Accelerating Transition to Virtual Research Organization in Social Science (AVROSS): First Results from a Survey of e-Infrastructure Adopters

Citation for published version (APA):
Barjak, F., Wiegand, G., Lane, J., Kertcher, Z., Poschen, M., Procter, R., & Robinson, S. (2007). Accelerating Transition to Virtual Research Organization in Social Science (AVROSS): First Results from a Survey of e-Infrastructure Adopters. In *Third International Conference on e-Social Science*

Published in:
Third International Conference on e-Social Science

Accelerating Transition to Virtual Research Organization in Social Science (AVROSS): First Results from a Survey of e-Infrastructure Adopters

Franz Barjak[1], Gordon Wiegand[1], Julia Lane[2], Zack Kertcher[2], Meik Poschen[3], Rob Procter[3], Simon Robinson[4]

[1] University of Applied Sciences Northwestern Switzerland

[2] National Opinion Research Center at the University of Chicago (NORC), USA

[3] National Centre for e-Social Science (NCeSS), UK

[4] empirica Communication and Technology Research, Germany

Email address of corresponding author: franz.barjak@fhnw.ch

Abstract. This paper presents the first results of a survey and case study analysis of early adopters of e-Infrastructure across the world, focusing on continental Europe, the UK and the USA. Information was gathered on a number of possible influences on the adoption and level of usage of e-Infrastructure in the social sciences and humanities, and comparisons are made by region. We find that the most important influences affecting whether and where e-Infrastructure is adopted in these disciplines is the availability of the necessary qualified staff and of the required funding are the most important influences on whether and where e-Infrastructure. There are indications that adoption could be accelerated were the most eminent scientists from the social sciences and humanities to more widely promote the benefits to research of e-Infrastructure adoption to their colleagues in the discipline. Last but not least, we see different types of projects in continental Europe, the UK and the US.

Introduction

The AVROSS study is being carried out in response to a European Commission call to report on the factors contributing to e-Infrastructure adoption in social science and humanities research in several fields. The aim of the study is to identify and analyze a selected number of the most promising applications of e-Infrastructure in these disciplines, those which have triggered transition to virtual research organizations and motivate sustained e-Infrastructure use. Special attention is paid to opportunities for computer-supported collaborative learning (CSCL). The ultimate goal is to provide recommendations to the Commission based on scenarios for large scale deployment of technologies and applications supporting virtual research organizations and novel services for students based on CSCL.

'Soft' science disciplines are acknowledged to have a key role to play in promoting e-Science uptake across all research (Edwards et al., 2007). However, these disciplines have to date not been the fastest adopters of e-Science techniques, especially in the case of advanced grid-

based e-Infrastructure (Catlett, 2006; Vanneschi, 2005). Our recommendations to EU policy-makers are intended to identify measures which the Commission and Member States could employ to begin to change this situation, to help promote e-Science in Europe in these disciplines and identify clear requirements for e-Infrastructure development. The policy recommendations can be expected to be adaptable to e-Infrastructure promotion in several other disciplines with related requirements, notably e-Health.

Research approach

Our research follows a social shaping of technology (SST) approach, which has proven its value in a number of science and technology studies, in particular in studies of the adoption of information and communication technologies in an academic environment (Williams & Edge, 1996; Kling & McKim, 2000). We use a broad conception of the notion of the social and have adopted a framework centred on four groups of factors:

- Technological frames and user requirements: Technological paradigms of developers and users, which are shaped by the capabilities of previous technologies and the demands of user communities, constitute frames which shape the introduction and spread of e-Infrastructure. "Innofusion", the concurrent realization of innovation and diffusion (Fleck, 1988), and aligning technical capacities and surrounding conditions, the particular culture, habits, customs and organizational setting of fields in the social sciences and humanities (Fry, 2004; Wouters & Beaulieu, 2006) are key issues.

- Scientific shaping of technology: Further progress in computer science and computer linguistics has been vital and is still required to produce usable and functionally adequate applications for the social sciences and humanities, as well as for solving problems of confidentiality and privacy, which are particularly acute in disciplines working with data on individual people. However, agendas of computer scientists and domain scientists are not always easy to reconcile (Lawrence, 2006).

- Economic factors: In addition to funding needed for the development of e-Infrastructure, an economic viewpoint highlights the influence of the e-Infrastructure adoption process on research output (e.g. research papers) and input (e.g. time needed to learn to use a new application). Other critical economic issues are missing solutions for assigning scientific credit and ownership rights for shared datasets, code and other tools, and the resulting disincentives to producing those (Woolgar & Coopmans, 2006).

