

Managing research publications: lessons learned from the implementation of a Current Research Information System

Cranfield University is in the process of implementing a Current Research Information System (CRIS). The system is intended to ensure a single and accurate source of Cranfield's research information to allow for effective planning, monitoring, reporting, communication and future modelling of its research activity. It will therefore play a key role in helping the University to prepare for the Research Excellence Framework (REF) in 2014. Phase one of the project, publications management, has recently been completed and the system is being rolled out to the University's research communities. The article briefly describes the institutional context, the business case and the requirements for the new system and reflects on the lessons learned from this implementation.



SIMON J BEVAN
Head of Information Systems
Kings Norton Library
Cranfield University



JOHN HARRINGTON
Head of Information Services
Kings Norton Library
Cranfield University

The institutional context

The origins of the decision taken at Cranfield University in October 2009 to acquire a Current Research Information System (CRIS) – quickly named the *Cranfield* Research Information System – can be traced back to the convergence of several institutional drivers. An analysis of the lessons learned from the data collection and management process for the 2008 Research Assessment Exercise (RAE) highlighted the need to put in place a much improved system for managing information generated from the University's research activity. At the same time the outcomes of the JISC-funded Embed Project¹ highlighted how a variety of barriers were inhibiting engagement between research communities and institutional repositories (IRs). The final report made a number of recommendations as to how IRs could become more closely integrated into the research publication workflow.

To assist the preparation of the RAE 2008 submission, an in-house publications database had been developed. A lot of work was put into its design and development and into efforts to quality-assure data. However, use of the database

was uneven across the five schools of the University, and ultimately the process of collecting and validating publication and the other research data required for the submission proved to be very labour intensive. A basic flaw was that data could not easily be exchanged between key existing information systems such as human resources (HR), finance, student information and CERES² the University's institutional repository.

Further reinforcement of the need for a CRIS was provided by the results of the Embed Project. One of the conclusions of the study was that a more efficient publication workflow aimed at reducing workload and integrating the IR with other internal and external information systems needed to be developed. The overall goal of the strategy is to move the IR more into the mainstream of research management within the institution, a process which in recent years has been referred to as 'routinization'³, that is, making sure the repository forms part of the daily working environment of academic, researchers and other relevant stakeholder communities.

The business case

A case for the acquisition of a University-wide research information management system was successfully presented and this highlighted a set of key business drivers including the need to:

- improve the institutional policy framework within which the University's information systems currently operate. The need to work with internal and external stakeholders to further develop the interoperation between the CRIS, CERES and other corporate information systems
- support the institution in meeting its mission in relation to future measurements of research quality generally, and in particular the preparations the University will be making for the REF
- provide a system which will allow academic staff to interact seamlessly and transparently with CERES. Deposit has to become an invisible part of the process by which research output is stored and managed
- enhance the user experience for depositors not simply by having recognizable interfaces, but also by the information that will be available to depositors, in particular generation of researcher profiles, usage statistics and key performance indicators relating to research income, milestone alerts and outputs for use in relation to REF
- increase the rate at which material is being deposited into CERES
- provide better integration of existing content which has been created in-house to support the provision of research skills and career progression
- develop a research profile generator to provide authors with a complete list of their publications which can be combined with related information such as esteem indicators and research and interest profiles.

While many institutions have similar business drivers, there are a number of different paths to enhance research information management. Currently there appear to be three clear choices: develop an in-house solution; re-engineer the IR; select a proprietary solution.

A number of institutions have chosen to develop their own system, most of which are able to link to already existing IRs, for example, Enrich at University of Glasgow⁴ and the Research Support

System at Trinity College.⁵ The University of Rochester reported that they have re-engineered the IR concept.⁶ A briefing document at e-prints.org described how the e-prints software could operate as a CRIS for smaller institutions.⁷ Recently a number of institutions have started to implement a proprietary solution, for example, Imperial College London selected Symplectic to link to their DSpace repository⁸ and St Andrews and Aberdeen have jointly selected Atira's Pure CRIS software.⁹

The decision at Cranfield to select a proprietary solution was based on a number of factors.

- an extremely short time frame set by the University requiring the project team to provide a publications management system within 12 months
- the importance of having a system that is able to interact with external stakeholders including funding bodies. To do this it was concluded that the system would need to be compliant with CERIF¹⁰, which currently appears to be the best available standard for the exchange of research information
- insufficient skills and resources in-house to deliver a system capable of meeting the University's requirements in the time specified
- a number of proprietary systems appeared to be approaching maturity.

Following a thorough evaluation, the decision was made to select the Converis system from Avedas¹¹ in January 2010.

