Integrating CRIS and repository – an overview of the situation in Finland and in three other Nordic countries

Author: Jyrki Ilva, National Library of Finland
E-mail: jyrki.ilva@helsinki.fi

ABSTRACT

Current research information systems (CRIS) are being implemented rapidly in many universities and other research organizations all around Europe. Although research information systems of some sort have been around for a long time, an important impetus for the recent development has been the arrival of a new generation of sophisticated, proprietary CRIS platforms. There are also political factors related to changing research evaluation and funding schemes that have increased the attractiveness of these systems.

Because of this development, the relationship between CRIS and repository is a timely topic in many countries. The integration of these systems would obviously provide many benefits, but there are many ways to do it, both locally and on a trans-organizational level. One potential option is to combine the functions of these two systems into one single system.

This paper looks at the situation in Finland and three other Nordic countries, Denmark, Norway and Sweden. The countries are in many ways very similar, and there have been similar motivating factors for the development of both local and national CRIS and repository infrastructure. However, because of slightly different circumstances in each country and the timing of the choices that have been made, they have all ended up with quite different outcomes.

BACKGROUND

Current research information systems are usually implemented primarily to serve the needs of research administration, although the library may also have a role in the management of the system. They are expected to contain reference information for all of the research publications produced within the organization. In addition, they may contain information on other activities, awards, projects, departments and funding. They may also provide personalized home pages for the researchers working at the organization, with the aim of increasing the visibility of the researchers and the research outputs produced by them.

On the other hand, institutional repositories usually concentrate on providing (open) access to full-text materials produced within the organization. They are usually run by the library.

There is obviously a lot of overlap between these two types of systems, and there are many potential benefits to be realized from combining their processes and workflows on a local, national or international level. However, it is worth noting that the basic motivations behind them are not identical. Following this, the integration of these two systems is not always a straightforward task.

The history of current research information systems is actually fairly long, going back to the pre-Internet era. The first version of the CERIF format (currently maintained by EuroCRIS) meant for data exchange between different systems was released as early as in 1991.[1] Although research information has been collected and used by research administration for a long time, in recent years the role of this information has become ever more important. Many European countries have established new evaluation methods and funding models for their research organizations. These generally place great weight on the number and quality of scientific publications produced within the organization. Although some of the research evaluation needs can be met by utilizing the information contained in international databases like Web of Science (Thomson Reuters) or Scopus (Elsevier), a local or

[1] History of CRIS, EuroCRIS web site, http://www.eurocris.org/Index.php?page=historyCRIS&t=1.

national CRIS connected to various information sources is usually considered necessary for the effective implementation of these schemes.

One of the key developments in this area has been the launch of a new generation of modern, proprietary CRIS platforms based on an updated CERIF data model. Within the last five years these systems (mainly Pure, Converis and Symplectic Elements) have been implemented in increasing numbers in many universities, most notably in the UK.[2] Most of these organizations also have an institutional repository. This has led to increasing discussion on the possibilities of integrating CRIS with repositories, or even, of one of these systems replacing the other. A lot of work has been done in various projects to find out the best practices for doing this, but it seems that there will be many different approaches even in the future.[3]

This paper concentrates on four Nordic countries, Finland, Denmark, Norway and Sweden. Some of these countries have been pioneers in adopting research information systems either on a local or on a national level. The development of repositories also started out relatively early in these countries, often with ETD programs that later expanded into full-fledged repositories and led to the adoption of one of the international repository platforms. Despite the many similarities between these four countries, when it comes to the integration of current research information systems and repositories, they have ended up taking quite different paths.[4]

COMPARING THE NORDIC COUNTRIES

Finland

It seems that Finland is currently alone among the Nordic countries in building two separate system infrastructures for repositories and research information systems. Just about all of the Finnish universities and universities of applied sciences and many of the state research institutes already have an institutional repository, which is either hosted at a local level or uses the centralized DSpace-based platform provided by the National Library of Finland.[5] Some of the state research institutes use DSpace for their publication data (including metadata for non-OA publications) as well, although it doesn't fit all of the CRIS qualifications.

