THE ESFRI ROADMAP AND ITS DEMANDS ON THE E-INFRASTRUCTURE

K Jeffery

Science and Technology Research Council, Rutherford Appleton Laboratory, Harwell Science and Innovation Campus, Chilton, Didcot, Oxfordshire OX11 0QX, UK
Email: keith.jeffery@stfc.ac.uk

ABSTRACT

The ESFRI Roadmap marked a turning point in the evolution of European thinking on research facilities, providing a catalogue of such facilities with their characteristics. In parallel, the ESF (European Science Foundation) completed a questionnaire-based survey of research facilities. Finally, the ERF (European Research Facilities) consortium representing national facilities with international access was formed to parallel EIROForum (the European laboratories funded by international subscriptions). It is becoming increasingly clear that management of these facilities and management of the research process require extensive ICT: for research managers that is provided by CRIS (Current Research Information Systems) and to give researchers additionally access to facilities to control experiments with associated modelling and simulation and access to research datasets and software.

1 INTRODUCTION

Research facilities have long provided researchers with the ability to 'see further and deeper' and thus push back the frontiers of human knowledge. As the large research facilities become more expensive to construct and operate, there have been two effects:
a) national governments have agreed to contribute to one or more facilities not in their own country but available to their researchers;
b) national governments have agreed to allow access to their national research facilities by researchers from other countries.
The extreme form of this is when a large number of national governments agree (usually via a treaty) to the joint construction and operation of an international facility open to all their researchers.

2 ESFRI

ESFRI, the European Strategy Forum on Research Infrastructures, is a strategic instrument to develop the scientific integration of Europe and to strengthen its international outreach. The 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.

The mission of ESFRI is to support a coherent and strategy-led approach to policy-making on research infrastructures in Europe and to facilitate multilateral initiatives leading to the better use and development of research infrastructures, at the EU and international levels.

ESFRI's delegates are nominated by the research ministers of the member and associate countries and include a representative of the commission, working together to develop a joint vision and a common strategy. This strategy aims at overcoming the limits due to fragmentation of individual policies and provides Europe with the most up-to-date research infrastructures, responding to the rapidly evolving science frontiers, advancing also the knowledge-based technologies and their extended use.

ESFRI defines research infrastructures as:
"They are facilities, resources or services of a unique nature that have been identified by pan-European research communities to conduct top-level activities in all fields."

There are essentially two classes or kinds of research infrastructure facilities considered:

EiroForum: European facilities (i.e. funded as Europe) E.g. CERN, ILL, ESRF, EMBL, ESA…. and

ERF: National facilities with international access / participation e.g. DESY/XFEL, RAL/ISIS…

ESFRI does not consider small, individual facilities, e.g., at a university or restricted to national use although any e-infrastructure must take account also of these facilities.

A typical research infrastructure facility is a large instrument that provides a method (e.g., through neutrons, electrons, photons) of observing directly natural phenomena (e.g. telescopes) or experimenting with matter to understand their properties (e.g. structure of proteins in cells).

Increasingly, however, led by the thinking behind the UK e-Science initiative (Jeffery, 1999a), research requires:

- Observation or experiment with observation;
- modelling or simulation; and
- comparison of the results.

Moreover, the research commonly is performed not by one individual in one place but by a team of individuals geographically distributed from each other and from the facility (although some representatives of the team may attend the facility to perform the experiments).

This leads to the need for ICT support for:

- cooperative working / intercommunication between researchers;
- interactive distance management of experiments on facilities (this happens most obviously on spacecraft);
- data collection, storage and preservation;
- data analysis and interpretation;
- production of scholarly publications and their distribution;
- scholarly discussion;
- innovation leading to patents and products; and
- synthesis and summarisation leading to inputs to research policy and strategy.

Clearly there is here a research workflow analogous to that in production industry or commerce. An advantage of workflow is that information is gathered incrementally and previously collected information is not re-entered but re-used thus gaining both efficiency and accuracy.

3 ESFRI ROADMAP

ESFRI was initiated following a conference in 2000 and subsequent declarations. It really got under way with the statement from the Competitiveness Council in 2004:
"In the context of developing research infrastructures of European interest, the Council of the European Union welcomes the development of a strategic roadmap for Europe in the field of research infrastructures and the role of the European Strategy Forum on Research Infrastructures (ESFRI) in this context. This roadmap should describe the scientific needs for Research Infrastructures for the next 10-20 years, on the basis of a methodology recognised by all stakeholders, and take into account input from relevant inter-governmental research organisations as well as the industrial community. The Council stresses that this roadmap should identify vital new European research infrastructures of different size and scope, including medium-sized infrastructures in the fields of humanities and bio-informatics, such as electronic archiving systems for scientific publications and databases, covering all scientific areas as well as existing ones that need to be upgraded."