- Political influences: The activities of political institutions and intermediaries – these are, for science, in particular, government departments of research and higher education, research foundations and scholarly societies – shape the spread and use of e-Infrastructure.

Our study has to date employed two empirical methods: an exploratory survey of e-Infrastructure adopters and enthusiasts, and eight case studies of e-Infrastructure development and use in social sciences or humanities. This paper draws mainly on the results of the survey.

Exploratory survey of early e-Infrastructure adopters

The AVROSS email survey was carried out in the spring of 2007 using a snowball sample of over 2,000 individuals identified as interested in and potentially involved in e-Infrastructure work. The aim was to cast the net wide, and ensure a useful number of responses. Of the 560 responses received, the UK and the US each made up one third; one fourth came from continental Europe and under 10% from other countries (in particular Australia, Canada, and New Zealand).

The survey collected information in five main areas: (1) the respondent's background, organization, and experience with e-Infrastructure; (2) the respondent's current or most recent e-Infrastructure project; (3) background about funding and results; (4) the respondent's views of catalysts and barriers to the development and implementation of e-Infrastructure projects; and (5) further e-Infrastructure projects and people who might be able to provide interesting information. For the present paper we selected some of the key findings on projects and on individual decisions to adopt e-Infrastructure.[1]

For the purposes of the study we adopted the e-infrastructure definition promoted by the e-Infrastructure Reflection Group (e-IRG), a multinational group that coordinates e-infrastructure activities in Europe. The e-IRG defined e-infrastructure as *integrated ICT-based research infrastructure* (Leenaars, Heikkurinen, Louridas, & Karayannis, 2005). Key elements include networking infrastructure, middleware and organization and various types of resources (such as supercomputers, sensors, data and storage facilities). Though the definition includes "old" components like supercomputers, the World Wide Web or e-mail, it requires that these be part of an integrated system before they become part of e-infrastructure. The basic requirement for any component is that it should be able to exchange information through a standardized interface or protocol.

Catalysts and barriers to e-Infrastructure adoption

The survey included a set of questions on past, current and future involvement in projects using e-Infrastructure. Those who were not current users of e-Infrastructure were classified as 'interrupters', 'drop-outs', 'future users' or 'non-users' depending on their past use and any expectations of beginning or resuming use. Current users are the largest group, making up nearly half of respondents. Interrupters, drop-outs and future users have shares of between 5% and 7% and non-users add up to about a quarter.

Survey respondents were asked to rate the importance of a number of sources of information from which they may have learned about e-Infrastructure and its applications. The category receiving most ratings of high importance was that of other scientists in their discipline or with whom they collaborated (Table I). Printed information is of comparatively little importance for the majority of scientists; those who did find importance were primarily those who collaborate at national and international, level rather than locally. This group might be less integrated in their local scientific communities and printed matter may substitute for local meetings and workshops.

Table I. Sources of information about e-Infrastructure (in % of responses)

Source	Very important	Somewhat important	Neutral	Somewhat unimportant	Not at all important
Meetings or workshops which provided information on e-Infrastructure	29.0%	29.0%	20.8%	9.7%	11.6%
Infrastructure or administration people at your own organization	31.6%	28.2%	13.6%	11.2%	15.5%
Infrastructure or administration people from other organizations	32.4%	38.1%	17.1%	4.3%	8.1%
Journal, magazine, or other printed or electronic information source	13.2%	30.4%	26.5%	12.7%	17.2%
Other scientists, colleagues, or collaborators	54.5%	32.9%	9.4%	1.9%	1.4%

Source: AVROSS early adopter survey.

---

[1]   A full report on the survey results can be obtained from the authors upon request.

Most early adopters in the survey agreed that a number of factors were important catalysts for e-Infrastructure adoption and use, in particular, seed funding, collaboration, and interesting research (Table II). Although organizational incentives were rated important by fewer scientists, responses to open-ended questions show that organizational environments, and the more general academic environment in a field, do play an important role in adoption and successful project realization. Around 30 percent of respondents who answered the open-ended questions mentioned these environments, for instance:

"Tool development is not particularly well-regarded within the social sciences – embarking on tool development is a risky career move, … A further problem with tool development in the social sciences (if you are also pursuing an academic career) is that you can be pigeon-holed as a 'technician' or technical support officer for your non-technical social science colleagues who are going to be using the tools. …"

"Senior leaders in most fields tend to look backward and value the modes of inquiry that shaped their own thinking while in graduate school."

