USING A CRIS FOR E-INFRASTRUCTURE: E-INFRASTRUCTURE FOR SCHOLARLY PUBLICATIONS

E Dijk[1*] and M van Meel[2]

*[*1] Dept of Research Information, Royal Netherlands Academy of Arts and Sciences (KNAW), P.O. Box 19121, 1000 GC, Amsterdam*
Email: elly.dijk@bureau.knaw.nl
[2] Dept of Research Information, Royal Netherlands Academy of Arts and Sciences (KNAW), P.O. Box 19121, 1000 GC, Amsterdam
Email: marga.van.meel@bureau.knaw.nl

ABSTRACT

Scholarly publications are a major part of the research infrastructure. One way to make output available is to store the publications in Open Access Repositories (OAR). A Current Research Information System (CRIS) that conforms to the standard CERIF (Common European Research Information Format) could be a key component in the e-infrastructure. A CRIS provides the structure and makes it possible to interoperate the CRIS metadata at every stage of the research cycle. The international DRIVER projects are creating a European repository infrastructure. Knowledge Exchange has launched a project to develop a metadata exchange format for publications between CRIS and OAR systems.

Keywords: CERIF, repositories, Interoperability, Metadata format, Open access, Repositories, CRIS, Scholarly publications, DRIVER, e-infrastructure

1 INTRODUCTION

The idea behind the e-infrastructure concept is that all researchers have shared access to unique or distributed scientific facilities (including publications, data, instruments, computing, and communications), regardless of their type and location in the world.

Scientific publications, for example patents, are an important part of the e-infrastructure. Such publications can be used as a basis for further research, and their authors and institutes will also use them to attract research funding.

Academic communication is changing. With the advent of the Internet, it is no longer enough to publish in printed books or journals. For greater visibility, researchers need to publish their work digitally. Following the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2003), more and more institutes have wanted such full-text publications to be freely available worldwide. One way to make the output available is to store the publications in institutional Open Access Repositories (OAR).

Managing the e-infrastructure requires these publications to be interlinked to research information, for example, researchers, research projects and programmes, and research institutes. A Current Research Information System (CRIS) that conforms to the standard Common European Research Information Format (CERIF) could be a key component in the e-infrastructure. A scholarly publication would then be related to one or more authors, to the organisational unit or units employing the author or authors, and to the project in which the research was done. The project can also be related to the funding programme of a funding organisation (organisational unit). A CRIS provides the structure and makes it possible to interoperate the CRIS metadata at every stage of the research cycle. The metadata in the administrative system (for example persons, organisational units, projects, and programmes) can be used in the publication repositories.

Two projects were carried out in Europe in order to create a European repository infrastructure. DRIVER I and II (Digital Repository Infrastructure) have encouraged the implementation of institutional repositories in Europe. Knowledge Exchange will establish a metadata format.

The rest of this article is organised as follows. Section 2 gives a brief definition of the e-infrastructure. Section 3 describes why a CRIS based on the CERIF data model could be a key component in the e-infrastructure. Section 4 looks at open access publishing and OARs as a part of the e-infrastructure. Section 5 describes international

projects to create a European repository infrastructure and a project to interlink CRISs and OARs. Section 6 offers a set of conclusions.

2 THE E-INFRASTRUCTURE

The Research Infrastructures component of the FP7 Capacities Programme (http://cordis.europa.eu/fp7/ict/e-infrastructure/home_en.html) supports innovative methods of conducting scientific research (referred to as e-Science) by creating a new environment for academic and industrial research in which virtual communities share, unite, and exploit the collective power of European scientific facilities.

The term e-Infrastructure refers to this new research environment in which all researchers, whether working in the context of their home institutions or in national or multinational scientific initiatives, have shared access to unique or distributed scientific facilities (including data, instruments, computing, and communications), regardless of their type and location in the world.

3 CERIF BASED CURRENT RESEARCH INFORMATION SYSTEM

A Current Research Information System - CRIS - could be a key component in the e-infrastructure (http://www.eurocris.org/fileadmin/upload/200709.pdf). The CRIS contains management information and has the ability to identify, authenticate, and provide access to end-users. The information in a CRIS can be re-used; the metadata in the CRIS can be used in the interoperation between the CRIS and repositories containing publications and research datasets with associated software.

