CRIS 2014

The gradual merging of repository and CRIS solutions to meet institutional research information management requirements

Pablo de Castro[a, *], Kathleen Shearer[b], Friedrich Summann[c]

a GrandIR Ltd, Edinburgh EH3 5ET, Scotland, UK
b Confederation of Open Access Repositories (COAR), 37073 Göttingen, Germany
c Bielefeld University Library, Bielefeld, Germany

Abstract

Much has been said in recent times about the alleged dichotomy between Institutional Repositories (IRs) and Current Research Information Systems (CRISs). According to this highly ideological argument, IRs would be the platforms to support the non-commercial initiative jointly carried out by HEIs – and specifically their Libraries – in order to freely disseminate their research outputs, whereas CRISs would support the whole institutional research information management (RIM) with special emphasis on projects and funding. RIM being an activity oriented towards reporting for research assessment exercises and thus tightly connected to the institutional funding, the support from the Management at HEIs for CRIS implementation and operation and for the Research Office traditionally in charge of such tasks would be much higher than for the much less relevant IR. Moreover, the awareness of researchers and scholars towards such platforms will usually be much higher for the CRIS – from whose accurate and complete depiction of their research activity their salaries will ultimately depend – and it won't be unusual to collect complaints on the need to ensure that both systems are simultaneously fed with the appropriate, often duplicated information. According to this conception, it is often hard to get the institutional Research Office and Library to work together for improving the end-user experience by enhancing their system interoperability.

While much of this may still be happening at a number of HEIs, the general landscape is swiftly evolving and it's not that accurate anymore to describe the RIM system configuration at institutions in such oversimplified terms. CRIS/IR interoperability is now a fairly widespread feature that will allow both platforms to efficiently exchange information and reinforce each other's features, and especially the borders between what each of these platforms is and does are becoming increasingly blurred. Commercial CRISs are gradually becoming compliant with the OAI-PMH protocol and thus becoming able to offer institutions an integrated repository functionality, while the main open source IR platforms have now developed extended data models that

---

*

 * Corresponding author. *E-mail address:* pcastro@grandir.com

will allow them to deliver features traditionally associated to CRISs such as project and funding management, hence becoming suitable solutions for research institutions where purchasing or developing a highly-sophisticated CRIS is not a top priority.

This paper aims to describe the areas where CRIS/IR interoperability is taking place, and will provide a set of use cases for institutional research information system configuration involving IRs, CRISs and a combination of both. These will show how both systems are now increasingly merging for best serving institutions and their researchers.

1.   Introduction

In recent years there has been widespread growth in the adoption of CRIS and repository systems at universities worldwide. The OpenDOAR Directory of Open Access Repositories[1] at the University of Nottingham currently lists over 2,500 Open Access repositories across the world. A similar euroCRIS initiative – the Directory of Current Research Information Systems or DRIS[2] – to list the available Research Information Management (RIM) systems or CRISs is under way, showing there are also a very large numbers of such systems, although they are more concentrated in Europe and North America.

Originally repositories and CRISs have somewhat different aims and have evolved fairly independently of each other. CRISs collect a wide range of metadata about all aspects of the research activity carried out at an institution. They have been developed to "assist the users in their recording, reporting and decision-making concerning the research process, whether they are developing programmes, allocating funding, assessing projects, executing projects, generating results, assessing results or transferring technology" [3]. Institutional repositories (IRs), on the other hand, have evolved as part of the open access movement and aim to collect and provide free access to the research outputs created at the institutions. To date, IRs have focussed mainly on collecting research articles as well as theses and dissertations, although there is growing interest in expanding their scope to collect research data sets. From the CRIS perspective, publications are the result of projects and related institutional activities. From the IR point of view publications are academic resources to be made available for reuse.

Despite these somewhat distinct missions of IRs and CRISs, there are a number of overlapping areas in the tasks that they perform and there has been a gradual convergence of these two types of systems. Many institutions have now fully integrated their repository and CRIS, either by enabling systematic metadata transfer between both systems (such as at the University of St Andrews[4]) or by allowing one of them to take over the features of the other one, thus delivering a single-system integrated functionality, such as at the University of Hong Kong, which has developed their IR to also perform "as a system for reputation management, impact management, and research networking and profiling, all of which are concepts included in the broad term Current Research Information System" [5].

