Learning How to Play Nicely: Repositories and CRIS

Nick Sheppard reports on the event examining integrated, systemic approaches to research information management organised by the Welsh Repository Network and supported by JISC and ARMA at Leeds Metropolitan University in May 2010.

Introduction

More than 60 delegates convened at the Rose Bowl in Leeds on 7 May 2010 for this event to explore the developing relationship and overlap between Open Access research repositories and so called 'CRISs' – Current Research Information Systems – that are increasingly being implemented at universities.

The Welsh Repository Network (WRN) [1], a collaborative venture between the Higher Education institutions (HEIs) in Wales, funded by JISC, had clearly hit upon an engaging topic *du jour*. The event, jointly supported by JISC [2] and ARMA (Association of Research Managers and Administrators)[3], was fully booked within just five days of being announced. In the main, delegates were either research managers and administrators, or repository managers, and one of the themes that came up throughout the day was the need for greater communication between research offices and libraries (where repository services are often managed.)

As well as JISC and ARMA, euroCRIS [4], a not-for- profit organisation that aims to be an internationally recognised point of reference for CRISs, was represented at the event. Delegates could also visit the software exhibition and speak with representatives of Atira, Symplectic Ltd and Thomson Reuters, among others.

Overview of CRIS and Repository Overlaps and Position Statements

Why a CRIS: Andy McGregor, JISC

As a JISC programme manager who has overseen more than 60 repository projects in the last 3 years, Andy is uniquely qualified to describe the repository landscape in the UK and asserted for us just how common repositories have become both in this country and internationally [5]. He also reminded us, however, that repositories are 'lonely and isolated'; still very much under-used and not sufficiently linked to other university systems. They are often under-resourced with low levels of full-text deposit and require continuous advocacy to academics and research staff, the end-users who would benefit most.

Andy went on to suggest that the ideal partner for a repository is, in fact, a Current Research Information System, or CRIS for short. Ostensibly they manage the same data for analogous purposes and for similar end-users; they have shared interests including reducing duplication, collating data for the Research Excellence Framework (REF) and feeding other institutional systems. Moreover, they must both integrate appropriately with the research lifecycle, recording bibliographic data and documenting grant-related information, for example, in the case of a CRIS, and archiving an appropriate full-text version of a research paper in the case of a repository.

This last comparison emphasises Andy's next point, that there are enough differences in the respective requirements that neither system can fully supplant the other and it makes sense to keep them separate. CRISs, for example, are more focused on monitoring rather than maximising impact, the latter being the goal of repositories which, in turn, tend to be more focused on preservation of full-text material rather than just bibliographic data. There are also likely to be significant differences between the administrative workflow and the research workflow itself that can be managed more appropriately by systems tailored to the specific requirements of each; it is also important to remember that institutions are different and there is no one-size-fits-all solution.

Andy finished his presentation by highlighting JISC-funded work in this area including Research Revealed [6], and Readiness 4 REF [7].

Institutional Repositories: Just a Bit of CRIS?: Simon Kerridge, ARMA

Simon [8] was representing ARMA, the professional association for research managers and administrators in the UK, and works in the central research support office at the University of Sunderland. He has designed and overseen the various stages of electronic research administration (ERA) systems at the University for the past 15 years and is currently working with the University Library Service to implement an Open Access repository and integrate it with the research office systems.

Simon emphasised the multi-skilled role of research managers as an interface with the academic community. They are in a position to give advice on the practicalities of submitting a research proposal, for example, and all the associated paraphernalia (costing, pricing, contracts, ethics) as well as contributing to governance, planning and other strategic objectives.

So what exactly is a CRIS? Like repositories, perhaps, it is not easy to offer a simple and definitive definition, however, they are 'loosely defined' by Rodman and Stanford [9] 'as improving [research] administrative processes through the application of technology, particularly computer technology'. CRISs go under a variety of names including the straightforwardly descriptive Research Management and Administration System (RMAS) or Sunderland's own Electronic Research Administration system (ERA) and are essentially tools for managing research information in several main areas:

- Staff (research)
- Publications (bibliographic data)

- Projects and proposals (funding information)

- Post-graduate research

- Impact

- Ethics

- Key Performance Indicators (KPI)

Such a system can be used for a range of management requirements including strategic planning and providing data to funders; notably HEFCE in the form of the Research Assessment Exercise (RAE) and its replacement, the upcoming Research Excellence Framework (REF). Internally they need to interact with a number of other university systems and processes including Human Resources and Finance, Library and Faculty systems and student databases; while externally they should be inter-operable with specialised funder systems including those for large-scale assessment exercises like the RAE and REF. Moreover, the exchange of research data is becoming increasingly important with Web-enabled systems and this has led to the development of the CERIF data-model [10] which is explored by the JISC-funded EXRI-UK (Exchanging Research Information in the UK) Project [11] and was discussed in more detail in the case study presented by Anna Clements (see below.)

