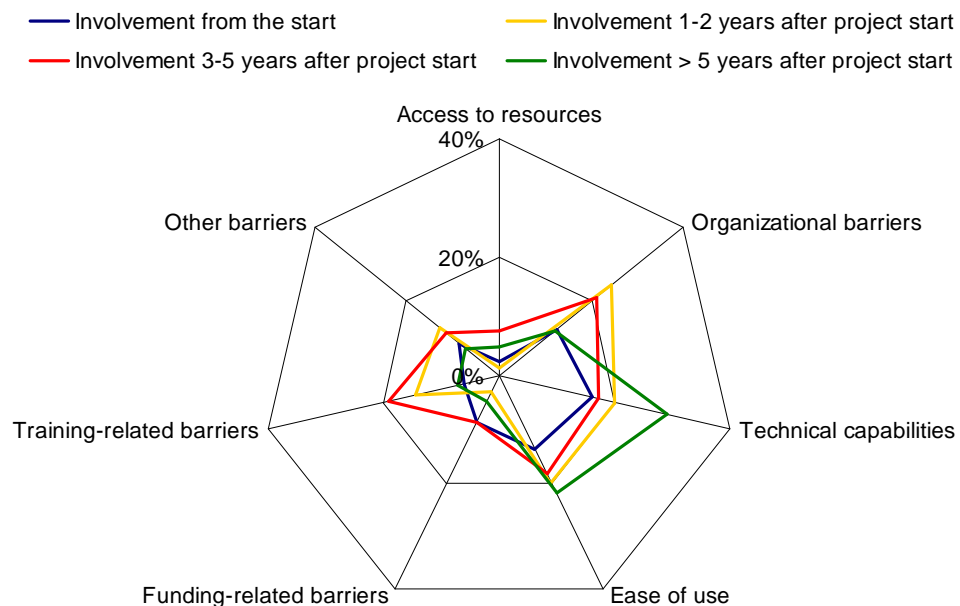**Figure 6-28: Catalysts by start of involvement with the selected e-infrastructure (in %)**

Figure 6-29: Barriers by start of involvement with the selected e-infrastructure (in %)
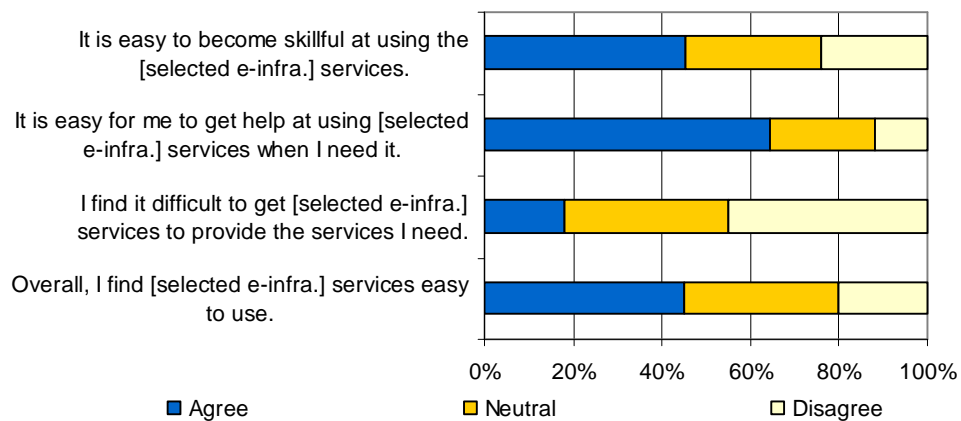
6.4.6 Usability

Respondents could assess the usability of the e-infrastructure which they had selected through four different questions:

- Statement 1: It is easy to become skilful at using the [selected e-Infrastructure] services.
- Statement 2: It is easy for me to get help at using [selected e-Infrastructure] services when I need it.
- Statement 3: I find it difficult to get [selected e-Infrastructure] services to provide the services I need.
- Statement 4: Overall, I find [selected e-Infrastructure] services easy to use.

Positive responses dominate to these questions: 45% agree that it is easy to become skilful and the services are overall easy to use and 64% answer that it is easy to obtain help when they need it. Only 18% agree that it is difficult to get the services working. However, on each question there is a sizeable share of undecided or neutral respondents who could not readily agree to the statements. This tarnishes the overall positive impression somewhat (see Figure 6-30).

Figure 6-30: Assessment of the usability of the selected e-infrastructure (in %)

Grouping the responses by type of e-infrastructure, we get a slightly surprising finding (Table 6-28): usability is more of a problem for national, disciplinary, data and community-driven e-infrastructures according to the responses to statements 1, 2 and 4 of the (matrix) question. However, *fewer* respondents agreed to statement 3 which was worded negatively. A possible explanation for this apparent inconsistency could be a response problem: respondents might have overlooked the negative wording and simply given the wrong answer. This explanation, however, seems unlikely, as we do get low but significant negative rank correlation coefficients between statement 3 and the other statements (see table 2-30 in the annex). So respondents answered consistently. Another more likely explanation is that the questions indeed point to different aspects of usability: becoming skilful addresses the initial learning efforts which are not evaluated too differently across the different types of e-infrastructures. The availability of help covers the existence of support services which is according to the survey responses more developed for international, multidisciplinary, computing and developer-driven e-infrastructures. Overall usability (statement 4) and needed service (statement 3) are in different dimensions, too. For instance, it may be that thanks to the provided help a computing e-infrastructure is overall easy to use, but this does not necessarily guarantee that its service matches one-to-one on the users needs. The fact that respondents more often disagree to statement 3, the earlier they have become involved in an

infrastructure and the more intensively they use it confirms our judgement regarding the consistency of the responses (see Table 2 31, Table 2 32, Table 2 33 in the annex).

Table 6-28: Assessment of the usability of the selected e-infrastructure by type of e-infrastructure (in %)

Assessment of the usability of the selected e-infrastructure	Geographic scope		Disciplinary scope		Type of service		Driver	
	National	Inter-national	Disci-plinary	Multidis-ciplinary	Com-puting	Data	Devel-oper	Com-munity
It is easy to become skilful at using the [selected e-infra.] services.	27.6%	39.0%	34.8%	36.3%	35.9%	38.5%	38.8%	30.0%
It is easy for me to get help at using [selected e-infra.] services when I need it.	31.0%	66.7%	34.8%	63.5%	62.9%	42.3%	65.4%	46.3%
I find it difficult to get [selected e-infra.] services to provide the services I need.	13.8%	18.8%	8.0%	20.8%	20.4%	7.4%	20.0%	12.2%
Overall, I find [selected e-infra.] services easy to use.	27.6%	39.8%	30.4%	38.0%	39.2%	28.0%	41.0%	27.5%

6.4.7 Involving others

The questionnaire also included a question that assessed whether respondents in addition to being involved themselves engaged in activities to summon the participation of others in the e-infrastructure on which they reported. Engaging others may be interpreted as a symbol of satisfaction and positive evaluation of the involvement.

We see that overall a majority of the respondents engaged in such activities. Only 14% answered that they have not engaged in any such activities. Most common were informal communication with colleagues at the same institution, talks and/or demonstrations and communication with colleagues at other institutions. Formal publications were slightly less common (see Figure 6-31).

Figure 6-31: Activities undertaken to involve others in the selected e-infrastructure (in %)
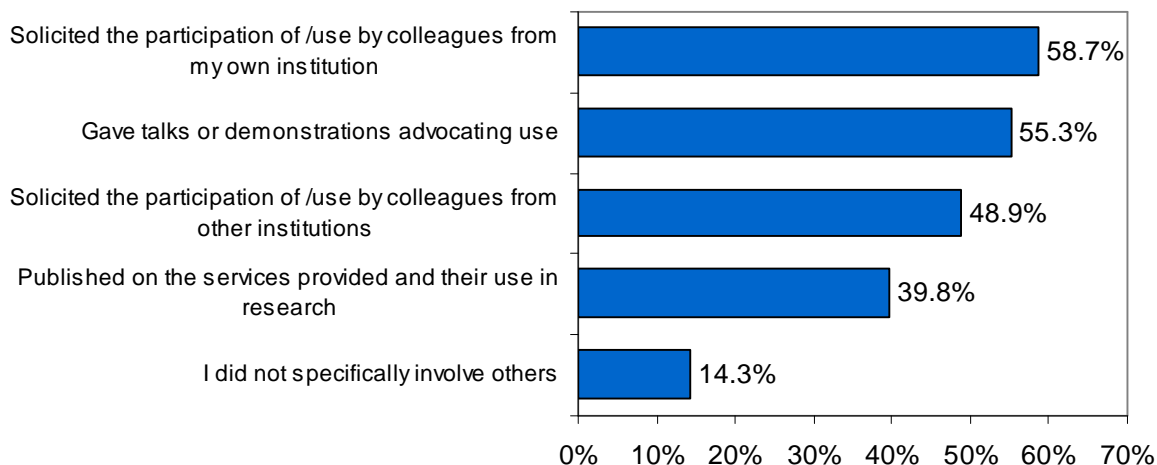
If we compare the responses by type of involvement in the e-infrastructure, i.e. whether respondents classified themselves as research users, other users, or developers, we see that the developers are the most active disseminators in particular when it comes to talks and publications (see Table 6-29). Among the developers those who are active in the areas of supercomputing and distributed computing excel the others (see table 2-34 in the annex). The other users are the least active group. We can deduce from this finding a certain bias of the outreach activities into the developers' communities.

Table 6-29: Respondents by activities undertaken to involve others and type of involvement in the e-infrastructure (in %)

Activities undertaken to involve others	Type of involvement in the e-infrastructure			
	Research user	Other user	Developer	Total
Gave talks or demonstrations advocating use	46.0%	37.8%	68.3%	55.3%
Published on the services provided and their use in research	32.6%	13.5%	52.5%	39.8%
Solicited the participation of /use by colleagues from my own institution	57.8%	43.2%	62.8%	58.7%
Solicited the participation of /use by colleagues from other institutions	43.3%	27.0%	59.0%	48.9%
I did not specifically involve others	14.4%	32.4%	10.4%	14.3%

Respondents involved in EGEE were the most active ones when it comes to talks and demonstrations and publishing on the infrastructure (see Table 6-30). Those involved in DEISA were the least active ones. Respondents working with EELA-2 specifically solicited participation from local peers. It will come as little surprise that those involved rather early in an infrastructure had more time and opportunities to involve others than those who became involved more recently (see annex table 2-35).

Table 6-30: Respondents by activities undertaken to involve others and selected e-infrastructure (in %)

Activities undertaken to involve others	E-infrastructure selected by the respondent					
	DEISA	EELA-2	EGEE	US NVO	Other	Total
Gave talks or demonstrations advocating use	35.0%	52.1%	70.9%	60.0%	55.6%	55.3%

Published on the services provided and their use in research	25.0%	42.5%	45.5%	40.0%	40.2%	39.8%
Solicited the participation of /use by colleagues from <u>my own</u> institution	42.5%	72.6%	60.0%	48.0%	57.9%	58.7%
Solicited the participation of /use by colleagues from <u>other</u> institutions	47.5%	53.4%	50.9%	28.0%	49.5%	48.9%
I did not specifically involve others	20.0%	8.2%	9.1%	20.0%	15.9%	14.3%

6.5   Impact of e-infrastructure involvement

6.5.1 General importance and effects of a lack of e-infrastructure

As a first assessment of the impact we let research users and other users rate the importance of the e-infrastructure they had selected for their research or work and then asked them what consequences would result if this or similar e-infrastructures were lacking (see questions 28, 29, 46, and 47 in the questionnaire in the annex on the exact wording). The importance rating shows little variance and more than 85% of the respondents rated the e-infrastructure as important or very important (see Table 6-31).[23] Similarly, only few respondents do not see their research or work programmes impaired if the e-infrastructure did not exist (see

Table 6-32).

Table 6-31: Importance of the selected e-infrastructure for the research or work of the respondents

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Very unimportant	3	.7	1.7	1.7
	Unimportant	5	1.2	2.8	4.5
	Neutral	14	3.4	7.9	12.4
	Important	55	13.5	31.1	43.5
	Very important	100	24.6	56.5	100.0
	Total	177	43.5	100.0	
Missing	System	230	56.5		
Total		407	100.0		

Note: As this question was not asked to developers a large number of missing values appears.

Table 6-32: Research or work programme would be impaired if the selected e-infrastructure or similar resources were lacking

|		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	not at all	13	3.2	7.6	7.6
	a little	40	9.8	23.4	31.0
	much	92	22.6	53.8	84.8
	totally	26	6.4	15.2	100.0
	Total	171	42.0	100.0	
Missing	System	236	58.0		
Total		407	100.0		

---

[23] This lack of variance somewhat limits the usefulness of this question for further analyses.

Note: As this question was not asked to developers a large number of missing values appears.

North-American respondents experience a higher risk of having their research and work programmes negatively affected than respondents from other continents (see table 2-36 in the annex).

We obtain an interesting result when we distinguish responses on the importance by e-infrastructure type and compare national with international, disciplinary with multidisciplinary, computing with data and developer-driven with community-driven e-infrastructures (seeFigure 6-32). The importance of e-Infrastructure is generally rated as higher for the border- and discipline-crossing infrastructures. Computing infrastructures are more often rated as important than data infrastructures, as well as developer-driven compared to community-driven infrastructures. We get similar results, if we use the other question on the consequences of a lack of e-infrastructure services (see annextable 2-37).

Figure 6-32: Importance of the selected e-infrastructure for the research or work of the respondents by type of e-infrastructure (in %)
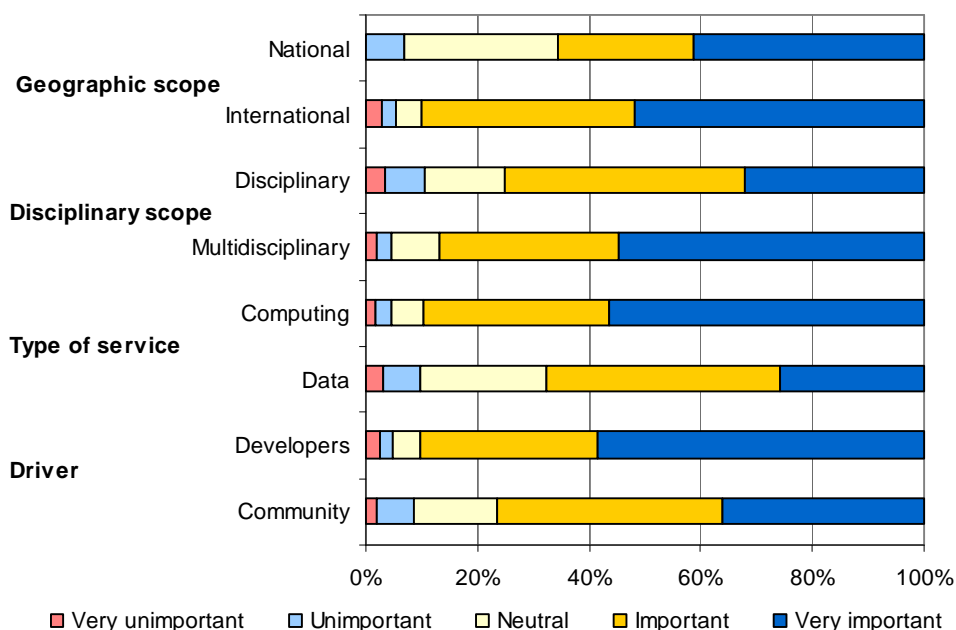
It is hardly a surprise to see that over time the availability of an e-infrastructure binds the respondents (see Table 6-33): Those who became involved in the 1990s would be significantly more affected than those who became involved in later years if the infrastructure or an appropriate substitute were not available.

Table 6-33: Research or work programme would be impaired if the selected e-infrastructure or similar resources were lacking by year of first involvement in the e-infrastructure (in %)

Lack of the e-Infrastructure or similar resources would impair my research/work programme	Year of first involvement			
	1990-1999	2000-2004	2005-2009	Total
Not at all	0.0%	9.1%	7.8%	7.6%
A little	11.1%	15.2%	26.6%	23.5%
Much	44.4%	51.5%	55.5%	54.1%
Totally	44.4%	24.2%	10.2%	14.7%
Total	100.0%	100.0%	100.0%	100.0%

Field differences are as expected from the involvement or use intensities (compare Table 2 38 with Table 2 25 and Table 2 26 in the annex): Earth scientists and astronomers would be most affected; the social scientists, life scientists and physical scientists among the respondents rather not. We classified above the fields of research and work of the respondents as established low collaboration, novel dynamic collaborative or dynamic competitive, depending on the patterns of collaboration and competition, maturity and field dynamics (see 6.2.3). Respondents from the fields classified as novel dynamic collaborative and dynamic competitive seem to be more sensitive to e-infrastructure availability than respondents from established low collaboration fields (see Figure 6-33 and Table 2 39 in the annex).

Figure 6-33: Importance of the selected e-infrastructure for the research or work of the respondents by type of field (in %)

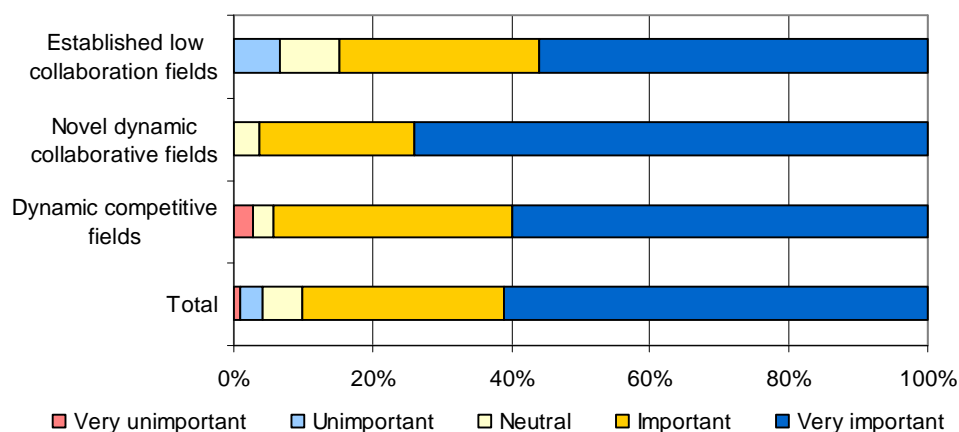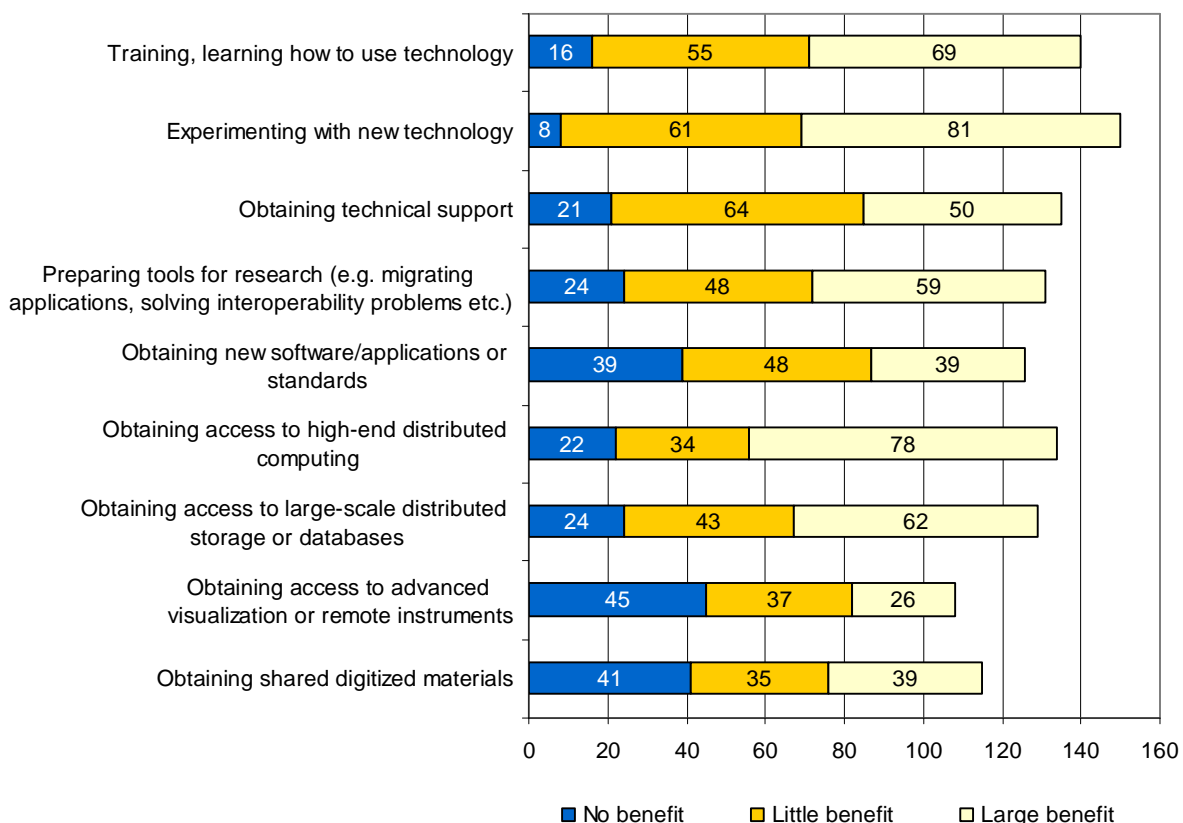Note: Not all respondents answered the questions on field characteristics and could be classified according to this; hence, the "total" bar is on a larger data set than the sum of the three field-specific bars.

6.5.2 Impact of e-infrastructure on research and other use

Another more detailed measure of the impact of e-infrastructures was built on the basis of a set of more technical benefits and the experiences of working with new technology, for instance learning how to use technology, obtaining access to high-end distributed computing, obtaining shared digitized materials and the like (see questionnaire in the annex on the full questions). Necessarily this question was asked with different wordings to research/non-research users and developers. As we see inFigure 6-34, the possibility to experiment with new technology, obtaining access to high-end distributed computing, obtaining access to large-scale distributed storage or databases and training and learning how to use technology were rated most often as the benefits. Obtaining technical support and preparing tools for research were slightly lower rated and obtaining access to data and other resources (new software/applications, standards, advanced visualization or remote instruments) received the fewest positive ratings. However, this result is clearly related to the distribution of responses across different types of e-infrastructure. As can be seen from the annex table 2-8, we have nearly twice as many respondents involved with computing infrastructures than data infrastructures and the former rate access to data and other resources less often as beneficial (see table 2-40 in the annex).

Figure 6-34: Respondents by degree and type of benefits that result from using the selected e-infrastructure[a]

Type of benefit	No benefit	Little benefit	Large benefit
Training, learning how to use technology	16	55	69
Experimenting with new technology	8	61	81
Obtaining technical support	21	64	50
Preparing tools for research (e.g. migrating applications, solving interoperability problems etc.)	24	48	59
Obtaining new software/applications or standards	39	48	39
Obtaining access to high-end distributed computing	22	34	78
Obtaining access to large-scale distributed storage or databases	24	43	62
Obtaining access to advanced visualization or remote instruments	45	37	26
Obtaining shared digitized materials	41	35	39

■ No benefit   □ Little benefit   □ Large benefit

a N_{max} = 150 as question only asked to research users.

Another not surprising finding is that it takes some time for the benefits to materialize and that they are not always obvious at the beginning. We see that the percentage of respondents who sees large benefits is higher for most of the types of benefits among respondents who became involved with an e-infrastructure in the first period 1990-99 than for those who became involved more recently between 2005 and 2009 (see Table 6-34).

Table 6-34: Percentage of respondents with large benefits from using the selected e-infrastructure and year of first involvement with the infrastructure (in %)

Type of benefit	Year of first involvement with the e-infrastructure		
	1990-1999	2000-2004	2005-2009
Training, learning how to use technology	66.7%	63.3%	44.0%
Experimenting with new technology	62.5%	59.4%	52.3%
Obtaining technical support	66.7%	41.4%	33.3%
Preparing tools for research (e.g. migrating applications, solving interoperability problems etc.)	62.5%	45.2%	42.9%
Obtaining new software/applications or standards	71.4%	36.7%	26.1%
Obtaining access to high-end distributed computing	57.1%	67.9%	56.1%
Obtaining access to large-scale distributed storage or databases	42.9%	67.9%	43.0%
Obtaining access to advanced visualization or remote instruments	57.1%	40.0%	16.0%
Obtaining shared digitized materials	50.0%	50.0%	26.3%

Another set of questions collected the respondents' opinion on how the selected e-infrastructure affected their research or work. Respondents who had classified themselves as "developers" for the infrastructure in a previous filter question were asked to judge the impact on the research of the e-infrastructure users as they experienced it (see the questions 26, 44, and 62 in the annexed questionnaire)[24].