IRs are evolving and creating new functionality and services. For example funding agencies worldwide are issuing open access policies which are driving the gradual development of new monitoring services built on top of repositories and repository networks. These services, most of which are in their early stages of development, aim to track adherence to open access policies and collect information about research funders and research projects, and link them with related research outputs. Examples of such initiatives are OpenAIRE in Europe and the SHARE project in the United States. Thus open access repositories have begun to take on a small role in research monitoring

and evaluation. In less developed countries, repositories may increasingly be used for this purpose given the lack of resources available for maintaining full fledged CRIS systems.

At the same time, CRISs and IRs are well aware of the role played by the other stakeholder and of the need to find ways for joint work via system interoperability. A very relevant step forward in this regard has been the recent release (March 2014) of the OpenAIRE Guidelines for CRIS Managers based on CERIF XML[6], which will allow CRIS systems to become data providers for the OpenAIRE repository infrastructure. A growing degree of system interoperability will enable institutions running a wide range of different system configurations to join the common effort for collecting and sharing institutional research information and offering access to research outputs.

2.   Areas of Integration between IRs and CRISs

CRISs and IRs share a similar interest on the scientific publication in the academic environment – and this is the main overlapping factor between them. On the other hand it has to be noticed that CRIS platforms and institutional repositories have a different approach, a different history of emergence and a different position in the academic environment. This leads to different views on publications and different key aspects of activities and functionality.

From the CRIS point of view publications are the result of projects and related institutional activities. They are tied to institutional relationships in multiple ways. From the IR point of view publications are academic resources with a certain local binding which are presented to the scientific community. IRs observe the complete publication output including extended material. Thus IRs have parallel interoperability channels – similar to and not less relevant than those at CRISs – with specific systems such as publication platforms for journals and monographs, digital collections of source material, research data processing and virtual research environments. Another relevant aspect is that repositories have a longer history from both a technical and an organizational perspective. Due to the efforts devoted to platform improvement, IRs have gradually added a broad range of functionality and services where opportunities may lie for enhancing CRIS/IR interoperability, among them:

-    optimizing the visibility of the documents and the repository
-    handling open access availability and restriction (embargoes etc.)
-    persistent identifiers, usage statistics and bibliometric figures
-    long-time preservation
-    publication management
-    interfaces (APIs, HTTP, REST, SRU, OAI-PMH)
-    end-user usability (searching, embedding)
-    metadata ingest from external providers
-    metadata quality and data curation activities
-    integration with external systems such as research data management platforms, virtual research environments, publishing platforms like journal systems (OJS) and CRISs
-    citation style support
-    linked open data support

A specific issue for CRIS/IR interoperability is the synchronization of metadata formats for CRISs (CERIF) and IRs (DC, MARC and MODS/METS). CERIF is not used yet on a broad scale among IRs, but rather these other library-specific bibliographic metadata formats as a result of the traditional library-related repository nature. Initiatives like the CERIF-XML scheme developed by OpenAIRE will be of much help for tackling these metadata-related interoperability issues.

3. Use cases for institutional RIM system configuration

In order to be able to provide services across the UK repository network onto a wealth of RIM system configurations at HEIs, the UKRepNet[7] needed to collect an accurate picture of these RIM systems (mainly involving IRs and CRISs). Aware that many institutions, especially smaller ones, do not run CRIS systems and do instead use their IRs as their main research information management tool, the project carried out an analysis of the most widespread use cases across UK HEIs which has been found to be generally applicable. The wide variety of RIM system configurations identified at HEIs has been synthesized into the following use cases:

1. IR-only

A significant number of HEIs are currently relying on their IRs as their sole RIM platform. These HEIs do not run a CRIS and have no plans to purchase or develop one, but aim instead to collect the institutional research output in the IR "as the only centralised record of research activity in the university". As a result of restrictive copyright policies that will not allow every full-text version to be offered open access, these IRs will often contain a high rate of metadata-only items. Specific repository enhancements such as the CRIS add-on for EPrints[8] or the DSpace CRIS open source solution recently released by CINECA[9] enable the IR-only use case to evolve into the IR-as-CRIS one, where repositories become able to manage research information traditionally associated with CRISs such as person or project data.