All of which brought Simon back to his original question: Is an institutional repository (IR) a subset of a CRIS? A show of hands indicated that audience opinion was split, which was appropriate as Simon's answer was an equivocal 'Yes...and No…' Like Andy, he observed that, generally, they are not managed by the same service and emphasised that research managers and administrators need to work closely with their counterparts in 'the library'.

Events such as this, he concluded, are certainly a step in the right direction.

Case Studies

The Ideal CRIS?: Anna Clements, University of St. Andrews representing euroCRIS

Anna is a data architect at the University of St. Andrews where she is responsible for establishing and leading a programme of information management improvement across the institution; she also represents euroCRIS and perhaps took a broader view of 'Who needs it and why?' than Simon [12]. While echoing many of his institutional-level drivers, she also emphasised 'political decision-makers', 'entrepreneurs and innovators' and 'media and general public,' citing researcher CVs, research bibliographies and commercial output reports as just some of the potential outputs from a CRIS.

Anna began with a broad definition of a CRIS as 'any information tool dedicated to provide access to and disseminate research information,' going on to describe the CERIF data model before providing an overview of the CERIF-CRIS that has been implemented at her home institution of St Andrews, based on Pure software from Atira.

In order to emphasise its constituent elements, Anna deconstructed the acronym favoured by euroCRIS before putting it back together in the historical context of the organisation:

- **Current** means timeliness and vitality (includes ongoing relevance not merely being contemporaneous)

  Also implicit is the dynamic nature of relationships, as staff move between institutions, or work on different funded projects for example.

- **Research information** comprises the various entity attributes required for comprehensive research evaluation (people, organisations, funding programmes, etc)

- **System** is the tools and, crucially, the data model to manage heterogeneous (meta)data from disparate systems

CERIF and euroCRIS have their antecedents in the mid-1980s when the need was identified to share information better about what research was being undertaken in institutions across Europe, in order to target funding more effectively, for example, and to facilitate collaboration. It was developed in two major phases between 1987-1990 and 1997-1999. The first iteration was relatively limited, based primarily on project information, but has become increasingly sophisticated with each release. In 2002 the European Commission authorised euroCRIS to maintain and develop CERIF and its usage [10] and the standard has become a European Union recommendation to member states.

CERIF 2008 recommends a 'core' of entities, attributes and relationships and allows entities to be separated from the semantic layer for greater flexibility when sharing local data with another institution as well as supporting multiple languages. The diagrams below are taken from Anna's slides and represent the different types of entities in different colours:

- Core entities (green)
- Result entities (orange)
- 2nd-level entities (blue)
- Link entities (purple)