"Lack of interest in developing humanities based digital projects on the part of administrators and colleagues at my home institution."

Table II. Catalysts for e-Infrastructure adoption (cases)

Catalyst	Very Important	Somewhat Important	Neutral	Somewhat Unimportant	Not at all Important
Seed funding from an outside agency	100	40	16	5	12
Seed funding from home institutions	59	52	26	16	18
Organizational incentives within your institution	44	53	38	11	22
Collaboration	117	45	14	3	0
Observation of successful projects in other areas	42	70	37	11	7
The computational requirements of your research	54	54	38	11	13
Contribution to interesting research expected	94	54	22	0	3
Support for teaching activities	26	49	42	29	25
Emerging standardization of available tools	39	60	33	22	14

Source: AVROSS early adopter survey.

Few differences were found between categories of respondent or project. Seed funding is seen as more important in the US and in "other countries" than in the UK, and least important in continental Europe (Figure 1). The computational requirements of research, on the other hand, are more important in the latter region.
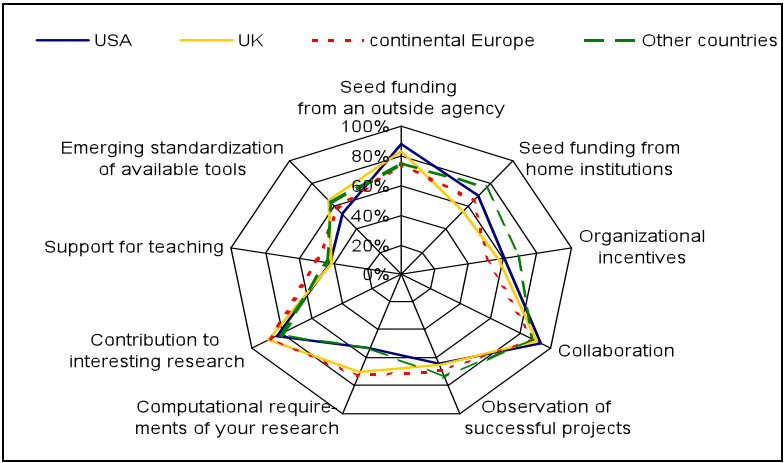

Figure 1. Catalysts for e-Infrastructure adoption by country
(% of respondents who considered this catalyst at least somewhat important).

The respondents identified a number of key barriers to e-Infrastructure adoption. Three factors were consistently rated most important regardless of geographic region, length of project, or date of adoption: lack of funding, costs, and lack of qualified staff (Table III). Budgetary issues referred to problems of obtaining long-term funding, inflexibility in managing funds, and larger development costs than expected, among others. On the role of staff, one respondent remarked:

> "Fund staff!! Applied projects which succeed best have paid committed staff. In Social Sciences, there are many social and methodological issues which are barriers to using even the current data networking technologies for research. We need recurrent funding for research assistants to engage with the research community and foster new ideas. We also need recurrent funding for archivists to help researchers use the technology. Equipment without the staff and expertise to run it is wasted."

This is in line with US and UK scientists' substantial concern about sufficient numbers of trained individuals for the full exploitation and maintenance of e-social science investments.[2] Already two years ago, the e-IRG proposed to increase efforts in the training of scientists and computer support personnel on working with grid environments (Leenaars, et al., 2005).

Table III. Barriers to e-Infrastructure adoption (cases)

Barrier	Very Important	Somewhat Important	Neutral	Somewhat Unimportant	Not at all Important
Lack of initial funding	77	55	16	12	10
Costs associated with e-Infrastructure dev.	66	70	22	8	6
Lack of information about usefulness	31	47	39	23	23
Lack of staff available to help with development	57	60	26	15	12
Insufficient applicability of existing technology to social science research problems	35	45	36	24	27
Problems with intellectual property rights	17	52	41	31	28
Lack of trust in sustainability	27	50	38	22	30
Problems with protecting confidentiality of data	22	47	43	27	30
Locked into other technologies	14	35	51	26	32

Source: AVROSS early adopter survey.

Interrupters and drop-outs were directly asked why they stopped or interrupted their involvement with e-Infrastructure. As Table IV reveals, the reasons cited most often were those concerning lack of funding or lack of staff.