First we note that there is widespread agreement about the positive impact of e-infrastructures. For all items the number of positive opinions exceeds the number of negative opinions by at least the factor three (see Table 6-34). For 7 out of 8 items more than 60% of the respondents agree to experience a positive impact. The main benefits, i.e. those that respondents agreed to most often, refer to the speed of doing research or work: accomplish tasks more quickly, access resources faster or better, produce process or analyse data faster or better. Equally important is the ability to work on new problems which could not be addressed with previously available technology. Slightly less frequently respondents agreed with positive effects on productivity ("Produce more output per year"), costs, and quality ("Do more accurate, higher quality research work"). The lowest number of positive responses was on the impact in regard to publications.

Figure 6-35: Respondents' agreement to statements on the impact of using the selected e-infrastructure (in %)
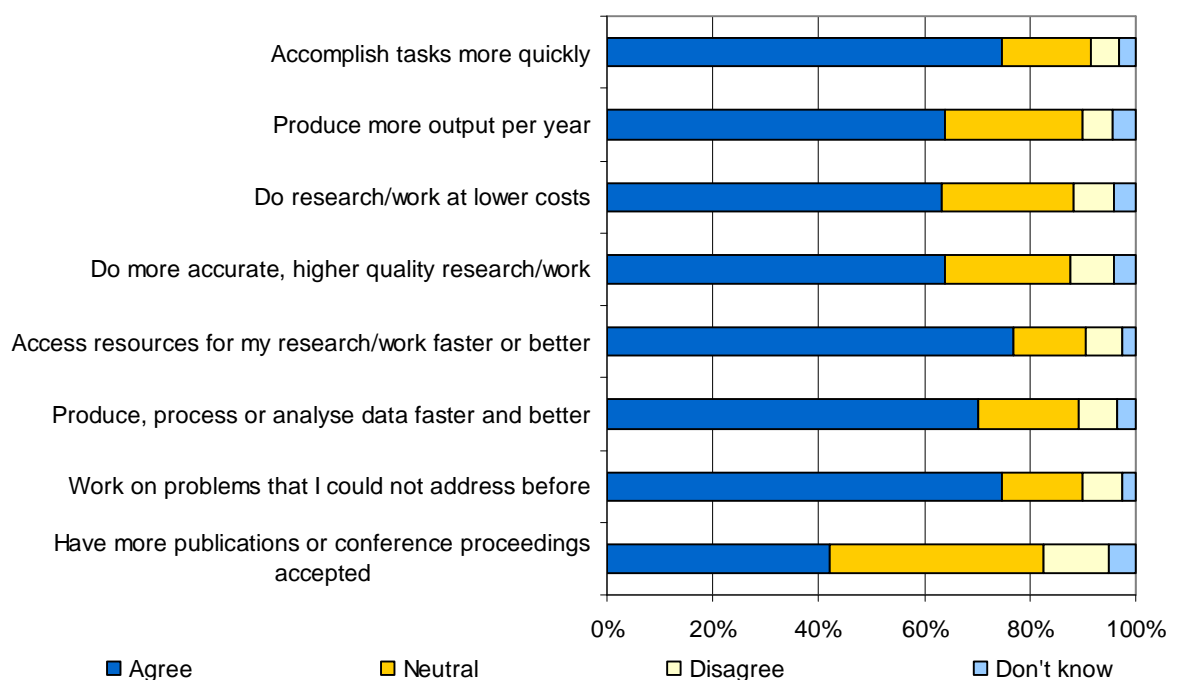
We differentiate these impact ratings now in regard to several other variables which characterize the respondents. How the ratings co-vary with the continents of the respondent is shown in annex table 2-41. We see that North-Americans are slightly more negative about all the impacts than Europeans, whereas Latin Americans and Asians are slightly more positive. But the differences are not too pronounced in most cases. They are a little bit larger if we group respondents not by continent but in a different manner and according to the development status of their countries of residence (see Table 6-35). Then we see that respondents from least developed, low and middle income countries more often agree to the positive impacts than respondents from developed high income countries.

---

24 This structural difference between the questions did not, however, affect the results. As can be seen in Table 6-29 research users, professional users and developers agree to more or less the same extent to the statements on the impact of an e-infrastructure.

Table 6-35: Respondents agreeing to statements on impact of using the selected e-infrastructure by development level of their country (in %)

The selected e-infrastructure has enabled me to ...	Least developed, low and middle income countries	Developed and high income countries	Total
Accomplish research tasks more quickly	77.1%	73.8%	74.6%
Produce more research output per year	66.2%	63.1%	63.9%
Do research at lower costs	63.0%	63.4%	63.3%
Do more accurate, higher quality research	73.6%	60.6%	64.0%
Access resources for my research faster or better	81.9%	75.2%	77.0%
Produce, process or analyse data faster and better	73.2%	69.3%	70.4%
Work on research problems that I could not address before	79.5%	72.9%	74.6%
Have more publications or conference proceedings accepted	49.3%	39.5%	42.1%

The differences are somewhat more pronounced if we look at the selected e-infrastructures (see Table 6-36). On several aspects EELA-2 receives the most positive ratings out of all projects. In regard to DEISA more people than on average stress the positive impact on research quality, which is even more striking as EGEE users agreed to this statement notably fewer times (and NVO users even less). EGEE users also agreed less often to access resources faster or better and do research at lower costs, but still half of the respondents give positive ratings for both questions. NVO users agreed to all statements less often than the total set of respondents, and they gave quite low ratings for the statements on the impact on research productivity, quality and acceptance of publications. The main impact of NVO is the faster or better resource access and the possibility to address new problems. However, these aspects can then be found again if we differentiate the projects by geographic scope, disciplinary scope, type of service and driver (see Table 6-37). NVO is the only national and disciplinary data infrastructure displayed in Table 6-36; the others are all international and multidisciplinary computing projects. So the NVO rating seems not to be an NVO particularity, but a feature of this type of e-infrastructure. For national, disciplinary, data and community-driven e-infrastructures positive impacts are seen less often than for international, multidisciplinary and computing e-infrastructures by our respondents. The only exceptions are resource access, where disciplinary and data e-infrastructures were rated slightly better than their corresponding counterparts.

Table 6-36: Respondents agreeing to statements on impact of using the selected e-infrastructure by e-infrastructure (in %)

The selected e-infrastructure has enabled me to ...	E-infrastructure selected by the respondent					
	DEISA	EELA-2	EGEE	US NVO	Other	Total
Accomplish research tasks more quickly	77.1%	71.4%	81.3%	65.0%	75.0%	74.6%
Produce more research output per year	72.7%	67.3%	71.0%	26.3%	64.1%	63.9%
Do research at lower costs	58.1%	72.0%	58.1%	50.0%	64.3%	63.3%
Do more accurate, higher quality research	79.4%	70.0%	53.3%	31.6%	64.8%	64.0%
Access resources for my research faster or better	79.4%	84.0%	62.5%	70.0%	78.1%	77.0%
Produce, process or analyse data faster and better	60.0%	75.5%	75.9%	42.1%	73.4%	70.4%
Work on research problems that I could not address before	85.7%	80.4%	64.5%	66.7%	73.1%	74.6%
Have more publications or conference proceedings accepted	50.0%	54.3%	43.3%	5.9%	40.3%	42.1%

Table 6-37: Respondents agreeing to statements on impact of using the selected e-infrastructure by type of e-infrastructure (in %)

The selected e-infrastructure has enabled me to …	Geographic scope		Disciplinary scope		Type of service		Driver	
	National	Inter-national	Disci-plinary	Multidis-ciplinary	Comp-uting	Data	Devel-oper	Com-munity
Accomplish research tasks more quickly	65.3%	77.0%	71.1%	74.5%	76.1%	67.9%	77.9%	67.1%
Produce more research output per year	50.0%	65.8%	47.6%	66.2%	71.1%	37.0%	68.5%	50.0%
Do research at lower costs	50.0%	62.5%	50.0%	64.0%	63.8%	47.3%	66.4%	50.7%
Do more accurate, higher quality research	39.6%	66.5%	51.1%	64.9%	67.3%	42.9%	75.2%	42.3%
Access resources for my research faster or better	62.0%	80.2%	82.6%	75.5%	73.7%	82.8%	86.8%	63.5%
Produce, process or analyse data faster and better	56.3%	71.2%	61.4%	69.0%	71.4%	57.4%	72.2%	58.0%
Work on research problems that I could not address before	68.0%	74.3%	64.6%	74.7%	78.2%	56.9%	77.8%	61.6%
Have more publications or conference proceedings accepted	27.9%	43.7%	18.9%	46.2%	49.0%	14.6%	48.5%	26.2%

Not only the type of e-infrastructure, but also the intensity with which an e-infrastructure is used correlates with the impact rating (see Table 6-38): The more respondents are involved in an infrastructure in regard to the number of services that they use and the frequency of use (see section 6.4.4 on the indicator), the more often they attribute a positive impact to this infrastructure. The same applies also to the time they spend on working with an e-infrastructure (see table 2-42 in the annex). However, this should not be interpreted as a causal relationship between involvement and impact. We cannot deduce from this result, that more involvement also produces more positive impact. It may well be that respondents who are more involved in an e-infrastructure are also more positive about its impact and feel the need to justify their involvement. There are limitations to the validity of such self-assessments of impacts which we cannot resolve.

Table 6-38: Respondents agreeing to statements on impact of using the selected e-infrastructure by intensity of e-infrastructure involvement (in %)

The selected e-infrastructure has enabled me to …	Intensity of involvement with services and resources of selected e-infrastructure			
	Small involvement	Medium involvement	High involvement	Total
Accomplish research tasks more quickly	63.3%	80.3%	86.8%	74.6%
Produce more research output per year	55.1%	66.4%	77.4%	63.9%
Do research at lower costs	51.9%	69.0%	75.5%	63.3%
Do more accurate, higher quality research	52.8%	66.1%	83.0%	64.0%
Access resources for my research faster or better	67.6%	83.3%	83.0%	77.0%
Produce, process or analyse data faster and better	62.4%	72.2%	83.0%	70.4%
Work on research problems that I could not address before	62.4%	82.5%	83.0%	74.6%
Have more publications or conference proceedings accepted	35.4%	42.3%	54.9%	42.1%

Given the results in Table 6-34 above, it is not surprising that respondents who became involved in the e-infrastructure in an early phase, 1990-1999 or 2000-2004, are more positive about the impact than those who became involved more recently between 2005 and 2009 (see table 2-43 in the annex).

There are some field differences, but the N for each of the fields is quite low and the reliability of the results is therefore debatable (see table 2-45 in the annex). Case numbers become less of a problem if we use field classifications (see Table 6-39). Respondents from fields classified as "dynamic competitive" or "novel dynamic collaborative" are quite similar in regard to their agreement to the statements on e-infrastructure impact. On the contrary, respondents from "established low collaboration fields" show less agreement to all statements. And this is *not* due to the fact that they use e-infrastructures less intensively than respondents from the other types of fields – as we have seen above, they use them more or less to the same extent (see section 1.4.4 and Table 2 27 and Table 2 28 in the annex).

Table 6-39: Respondents agreeing to statements on impact of using the selected e-infrastructure by type of field (in %)

The selected e-infrastructure has enabled me to …	Type of field			
	Established low collaboration	Novel dynamic collaborative	Dynamic competitive	Total
Accomplish research tasks more quickly	61.1%	85.0%	86.4%	74.6%
Produce more research output per year	55.9%	70.0%	71.9%	63.9%
Do research at lower costs	61.1%	63.3%	72.3%	63.3%
Do more accurate, higher quality research	53.8%	61.0%	75.8%	64.0%
Access resources for my research faster or better	66.7%	83.1%	84.8%	77.0%
Produce, process or analyse data faster and better	62.2%	77.2%	78.1%	70.4%
Work on research problems that I could not address before	72.0%	71.2%	77.6%	74.6%
Have more publications or conference proceedings accepted	36.0%	47.3%	45.9%	42.1%

6.5.3 Impact of e-infrastructure on collaboration networks

E-infrastructures cannot only affect how research is done and what results are produced, but also who works with whom in the process. In order to cover this, the questionnaire included a question on the relationship between working with a selected e-infrastructure and certain trends in the respondents' collaboration networks: Overall growth, geographical range, collaboration with colleagues from developing countries, collaboration with commercial firms, academic institutions and colleagues from other fields of science.

The respondents stated in three quarters of the cases agreement to the statements that a) they collaborate more, b) they collaborate more widely in the geographical sense, and c) they collaborate more with academic institutions thanks to the influence of the e-infrastructure they work with (see Figure 6-36). Less agreement received the statement on more collaboration with colleagues from other fields of science/work, labelled here shortly "interdisciplinary collaboration". The positive and negative opinions on the statement on more collaboration with colleagues from developing countries are nearly balanced. Only few respondents agreed with having more collaboration with commercial firms due to their e-infrastructure involvement.

Figure 6-36: Respondents' agreement to statements on the influence of using the selected e-infrastructure on their collaboration networks (in %)
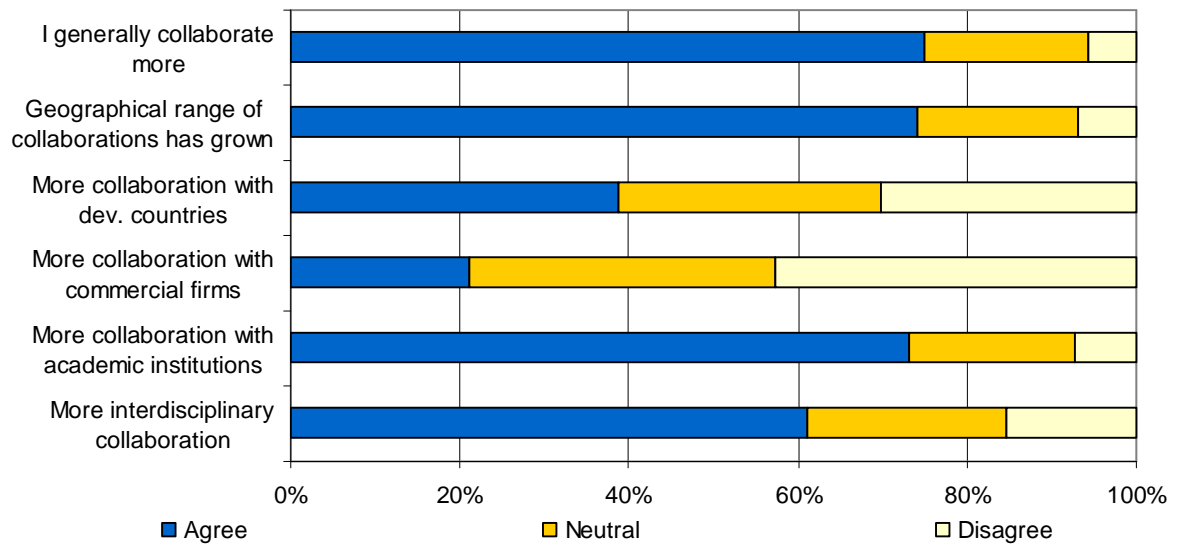
Differentiating the responses by continent we get a similar result as on the impact variables in the previous section: North-Americans are more negative about the influence of e-infrastructure on their collaborations than Europeans, whereas Latin Americans and Asians are more positive (see table 2-46 in the annex). An interesting result appears if we classify the respondents in a different manner and distinguish between developed high income countries and less developed, low and middle income countries. The influences on the collaboration network do not differ between both groups except for one variable: the influence on collaboration with developing countries. 53.5% of the respondents from developing countries agreed that their collaboration with colleagues from developing countries has grown, whereas only 33.5% of the respondents from developed countries agreed to the statement. So the positive influence of e-infrastructure on collaboration with developing countries is felt more in the form of South-South collaboration, than North-South collaboration.

We see differences across the selected e-infrastructures in a similar manner as for the overall impact as illustrated above (see Table 6-36 above). DEISA and NVO users agreed to all statements less often than the total set of respondents (see Table 6-40). No DEISA user agreed to have more collaborations with commercial firms or colleagues from the South. Working with an infrastructure that aims to foster the collaboration between Europe and Latin America EELA-2 users most felt the influence on collaboration with developing countries. EGEE users made quite often positive statements about collaboration with academic institutions and interdisciplinary collaboration. We saw above that international, multidisciplinary, computing and developer-driven e-infrastructures received more often positive impact assessments by our respondents. In regard to the influence on the collaboration networks we see the same for international compared to national e-infrastructures (see

Table 6-41): their influences on the collaboration network are more often rated as positive. For the other three categorizations these differences are not very pronounced, but we note an even opposite result regarding collaborations with commercial firms: Respondents on disciplinary, data and community-driven infrastructures more often stated that their involvement also had brought them into contact with colleagues at commercial enterprises than respondents on interdisciplinary, computing and developer-driven infrastructures.

Table 6-40: Respondents agreeing to statements on the influence of using the selected e-infrastructure on their collaboration network by e-infrastructure (in %)

My involvement with the selected e-infrastructure has influenced my collaboration network …	E-infrastructure selected by the respondent					
	DEISA	EELA-2	EGEE	US NVO	Other	Total
I generally collaborate more	53.1%	80.0%	80.0%	59.1%	78.7%	74.8%
Geographical range of collaborations has grown	51.6%	75.0%	82.4%	54.5%	79.3%	74.1%
More collaboration with colleagues from dev. countries	0.0%	64.7%	41.2%	27.3%	38.4%	38.7%
More collaboration with commercial firms	0.0%	15.7%	27.3%	23.8%	25.5%	21.1%
More collaboration with academic institutions	59.4%	80.0%	88.6%	59.1%	71.9%	73.1%
More interdisciplinary collaboration	20.7%	76.8%	85.3%	31.8%	61.6%	61.0%

Table 6-41: Respondents agreeing to statements on the influence of using the selected e-infrastructure on their collaboration network by type of e-infrastructure (in %)

My involvement with the selected e-infrastructure has influenced my collaboration network …	Geographic scope		Disciplinary scope		Type of service		Driver	
	National	Inter-national	Disci-plinary	Multidis-ciplinary	Computing	Data	Devel-oper	Com muni
I generally collaborate more	61.2%	75.1%	72.2%	71.8%	73.0%	68.8%	75.0%	69.
Geographical range of collaborations has grown	62.5%	76.1%	70.2%	73.5%	73.9%	70.6%	73.1%	70.
More collaboration with colleagues from dev. countries	22.4%	41.3%	29.4%	40.3%	39.7%	30.6%	43.6%	29.
More collaboration with commercial firms	18.8%	17.2%	24.5%	15.6%	15.4%	22.0%	13.8%	24.
More collaboration with academic institutions	59.2%	77.4%	66.7%	76.4%	76.1%	65.6%	74.6%	72.
More interdisciplinary collaboration	36.7%	67.6%	56.4%	61.7%	62.2%	56.1%	61.2%	60.

* Percentages for the total may be higher than those of either of the categories, as not all e-infrastructures could be classified and responses on non classified e-infrastructures were only included in the total.

The intensity of involvement with an e-infrastructure also correlates with the respondents' opinion on its influences on their collaboration network: the more they are involved, the more often they see positive influences on their collaborations (see Table 6-42 and table 2-47 in the annex for the indicator on working time). We do not want to rule out that there is indeed a positive influence of using e-infrastructure on collaboration activities, but we would like to add the same disclaimer as before: We do not have sufficient information to prove a causal relationship between involvement in an e-infrastructure and collaboration and we cannot entirely rule out that respondents give better ratings to justify their involvement.

Table 6-42: Respondents agreeing to statements on the influence of using the selected e-infrastructure on their collaboration network by intensity of e-infrastructure involvement (in %)

My involvement with the selected e-infrastructure has influenced my collaboration network ...	Intensity of involvement with services and resources of selected e-infrastructure			
	Small involvement	Medium involvement	High involvement	Total
I generally collaborate more	65.2%	78.2%	87.0%	74.8%
Geographical range of collaborations has grown	65.5%	76.4%	86.8%	74.1%
More collaboration with colleagues from dev. countries	31.7%	36.2%	60.0%	38.7%
More collaboration with commercial firms	14.2%	22.5%	32.7%	21.1%
More collaboration with academic institutions	66.7%	79.5%	73.6%	73.1%
More interdisciplinary collaboration	55.0%	61.5%	71.2%	61.0%

The relationship between the year in which respondents became involved in the e-infrastructure and their opinions on trends in the collaboration network is less clear than on the general impact of the e-infrastructure (see annex table 2-48). It may be that either the effect or the awareness of it wears off over time.

We see some differences between academic fields but we do not wish to interpret them because of low case numbers (see table 2-49 in the annex). As the case numbers are somewhat higher if we classify the respondents into other, field-related groups, we interpret these instead. As we would expect, less influence on collaboration is felt in the established low collaboration fields and more is felt in the novel dynamic collaborative fields (see Table 6-43). Respondents from dynamic competitive fields tend to be close to the average except for collaboration with colleagues from developing countries, where they particularly few times experienced growth.

Table 6-43: Respondents agreeing to statements on the influence of using the selected e-infrastructure on their collaboration network by type of field (in %)

My involvement with the selected e-infrastructure has influenced my collaboration network ...	Type of field			
	Established low collaboration	Novel dynamic collaborative	Dynamic competitive	Total
I generally collaborate more	65.7%	80.3%	75.8%	74.8%
Geographical range of collaborations has grown	63.5%	85.0%	76.1%	74.1%
More collaboration with colleagues from developing countries	31.6%	55.2%	25.0%	38.7%
More collaboration with commercial firms	17.0%	37.9%	16.9%	21.1%
More collaboration with academic institutions	60.2%	79.7%	73.5%	73.1%
More interdisciplinary collaboration	52.0%	68.3%	56.9%	61.0%

6.5.4 Impact clusters

Using the questions from the previous sections we constructed a compound indicator for assessing the perceived impact of an e-infrastructure and in order to obtain a more concise overview of the different groupings discussed above. This compound indicator was built by using the Two-step clusters (TSC) procedure in SPSS which is suitable for clustering ordinal variables such as the responses to the impact-related statements. For a set of 207 cases out of 407 (51%) we had impact ratings for all variables shown in Table 6-44. The TSC procedure was employed with an automatic cluster determination based on the Bayesian Information

Criterion. Cases were ordered by case ID which was automatically given to a case when the respondent clicked on the link to the survey site and started filling in the questionnaire. This ID number stems from the management of the survey and is not statistically related to any of the analytical variables. However, as the TSC is not independent of the order of the cases, we controlled the robustness of the results by running the analysis on five different case orders. The obtained results overlap to at least 87% (180 out of 207 cases). Out of the up to 27 cases that were allocated differently when the order was changed all except for one referred to clusters 1 (strong positive impact) and 2 (positive impact). Cluster 3 (no impact) was unchanged except for one case in spite of the different case orders. This indicates that the presented results can be considered as robust to changes of the case order.

We see that the impact clusters differ quite notably for the different impact variables included in the TCS (see Table 6-44). In the cluster with a strong positive impact (74 cases) the majority of respondents stated strong agreement with nearly all research impact measures and most collaboration impact measures. In the cluster with a positive impact (101 cases) the median value is still positive and indicating agreement to most of the statements. In the third and smallest cluster "No impact" (32 cases), however, respondents more often than not *dis*agreed with the statements on research impact and gave neutral answers to the statements on collaboration impact.