The coloured loops indicate that there is recursive logic within the entities themselves with hierarchical relationships between people or within an organisational unit:
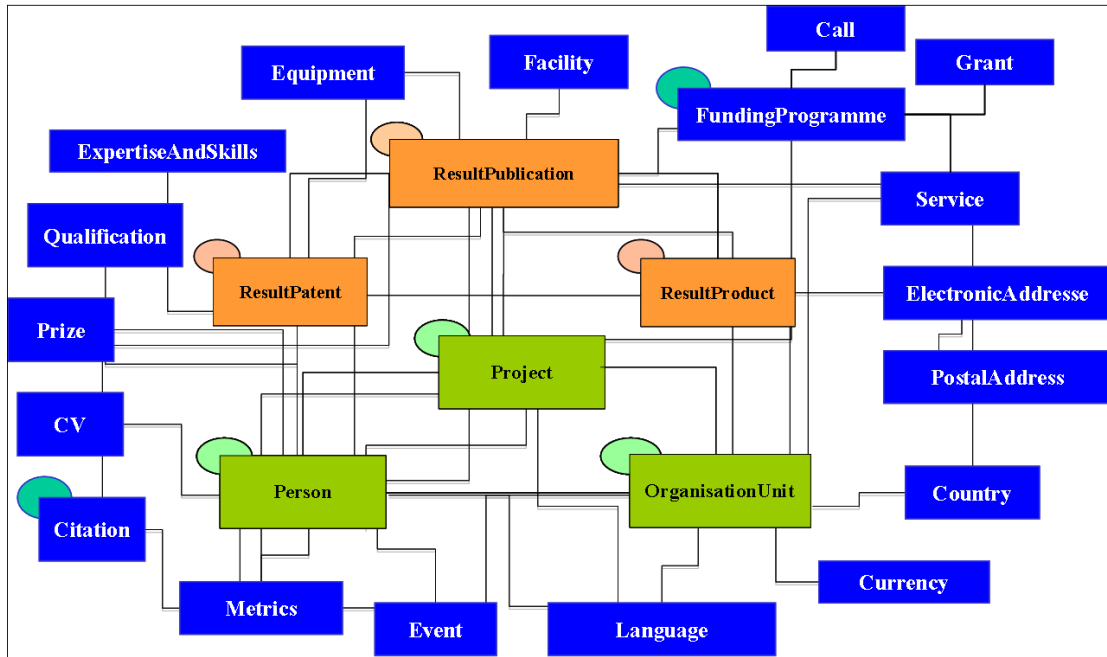
Figure 1: CERIF 2nd Level Entities (Image © euroCRIS 2010)

Link entities record the precise, time-bound relationships between and within entities, so a person might be a member of both a project and an organisational unit, for different periods of time. Such relationships can become extremely complex, and the strength of CERIF is that it is able to capture such web-like complexity with a data model that is essentially very simple.
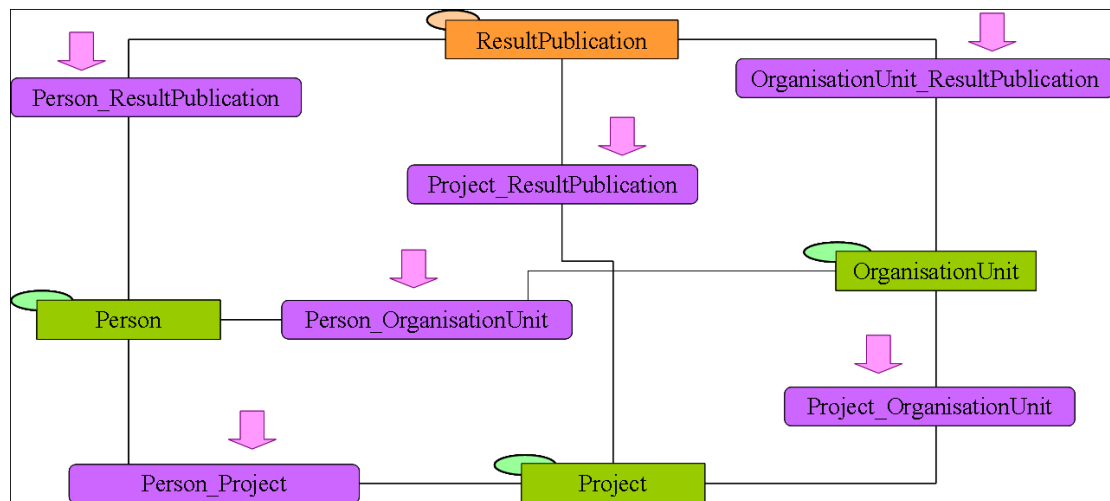

Figure 2: CERIF example Link Entities (Image © euroCRIS 2010)

Anna gave some examples of the types of questions that can be answered when research data are captured within such a sophisticated data model, pertaining, for example, to individual authors (how many articles author X published in 2007 as a first author; whether author X publishes with institutionally external authors) and/or

specific projects (how many publications have resulted from project Y; how many women have been involved in FP6 projects).

CERIF is currently most widely used in Northern Europe, in both institutional and national systems in Denmark, Ireland, the Netherlands and Norway, but uptake has not been so great in the UK. After the difficulties encountered collating data for the RAE in 2008, however, it is increasingly on the agenda as universities aim to ensure the process is easier for the REF in 2012. This was the impetus behind the JISC funded EXRI-UK Project [11] which aimed to explore 'current and future scenarios for the exchange of research information in the UK' with the specific objective 'to appraise the options and, specifically, whether any particular format for exchanging research information (eg CERIF) would be suitable' though 'linked data' and 'known Semantic Web approaches to modelling the research domain' were also considered.