The EC-recommended data model CERIF (Common European Research Information Format) offers a suitable standard for CRIS operation. CERIF was developed by EuroCRIS (http://www.eurocris.org/), a European organisation designed to be the internationally recognised point of reference for all matters relating to Current Research Information Systems. CERIF allows interoperability between CRISs and can be used to interlink CRISs and repositories. A scholarly publication is related to one or more authors, to the organisational unit or units employing the author or authors, and to the project in which the research was done. The project can also be related to the funding programme of a funding organisation (organisational unit). A CRIS provides the structure and makes it possible to interoperate the CRIS metadata at every stage of the research cycle. The metadata in the administrative system (for example persons, organisational units, projects, and programmes) can be used in the publication repositories.

This may be of use in a single organisation or country, but the widespread use of a CRIS would give researchers and research managers the national and international overview required for decision-making and in a European research environment.

It was against this background that ideas concerning the interoperation of systems supporting research have gradually developed.

4 OPEN ACCESS

Academics and scholarly organisations worldwide want to provide free access to scientific results, not only publications but also other results such as datasets, videos, etc. The Open Access (OA) movement (http://www.earlham.edu/~peters/fos/overview.htm) advocates free availability and unrestricted use: no price barriers, for example, subscriptions, licensing fees, or pay-per-view fees or permission barriers, for example, most copyright and licensing restrictions. This movement is supported by various declarations.

In December 2001, the Open Society Institute (OSI) (http://www.osi.hu/infoprogram) called a meeting in Budapest of leading proponents of open access for scientific and scholarly journal literature. This led to the Budapest Open Access Initiative (February 2002), which has now (March 2009) been signed by 5,431 institutions. The goal of this Initiative is to achieve open access to scholarly journal literature.

The Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2003) calls upon researchers to make their scientific material freely available to one and all. This means not only articles, but also raw data and other research material. At the moment (April 2010), 274 institutions from all over the world have signed this declaration.

The Open Access movement has spread all over the world. It now covers not only open access to publications but also to research data, as we can see in the OECD Declaration on Open Access to publicly-funded data (2004). In 2007, the OECD published its Principles for Access to Research Data and Public Funding, in which it recommended open access of research data.

The European Strategy Forum on Research Infrastructures (ESFRI), founded to promote scientific integration within Europe and to strengthen its international outreach, also embraces the concept of open access to high-quality research infrastructures.

The ERC Scientific Council's Statement on Open Access of December 2006 stressed the fundamental importance of peer-review in ensuring the certification and dissemination of high-quality scientific research, as well as the importance of wide access and efficient dissemination of research results. In December 2007, the ERC Scientific Council launched the Guidelines for Open Access (http://erc.europa.eu/pdf/ScC_Guidelines_Open_Access_revised_Dec07_FINAL.pdf). In these Guidelines the ERC requires that all peer-reviewed publications from ERC-funded research projects be deposited on publication into an appropriate research repository where available, such as PubMed Central, ArXiv or an institutional repository, and subsequently made Open Access within 6 months of publication.

In August 2008 the European Commission launched a pilot initiative on open access to peer reviewed research articles in its Seventh Research Framework Programme (FP7). It requires grant recipients in seven areas (Health, Energy, Environment, Information & Communication Technology, Research Infrastructures, Socio-economic sciences & Humanities and Science in Society) to "deposit peer reviewed research articles or final manuscripts resulting from their FP7 projects into an online repository and make their best efforts to ensure open access to these articles". This pilot will run until the end of FP7.

4.1 Two routes to open access publishing

There are two routes to open access publishing: the Golden Route and the Green Route.

4.1.1 The Golden Route

We will give only a brief description of the Golden Route because our focus in this article is on repositories. The Golden Route is when the researcher publishes in OA journals, for example, PLoS (http://www.plos.org/) for the biomedical sciences. PLoS publishes seven peer-reviewed open-access journals, and the author retains the copyright in his or her publication. An overview of open access journals is available in the Directory of Open Access Journals (DOAJ) (http://www.doaj.org/). There are currently (April 2010) 5,000 OA journals worldwide.

