

Report: Workshop on European Infrastructure for Repositories of Scientific Information: 8—9 June 2006, Brussels

CJ Wright

University of the Witwatersrand, Johannesburg

This workshop was arranged by the European Commission DG INFSO F3 in order to obtain input from the scientific and information science community into the FP7 planning process.

The workshop was attended by more than 100 participants from 27 countries: 16 EU member states, 4 participating in the Framework Programme, and 7 outside Europe. The 15 presentations were followed by lively and intensive debate. Besides myself, there was one other South African registered delegate, who unfortunately failed to arrive. Two other African delegates attended, both from Congo.

EC Background

The FP7 programme will cover the seven-year period 2007—13. The draft FP7 budget for e-infrastructures is €650 million. The work-plan for research infrastructures is to be discussed at the Management Committee on 7 July 2006, and input from this workshop could inform the shaping of the work-plan. Areas where proposals for relevant actions were invited, are:

- models of data sharing
- stakeholder interests
- cost models
- common approaches to standards, protocols and interfaces
- policy level actions to encourage development of digital repositories and sharing of data.

Cross-border research is of growing importance in the EU, in both scientific and economic terms. Under FP5 and FP6 there has been a very successful deployment of e-infrastructure at the network and grid layers. Under the FP7 roadmap, an additional *data layer* will be built on these foundations. Of particular relevance to the workshop was the proposal that this layer would deploy and support data repositories for scientific communities.

Aims of the Workshop

The aim of the workshop was to discuss the development of Scientific Digital Repositories as part of the European e-Infrastructure. Scientific Digital Repositories form an essential component of the Knowledge Infrastructure envisaged under FP7 and are of growing strategic importance to the concept of a Single Information Space within the European Research Area. The workshop was asked to consider the following aspects:

- Prospective views on Knowledge Infrastructures and their impact in different scientific areas of research and education.
- Review of different sides of a Knowledge Infrastructure: technological, operational and organisational and in the context of purpose and usage.
- Lessons learned from different initiatives in the EU member states (and other regions) and how these could help in tailoring pan-European initiatives.

Programme

Welcome and introduction: *Kyriakos Baxevanidis*, Deputy Head of Unit, DG INFSO F3, EU Commission

Session 1: Looking into the future—vision on the future of knowledge infrastructure. *Chair: Kyriakos Baxevanidis*, Deputy Head of Unit, DG INFSO F3, European Commission

Miron Livny, Professor, Computer Sciences Department, University of Wisconsin

Jens Vigen, Head of the CERN library

Herbert Van Der Sompel, Digital Library Research and Prototyping Team, Research Library, Los Alamos National Laboratory

Session 2: User and technology requirements—analyse the cycle of e-Science and model future Data Repositories in the light of its requirements. *Chair: Liz Lyon*, Director, UKOLN

Ms. Simone Rieger, Max Planck Institute for the History of Science

Francoise Genova, Director, CDS (ULP-CNRS)

Hans Pfeiffenberger, Alfred Wegener Institut

Wolfgang Genztsch, Coordinator of D-Grid

Session 3: Current best practices—best practices in Member States and research communities as well as possible approaches to pan-European knowledge infrastructure. *Chair: Juergen Renn*, Max Planck Institute for the History of Science, member of ESFRI group

Leo Waaijers, SURF, Platform Manager, Project Driver / *Peter Doorn*, Director, Data Archiving and Networked Services (DANS), The Hague

Liz Lyon, Director, UKOLN

Norbert Lossau, University Librarian, Bielefeld University

José Fernandes, Foundation for National Scientific Computing, Biblioteca on-line, Portugal

Session 4: Drafting the key messages for FP7 and policy actions

Moderators of the debate : *Yannis Ioannidis*, Professor, University of Athens and *Donatela Castelli*, Researcher, CNR - ISTI

Session 5 - Wrap-up and conclusion. *Chair: Ulf Dahlsten*, Director, DG INFSO F, European Commission
Presentation of the conclusions - Rapporteur *John Martin*, ENPG Secretary

Conclusions and Recommendations

The workshop received presentations on several successful digital repository deployments in Europe. These span a number of thematic areas and demonstrate the federation of institutional repositories into a national knowledge infrastructure. These working repositories show how scientists, scholars, and others can benefit from bringing the results of research and scholarship into an integrated system of online holdings.