Table 6-44: Median values for respondents' agreement to statements on the impact of the selected e-infrastructure by impact cluster

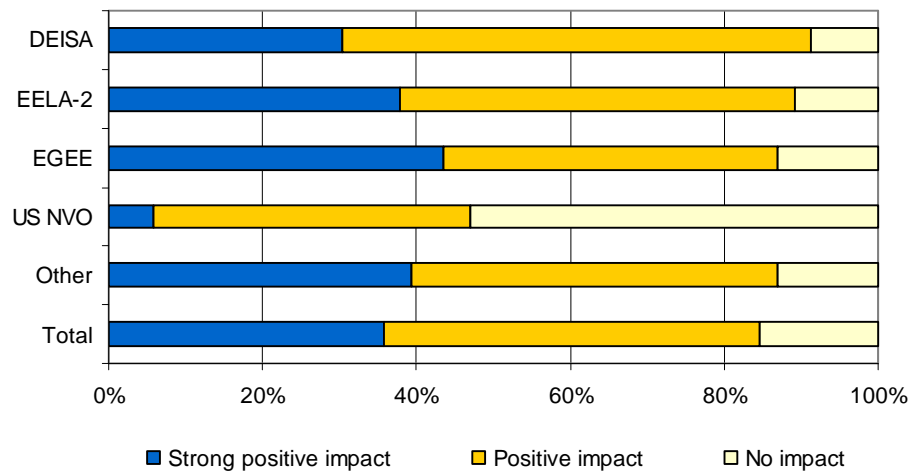|	Impact clusters			Total
	Strong positive impact	Positive impact	No impact	
Research impact measures				
Accomplish research tasks more quickly	5	4	2	4
Produce more research output per year	5	4	2	4
Do research at lower costs	5	4	2	4
Do more accurate, higher quality research	5	4	2	4
Access resources for my research faster or better	5	4	3	4
Produce, process or analyse data faster and better	5	4	2	4
Work on research problems that I could not address before	5	4	2	4
Have more publications or conference proceedings accepted	4	3	2	3
Collaboration impact measures				
I generally collaborate more	5	4	3	4
Geographical range of collaborations has grown	5	4	3	4
More collaboration with colleagues from developing countries	4	3	2	3
More collaboration with commercial firms	3	3	2	3
More collaboration with academic institutions	5	4	3	4
More interdisciplinary collaboration	4	4	3	4

5 = strongly agree, 4 = agree, 3 = neutral, 2 = disagree, 1 = strongly disagree

Differentiating the clusters further by the variables on respondent and project characteristics additional missing values further reduce the case numbers. This should be kept in mind in the reading of the following paragraphs and the results should be taken as indicative but not more. The strongest relationship appears for the different e-infrastructures on which the respondents reported (see Figure 6-37). We see that in particular the NVO users and developers (17 respondents) are often classified in the "No impact" cluster according to their responses. For all the other e-infrastructures this share is considerably lower. The responses

on NVO also affect the results for the groupings by continent and type of e-infrastructure, as most of the NVO respondents are from North-America and NVO was classified as a national, disciplinary, data, and community-driven e-infrastructure. Therefore, for each of these categories we have large shares of respondents in the "No impact" cluster (see table 2-50 in the annex).

Next, we also see that the duration and intensity of involvement in an e-infrastructure correlate with cluster membership and the impact rating: people who have been involved longer time with an e-infrastructure and who are involved with more of its functions and services more strongly agree to the impact statements (= are in the clusters of positive impact) than respondents who have been involved just a short time and use few functions and services (see Figure 6-38 and Figure 6-39).

Figure 6-37: Respondents' by impact cluster and selected e-infrastructure (in %)

Note the low case numbers: DEISA: 23, EELA-2: 37, EGEE: 23, US-NVO: 17, Other: 107, Total: 207

Figure 6-38: Respondents' by impact cluster and involvement in the selected e-infrastructure (in %)
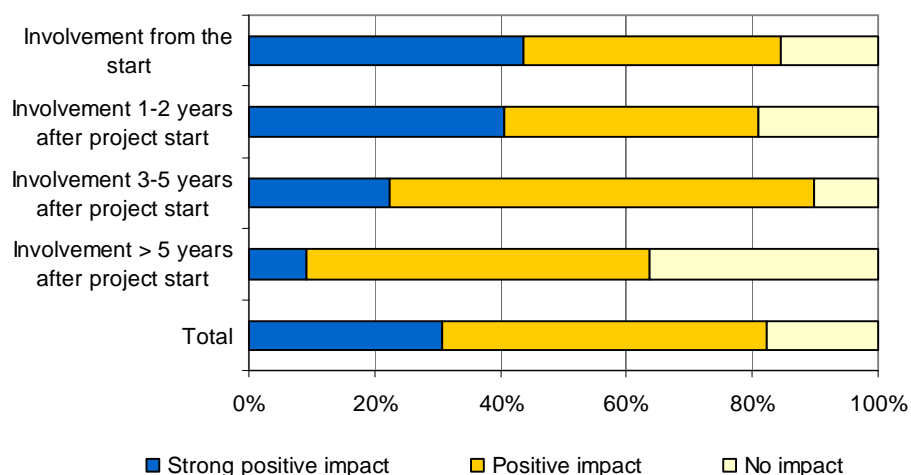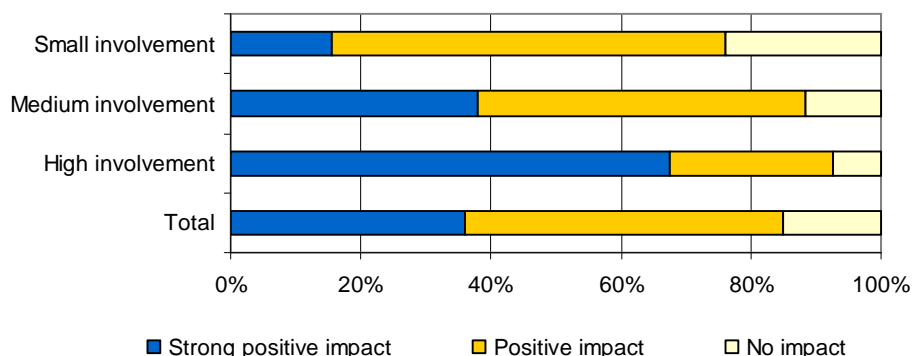
Figure 6-39: Respondents' by impact cluster and degree of involvement in the selected e-infrastructure (in %)

If we differentiate the cluster membership by type of field in which the respondents work, we see that those respondents who were classified as working in novel dynamic collaborative fields and dynamic competitive fields are more often to be found in the two clusters with rather positive impact ratings, whereas those who are classified into established low collaboration fields are more often in the "No impact" cluster. This reconfirms the results which we obtained above for the different field types.

Table 6-45: Respondents' by impact cluster and type of research field (in %)

	Field characteristics			
	Established low collaboration	Novel dynamic collaborative	Dynamic competitive	Total
Cluster "Strong positive impact"	25.7%	50.0%	33.3%	34.8%
Cluster "Positive impact"	54.1%	42.0%	57.4%	51.7%
Cluster "No impact"	20.3%	8.0%	9.3%	13.5%
Total	100.0%	100.0%	100.0%	100.0%
Case numbers	74	50	54	178

6.6   Trends and policy issues

Two sections in the questionnaire asked respondents on their opinions on certain trends and policy issues. This included questions on:

- The adoption and contribution to scientific progress of new resource delivery models such as Software as Service, Cloud Computing or Utility Computing,
- The contribution and role of National Grid Initiatives (NGI) and International Grid Initiatives such as the European Grid Initiative (EGI)
- Recommendations to policy makers (open-ended)

We will present the results on these questions in this section of the report.

6.6.1 Adoption and contribution of new resource delivery models

Roundabout two thirds of the respondents also answered the questions on the expected adoption and contribution to scientific progress of new resource delivery models such as Software as Service, Cloud Computing or Utility Computing in the next five years. A large majority of 80% of those responding here find it likely or very likely that these new developments will spread and have a significant impact in science in the near future (see

Figure 6-40).

Figure 6-40: Respondents' agreement to statements on the role of new resource delivery models (in %)
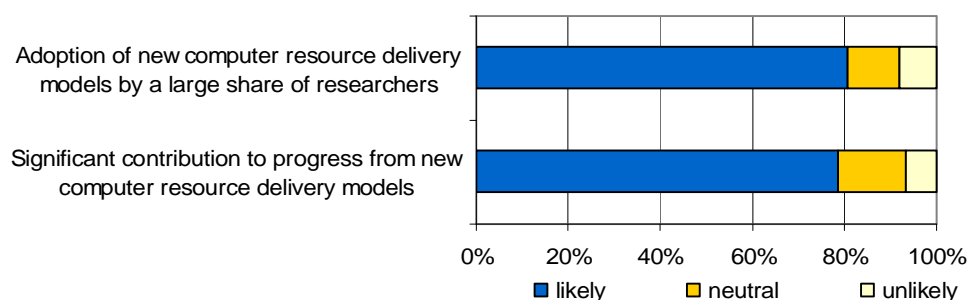
In particular Asians and South-Americans and respondents from the private sector were positive about the strength of these new resource delivery models (see

Table 6-46 and

Table 6-47, note however the low case numbers). Moreover, we see higher expectations among those working in novel, dynamic and competitive fields (see

Table 6-48).

Table 6-46: Respondents agreement to statements on the role of new resource delivery models by continent (in %)

Role of new resource delivery models	Europe	North-America	Latin America	Asia	Other	Total
Adoption of new computer resource delivery models by a large share of researchers	76.7%	73.1%	91.5%	93.8%	100.0%	80.8%
Significant contribution to progress from new computer resource delivery models	75.4%	75.0%	88.1%	93.3%	60.0%	78.8%
Cases (without missing values)	172	26	59	16	3	276

Table 6-47: Respondents agreement to statements on the role of new resource delivery models by primary institutional affiliation (in %)

	Primary institutional affiliation			
Role of new resource delivery models	Academia	Government and international org.	Private sector	Total
Adoption of new computer resource delivery models by a large share of researchers	80.6%	73.5%	91.7%	80.8%
Significant contribution to progress from new	79.9%	65.6%	92.3%	78.8%

| computer resource delivery models				
Cases (without missing values)	216	34	12	262

	Field characteristics			
Role of new resource delivery models	Established low collaboration	Novel dynamic collaborative	Dynamic competitive	Total
Adoption of new computer resource delivery models by a large share of researchers	78.5%	96.7%	75.4%	80.8%
Significant contribution to progress from new computer resource delivery models	77.1%	93.4%	73.0%	78.8%
Cases (without missing values)	93	61	61	215

6.6.2 Contribution and role of National Grid Initiatives and International Grid Initiatives

Before asking the questions on National Grid Initiatives (NGIs) and International Grid Initiatives (IGIs) the respondents were asked whether they are familiar with, involved in the establishment or expecting to benefit from such efforts. Out of 331 respondents who answered this question – the large amount of 76 missing cases is probably due to the fact that the question was asked close to the end of the questionnaire – two thirds (219 respondents) stated that they were involved and one third (112 respondents) that they were not involved. Respondents involved in distributed grid computing infrastructures like EELA-2 or EGEE more often agreed to this statement, respondents on other infrastructures agreed less often (

Figure 6-41). Moreover, developers (75%) agreed more often than research users (59%), and those using e-infrastructures intensively (see section 6.4.4) agreed more often (82%) than those using them just infrequently (61%).

Figure 6-41: Respondents' agreement to being familiar with, involved in the establishment or expecting to benefit from National or International Grid Initiatives (in %)
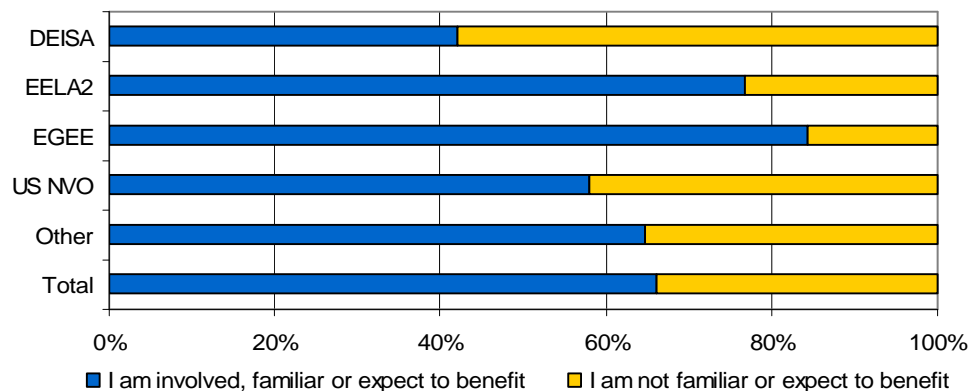
We see also wide agreement from the respondents to statements on the necessity and benefits of NGIs and even more so of IGIs (see

Figure 6-42 and Figure 6-43). In particular statements on their necessity as coordination bodies and for optimising operation and support of distributed computing services are acknowledged by at least four out of five respondents. We see generally a larger percentage of agreement among Latin American and Asian respondents and a smaller percentage among North-American respondents (see Table 6-49, note however the low case numbers).

Figure 6-42: Respondents' agreement to statements on National Grid Initiatives (in %)
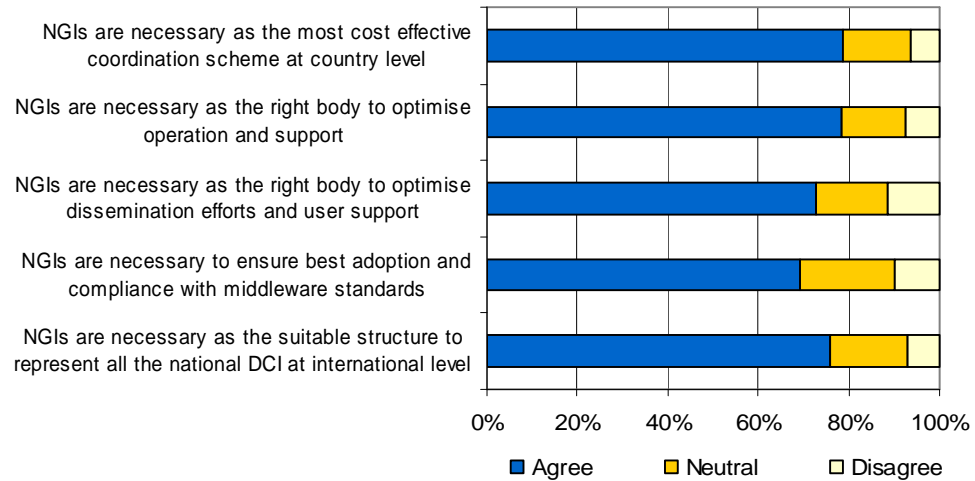

Figure 6-43: Respondents' agreement to statements on International Grid Initiatives (in %)
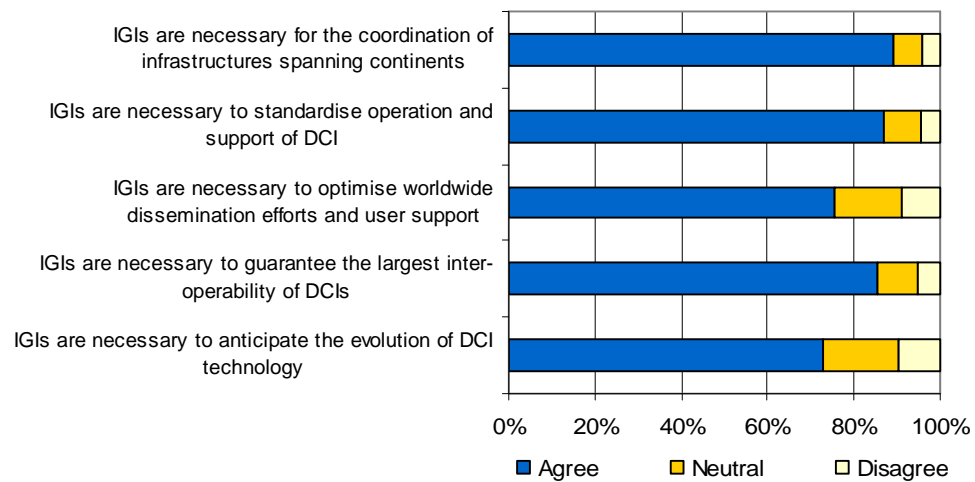
Table 6-49: Respondents' agreement to statements on National and International Grid Initiatives by continent (in %)

	Continent of respondents					
	Europe	North-America	Latin America	Asia	Other	Total
National Grid Initiatives						
NGIs are necessary as the most cost effective coordination scheme at country level	72.5%	57.9%	95.7%	100.0%	100.0%	78.7%
NGIs are necessary as the right body to optimise operation and support	76.9%	42.1%	90.9%	100.0%	100.0%	78.4%
NGIs are necessary as the right body to optimise dissemination efforts and user support	66.7%	47.4%	93.3%	100.0%	66.7%	72.8%
NGIs are necessary to ensure best adoption and compliance with middleware standards	63.0%	63.2%	88.4%	71.4%	66.7%	69.2%
NGIs are necessary as the suitable structure to represent all the national DCI at international level	73.1%	36.8%	93.5%	91.7%	100.0%	75.9%
International Grid Initiatives						
IGIs are necessary for the coordination of infrastructures spanning continents	88.4%	72.2%	95.3%	100.0%	100.0%	89.4%
IGIs are necessary to standardise operation and support of DCI	84.6%	83.3%	90.5%	100.0%	100.0%	86.9%
IGIs are necessary to optimise worldwide dissemination efforts and user support	70.6%	50.0%	95.2%	100.0%	66.7%	75.6%
IGIs are necessary to guarantee the largest inter-operability of DCIs	86.8%	66.7%	90.2%	81.8%	100.0%	85.6%
IGIs are necessary to anticipate the evolution of DCI technology	69.7%	44.4%	88.4%	92.3%	66.7%	73.0%
Case numbers	121	18	43	13	3	198

As we would expect, respondents involved in distributed computing initiatives like EGEE and EELA-2 agree more often to the statements than respondents involved in other types of e-infrastructures (see annex table 2-51). Last but not least we also see that respondents who described their fields as novel, dynamic and collaborative see a larger value of NGIs and IGIs than other respondents (see annex table 2-53). This should not come as a surprise, however, as we find among those a large share of developers of distributed computing (see Table 6-8, page 143 above).

6.6.3 Recommendations to policy makers

Last but not least respondents were also given the opportunity to make recommendations to policy-makers in an open-ended question at the end of the questionnaire. Out of the 407 respondents who answered enough questions in the questionnaire to be included in the analysis 30.2% (123 respondents) made at least one recommendation. The recommendations are shown verbatim in annex table 2-52.

In order to include the recommendations in the analysis, they were coded independently by two different researchers from the team. For the coding the same categories as for the catalysts and barriers were used (see chapter 6.4.5), plus one additional category for awareness-raising measures (see Table 6-50 on examples). We obtain the distribution of

respondents on these categories as shown in Figure 6-44. Most important among the recommendations are those addressing organizational or funding issues which were put forth by more than 10% of the respondents. Any of the other categories was mentioned by just 5 or less percent of the respondents to the survey.

**Table 6-50: Examples for answers on policy recommendations**

Category	Response examples
Access to resources	– Make it institutionally and ubiquitously available as if it were the telephone, mobile phone, electricity, or air we breathe. – Policy maker should push for a flexible and open GRID access to a variety of computational resources, both HPC and High Throughput oriented, stressing the – by providing tools allowing reallocation of resources for a given group of scientists on demand
Organizational	– Support software applications design and provide career and career plans for whole generations of developers rather that living from hand to mouth on short term contracts well into their forties and fifties. – Provide clear national strategy around einfrastructure, outlining drivers and strongly connected research communities, and lead agencies and organisations; Facilitate the aggregation of research agendas towards developing and sustaining einfrastructure developments – A grid services brokerage company is required. Infrastructure use grants could be given.
Technical capabilities	– Focus on alternatives to "Grid", especially on web service standards. These have proved far more effective in promoting interoperability and integration of data-dependent services. – Creating standards and study previous cases such as the Internet evolution
Ease of use	– By paying more attention to the needs of end users and less to the claims of those promoting technologies – Improve the simplicity and accessibility of the user interface layer. – participation should be easier and encouraging
Funding-related	– 1) by rewarding and funding the development and evaluation of production-ready technology; 2) by providing stable funding for user support and training – By making clear decisions on sustained funding, not just funding projects. Basic for advancing e-infrastructures is the long-term maintenance.
Training-related	– Making the e-infrastructure familiar for more people, with workshops for the older and introducing or building e-infrastructure in public schools, for the children. Also teachers should enhance their knowledge to keep on with new technologies and teaching strategies. – In countries where the technology is not widespread, I believe that most of the effort should be placed in training people to use new scientific methodologies that can profit from the massive amounts of computing and storage available and that can be put together thanks to these e-Infrastructures.
Awareness raising	– There must be identified applications that will create impact in the country's economic value, to make policy makers in the national level to support and sustain the investment and advance use of e-infrastructure. In developing countries, immediate problems have priority. – Promoting through events and tutorials the use of grid, at least once a year in all the involved countries. – by showing good examples (pilot projects); by making it easy and relatively cheap to access the e-Infrastructure; by taking away the (emotional and political) barriers – funding and articulation of a global vision explaining goals, plans and motivations
Other recom-mendations	– Just remembering that, nowadays, e-Infrastructures are becoming a necessary condition for development, i.e. for independence, in a e-Society.

Figure 6-44: Respondents' recommendations to policy makers (in % of all respondents)
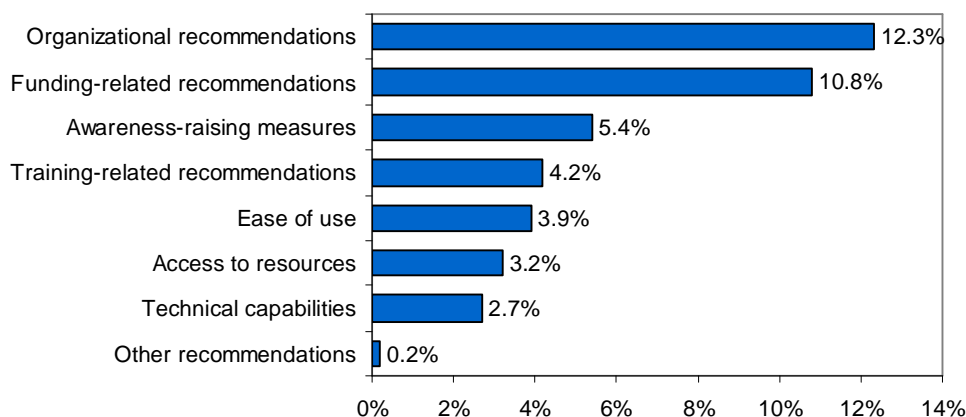
The number of responses on policy recommendations is small and the following statements should just be taken as intimations of issues that need to be discussed with the stakeholders in more detail before any measures are designed. Differentiating the recommendations by e-infrastructure, we see that respondents on DEISA point more often than we would expect to measures related to resource access (see annex table 2-54). Next, we see in Figure 6-45 that funding recommendations are given nearly twice as often by those who have been involved in the e-infrastructure for considerable time. We would interpret this as a hint to the severity of funding problems: those who recently received funding for becoming involved in an e-Infrastructure may tend to lose sight of funding restrictions, but those involved for a longer time perceive it as a much more permanent threat.

Figure 6-45: Respondents' recommendations to policy makers by start of involvement in the selected e-infrastructure (in % of all respondents)
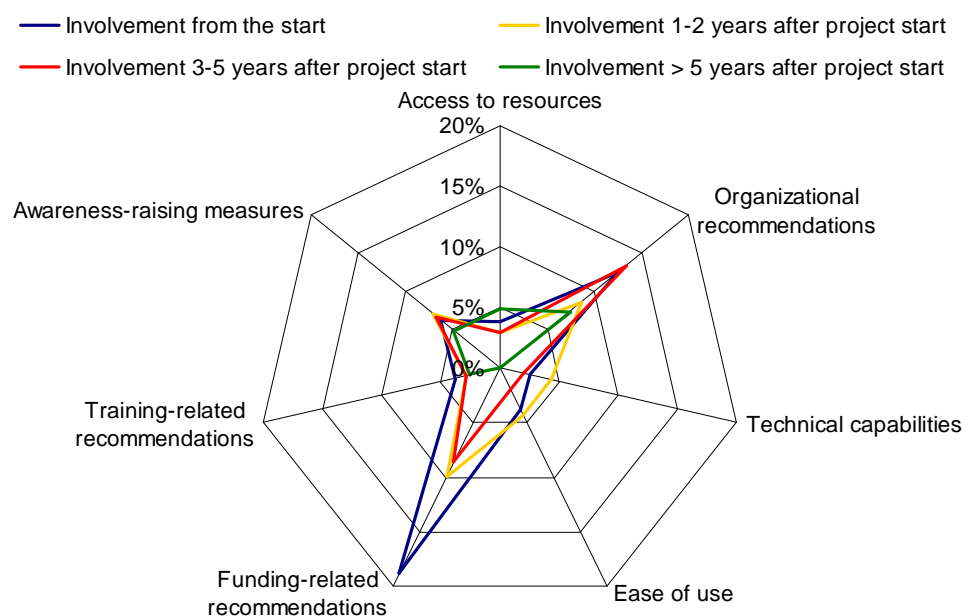
Another interesting finding is that among the non-research users and those using e-infrastructures for professional work twice as many (10.8%) as the average (5.4%) ask for awareness-raising measures. It seems that a lack of awareness about the usefulness of e-infrastructure is even more an issue in non-academic areas which potentially benefit from it than inside academia. Another distinction can be made according to the intensity of

involvement in selected e-infrastructure, i.e. the number of services that a respondent uses and the frequency of use (see chapter 6.4.4). We find that more intensive involvement correlates also with more recommendations in nearly all areas. We may interpret this as a measure of concern: those who work with e-infrastructures a lot care more for their sustainability and adequate measures for achieving this.

