

Research information meets research data management ... in the library?

Research data management (RDM) is a major priority for many institutions as they struggle to cope with the plethora of pronouncements including funder policies, a G8 statement, REF2020 consultations, all stressing the importance of open data in driving everything from global innovation through to more accountable governance; not to mention the more direct possibility that non-compliance could result in grant income drying up. So, at the coalface, how do we become part of this global movement?

In this article the author explains the approach being taken at the University of St Andrews, building on the research information management infrastructure (data, systems and people) that has evolved since 2006. Continuing to navigate through the rapidly evolving research policy and cultural landscape, they aim to establish services to support their research community as it moves to this 'open by default' requirement of funders and governments.

CRIS-IR from the outset

At the University of St Andrews in the UK, we have had an integrated research information infrastructure since 2006. The overall architecture has remained unchanged with a current research information system (CRIS) providing tools for managers and researchers to access all research-related institutional data from corporate systems such as human resources, student records, research grants and finance. In addition, the CRIS stores research outputs, outcomes, impacts and activities either via harvesting from third-party sources, such as Scopus¹ and Web of Science², or via manual data entry by researchers.

The technology has been updated over the years, with an in-house CRIS being replaced by Pure³ CRIS from Atira (part of Elsevier) in 2010.

From the outset, the CRIS has been integrated with our open access (OA) institutional repository (IR) running in the DSpace platform. The CRIS is the single 'golden' data source for the research publication metadata and, where a full-text version can be made OA, these metadata are pushed through to the institutional repository together with the full text. All workflow on copyright clearance and embargo periods is done in the CRIS.

Thus the IR acts as a genuine *repository* of openly accessible documents.

Information management principles

The CRIS-IR is a prime example of the successful practical application of the principles of good information management:

- data is entered once, as close to source as possible, and reused
- data stewards keep control of the data within their domain of expertise
- data is available to (only) those who need it, in the format needed and when needed
- data standards, such as Common European Research Information Format (CERIF)⁴ and existing data sources, such as Web of Science and Scopus, are used.



ANNA CLEMENTS
Enterprise Architect
University of St
Andrews

"... the IR acts as a genuine *repository* of openly accessible documents."

299 Examples of data reuse include the preparation and submission of data for research assessment, e.g. RAE2008 and REF2014⁵, and to the Research Councils UK (RCUK) research outcomes system (ROS)⁶, introduced in 2012. Within the University, data from the CRIS is reused on individual departmental websites and several projects are now under way to aggregate St Andrews' data with data from other university CRISs, to present collaborations across the sector.

Figure 1 shows the CRIS architecture at St Andrews, with the red items indicating the areas for extension over the next phase of development.

This approach has also resulted in close co-operation across several university functions, principally, the Research Policy Office, the Library and IT Services (ITS). This strong sense of co-ownership of the CRIS-IR infrastructure has resulted in clear and co-ordinated communication to researchers about services and tools available and an effective mechanism for gathering feedback to improve these services. It has also ensured that we have a very strong technical and support infrastructure to build on new services, such as research data management.