The RAE, in fact, was the driver for St. Andrew's in-house (non-CERIF) research information system implemented in 2002 and linked to its DSpace repository; Anna emphasised, as had Andy and Simon, the value of working together as well as the importance of good data management and re-using data that had already been gathered. This approach naturally led to a model with the research information management system at the centre of the process fed by data from other systems:
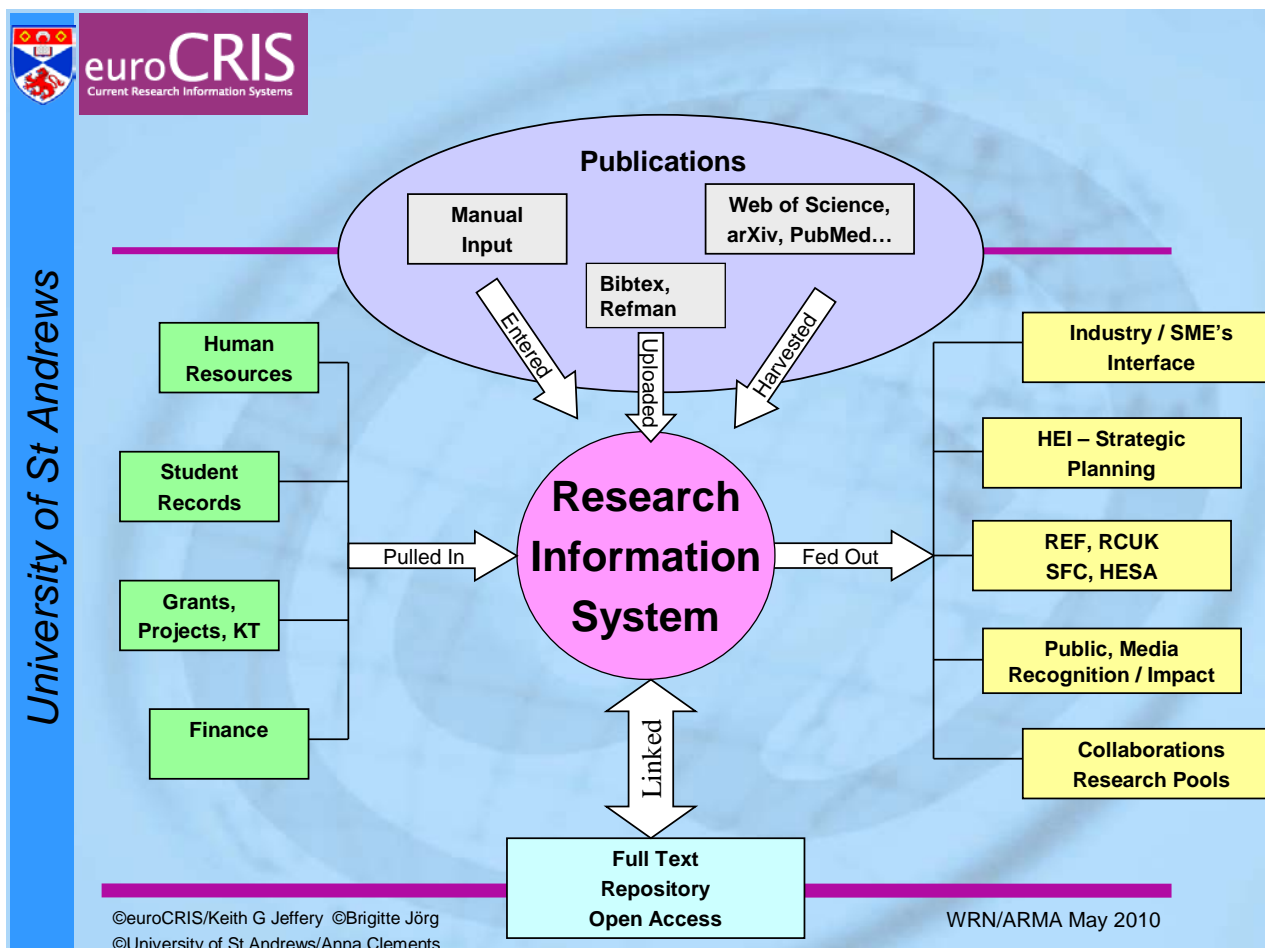