Table 6-51: Respondents' recommendations to policy makers by intensity of involvement in the selected e-Infrastructure (in % of all respondents)

	Intensity of involvement with services and resources of selected e-infrastructure			Total
	Small involvement	Medium involvement	High involvement	
Access to resources	1.9%	4.0%	6.6%	3.2%
Organizational recommendations	11.9%	12.7%	18.0%	12.3%
Technical capabilities	1.3%	3.3%	6.6%	2.7%
Ease of use	4.4%	2.7%	8.2%	3.9%
Funding-related recommendations	9.4%	12.7%	14.8%	10.8%
Training-related recommendations	4.4%	3.3%	8.2%	4.2%
Awareness-raising measures	8.8%	3.3%	4.9%	5.4%
Other recommendations	0.6%	0.0%	0.0%	0.2%
Cases	160	150	61	371

6.7   Survey summary

Role of e-infrastructures in research communities

In order to shed further light on the role of e-infrastructures in research communities we developed an online questionnaire that asked about the involvement of different groups of stakeholders (research users, non-research users, developers) in e-infrastructures and the impact of this involvement. A large part of the questions was oriented to a particular e-infrastructure which the respondents could select from a drop-down list (with the possibility to add a different non-listed e-infrastructure). It was not possible to prepare a sample frame that would have permitted representative sampling from a survey population. Instead we opted for a snowball approach asking contact persons of selected e-infrastructures to distribute a link to the questionnaire to their constituencies and communities. Thus the survey responses can only be used for illustrative and descriptive purposes.

Virtual research communities. The respondents to the survey were asked about the amount, geographic distribution and institutional affiliations of peers in their fields who use or participate in the selected infrastructure in the same way as they do. According to the responses these virtual research communities are to two thirds rather small communities of 100 people or less (including those who were not able to estimate the number of peers). Only 8.5% of the respondents pointed to "big science" communities with more than 500 members. The geographical scope of these communities is in one third of the responses national and two thirds international. Involvement of non-academics (government, international organisations, private business) is seen by 60% of the respondents; in most cases as mixed communities together with academics.

Involvement in e-infrastructures. From a funding perspective of e-infrastructure involvement it is interesting to note that the largest share of respondents (44%) was funded through national sources. International funding and institutional funding from the respondents own organizations were each named as major funding sources by one quarter of the respondents.

The role of private businesses is negligible. Grid computing was the most widely used service in our set of responses; data-related tools and services come next in line. Heavy users of e-infrastructure are rather uncommon; most respondents are only involved to small extent. Still, the most important driver to becoming involved with an e-infrastructure was access to resources followed by organizational catalysts and (enhanced) technical capabilities through e-infrastructure. Making new resources accessible to research and good accessibility of the e-Infrastructure itself are therefore important inputs to gaining wider use. Among the barriers (low) technical capabilities, organizational barriers and low usability are of more or less equal importance. This reinforces what we have seen in the case studies: socio-cultural barriers are at least as important as technical limitations in the adoption process and adapting an e-Infrastructure to its users' abilities will be key for establishing it successfully. Contrary to our expectation a lack of funding does not belong to the most important barriers, neither in the case studies nor in the survey responses. Most of the people involved in an e-infrastructure also try to involve others, in particular through talking to their local peers and giving talks and making demonstrations. Developers are particularly active in this area.

Impact of e-infrastructure. It is striking to see that most respondents have become quite attached to the e-infrastructure on which they reported: 85% evaluate it as important for their research or work, and nearly 70% would see their research programmes impaired without e-infrastructure. However, we should bear in mind that the respondents are not a random selection of researchers but rather an e-science-savvy sample. They also give predominantly positive ratings of the impact of the e-infrastructure on their (research) productivity and the quality of their work results. For instance 75% agree that it has helped them to work on problems which they could not address before; the same percentage agrees to accomplish research tasks faster; still 65% agree that they do more accurate and higher quality research, have become more productive or save costs. A positive impact is also seen on research collaboration which is intensified through involvement in an e-infrastructure.

Trends and policy issues. A large majority of 80% of those responding find it likely or very likely that new resource delivery models such as Software as Service, Cloud Computing or Utility Computing will spread and have a significant impact in science in the next five years. We see also wide agreement from the respondents to statements on the necessity and benefits of National and international Grid Initiatives. In particular statements on their necessity as coordination bodies and for optimising operation and support of distributed computing services are acknowledged by at least four out of five respondents.

Roughly 30% of the respondents also made policy recommendations. Most important among the recommendations are those addressing organizational or funding issues which were put forth by more than 10% of the respondents.

In particular the measures of involvement and impact co-vary with several characteristics of the respondents and properties of the projects.

Patterns according to respondents' field, geographical provenance, affiliation, activities and time of adoption

Fields. The importance of (research) field conventions and practices for the use of the internet and other information and communication technologies is an undisputed result of previous analyses. Hence, it does not come as a surprise that field characteristics shape e-infrastructure involvement (and its impact) in our dataset. Case numbers are too few to interpret the differences between research domains, fields of work or areas of development activities in which the respondents work. Instead we clustered responses in regard to patterns of collaboration, competition and dynamics of the respondents' main fields separating three field clusters:

- Established low collaboration (ELC) fields in which collaboration is still the dominant mode of work, but less so than in other fields; respondents agreed more often than in the other clusters that work is typically done by individuals.

- Competition and the change of research problems, paradigms, approaches or methods are denoted as low (respondents frequently work in chemical and material sciences, computer and information sciences, social sciences, humanities, and application development).

- Novel dynamic collaborative (NDC) fields stick out by the fact that they are described as not yet established and with a comparatively fast pace of change of research problems, paradigms etc. In addition, collaboration is deemed as essential for achieving progress in these fields and work is more often done in large-scale collaborations of more than ten people (often respondents form earth and other natural sciences, physics).

- Dynamic competitive (DYC) fields is a cluster characterized by the high intensity of competition in combination with a fast pace of change. The importance of collaboration is average, but more in small than in large groups of collaborators (often engineers, physicists, supercomputing and grid computing developers).

Depending on their type of field respondents also show some notable variation in their e-infrastructure involvement. In particular, we see that a "one size fits all" approach is not realistic at this level of infrastructure, and that users differ in regard to their service and resource needs. Respondents from NDC fields use distributed computing and collaboration tools more than respondents from the other two categories of fields. Data-related tools and online storage are most often used in ELC fields (i.e. those with low levels of competition and dynamics). In DYC fields several services are less widely used – exceptions are supercomputing, simulation applications, and the respondents' own applications ported on the e-infrastructure. This corroborates a result from the multi-case comparison, that e-science is more frequently related to working on grand scientific challenges than to dealing with fast changing problems, paradigms and approaches.

Respondents from ELC fields also differ from the more dynamic fields as they rely more on funding from their own organization, are less sensitive to the availability of e-infrastructures, see less often a positive impact of e-infrastructures on research productivity and quality, and experience less often growing collaboration networks thanks to e-infrastructure. In sum, they seem to be the less engaged users.

Geographical provenance. Respondents to the survey were also classified according to a) the continent and b) the development status of the country (according to the OECD DAC list). According to the sampling approach we find in particular respondents from Latin America in this group (more than 75% of the respondents).

Ad a) Continental patterns. The continental differences are diverse and it is not possible to sketch a clear picture. European respondents to the survey are more involved with European and purely academic communities, respondents from the US with global and mixed (academic and non-academic) communities. As to be expected, the funding patterns of the e-infrastructure involvement also differ: North-American respondents rely primarily on national governmental sources and Europeans get funding to similar percentages from national and international (EU) programmes. Respondents from North-America and Asia are more intensive users than respondents from Europe or Latin America (measured as the number and frequency of services and resources used). Last but not least, catalysts and barriers also vary by continent: North-Americans mentioned access to resources most often among the catalysts, Latin Americans organization-related issues and Asians technical capabilities; Europeans point to the three types of catalysts equally often. Organizational, training- and funding-related barriers were particularly important among Latin American respondents. Among North-Americans technical capabilities, usability and organizational barriers were the most important. European and Asian respondents mention most barriers less often than respondents from the American continent.

Ad b) Patterns by development status of the country. The financial and technical situation of universities and research organisations in developing countries is worse than in developed

countries. This shapes their involvement with e-infrastructure and financial constraints constitute the most important barrier (though funding from international sources has filled some gaps). However, e-infrastructures contribute to reducing technical constraints in developing countries and they are rated more often positively than in developed countries when it comes to their effects on research productivity and quality. A positive impact of e-infrastructure on collaboration is more frequently felt in the form of South-South collaboration, with new collaborators from other developing countries, than in the form of North-South collaboration.

Affiliation and activities. More than 80% of the survey respondents work in the academic sector, 13% in government agencies and international organizations, and 6% in firms. Taking the main activities as criteria, we can classify 30% as scholars, 40% as researchers, 20% as professionals and 10% as administrators.

Respondents' affiliations and activity profiles correlate with the resources they use or develop: Respondents from governments and international organizations less often work with computing and simulation resources, and more often with data-related tools and online storage. Along the same lines professionals, i.e. respondents who spend a large share of their time on professional (and not academic) work, are less involved with computing resources – in particular supercomputing – and the services and resources which support analytical tasks (data analysis tools, simulations, remote access to research instruments, or own applications ported on the e-infrastructure). In contrast, they particularly often work with data management tools and data collections. Service and resource use patterns also vary in the dataset: those using e-infrastructure for research require a broader set of applications without being able to spend too much time with each; other users require only very few applications which they then use more frequently.

Time of adoption. We constructed two measures for the time of adoption: a) Calendar year; the large majority of respondents (70%) became involved with the e-infrastructure on which they reported in 2005 or later. b) Relating the first involvement with the start year of the project we get roundabout 50:50 shares of those who were involved from the start till 2 years after the start and those who became involved at later stages. The responses to several questions co-varied with these two different measures of the time of adoption.

Ad a) Calendar year. We find that the longer respondents are involved with an e-infrastructure, the more intensively they work with it, the more they would be affected if the infrastructure or an appropriate substitute were not available, and the more positively they evaluate the benefits and impact of an e-infrastructure for their research or work in general. We can offer two explanations for this: benefits take some time to materialize and less satisfied users discontinue use after some time. Both mechanisms may explain higher satisfaction among the early than the more recent adopters.

Ad b) Time period after project start. Taking the second indicator of involvement we first see an interesting pattern relating to whether respondents are involved as research users or developers: Research users are more often latecomers to the projects. In contrast, developers have often been involved from early on in the project. This result points to a rather traditional model of technological innovation, in which most users are involved at late stages of development – probably not the best way of ensuring good usability and frictionless matching of services and users' needs. Another interesting result is obtained for the funding sources: those who have become involved in an infrastructure in its early phases received funding more often from the EU or other international funding agencies, whereas those who became involved later were funded more often by national agencies and their own institutions. From this we may gather that initial international funding has an important enabling function.

Patterns according to characteristics of the projects on which respondents reported

The respondents were asked to select a project with which they have been involved and on which they wanted to report. Due to the considerable breadth of this list the frequencies are rather small and we only report on four of the selected projects in detail. Furthermore, we grouped the projects according to their geographical scope (national versus international), disciplinary breadth (disciplinary versus multi-disciplinary), main type of service (computing versus data) and the drivers of the projects in its early phases (developer- versus community-driven). The responses on involvement and impact also vary for these groupings.

Selected projects. The numbers of responses permitted us to single out and compare four e-infrastructures (DEISA, EELA-2, EGEE and US NVO). We find that they are important correlates of the size, extension and affiliation of the virtual research communities. Participants of DEISA point to small European communities working with DEISA in the same way as they do. EELA-2 participants are also aware of small communities; however, as EELA-2 specifically fosters collaboration with South-America, the communities also reach out to the American continent. Respondents on EGEE point to large and global communities of peers with similar interests. In the same vein, participants to US NVO see a large community of peers involved in the infrastructure, but they come mainly from the US and there is a strong non-academic component. All in all, we see that each infrastructure caters to a set of clients with specific characteristics. Another distinction refers to funding: DEISA and NVO participants are most often funded by their national governmental funding agencies, EELA-2 participants by their own institutions, and EGEE participants by international governmental funding agencies (e.g. EU).

Geographical scope. It will come as little surprise that people involved in national e-infrastructures need to rely more on national funding than those involved in international infrastructures. National infrastructures also cater more to nationally-bounded communities than international e-infrastructures. Another distinction refers to the services and resources used or developed: respondents on an international e-infrastructure more often point to computing services and related tools, whereas respondents on national e-infrastructures point to data-related services. When it comes to assessing usability and effects, international e-infrastructures fare also better than national e-infrastructures in regard to ease of use and help provided to users, their overall importance for research, impact on research productivity and quality, and impact on collaboration networks. One notable exception are collaborations with commercial firms, which are not influenced very much by e-infrastructure in general and where we do not find a difference between international and national e-infrastructures.

Disciplinary breadth. An important feature of disciplinary e-infrastructures is that they provide more data-related services than computing services, whereas for multidisciplinary e-infrastructures it is the other way around. The effects of multidisciplinary e-infrastructures on research are rated more positively than those of disciplinary e-infrastructures except for collaboration with firms which is more often achieved in a disciplinary setting.

Type of service. Classifying the e-infrastructures into computing respectively data infrastructures also reveals some interesting patterns. Respondents involved in computing infrastructures point out more often than those involved in data infrastructures, that only few peers from the same field work with the e-infrastructure. This suggests that data infrastructures tend to involve a larger number of people in the same manner and with similar needs, whereas computing infrastructures rather serve small groups in different ways. Moreover, infrastructures offering computing services cater more to academic communities, non-academic communities are not that important. Data infrastructures, on the other hand, more often deal with mixed, academic and non-academic communities. The funding structures of the users' involvement with the e-infrastructure – not the e-infrastructures themselves – also vary by service provided: Those involved in computing infrastructures depend to nearly half on national government funding whereas the participants in data infrastructures rely to a larger degree on institutional funding. Computing e-infrastructures

were more often rated as important and received better ratings on the impact on research than data infrastructures. However, nearly three times as many respondents found it more challenging to get the needed services out of a computing infrastructure than a data infrastructure. However, this did not appear to be a problem of usability but rather of matching needs and services.

Drivers. The patterns that we obtain resemble somewhat the pattern for the previous classification on type of service though the overlap in the classification is only 65%. Respondents on developer-driven infrastructures also point less often to larger communities of peers involved in the e-infrastructure in the same way than respondents on community-driven infrastructures. Non-academic community participation is more important for community-driven e-infrastructures. Developer-driven e-infrastructures were more often rated as important and received better ratings on the impact on research than community-driven infrastructures. We find the same problem of matching needs and services as stated in the previous paragraph for computing infrastructures. Last but not least , 25% of the respondents on community-driven e-infrastructures agreed to rising collaboration with the private sector compared to less than 15% on developer-driven infrastructures.

References

Atkins, D. E., Droegemeier, K. K., Feldman, S. I., Garcia-Molina, H., Klein, M. L., Messerschmitt, D. G., et al. (2003). Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure: National Science Foundation.

AVROSS - Accelerating Transition to Virtual Research Organisation in Social Science (2007a). M2 Interim Report 1. Unpublished deliverable to the European Commission, DG Information Society & Media.

AVROSS - Accelerating Transition to Virtual Research Organisation in Social Science (2007b). M3 Interim Report 2. Unpublished deliverable to the European Commission, DG Information Society & Media.

Axelsson, A.-S. and Schroeder, R. (forthcoming 2009). 'Making it Open and Keeping it Safe: e-Enabled Datasharing in Sweden', Acta Sociologica (previous version in Proceedings of the Third International Conference on e-Social Science, October 7-9, 2007, Ann Arbor, Michigan, US.

Barjak, F., Lane, J., Kertcher, Z., Poschen, M., Procter, R., & Robinson, S. (2009). Case Studies of e-Infrastructure Adoption. Social Science Computer Review, 27(4), pp. 583-600.

Bauer, M. & Gaskell, G. (2002). 'The biotechnology movement', in Martin Bauer and George Gaskell (eds.), Biotechnology: The Making of a Global Controversy. Cambridge: Cambridge University Press, 379-404.

Bégin, M.-E. (2008). *An EGEE comparative Study: Grids and Clouds - Revolution or Evolution.* https://edms.cern.ch/file/925013/4/EGEE-Grid-Cloud-v1_2.pdf.Bel, N., S. Bel, et al. (2008). "The CLARIN project: a scientific research infrastructure for the humanities and social sciences." *DIGITHUM* 10.

Berman, F., & Brady, H. (2005). Final Report: NSF SBE-CISE Workshop on Cyberinfrastructure and the Social Sciences. Retrieved December 7, 2006 from http://vis.sdsc.edu/sbe/reports/SBE-CISE-FINAL.pdf

Blanke, T. and M. Hedges (2008). Providing linked-up access to Cultural Heritage Data. Proceedings of the ECDL 2008 Workshop on Information Access to Cultural Heritage. Aarhus, Denmark.

Blanke, T. H., M.; Dunn, S.; (2009). "Arts and humanities e-science - Current practices and future challenges." *Future Generation Computer Systems 25*: 474-80.

Borgman, C. (2007). Scholarship in the Digital Age. Cambridge MA: MIT Press.

Bosa, K. & Schreiner, W. *Report on Experiments with Globus 4 and gLite, Technical Report of the Research Institute for Symbolic Computation (RISC) at Johannes Kepler University Linz.* http://www.risc.uni-linz.ac.at/publications/download/risc_3400/AG-D4-1-2007_2.pdf [accessed 20 July 2009]

Brasileiro, F., Duarte, A., Carvalho, D., Barbera, R. & Scardaci, D. (2008). An Approach for the Co-existence of Service and Opportunistic Grids: The EELA-2 Case. II Latin-American Grid Workshop (LAGrid 2008), Campo Grande, Brazil, November 2008.

Broeder, D., D. Nathan, et al. (2008). Building a federation of Language Resource Repositories; the DAM-LR project and its continuation within CLARIN. 6th International Conference on Language Resources and Evaluation (LREC 2008). Marrakech.

Burk, D. (2007). Intellectual property in the context of e-science. Journal of Computer-Mediated Communication, 12(2), article 13. http://jcmc.indiana.edu/vol12/issue2/burk.html

Bush, V. (1996). As We May Think. *Interactions, 3*, 35-46.

Candela, L.; Castelli, D.; Pagano, P. gCube: A Service-Oriented Application Framework on the Grid. In: ERCIM News 72 Special theme: The Future Web, January 2008, pp 48-49.

Castelli, D. e-Infrastructures designed for demanding science. In: *eStrategies Europe* Vol 2 No 4 on "Europe's new flagship for innovation", British Publishers, November 2008.

Castelli, D.; Michel, J. D4Science - Developing Virtual Research Environments. In: ERCIM NEWS 74, July 2008. p. 8-9.

CLARIN (2009). CLARIN Newsletter 5.

CLARIN: Survey of liaisons with other European projects and Initiatives (2009), Document ID: D5C-1, http://www.clarin.eu/system/files/private/D5C-1OutcomeLiaison.pdf, accessed 3 July 2009

Crane, G., B. Fuchs, et al. (2007). *The Humanities in a Global e-Infrastructure: A Web-Services Shopping List.*

Cummings, J. N., & Kiesler, S. (2005). Collaborative Research across Disciplinary and Organizational Boundaries. *Social Studies of Science, 35(*5), 703-722.

DARIAH (2008). DARIAH Newsletter 1.

DARIAH (2009). DARIAH Newsletter 2.

David, P. A., & Spence, M. (2003). Towards institutional infrastructures for e-science: The scope of the challenge. OII Research Report No. 2. Retrieved June 5, 2009 from http://www.oii.ox.ac.uk/resources/publications/RR2.pdf.

David, Paul; den Besten, Matthijs & Schroeder, Ralph. (2006). 'How Open is e-Science?' Proceedings of IEEE e-Science, Amsterdam, December 4-6.

Daw, M. et al., (2007). 'Developing an e-Infrastructure for Social Science', in Proceedings of the Third International Conference on e-Social Science, October 7-9, 2007, Ann Arbor, Michigan, US.

Demographic Database (2009). http://www.ddb.umu.se/index_eng.html. Accessed on March 18, 2009.

Den Besten, M., Schroeder, R. and Thomas, A. (2009). 'Life Science Research and Drug Discovery at the Turn of the 21st Century: The Experience of SwissBioGrid', forthcoming in Discovery and Collaboration.

DISC (2009). http://www.disc.vr.se/ Accessed on March 18, 2009.

Doorn, P. (2007). Data sharing Infrastructures in the ESFRI Roadmap: A Perspective from the Social Sciences and Humanities, Università di Padova, Palazzo del Bo, Via VIII Febbraio 2, Padova, Italy http://www.aepic.it/conf/

Driver: Architectural Specification. Deliverable code: DRIVER-03-D2.0-4.7

Driver: DRIVER-II Factsheet. http://www.driver-repository.eu/PublicDocs/FACT_SHEET_ I3_driver_ii.pdf

Driver: European Network Plan. Deliverable D2.1

Driver: Functional Specification. Deliverable code: DRIVER-03-D1.0-1.1

Driver: Organisational Models. Deliverable 2.3b. Deliverable code: DRIVER-02-D2.3b

Drori, G., Meyer, J., Ramirez, F. & Schofer, E. (2003): Science in the Modern World Polity: Institutionalization and Globalization. Stanford: Stanford University Press.

Dutra, I, Gutierrez, J. M., Hernandez, V., Marechal, B., Mayo, R. & Nellen, L. (2007). e-Science Applications in the EELA Project. IST-Africa 2007 Conference Proceedings, Paul Cunningham and Miriam Cunningham (Eds), IIMC International Information Management Corporation, 2007. Retrieved 16.03.2009 from: http://documents.eu-eela.org/getfile.py?recid=711.

Dutton, W. H., & Meyer, E. T. (2008, 18-20 June). 'The Diffusion of e-Research: The Use and Non-Use of Advances in Information and Communication Technologies across the Social Sciences' presented at the *4th International Conference on e-Social Science*, Manchester, UK.

Eccles, K. et al. (2009). 'The Future of e-Research Infrastructures', Proceedings of NCeSS International Conference on e-Social Science, Cologne, June 24-26.

Edge, D. (1995). The social shaping of technology. In N. Heap, R. Thomas, G. Einon, R. Mason & H. Mackay (Eds.), *Information technology and society: a reader* (1 ed., pp. 14-32). London, Thousand Oaks, New Delhi: Sage.

Edwards, P.N., Jackson, S.J., Bowker, G.C. & Knobel, C.P. (2007) Understanding infrastructure: Dynamics, tensions, and design. Report of a workshop on 'History & theory of infrastructure: Lessons for new scientific cyberinfrastructure'. http://www.si.umich.edu/InfrastructureWorkshop/documents/UnderstandingInfrastructure2007.pdf

e-Science Directors' Forum Strategy Working Group (2009), 'Century of Information Research (CIR): A Strategy for Research and Innovation in the Century of Information', Prometheus, vol.27, no.1, pp.27-45.

Etzkowitz, H., & Leydesdorff, L. (2000). The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university-industry-government relations. Research Policy, 29(2), 109-123.