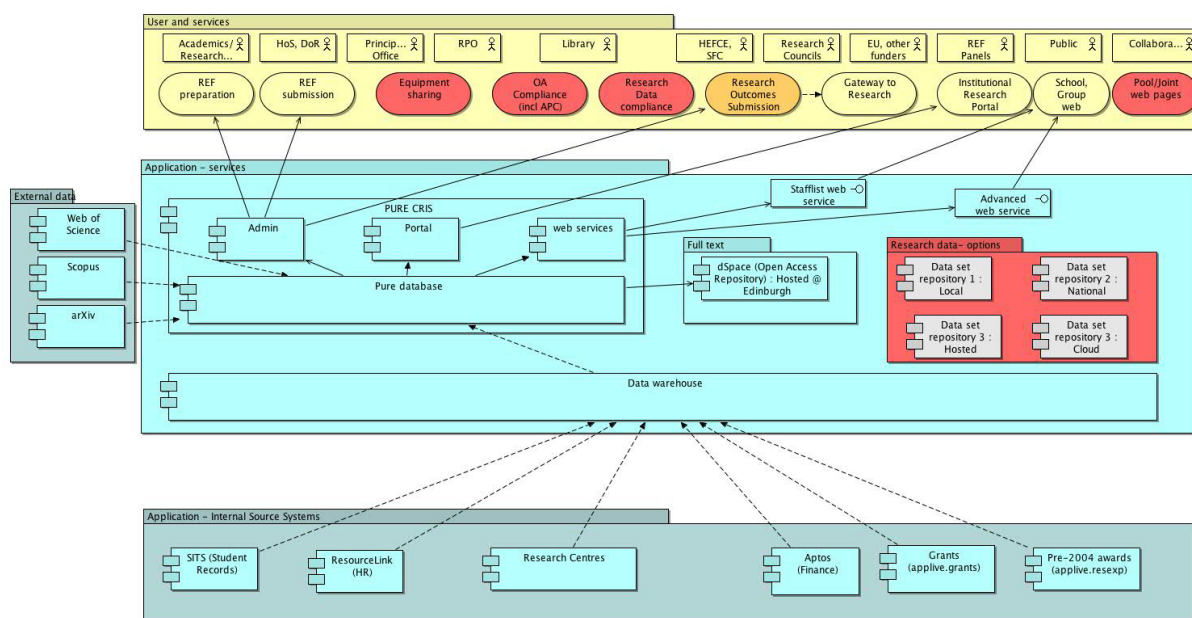


Figure 1. CRIS architecture at the University of St Andrews

Extending the CRIS-IR

In the UK, RDM (that is, the management of the data that researchers produce or use) is high on the agenda of funders, and therefore, institutions. A new joint OA policy from the seven research councils came into force on 1 April 2013⁷, which not only requires the article to be published in an RCUK OA-compliant journal but 'must include ... a statement on how the underlying research materials – such as data, samples or models – can be assessed'.

In addition, data management plans or similar are required for all grant applications to research councils, and the Engineering and Physical Sciences Research Council (EPSRC) in particular has set a deadline of 1 May 2015⁸ for institutions to have the policies, processes, infrastructure and tools in place to satisfy the main principles 'that publicly funded research data should generally be made as widely and freely available as possible in a timely and responsible manner; and, secondly, that the research process should not be damaged by the inappropriate release of such data'.

300 This move towards 'open data' is a global trend, with the G8 releasing an open data charter⁹ at the London meeting in June 2013. It supports the principle of 'open by default' and recognizes the central role open data can play in stimulating worldwide growth through the innovative and transparent use and reuse of data.

The practical consequences are far reaching, both for researchers and research organizations. Researchers need to adjust to an environment where there is the presumption that the data that they gather or generate during the research process will need to be made available with adequate metadata, software tools and related resources, to ensure that other researchers can understand the potential for further research and reuse of the data. Some provision to protect intellectual property is provided; for example, the EPSRC policy framework on research data stipulates that 'researchers should be entitled to a limited period of privileged access to the data they collect to allow them to work on and publish their results'.

It ... recognizes the central role open data can play in stimulating worldwide growth ..."

As research organizations, we need to know what data our researchers generate (who funded it, formats used, quantities, how sensitive it is, where it is, which publications has it generated, should it be preserved or can it be reproduced, can it be freely available or are there conditions of (re)use, and so on). The minimum requirement in all of this is a catalogue of research data sets, and that is where our existing CRIS-IR infrastructure comes in. We can link research data sets to the information on people, organizations, projects, funding, outputs, impacts and activities that we already have in the CRIS. The data sets themselves will be stored in various locations: internal data repositories, external discipline-based repositories, funder repositories, and so on. The key though is that the metadata for all these data sets can be collected in our CRIS and made available together with the other contextual information, and so maximize the visibility of St Andrews research activity. An added benefit is that because our CRIS is CERIF compliant, we are well placed to support interoperability with funder systems, e.g. for reporting outcomes and compliance, and with the planned national data register being developed by the Digital Curation Centre (DCC)¹⁰. As a related development we are also looking to add equipment and facilities information to the CRIS and so build further links between these and the resulting data and publications that are produced; thus providing evidence of the impact and cost-effectiveness of our infrastructure.

Sticks or carrots?

