

Berichte und Mitteilungen

Research Information Systems – fit for the future?

A report on the situation and plans of the University of Sheffield Library

1. Introduction

This article discusses the University of Sheffield's rationale for joining, and the experience of participating in, the Development Partner Programme for Ex Libris' Esploro product. It should be noted that the article expresses the opinions of the University of Sheffield Library staff writing the article, and not the Ex Libris company.

The University of Sheffield is located in South Yorkshire in the North of England. It was founded in 1905 from a merger of local technical and medical colleges and has a prominent role in the region and city as a civic university. Sheffield is a research-intensive institution with a background in research-led teaching and learning. There is a strong relationship with local industry and the regional economy, most notably manifest in the Advanced Manufacturing Research Centre located just outside Sheffield.

The University has 25.000 full-time students, almost 3.000 undertaking part-time study. Of these, over 2.700 are studying for a degree by research (MPhil or PhD). There are approximately 1.400 academic staff across five subject-based faculties: Arts & Humanities, Social Sciences, Science, Engineering, and Medicine, Dentistry & Health. Additionally, the International Faculty is based in Thessaloniki, Greece. The University has a strong track record in research and was joint 14th on grade average in the most recent UK Research Excellence Framework (assessing the quality of research outputs).¹ In the QS World University Rankings, Sheffield ranks 82nd in the world and 13th in the UK.²

2. Current Research Environment

The University teaches and researches across a wide range of disciplines in addition to significant interdisciplinary research and collaborations from researcher level through the regional UK N8 group to strategic international partnerships. All this research activity leads to diverse requirements in terms of research information management (RIM) and research data management (RDM).

A significant proportion of research is funded through UK Research Councils (RCUK, now newly amalgamated under UKRI, UK Research and Innovation) or large charity funders such as the Wellcome Trust. This complex funding landscape leads to specific, and occasionally diverse, requirements for that impact on RIM and RDM. Policy and mandates underpinning grant applications, data management plans, persistent researcher identifiers (one variant being ORCID), institutional identifiers, and

1 "Research Excellence Framework 2014: Overall ranking of institutions," Times Higher Education, accessed 20 April 2018, <https://www.timeshighereducation.co.uk/sites/default/files/Attachments/2014/12/17/k/a/s/over-14-01.pdf>.

2 "QS World University Rankings," QS Top Universities, accessed 20 April 2018, <https://www.topuniversities.com/university-rankings/world-university-rankings/2018>.

open access publishing requirements can vary to a significant degree between funders. In addition to funder requirements, universities and researchers must comply with emergent government policy around open science and open access to research outputs.³ Hitherto, this has not necessarily been consistent with individual funder policy, the most prominent example of this being the divergent routes to open access research articles taken by RCUK and HEFCE (Higher Education Funding Council for England), though with RCUK transitioning to become UKRI, it may be converging.⁴ Open access is generally seen as desirable but the routes taken to achieve this goal are not necessarily the same. Over and above this, the University has to take into consideration the international landscape, most notably the EU, where policy although consistent in the general aims can be subtly at variance with regard to the detail around open access and research data.

Another key part of the UK higher education landscape is the Research Excellence Framework (REF)⁵ which is an exercise held approximately every six to seven years which monitors the quality of research outputs from institutions and determines apportionment of government research funding for institutions. It is imperative that institutions with a research focus do as well as possible in this exercise. In order for research outputs to be considered for the REF exercise they must be openly accessible to the public, either by paying for 'gold' open access, or by making the author-accepted manuscript available in a repository within a certain timescale, i.e. 'green' open access. All of this, along with necessary exceptions, requires managing in a systematic way.

3. Existing Systems Infrastructure

The research environment and diverse research outputs created (such as pre-print articles, published articles, theses, conference papers, monographs and datasets) has led the University to implement a range of systems to help manage these objects and provide deposit workflows and end-user discovery. Outputs have multiple facets that require coordinated management including discovery, financial management, funder compliance, links to grants and attributions to both researchers and institutions. In the past the systems infrastructure, processes and workflows used to manage outputs have developed in response to policy and the external environment. Inevitably, this has resulted in integration challenges, a fragmented overall architecture and in some cases gaps where key functionality is now mission critical.

3 "Government to open up publicly funded research," UK Government, published 16 July 2012, <https://www.gov.uk/government/news/government-to-open-up-publicly-funded-research>

4 "Open access," UK Research and Innovation, accessed 20 April 2018, <https://www.ukri.org/funding/information-for-award-holders/open-access/>.

5 "REF2021," HEFCE, Research England et al., accessed 20 April 2018, <http://www.ref.ac.uk/>.