There are a number of institutions where an original EPrints repository has been upgraded into a CRIS system by extending its data model, thus providing examples for the IR-as-CRIS use case. By being able to host information about projects and funding, institutional repositories like Enlighten at the University of Glasgow or Sussex Research Online (SRO) at the University of Sussex in the UK have already been used as data providers for the recent REF2014 research assessment exercise. A short presentation by Chris Keene, SRO repository manager, at the Repository Fringe event in Edinburgh last Aug 2013, described the impact of this upgrade on the repository working procedures:

"It all comes to whether you have a CRIS or you don't. We [at University of Sussex] don't. We were doing this nice Open Access thing when earlier this year, driven by both the REF and the RCUK policy compliance needs, we had to suddenly start worrying about how aspects such as metadata quality would impact on the financial stability of the institution. Before that, it was our personal pride as librarians that would drive the effort for collecting accurate metadata, but as a result of using the REF plug-in on our IR for delivering the research outputs information required for the REF2 section, metadata quality became a really serious affair which millions of pounds and job positions could eventually depend on.

As a result of this involvement in the REF reporting effort, the IR became risk-averse, very much like other areas of institutional administration such as Accounting or Finance. People at the University Schools have started talking of the IR as "the REF System", and even if what gets the Uni the millions of pounds and the reputation is actually the quality of the research –something the IR cannot help with– mistakes in the way it gets described may cause problems in the reporting process".

Another relevant example for this IR-as-CRIS use case is provided by the Hong Kong University (HKU), whose DSpace-CRIS-based Scholars Hub[10] was developed from a DSpace repository by adding extra CERIF entities to its data model which allowed it to cover a far wider metadata structure than the one needed to describe publications, extending into areas such as organisational units, projects or funding. The DSpace enhancement work jointly carried out by HKU and the Italian consortium CINECA resulted in the release of the new DSpace-CRIS platform, an open source CRIS which is gradually being adopted by institutions worldwide.

2. IR + CRIS

This use case covers the joint operation of an in-house built or commercial CERIF-based CRIS (such as Pure or Converis) and an IR at a single HEI. The two systems may run independently from each other supporting completely detached institutional processes for reporting on the research activity (usually carried out from an institutional Research Office) and disseminating its research outputs (usually from the Library). However, due to efficiency considerations and to the gradual availability of the technology that will make it possible, both systems are usually linked to each other so that a systematic data exchange process takes place between them: usually the CRIS will automatically submit all publication metadata collected from external databases to the IR by means of a basic CERIF/DC or CERIF/MODS mapping. The CRIS acts thus as an internal CERIF-based management system for research outputs (plus other entities such as projects, people, etc) whereas the OAI-PMH-compliant IR disseminates such outputs to the outside world, offering Open Access to the full-text contents whenever possible.