Figure 3: Schematic representation of the research management infrastructure at the University of St Andrews implemented in 2002 (Image © euroCRIS 2010)

After the RAE in 2008, the issue of a highly functional CRIS was climbing up the agenda at St. Andrews; the functionality of the in-house system was relatively basic and though it provided a system to aggregate the various data, it did not adequately facilitate processing and retrieval of those data. Moreover, as a small institution, there were not the resources available to invest in ongoing development. It had become increasingly apparent that the requirements were similar across the sector and research into available systems and links with the University of Aberdeen ultimately resulted in a joint project to tender and implement a CERIF-CRIS to serve the two institutions.

The system that has now been implemented is Pure, a commercial CRIS from Atira which is also linked to the Institutional Repository; Pure itself does not preserve full-text research outputs but is able to use the CERIF data model to link to external systems like the IR which, in turn, provides the technology to preserve full text and ensure metadata are harvested by OAI-PMH. In addition, full-text deposit to the repository is mediated through the Pure interface itself giving an integrated system for the user.

Anna's concluding message was that the repository at St. Andrews is not being subsumed but rather put into context within the broader electronic research management infrastructure. The expertise and work already invested in repository development at universities are still essential. Moreover, there is an ongoing need for Open Access advocacy, but ultimately a repository linked to a CRIS is greater than the sum of its parts and will make the job of both repository managers and research administrators easier.

An Enlighten-ed View of Repository and Research System Integration: William Nixon & Valerie McCutchean, University of Glasgow

The next case study once again illustrated the importance of appropriate liaison between research management and repository development.

William is the service manager for Enlighten, the IR at the University of Glasgow; he was the project manager for the recently completed JISC-funded project Enrich [13] and works closely with Valerie, operations manager in the Department of Research and Enterprise at the University.

The broad aims of the Enrich Project were to establish Enlighten as a comprehensive, University-wide repository and central publications database, and to improve staff profiles by linking data from core institutional systems. In addition, it sought to ensure compliance with funders' open access policies and reporting requirements as well as improving publicity for research activity and outputs.