Lessons learned from the implementation

The implementation project was divided into two phases. The first focused on publications management and professional activities or 'esteem indicators' in RAE terms. The starting point for the implementation was the standard Converis data model¹² which was then adapted to reflect local data requirements. Much of this opening period was therefore taken up by issues relating to data quality and data import.

Data quality

The basic data model describes the association between information about people, their organizational affiliation and their publications. Ideally, the team would have spent more time checking and

cleaning the data prior to import in order to optimize the consistency of the data. The import had to deal with the complexity of combining records from three separate systems but which contained an unknown number of duplicate items. The first system developed in-house for the RAE submission in 2008 held around 9,000 publication records; the second system, the School of Management's publications database, contained around 6,000 articles. The third system, the University DSpace institutional repository, CERES, contained around 1,000 publication records. Not only was there overlap between these systems but also the structure and quality of the data stored in each of the systems varied.

Data import

The most important part of phase one was the integration with the University HR system, which would enable the mapping of the publication records to University authors. However, this also presented a number of unforeseen issues.

The structure of the departmental hierarchy in the CRIS showed up an inconsistency with the University departmental structure in the HR system, so before personnel data was loaded into the CRIS, this issue had to be resolved within the HR system. Moreover, during the implementation, two of the five University Schools carried out major restructuring which necessitated further changes to the organizational hierarchy. Another problem was that legacy data contained details of publications written by staff who had left the University and therefore would not be eligible for submission to the REF. These articles were identified as exceptions once the integration with the HR system had taken place.

While the vast majority of publications were successfully mapped to individuals, it was inevitable that some inaccuracies would occur. It was also inevitable that while the mapping enabled the creation of publication lists for authors, these were by no means complete. This meant that as part of the roll-out academics and researchers have been asked to carry out a number of checks on their data. These include confirming and, if necessary, editing their own personal data, checking publication lists, dissociating themselves from incorrectly attributed items and where necessary adding those papers that were not contained in either of the three legacy databases.

Validation and workflow design

Following the data modelling and subsequent import, the next step was the design of the workflow, the definition of roles and the associated permissions. Here there was a clear trade-off between the desire to minimize workload while ensuring a sufficiently robust set of validation procedures to preserve data quality. The Embed study had identified fears over increased workload as one of the major barriers constraining submission to the University's IR and the project team were concerned that the same issue could arise with the CRIS. Failure to incorporate sufficient data checks on the other hand could compromise the quality of the information added post launch. Finding the right balance presented a number of issues.

As part of the roll-out strategy it was decided to present the manual addition of new records very much as a last resort. Instead, authors have been encouraged to import records automatically from three external online resources which have been integrated with the CRIS. This approach not only reduces the risk of keystroke error on input, but also saves time and effort when records are being validated.

A further policy decision which has also focused attention on this trade-off between workload and data quality has been whether or not to validate external co-authors and their institutional affiliations. The options offered to the Schools were to validate all, to validate none or to validate selected co-authors/institutions. When these options were presented to the REF strategy group which is steering the project, the recommendation was to validate all. The disadvantages of the additional workload would be outweighed by the benefits of being able to map collaborative research across different institutions as well as at an individual author level. This information could then be used to inform future research strategy planning and evaluation. The main difficulty relates to the quality of person data and the ability to apply some kind of authority control on external authors. It is to be hoped that developments from both the NAMES project¹³ and the Open Researcher and ContributorID (ORCID) initiative¹⁴ can, in the future, be integrated into the CRIS to help to address these problems.

The acquisition of the CRIS has already brought a number of important benefits to the University,

not least the positive engagement of all five Schools and the decision to adopt a common publications workflow. There is already evidence, however, that the Schools are adopting different roll-out strategies. For example, two Schools have already issued clear policy statements to indicate that they expect individual academics and researchers to be responsible for adding their own publications. In one department in another School, academics have decided to delegate responsibility to an administrator. There have also been differences in the way that Schools have applied the departmental validator role. Some have given this role to departmental administrators, some to academics, one is using a research support officer, who is also part of the project team, as a catalyst for cascading knowledge and skills, and one has been considering appointing an ex-academic member of staff for a temporary period to assist with the School roll out because they believe someone with this background has the necessary skills set.

Significantly for the Library Service, the CRIS implementation represents a shift away from the existing mediated repository service to a distributed submission process which drives the population of the repository. For the repository team who have now become library validators in the workflow this represents a change in emphasis from metadata creation to a role focused on quality and copyright checking.

Engagement and sustainability

It is still too early in the project to be totally confident that the key issues of engagement and the sustainability of the CRIS have been fully understood and resolved. It is possible, however, to identify a number of factors which are likely to affect the long-term success of the CRIS.