At St Andrews, we are still at the early stages of the research data management journey: concentrating on working with researchers to understand what support they require to move into this new 'open data' world. There are important external policy drivers, as mentioned earlier, and these help focus policy development and planning within the University. But winning the hearts and minds of the researchers has to be based on explaining and demonstrating the benefits of open data to them, whilst ensuring that whatever system and metadata infrastructure we provide to support this new world should, as with our current CRIS-IR, capture metadata once at the appropriate point in the research process and allow reuse throughout the research lifecycle.

We already have the means to manually register data sets as metadata-only records within this infrastructure and are encouraging researchers to do so as part of complying with the RCUK OA policy. However, this is really the end-point of the process, i.e., the registration of the data set at the point of publication. Our challenge is to work out how best to support researchers from the start of the research data lifecycle, without adding unnecessary administrative overhead. As well as working closely with the researchers in the different disciplines, the development and delivery of these support services require co-ordination across the research office, library and ITS; the same cross-functional team as already in place for the CRIS-IR infrastructure.

So what part can or should the library play?

The recent RCUK policy on OA and the accompanying pots of money available – an initial one-off amount from Department for Business, Innovation and Skills (BIS)¹¹ in early 2013 and now a recurrent grant for 2013/14 and 14/15 from RCUK¹² primarily to promote gold OA – have made it clear that those supporting OA (overwhelmingly based in university libraries) need access to detailed, accurate and up-to-date research information, including project details such as funder, start and end dates, external and internal grant references and researchers involved in the project, including collaborations with other institutions. But the RCUK OA policy is not just about making the textual article open access. As mentioned earlier, the author ‘must include ... a statement on how the underlying research materials – such as data, samples or models – can be assessed’. There is also the presumption that these data should be openly available unless there are ‘**compelling** reasons to protect access to the data, for example commercial confidentiality or **legitimate** sensitivities around data derived from potentially identifiable human participants’.

“... there has been a shift towards institutional responsibility ...”

Researchers will be looking for a single point of contact (or at least a single web portal) providing help, support and training as they struggle to understand such things as how to develop data management plans, what metadata is important to capture, how to ensure their data is secure, how they ensure their data can be understood and reused successfully, why they should share it with potential rivals and what are exemptions that can be applied, which data they should keep and which they can get rid of, which data repository is best for their data, whether subject, funder, institutional, publisher, and so on.

Although researchers themselves need to be primarily responsible for complying with the policies of those funding their research, there has been a shift towards institutional responsibility both with the OA grants provided to institutions and the EPSRC policy framework on research data. For the former, institutions will be required to report on compliance, not individual researchers. In the latter case, the institution is responsible for policy development, awareness raising, standardized metadata collection and the curation and preservation (for at least ten years for EPSRC) of research data.

Specifically, ‘Research organizations will ensure that effective data curation is provided throughout the full data lifecycle, with ‘data curation’ and ‘data lifecycle’ being as defined by the Digital Curation Centre¹³. The full range of responsibilities associated with data curation over the data lifecycle will be clearly allocated within the research organization, and where research data is subject to restricted access the research organization will implement and manage appropriate security controls; research organizations will particularly ensure that the quality assurance of their data curation processes is a specifically assigned responsibility¹⁴.

“There is a need to understand and engage with the research process from the start ...”

The Library already deals with metadata standards and the curation and preservation of resources useful to teaching and research so it is an obvious step to extend these skills to cover resources such as research data and associated documentation that are created by our researchers during the research process.

But research data management is not just about curation and preservation. There is a need to understand and engage with the research process from the start and not just pick up the research results at the end of the project, as traditionally happens with the deposit of full-text articles into the IR. We need to know what support our researchers require to allow them to better manage their data from creation to curation and then develop services to deliver this support in as effective and efficient a manner as possible. Although the Library may be the place where much of the support could sensibly be based, there will still be an ongoing need for co-ordination with others in the research office and ITS, as well as with researchers and support staff throughout the schools and departments.