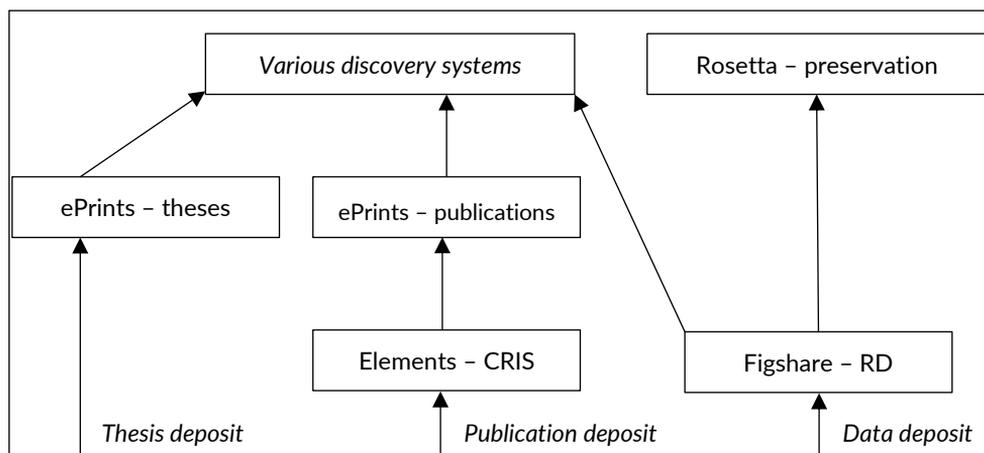


Figure: Simplified diagram of data/metadata flows in University of Sheffield systems discussed in this article

3.1. Institutional Repository (IR)

The University Library manages the institutional repository (IR) which in the case of Sheffield is an ePrints instance jointly run as a shared service with the regional White Rose Consortium.⁶ The IR was originally an output of the Jisc SHERPA Repositories Support programme and since its inception has now become a business critical part of the research systems environment. Deposits in Symplectic Elements are harvested into the repository, where Library staff check open access compliance and update metadata accordingly. The articles can then be published in the repository, fulfilling a range of public open access mandates and meeting compliance targets. Thereafter outputs are then harvested by other discovery tools including the University Library's Primo resource discovery solution and repository aggregator services.

3.2. Electronic Theses Repository

Institutional eTheses are also held in a shared service repository with White Rose. Deposit of theses (post-examination and pre-degree award) is handled by a different instance of ePrints (White Rose eTheses Online - WREO). This has a customised deposit workflow to handle specific eTheses requirements, such as supervisor details and embargoes at multiple content levels.

3.3. CRIS

For deposit of open access articles (in particular the author's accepted manuscript), the University uses Symplectic Elements.⁷ Deposit can be a manual process, though Elements has integrations into Web of Knowledge, Scopus, arXiv and with ORCID, in order to harvest metadata and allow researchers to 'claim' publications removing the necessity to re-key metadata. The University's research office is the lead stakeholder for this system as Elements is used to support managing the REF exercise.

6 "Eprints," Eprints, accessed 20 April 2018, <http://www.eprints.org/uk/>.

7 "Elements," Symplectic, accessed 20 April 2018, <https://symplectic.co.uk/products/elements/>.

3.4. Research Data Management

Some funders now require management and publishing of the data underlying the research that is undertaken.⁸ The first step in managing research data is the production of a Data Management Plan (DMP). The Library encourages researchers to use tools such as DMPonline⁹ to create DMPs, but at present there is no integration with any local systems.

Once the dataset is ready for publishing, researchers have a number of options. There are many subject-based data repositories available (for example, the UK Data Service, or the Archaeology Data Service) that the Library encourages researchers to use if they are suitable. These meet funder requirements, but make it very difficult for the University to gain a good picture of the range of its research outputs or to showcase them. In addition, there are many subjects for which there is no suitable external data repository. In response to this fragmented picture, to meet compliance mandates and to underpin showcasing of data the University acquired figshare for institutions.¹⁰ Researchers can deposit datasets into figshare, or create metadata-only records if the dataset is held in an external repository or if the data is too large or sensitive for holding in an online repository. Currently there is no way to harvest external data repository records into figshare, meaning that researchers must duplicate their efforts.

In order to improve the showcasing of the University's research outputs, the Library has collaborated with the IT department to create a custom front-end to figshare that is called ORDA. This displays research data outputs in an attractive way and is styled to match University branding. ORDA is a thin web front-end built using figshare APIs. It has also been possible to add additional services in ORDA. For example, if a researcher has created a visualisation of their dataset (using plot.ly for example), they can add a metadata file to their dataset which indicates the location of the visualisation and ORDA can then display this alongside the dataset.