Whilst these repositories contain very large holdings, often many hundreds of thousands of digital objects, and whilst a great diversity of multimedia objects—documents, datasets, images, video, publications, transcriptions, reference works, etc—are included, they represent only a tiny fraction of what will be possible in a pan-European knowledge infrastructure of the future.

The workshop identified a number of success factors which characterise well-regarded, operational digital repositories: engagement of users; participation of users as both contributors and consumers; user perception of the repository as an environment to do research; organisation and support of an appropriate professional standard; handling of rights issues.

Many technical and organisational aspects of the digital repository infrastructure are complex and under continual development. This is not to say that successful implementations are not possible today but, rather, that much development work is envisaged to achieve the goals of a services-rich infrastructure covering all disciplines and regions. Aspects identified: business models; access control; copyright; organisation; adding content; motivating users; meta-data; persistent identifiers; harmonisation; standards; longevity; sustainability.

In the context of the EU Framework Programme 7, which envisages a pan-European knowledge infrastructure as part of the future e-infrastructure, a number of specific recommendations were made:

1. The pan-European knowledge infrastructure should be built on existing national and trans-national knowledge infrastructures.
2. Encouragement should be given under FP7 to creating EU thematic communities who can participate in the building of those parts of the knowledge infrastructure relevant to their research.
3. Impetus should be given to the development of the knowledge infrastructure under FP7 by launching two 'grand challenge' projects in contrasting thematic areas to demonstrate the benefits of creating highly functional and richly populated knowledge infrastructures for research.
4. Within the FP7 programme, interoperability should be encouraged or mandated for all digital repository development and deployment.
5. The design of the knowledge infrastructure should avoid any sharply defined boundaries. It should be easily federated with other knowledge infrastructures in other parts of the world, and it should be easy to open its benefits to other potential user areas such as e-health, e-learning, e-government and others.
6. Future knowledge infrastructure should be sustainable.
7. Consideration should be given by the EU to making a declaration in support of the principle of Open Access.

Relevance to South Africa

Large scale digital data repositories are becoming important in South Africa. These occur in different manifestations: data intensive high performance computing; curating, managing accessing, analysing large complex data objects are but two examples. This EU initiative could serve as a role model to a future South African initiative. Already Meraka has noted the need in this regard.

The European pathway prior to FP7 has been to:

- agree on co-operation,
- establish fast, robust networks across Europe;

- establish regional high performance computer centers comprising supercomputers, grid and other clusters
- create human resource and software infrastructures to allow transparent computing across the network
- expand the high performance network to the US, South America and into the East.

Under FP7, besides expanding the current infrastructure, it is intended to put in place the infrastructure to permit effective curatorship, harvesting, exploitation and accessibility to/of digital data repositories.

Currently, South African networks and other infrastructure are not capable of supporting activities at the same intensity. However the recent 2006 budget announcement by the Minister of Science & Technology that:

- “...one of our flagship projects is the development of a high-speed broadband network in the region. This is the UbuntuNet project, which we will link through the South African National Research Network to Europe, giving South Africa and its research community a high-speed network...”
- “...we will establish the Centre for High Performance Computing. This national resource will support a diverse base of researchers and scientists, and facilitate the collaboration and multidisciplinary approach needed to solve today’s complex computational problems. Its research objectives are to provide high-end computing and computing expertise for all research in South Africa in natural science, medicine, engineering and social sciences.”

provides a foundation for a future South African large scale digital data repository plan. Further, the South African grid computing fraternity continues to grow.

It is evident then that South Africa is following the same path—not necessarily with the same intensity—as Europe. Because of the ESASTAP program and its links with FP7, it is feasible for South Africa to:

- learn from the European experiences;
- benefit directly by linking into the European infrastructure;
- obtain support for a South African data repository initiative.

Acknowledgement

I must thank the Department of Science and Technology for having afforded me the opportunity to attend this important workshop.

CJ Wright
29 June 2006