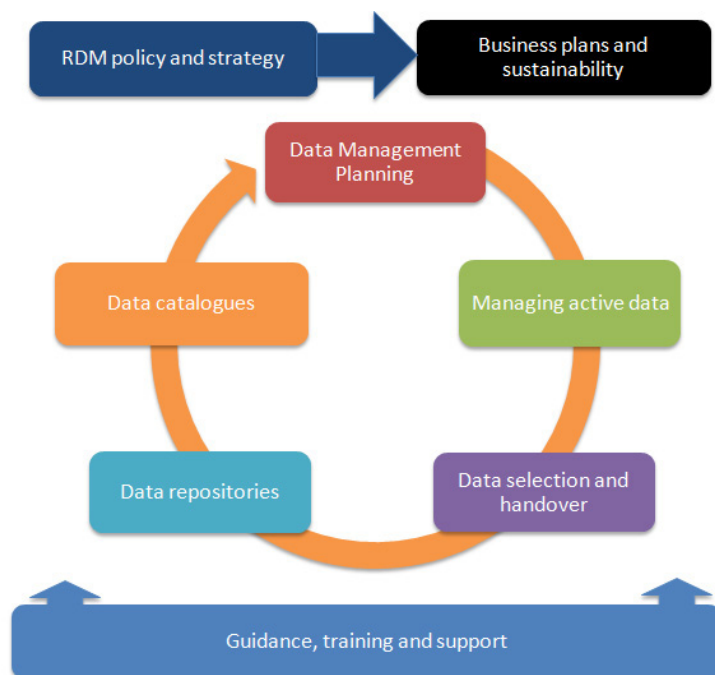


Figure 2. Components of an RDM service, Digital Curation Centre¹⁵

Systems or services; institutional or shared?

I often get asked to talk about CRIS-IR and where the responsibility for such systems should lie within an institution; or, more controversially, perhaps, 'Do we need an IR if we have a CRIS?', or vice-versa. My answer is to stop thinking about systems and think instead about services and what people, processes, tools and standards are available internally or externally that can best deliver the necessary services now and in the future. At St Andrews, we have certainly benefited from this approach and concentrated our efforts into delivering joined-up services to our researchers and research managers, whatever and wherever the system.

As we move into the new and complex world of managing research data sets, we now have data repositories thrown into the mix. Given the quantity, size and complexity of these data sets and the substantial differences in how data are understood and managed between disciplines, we will need to continue to build on this model of interoperation at the technical, metadata and organizational levels, and I expect that the boundaries between systems will become even more blurred and irrelevant. What will become important is developing cost-effective models for managing data, and in particular discussions across the sector to pool resources – between funders and research organizations – to develop sustainable solutions.

“... stop thinking about systems and think instead about services ...”

One of the key drivers for the rapid investment in CRIS in the UK¹⁶ following the previous research assessment exercise (RAE2008) was the recognition that we should not all duplicate effort building our own systems and processes to manage the next exercise (REF2014).

The experience we have had with using Pure and working with the 20 or so other institutions in the Pure UK User Group is the successful delivery of complete Research Excellence Framework (REF) preparation and submission functionality – developed collaboratively – not duplicated 20 times across the sector.

I would argue that we are in a similar position now and need to find ways of sharing the effort to deliver research data management across the sector. Much work has already been done by the Jisc Managing Research Data Programme (JISCMRD17), which has successfully

supported a substantial number of institutions across the UK as they define and develop pilot RDM services, including some very useful training materials¹⁸.

In addition, the DCC organizes many events up and down the country and produces excellent materials on case studies and best practice in RDM. Currently, it is working on the development of a national data register, similar to the Australian National Data Service (ANDS)¹⁹ which will drive the standardization of metadata requirements for research data.

What may still be missing is a national strategy on data repositories; some funders, such as the Natural Environment Research Council (NERC)²⁰ and the Economic and Social Research Council (ESRC)²¹ already provide a trusted repository infrastructure and the services to preserve that data in accordance with their own research data policies. In other subject-specific areas, such as archaeology, there is a well-established service provided by the Archaeology Data Service²². However, this is not the case for all funders.

I would argue that such shared services should be considered for all research councils as it must surely be more cost effective to design, run and sustain such services at the national rather than the institutional level.

Whatever the eventual outcome of such a debate, if we continue to develop our research information and data management infrastructure in a co-ordinated approach (data, systems and people) and to adhere to the principles of good information management, we give ourselves a fair chance of being able to cope with the next seven years of changing research policy landscape (which incidentally takes us up to the next REF exercise, due in 2020), just as we have done for the previous seven.