There is an ever growing number of examples for this kind of CRIS/IR interoperability running at many institutions worldwide. The integration between the Pure CRIS system and the DSpace-based Research@StAndrews institutional repository provides one of the most consolidated examples for such system architecture[11].
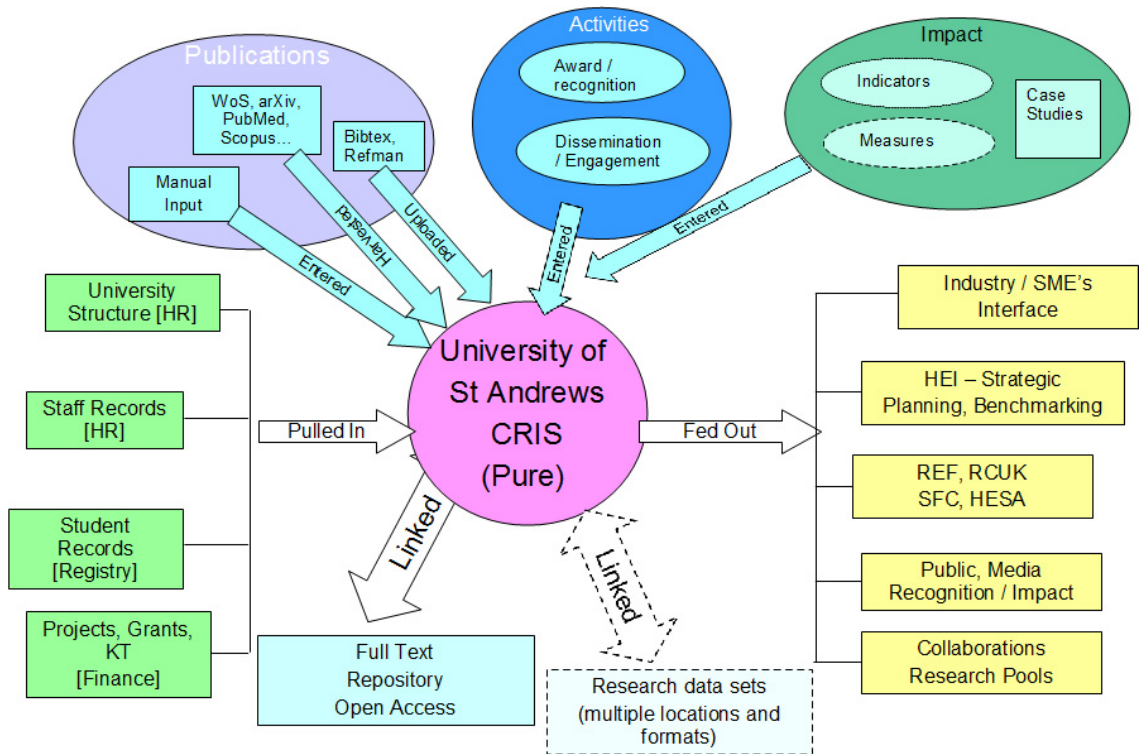

Fig. 1. Information workflows around the University of St Andrews Pure CRIS.

By relieving authors from the time-consuming task of typing in the metadata, the integration between the Pure CRIS and the DSpace repository at the University of St Andrews will allow the level of contents stored in the latter to become much larger. Metadata are automatically fed into the IR through an internal gateway and all authors are required to do is to validate the references as their own and to add in the full-text version that complies with the publishers' policies. These are later checked by the Library and released from the repository, while the CRIS will store the link to the full-text publication together with comprehensive context information. This way both platforms

– plus the institutional units that run each of them, ie the Research Office and the Library – are able to effectively work together, thus breaking the perceived dichotomy between CRISs and IRs.