Valerie began by giving an overview of the systems underpinning research management infrastructure at the University of Glasgow which has had a 'data-rich' research system since 1994 with links to the Human Resources, Finance and Student systems; part of Enrich has been to also link to the repository. She illustrated how the various systems interact throughout the research lifecycle to facilitate an integrated process encompassing pre-award through to post-award and project completion, whereupon an automatic email will request full text for the repository, for example:
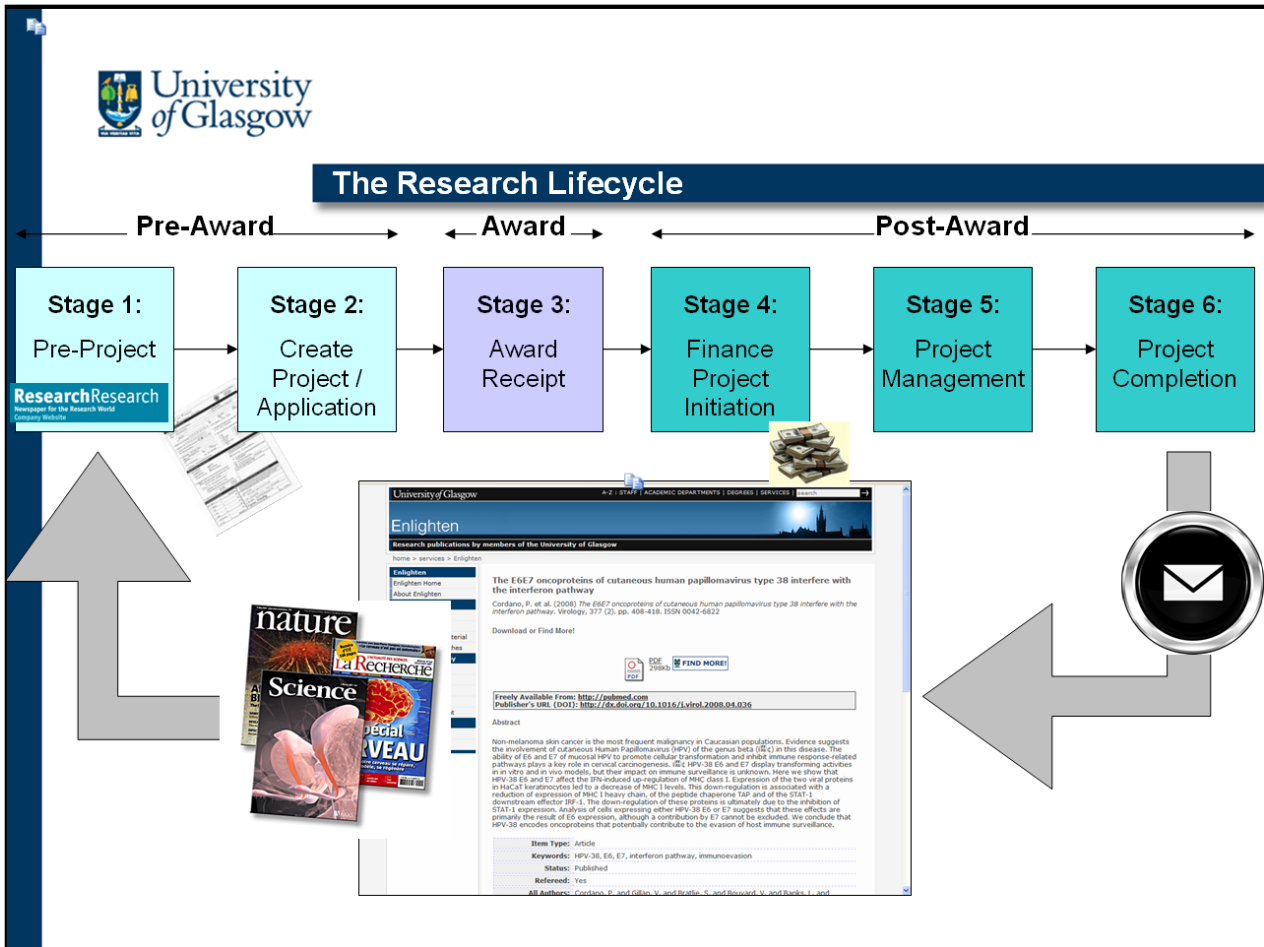
Figure 4: The Research Lifecycle (Image © University of Glasgow 2010)

The system, though based on relatively 'old' technology, broadly fulfils the requirements of research administration at the University and enables the Department of Research and Enterprise to collate and coordinate a large amount of data centrally, including staff and student information (from the HR system), funder details, internal and external collaborations, costings, ethics and awards. Ongoing development is overseen by the Research Systems User Group which enables the different stakeholder groups to communicate effectively what is required from the interoperable systems and, after Enrich, now includes members from the library as well as from faculties, HR and Finance. Staff are also trained across the different systems so that repository staff can administer the research system and vice versa.

After Valerie's introduction, William took over and spoke in more detail about their EPrints repository – branded as Enlighten - and the Enrich Project.

Rather than a formal mandate, Glasgow has a Publications Policy which essentially requests that, where copyright allows, staff deposit a copy of peer-reviewed, published journal articles and conference proceedings into Enlighten as soon as possible after publication. In addition, it also attempts to capture bibliographic metadata for all published outputs.

One of the crucial elements of Enrich has been to integrate Enlighten with the University LDAP system using a Glasgow unique identifier (GUID), meaning not only that users do not need to register a separate account to begin depositing their outputs, but also that those outputs are tied to the same unique identifier in all systemic components of the institutional research infrastructure. This makes it easier to pull together all publications for a given author and to share and reuse the data across different systems. The technique also makes it easier to feed repository data to other areas of the University Web site, so that dynamic publication lists can be added to departmental or individual academics' Web pages, for example.

An additional benefit is the way that an author view can be constructed when author records are disambiguated in this way. Many repositories (and other research systems) lack a 'name authority file' which leads to several problems. When originally entered into a system, names are likely to be keyed in different ways by different users and common names may be duplicated within an institution. Academics may also change their name, due to marriage, for example. Until this development, author listings in Enlighten (as in many other repositories) were based on the author name in the metadata record, giving an ungainly mix of 'Jim Smith,' 'J. Smith,' 'Smith, James,' etc, which were all listed separately. However, users may now browse for all of their publications under their official University designation, irrespective of how their name is entered in individual records and also including an honorific (eg 'Smith, Dr James').