European Commission. (2003). Third European Report on Science & Technology Indicators 2003 - Towards a knowledge-based economy. Brussels: European Commission.

Finholt, T. A., Rocco, E., Bree, D., Jain, N., & Herbsleb, J. D. (1998). NotMeeting: A field trial of NetMeeting in a geographically distributed organization. SIGGROUP Bulletin, 20(1), 66-69.

Frischer, B., Unsworth, J., Dwyer, A., Jones, A., Lancaster, L., Rockwell, G. & Rosenzweig, R. (2006) Summit on digital tools for the humanities: Report on summit accomplishments. http://www.iath.virginia.edu/dtsummit/SummitText.pdf

Fry, J., Den Besten, M. and Schroeder, R. (2009). 'Open Science in e-Science: Contingency or Policy?' with Journal of Documentation, vol.65, no.1, pp.6-32.

Funtowicz, S., & Ravetz, J. (1993). Science in the post-normal age. Futures, 25, 739-756.

Gardner, D., Toga, A.W., Ascoli, G.A., Beatty, J., Brinkley, J.F., Dale, A.M., Fox, P.T., Gardner, E.P., George, J.S., Goddard, N., Harris, K.M., Herskovits, E.H., Hines, M., Jacobs, G.A., Jacobs, R.E., Jones, E.G., Kennedy, D.N., Kimberg, D.Y., Mazziotta, J.C., Miller, P., Mori, S., Mountain, D.C., Reiss, A.L., Rosen, G.D., Rottenberg, D.A. Shepherd, G.M., Smalheiser, N.R., Smith, K.P., Strachan, T., Van Essen, D.C., Williams, R.W. and Wong S.T.C. (2003). Towards Effective and Rewarding Data Sharing, *Neuroinformatics*, vol. 1, pp. 289-295.

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (1994). *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies.* (1 ed.). London; Thousand Oaks; New Delhi: Sage Publ.

Gibbons, S. M. C, Kaye, J., Smart, A., Heeney, C. and Parker, M. (2007). Governing Genetic Databases: Challenges Facing Research Regulation and Practice, *Journal of Law and Society*, vol. 34, no. 2, pp. 163-189.

Gläser, J. (2003). What internet use does and does not change in scientific communities. *Science Studies, 16*(1), 38-51.

Grimme, C., Langhammer, T., Papaspyrou, A., Schintke, F.: "Negotiation-based Choreography of Data-intensive Applications in the C3-Grid Project", German e-Science Conference 2007 Baden-Baden, May 2007

Guidetti, V. Earth Science as an e-Infrastructures Application: Practices at the ESA. In: *Zero-In* second edition eMagazine, Zero-In Issue 2, 2009. http://www.beliefproject.org/zero-in/zero-in-second-edition-emagazine/earth-science-as-an-e-infrastructures-application-practices-at-the-esa.

Hermerén, G. (1986). *Kunskapens pris: forskningsetiska problem och principer i humaniora och samhällsvetenskap* [The Price of Knowledge: Problems in Research Ethics and Principles within Humanities and Social Sciences], HSFR, Stockholm.

Hey, A., & Trefethen, A. 2003. The data deluge: An e-science perspective. In F. Berman, G. C. Fox, & Anthony Hey (Eds.), *Grid Computing: Making the Global Infrastructure a Reality* (pp. 809-824). Chichester, UK: John Wiley & Sons, Ltd.

Hilgartner, S. (1995). Biomolecular database - new communication regimes for biology? *Science communication, 17*(2), 240-263.

Hughes, T. P. (1994). 'Technological momentum', in L. Marx & M. R. Smith (Eds.), *Does technology drive history? The dilemma of technological determinism* (pp. 101-113). Cambridge, MA: MIT Press.

Hughes, T. P. (1987). 'The Evolution of Large Technological Systems.' In Wiebe Bijker, Thomas Hughes and Trevor Pinch (eds.), *The Social Construction of Technological Systems.* Cambridge, MA: MIT Press, 51-82.

Jackson, S. J., Edwards, P. N., Bowker, G. C., & Knobel, C. P. (2007). Understanding Infrastructure: History, Heuristics, and Cyberinfrastructure Policy [Electronic Version]. First Monday, 12 from http://www.firstmonday.org/issues/issue12_6/jackson/index.html.

Jones, B. (2009). EGEE-III Status. EGEE-III First Review (CERN), 24-25 June 2009. http://indico.cern.ch/getFile.py/access?contribId=2&resId=0&materialId=slides&confId=53198 [accessed 15 October 2009]

Justitiedepartementet. (2007). Skyddet för den personliga integriteten. Kartläggning och analys. [The Protection of the Personal Integrity. Mapping and analysis]. *Statens offentliga utredningar*, SOU 2007:22 [The Swedish Government Official Reports, SOU 2007:22].

Kemps-Snijders, M., A. Klassmann, et al. (2008). Exploring and Enriching a Language Resource Archive via the Web. 6th International Conference on Language Resources and Evaluation (LREC 2008). Marrakech.

Kindermann, S., Ronneberger, K. (2006). "Grid technology projects at DKRZ" *TerraFLOPS, Newsletter of DKRZ and M&D.* Ausgabe April 2006 http://www.dkrz.de/pdf/tf/TerraFlops_7_04.pdf?dkrzsid=34f0a0d0934836e91eba2c05a92f0234 (accessed May 4th 2009)

Kindermann, S., Stockhause, M., & Ronneberger, K., (2007): "Intelligent Data Networking for the Earth System Science Community" German e-Science Conference 2007 Baden-Baden, May 2007 http://edoc.mpg.de/get.epl?fid=36067&did=316512&ver=0, (accessed May 4th 2009)

Kindermann, S. (2006). "Climate Data Analysis and Grid Infrastructures: Experiences and Perspectives" Grid-Enabling Legacy Applications and Supporting End Users Workshop (GELA), Paris, France, 20 June 2006: within the framework of the 15th IEEE International Symposium on High Performance Distributed Computing

Kindermann, S. (2006) "Klimadaten und Grid-Infrastrukturen" Jahrbuch der Max-Planck-Gesellschaft 2006 http://www.mpg.de/bilderBerichteDokumente/dokumentation/jahrbuch/2006/dkrz/forschungsSchwerpunkt/pdf.pdf (accessed May 4th 2009)

Knorr-Cetina, K. (1999). *Epistemic Cultures: How the Sciences Make Knowledge.* Cambridge, MA: Harvard University Press.

Krawer, S. (2009) 'CLARIN-EU: Where do we stand?' http://www.hum.uu.nl/clarin-nl/events/kickoff/presentations/Krauwer-CLARIN-NL-launch.ppt, accessed 3 July 2009

Lamanna, M. (2006). The LHC computing grid project at CERN, *Nuclear Instruments and Methods in Physics Research,* A 534, 1-6.

Lederer, H. DEISA2: Supporting and developing a European high-performance computing ecosystem, Journal of Physics: Conference Series 125 (2008).

Lederer, H.; Alessandrini, V. (2008). DEISA: Enabling Cooperative Extreme Computing in Europe, Parallel Computing: Architectures, Algorithms and Applications, Volume 15 Advances in Parallel Computing; Eds. C. Bischof et al., IOS press, p. 689.

Leydesdorff, L. & Wagner, C. (2007). Is the United States losing ground in science? A global perspective on the world science system in 2005. In *Proceedings of ISSI 2007, Volume 1.* 11th International Conference of the International Society for Scientometrics and Informetrics, CSIC, Madrid (pp. 499-507).

Leydesdorff, L., & Etzkowitz, H. (1997). A Triple Helix of University-Industry-Government Relations. In H. Etzkowitz & L. Leydesdorff (Eds.), Universities and the Global Knowledge Economy A Triple Helix of University-Industry-Government Relations (1 ed., pp. 155-162). London: Pinter.

Leydesdorff, L., & Zhou, P. (2005). Are the contributions of China and Korea upsetting the world system of science? Scientometrics, 63(3), 617-630.

Lichtenstein, P., deFaire, U., Floderus, B., Svartengren, M., Svedberg, P. and Pedersen, N. L. (2002). The Swedish Twin Registry: A Unique Resource for Clinical, Epidemiological and Genetic Studies, *Journal of Internal Medicine*, vol. 252, no. 3, pp. 184-205.

Marechal, B. & Gavillet, P. (2008). EELA-2 and beyond it. Retrieved 07.05.2009 from: http://documents.eu-eela.org/record/1225/files/.

Marechal, B. (2008). Grid infrastructure for e-Science: a use case from Latin America and Europe. Presentation at the EuroAfriCa-ICT FP7 Awareness Workshop, Kampala, Uganda – 20-21 October 2008. Retrieved 07.05.2009 from: http://documents.eu-eela.org/record/1216/files/.

Marechal, B., Gavillet, P. & Barbera, R. (2009). Long-term sustainability of e-infrastructures in LA: the EELA-2 model. *Presentation at the CCICT Conference,* Kingston, Jamaica, 16.03.2009. Retrieved 07.05.2009 from: http://documents.eu-eela.org/record/1265/files/.

Mark, G., Grudin, J., & Poltrock, S. E. (1999). Meeting at the Desktop: An Empirical Study of Virtually Collocated Teams. In *Proceedings of ECSCW'99*, the 6th European Conference on Computer Supported Cooperative Work (1 ed., pp. 159-178). Copenhagen, Denmark.

McLoughlin, I. (1999). *Creative technological change. The shaping of technology and organisations* (1 ed.). London; New York: Routledge.

MONARC Study Group (1999). *Models of Networked Analysis at Regional Centres for LHC Experiments (MONARC Study)*, http://www.cern.ch/MONARC/docs/progress_report/Welcome.html [accessed 20 July 2009]

Narin, F., Stevens, K., & Whitlow, E. S. (1991). Scientific collaboration in Europe and the citation of multinationally authored papers. *Scientometrics, 21*, 313-323.

National Science Board. (2004). *Science and Engineering Indicators 2004* (1 ed.). Arlington, VA: National Science Foundation.

Nentwich, M. (2003). *Cyberscience. Research in the age of the Internet* (1 ed.). Vienna: Austrian Academy of Science Press.

NIH (2003). http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm Accessed on March 18, 2009.

Olson, G., & Olson, J. (2002). Distance matters. In J. M. Carroll (Ed.), *Human-Computer Interaction in the New Millennium* (1 ed., pp. 139-179): Addison-Wesley.

Otjacques, B., Hitzelberger, P. and Feltz, F. (2006). Identity Management and Data Sharing in the European Union, *Proceedings of the 39th Hawaii International Conference on System Sciences*, January 04 - 07, vol. 4, IEEE Computer Society, Washington, DC, USA, p. 70.1.

Podvinec, M., Maffioletti, S., Kunszt, P., Arnold, K., Cerutti, L., Nyffeler, B., Schlapbach, R., Türker, C., Stockinger, H., Thomas, A.J., Peitsch, M.C., Schwede, T. (2006). The SwissBioGrid Project: Objectives, Preliminary Results and Lessons Learned. In *2nd IEEE International Conference on e-Science and Grid Computing*, Amsterdam, the Netherlands.

Pringle, G. J.; Bournas, O.; Breitmoser, E. et al. (2007). Code Migration within DEISA, *Proceedings of ISC'07*, Dresden, June, 2007, http://www.epcc.ed.ac.uk/docs/2007-jul/Pringle2007.pdf.

Ronneberger, K., Kindermann, S., Biercamp, J. (2006). "Grid Based Climate Data Analysis" *Conference EGEE06*, Geneve, Switzerland, 25-29 September 2006 http://www.c3grid.de/index.php?id=54&L=1 (accessed May 4th 2009)

Sanderson, D. (1996). Cooperative and collaborative mediated research. In T. M. Harrison & T. D. Stephen (Eds.), *Computer Networking and Scholarly Communication in the 21st Century University* (1 ed., pp. 95-114). Albany, NY: State University of New York Press.

Scheuch, E.K. (2003). "History and visions in the development of data services for the social sciences." *International Social Science Journal 55*(177): 385-399.

Schindler, U., Bräuer, B., Diepenbroek, M. (2007). "Data Information Service based on Open Archives Initiative Protocols and Apache Lucene", *German e-Science Conference 2007* Baden-Baden, May 2007 http://hdl.handle.net/10013/epic.26667 (accessed May 4th 2009)

Schroeder, R. (2007). e-Research Infrastructures and Open Science: Towards a New System of Knowledge Production?', *Prometheus, 25*(1), pp.1-17.

Schroeder, R., den Besten, M. and Fry, J. (2007). Catching Up or Latecomer Advantage? Lessons from e-Research Strategies in Germany, in the UK and Beyond', in *Proceedings of German e-Science 2007*, Baden-Baden, http://edoc.mpg.de/display.epl?col=100&grp=1414

Shimizu, T., Shirai, D., Takahashi, H., Murooka, T., Obana, K., Tonomura, Y., et al. (2006). International real-time streaming of 4K digital cinema. *Future Generation Computer Systems, 22*(8), 929-939.

Shirai, D., Kawano, T., Fujii, T., Kaneko, K., Ohta, N., Ono, S., et al. (2009). Real time switching and streaming transmission of uncompressed 4K motion pictures. *Future Generation Computer Systems, 25*(2), 192-197.

Simeoni, F.; Castelli, D.; Pagano, P.; Simi, M.; Connor, R. (2008). Application-level Research e-Infrastructures: the gCube Approach, *UK e-Science All Hands Meeting 2008*, September 2008.

Smarr, L., Herr, L., DeFanti, T., Ohta, N., & Otto, P. (2007) CineGrid: A New Cyberinfrastructure for High Resolution Media Streaming, *CTWatch Quarterly, 3*(2), [Electronic Version] from http://www.ctwatch.org/quarterly/articles/2007/05/cinegrid/

Statistics Sweden (SCB). (2009). www.scb.se/eng. Accessed on March 18, 2009.

Stenberg, S-Å and Vågerö, D. (2006). Cohort Profile: The Stockholm Birth Cohort of 1953, *International Journal of Epidemiology*, 35, pp. 546-548.

Svensk Nationellt Datatjänst (SND) (2009).http://www.ssd.gu.se/?lang=en. Accessed on March 18, 2009.

Swedish Twin Registry (STR), (2009). http://ki.se/ki/jsp/polopoly.jsp?d=9610&l=sv. Accessed on March 18, 2009.

Ulbrich, U., J.G. Pinto, H. Kupfer, G.C. Leckebusch, T. Spangehl and M. Reyers (in press). "Changing Northern Hemisphere Storm Tracks in an Ensemble of IPCC Climate Change Simulations", *J. Climate.*

United Nations Development Program (2008). *Human Development Report 2007/2008.* Retrieved 27.05.2009 from: http://hdr.undp.org/en/media/hdr_20072008_table_1.pdf.

Vetenskapsrådet (VR), (2005). Strategi och infrastruktur för världsledande forskning på svenska register [Strategy and Infrastructure for World Leading Research on Swedish Registers]. *Internal report based on an investigation concerning the possibilities of improving the conditions for research on Swedish register data.*

Vetenskapsrådet (VR), (2009). www.vr.se. Accessed on March 18, 2009.

Widmann, H., Kindermann, S. (2006). "Data Discovery and Basic Processing in C3-Grid" GO-ESSP Workshop June 19-21, 2006 at the Lawrence Livermore National Laboratory LLNL, Livermore. http://data1.gfdl.noaa.gov/%7Eck/goessp/presentations/06_19_06/agenda_prese ntations.html%29 (accessed May 4th 2009)

Wooley JC, Lin HS (Eds). (2005). *Catalyzing Inquiry at the Interface of Computing and Biology*. Washington, DC, USA: The National Academies Press 2005.

Wouters, P. (2002). Policies on Digital Research Data – An International Survey. Amsterdam: NIWI-KNAW.

Wouters, P., & Schröder, P. (Eds.). (2003). Promise and Practice in Data Sharing. Amsterdam: NIWI-KNAW.

Wuchty, S., Jones, B. & Uzzi, B. (2007). The Increasing Dominance of Teams in Knowledge Production, *Science, 316,* 1036-1039.

PART 2 -A ROADMAP TO 2020 AND BEYOND

Authors:

Ralph Schroeder, Kathryn Eccles, Eric Meyer

Oxford Internet Institute

With contributions from the eResearch2020 consortium and external reviewers

1   Introduction and objectives

This roadmap is part of the study eResearch2020[25], commissioned by the European Commission which aims to examine the role of e-Infrastructures in the creation of global virtual research communities. The study was undertaken to identify how e-Infrastructures can support research and society in Europe and beyond in the coming years. The roadmap provides recommendations (listed at the end, they are also highlighted throughout the report) to the Commission on how e-infrastructure development can best be promoted through EU policy and the larger community of stakeholders. The study's objectives have been reached through extensive research (including case studies and a survey) on which this roadmap and its recommendations are based. This roadmap and the research for the study provide a contribution to policy making and to raising awareness about the future implications of e-Infrastructures.

1.1   Definitions and Key questions

e-Infrastructures can be **defined** as networked tools, data and resources that support one or several communities of researchers, broadly including all those who participate in and benefit from research. The impact of e-Infrastructures on virtual research communities will be affected by:

- harmonization of regulation and governance of e-Infrastructures, and integration of national and disciplinary e-Infrastructures
- organizational models and models for sustainability
- developing strategies for engaging research communities

As our concern involves both e-Infrastructure providers (defined here as socio-technical organizations which provide support services for research using digital tools and data), and their respective virtual research communities (the users of this support service)[26], **key questions** addressed in the study on which this roadmap is based included:

- What kinds of e-Infrastructures are successful and less successful in anticipating and catering to the needs of virtual research communities?
- How well do e-Infrastructure providers define, consult, plan for, engage with and overcome bottlenecks in scaling up to match growth in their user community?
- How do e-Infrastructures coordinate with other complementary tools and resources to maintain a unique profile while also integrating with other synergetic efforts?
- How do e-Infrastructures implement a strategy to ensure that they make an essential contribution to their community of beneficiaries?
- What kinds of instruments do e-Infrastructures need to gauge and adjust their provisions on an ongoing basis in order to cater to their communities?
- How do e-Infrastructures contribute to the integration of the European research area (ERA) and its integration globally?

---

[25] See http://www.eresearch2020.eu/index.htm
[26] The distinction between providers and users may be blurred in practice, as providers can also be users.

2   e-Infrastructure and its Potential Impacts

Specifying future developments in e-Infrastructure requires recognition of the dual value of e-Infrastructure to research communities, both as a resource in its own right and as a means of access to other resources. The Green Paper "The European Research Area: New Perspectives" (European Commission 2007) distinguishes e-Infrastructure from "Science &Technology infrastructures" (which tend to be physical large-scale facilities), pointing to the near-universal added value of e-Infrastructure by declaring a need to ensure "coherent planning, parallel development and integration between European S&T infrastructures and new generations of electronic infrastructures".

Whereas e-Infrastructure (specifically: links in the network layer) is fundamentally distributed geographically, other research infrastructure - particle accelerators, telescopes etc. - are inherently located at one or a few points in **geographic space**. As well as being a research infrastructure in its own right, e-Infrastructure provides added value to fixed-location research infrastructures. e-Infrastructure can provide many modes of access, independently of distance to fixed research infrastructures such as particle accelerators, telescopes, marine research ships etc. which previously could be used in research only at the specific location concerned. It is worth pointing out that the location & access dichotomy lies within e-Infrastructures themselves. Even large databases are inherently located; they involve data stored on specific devices. However, caching, replication and remote management techniques implemented in a grid layer enable these to be used in many ways as if they were local. This can enormously reduce time and effort by researchers in a distributed community. Further, it can enable many different collaborations apart from in distributed communities; for example remote access to instruments, Wiki-style aggregation of data and content that can analyzed using semantic web-type technologies, video-conferencing combined with research meetings and sharing of results, and many more besides.

This collaboration also includes access to supercomputers (also known as high-performance computers), which provide resources for a range of high-end users and research teams.

e-Infrastructures are commonly conceived in **layers**, from the networks up through grid integration to data, applications and users. All these layers are essential to achieving a pay-off in research communities. However, each brings with it specific issues to be taken into account in policy. High-speed networks are essential to bridging distance in a way that data transport is quasi instantaneous for many applications. A key issue is the major investment needed to create these networks, and the charges or cost-attribution for their use. Grid technologies are essential to remove the costs associated with adapting research applications to various specific sets of underlying throughput and capability resources, masking the fact that computational and storage resources are usually not local to the researcher. And some important grid technologies are in proprietary hands, raising issues of cost to public research.

Access to, **sharing of**, and curation of **data** (in a broad sense) over the longer term is a major issue for e-Infrastructures that will deserve special consideration in its impact on research communities (see also below, under 'Bottlenecks'). Indeed, the issue of sharing and re-use of data is moving to the forefront of discussions of e-Infrastructures (Nelson 2009, and other articles in the special issue of the journal Nature devoted to data 2009). The reasons why researchers are reluctant to share data are by now well-known (see also Borgman 2007). They vary by discipline and include, foremost, the competition by researchers to be first with their results and the sheer effort involved. Incentives by funding bodies and recognition for contributions to shared databases are often seen as possible measures to overcome these obstacles. Further, there is a distinction between e-Infrastructure requirements for those disciplines dealing with sensitive data about human participants and populations (Axelsson

and Schroeder 2009), as against researchers who share data without human participants (i.e. astronomy, environmental data, and artistic works) or where issues of intellectual credit, curation, and copyright arise (on the IP of databases, see Wouters 2002; Wouters and Schröder 2003). It is also important to expose problems of timing or temporal fit between ever more rapid ability to achieve research results and the relatively slow approval and publication process, and to ask if these processes can also be speeded up *(see Recommendations 4 & 5).*

2.1   How useful is the term e-Infrastructures?

The use of the term 'infrastructures', and the framing of some e-Infrastructure projects, invites comparison with the large technological systems that were built to support societies over the course of the late 19th and 20th centuries such as transport, electricity and communication. A recent workshop of the US National Science Foundation (Edwards et al. 2007) reported on the tensions in the evolution of e-Infrastructures, based partly on analogies with these previous infrastructures. Key lessons include addressing 'reverse salients' (or bottlenecks) and the implications of path dependence (lock-in effects deriving from the selection of certain technologies over others early on – which then have massive consequences since one is locked in to particular technologies).

But these comparisons can also be misleading. Unlike the large historical infrastructures, e-Infrastructures are aimed at target audiences of users mostly in the tens of thousands or less, and from specialist communities, and often serve highly specialized research aims. They are also quite diverse, and while some serve a single community, others will serve several. It is important, therefore, not to take the comparison with traditional infrastructures too far, since many of these socio-technical ensembles are more like temporary technical networks, serving smaller and larger numbers of users who may use a number of such networked tools in an overlapping way, but without becoming an essential support for the whole community of researchers – which the term 'infrastructures' implies.

When we think about e-infrastructures, if we reject this image of an 'infrastructure', with one fundamental common top layer, supporting a middle layer of standardized tools and data and resources (or services) or 'middleware', and the user on the bottom layer accessing the other two infrastructure layers, we might perhaps replace it with a jumble of overlapping and intersecting networks connecting in various ways to sets of overlapping and intersecting communities of users*.* For example, there is no one-to-one mapping between providers and communities – but rather there could be multiple links between both (i.e. one provider serving multiple communities, and vice versa.) *(see Recommendation 11).* Hence, too, we use the plural 'infrastructures' to indicate this heterogeneity. This more accurate and realistic view of e-Research infrastructures projects invites a number of useful reflections, key amongst them that there is no 'one-size-fits-all' approach in developing both the tools and the policies which support them.

2.2   e-Infrastructure in 21st century research

Changes are taking place in European and global science which require analysis to secure the position of European research in the global science and technology system. Among these developments, the following are considered highly relevant for the issues addressed in this roadmap:

1. *Increased demands on science from other areas of society*, in particular from industry, politics and social interest groups in the wake of "Post-Normal Science" (Funtowicz & Ravetz, 1993), "Mode 2" (Gibbons et al., 1994), or the "Triple Helix" (Etzkowitz & Leydesdorff, 2000; Leydesdorff & Etzkowitz, 1997). The pressures for researchers to respond more quickly and on a larger geographical scale to

tackle global problems such as climate change, the monitoring of environmental problems, diseases and epidemics, or social and political crises have grown, as have the demands for economic and social returns from scientific work and the financial squeeze of science funding *(see Recommendation 13)*.