4.1.2 The Green Route

The second route to open access publishing is the Green Route, or self-archiving. This means that the researcher archives his or her publications in repositories. Repositories are organised mainly by institution, subject, or discipline. They can also contain preprints, post-prints, and non-peer reviewed material, such as dissertations, audio and video files, etc.

When repositories conform to standards created by the Open Access Initiative, they can be “harvested” and, for example, made accessible through a single portal, such as the Dutch portal NARCIS (http://www.narcis.nl).
This portal provides access to the scientific publications of all the Dutch universities, the Royal Netherlands Academy of Arts and Sciences, the Netherlands Organisation for Scientific Research (NWO) and a number of scholarly institutes. Users need not know which repositories there are or where they are located in order to find and make use of their contents.

Today (April 2010) there are over 1,500 OARs worldwide, and the number is growing at a rate of around one a day. A list of OARs can be found in the Directory of Open Access Repositories, OpenDOAR (http://opendoar.org/).

4.2. Self-archiving: advantages and problems

The Green Route to open access publishing has various advantages for the researchers, the research institute, and researchers elsewhere in the world. On the other hand, the problems associated with self-archiving make researchers reluctant to place the full text of their publication in a repository.

4.2.1 Advantages of self-archiving

Archiving in institutional or other repositories has many advantages (Jeffery, 2006). The researcher can maximise the research impact because his or her work will be read and cited more often. The institutions can arrange for the long-term preservation of their own publications in the repositories. For example, the National Library of the Netherlands has made arrangements about preservation with the Dutch universities. The National Library harvests the publications in the repositories and guarantees that the full-text publications remain accessible for at least a hundred years.

There are also ethical arguments. Research funded by the public should be available to the public, and the research results in repositories are also available to third-world countries, where many research institutions cannot afford expensive journal subscriptions. Institutions everywhere face the problem that the cost of journal subscriptions is increasing. If researchers all put their research results into a repository, everyone benefits. Depositing research results in a repository makes them available faster than publishing them in a journal article. And because the research results are available on the Internet, it is possible to crosslink the publication to any research datasets and software used in producing the paper (enriched publications).

The Open Access Initiative - Protocol for Metadata Harvesting facility makes it possible to harvest all repositories adhering to this protocol and to make their contents freely available. Service providers can build a range of services based on the repository contents, such as the generation of dynamic publication lists for a researcher's personal home page.

4.2.2 Problems associated with self-archiving

Of course there are also problems associated with self-archiving. Once an institution has developed a repository, the biggest problem is collecting the content. The self-archiving rate is still low; the overall open access rate is 15-18% of all publications. Researchers want to publish in reviewed journals and are not interested in self-archiving.

Another issue pertains to copyright. Nowadays, however, more and more publishers permit self-archiving, usually after a period of six months.

4.3 Advocating self-archiving

A growing number of funding councils are recommending open access, for example, EUROHORCS (the European organisation of research councils). We also see governments recommending open access publishing, for example, the central governments of Ireland and Australia.

Open access is also being promoted more vigorously, for example, in the EurOpenScholar initiative, launched by European University Rectors. Some institutions have even made self-archiving mandatory, such as Harvard University's Faculty of Arts and Sciences or the Dutch universities' e-theses programme.

In the Netherlands, SURFFoundation, acting in close cooperation with the Royal Netherlands Academy of Arts and Sciences, the National Research Council, and all the Dutch universities, proclaimed 2009 Open Access Year in a bid to promote open access publishing among the various stakeholders, researchers, publishers, and so on.

One of the objectives of DRIVER II (See 5.2) has been the building of an organisation around the DRIVER infrastructure, capable of maintaining it over time. COAR - Confederation of Open Access Repositories - has been established to continue this work (http://coar-repositories.org).The goal of COAR is to enhance greater visibility and application of research outputs through global networks of Open Access digital repositories. COAR was launched during Open Access Week October 2009 by representatives of 28 international

organisations, as founding members. It is an international not-for-profit association with the host seat in Göttingen, Germany.