The case study provided above features a CERIF-compliant CRIS system (Pure) as the metadata source in the CRIS/IR integration. Having a standard CERIF/DC or CERIF/MODS metadata gateway will of course offer great opportunities for standardizing CRIS/IR interoperability. However, there are also numerous examples of CRIS/IR interoperability where the CRIS is not a CERIF-compliant platform. The figure below shows the impact in terms of repository content storage of the integration between the non-CERIF-compliant Drac CRIS and the EPrints repository at the Polytechnical University of Catalonia (UPC)[12], with direct deposits into the IR featured in dark green and the increase in openly available (light green) or restricted-access (red) items resulting from system integration.
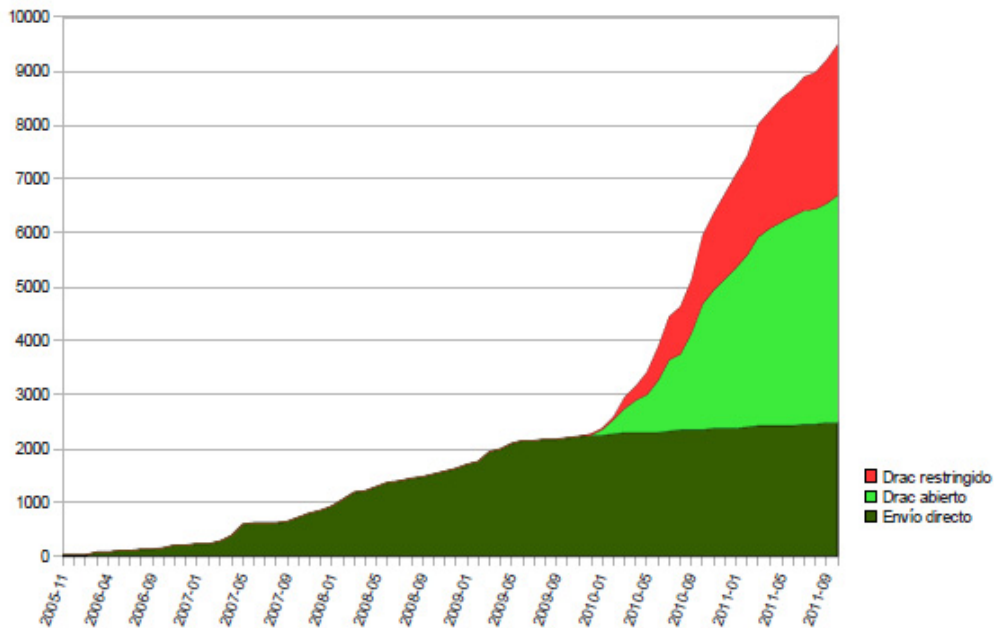

Fig. 2. Increase in repository content resulting from CRIS/IR integration at UPC.

3. CRIS-only

Some commercial CRIS platforms have recently developed IR features, including compliance with the OAI-PMH protocol. As a result of this, there is a trend in countries with an advanced research information management infrastructure towards CRISs and IRs becoming a single merged system with the research outputs held in the CRIS where they are also made Open Access available from. This case where a CRIS takes over the IR role for institutional research output dissemination (including item harvesting by aggregators via OAI-PMH) is the sole one where a reference to the system dichotomy could be made, even if an independent IR will sometimes remain in operation for disseminating other kinds of institutional output such as dissertations and/or grey literature.

This is the use case known as CRIS-as-IR, where a given HEI relies solely on its CRIS for managing its research outputs (plus all additional information related to its research activity) without operating an independent IR, whose functionality is actually embedded in the CRIS. This use case extends the boundaries of the Open Access repository

definition, and examples of such implementations for the Pure platform are increasingly featured in the abovementioned OpenDOAR repository directory (see below the entry for the Pure-based 'QUB Research Portal' institutional repository for Queen's University Belfast).

Fig. 3. An example for system merger: Pure CRIS at Queens University Belfast featured in OpenDOAR repository directory.

CRISs were however not originally design for the purpose of disseminating research outputs towards the outside, but rather for internally managing institutional research information for meeting the reporting needs. As a result, although increasingly becoming OAI-PMH-compliant, the repository functionality offered from CRIS platforms is still far from the sophisticated features IRs will offer in areas like integrated usage or automatic content transfer via the SWORD protocol. However, at a time where the economic inefficiencies of simultaneously running two potentially overlapping systems are very much taken in consideration by HEI administrators, the design and implementation of new, specific repository services is critical for this infrastructure to survive, together with the realisation that they are the best placed systems to meet the open access policies issued by research funders at those institutions where no urge is felt to run a CRIS.

4.   Conclusions

There is a wide range of system configurations running at HEIs nowadays in order to meet the institutional requirements in terms of research information management. These cover stand-alone institutional repositories and current research information systems or CRISs, plus different combinations of both. Certain traditional opposition between these platforms is gradually disappearing as the emphasis shifts towards enhancing interoperability among all available institutional systems for an optimal exploitation of their joint functionality. As a result, we have now repositories acting as CRISs, CRISs acting as repositories and CRISs and IRs working together through systematic data exchange.

There are wide differences in the CRIS/IR landscape in different countries, often depending on how deeply felt the need is for meeting reporting requirements from the Government or the research funders. In geographical areas where reporting is perceived as a key element of an effective institutional research information management, the variety in system configurations at HEIs tends to become much more complex, with CRISs gradually adopting predominant roles. Many institutions will however choose to keep their institutional repositories, either as their main platform for managing research outputs where there is no intention of running a CRIS, or as a specific system for

externally disseminating the institutional research outputs working in combination with a more internally-oriented research information system.