3.5. Preservation of Research Data Outputs

The University Library is also concerned about preservation of digital assets including research outputs. The University's Digital Preservation function works within the Library, and the University has procured the Rosetta digital preservation system from Ex Libris.¹¹ This already preserves much material from the Library's Special Collections and archives, but work is now underway to link Rosetta with figshare and with ePrints to preserve research outputs (articles, theses and datasets). A number of challenges remain, especially with large datasets, but significant progress has already been made.

8 "Research Funder Policy Summaries," University of Sheffield Library, accessed 20 April 2018, <https://www.sheffield.ac.uk/library/rdm/funders>.

9 "DMPonline," Digital Curation Center, accessed 20 April 2018, <https://dmponline.dcc.ac.uk/>.

10 "figshare for institutions," figshare, accessed 20 April 2018, <https://figshare.com/services/institutions>.

11 "Rosetta," Ex Libris, accessed 20 April 2018, <http://www.exlibrisgroup.com/products/rosetta-digital-asset-management-and-preservation/>.

4. The Challenge

With this complex patchwork of systems come a number of problems. The connections between systems and the data flows have become very complex and difficult to manage. There are still some gaps in functionality, and areas where there is no way of managing all aspects of a process in one place. The channels of information about research, for example grants, opportunities, bibliometric data, are also limited.

For example, staff in the Library who support open access work have no way of fully managing the customer process, from the point that a researcher may contact them to the final publication of the open access copy. Different systems are involved in different parts of the work and the integrations are not good enough to allow an enquiry to be followed without staff doing significant work.

A particular difficulty with open access is the management of Article Processing Charges (APCs). Many research articles are published open access by the payment of a publication charge to the publisher (this is known as 'gold' open access). This can happen in hybrid journals where gold open access articles sit alongside articles that reside behind a subscription paywall. Alternatively, gold articles can be within journals whose only commercial model are the charges received for publication of gold open access articles. In recent years, the University Library has managed a block grant to support payment of these charges, but there are also other sources of funding. An important question for institutions is what this means for the total cost of ownership of the subscriptions.¹² Is the Library paying twice – both to publish articles and to access them? Subscriptions and the payments for access are managed via the Library's management system (Ex Libris Alma). However, there is no system which manages APCs and sources of APC funding. This is currently managed via spreadsheets. This makes analytics about the full cost of subscriptions extremely difficult. The fragmented nature of this workflow also builds complexity into the system which in turn results in higher overheads than those that would typically result from a unified and integrated system.

Additionally, links to external sources of data are not always available for some systems. In particular, it is not currently possible to harvest records about datasets published in repositories outside the University. There is a vast amount of information available that could be used to underpin the University's research enterprise, but at present there is no structured way to pull this into existing systems and to make it usable and meaningful in a resilient and sustainable way to key internal stakeholders – most importantly researchers.

Gaps in functionality also exist around linking outputs to grants, sources of funding, potential collaborations and identifying research strengths and weaknesses elsewhere. There is no one system where researchers can go to establish bibliometric information. They are faced with a variety of

¹² Cf. Stephen Pinfield, Jennifer Salter and Peter A. Bath, "The 'Total Cost of Publication' in a Hybrid Open-Access Environment: Institutional Approaches to Funding Journal Article-Processing Charges in Combination with Subscriptions," *Journal of the Association for Information Science and Technology*, 67, no. 7 (July 2016): 1751–1766, <https://doi.org/10.1002/asi.23446>.

metrics produced by different sources which they have to consult individually – this, as with all overly complex spaces, is time consuming.

There are difficulties in finding open access copies of articles, even when these exist. Tools, such as OA Button, aim to help users find copies of articles that they can access. However, the challenges ahead were highlighted when one of the authors participated in a recent test of several common tools. This was used on a test repository built for the purpose containing records from a wide array of sources, 100 % of which had an open access copy available somewhere. Results were not encouraging and suggest that there are problems to be solved. This is not just the traditional library issue of enabling users to find content, but ensuring that the institution's outputs can be found by users around the world.

5. Esploro

For all the reasons outlined above, there is a desire to bring together a number of different systems – exactly which ones depends upon what is technically and organizationally possible. The hope is to increase efficiency, and therefore free up resources for new initiatives (the data visualisation service being a recent example). It should make essentials, such as links between different types of research outputs (for example, a dataset and the publications that cite it) much easier than is currently possible. In addition, it should be possible to take workflows, e.g. for APCs, which are not managed within current systems, and integrate them properly.