The 2006 roadmap (ESFRI, 2006) covered existing facilities, policies, size, capabilities and planned facilities, policies, size, and capabilities (including upgrades). It essentially provided a catalogue of facilities existing or planned in near-time.

The 2008 (updated) roadmap (ESFRI, 2008) stressed research infrastructures in new areas such as energy, biological and medical science, and environmental science. However the roadmap also states (p11):

"e-Infrastructures are critical to all projects in this roadmap. The e-infrastructure aspect of each project has therefore been assessed in much greater detail than before."

4 E-INFRASTRUCTURE

The e-Infrastructure can be envisaged at several levels as in the original GRIDs architecture (Jeffery, 1999a) namely:
1. a lower computation and data layer consisting of the network, computational facilities, data storage (and preservation) facilities, and detectors or instruments;

2. a middle information layer where the data is structured as information and there managed, analysed, and displayed; and

3. a higher knowledge layer where – either by humans or by computers using data mining techniques – the information is converted to knowledge (justified belief).

The knowledge from the upper layer can, of course, be used to control the lower layers: for example, due to deduction from information to knowledge indicating that the range of interest of wavelengths of electromagnetic radiation is between x and y. Then the detectors can be set (or tuned) to maximum reception in the x-y range.

The lower layer has similarities to the US-developed GRID concept (Foster & Kesselman, 1998).

Such an architecture meets the requirements of: data collection, data storage and distribution, data preservation (all including metadata), and all with associated capability for processing and all within constraints of trust, security, privacy, and rights of use. Indeed, the metadata (Jeffery, 2000) is the crucial component because it provides a way to inter-relate all the various data sources and ensure their appropriate management using a CRIS, particularly if it uses the EU recommendation to member states CERIF (Common European Research Information Format) (CERIF).

However, the architecture also supports the research infrastructure facilities by providing the capabilities for: experimental control (steering), associated simulation/modelling (understanding), comparison of model with experimental results (gaining insight), and advanced analysis / visualisation tools (aiding interpretation). Alongside the experimental research the e-infrastructure - through digital libraries - also provides access to scholarly publications and thus to the wealth of past and contemporaneous research.

This describes the 'researcher's workbench.' However, the researcher also has management or administration tasks and – in the e-infrastructure – these should also be integrated. Typically, these include systems for project management, financial management, research proposals, authoring scholarly publications, etc. Using a CRIS to assist in driving this workbench is a very effective way of reducing the workload of the researcher because the CRIS contains much information that is requested time and time again from the other systems, so the CRIS can be used to produce this information in place of repeated input by the researcher.

In a complementary way, research managers – both in research institutions and in funding organisations – will wish to analyse research information in order to support strategic decision-making and management of research. This is usually based on output metrics from the research such as publication citations, license fees from patents, spin-out company turnover, invited presentations, work for the community (e.g., reviewing publications or research proposals), etc. A CERIF-CRIS contains exactly such information in a structured way so that valid statistical analyses (with appropriate graphic outputs as required) can be performed.

The use of CERIF-CRIS for scientific (research) metadata was discussed in (Jeffery, Lopatenko, & Asserson, 2002).

5 RELATIONSHIP TO CRIS

The ESFRI Roadmap 2008 (ESFRI, 2008) document states (p10):

"This definition of research infrastructures, including the associated human resources, covers major equipment or sets of instruments, in addition to knowledge-containing resources such as collections, archives and data banks. Research infrastructures may be "single-sited", "distributed", or "virtual" (the service being provided electronically). They often require structured information systems related to data management, enabling information and communication. These include ICT-based infrastructures such as Grid, computing, software and middleware. In all cases considered for the roadmap, these infrastructures must apply an "Open Access" policy for basic research, i.e. be open to all interested researchers, based on open competition and selection of the proposals evaluated on the sole criterion of scientific excellence by international peer review."