On the other hand, none of the research universities use the repository to manage their publication data. While many of the universities are still using older research information systems,[6] the University of Helsinki was the first Finnish organization to adopt one of the international CRIS platforms when it launched its own Pure instance, Tuhat (https://tuhat.halvi.helsinki.fi/portal/en/), in 2010. Tuhat is also used as a submission system to Helda (https://helda.helsinki.fi), the DSpace repository of the university. The researchers are expected to upload the full-text files of their scientific publications to Tuhat, from where they are then automatically copied to the repository.

The new funding model of the Finnish universities which came into effect in 2013 has escalated the development towards the adoption of new research information systems. In the model as much as 200 million euros a year (13% of the total state funding for the universities) is distributed according to the number and quality of the publications produced at each university.[7] Because of this the universities have more or less suddenly become

[2] For the early UK discussion see Nick Sheppard: Learning How to Play Nicely: Repositories and CRIS, Ariadne 64, 2010, http://www.ariadne.ac.uk/issue64/wrn-repos-2010-05-rpt, and for the current situation, RepNet map for CRIS/IR implementation at UK HEIs, http://goo.gl/maps/aqB1b.

[3] One good starting place is Mikael K. Elbæk: Knowledge Exchange CRIS-OAR metadata interoperability project (2012), https://infoshare.dtv.dk/twiki/pub/KeCrisOar/ProjectDocuments/Knowledge_Exchange_CRIS-OAR_final_report_26032012.pdf

[4] The early history of current research information systems and repositories in Nordic countries has been charted in Ingegerd Rabow: Research Information Systems in the Nordic Countries - Infrastructure, Concepts, and Organization (2010), http://hprints.org/hprints-00433868.

[5] The services provided by the National Library are currently used by 40 organizations, whose collections are located in five DSpace instances, Doria, Theseus, Julkari, Jukuri and Tampub. The Helsinki University Library also provides repository services for a number of partner organizations.

[6] SoleCRIS (http://www.solenovo.fi/en/solutions/solecris/), a product of Finnish company Solenovo, is currently the market leader.

[7] By comparison, the amount of money the same universities spend on the licensing of electronic materials is currently in the range of 20 million euros a year.

very interested in acquiring one of these systems, and although many of the universities are still in the acquisition or planning phase, it is estimated that by 2016 all of the universities will have either one of the new CRIS platforms or at least an upgraded version of their current CRIS.

CSC – the Finnish IT Centre for Science collects the publication data from the universities for the purposes of the Ministry of Education and Culture. Currently the data is collected once a year in .csv format, but there are plans to move to an automated system once the universities have implemented research information systems capable of providing the information. The research publication data has been made publicly available in Juuli (http://www.juuli.fi), a VuFind-based user interface built by the National Library. [8] However, since the process of data collection doesn't involve repositories, the information on the Open Access status of the publications is currently lacking in quality.

Denmark

Pure, one of the top CRIS platforms, originates from Denmark. It was developed by a small start-up company (Atira A/S) originally hired by Aalborg University to develop their research database. In 2005 Denmark's Electronic Research Library (DEFF) provided support for the implementation of Pure in four universities, and by 2010 it had been adopted in all Danish universities.[9]

In Denmark Pure is used both as a CRIS and also as a repository platform, instead of utilizing one of the open source repository solutions. A good example of a combined CRIS/repository is DTU Orbit (http://orbit.dtu.dk/en/), the research information system of the Technical University of Denmark. The university had been developing a Fedora-based solution, but ended up migrating to Pure.

One of the advantages of the wide use of Pure is the possibility of harvesting rich publication data automatically to a Danish National Research Database, Forskningsbasen.dk (http://www.forskningsdatabasen.dk/). Partly due to relatively high demands on metadata compatibility (the use of XML-based DDF-MXD format is required) it is not quite comprehensive, but it contains more than half a million records from 13 organizations. It also does a good job of providing an overview on the prevalence of Green Open Access in these organizations.

Norway

Norway has been one of the pioneers in building a comprehensive national research information system which is also integrated with repositories. The current version of the national CRIStin system (http://www.cristin.no/english/system/) has been developed specifically for this purpose, and it works as a national CRIS, which is also the primary data entry point for all of the Norwegian research organizations.