5 EUROPEAN E-INFRASTRUCTURE: DRIVER, KNOWLEDGE EXCHANGE, AND OPENAIRE PROJECTS

Part of the e-infrastructure consists of a repository infrastructure. Two projects were carried out in Europe to create a European repository infrastructure: the DRIVER I and DRIVER II projects (http://www.driver-community.eu). Another project to interlink CRISs with OAR systems was set up by Knowledge Exchange.

5.1 DRIVER I

DRIVER (Digital Repository Infrastructure Vision for European Research) was co-funded by the European Commission. The goal was to set up a European repositories infrastructure that will provide easy access to the results of scientific/scholarly research, such as papers, reports, and research data, including models and visualisations.

DRIVER is modelled on the Dutch DAREnet (now integrated into the scholarly portal NARCIS), the Dutch network of repositories, which provides access to academic results.

At the end of the DRIVER I project in November 2007, a test bed system was delivered, which offered a search portal with open access content from over 70 repositories while focusing on the infrastructure aspect. Dutch, German, French, English, and Belgian institutions participated in the test bed.

5.2 DRIVER II

The successor to DRIVER, the DRIVER II project, started in 2007 and ended in December 2008. DRIVER was integral to the suite of electronic infrastructures that has emerged in the worldwide GÉANT network and was hence funded under the e-Infrastructures component of the European Commission's 7th framework programme. DRIVER II involves extending the repository infrastructure into a European confederation of repositories. The test bed has been converted into an operational environment in which 15 countries participate. In June 2008, DRIVER launched D-NET version 1.0. This open source software offers a tool-box for deploying a customisable 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, and profiling to innovative tools for repository manager users.

The focus of DRIVER II was on building repositories of “compound objects” or “enhanced publications” for various academic disciplines. This involves interlinking traditional publications to related research descriptions, datasets, or simulations. One part of the project coordinated by the Royal Netherlands Academy of Arts and Sciences and the Data Archiving and Networked Services (DANS) institute involved creating several sample datasets of enhanced publications that were then used as input for a demonstrator of a repository of enhanced publications (http://driver2.dans.knaw.nl/demonstrator/html). The enhanced publications for the demonstrator were chosen to reflect material typical of the three academic disciplines: a Hebrew database representing the humanities, measurements representing the sciences, and survey data representing the social sciences. Open Archives Initiative Object Reuse and Exchange (OAI-ORE), and Resource Description Framework (RDF) were used to aggregate and serialise these enhanced publications. The point of all these efforts was to achieve a better e-infrastructure in the end.

5.3 Exchange format for publications between CRIS and OAR systems

As mentioned earlier, a CRIS based on CERIF could be the key component in the e-infrastructure. It can interlink research information (programmes and projects, researchers and research institutes) with publications in repositories.

At the end of 2008, Knowledge Exchange (http://www.knowledge-exchange.info/) launched an international project that proposed a metadata exchange format for publications between CRIS and OAR systems. The

purpose is to reuse metadata, resulting in more efficient data use. CERIF will be taken as starting point for the data model in this project.

Knowledge Exchange (KE) is an international partnership that supports the use and development of ICT infrastructures for higher education and research. The partners in KE are Denmark's Electronic Research Library (DEFF), the German Research Foundation (DFG), the Joint Information Systems Committee (JISC) in the United Kingdom, and the SURFfoundation in the Netherlands.

At the moment there is neither international consensus on nor any existing international exchange format between CRIS and OAR platforms. All the KE partners and all other interested parties, for example, universities all over Europe and EU projects such as DRIVER, will benefit from a metadata exchange format because it will make metadata input more efficient and improve its quality, reliability, and reusability. It will ***avoid or reduce*** duplicate inputs of the same metadata on different platforms. A side effect will be a reduction in the cost of metadata handling and exchange.

The purpose of the project is to increase the interoperability of the metadata used in CRIS and OAR systems by defining a metadata exchange format and a common vocabulary that can be used by both platforms. CERIF will be used as the starting point for the metadata exchange format in this project. The results will be presented at the CRIS 2010 conference (http://www.cris2010.org) to be held in Aalborg, Denmark, in June 2010.