Another important development is that funder information can now be extracted from the research system, incorporated into individual repository records and facilitating 'Browse by Research Funder name'. The link is bi-directional, allowing browse by funder code in the research system, so that the Department of Research and Enterprise can more easily discover publications attached to a specific project. Valerie stressed that one of the main reasons for linking awards to outputs in this way is due to the rising profile of the impact and output agenda across both the institution and the HE sector; they are keen to integrate effort across the University to identify users of research outputs and to gain feedback to illustrate impact. Work is also ongoing to ensure that Enlighten can manage other types of research outputs (exhibitions, broadcasts, artwork, etc) as well as, from an OA perspective, an ongoing focus on increasing full-text content in the repository.

The take-home message from William, Valerie and Enlighten was encapsulated in the 'three Ps':

- People: good working relationships across stakeholder groups are essential to developing an integrated research management system

- Processes: developing synergies between the different workflows of the research system and the repository

- Policies: developing a coherent institutional publications policy and working with funders' policies (eg Wellcome Trust mandate)

Where Did It All Go Wrong?: Confessions of how not to do it and lessons learnt:
Jackie Knowles, Project Manager, Welsh Repository Network

The final session of the morning was composed from anonymous 'confessions' from the community. By providing evidence from real-life situations, the aim was to address the issue of 'how not to do it/ pitfalls to avoid' when developing a repository, research management system or CRIS.

An overarching theme was 'they just don't get it', but it was far from clear precisely who and what this referred to; Jackie suggested 'they' and 'it' had, in fact, become less vague throughout the morning sessions with the overall emphasis on the different stakeholder groups and the importance of communication. The value of training researchers and managers to understand the research lifecycle better, for example, was highlighted; one should not assume that people are already aware of the issues.

Other 'top tips' were to pursue simplicity – avoid reinventing the wheel - but also recognise that there is no one-size-fits-all solution to the disparate requirements of research management and associated workflows. It is valuable to map them in some detail and to elucidate where the respective elements fit into the 'bigger picture'. Detailed user scenarios were cited as a useful way of specifying functionality; what does a user actually require from the system?

The power of statistics as an advocacy tool was emphasised and the importance of advocacy in general. It was better to launch early rather than trying to perfect the system first, and to spread the word and address non-engagement.

'Lessons learned' again emphasised the importance of communication: one story described how, during the tendering process for a CRIS, repository staff had not been involved and the repository had therefore not been sufficiently prioritised. As a result, it took 2 years' work subsequently to tailor the systems to work together.

Another interesting point was to avoid making assumptions about other stakeholders' knowledge and perspectives; not everyone will automatically and unreservedly think Open Access is a good thing, for example. Moreover, there may well be specific barriers associated with the institutional context; an OA mandate, for instance, is powerless without the means to enforce it.

By way of conclusion, Jackie offered generalised personality types that she had encountered in her own work with repositories and that may well have a bearing on developing effective relationships. These included 'the obstinate preservationists' who manage their own data and see no need to work differently; 'the endless debaters' who like to attend events and engage in discussion around the philosophy of the approach but without actually changing their working practices, and 'the non-communicators' who may be working on similar systems without engaging with related projects already underway elsewhere in the institution - or even in the same department.

Café Society Discussions

After lunch, four topics were explored in café society discussions with delegates moving between sessions throughout the afternoon; session aims are summarised below and individual facilitators have posted full reports on the WRN blog [14]

Topic 1: Drivers:

Facilitator: Andy McGregor

The session [15] was designed to explore the issues that are driving the development of research management systems, processes and policies in universities. The range of drivers was examined as well as the ways in which institutions were choosing to address the various issues. These approaches were used to develop a rough-and-ready action plan for institutions wishing to look at research management.

Topic 2: DIY vs Commercial Solutions

Facilitator: Anna Clements

The session [16] explored the pros and cons of either developing a system in-house or implementing a commercial system (or systems). It also examined the implications of moving from one to the other.

Topic 3: Stakeholder Engagement

Facilitator: William Nixon

This session [17] asked the question 'Who are the main stakeholders and how do we engage them?' The focus was on researchers, research office and repository staff – many other stakeholders were also identified including funding bodies, university management, JISC and HEFC.