2. *Increasing presence of non-triadic (Europe, North-America, Japan) countries* like China and India, but also other East Asian or South American countries, on the global scientific market (Leydesdorff & Wagner, 2007; Leydesdorff & Zhou, 2005).

3. *Rising importance of team-based research and collaboration.* Research is increasingly taking place in larger teams, and team efforts have a greater impact compared with individual efforts or those of smaller groups as measured by citations. This applies not just to natural science, but also to social sciences and only to a somewhat lesser extent, to the humanities (Wuchty, Jones and Uzzi 2007). Several reports have shown that the importance of scientific collaboration has grown in the last 25 years (European Commission, 2003; Narin, Stevens, & Whitlow, 1991; National Science Board, 2004).

4. *The increased complexity of science.* This includes the growing scale of challenges such as climate change which require more sophisticated models and pooling data from medical trials which require large-scale population samples. Other examples include larger instruments for capturing astronomy data or sensor networks that aim at extensive geographical coverage *(see Recommendation 1).*

5. *Growing collaboration between disciplines.* The increasing scale of terms (point 3) has already been noted, but there are trends towards the combinations of disciplines and fields including not only the hyphenated sciences (bio-physics, bio-chemistry etc.) but also new specialisms (arts computing). Apart from this, e-Infrastructures often require coordination and collaboration not just between computing and other disciplines, but also multi- and interdisciplinary teams that include a range of disciplines.

> These trends have created pressure for significant investment in technologies that support distributed research and collaboration. e-Infrastructures are being rapidly developed and deployed worldwide and across the European Research Area (ERA) to support team-based research and resource sharing over e-Infrastructures in the form of virtual or distributed organizations. The European Commission is the main driver of these infrastructures in Europe through the Framework Programmes. In FP6, this included networking and Grid infrastructures such as GEANT and EGEE, but also domain specific ones such as BioinfoGRID and those developing high-performance computing such as DEISA. Some of these initiatives carried into FP7, such as ESFRI (The European Strategy Forum on Research Infrastructures), EGEE III, and many others. There is now a transition from EGEE to EGI, the European Grid Initiative, which will link national grids across Europe, provide access from and to e-Infrastructure projects, and link e-Infrastructures more globally. There is now, since August 28, 2009, also now a new European legal instrument, ERIC (European Research Infrastructure Consortium), which can provide support for e-Infrastructure collaborations (Thies 2009).
>
> **Outside of Europe**, the initiatives of the US National Science Foundation through its Office of Cyberinfrastructure gained considerable momentum after the publication of the NSF blue-ribbon panel report in 2003 (Atkins et al., 2003, Berman and Brady 2005). Significant funding in the US is directed to Cyberinfrastructure development and deployment in such infrastructures as the TeraGrid and the Open Science Grid as well as a variety of global virtual organizations for research such as the Open Grid Forum. In the US, the continued commitment of the NSF to cyberinfrastructure has recently been confirmed with the announcement of FutureGrid[27], which is set to be integrated with Teragrid in the coming years. Outside of the US and Europe smaller, but by no means negligible efforts are being undertaken for instance in China, Japan, Australia, New Zealand, and Canada. The European efforts are already being coordinated with some of these (for example, EUChinaGRID).

---

[27] http://futuregrid.org/

2.3   Current EU policy on research infrastructures

The 2007 Green Paper "The European Research Area: New Perspectives" (European Commission 2007) points to a number of requirements for future roadmaps of research infrastructures, of which the current roadmap is part. Reference is made to some apparent shortcomings of the first ESFRI roadmap (2006), one being that coverage of the appropriate range of research infrastructures may not be complete. The question of coverage also emerges from the study on which this roadmap is based, since different fields and disciplines have quite variable provision of e-Infrastructures – in terms of current provision as well as future anticipated needs and provision.

The first **update of the ESFRI Roadmap** was published in late 2008, as requested by the European Council of Research Ministers. This update reported on 'new facilities tackling challenges in Environment, Energy and Health' and gave progress reports on, and endorsed '*almost all* the previous projects in the first edition' (my italics). Crucially what this update does not offer is any reflection or analysis of the projects that did not succeed following inclusion on the first ESFRI Roadmap; neither does it provide any discussion on the comparative prospects of newly included projects. The update therefore eclipses the original roadmap and its predictions, rewriting the future of e-Infrastructures without analytically probing their past.

The 2006 ESFRI roadmap has faced a number of other challenges which hold lessons for policy makers. There is the suggestion in the Green Paper that policy endorsement of the roadmap may not be complete, for example in terms of national funding bodies agreeing to co-funding of the ESFRI projects, and that (consequently) the € 14 bn funding over 10 years required to implement the roadmap has not yet been made available. Current European Union policy is to focus resources on supporting open access to infrastructures of interest and stimulating their coordinated development and networking, rather than providing core funding for new infrastructures *(see Recommendation 10)*. The Green Paper also notes that several infrastructure projects proposed by ESFRI are on such a scale and scope that they would, if adopted as policy, require cooperation at global level. In any event, the ESFRI roadmap and projects can be seen as a central plank in European planning for future e-Infrastructures provision, and thus ensuring that this raft of projects goes forward in some form must be one major policy goal.

The Green Paper raises funding sustainability as one of a number of questions to be addressed in e-Infrastructures policy *(see Recommendation 2).* It is asked how the EU can fund infrastructures and with what combination of specific Community funding, Member State contributions and synergy with policy instruments. There is also mention of the European Investment Bank and other financial institutions, but though these can provide an important contribution by enabling expenditure to be brought forward in time, they are not net sources of funds and they will require assurance that the costs of the e-Infrastructures they may finance will be met from other sources in due time. The 2008 update to the ESFRI roadmap reports that increased integration with national roadmaps and funding priorities has seen greater financial security for e-Infrastructures, but also points to future efforts (such as the development of a new legal framework and the integration of the national budgets with EU funds) *(see Recommendation 3).*

A particularly important issue is how **private sector resources** can be mobilised. The Green Paper reported that there has been little success in mobilising investment in ESFRI roadmap infrastructures from industry, even where a strong business interest might be supposed. As industry has disappointed to date as a source of funds, discovering what policy and legal changes might be necessary to unleash private sector investment is vital. The ERIC framework does not exclude participation by the private sector, but is clearly intended as an instrument for non-commercial entities in the first instance *(see Recommendation 13).*

The low level of **industry involvement** is surmised to be due in part to a current lack of appropriate legal structures. One of the cases in our study, SwissBioGrid, demonstrates that legal issues around industry involvement can be successfully resolved, even if they are time-consuming. One question that remains is whether a (new) European legal framework and/or by common and transparent principles for the management of, and access to, European e-Infrastructures would facilitate the emergence, operation and continuous improvement of new (electronic) infrastructures? Such a framework would enable each partner in a joint venture to reap the rewards which motivate the contribution of resources and sharing of costs *(see Recommendation 5)*, and would need appropriate governance structures which give the partners the appropriate level of control and capability for intervention. In terms of partnerships, it also needs to be mentioned that there has been (in FP7), and should continue to be openness to the involvement with industry *(see Recommendation 13).*

The geographical reach of future e-Infrastructures is a key topic. The Green Paper, apart from pointing to the need for e-Infrastructures to reach peripheral regions, exhorts Europe to 'continue with the extension to other continents of GEANT and grid electronic infrastructures', giving the reason that these 'constitute powerful instruments for international cooperation and the establishment of **global research partnerships**'. Clearly, international collaboration, partnerships and communities are at the centre of the study, and the report points to a number of interesting geographical features of current e-Infrastructures provision and of research communities. This more collaborative and global and intensive use of e-Infrastructures for research as essential to innovation in the ERA is also a key point in the most recent council statement of the Council of the European Union (2009) *(see Recommendation 1).*

2.4   How can roadmaps support e-Infrastructures?

It may be useful to reflect briefly upon the role of roadmaps in guiding policy, and how this roadmap complements and goes beyond others. Technology roadmaps began to gain acceptance in industry and government circles in the late 1990s, and have from that time onwards become increasingly common in science more generally (Galvin, 1998). Galvin's definition of a roadmap is 'an extended look at the future of a chosen field of inquiry composed from the collective knowledge and imagination of the brightest drivers of change in that field. [They] can comprise statements of theories and trends, the formulation of models, identification of linkages among and within sciences, identification of discontinuities and knowledge voids, and interpretation of investigations and experiments. Roadmaps can also include the identification of instruments needed to solve problems, as well as graphs, charts, and showstoppers.' Roadmaps from different sectors (industry, government, academia) can have very different agendas and motivations (Kostoff and Schaller, 2001), comprising different combinations of the components identified by Galvin. The extent to which e-Infrastructures roadmaps can be useful in driving policy is therefore clearly dependent on the nature of the organisations producing and applying the roadmaps.

Two key organisations producing e-Infrastructures roadmaps are the European Strategy Forum on Research Infrastructures (ESFRI), discussed above, and the e-Infrastructure Reflection Group (e-IRG) whose most recent roadmap was published in 2006, with an update in 2007 and a further update currently under construction.   ESFRI, formed at the behest of the European Council in 2002, is intended to function as 'a strategic instrument to develop the scientific integration of Europe and to strengthen its international outreach'. It states that 'competitive and open access to high quality Research Infrastructures supports and benchmarks the quality of the activities of European scientists, and attracts the best researchers from around the world'[28]. ESFRI delegates are nominated by Research Ministers of European Member and

---

[28] From http://cordis.europa.eu/esfri/, consulted 07/08/09

Associate Countries, and include a representative of the Commission, all of whom are charged with judging and reporting on the latest developments in science, in research infrastructures, and in the application and use of knowledge-based technologies. They are also responsible for working to overcome potential problems, caused by the fragmentary nature of national efforts and policy, in coordinating efforts in these areas.

e-IRG was formed in 2003, in order to 'support the creation of a political, technological and administrative framework for an easy and cost-effective shared use of distributed electronic resources across Europe'[29]. A major focus of e-IRG's work is on grid computing and general IT infrastructures, encompassing networking, highly advanced computing, grids and storage. The pesidency of the e-IRG rotates alongside chairmanship of the European Union. The most recent policy guidance of e-IRG is the White Paper (2009) and Roadmap (2009, consultation version). This White Paper gives a broad overview of current e-Infrastructures, the issues and challenges they face, and makes a range of policy recommendations.

The ESFRI roadmap is concerned with a specific raft of e-Infrastructure projects while the e-IRG has tackled specific issues such as open access, interoperability and standards. The 2020 roadmap, in contrast, deals with broader and more comprehensive issues than e-IRG, but also covers the whole spectrum of science and research on the other. In particular, this roadmap provides a foundation for the more future-oriented scenarios described later. Apart from the case studies and survey results of the 2020 study that are drawn upon below, one of the most important challenges relating to current e-Infrastructures road mapping is that there is little empirical data or information to go on. Among the gaps that should be highlighted are:

- Lack of a comprehensive understanding of the implications of e-Infrastructures for research and knowledge (see Meyer and Schroeder 2009a,b)

- Tendency to focus on technology savvy disciplines and users from these technology savvy disciplines, rather than also including those who are less technologically savvy or have little awareness and/or interest in e-Infrastructures (but see Dutton and Meyer 2009, AVROSS 2008)

- Need for more evidence-based analysis of the impact of e-Infrastructures, for example via bibliometrics and webometrics (but see Park, Meyer, and Schroeder 2009)

- The tendency to build e-Infrastructures due to capability and/or enthusiasm for certain kinds of structures, rather than allowing research/field specialists to influence technology development, has been a problem that has been recognized, for example, in the case of the UK (see e-Science Director's Forum Strategy Group 2009), where e-Science is now almost a decade old. This shift of balance away from technology developers to focusing on users or research communities will only take place with the maturing of e-Infrastructures.

This study has attempted to overcome some of these problems by means of case studies, a survey, and related research. But it should be stressed that much more research on this topic will be needed *(see Recommendation 14)* as large-scale investments are deployed and further evidence which could address the limitations of current road mapping become available.

---

[29] From http://www.e-irg.eu, consulted on 07/08/09

3   Foundations of the Roadmap

3.1   Case Studies from the 2020 Report

The case studies conducted for this study, the information from which is used to construct this roadmap, represents the first attempt to develop a systematic understanding of the range of technological and organizational outcomes in the transition to e-Research infrastructures, and the implications of these findings for guiding policy regarding their future development. Although several studies have recently examined a single e-Research infrastructure (see for example Olson, Zimmerman and Bos, 2008), much less studied but of considerable empirical and conceptual significance is research that identifies some defining characteristics that distinguish between different types of infrastructures and across different fields of research.

Methodology: Selecting the sample case studies

In order to capture different levels of involvement, services offered and developed, and organizational objectives, this study distinguished between:

- *Providers*: including distributed organizations that offer e-Infrastructures to virtual user communities. Among the services offered are dedicated high-bandwidth networks, supercomputing and Grid computing facilities - including data Grids - community portals, training and technical support.

- *User communities*: virtual communities that utilize and further develop e-Infrastructures applications and instruments that are specific to their domain. This study analyzed communities from diverse disciplines and fields, including the life sciences, hard sciences, social sciences and the humanities.

Sample Case Studies		
	Case study	**ESFRI category**
Providers	DEISA	e-Infrastructure
	EELA-2	e-Infrastructure
	EGEE	e-Infrastructure
	GÉANT	e-Infrastructure
	OSG	e-Infrastructure
	Teragrid	e-Infrastructure
	Swedish National Data Service	Social Sciences and Medical Sciences
User communities	C3-Grid	Environmental Sciences
	CineGrid	e-Infrastructure
	CLARIN	Social Sciences and Humanities
	D4science	Environmental Sciences
	DARIAH	Social Sciences and Humanities
	DRIVER	e-Infrastructure
	ETSF	Materials and Analytical Facilities
	MediGrid	Biological and Medical Sciences
	NVO	Physical Sciences and Engineering
	Swiss BioGrid	Biological and Medical Sciences
Standards	OGF – Open Grid Forum	e-Infrastructure

* Note: Non ESFRI-projects were classified into the ESFRI categories by the authors.

The report details the selection of these case studies and includes considerable additional detail about each.

3.2   2020 Survey of e-Infrastructures and research communities

This study has investigated research activities and e-Infrastructures use across a range of research communities. Both individual researchers and research communities have been asked to provide essential insight into the research process using e-Infrastructures. The surveys conducted for this project have been designed to address both sets of informants - surveys of both providers and the research communities they serve. Provider information is the key to the *e-Infrastructures Survey*, whereas the point of view of researchers is at the centre of the *Research Communities Survey*.

e-Infrastructures providers are addressed as those responsible for the characteristics of the technologies that undergird e-Infrastructures and the research communities using it. e-Infrastructures service providers are well positioned to help evaluate usage scenarios of various research communities, as well as to provide a coherent account of some of the challenges that have arisen over time. However, service providers often do not have a detailed insight into the extent of collaborative research activity or into many aspects of research community behaviour relevant to this study. Also, it would be a mistake to neglect the possible conflict of interest there may be in some cases between an honest assessment of history and current situation of an e-Infrastructure, and the promotion of the economic success of the provider organisation.

3.3   Typologies emerging from the 2020 report

The e-Infrastructures reviewed in this study can be categorized in many different ways. Some are relatively small, starting from single institutions (SND) or countries (Swiss BioGrid), others such as CineGrid could be thought of as a medium-scale international collaboration, and large-scale projects would describe projects such as OSG and EGEE, with more than 50 partner organizations. Some of these projects, while small in scale, have actively marketed themselves to more than one country, while other larger projects have thus far limited their activities to one country. It does not necessarily follow, therefore, that if an infrastructure has a large number of partners it is likely to be more international in its outlook, nor that smaller scale efforts limit their participation to local partners.

It can also be a useful to think about disciplinary boundaries as a means of categorizing e-Infrastructures. Some projects originate from within one discipline or field (MediGrid), or indeed a sub-field within a discipline (Swiss BioGrid), others span multiple disciplines (CLARIN, SND) and even commercial fields (CineGrid), while other projects, largely the 'grid' infrastructures presented in this study, originate less from a single discipline and seek rather to offer technological power to a range of academic users (OGF, TeraGrid). These disciplinary differences can often be effectively mapped on to an analysis of 'developer-driven' versus '(user) community-driven' growth factors, with community-driven efforts often emerging from a single discipline or field (which are then frequently applied to other disciplines or fields when viability of methods or technology has been established) and developer-driven efforts more likely to emerge from technological specialists who seek to apply their developments across disciplines. Funding is also a factor that can be considered here, as community-driven efforts have historically been funded on a small scale and through disciplinary or institutional channels, whereas large-scale developer-driven technological infrastructures have tended to emerge through government or EC driven funding priorities.

3.4   Governing e-infrastructures

In terms of the organization of governance, there is a scale from the small and informally organized (CineGrid is an example) to larger multi-tiered and more elaborate and complex structure (Geant). One feature that is common to all larger projects is an advisory or steering committee of some sort (in some cases both, such as for Clarin) – a group which oversees the project and guides the management level. These are sometimes internal, sometimes external. They are also sometimes constituted so as to provide guidance, sometimes more to ensure 'democratic' representation from among all project members or stakeholder groups. We could call this 'metagovernance', which can be defined as an external layer on top of organizations using technology to mediate between them. This can be very thin, informal and flexible (Swiss BioGrid is an example) or highly complex, formally institutionalized and hierarchical (EGEE serves as an example). It is interesting to think of this 'metagovernance' layer, enabling technological development and use, as being one key to enabling successful e-Infrastructures *(see Recommendation 7).*

Further, it is noticeable that in some cases, both the advisory or steering committees and the management group seem to come from among the researchers and from within the disciplines themselves (NVO), whereas in other cases a broad constituency from across disciplines is represented (Swedish National Data Service). A further dimension is whether the governing bodies are permanently constituted and include core staff that is constantly occupied by governance tasks, or if there is only episodic governance by means of regular face-to-face or teleconferencing type meetings.

More versus less centralization is of course a key factor in governance, but it seems that, unless projects are so small as to be a 'one man show' (SwissBioGrid), there are either one or a few coordinators who delegate tasks, with only the larger projects in addition having a larger more representative body which coordinates and delegates tasks. There is a split between more straightforwardly organized computing intensive projects such as DEISA - versus more sensitive data oriented and organizationally 'messy' projects (such as SND). Only in a few cases (OGF, TeraGrid, OSG) is there a move away from a centralized towards a more federated or 'flat' organization which has multiple coordinators for different tasks (though some projects have such a body underneath the centralized coordinator or coordinating body). Apart from centralization, the main variety in governance comes from the high or low degree of division of labour. Whether there is good match between governance structure and the project functioning is difficult to generalize about. What is clear is that a variety of governance styles is possible, and that oversight and strategy as against management are separated in all the cases of larger infrastructures *(see Recommendation 7).*

3.5   Key Bottlenecks – technical and social

Bottlenecks can occur at many stages of the life-cycle of an e-Infrastructure and for a variety of reasons. A common scenario is when there has been limited collaboration between technical and domain specialists in the development of an e-Infrastructure, or difficulties in establishing the technical requirements of a project. The speed of development is therefore slowed by the capacity for both parties to communicate and collaborate.

Fewer bottlenecks occur when disciplines and fields are much more technically adept, and therefore domain specialists are more likely to develop requirements for e-Infrastructures ('bottom-up' growth) and/or to have the language and technical ability to collaborate with technical specialists on requirements. A good example of this kind of project would be Swiss BioGrid, in which a grid network was developed to support scientific research with few

bottlenecks, due to the sympathy between technical and domain specialists, and their ability to collaborate to solve problems when these arose.

Organizational barriers also appeared where both - collaboration across different organizations and across different countries or continents - was hampered by regulations at local or national levels, cultural (field differences, strong identities) or technical particularities. Another key factor influencing the development of bottlenecks is the financial status of a project. Projects with sustained funding can afford to keep the development of a project flowing, whereas those with start-up funding only or with more precarious financial arrangements, can find that bottlenecks develop when technical and social resources and enthusiasm for a project constitute a strong force that cannot then feed into its successful realization *(see Recommendations 2 & 3).*

Finally, as mentioned, a key bottleneck are data repositories or library-type facilities for **data sharing**: As a recent special issue of the journal 'Nature' (2009) has pointed out, the contributions and uptake of these facilities has been disappointing. Factors include not just unwillingness to share data because of scholarly competition (scenario 3 below), but also simply that the incentives are not available, that contributing to shared archives is not rewarded, that the governance mechanisms are uncertain, that the outlook for the longevity of the data is uncertain (it makes little sense to make the effort to contribute unless you are contributing to a long-lasting resource), that a lot of work is involved for which there are no resources, and the Catch-22 problem that these resources are not being built up unless there are users but no users unless there are useful and well-developed resources *(see Recommendations 4&6)*.

This roadmap is not the place for a complete list of bottlenecks. The report of the 2020 project identifies many more from our case studies and survey, and there is now also sizeable body of academic and policy research on the topic of bottlenecks. One point to emphasize here is that currently at the top of list of bottlenecks is perhaps sharing data[30], especially as this has become (and will continue to be) one of the major expenses, also in terms of effort, of e-Infrastructures, is likely to be one of the main – if not *the* main – challenge for e-Infrastructures development and uptake. It can be mentioned finally that bottlenecks can, of course, also be enablers, especially if these bottlenecks can be translated into points of action to be taken.

3.6 User Profiles and Use Profiles

Our case studies revealed many inconsistencies in the way that projects defined 'users' and 'usage', such as the measurement of single users rather than user organizations. We also found that there can be practical problems in attempting to trace and measure usage at the e-Infrastructure level:

- Users connect through gateways or portals which then do not appear as distinguishable organizations or individuals to the e-infrastructures providers;
- Registration and authentication are handled at a higher level (organization) and the individual user's identity is not revealed at log-in;
- Users log-into the e-infrastructure and then there is little monitoring of what tools and applications they actually use;
- It is impossible to distinguish between a former user who stopped use, e.g. because of a more suitable alternative, and someone who interrupted use, and e-infrastructures frequently lack the knowledge of which past users will return in the future.

---

[30] Tthis is also highlighted in the current e-IRG Roadmap 2009, see: http://www.e-irg.eu/index.php?option=com_content&task=view&id=39&Itemid=38; last accessed 20.1.2010

For these reasons, the numbers of users can vary widely between different e-Infrastructures.

There are other factors relating to the number of users, and the profile of users of e-Infrastructures. Some infrastructures reach very large and multidisciplinary user communities, others deal with a rather narrow set of 50 to up to 200 people. Those with small user communities are often restricted to pilot users, i.e. users from organizations participating in the project and a few scientists from other organizations, or have been deliberately developed for a specific set of users or a specific project or set of projects (C3-Grid, D4Science, Swiss BioGrid).

The number of users raises the questions of indicators, and how to judge the success of e-Infrastructures. It should be pointed out immediately that this is not just a question of the number of users: a variety of other indicators is possible, which include the outputs of research, enhancing the quality of research, providing access (also to those outside of research), mapping the number of users onto the number of potential users (for example, in the most extreme cases, an e-Infrastructure which consists of a large remote instrument, or of a very rare electronic manuscript, which is used only by a single research group – could be highly successful while an e-Infrastructure with tens of thousands of registered users who have no discernible benefit could be unsuccessful). The question of well-rounded, useful and valid indicators is inevitably moving high onto the agenda of research policymakers and e-Infrastructures developers, and it can be recommended that providers should be obliged to address the issue of valid measures (including number of effective users) explicitly in order to maximize their impact *(see Recommendation 12)*

Recruitment of new users is an important step towards creating a sustainable resource, and many of our studied cases have employed multiple measures in this regard. The ability to attract new users depends on many factors, including: security and ease of access, awareness of demand for the resource, interoperability, and willingness to invest time in the resource *(see Recommendation 11).* The extent to which infrastructures are embedded within the research communities they serve has a large impact on awareness among potential users. Nascent projects such as CLARIN, which is firmly established within its primary target research community and has grown largely from the enthusiasm and vision of this strongly networked community, have an informed notion of their potential user base. Infrastructures that are less well established within their target research communities are likely to have to work harder to develop user profiles and to integrate them into the research agenda. Technical accomplishment within the field also makes a crucial difference here, as developer-driven or 'top down' infrastructures are more likely to succeed within communities that are pre-disposed to technology. Some areas of the Humanities and Social Sciences, for example, may require a heavy investment in training and development on both sides, which may have an impact on uptake *(see Recommendation 8).*

An emerging trend in e-Infrastructures is the increasing investment in Humanities and Social Science projects, which have until recently (particularly in the case of Humanities) been somewhat overlooked by the drive for e-Research. Some areas of research within these fields are more technically assured than others, and efforts have so far focused on these disciplines and sub-fields. Investment in projects such as DARIAH, DRIVER and Europeana will have a significant impact on these fields and will facilitate much comparative and inter-disciplinary work. These fields may, however, require more investment in user support, since technical knowledge is not necessary to world-leading research in many of the disciplines in this area.