5.4 OpenAIRE

OpenAIRE - Open Access Infrastructure for Research in Europe - is a three-year project funded under the 7th Framework Programme of the European Commission (http://www.openaire.eu). It started in December 2009.
The main goal of OpenAIRE is to support the Open Access pilot, launched by the European Commission in August 2008. This Open Access pilot, which covers about 20% of the FP7 budget, commits researchers from 7 thematic areas to deposit their research publications in an institutional or disciplinary Open Access repository, to be made available worldwide in full text.

It will establish the infrastructure for researchers to support them in complying with the EC OA pilot and the ERC Guidelines on Open Access. OpenAIRE will provide an extensive European Helpdesk System to ensure localized help to researchers within their own context. It will build an portal and e-Infrastructure for the repository networks and explore scientific data management services together with 5 disciplinary communities. It will also provide a repository facility for researchers who do not have access to an institutional or discipline-specific repository.

6 CONCLUSION

With respect to the e-Infrastructure for scholarly publications, it is evident that institutional and other repositories are becoming more firmly established. Many academic institutions set up repositories so that their researchers can fill them with open access publications. Software such as DSpace, Eprints, or Fedora is also available now, and standard metadata models and protocols have been developed. Unfortunately, repositories are not really embedded in scientific practice. It is difficult for repository managers to get the repositories filled, the field of high energy physics being an exception; there, it is quite common to archive publications in repositories.

On the other hand, institutions, governments, and funding organisations have advocated or even mandated open access, not only in Europe but all over the world. In Europe, the aim of the two DRIVER projects was to establish a repository infrastructure, with the focus in the DRIVER II project being on "compound objects". The main goal of project OpenAIRE is to support the Open Access pilot, launched by the European Commission in August 2008, by establishing an infrastructure for researchers.

In terms of placing publications in OARs in a research context of information about researchers, research institutions, research programmes, and projects (a CRIS), it is clear that CERIF can act as the core of the e-infrastructure, supporting all aspects of research activity.

Knowledge Exchange has launched a project in Europe to develop a metadata exchange format for publications between CRIS and OAR systems. The data model CERIF will play an important role in this project.

7 REFERENCES

COAR website. Retrieved April 23, 2010, from the World Wide Web: http://coar-repositories.org/

CORDIS website. Retrieved March 31, 2009, from the World Wide Web: http://cordis.europa.eu/fp7/ict/e-infrastructure/home_en.html

CRIS and European e-infrastructure - enabling European Research, Report, September 2007. Retrieved March 31, 2009, from the World Wide Web: http://www.eurocris.org/fileadmin/Upload/200709.pdf

Demonstrator of Enhanced Publications website. Retrieved April 6, 2009, from the World Wide Web: http://driver2.dans.knaw.nl/demonstrator/html/

DOAJ website. Retrieved April, 2010, from the World Wide Web: http://www.doaj.org/

DRIVER website. Retrieved March 31, 2009, from the World Wide Web: http://www.driver-community.eu/

Dutch portal NARCIS website. Retrieved April, 2010, from the World Wide Web: http://www.narcis.nl

ERC Scientific Council Guidelines for Open Access. Retrieved 23 April, 2010 from the World Wide Web: http://erc.europa.eu/pdf/ScC_Guidelines_Open_Access_revised_Dec07_FINAL.pdf

EuroCRIS website. Retrieved April 6, 2009 from the World Wide Web: http://www.eurocris.org/

Jeffery, K.G. (2006). Open Access: An Introduction. ERCIM NEWS, 64. Retrieved March 31, 2009 from the World Wide Web: http://www.ercim.org/publication/Ercim_News/enw64/jeffery.html

Knowledge Exchange website. Retrieved March 31, 2009, from the World Wide Web: http://www.knowledge-exchange.info/

Open Society Institute website. Retrieved March 31, 2009, from the World Wide Web: http://www.osi.hu/infoprogram

OpenAIRE website. Retrieved April 23, 2010, from the World Wide Web: http://www.openaire.eu/

*Open*DOAR website. Retrieved April 23, 2010, from the World Wide Web: http://www.opendoar.org/

Peter Stuber's website – for Open Access overview. Retrieved March 31, 2009, from the World Wide Web: http://www.earlham.edu/~peters/fos/overview.htm

PLoS website. Retrieved March 31, 2009, from the World Wide Web: http://www.plos.org/