Research data management: a case study

This item was submitted to Loughborough University's Institutional Repository by the/an author.


Additional Information:

- This article was published in Ariadne at: http://www.ariadne.ac.uk/issue74/brewerton

Metadata Record: https://dspace.lboro.ac.uk/2134/19443

Version: Accepted for publication

Publisher: © Ariadne

Rights: This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: https://creativecommons.org/licenses/by-nc-nd/4.0/

Please cite the published version.
Research data management: A case study

In April 2014 Loughborough University launched an innovative cloud-based platform [1] to deliver long-term archiving and discovery for its research data. The platform was based upon the Arkivum/100 [2] digital archiving service from Arkivum and the figshare for institutions solution from Figshare [3]. This article discusses the background and implementation of this new platform at the University.

Background

Like many other Universities, Loughborough faced a number of challenges in meeting the expectations of its research funders, in particular:

- publishing the metadata describing the research data it holds
- where appropriate providing access to the research data
- preserving the research data for a minimum of ten years since last accessed.

In addition to these challenges, the University also wanted to further promote its world-leading research and believed that one method of doing so, was by exposing the underlying data that supported the research to its peers, future collaborators and the public at large.

In autumn 2012 the University formed a Steering Committee to engage with both local stakeholders and funders to identify the issues around research data management (RDM) and recommend to the University, at large, a solution. The committee was formed of active researchers drawn from varying disciplines and at different points in their academic career, it was chaired by the University’s Research Office and included representatives of IT Services and the University Library.

One of the first actions of the steering committee was to undertake a survey of Loughborough’s research groups to determine their existing data management practices and storage requirements. The storage requirements were of particular note to the committee as unlike traditional research outputs, such as articles and reports which are mostly text, the underlying data could take a variety of formats and therefore varying dramatically in size. It was also recognised that not all the data collected by the researchers would need to be preserved, particularly where this data was derivable or not relevant to the findings of the research. Both of these factors made it exceptionally difficult to predict the storage required with any degree of certainty.
The committee also undertook a survey [4] of other Universities’ preparations for managing their research data. From the survey responses it was clear that many other institutions were either at the same stage as Loughborough, or were awaiting a commitment for additional resources (e.g. project managers, data storage) before being able to progress their plans.

![Figure 1. Chart showing status of RDM service at other institutions](image)

The steering committee were also active in advocating changes in the research environment to local academics and researchers (including PhD students). In this activity the committee were greatly assisted by the Digital Curation Centre (DCC) [5] who provided materials and speakers for local events aimed at researchers and support staff. In-house training was subsequently developed and delivered by the Library based upon these initial sessions.

With a better understanding of the business needs of the University and through engagement with local researchers, the committee developed a draft RDM policy [6] with assistance from the DCC. Key to this policy was the need for researchers to produce RDM plans in support of the expectations of their funders. The policy also reiterated the general requirements of research funders and encouraged the deposit of data that had value to the wider research community or was potentially of historical interest, even if not mandated by the funder.

The next step for the steering committee was to evaluate possible solutions to recording and managing its research data. The University already had an established institutional repository [7] for other forms of research output (e.g articles, reports, theses, posters). However the, at times, laborious nature of the deposit process and concerns over digital storage requirements, meant
this was unlikely to be suitable. Instead the committee undertook a review of the embryonic market for archiving and discovery solutions and short-listed two possible candidates:

- **Arkivum/100**: a digital archiving service from Arkivum guaranteeing long-term preservation of data deposited
- **figshare for institutions**: a new institutional offering from the figshare cloud-based research sharing repository.

Each solution answered a different aspect of the RDM problem for the University. Arkivum could provide the storage and preservation required, whilst figshare addressed the light-touch deposit process and discoverability of the research data. The steering committee therefore decided to approach both suppliers and ask them to work together to develop a platform to meet all the University’s needs.

**Proposed solution**

A proposal was submitted by the steering committee to the University in February 2014, asking for funding to develop a platform to manage research data and two posts (a research data manager based in the Library and a post in IT Service to manage the research data system) to support ongoing operation of the platform. The proposal went through a number of committees and iterations before approval and funding was provided to implement the new platform and a two tier implementation group (management board and working group) was established to replace the steering committee, chaired by the University’s Chief Operating Officer.

Key requirements identified for the platform included:

- **use the University’s existing Current Research Information Systems (CRIS) [8]** to act as a register of all research data even if not held locally
- **provide an intuitive and light-touch, user interface for depositing research data**
- **allow for metadata only records where research data is stored elsewhere (e.g. in the UK Data Archive [9])**
- **allow for the embargo of research data until a given date, or after a specific time period**
- **mint a Loughborough branded Digital Object Identifier (DOI) [10]** upon publication of the research data
- **preservation of the research data for a period of at least ten years**
- **allow easy searching of metadata and direct access to available research data**
- **provide metrics on access to the metadata and downloads of the associated research data**
to go live before the 1st May 2015.

Implementation

In September 2014 a meeting was held between Arkivum, Figshare, Symplectic (suppliers of the CRIS) and the University to confirm the scope of the project, determine responsibilities and agree milestones. As part of these discussions it was decided to manage the project using the cloud-based Basecamp [11] application and so a number of “to-do” lists were setup and tasks assigned among the development partners. For the University, the assigned tasks included:

- determine storage quotas for researchers and other depositors
- provide branding material (e.g. logos, pictures) for the interface
- configure local SAML 2.0 [12] identity provider (IdP) to work with Figshare
- provide a feed of users details.