Table IV. Importance of reasons for interrupting or ending participation in humanities or social science e-Infrastructure projects

	Very important	Somewhat Important	Neutral	Somewhat unimportant	Not at all Important	All valid N
Lack of sustainability of funding	31.4%	40.0%	11.4%	8.6%	8.6%	35
Lack of staff available to help with development and deployment	21.1%	39.5%	21.1%	2.6%	15.8%	38
Not enough scientific pay-off	13.5%	21.6%	29.7%	16.2%	18.9%	37
Technology was not mature enough	11.1%	22.2%	30.6%	16.7%	19.4%	36
Other reasons	60.0%	20.0%	10.0%	0.0%	10.0%	10

Source: AVROSS early adopter survey.

---

[2] Unpublished summary reports NSF/SBE cyberinfrastructure workshops Sept 18, 2004 and Oct 22, 2004; survey results from ESRC review of NCeSS hub, 2005.

To reveal time effects, we differentiated between (very) early adopters, those whose first e-Infrastructure involvement was before 2000, and (relatively) late adopters, those who had started in 2000 or later. Most barriers were rated similarly by both groups and so can be seen as stable over time. However, three barriers seem to have increased in importance, receiving more ratings of importance by later adopters compared to earlier adopters. Two of these barriers of growing importance, "lack of information on usefulness" and "insufficient applicability to research problems", might be a reflection of the growing proportion of scientists who are exposed to existing e-Infrastructure and in this way prompted to reflect on its relevance to their work. The more prevalent importance of "data confidentiality" among later adopters possibly reflects increasing awareness of these issues.

Another important issue, addressed by more than half of the respondents to open-ended questions, is the involvement of and interaction with users in the development and deployment of e-Infrastructure. The following quotes from the responses illustrate this:

"Keep users involved in all stages and find 'champions' among domain scientists"

"Leadership must come from members of the domain community (e.g., a humanities or social science faculty member) -- and not from a computer or computational scientist. Relying on CI centers (e.g., NCSA or SDSC) only engenders 'learned helplessness.' It is better to adopt less ambitious technology that can be controlled/customized by humanities/social scientist users then to depend on the latest thing from the centers (which produces a state of dependency)."

"Keep it practical and applied. Developing a tool is applied work for the community, it is NOT your ticket to a long ride on the academic granting gravy train. People who use these programs to advance their academic career rather than produce robust tools in a timely manner are destroying some schemes. Equally, technologists who have little idea about what researchers need are responsible for many expensive projects which are never used."

"Don't wait for the tool to be 'perfect' - get using it for research as soon as possible because the development of the tool should obviously be in the context of particular research projects. The tool is useless if it isn't being used to generate research outcomes that are being published in respected social science journals."

e-Infrastructure projects

We found a great deal of variation in the number of e-Infrastructure projects in which an individual had been involved: 72 had been involved in just one project, 51 in two, 32 in three, 18 in four, 5 in five, and 66 in more than five such projects.

There are some regional differences in length of experience with e-Infrastructure (Figure 2). Most strikingly, US participants are on average more experienced than their colleagues from other regions, with an average over more than 10 years experience and more than 4 projects. Despite the fact that there are currently numerous e-Infrastructure projects in the UK, the relatively recent nature of this phenomenon is evidenced by the fact that the typical e-Infrastructure user has a relatively short experience with e-Infrastructure, and has worked on relatively few projects.