CRIS/IR interoperability is based on a systematic metadata exchange between both systems that will allow the "one input, many outputs" concept to be gradually realised: metadata are to be ingested (either manually or automatically) just once into one given system and then automatically transferred to all the other ones. CRIS/IR interoperability will benefit institutional repositories, which will collect a much larger level of contents as a result of metadata being automatically made available, and also CRISs, which will have a complementary platform where the full-text research outputs are showcased from and where indicators such as content usage are collected at.

There is an increasing trend towards system merger and towards a situation where the functionality to deliver is far more important than the systems that enable its delivery. Solutions like the ever more frequent IR-as-CRIS and CRIS-as-IR integrations point at an evolution towards single-system solutions which manage to tie up external dissemination and research information management functionality.

Research data management (RDM) is another emerging area where CRISs and IRs will need to exploit their interoperability. Right now some HEIs use their repositories – either the existing ones for research publications or data repositories specifically set up for the purpose – while others are planning to store the information on research data into their CRIS systems. Taking into account that researchers will often choose to store their research data in subject-specific platforms external to institutions, both IRs and CRISs are likely to become suitable systems for managing institutional research data outputs by defining the appropriate metadata standards and linking dataset descriptions to externally available files. In order for the resulting content collections to be aggregated by funders issuing data deposit mandates[13], some harmonised data description standard will need to arise that fits both platforms, allowing institutions to report on their RDM activity regardless of what specific system they use.

Finally, it should also be kept in mind that repositories are much lighter and easy-to-configure platforms than CRISs. As a result, there are whole geographical areas in the world with a fairly well-established open access repository network, very little CRIS implementation and barely any mechanisms in place yet for ensuring CRIS/IR interoperability. These areas are likely to keep extensively relying on their repository networks for meeting their research information management requirements, maybe eventually turning their eyes to open source CRISs when comprehensive research reporting starts to be perceived as an institutional need.

References

1. OpenDOAR, http://opendoar.org/
2. euroCRIS Directory of Research Information Systems (DRIS), http://eurocris.org/DRISListAll.php?order=cfTitle
3. "CRIS concept and CRIS benefits", euroCRIS website, http://www.eurocris.org/Index.php?page=concepts_benefits&t=1
4. Jackie Proven, Janet Aucock, "Increasing uptake at St Andrews: Strategies for developing the research repository". ALISS Quarterly 2011, 6(3): 6-9, http://hdl.handle.net/10023/1824
5. David T. Palmer, "Moving From An Institutional Repository To A Current Research Information System: The Why & How". CNI Spring 2013 Project Briefings, http://www.cni.org/topics/repositories/moving-from-an-institutional-repository-to-a-current-research-information-system-the-why-how/
6. OpenAIRE Guidelines for CRIS Managers based on CERIF XML, https://guidelines.openaire.eu/wiki/OpenAIRE_Guidelines:_For_CRIS
7. UK RepositoryNet+ project, http://repositorynet.ac.uk/
8. CRIS extension for EPrints, http://bazaar.eprints.org/154/
9. DSpace-CRIS, http://cineca.github.io/dspace-cris/index.html
10. HKU Scholars Hub, http://hub.hku.hk/
11. Pablo de Castro, Jackie Proven, "The STARS Shared Initiative: Delivering Repository Services in an Advanced CRIS/IR Environment". Repository Fringe 2013, Edinburgh, Aug 2013, http://www.slideshare.net/repofringe/stars-slides-rfringe13final
12. Toni Prieto, "Interoperability experiences between CRIS systems and repositories in Catalunya" [Spanish]. GrandIR Technical Session on CRIS/IR Interoperability, Barcelona, Nov 2011, http://hdl.handle.net/10609/10881
13. EPSRC Policy Framework On Research Data, http://www.epsrc.ac.uk/about/standards/researchdata/Pages/policyframework.aspx