During the meeting it was agreed that the initial launch of the RDM platform would be mediated by the University Library, to better support researchers during the transition from development project to live service. However, it was also discovered that integration between the proposed platform and the CRIS was not feasible with such a tight implementation timescale. Therefore, it was decided that this development would be better left until after the launch of the service.

This initial meeting set the scene for future face-to-face meetings which were held every two months throughout the project; of particular importance to the attendees was the continued provision of lemon drizzle cake. More regular communication between the partners occurred via the forums on Basecamp, email and through fortnightly catch-up calls using Google Hangouts [13].
The first project milestone was achieved a month ahead of schedule, with the early installation of a server at Loughborough by Arkivum. The server acted as a cache to aid with the rapid uploading and downloading of content from the University. The next milestone involved Figshare working directly with Arkivum to utilise their digital archive, as opposed to the Amazon AWS storage [14] they commonly use, for deposit of the University’s research data.

During this development and implementation phase there were still some outstanding issues in need of resolution. Chief among these issues was whether the platform needed to include an approval stage before any research data was published. Researchers consulted by the working group varied in their views on whether the approval stage was needed and importantly who should be doing it (e.g. the Research Office, a departmental administrator, Dean of School, the Library, etc). In the end it was decided that rather than create a potentially burdensome extra process, responsibility for any approval would lie with the depositor. However, it was thought that a check of the metadata by Library staff, at least in the first six months after launch, would be a good idea.

The third and most significant milestone was the delivery of the figshare for institutions service which would in effect be the interface to the RDM platform. Whilst the University awaited access to this new interface, which was unfortunately delayed by approximately five weeks, advocacy
with local researchers took a back step to other activities. One of these activities was the realisation that by agreeing for the Library to initially mediate the service, there was a need for library administrators to be able to “act as” other users of the system so as to be able to upload content on their behalf, which hadn’t been previously considered.

Functionality to allow users to “act as” other users was on the development roadmap for figshare for institutions but wasn’t planned for delivery until well after the launch of the RDM platform at the University. Rather than disrupt the existing development effort the University instead choose to implement this “act as” functionality within its IdP authentication service [15].

The interface was made available to the University in mid-April 2015 for testing. The working group uncovered a few minor issues (e.g. missing branding, licence not displayed) during testing which were quickly remedied by Figshare.

With testing complete the RDM platform was wiped clean of test content and a file of active researchers manually uploaded. Figshare were able to import some pre-existing research data sourced from the Public Library of Science (PLOS) [16] into the platform, ready for the launch.

Launch and beyond

The RDM platform was launched as a mediated service towards the end of April, four days before the 1st May project deadline. A number of workshops were organised for researchers both in
central locations, such as the Library, and in academic departments. From these initial workshops it soon became evident that having taken a back step in our advocacy during the project’s implementation phase the University now had to remind its researchers of their funder’s expectations before demonstrating the solution.

Reaction to the platform has been very positive with researchers impressed with the elegant interface and simple deposit workflow. The ability to mint, and even reserve, DOIs for their research data is valued by the researchers, as is the ability to easily “drag and drop” citations into documents.

Development of the RDM platform continues apace with new features such as projects, which are collaborative spaces for groups of researchers, appearing. Recently steps were taken to integrate the platform with the University’s CRIS whereby the metadata of research data deposited is automatically harvested and imported into the CRIS.

The Research Office and University Library are working together to embed RDM into the ongoing workflow of all researchers. In particular, the creation of RDM plans either pre-award or shortly after funding is approved is seen as a key means of promoting what is potentially a disruptive process into simply business as usual.
The future

Mediation of the RDM platform by the University Library will soon come to an end, with researchers in future taking responsibility for depositing their own research data. However, it is clear that the researchers have appreciated the second pair of eyes provided by the Library in providing a “sanity check” of their deposit before it is published to funders, peers and the world at large. Therefore the University is working again with its development partners to develop this concept into a light touch review/approval process.

Conclusions

The RDM platform was delivered on-time, within budget and has exceeded the expectations of both the University and its research funders. It is emerging as a showcase of Loughborough research and means that the University is in a fantastic position to take advantage of funding opportunities and hopefully attract future collaborators. The project is a great example of what public/private partnerships can achieve and this platform is one that other institutions could readily adopt.

However, it remains to be seen how researchers will engage with the platform in the mid- to long-term, but it is clear that advocacy will need to remain an ongoing process if the platform is going to achieve continued success.

Acknowledgements

Any project of this sort requires a legion of people to make it a success and if I was to list them all it would undoubtedly double the length of the article! So just a few thank yous. Thank you to our development partners, Arkivum and Figshare and Symplectic. Thank you and good luck to Dr Gareth Cole our recently appointed Research Data Manager. And lastly, but certainly not least, special thanks go to Dr Sue Manuel (RDM Project Manager) whose sterling work as part of the University’s Steering Committee made the following implementation project so straightforward.

References


   http://dx.doi.org/10.6084/m9.figshare.817926


   http://www.lboro.ac.uk/service/research/offcampus/docs/ResearchDataManagementPolicy-Draft.pdf

7. Loughborough University’s Institutional Repository. http://dspace.lboro.ac.uk


10. Cite your data, DataCite. https://www.datacite.org/services/cite-your-data.html