We found that research foundations and councils were the dominant source of funding across the board. The median project was initially funded at just over 335,000 Euros; the median annual budget was just over 108,000 Euros. The projects in continental Europe and the USA are larger than projects in the UK, both with respect to funding and staff (see Figure 3). The scheduled funding period also differed among the regions, with continental European projects lasting the longest at an average of 60 months, compared with 47 months in the USA, 26 months in the UK, and 35 months in the rest of the world.
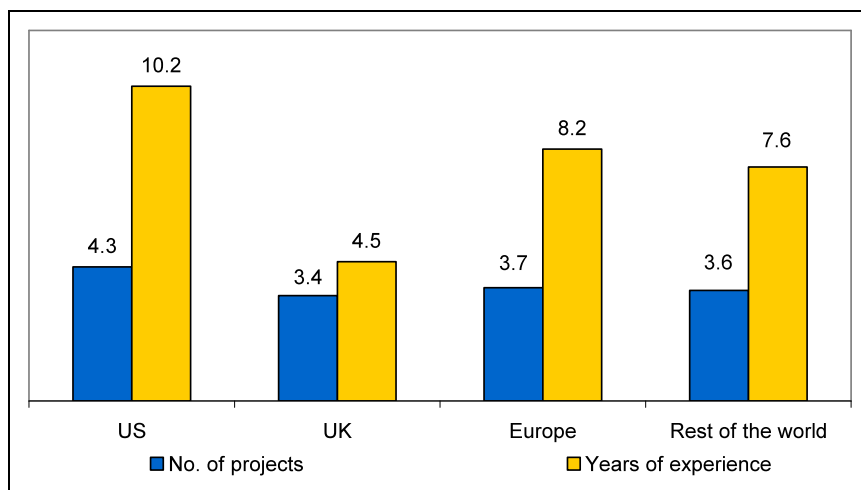
Figure 2. Experience in e-Infrastructure projects by region of the respondent (arithmetic means).
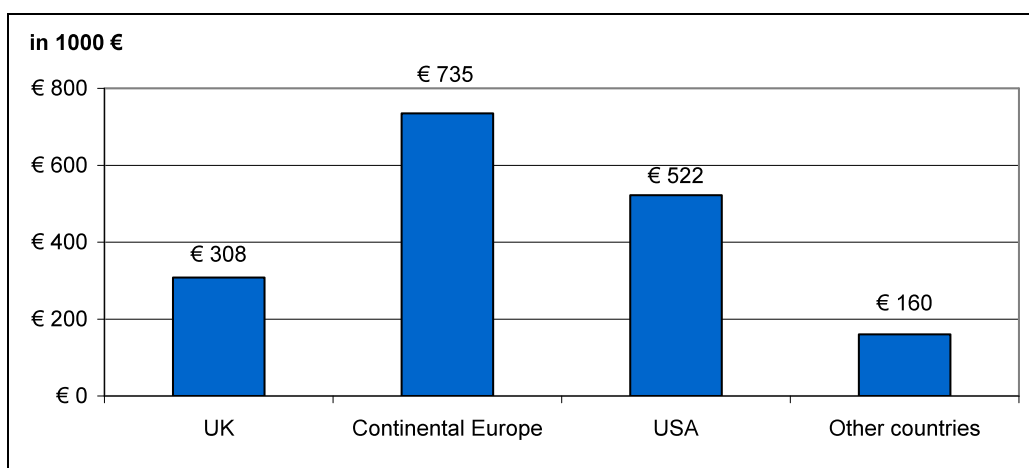

Figure 3. Initial funding of the projects in Euro (median values).

Our respondents also provided information about the number of people working on their projects. Again, this differed by region of the respondent (Figure 4). US projects tend to be quite large: on average 20 people - 7 scientists, 5 graduate students and a substantially larger number of non-scientific staff than their European counterparts. The UK projects reported are quite small, averaging around 10 staff, with 4 scientists and just 1 graduate student. The continental European respondents reported average staff sizes of 15 - 8 scientists and 4 graduate students.

The e-Infrastructure items used most frequently by the projects are communication and collaboration tools, distributed data storage, and high bandwidth networks. High performance computing, which is a feature of e-science in other disciplines, is not as important, nor are innovative data collection methods. Some level of variation is visible by country (Table V): learning environments and virtual/3D environments play a larger role in US-based projects; continental European projects more often use large data repositories or distributed data. Videoconferencing is used more than twice as often in UK-based projects compared to their continental European counterparts.
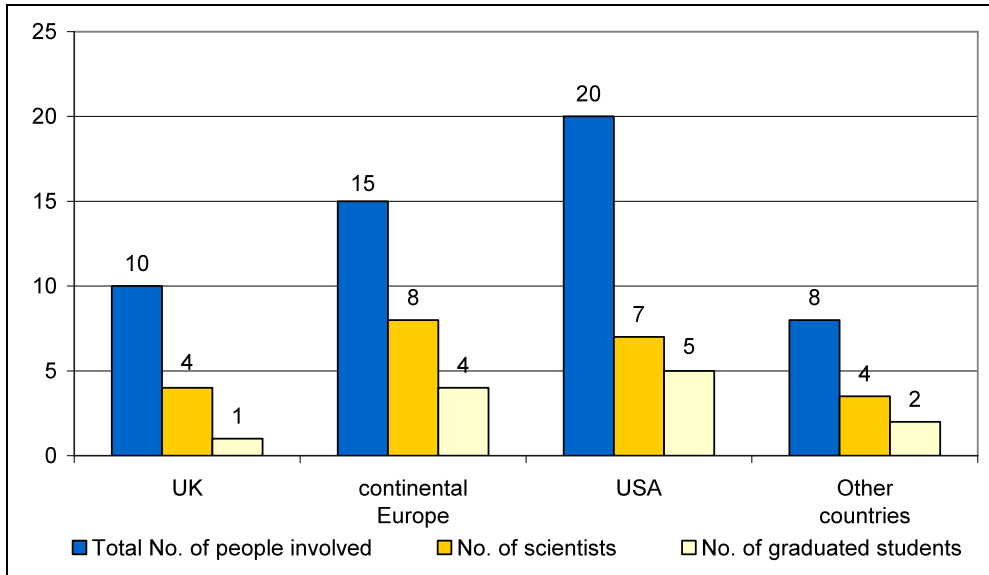
Figure 4. Size of the projects grouped by regions (median personnel data).