Topic 4: Data Quality

Facilitator: Simon Kerridge

This session [18] framed the issue of Data Quality as 'How do we ensure data quality in our systems? What are the best methods for getting data out of legacy systems?'

Panel Q&A and Concluding Discussions

At the end of the afternoon, facilitators fed back from their respective discussions and delegates had an opportunity to raise any final questions. The session was filmed and along with all presentations from the day can be viewed online [19].

Conclusion

Institutions are all different and there is no one-size-fits-all solution. A Russell Group university, for example, will have very different requirements to a Million+ institution, and it is important to focus on precisely what is required from a research management system in your particular context. Institutions should also, perhaps, take a broader view of their requirements and be circumspect about focusing too closely on one particular driver. With many institutions, for example, concentrating on the REF, there is a danger of developing systems that are too narrowly focused on the specific requirements of that one exercise.

The availability of human and other resources, in-house expertise and (as illustrated by the contrasting case studies at Glasgow and St Andrews) the existing infrastructure of an institution will all have a huge impact on the most appropriate course of action. That course of action may include whether existing systems should be developed in-house or tenders submitted for entirely new systems. If starting with a blank slate, it probably makes sense for a CRIS to be the central system with the repository as a linked peripheral component; but, of course, very few are actually starting from this point and different models can be just as effective.

The over-riding conclusion that was reinforced throughout the day was the need for effective communication channels between research administrators and repository managers in particular, but also among the full range of stakeholders at a given institution. By working together, the disparate systems of institutional research management infrastructure can become more effectively integrated.

References

1. Welsh Repository Network - http://www.wrn.aber.ac.uk/
2. JISC http://www.jisc.ac.uk/
3. ARMA - http://www.arma.ac.uk/
4. euroCRIS - http://www.eurocris.org/
5. Why a CRIS? http://www.wrn.aber.ac.uk/events/cris/presentations/andy_mcgregor.ppt
6. Research Revealed http://www.jisc.ac.uk/whatwedo/programmes/inf11/sue2/researchrevealed.aspx
7. Readiness for Ref http://www.jisc.ac.uk/whatwedo/programmes/inf11/sue2/r4r.aspx
8. Institutional Repositories, just a bit of a CRIS? http://www.wrn.aber.ac.uk/events/cris/presentations/simon_kerridge.pptx

9. Electronic Research Administration (ERA) : Commencement, Practice, and Future, John A. Rodman and Brad Stanford, in "Research administration and management" Elliott C. Kulakowski, Lynne U. Chronister (2006), Chapter 30, p. 297.
10. CERIF http://www.eurocris.org/cerif/introduction/
11. EXRI-UK http://ie-repository.jisc.ac.uk/448/
12. The ideal CRIS? - http://coursecast.aber.ac.uk/CourseCast/Viewer/Default.aspx?id=9a976666-5734-4211-997f-e16d016d3b17
13. Enrich - http://www.jisc.ac.uk/whatwedo/programmes/inf11/sue2/enrich
14. WRN blog - http://welshrepositorynetwork.blogspot.com/
15. CRIS Event Cafe Society Write up - Group 1: Drivers http://welshrepositorynetwork.blogspot.com/2010/05/cris-event-cafe-society-write-up-group.html
16. CRIS Event Cafe Society Write up - Group 2: DIY v. Commercial Solutions http://welshrepositorynetwork.blogspot.com/2010/05/cris-event-cafe-society-write-up-group_17.html
17. CRIS Event Cafe Society Write up - Group 3: Stakeholder Engagement http://welshrepositorynetwork.blogspot.com/2010/05/cris-event-cafe-society-write-up-group_7100.html

18. CRIS Event Cafe Society Write up - Group 4: Data Quality
http://welshrepositorynetwork.blogspot.com/2010/05/cris-event-cafe-society-write-up-group_9742.html
19. Presentations and videos
http://www.wrn.aber.ac.uk/events/cris/presentations.html

Author Details

Nick Sheppard
Repository Development Officer
Leeds Metropolitan University

Email: n.e.sheppard@leedsmet.ac.uk
Web site: http://repositorynews.wordpress.com/

_____

Keywords: please list in order of importance the keywords associated with the content of your article (usually up to 8). These will not be made visible in the text of your article but may be added to the embedded metadata:

1. *Research*

2. *Repository*

3. *CRIS*

4. *Information*

5. *Data*

6. *euroCRIS*

7. *JISC*

8. *ARMA*