3.7   The role of e-Infrastructures in supporting researchers versus supporting society-at-large

Industry/commerce

A number of the projects that form our case studies interact, or have interacted, with industry. Swiss BioGrid, for example, was able to establish a successful working partnership with a research laboratory attached to the pharmaceutical company Novartis, in collaboration with the institutional partners that formed the grid. This public-private collaboration is highly unusual. Novartis further contributed a promise to validate any simulated results with experimental data, and to distribute any drugs that resulted from this research in the developing world at cost. The compounds are now being screened by Novartis.

A further example of this kind of successful public-private partnership is that operated by CineGrid, which took its lead from developments in research networking and scientific visualization and seeks to transfer these to digital cinema. Links between the two, therefore, are vital to the success of the enterprise, and have brought benefits on both sides. CineGrid maintains active links with commercial sponsors and partners, and the practical problems and revenue derived from these aids the technological development of this community as well as forming an important part of CineGrid's sustainability plans.

Both Swiss BioGrid and CineGrid have shown that these public-private partnerships can work, and that they can yield essential research, development, technical competence, financial security and sustainability for the projects involved. If these partnerships were to be more widely advertised and encouraged, however, it would be important to consider what benefits are to be gained from such collaboration. Some projects would be unable to collaborate in this way, others would simply choose not to for valid reasons. Policy makers must be clear about their goals in this regard, and should neither punish those e-Infrastructures that could or should not engage with industry, nor overly prioritise those who have an explicit link with industry *(see Recommendation 13.)*

Government

What do current projects offer government, directly and indirectly? Some of our e-Infrastructures case studies have had a profound effect on shaping governmental policy on e-Infrastructures and investment in research. Swiss BioGrid, for example, did not receive any central funding for its infrastructures, the entirety of the cost being borne by the institutions involved and project developers. Reluctance to invest in this type of project was largely related to the fact that this project was highly speculative and did not relate to a specific scientific goal (although it achieved such goals in its successful application). Having proved a successful test bed, lessons learned from the Swiss BioGrid project fed directly into a national initiative, the Swiss National Grid project (SwiNG), and therefore potentially have a huge impact on research in Switzerland more generally, though specific evidence for this remains to be seen. Another project that is relevant in this context is SND (the Swedish National Data Service), which has the potential for informing government policy with data, or C-3Grid which can inform government environmental and disaster policies. These kinds of synergies can be found in a range of our case studies.

Health

Health is becoming an increasingly urgent issue for research for a variety of reasons (ageing populations, more sophisticated health technologies, availability of large amounts of data). e-Infrastructures play a large role in health challenges via the use, for example, of physiological sensors, new techniques in the management of health records, predictive modelling for

disease and responses to therapy, and control of real-time therapeutic devices. These new techniques raise new challenges, such as data privacy, knowledge curation, integrating between European health services and between these services and researchers. FP7 and its successors and other initiatives are making massive investments in health research e-Infrastructures, and hence this area will require prioritisation – if only in coordinating this commitment of resources and ensuring that the unique bottlenecks in the domain of health e-Infrastructures (again, privacy and security of data foremost among them) are addressed early and effectively.

Education

e-Infrastructures have potentially a variety of impacts on (higher) education. In identifying such benefits, it is important to distinguish between research-led teaching and training taking place in institutions who are partners in or users or particular infrastructures, which we might expect to develop alongside the research facilitated by the infrastructure, and other educational uses or applications that result from the availability of e-Infrastructures. The latter might include, for example, ICE-Age[31] which has provided specific training. Some of the projects among our case studies had a more direct interaction with education than others. The project that had perhaps the most contact with education is Géant, which, through its contact with TERENA, the European association of research and education networking organisations, handles a number of educational outreach activities. In particular, it encourages the common exploration of new technologies between project partners and other groups that are active in technical development of particular relevance to research and education networking, through the continued operation of TERENA task forces. Users are connected to the national NRENS and using the network structure. These users are frequently unaware that they are using Géant.

TERENA also supports development of research and education networking in less-advanced regions in and around Europe and undertakes specific actions in support of the research networking organisations in the countries concerned. Geant and TERENA have also supported 'educonf'[32] a development activity designed to provide networked services across education and research institutions, and 'eduroam'[33] an educational network roaming infrastructure that allows users of participating institutions to access a wireless LAN at other participants' locations using their home institutions' credentials.

For the former type of educational benefits we find examples in nearly all of the cases that we looked at. Often this training is targeted at increasing users' abilities to use the e-Infrastructure, for instance by holding 'Grid schools' as the EELA-2 project does or training workshops as in CineGrid, NVO, EGEE and several other projects. In addition to that, the projects involve postgraduate students who receive part of their postgraduate education through doing research with the e-infrastructure. In regard to the educational use of e-infrastructure applications going beyond the project consortia we find little mention in the cases. Such educational use could yield considerable benefits, for instance when content is visualized and transmitted with high-resolution video and high-quality audio as would be possible with CineGrid, or cultural artefacts such as those provided by DARIAH and social data such as those provided by SND are provided to students in their education programmes. Some of the projects have not yet reached the stage of maturity to realise this link, but we got the impression that closer links to the higher education sector could help them in becoming more mature. There is one exception, of course, among the projects that we analysed: through the NRENs which it connects, Géant serves the education sector all over Europe from primary to secondary and tertiary levels.

---

[31] http://www.iceage-eu.org/v2/index.cfm
[32] http://educonf.geant2.net/
[33] http://www.eduroam.org/

Large-scale Arts and Humanities projects such as DARIAH and DRIVER are set to have a considerable impact upon education, as they will provide an infrastructure for the discovery of a huge range of European cultural artefacts and data. DARIAH seeks to create 'a common understanding of the cultural diversity and its history in Europe'. Arts, Humanities and Social Sciences projects of this kind are more accessible to schools and non-traditional spheres of education, and are therefore likely to have more of an impact in these areas. But public education via engagement is not limited to arts, humanities and social sciences, as the Galaxy Zoo project[34] has shown in the case of astronomy.

Finally, training with e-Infrastructures requires a balance between the need to reach a level of standardisation and maturity without a lock-in effect to a particular technology *(see Recommendation 8).*

Cultural heritage

Projects currently under development such as DARIAH, CLARIN and DRIVER, together with existing EU projects such as Europeana[35], are creating new and exciting possibilities for discovering and accessing cultural heritage across Europe. These projects are not only developing infrastructures to facilitate access to high quality national collections across Europe, but are prioritising interoperability, hugely increasing the impact of single projects, and creating powerful incentives for custodians of these resources to get involved. These projects have the potential to transform research, by linking up disparate collections, allowing new research questions to be posed, overcoming the language barriers that might prevent comparative research being undertaken, and creating new research tools that allow researchers to work collaboratively in new environments.

e-Infrastructures in the arts and humanities have a number of additional benefits that are often overlooked in comparison with e-Infrastructures in sciences and social sciences. First, they provide education resources for younger scholars, together with community or distant learning groups, with new opportunities for resource discovery. Like the education benefits discussed in the previous sections, such projects have the added benefit of mass appeal for a variety of users, vastly increasing the potential return on this investment. Second, these e-infrastructures developments frequently involve digitisation of cultural artefacts of substantial import and interest to the wider community. The democratisation of these resources therefore has a more easily understood and direct impact on the public at large than similar efforts in the sciences. In addition, opening up the cultural heritage of the European Union to every one of its citizens could have a profound effect on the unity of the community, and in promoting understanding between the different countries that comprise the Union.

---

[34] http://www.galaxyzoo.org/
[35] http://www.europeana.eu

4   Key Patterns from the Case Studies and Survey

We can now turn to some lessons from the case studies and survey combined. First, as for disciplinary and organizational differences:

- e-Research communities tend to either sit within well-defined disciplines, or they are more diffuse and need to adapt to a larger changing environment. Here we can think of physicists and astronomers in the former category, and large heterogeneous projects like EGEE (or EGI) or EELA on the other.

- There is a split between more straightforwardly organized computing intensive projects such as DEISA - versus more sensitive data oriented organizationally 'messy' projects (such as SND).

- Maturity brings out differences between disciplines. In mature disciplines, challenges have become clear. Latecomers to e-Research are interested in access to organizations, and resources and training - and challenges are yet to be identified. This is a key point for future policy: lessons from mature disciplines can be transferred to latecomers, but this also depends on whether disciplines (or transdisciplinary efforts) face the same or similar challenges.

Larger patterns emerging from our study:

- There is an interesting mix of clusters of projects with similar features, which points to variety. Is this a healthy pluralism? Or does it reflect the various stages different crystallizations of e-Infrastructures?

- There is a subset of national, data-centric e-Infrastructures with organizational barriers. Here the view of the research community can be summarized as: 'set the data free!'

- There is a subset of global, compute- and technology- centric e-Infrastructures without barriers. Here the main perceived threat is technological, for example cloud computing.

- All e-Infrastructures entail, in addition to their organizational or governance, a new organizational or governance form which can be called 'metagovernance', which (as noted earlier) can be defined as an external layer on top of organizations using technology to mediate between them. This can be very thin, informal and flexible (Swiss BioGrid is an example) or highly complex, formally institutionalized and hierarchical (EGEE serves as an example). It is interesting to think of this 'metagovernance' layer, enabling technological development and use, as being one key to enabling successful e-Infrastructures.

- There is still no clear mapping of disciplines and transdisciplines to different e-Infrastructures 'types', nor measurement of the impact of e-Infrastructures by means of indicators. The measurement problem is addressed in the Work Programme for 2010 for e-Infrastructures (Infra -2010-3.3), but this will be the single most important support for policy in the coming years *(see Recommendation 12).*

- There is a mix of risk levels identified among case studies, in the survey and from other parts of this Roadmap: some of the e-Infrastructures developments and investment are much larger-scale, more long-term and more leading-edge than others which are well-established and will require limited resources within a foreseeable time-scale. The mix of more pioneering and more 'conservative' e-Infrastructures organisations require a balance and needs to be monitored.

4.1   Emergent Patterns

Any roadmap for the next decade of e-Infrastructures must highlight the special role of the **life sciences**. But this area also highlights the varieties of e-Infrastructures, and technologies which are not e-Infrastructures but closely related to it (scientific publishing and reading, Wikis, and the semantic web) but which will intersect with e-Infrastructures. For example, as Renear and Palmer (2009) show, the rapidly increasing number of papers of biomedicine that researchers in this area must read means that new tools for annotation are required that allow researchers to structure and organize the information they have to cope with, include via shared annotation databases. Other examples come from the semantic web, such as the European 'Large Knowledge Collider' project (http://www.larkc.eu/).

Another key issue is the **management** of these large-scale and distributed projects. In the UK, the fact that not enough attention is paid to this (relative to the research aspect) in e-Science, and that distributed projects have additional challenges has resulted in a special project that is devoted to managing and ensuring the usability of e-Science Embedding e-Science Applications: Designing and Managing for Usability[36], which includes reports on how to manage these projects. It is foreseeable that project management will also need to be part of European-wide and global e-Infrastructure efforts.

Governance has already been discussed in terms of policy, but it is also important that there are a variety of bodies and roles, and larger projects typically have more complex forms, while in some small projects governance is 'lightweight'. It is not possible to say how strong – or laissez faire – these bodies are. There is also a range in the roles of governance, from guidance and steering to 'democratic representation' among various members or stakeholders. What we see organizationally is a range from highly centralized and hierarchical governance to more 'flat' or federated governance which may have multiple centres. A key point here, as elsewhere, is that 'not one size fits all' *(see Recommendation 9).* Furthermore, as technologies evolve, a key question will be how there can be greater integration or operation across a range of systems which will potentially bring definite advantages to users who will have access to this range of systems. Thus the question will be how they secure resources across this range as part of a complex workflow which will become the basis for research in many areas, especially for example in the life sciences.

One threat or opportunity, depending on one's point of view, in e-Infrastructures, as we shall see, is thought to be cloud computing. But another, perhaps more important, is the emergence of bottom-up **Web 2.0** (or 3.0) tools and datasets. Dutton and Meyer (2009), who surveyed e-Social Scientists, found that many social scientists built their own tools and datasets, often in idiosyncratic ways, to meet their particular needs and because no other tools and datasets were available to meet these needs. With the growing popularity of Web 2.0 or Wiki-style forms of collaboration, this type of tool and data development has become widely accessible. And social scientists are not the only ones engaging in this type of bottom-up activity, as the bioinformatics community (discussed in the previous paragraph) is also moving in this direction. Unless e-Infrastructures monitor, engage with, and either focus elsewhere or directly embrace these developments, this could lead to either of the scenarios in which there is little uptake (see below).

The **commercial sector**, and especially software providers, will play important role in future scientific developments. The Microsoft Report, 'Towards 2020 Science' (Microsoft 2006) highlighted a number of projects and future opportunities. And Microsoft (again, to pick just one prominent example, has a number of groups (Computational Science, Computational and Systems Biology, e-Science) that are all working in areas that are close to those that academic e-Infrastructures researchers are working in. It is important to recognize that these are

---

[36] See http://www.oerc.ox.ac.uk/research/embedding-e-science, last accessed on 25.9.2009

parallel efforts and that these commercial efforts will both compete (for example, in developing software for the annotation of scientific texts) and collaborate.

This brings us the role of 'clouds' and data. Here we can mention a project at Google called Data Liberation Front, which, according to the website[37], has as its mission, that: 'Users should be able to control the data they store in any of Google's products. Our team's goal is to make it easier for them to move data in and out.' The team leader, Fitzpatrick, when asked about how this idea will make money for Google, told an interviewee the following: Eric Schmidt, Google's chief executive…"He keeps telling us, the way to not be evil is to not lock users in," Fitzpatrick says. "He tells us, just get the users and we'll figure out how to make money."[38].This quote could equally serve e-Infrastructures, if we replace 'make money' with 'demonstrate the value of e-Infrastructures'.

Uptake is thus critical in the future, and reasons for successful e-Infrastructures development and uptake include:

- Low task uncertainty and high mutual dependence among researchers and research communities (see Whitley 2001). Notice however that this kind of technological and social organization does not fit all fields or disciplines, and may be 'alien' to some fields or disciplines.
- Strong social movements, as conceived in socio-technical interaction network (STIN) theories (Kling, McKim and King 2006), which enrols actors around common technological platforms and thus compatibility with other researchers, rather than a superior technology, play a key role.

---

[37] , http://www.dataliberation.org/
[38] http://www.guardian.co.uk/technology/2009/sep/09/google-data-liberation-export

5   Four Scenarios, with Two Dimensions

Against this background, we now present four scenarios for the future of e-Infrastructures developments:

- Scenario 1: Research Revolution
- Scenario 2: Winners and Losers
- Scenario 3: A Many-Headed Beast
- Scenario 4: European e-Infrastructures overtaken in the fast lane

It should be noted that the four scenarios are likely to be mixed in practice. However, separated analytically, they provide a way to think about different developments towards 2020. Moreover, we can map the difference between the four scenarios onto two dimensions: the vertical dimension is whether there is large or small uptake by virtual research communities, and the horizontal whether the impacts of e-Infrastructures are spread across all areas of technology and its effects on communities, or whether the effects are felt only in certain areas and not in others (or quite differently in different areas). This yields the following four quadrants in the diagram below:
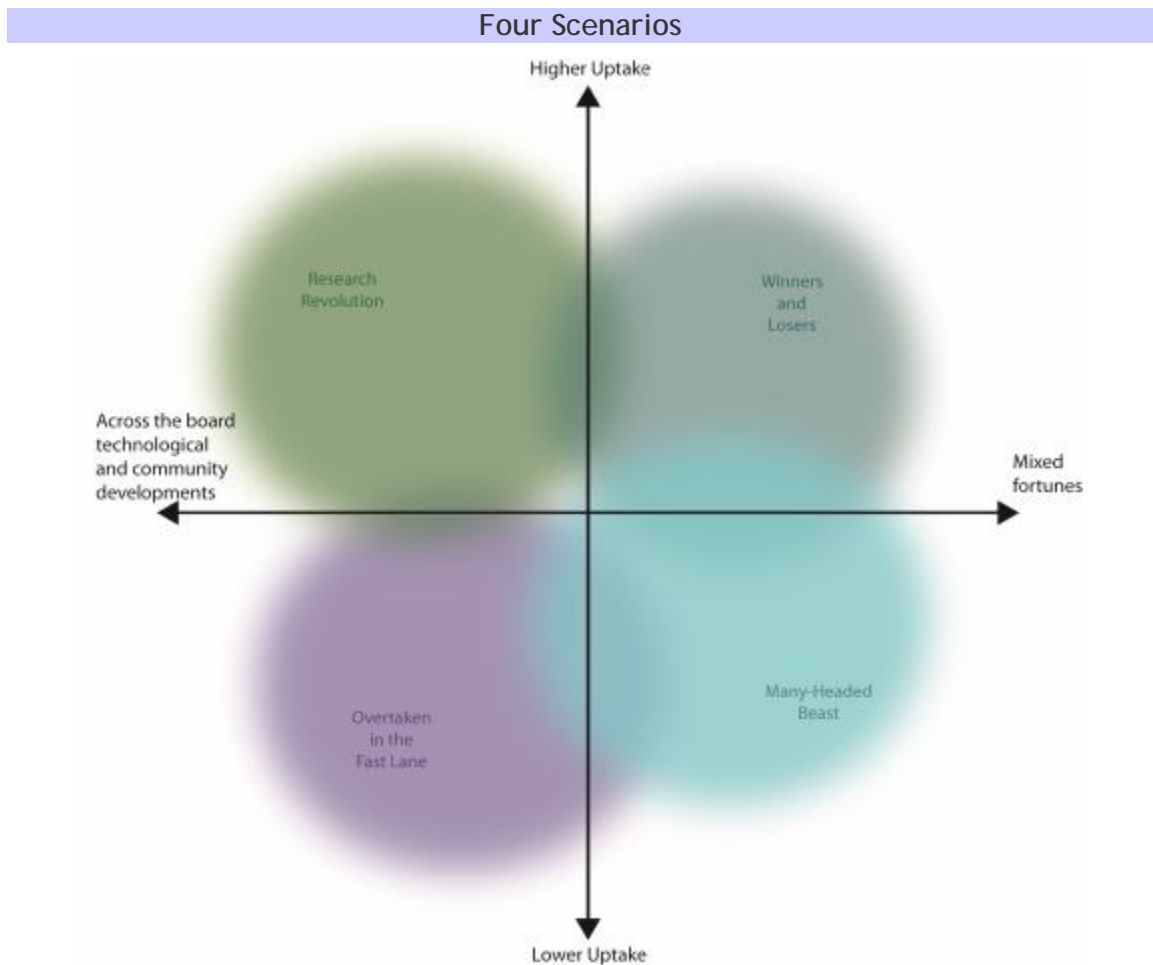
Four Scenarios

At this point we can highlight the key features of all four scenarios in more detail. Of course, these are not mutually exclusive, and may in fact overlap (as indicated by the overlapping circles in the diagram).

5.1   Scenario 1: Research Revolution

- Large-scale collaboration, data- and tool- intensive
- The nature of research is fundamentally transformed and carried out in distributed mode
- Change takes place across all disciplines and there is cross-disciplinary fertilization
- Change takes place on all levels of research (infrastructures, applications, daily practices) and all levels, including in schools
- Industry joins up with the research community and there are links to e-Government, e-Health and the public
- Public funding is complemented by private funding, an 'open science' ethos prevails

5.2   Scenario 2: Winners and Losers

- Some disciplines have strong uptake, succeed in creating strong communities, and move to new research questions
- Other disciplines have weak uptake, fall behind in creating collaborative communities, retreat into disciplinary silos
- Some disciplines and transdisciplinary communities mature rapidly, others don't get beyond planning
- Some fields gain via data- and resource-sharing, others are unable to benefit
- Winners move forward and e-Research supports collaboration and healthy competition in the field, losers are left behind

5.3   Scenario 3: A Many-Headed Beast

- Only certain fields develop e-Infrastructures - others concentrate on large facilities, still others focus on Web 2.0, e-Research is ignored in some areas – a plethora of directions
- Some areas duplicate efforts, in others there are no e-Research efforts or different directions
- A mixture of private and public funding, neither across the board, and funding is concentrated in pockets
- There are enormous disparities between sciences, social sciences, and humanities in funding (with little for humanities, even though there is much potential for cross-pollination with cultural heritage, educational outreach, and public access)
- A mixture of strong and weak research identities, large geographical variation, efforts separated by technologies and possibilities for collaboration

5.4 Scenario 4: European e-Infrastructures overtaken in the fast lane

- EU e-Infrastructures are overtaken by developments in the US and Asia, where there is more uptake of newer technologies other than e-Infrastructures

- Technological and social developments (clouds become a commercial Google or Amazon service in the US, petabyte libraries on mobile phones become common in Asia) overtake Grids, supercomputing and other research infrastructures – enabling computing-based research to move onto different terrain

- Data storage and compute resources become a commodity outside of research, so that shared public e-Infrastructures have little uptake outside universities

- Within research, e-Infrastructures investment atrophies

- Research quality and competitiveness in the EU suffers decline compared to Asian and US research

The scenarios allow us think about **different paths with different risks**:

- **Scenario 1:** Research Revolution seems least risky, but is likely to require the largest amount of funding and researcher effort. The benefits, for the research community and for society-at-large, are potentially enormous, but as with many innovations, it is possible that these benefits will only become realized after a considerable time. This is the main risk of scenario 1, which also entails that critical grand societal challenges (climate, energy, disease) that need to addressed will not be addressed quickly enough by an e-Infrastructure research revolution.

- **Scenario 2:** Winners and Losers represents risks for certain research communities rather than others. The benefits for some fields or disciplines will be balanced against the losses for others, so that researchers and society-at-large must for example bear the cost of lacking an e-Infrastructure that would provide cultural heritage while having one for particle physics, or vice versa – with all that this entails for the research community and the public.

- **Scenario 3:** A Many-Headed Beast, suffers from a different main risk; namely, that the benefits of coordination and potential synergies between research communities would not be realized. This would apply both to geographic spread and to spread within and between fields: some would be well-provided for (but without the possibility of linking to other e-Infrastructures since different technologies would not interoperate), others would be overprovided because of parallel efforts, and yet others would be left out altogether. One way to avoid this risk is to implement a policy whereby any funding allocated for infrastructure is granted on the condition that the e-Infrastructure must be open and must interoperate with other systems.

- **Scenario 4:** EU e-Infrastructures overtaken in the fast lane means that the research initiative passes to non-EU researchers. This includes newer technologies and possibilities of benefits to the wider society, and isolates Europe from the connected world of research outside the EU. The pay-offs from e-Infrastructures investment is not realized, and the status of European research declines in relation to that of other parts of the world.

The obvious solution is a balanced risks approach, which includes not only a mix which takes into account the heterogeneity that we have identified on a number of occasions, but also entails funding e-Infrastructures that provide a maximum of technological flexibility (for example, a mix of leading edge or 'moon shot' technologies with building on well-established paradigms, a mix of bottom-up approaches including Web 2.0 with top-down approaches where centralization and standardization may be required, and a mix of technologies which address specialized niches with services that operate wide-ranging communities with different skills levels) *(See Recommendation 10).* An additional point is worth making: the developing world has not been elaborated for all scenarios, but as the benefits for the South are likely to

be relatively larger than for the North, the risks of leaving out this part of the world are also disproportionate.