Table V: Use of e-Infrastructure technological items by countries

	USA		UK		Continental Europe		Other countries	
	N	%	N	%	N	%	N	%
High performance computing	30	45%	23	39%	18	38%	6	38%
High performance communication	40	62%	27	46%	22	45%	12	71%
High bandwidth	50	76%	32	53%	40	77%	11	65%
Distributed data, data repository	49	75%	54	82%	50	93%	14	82%
Collaboration tools/systems	59	83%	51	77%	47	84%	16	89%
Learning environments	34	53%	22	36%	23	45%	5	31%
Grid-enabled videoconferencing	23	37%	24	44%	10	21%	7	44%
Virtual/3D environments	15	24%	9	18%	8	18%	2	13%
Innovative data collection methods	18	45%	14	39%	15	43%	8	53%

Source: AVROSS early adopter survey.

When asked about the main outcomes of the projects they had been involved in, these were reported to be publications (148 cases), new methods (129), new data (114), follow-on collaborations (143) and new tools (143). The prevalence of different outputs varies little by country or region, but publications and new methods were less often reported as an outcome of projects in the 'other countries' (Canada, Australia, New Zealand etc.), new data and new collaboration less often for UK projects.

Of the 180 who answered the question on user constituency of the project, 129 reported that there was a constituency for their work, 58 reported that there was not. Interestingly, almost all user constituencies that are reached are reached by projects which include participants from the same discipline as the user constituency (Table VI). There are a number of possible interpretations of this intriguing result. It could be that projects are developed by researchers in given disciplines because they have specific disciplinary needs in mind. It could also be that researchers in a project already have a dissemination network in place that is discipline specific, and that knowledge about the project is transmitted through such disciplinary networks.

Table VI: The interaction between project disciplines and the disciplines of user constituencies

	Proportion of identified project disciplines with constituency in same discipline*	Proportion of constituencies identified with the same discipline as the project**
Humanities		
Archaeology	50.0%	72.2%
Art (arts, history of arts, performing arts, music)	57.1%	70.6%
History	47.8%	66.7%
Languages and literature (excluding linguistics)	54.3%	70.4%
Linguistics (including computational linguistics)	44.4%	74.1%
Philosophy, ethics, religion	31.3%	55.6%
Other Humanities	38.5%	51.7%
Social sciences		
Economics and business	31.1%	70.0%
Educational sciences	50.0%	60.0%
Law	33.3%	40.0%
Political science	35.1%	52.0%
Psychology	40.0%	46.2%
Social and economic geography, regional science	48.4%	72.1%
Sociology	45.8%	70.2%

* Read as follows: 50% of the projects with agricultural scientists on the team had also agricultural science as user constituency.

** Read as follows: 72.2% of the projects with agricultural science as the user constituency also had agricultural scientists on the team.
Source: AVROSS early adopter survey.

Conclusions

What can we learn from these preliminary findings on the adoption of e-Infrastructure in the social sciences and humanities in general and in different regions? Three findings seem noteworthy.

1. Funding and staff issues are of key importance all across the board, no matter whether we look at continental Europe, the UK, the US or other countries. Seed funding is widely seen as necessary to investigate and develop new ideas. There are several institutional barriers in science which stand between an idea for innovative e-Infrastructure and its receiving funding: notably, lack of interest and lack of understanding on the part of decision makers at the field-level, in local universities, and in funding bodies.