“... shared services should be considered for all research councils ...”

References

1. Scopus:
<http://www.elsevier.com/online-tools/scopus> (accessed 2 September 2013).
2. Web of Science:
<http://thomsonreuters.com/web-of-science/> (accessed 31 August 2013).
3. Pure:
<http://info.scival.com/pure> (accessed 31 August 2013).
4. Common European Research Information Format (CERIF):
<http://www.eurocris.org/Index.php?page=CERIFreleases&t=1> (accessed 31 August 2013).
5. Research Excellence Framework (REF), 2014:
<http://www.ref.ac.uk/> (accessed 2 September 2013).
6. Research Councils UK Research Outcomes System (ROS):
<http://www.rcuk.ac.uk/research/researchoutcomes/pages/ROS.aspx> (accessed 2 September 2013).
7. Research Councils UK (RCUK) policy on Open Access:
<http://www.rcuk.ac.uk/documents/documents/RCUKOpenAccessPolicy.pdf> (accessed 31 August 2013).
8. Engineering and Physical Sciences Research Council (EPSRC) Policy framework on research data:
<http://www.rcuk.ac.uk/documents/documents/RCUKOpenAccessPolicy.pdf> (accessed 31 August 2013).
9. G8 Open Data Charter:
<https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex> (accessed 31 August 2013).
10. Digital Curation Centre (DCC):
<http://www.dcc.ac.uk> (accessed 2 September 2013).
11. UK Government, Department for Business, Innovation and Skills (BIS):
<http://news.bis.gov.uk/Press-Releases/Government-invests-10-million-to-help-universities-move-to-open-access-67fac.aspx> (accessed 2 September 2013).
12. Research Councils UK (RCUK), Block grants for universities to aid drives to open access to research outputs:
<http://www.rcuk.ac.uk/media/news/2012news/Pages/121108.aspx> (accessed 2 September 2013).
13. Digital Curation Centre, DCC Curation Lifecycle Model:
<http://www.dcc.ac.uk/resources/curation-lifecycle-model> (accessed 2 September 2013).
14. EPSRC Policy Framework on Research Data:
<http://www.epsrc.ac.uk/about/standards/researchdata/Pages/policyframework.aspx> (accessed 17 September 2013).

15. Jones, S, Pryor, G and Whyte, A, How to Develop Research Data Management Services - a guide for HEIs. In: DCC *How-to Guides*, 2013, Edinburgh, Digital Curation Centre:
<http://www.dcc.ac.uk/resources/how-guides> (accessed 17 September 2013).
16. UKOLN, Russell, R, *Adoption of CERIF in Higher Education Institutions in the UK: A Landscape Study*, March 2012, London, UKOLN:
<http://www.ukoln.ac.uk/jisc/reports/cerif-landscape-study-2012/CERIF-UK-landscape-report-v1.1.pdf> (accessed 17 September 2013).
17. JISC Managing Research Data (JISCMRD) programme:
<http://www.jisc.ac.uk/whatwedo/programmes/mrd.aspx> (accessed 2 September 2013).
18. JISC JORUM Research Data Management:
<http://dspace.jorum.ac.uk/xmlui/handle/10949/18267> (accessed 2 September 2013).
19. Australian National Data Service (ANDS):
<http://www.ands.org.au> (accessed 2 September 2013).
20. Natural Environment Research Council (NERC) Data Centres:
<http://www.nerc.ac.uk/research/sites/data/> (accessed 31 August 2013).
21. Economic and Social Research Council (ESRC) Economic and Social Data Service:
<http://www.esrc.ac.uk/research/data/esds.aspx> (accessed 31 August 2013).
22. Archaeology Data Services (ADS):
<http://archaeologydataservice.ac.uk> (accessed 31 August 2013).

Article © Anna Clements

Anna Clements, Enterprise Architect
University of St Andrews, UK
E-mail: akc@st-andrews.ac.uk

To cite this article:

Clements, A, Research information meets research data management ... in the library?, *Insights*, 2013, 26(3), 298–304; DOI: <http://dx.doi.org/10.1629/2048-7754.99>