6   Conclusion

At this stage, we can make some recommendations that can contribute to getting to Scenario 1 rather than Scenarios 2, 3, and 4. A number of tools and mechanisms are available to the Commission and to research policy makers to enable this, which include:

- Obtaining **better knowledge** of the impact of e-Infrastructures on research communities. This roadmap has stressed that many e-Infrastructures are still in-the-making, and so evaluation of their **impact and engagement** may be premature. However, the UK has already had a number of projects to measure uptake and engagement (see, for example, the e-Infrastructures Use Cases and Service Usage Models http://www.eius.ac.uk/), and a number of interesting findings are already available. Such studies and reports will be highly valuable, but will need to take place on a European-wide basis and beyond (and not just for individual countries). These should provide recommendations on a range of more detailed issues than are contained in this report and roadmap, especially concerning data sharing, life sciences, governance, and the like.

- Is it possible to obtain **metrics** for 'customer satisfaction' among e-Infrastructures communities? How else can it be ensured that they are obtaining the benefits they need? And should such metrics be built into the business plans of e-Infrastructures, also to ensure their sustainability? *(See Recommendation 2).*

- many e-Infrastructures are fragile: promising early efforts may not be sustainable, and **resources** will be required to move into sustainable structures. There is diversity in models for future resources and there are many uncertainties, so much more planning is needed.

- More information and dissemination about **best practices** is needed. A small-scale study such as this one cannot tackle this – much more systematic analysis, and also simply bringing together expertise across disciplinary and domain boundaries (for example, from libraries studies or from the private sector, two areas which don't often intersect in academic or even research policy gatherings) – may be useful.

- **Standardization and harmonization**. This is especially applicable on the side of data and data sharing, but also in other domains (software). As we have seen, especially in relation to data collections that may not currently transcend national boundaries (for example, the policies of the Swedish National Data Service) or the boundaries of fields (say, between the life sciences and national offices of statistics), the ERA will need to engage in standardization and harmonization, perhaps not all at once – but beginning with a few countries where barriers may be low and then expanding and inviting other member states and beyond to join. (As an aside, it will be necessary to standardize data access and formats, because the heterogeneity of data is a major obstacle.) But such standardization and harmonization is also needed because users will have access to a range of e-Infrastructures, and will need to integrate them into their workflows and practices. Seamless access and integration will be a key requirement..

- As discussed above, **indicators and measurements** are both becoming more powerful with new techniques and the unique online visibility of e-Infrastructures enabling a gauging of the success of e-Infrastructures (outputs, user numbers, quality of research). A multi-sided approach to indicators, and also ongoing development of new techniques, is required, but also combination with existing techniques (in-depth qualitative studies, bibliometrics, and the like). In addition to recommending the deployment of these approaches, they need to be complemented by efforts on the part of e-Infrastructures to build such measurement (or at least the possibility of such measurement, by means of record-keeping and providing data) into their ongoing work and requests for further funding. One special indicator that can be used is the extent to which e-Infrastructures foster or further European integration.

- The Commission and other research policy making bodies can support e-Infrastructures with **legislation and regulation**. This relates not just to data-sharing, harmonization of laws abut data, and governance arrangements (ERIC), but also to issues such as cross-border information flows, for example of private or financial data. Here laws and regulatory frameworks can play an enabling role.

- The Commission and other research policy making bodies can **mandate certain policies,** for example in relation to data sharing in funded e-Infrastructures or projects that contribute to them, in relation to funding programmes and how they reward contributing tools and data, and the like.

The next generation of e-Infrastructures will partly be driven by new technological developments that cannot be foreseen (we have mentioned wireless, clouds, next generation high-performance computing and the like). These new technologies will, however, not overcome existing e-Infrastructures of various levels of maturity, which will rather be forced to adapt – to providing these newer technologies to users, and to become service-oriented while they do so. Put differently, e-Infrastructures should be 'technology agnostic' rather than locked in around a particular technology, and they will be able to best serve users, including future users, if they can quickly adapt to new technologies that serve their communities. What e-Infrastructures will do is to adopt the best technologies suited for their users. The responses to the Report and the Roadmap have stressed the commercial services that are on the horizon (clouds, web applications for researchers, collaboration tools), and these too will be adapted to and incorporated if these services provide the best solutions for user communities.

Risk management will focus on these risks concerning adaptability, but will more importantly anticipate how the research community's needs can best be met, with certain technologies and flexibilities, but also in relation to changes in access to data and tools, a changing research landscape with new demands or requirements.

It is too early for this report to say what effect ERIC will have, but it can be noted that this provides a means of overcoming the barrier that e-Infrastructures among member states should have a mechanism so that they ensure the long-term agreements to enable them to govern themselves (and also to avoid value-added tax like other national educational and research institutions can) across borders.

The public-private partnerships in the ERA promoted by the Commission are still in an experimental stage, but they may also deepen e-Infrastructures into society – especially since, as this report and roadmap have stressed, the embedding of e-Infrastructures within education (below the higher education level), in business, government and in the third sector – is bound to progress further in the coming years. These kinds of collaborations – across the public/private divide – will be necessary because the grand challenges that have been identified here – tackling disease, sustainable energy, aging populations – will require collaborations across different sectors of society and societal actors.

Ongoing e-Infrastructures will need more embedding in mechanisms for sustainability. To take just the example of EGI: what is the business model for EGI? What will be the assignments of responsibilities? What are the details of the resource contributions and forms of engagement with user communities so that they can be put on a long-term structural footing? For the sake of persistence, these will need to be ironed out. Ad hoc governance and structural models, which we found in a number of our case studies, need to be put on firmer foundation if the infrastructures are to have a long-term future.

6.1   Priorities in e-Infrastructures policy

e-Infrastructures are fostering European integration and also integration across the globe, deepening the ties among European researchers and undergirding their collaboration with

socio-technical systems on different levels. There are important economies of scale that can be achieved by means of this integration. Again, both European integration and economies are possible benchmarks for e-Infrastructure measurement. For example, each Euro that is spent on e-Infrastructures is also a Euro spent to achieve the integration of European institutions. Our survey also indicates that the impact in terms of collaboration with developing countries is especially important, and this needs to be a future priority. The same applies to continent spanning projects that link Europe to the South. Further, current e-Infrastructure development follows a sequential model of innovation – earlier involvement of domain researchers as drivers rather than pilot users should be encouraged.

The challenge of e-Infrastructures policy development is partly to recognize the diversity and the common issues in problem-solving across and the linking of disciplines. The various potential social, institutional and technical challenges to the formation of effective e-Infrastructures collaborations do not pose uniformly serious obstacles or impinge with equal severity upon all branches of scientific inquiry. Similarly, the potential transformative impacts of an enhanced e-Infrastructure are not likely to be felt equally across all the domain sciences and emerging interdisciplinary fields. Gaining a better sense of the policy priorities will enhance the support of global research communities as e-Infrastructures become more complex and at the same time critical to the quality of research outputs as well as to productivity. But to measure productivity, a crucial input is measurement **and indicators**, which are becoming ever more important in the world of research in general. Importantly, it is possible to measure e-Infrastructures not just by means of citations, but also using webometrics and measuring online 'visibility' (Park, Meyer and Schroeder 2009). This type of measurement will be critical to future policy in this area. Of course, these measures of impact are only one way to gauge the importance of e-Infrastructures. Others include measures of uptake and distribution both geographically and across domains. Further, one might look at the degree to which computing resources are utilized and data is stored and used. Our conclusions and discussion of some priorities thus lead to a number of recommendations.

6.2   Recommendations for e-Infrastructures Policy Action

1.   European and other researchers will increasingly depend on the most technically and socially advanced e-Infrastructures for research, to compete in a more globally competitive world and meet increasingly urgent challenges. **e-Infrastructures development**, which underpins the future of meeting these challenges, should be a key priority for policymakers.
2.   **Sustainability** has already emerged as a key issue, and should be considered in a much longer-term perspective. Whether resources are sustained at the national or EU or other level, they must be committed for extended (10+ years) periods in order to be made an integral part of research planning so this commitment provides a reliable and well-integrated platform for the research community and beyond.
3.   The uncertainties around **funding** are the single-largest perceived barrier among providers, virtual research communities, and the yet-to-be-engaged. Clearer plans and funding agendas could overcome these uncertainties.
4.   While **data** is not scarce, having data available as needed and in formats that benefits the widest possible communities is still a major challenge. The key challenge has moved on from being the 'data deluge' to being the coordination, proper safeguarding, sharing and re-use of data beyond its initial purposes. Mandating clear policies to share software and make data interoperable in re-usable ways are essential.
5.   There are currently few **rewards** for researchers both inside communities and among providers for their contributions to e-Infrastructures development, or for sharing data and tools. Reward mechanisms need to be promoted that recognize and reward researchers to do this.
6.   'Openness' has been a much vaunted principle in e-Infrastructures development, but while open source software and open publishing can already show successes, much more by way of

coordination is needed to apply openness to standards and interoperability in systems and collaboration platforms.

7. **Governance** and metagovernance (governance which coordinates the governance of individual efforts) strategies are still emerging in many ad hoc forms. Although ERICs are emerging as a possible single legal mechanism for the future, there is still a great deal of uncertainty among the e-Infrastructures communities, and this is where there are many policy mechanisms at the EU level especially that can be put in place to overcome this uncertainty.

8. Researchers young and old are rapidly changing the ways they search for and access information and data. **Education and training** efforts for e-Infrastructures have lagged behind e-Infrastructures development, but offer an excellent route for much more widespread engagement with the novel research possibilities opened up by e-Infrastructures, and should thus be among the highest priorities in future planning and funding.

9. Though there is currently a series of different, sometimes overlapping efforts in a variety of **disciplines**, and no 'one size fits all' model should be imposed, e-Infrastructures should be open to supporting all fields and subfields as well as collaborations between and among them. Many opportunities for shared best practices and for sharing resources are currently unexploited in this respect, and could be fostered by more funding that favours cross-disciplinary teams and efforts.

10. Future efforts must also focus on generating and enabling completely **novel applications** to problems in which distributed computing has not yet been applied. These fairly high risk actions should be complemented by support for existing productive and mature e-Infrastructures, so that a flexible balance (flexible, among other things, in being subject to constant monitoring revision) is achieved. One specific action point here might be dedicating a Future and Emerging Technologies (FET) call specifically to such novel areas.

11. **Standards** are becoming critical, not just in software but also in the interlinking and accessibility (metadata) of data. Standardization in some requires a balance with flexibility, but otherwise, the more open and interoperable e-Infrastructures remain in relation to communities, to new technologies and towards other e-Infrastructures and tools and data, the better. In this case, again, mandating standards will often be useful.

12. **Indicators** of success and impact and quality are required in view of the need for coordination and resource planning that has been highlighted in this report. Powerful new tools for measurement are becoming available, and high priority should be given to providing resources for projects which undertake such measurement. e-Infrastructures should also be mandated to implement means whereby such indicators and measurement is facilitated by the e-Infrastructures themselves and by research from outside the e-Infrastructures to enable comparison.

13. It needs to be ensured that any barriers to participation by **industrial research partner** participants are removed. The effective exploitation of e-Infrastructures offers many potential benefits both to firms with sizeable R&D organizations and to SMEs, and the 'red tape' and other barriers such as lack of open standards need to be minimized.

14. **Further research** into the bottlenecks, effectiveness, and future potential of e-Infrastructures will be a vital for their effective governance and for policymaking. While it is a somewhat clichéd recommendation to call for more research, e-Infrastructures – as a relatively novel, still protean, and absolutely vital platform for research in the ERA and beyond – is still largely unexplored territory in terms of its social dynamic, especially in relation to Recommendation 12. Such research has enormous potential pay-offs.

7   References

Atkins, D. E., Droegemeier, K. K., Feldman, S. I., Garcia-Molina, H., Klein, M. L., Messerschmitt, D. G., et al. (2003). Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure: National Science Foundation.

AVROSS Accelerating Transition to Virtual Research Organization in Social Science (2008). M4 Final Report. Brussels: European Commission. Retrieved 08.07.2008, from: http://web.fhnw.ch/plattformen/avross/papers-and-prensentations/final-report/.

Axelsson, A.-S. & Schroeder, R. (2009). Making it Open and Keeping it Safe: e-Enabled Datasharing in Sweden and Related Issues, forthcoming in Acta Sociologica.

Barjak, F. (2006). Research productivity in the internet era. *Scientometrics, 68*(3), 343-360.

Borgman, C. 2007. *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. Cambridge MA: MIT Press.

Bos, N., Zimmerman, A., Olson, J., Yew, J., Yerkie, J., Dahl, E., et al. (2007). From Shared Databases to Communities of Practice: A Taxonomy of Collaboratories. *Journal of Computer-Mediated Communication, 12*(2), 652-672.

Council of the European Union (2009). The Future of ICT research, innovation and infrastructures: Adoption of Council Conclusions. Council of the European Union Document No. 16128/9, available at http://europa.eu/documentation/official-docs/index_en.htm

Dutton, W. H. and E. T. Meyer (2009). "Experience with New Tools and Infrastructures of Research: An exploratory study of distance from, and attitudes toward, e-Research," *Prometheus*, vol. 27, no. 3, pp. 223-238, 2009.

Edwards, P.N., Jackson, S.J., Bowker, G.C. & Knobel, C.P. (2007). *Understanding infrastructure: Dynamics, tensions, and design. Report of a workshop on 'History & theory of infrastructure: Lessons for new scientific cyberinfrastructure'.* http://www.si.umich.edu/InfrastructureWorkshop/documents/UnderstandingInfrastructure2007.pdf.

e-Infrastructure Reflection Group (e-IRG) (2009). *e-IRG White Paper*, available at http://www.e-irg.eu/index.php?option=com_content&task=view&id=40&Itemid=39 (last accessed Sept.20, 2009).

e-Science Directors' Forum Strategy Working Group (2009). 'Century of Information Research (CIR): A Strategy for Research and Innovation in the Century of Information'. Prometheus, vol. 27, no.1: 27-45.

Etzkowitz, H., & Leydesdorff, L. (2000). The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university-industry-government relations. *Research Policy, 29*(2), 109-123.

European Commission. (2007). *GREEN PAPER. The European Research Area: New Perspectives.* Retrieved 24. April 2007, from http://ec.europa.eu/research/era/pdf/ era_gp_final_en.pdf.

European Commission. (2003). *Third European Report on Science & Technology Indicators 2003 - Towards a knowledge-based economy.* Brussels: European Commission.

Funtowicz, S., & Ravetz, J. (1993). Science in the post-normal age. *Futures, 25*, 739-756.

Galvin, R. (1998). Science Roadmaps. *Science*, 280, 803.

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (1994). *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies.* (1 ed.). London; Thousand Oaks; New Delhi: Sage.

Hey, T., Tansley, S. and Tolle, K. (eds.). *The Fourth Paradigm: Data-Intensive Scientific Discovery.* Redmond WA: Microsoft Research.

Kostoff, R. & Schaller, R. (2001). Science and Technology roadmaps. *IEEE Transactions on Engineering Management*, 48(2), 132-143.

Leydesdorff, L., & Etzkowitz, H. (1997). A Triple Helix of University-Industry-Government Relations. In H. Etzkowitz & L. Leydesdorff (Eds.), *Universities and the Global Knowledge Economy A Triple Helix of University-Industry-Government Relations* (1 ed., pp. 155-162). London: Pinter.

Leydesdorff, L. & Wagner, C. (2007). I sthe United States losing ground in science? A global perspective on the world science system in 2005. In *Proceedings of ISSI 2007, Volume 1.* 11th International Conference of the International Society for Scientometrics and Informetrics, CSIC, Madrid (pp. 499-507).

Leydesdorff, L., & Zhou, P. (2005). Are the contributions of China and Korea upsetting the world system of science? *Scientometrics, 63*(3), 617-630.

Meyer, E. and Schroeder, R. 2009a. The World Wide Web of Research and Access to Knowledge, Knowledge Management Research and Practice (7): 218-33.

Meyer, E. and Schroeder, R. 2009b. Untangling the Web of e-Research, Journal of Informetrics (3): 246-60.

Microsoft (2006) 'Towards 2020 Science', available at http://research.microsoft.com/en-us/um/cambridge/projects/towards2020science/background_overview.htm (last accessed 25.9.2009)

Narin, F., Stevens, K., & Whitlow, E. S. (1991). Scientific collaboration in Europe and the citation of multinationally authored papers. *Scientometrics, 21*, 313-323.

National Science Board. (2004). *Science and Engineering Indicators 2004* (1 ed.). Arlington, VA: National Science Foundation.

Nentwich, M. (2003). *Cyberscience. Research in the age of the Internet* (1 ed.). Vienna: Austrian Academy of Science Press.

Nelson, Bryan. (2009). Empty Archives. *Nature*, vol. 461, issue 7261, 160-63.

Olson, G.M., A. Zimmerman, and N.Bos. (eds). (2008) *Scientific Collaboration on the Internet.* Cambridge: MIT Press.

Park, H.W., Meyer, E. and Schroeder, R. 2009. Mapping Global e-Research: Scientometrics and Webometrics', *Proceedings of the 5th International Conference on e-Social Science*, 24-26 June 2009, Cologne.

Renear, A. and Palmer, C. 2009. Strategic Reading, Ontologies, and the Future of Scientific Publishing. *Science.* 24: 828-32.

Special issue 'Data Sharing' (2009), Nature, 461: 145 (10 September).

Thies, Annika. 2009 The EC legal framework & how will it work in the future, presentation at the EGEE'09 Conference, Barcelona 22 Sept. http://indico.cern.ch/contributionDisplay.py?contribId=187&sessionId=67&confId=55893 (last accessed 29 Sept. 2009)

Trans-European Research and Education Networking Association (TERENA). (2007). *TERENA Compendium of National Research and Educational Networks in Europe. 2007 Edition.* Amsterdam: Terena.Whitley, R., 2000. *The intellectual and social organization of the sciences.* 2 ed. Oxford: University Press.

Wouters, P. (2002). *Policies on Digital Research Data – An International Survey.* Amsterdam: NIWI-KNAW.

Wouters, P. & Beaulieu, A. (2006). Imagining e-science beyond computation. In C. Hine (Ed.), *New Infrastructure for Knowledge Production: Understanding E-Science* (pp. 48-70). Hershey: Idea Group.

Wouters, P., & Schröder, P. (Eds.). (2003). *Promise and Practice in Data Sharing*. Amsterdam: NIWI-KNAW.

Wuchty, S.; Jones, B.; and Uzzi, B. (2007)..The Increasing Dominance of Teams in Knowledge Production, *Science*, vol.316, pp.1036-1039.

PART 3 – WORKSHOP REPORT

1   Workshop Report

Workshop report of the final study workshop, held in Brussels, Avenue de Beaulieu 25 on 24 February 2010.

Agenda:

- Welcome (Jean-Luc Dorel, European Commission and Simon Robinson, empirica)
- BELIEF and the BELIEF Brainstorming event in the afternoon (Stephen Benians, BELIEF)
- Keynote presentation by John Wilbanks, Vice President Science, Creative Commons
- Introduction and Study Approach (Tobias Hüsing, empirica)
- Empirical results of eResearch2020 case studies and surveys (Franz Barjak, FHNW)
- The eResearch2020 Roadmap (Ralph Schroeder, OII)
- Panel feedback and plenary discussion
- Concluding remarks (Kostas Glinos, Head of Unit "GÉANT & e-Infrastructure")

Panelists:

- Chair: Simon Robinson (empirica)
- Steven Newhouse, EGEE Technical Director, CERN
- Matthew Scott, General Manager, DANTE
- John Wilbanks, Vice President Science, Creative Commons
- Paul Wouters, Programme Leader, The Virtual Knowledge Studio for the Humanities and Social Sciences – VKS
- Rapporteur BELIEF: Stephen Benians, BELIEF

Contributions:

Jean Luc Dorel (European Commission) opened the meeting, describing the aims of the study and showing a video presentation of eInfrastructure activity in the EU. The keynote speech by John Willbanks (VP Science, Creative Commons) was well received. John contrasted the slick consumer product iPhone, with content controlled by Apple and closed to developers, with the openness of a PC connected to the Internet. In the open world, one in a hundred applications may succeed, failures - screwing it up - are an essential part of the creative process, or generativity. For eInfrastructure in research, generativity is key. Openness includes the freedom to screw it up.

Tobias Hüsing (Empricia, study team) outlined the study approach, the steps taken, quality measures and the interactioni with the Study's Network of Interested Parties and Steering Committee. Franz Barjak (FHNW, study team) presented results from the case studies and user survey. Franz made it quite clear that the user survey was not to be misinterpreted as a representative study of any research community, and agreed with comments from the floor that though the vast majority of respondents had declared that eInfrastructure was important or essential in their work, this could not be taken as a measure of demand for eInfrastructure in the research community. Nevertheless for those users covered, withdrawal of the facilities would constitute a serious setback in their work. Ralph Schroeder (Oxford Internet Institute, study team) presented the scenarios developed in the Study and the 14 recommendations

made for further action by the European Union in the field of eInfrastructure. In response to a question from the floor, Ralph clarified that the dimensions used - degree of uptake and uniformity of uptake across disciplines - were the result of a brainstorming approach.

In the panel discussion Matthew Scott (DANTE, Cambridge) welcomed the study as an interesting and valuable investigation bringing key issues and challenges to the fore. From the viewpoint of a representative of GEANT, he added some observations and recommendationss with regard the 'supply' side of e-infrastructure, namely in terms of accessibility, reliability, affordability and engagement with users. With regard to accessibility, digital divide issues have to be resolved to maximise research potential across Europe and attention needs to be paid to openness in terms of complexity, interoperability, consistent user interfaces and common access policies. Reliability means technical (24/7 production quality, meeting also exceptional demands) as well as financial (researchers invest time and build their academic career on these) and governance reliability. With regard to affordability, Matthew raised the question whether financing should be fully centrally funded or (partially) funded by users at point of use. Different solutions may be applicable to this.

Matthew pointed to some open questions, such as measurability and metrics and which data collection tools are there, such as 'webometrics' and which advantages and disadvatnages accrue to these. Experience has shown that statistics can be misleading and biased. Matthew warned to be over optimistic about public private partnerships mentioned in the study, as lessons learned at National levels has in some cases lead to a more sober assessment of its potentials. With regard to the request to remove of barriers to participation by industrial research partners he Matthew is concerned that it needs a decision making process as to who participates in what ways and whether a precondition might be to make results available in public domain. Matthew closed by the question whether longer term investments in e-infrastructures are actually to be seen as a high risk strategy or not rather as a pre-requisite to achieve the 'Research revolution'.

Steven Newhouse (EGEE Technical Manager, CERN, Geneva) •made the point the e-infrastructures will go the same road as the web did. 15 years ago, it was not necessary in the sense of being an indispensible requirement to be connected to the web. Today, it may not yet be necessary to be connected to e-Infrastrucutre for a researcher, but it is likely that it will be the case in the future. The DCI community today provides an easy and simple to use computing environment, but the result has not become yet the "iPhone" of research; it is (still) focussed on computer-literate people; now the question is what readjustments need to be made to reach this iPhone-status; the "creative chaos" that currently persists is necessary to reach this stage

Paul Wouters (VKS) welcomed the study as a good pilot exercise and starting point for a robust, large scale and representative study of eInfrastructure use in the research community. He called for investigation of research communities and their requirements for eInfrastructure which was representative - including non-users - and longitudinal, to capture trends. Paul welcomed the study findings as a valuable contribution to understanding users of eInfrastructures, but strongly criticised the method used in the study to generate scenarios, saying this fell short of state of the art in this field. Simon Robinson (empirica, chair) picked up on the theme of usability from Matthew's presentation, pointing out that there was tension between the idea of carefully designed applications and the unpredictability of new knowledge. John took this up and commented that research increasingly requires programming skills, pointing to the success of change in approach by MIT to focus on skills in programming networked applications. Simon asked Paul to respond from a social science and humanities perspective to John's suggestion that research using eInfrastructure should have access to programming skills. Paul pointed to the continuing difficulty of dialogue between social scientists and computer scientists, and thought this was not going to change.

Kostas Glinos (Head of Unit "GÉANT & e-Infrastructure", European Commission, DG Informaion Society) wound up the meeting, thanking all participants and declaring that the recommendations the Study has made were already being reflected in planning for the next call for proposals under FP7. The study brochure would be presented at the interministerial conference in Barcelona the following month.

Documentation:

The presentation slides can be downloaded at http://www.eresearch2020.eu/workshop/index.php#page=program.