   Staff issues are of similar importance, as e-Infrastructure in the social sciences and humanities is at least as much about technology as it is about robust and small scale services that support the use of the technology. There continue to be few advanced computer users among social scientists and humanities researchers; these domains appear particularly reluctant to change functioning work routines and see the pay-off to the investment of time needed as uncertain.

2. The ability of e-Infrastructure to support collaboration and the benefits it provides to research outcome should of itself draw social scientists into using e-Infrastructure. However, it is likely to be necessary to publicize these advantages better in the scientific community. The survey identified other scientists in the same discipline as possibly the most effective source of information about e-Infrastructure, suggesting that encouraging

highly regarded scientists in social science and the humanities to adopt and promote the adoption of e-Infrastructure in their discipline will be an important mechanism towards more widespread adoption. We also found that projects more easily connect to a user community when scientists from that discipline are active participants in the project, a finding consistent with our technological frames approach.

3. The current structure of e-Infrastructure involvement in the social sciences and humanities differs between continental Europe, the UK and the US. US scientists have the longest experience with e-Infrastructure and have some of the largest projects in terms of funding volume and team size. Continental Europe appears to be catching up with the scale of US projects and the UK e-social science program currently encompasses relatively small projects. The UK policy seems to be more in line with the finding that social scientists and humanities researchers may be more likely to seek involvement in small projects deploying practical tools which are easy to master, along with arrangements to support established work routines, than in large-scale projects demanding entirely new ways of doing research. It is difficult to be sure at this stage whether the small-scale or large-scale strategy for promoting e-Infrastructure uptake in the social sciences and humanities will prove the more successful.

Acknowledgments

References

Catlett, C. (2006): '*The State of TeraGrid - A National Production Cyberinfrastructure Facility.*' Retrieved 29 September, 2006, from http://www.teragrid.org/about/docs/ StateOfTeraGrid-June2006.pdf.

Edwards, P. N., Jackson, S. J., Bowker, G., & Knobel, C. P. (2007): 'Understanding infrastructure: Dynamics, tensions, and design. Report of a workshop on 'History & theory of infrastructure: Lessons for new scientific cyberinfrastructure'. Retrieved 29 July, 2007, from http://www.si.umich.edu/InfrastructureWorkshop/documents/ UnderstandingInfrastructure2007.pdf.

Fleck, J. (1988): 'Innofusion or Diffusation? The nature of technological development in robotics.' Edinburgh PICT Working Paper No. 7, Edinburgh University.

Fry, J. (2004): 'The Cultural Shaping of ICTs within Academic Fields: Corpus-based Linguistics as a Case Study.' *Literary and Linguistic Computing*, vol. 19, no. 3, pp. 303-319.

Kling, R., & McKim, G. (2000): 'Not just a matter of time: Field differences and the Shaping of Electronic Media in Supporting Scientific Communication.' *Journal of the American Society for Information Science*, vol. 51, no. 14, pp. 1306-1320.

Lawrence, K. (2006): 'Walking the Tightrope: The Balancing Acts of a Large e-Research Project.' *Computer Supported Cooperative Work (CSCW)*, vol. 15, no. 4, pp. 385-411.

Leenaars, M., Heikkurinen, M., Louridas, P., Karayannis, F. (2005): '*e-Infrastructures Roadmap.*' Retrieved 23 November, 2006, from http://www.e-irg.org/roadmap/eIRG-roadmap.pdf.

Vanneschi, M. (2005): *'Survey of Activities in Universities and Research Labs. Deliverable D.3.1.2 of GridCoord.'* Retrieved 2 Oktober, 2006 from: http://www.gridcoord.org/ grid/portal/information/public/D.3.1.2_V.1.2_181105.doc.

Williams, R., & Edge, D. (1996): 'The social shaping of technology.' *Research Policy*, vol. 25, no. 6, pp. 899.

Woolgar, S., & Coopmans, C. (2006): 'Virtual Witnessing in a Virtual Age: A Prospectus for Social Studies of E-Science' In C. Hine (Ed.), *New Infrastructure for Knowledge Production: Understanding E-Science* (pp. 1-25). Hershey: Idea Group.

Wouters, P., & Beaulieu, A. (2006): 'Imagining e-science beyond computation.' In C. Hine (Ed.), *New Infrastructure for Knowledge Production: Understanding E-Science.* Hershey: Idea Group.