Embedding new skills and provision of ongoing support

Although phase one is coming to an end with the roll out of publications management, implementation of the remaining modules in phase two is still being driven by the project team. For the CRIS to be ultimately sustainable, support in terms of knowledge and skills which until now have been provided by the system supplier will need to be transferred and successfully embedded.

It is still uncertain to what extent this will impact on the work of the Library's team of information specialists who on top of their existing responsibilities for academic liaison and information literacy teaching might yet be required to become CRIS advocates and trainers.

Cultural change

To what extent the CRIS will be a catalyst for cultural change is yet to be seen but there are strong indications that the system will be used to inform key internal procedures relating, for example, to annual review, promotion and, ultimately, REF submission planning.

Resource licensing issues

The CRIS relies heavily on external providers for a variety of publication metadata, bibliometric data and journal rankings. If these data providers impose subscription charges above and beyond what institutions are already paying to be able to re-use information, this could have significant cost implications. This is something that agencies responsible for negotiating licences might want to address.

Next steps

The second phase of the project, which is scheduled to run through the first six months of 2011, will focus on the remaining integrations with the student information system to pull in data relating to research students, and with the finance system. The latter will enable the incorporation of information on funded contracts, other projects and IPR and the relation of this data to the existing information on people, organizational affiliations, publications and professional activities. Other data to be added includes internal journal quality rankings and bibliometrics. The increase in the amount of data available in the CRIS will enable the focus of the project to shift away from inputs towards optimizing outputs from the system. This involves the specification of a suite of reports and on the implementation of web services to enable information to be exported from the CRIS to dynamically update researcher profile data on the University's intranet and the externally facing web pages. The final module will ensure that the system can deliver the information required for REF submission in 2014.

References

1. Embed Project:
<http://www.jisc.ac.uk/whatwedo/programmes/reppres/sue/embed.aspx> (accessed 23 December 2010).
2. CERES:
<https://dspace.lib.cranfield.ac.uk/> (accessed 23-Dec-2010).
3. Sale, A, Advice on filling your repository. *Serials*, 2010, 13(3), 207–211; to access via DOI:
<http://dx.doi.org/10.1629/23207> (accessed 23-Dec-2010).
4. Enlighten and Research Integration (Enrich) project plan, Glasgow University:
http://ie-repository.jisc.ac.uk/348/1/enrich_plan_21may09_public.pdf (accessed 23-Dec-2010).
5. Getting content into TARA Research Support System at Trinity College:
<http://www.tcd.ie/Library/riss/tara/getting-content.php> (accessed 23 December 2010).
6. Bell, S, Case Study: Re-Engineering an Institutional Repository to Engage Users, *New Review of Academic Librarianship*, 2010, 16(Supp 1), 77–89.
7. Briefing Note: EPrints, CRIS and Research Impact:
<http://www.eprints.org/resources/CRISEPrints/Briefing.pdf> (accessed 23 December 2010).
8. Symplectic at Imperial College London:
<http://www3.imperial.ac.uk/library/find/spiral/addpaper> (accessed 23 December 2010).
9. A new research tool for the University of Aberdeen:
<http://www.abdn.ac.uk/documents/Pure-launch-email.pdf> (accessed 23 December 2010).
10. CERIF standard:
<http://www.eurocris.org/Index.php?page=CERIFintroduction&t=1> (accessed 23 December 2010).
11. Avedas:
<http://www.avedas.com/> (accessed 23 December 2010).
12. Converis data model:
http://www.avedas.com/en/core_functionality.html (accessed 23 December 2010).
13. NAMES project website:
<http://names.mimas.ac.uk/> (accessed 6 January 2011).
14. ORCID initiative website:
<http://www.orcid.org/> (accessed 6 January 2011).

Article © Simon J Bevan and John Harrington

■ **Simon J Bevan**
Head of Information Systems
Kings Norton Library
Cranfield University
Cranfield
Bedfordshire MK43 0AL, UK
www.cranfield.ac.uk
Tel: +44 (0)1234 754445
E-mail: s.bevan@cranfield.ac.uk

■ **John Harrington**
Head of Information Services
Kings Norton Library
Cranfield University
Cranfield
Bedfordshire MK43 0AL, UK
Tel: +44 (0)1234 754477
E-mail: J.Harrington@cranfield.ac.uk

To view the original copy of this article, published in *Serials*, click here:

<http://serials.uksg.org/openurl.asp?genre=article&issn=0953-0460&volume=24&issue=1&spage=26>

The DOI for this article is 10.1629/2426. Click here to access via DOI:

<http://dx.doi.org/10.1629/2426>

For a link to the full table of contents for the issue of *Serials* in which this article first appeared, click here:

<http://serials.uksg.org/openurl.asp?genre=issue&issn=0953-0460&volume=24&issue=1>