Accelerating Transition to Virtual Research Organization in Social Science (AVROSS): the Results of a Study for the EC

Franz Barjak[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

Though it is clear that social sciences and the humanities have both much to gain and a key role to play in promoting e-Infrastructure uptake across the disciplines, they have to date not been quick to adopt advanced grid-based e-Infrastructure. Our recommendations to EU policy-makers point the way to changing this situation, promoting e-Infrastructure development and application in these disciplines, with expected impact in areas with related requirements. The empirical part of the study had two components. The first of these was an exploratory survey of early adopters of e-infrastructure. The second was an investigation of eight case studies.

Survey Analysis

The first component consisted of a web-based survey sent to early adopters of e-Infrastructure in the social sciences and the humanities. The intention was to provide a stock-taking of e-Infrastructure projects in Europe and beyond, and was sent in the spring of 2007 to roughly 2,000 individuals who had been identified as potentially involved in e-Infrastructure work. The aim was to cast a very wide net to generate the maximum number of responses, and over 560 responses were received - 448 usable responses (23.4% of the sample).
The survey yielded several striking findings. One set of findings concerned respondent perceptions about the factors leading to e-Infrastructure adoption in social sciences and humanities. The second set of findings concerned the structure of e-Infrastructure projects carried out so far.

Respondent perceptions

First, survey respondents stressed the importance of other scientists in spreading information. Printed information is of comparatively little importance, and is only important for those scientists predominantly collaborating at the non-local, national and international, levels.

Second, the respondents highlighted a number of key catalysts for infrastructure adoption: seed funding, interesting research, and collaboration. Seed funding was identified as more important in the US and in other countries than in the UK, and least important in continental Europe. The computational requirements of the research, on the other hand, are more important in the latter regions.

Third, the respondents identified a number of key barriers to e-Infrastructure adoption. Three factors were uniformly most important, regardless of the discipline and experience of the respondents: lack of funding, costs, and lack of qualified staff.

Fourth, the survey revealed that any funding strategy should take the needs of users and other stakeholders into account. Early adopters frequently remarked that community-building is a critical factor in a successful e-Infrastructure project. The ability of a project to connect to a user community appears to be easier when that discipline is also represented in the project. User feedback should be sought early; indeed, it may be advisable to ensure that tool development be user-led in order to secure the uptake of the results.

Fifth, it is important to have a supportive institutional and scientific environment, namely, supportive local IT staff and university administrations, as well as broader support in the disciplinary domain.

Finally, the technological limitations of e-Infrastructure include their reliability and user-friendliness. These tend to be exacerbated if the service models of computing services are deficient, particularly if the models' technical solutions are inflexible, there is a lack of openness to software revisions and information exchange and a failure to incorporate mutual learning across e-Infrastructure projects.

Project structure

We found that research foundations and councils were the dominant source of funding across the board. Projects in continental Europe and the USA are larger than projects in the UK, both with respect to funding and staff.

The most frequently used e-Infrastructure items included communication and collaboration tools, as well as distributed data, and required high bandwidth. High performance computing, which is a feature of other sciences, was not as important, nor were the innovative data collection methods. Some level of variation was visible by country of the project: learning environments and virtual/3D environments play a larger role in US-based projects. Continental European projects more often contain data repositories, whereas videoconferencing is relatively unimportant – it is used more than twice as often in UK-based projects. The items varied also by project length: virtual/3D environments were of notably higher relevance in long-term projects, lasting for three years or longer. This is consistent with a view that the provision of interfaces for learning and practice becomes more important when the development phase is completed and the actual user involvement gets more and more critical.

Respondents reported a variety of outcomes from their projects, including publications, new methods, new data, follow-on collaborations, and new tools. They also reported a very broad user constituency. Interestingly, almost all disciplinary constituencies that are reached are reached by a project that includes participants on the team with the same discipline as the user constituency.

Case Study Analysis

The second empirical contribution of AVROSS was based on an analysis of eight case studies identified as promising e-Infrastructure projects in the social sciences and humanities. Data on the cases was obtained through semi-structured interviews with developers, principal investigators, and users of the infrastructure; in addition, both published and internal project material was obtained from different sources such as the interview partners, project websites or other sites containing project descriptions and presentations. As with the survey, several striking findings emerged.

The first was the importance of technology. The main technological problems in the cases resulted from concerns about data security and reliability of the technology. Protecting new types of data and controlling access to it required the developments of new tools and applications. The second key technological constraint was the often negative experiences of (pilot) users when using the applications. These negative experiences resulted from complex user interfaces (UI), unstable applications, and difficulties in integrating existing applications and standards into the new environment. However, lack of computational power was not a major barrier. Interviewees from the case studies remarked that it is very difficult to align the approaches to computing followed by social scientists with those followed by high energy physicists or other more developed Grid user communities. Since the SSH approaches are ingrained in field-specific cultures and practices, SSH researchers are typically unwilling to adjust their practices in order to use the grid.

The second set of findings concerns key challenges. Most projects had difficulty in developing a user community. Only two of the eight projects currently have large user communities. The strategies for recruiting users are rather weak and underdeveloped: projects tend to rely on what is offered by their funding or institutional environment. A related challenge was the difficulty of developing sources of sustainable funding: For example, while the Access Grid project has been successful, a key ingredient to this success seems to be that the service is offered free or close to free of charge for the users. However, in contrast to the survey, the case study projects did not identify the recruitment of staff as a major issue.

The third set of findings centred around the disruptive nature of the technology. Some cases found difficulty in relating their work to existing practices and the culture within their field. Data privacy and access restrictions, high costs of producing metadata and making data usable for third parties were particularly problematic. A related challenge was the difficulty of finding appropriate academic compensation for tool developers, data producers, or contributors to methodology, since the assignment of academic credits and rewards for such tasks is not common in SSH. However, the case studies suggest that the problems of communication and collaboration across disciplinary borders was more sophisticated, possibly as these were more pressing and disrupting project progress.

The final set of findings had to do with the actual impact on research and teaching which has been, by and large, rather modest. This can partially be explained by the fact that some of the projects are either pilot projects or still in an early phase of development. However, this cannot hide the fact that making a measurable impact on the field is actually one of the main challenges for any e-Infrastructure project in SSH. Since publications are the key output measure in SSH as in other academic domains, and because data sources and tools are still rather neglected in research publications, it will be difficult to prove the impact through this channel. The teaching and learning activities are not very well developed in any of the investigated projects. Graduate students were mentioned as users in some projects; however, except for one project, they do not receive special attention, for instance through courses that teach the use of the infrastructure.

Policy recommendations

Although the survey was not scientific in nature, and case study analyses are inherently illustrative rather than definitive in scope, the study did identify several key issues which could be used for policy purposes. If policy is to be focused on increasing e-infrastructure adoption, the study supports the following measures (among other which will be included in more detail in the paper):

- Step up the role of e-Infrastructure in graduate education.
- Involve users at all stages: conceptualization, design and development, diffusion.
- Support the development of service-oriented business models.
- Invest in addressing the data confidentiality concerns that inhibit adoption.
- Institute a research program on e-Infrastructures in SSH.

It is not within the remit of the study to provide recommendations about the merits of a pan-European strategy versus a nationally based strategy to promote e-infrastructure adoption. However, we do note that a European approach could be extremely useful if it could: support international exchange and collaboration; transfer competence among countries; reduce problems of interoperability and coordination costs; achieve economies of scale; deal more effectively with collaboration extending beyond European boundaries and the academic sector; and create a joint technological space for the social sciences and humanities.