We demonstrate above how a CERIF-CRIS can support the researcher's workbench or the research manager performing research evaluation for strategic reasons or for governance. A CERIF-CRIS can be used to manage a research infrastructure facility and to interoperate among them. The CERIF-CRIS has a comprehensive and flexible data structure to record research activity including scholarly publications (Asserson & Jeffery, 2004; 2005). A CERIF-CRIS can be used to 'drive' other systems such as authentication and authorisation, directories, and web pages and can link to scholarly publication repositories, dataset and software repositories, finance, HR, procurement, project, and other systems (Jeffery & Asserson, 2006a) and provide the basic data required for a research workflow (Jeffery & Asserson 2006b). The CERIF-CRIS can thus drive the e-infrastructure for a research facility, within a research laboratory, across a whole research institution (academic or commercial), or in a research funding organisation. Furthermore, use of a CERIF-CRIS in each of these (or use of a CERIF to wrap legacy systems in each of these) permits interoperation leading to open access (i.e., on merit and without direct payment by the researcher) both to the facilities and to the products or outputs of the research.

6 CONCLUSIONS AND RECOMMENDATIONS

It is clear that research infrastructure facilities are shared increasingly among researchers from many nations. There is a need for e-infrastructure to support these research activities and subsequent management and use of the research outputs.

In order for the research facilities to be managed and used effectively, there is a need for information about the organisation, the people, the projects, the facility itself, the equipment, and the research outputs. This is the province of the CERIF data model used either or both for storage and for interoperation.

7 REFERENCES

Asserson, A. & Jeffery, K.G. (2004) Research Output Publications and CRIS. In A Nase, A. & van Grootel, G. (Eds.) *Proceedings CRIS2004 Conference*, Leuven University Press ISBN 90 5867 3839, pp 29-40.

Asserson, A. & Jeffery, K.G. (2005) Research Output Publications and CRIS. *The Grey Journal* 1 (1) TextRelease/Greynet ISSN 1574-1796, pp 5-8.

CERIF data model documentation: Retrieved from the World Wide Web April 20, 2010: http://www.eurocris.org/cerif/introduction/

ESFRI 2006 Retrieved from the World Wide Web April 20, 2010: ftp://ftp.cordis.europa.eu/pub/esfri/docs/esfri-roadmap-report-26092006_en.pdf

ESFRI 2008 Retrieved from the World Wide Web April 20, 2010: ftp://ftp.cordis.europa.eu/pub/esfri/docs/esfri_roadmap_2008_update_20090123.pdf

Foster, I. & Kesselman, C. (Eds.) (1998) *The Grid: Blueprint for a New Computing Infrastructure.* Morgan-Kauffman.

Jeffery, K.G. (1999a) Knowledge, Information and Data. Paper submitted to Director General of Research Councils. Retrieved from the WWW, April 13, 2010: http://www.semanticgrid.org/docs/KnowledgeInformationData/KnowledgeInformationData.html

Jeffery, K G. (1999b) An Architecture for Grey Literature in a R&D Context. In *Proceedings GL'99 (Grey Literature) Conference*, Washington DC Retrieved from the WWW April 20, 2010: http://www.emeraldinsight.com/Insight/viewContentItem.do?contentType=Article&contentId=861814

Jeffery, K G. (2000) Metadata. In Brinkkemper, J., Lindencrona, E., & Solvberg, A. (Eds.) *Information Systems Engineering*. Springer Verlag:London. ISBN 1-85233-317-0.

Jeffery, K.G. (2004) The New Technologies: can CRISs Benefit. In Nase, A. & van Grootel, G. (Eds.) *Proceedings CRIS2004 Conference*, Leuven University Press ISBN 90 5867 3839, pp 77-88.

Jeffery, K. G. (2005) CRISs, Architectures and CERIF. *CCLRC-RAL Technical Report* RAL-TR-2005-003.

Jeffery K. G. & Asserson, A. (2006a) CRIS Central Relating Information System. In Asserson, A. & Simons, E. (Eds.) Enabling Interaction and Quality: Beyond he Hanseatic League. *Proceedings 8th International Conference on Current Research Information Systems CRIS2006 Conference*, Bergen, Leuven University Press ISBN 978 90 5867 536 1, pp109-120.

Jeffery, K. G. & Asserson, A. (2006b) Supporting the Research Process with a CRIS. In Asserson, A. & Simons, E. (Eds.) Enabling Interaction and Quality: Beyond the Hanseatic League. *Proceedings 8th International Conference on Current Research Information Systems CRIS2006 Conference*, Bergen, Leuven University Press ISBN 978 90 5867 536 1, pp 121-130.

Jeffery, K.G., Lopatenko, A. & Asserson, A. (2002) Comparative Study of Metadata for Scientific Information: The Place of CERIF in CRISs and Scientific Repositories. In Adamczak, W & Nase, A (Eds) *Proceedings CRIS2002 6th International Conference on Current Research Information Systems*, Kassel University Press ISBN 3-0331146-844 pp 77-86.