What is unique in the Norwegian approach is the way the local repository workflows are connected with the national system. Although Norway has a distributed system of local organizational repositories (many of them DSpace instances hosted by Bibsys), the self-archived articles are uploaded to the national system and transferred from there to institutional repositories.[10] This makes it easier to keep track of the prevalence of Green Open Access.

Sweden

The Swedish situation differs to some extent from that of Finland, Denmark and Norway. The main difference compared to Denmark is that the proprietary CRIS platforms haven't so far made much of an inroad in Sweden. This is largely due to the dominant position of the local DiVA platform, which was originally developed nearly fifteen years ago by the University of Uppsala. DiVA is currently in use in more than 30 Swedish organizations, and it provides both a centralized portal (http://www.diva-portal.org/) for all participating organizations and a separate

[8] Jyrki Ilva: Juuli – a national portal for publication data, CSC News 3/2013, http://urn.fi/URN:NBN:fi-fe201402121466
[9] Elbæk (2012).
[10] Nina Karlstrøm: Bridge over troubled water - using the CRIS in building Open Access infrastructures (Presentation, 15.10.2013), available at http://urn.fi/URN:NBN:fi-fe201310166685

user interface for each of them. It can be used both as a full-text repository and also for the management of reference information on all publications produced at the organization. On a technical level, DiVA has been upgraded to use Fedora as a storage layer. However, some of the other Swedish universities have built their own repositories independently using other open source platforms.

One of the outcomes of the Swedish national OpenAccess.se project has been the creation of SwePub (http://swepub.kb.se/), which is a combined search interface for the research publications from all of the Swedish universities. SwePub is hosted by the National Library of Sweden and it uses OAI-PMH and the MODS metadata format to harvest publication data automatically from the Swedish repositories. While SwePub is originally a product of the Swedish repository community, it is gaining wider acceptance and is likely to be used for research evaluation purposes as well.

It remains to be seen whether collecting publication data will be enough in Sweden even in the long run, or whether there will be pressure to move towards the adoption of bona fide current research information systems at some point in the future.

CONCERNS FOR THE FUTURE

CRIS and repository as distinct concepts are probably likely to co-exist in the future, but it seems that the integration of these two systems will remain a popular target. Although there are projects trying to develop open source CRIS (it is too early to tell whether these will be able to gain large enough community support), currently the major CRIS platforms are all proprietary systems. At least so far, the emergence of modern CRIS seems to be mostly a European phenomenon, which may explain why the international repository community has been relatively slow to react to this development.

At the moment it seems that some of the research organizations are moving from open source repository platforms to proprietary current research information systems rather than vice versa. While the repositories do a good job of providing access to publications and other digital materials, the modern research information system seems to be offering more sophisticated submission forms, metadata formats and workflows tailored to meet the wider needs of research administration, libraries and scholars.

Are repositories (as we know them now) bound to be losing ground to these new systems? There should be no reason for doom and gloom. Of course, one of the key strengths of the current repository platforms is that they are open source software, and they can be adapted to the changing needs of our stakeholder communities, at least if there is willingness to put effort into their development. However, to do that, we must also be prepared to re-think what we consider to be the domain and concept of the repositories.

It should be pointed out that there are some concerns about the future development of the current CRIS products. Atira (Pure) and Avedas (Converis) were both originally small independent companies, but Atira was bought by Elsevier in 2012, and Avedas joined Thomson Reuters in November 2013. This means that the Nordic organizations looking for a modern CRIS will probably have to choose between the products of these two big companies (as far as I can tell, Symplectic has not been active in the Nordic countries).

As Elsevier and Thomson Reuters are also the leading players in the international research information field, it remains to be seen what kind of choices they will eventually make to consolidate the newly bought systems with their existing products (including Elsevier's SciVal and Thomson Reuters's InCites) and how well this will cater to the long-term needs of customer organizations.

On the other hand, as far as Green Open Access is concerned, the integration of CRIS and repository has not been an easy shortcut to unqualified success. While it makes sense to combine the ingest workflows of these systems, the advancement of Open Access tends to be secondary to the primary purpose of a CRIS. In practice it is often very easy for the researchers to bypass the self-archiving option available in the CRIS submission interface. The optimal time scales and workflows for reporting publications to a CRIS and uploading them to a repository do not always match each other, either. Even with integrated systems, campus level advocacy and practical guidance for the researchers are still indispensable.