The University is already a customer of Ex Libris, taking products such as the Alma Library Management System, the Primo Discovery System and the Rosetta Digital Preservation System. Therefore, there was interest when Ex Libris announced that they would be developing a research platform, now known as Esploro.

Ex Libris aim to develop a system that brings the workflows for different research objects (such as articles, datasets and theses) together. Whilst each type needs to be handled differently, bringing them together enables links between them to be easier to manage and more explicit. Ex Libris also hope to leverage the content that they have access to via ProQuest in order to make Esploro able to do as much automated discovery of research outputs as possible, reducing the effort required to deposit within the system. In particular, discovery of externally deposited datasets is a facility the University's systems do not currently offer.

Esploro would also be built with open access compliance in mind, and Development Partners from UK were recruited to bring the particular experience of UK compliance and feed that into the development. As a library systems company, Ex Libris has experience in the area of discovery, both with institutional-based discovery systems and harvesting of institutional records for external discovery systems. The discovery of the research outputs placed into Esploro will be a key part of the system.

Providing report and analytics from research systems is an essential requirement, both for reporting compliance to external institutions and for measuring performance internally. Ex Libris wanted to

work with their Development Partners to understand what reports and metrics were required in such a system and pressures on researchers are high.

As Esploro promises to meet a number of challenges the University faces with some of its research systems, and as Development Partners have an opportunity to influence development requirements to ensure that the system is fit for purpose, the University decided to join the Development Partner programme.

6. Development Partners and the Development Process

The Development Partners for Esploro are the Universities of Sheffield and Lancaster in the UK, and the Universities of Iowa, Miami and Oklahoma in the USA. The programme started in mid-2017 with Ex Libris presenting their vision for the product and doing an initial requirements gathering exercise. This was followed up by a number of webinars in different 'tracks' (i.e. topics) to discuss system requirements in more detail with staff who specialise in different areas of the research publication support.

Late in 2017, Ex Libris presented their proposal for what requirements should be included within the initial version of the product, i.e. the key elements that would make a system usable in at least one area. Development of this is underway and Partners expect further webinars and email conversations with Ex Libris analysts to clarify earlier comments and suggestions. Key priorities for the initial release include the deposit workflows for publications and datasets, automatic capturing (and claiming) of metadata records for publications, APC management, institutional repository functionality and publication to discovery services. As at the time of writing, Development Partners are expecting to be able to start testing releases of code from June 2018 onwards.

Ex Libris hope to have a system that Partners can implement early in 2019. Meanwhile, Ex Libris will be launching an Early Adopter programme, to expand the number of institutions that will contribute ideas and will help with the continuing development of the product.

As part of its contribution to the Development Partner Programme, the University of Sheffield has brought together a project team with expertise from the Library, the IT department and the research office, and experience of handling different types of research outputs. Some staff have experience of the deposit workflows and managing ingest of material, and others with discovery. Some staff have attended face-to-face meetings that have been held in the UK with Sheffield, Lancaster and Ex Libris, while others have only attended webinars for their specialist areas (the webinars feature participants from all Partners, plus Ex Libris).

Unlike many projects, the time required for this project is very uneven. During periods of requirement gathering, there can be quite a few meetings and webinars, and at other times there can be very little to do. Also, there have been very few meetings of the local project team: Most of the work so far has been to share Sheffield expertise with Ex Libris. There have just been a few local meetings to make sure everyone is aware of how the overall development is progressing.

There is a benefit just being involved as a Development Partner. There is the opportunity to reflect upon local practices, especially where workflows cross departmental boundaries. In a few cases, this has led to further meetings being set up outside the Esploro project to improve current procedures, without even requiring a new system. It is also possible to gain ideas from fellow Partners as they often have different ways of working, and Sheffield can look to adapt current procedures.

7. The Future

In the longer term, it is hoped that Sheffield's contribution will help Ex Libris to produce a research object management system that will meet requirements better than the current range of systems. As Esploro develops, Sheffield will be able to decide if it wishes to procure the system. This would probably involve replacing just one or two systems initially (integrating with the rest) with further expansion possible over time. Symplectic Elements plays a key role in the way the University manages its submission to the REF exercise, so this system could not possibly be replaced during the current cycle (due to end 2021), however, Esploro could integrate with it as ePrints currently does to provide a repository. Research data management and APC handling are other possible features to implement should the University decide to do so. It is hoped that the University of Sheffield's management of research outputs can then become increasingly efficient and that it will be possible to continue to deliver innovative services in this area.

Tracey Clarke, Andy Bussey, The University of Sheffield Library

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/2018H2S214-221>