

# Loughborough leads in implementing cutting-edge research data management

Gary Brewerton is one of the project leads for a group implementing a new research data management solution that combines *figshare for institutions*, *Arkivum* and *Symplectic Elements*. The solution will enable the institution efficiently to comply with funder mandates to make research data publicly available, and in the process increasing that data's discoverability. This will serve to showcase the institution's excellent research output, thereby enhancing its reputation globally.





Loughborough University is one of the UK's leading universities, with an international reputation for research that matters.

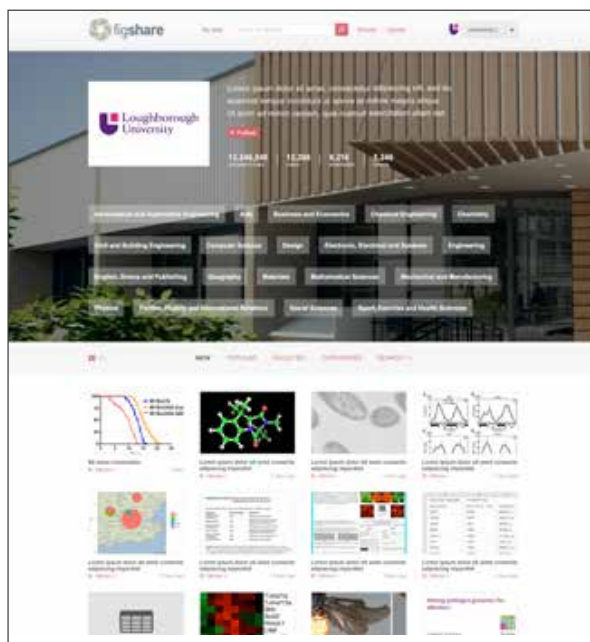
When the University started looking to implement a best-of-breed research data management solution it faced two key challenges. The first challenge was to satisfy a mandate to make research data from projects funded by Research Councils UK publicly available by May 2015. The second challenge was to further raise the profile of Loughborough research. According to Gary Brewerton, Middleware & Library Systems Manager, there was a clear and present need to improve the discoverability of Loughborough research. An investment in showcase infrastructure for research data is part of a considered effort to do that. Gary has worked with a cross-university steering committee to evaluate the most suitable solution to help address the two aforementioned challenges.

Gary stressed that there were other solutions that the University could have used to store research data to meet mandates, but it was the second challenge — making data easily discoverable — which *figshare for institutions* could provide and which was not readily available in other solutions.

### The solution selection process

Throughout the evaluation process, there was a Steering Committee with representation from each of the major stakeholders. The committee was chaired by the Director of the Research Office and also included a Project Manager and Policy and Planning Officer from the department. From the Library, there was the University Librarian and Academic Services Manager. From IT there was the Deputy Director and Research Computing Manager.

Input from faculty colleagues was also vitally important for the project and a number of academics, drawn from different disciplines, were involved in the steering committee's deliberations.



**Gary Brewerton** is Middleware & Library Systems Manager at Loughborough University. Gary leads a team that supports the growing area of middleware at the institution, specifically by writing code that joins information systems together and extracts business intelligence from the associated data. The team also maintains the various library systems such as the library catalogue, reading list management system and institutional repository.

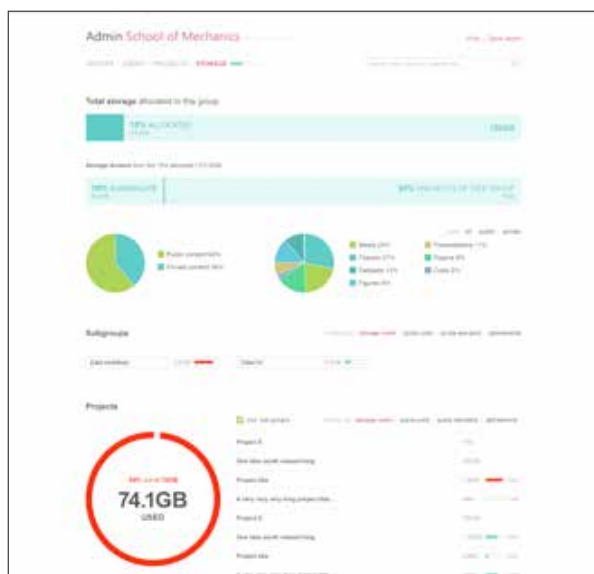


fig.1: institutional administrators can easily allocate storage to departments and researchers in *figshare for institutions*

### Researcher-centric approach

The steering committee had a variety of academics feed into the process from numerous departments including: Aeronautical and Automotive Engineering, Civil and Building Engineering, School of Science, Sport, Exercise and Health Sciences and Social, Political and Geographical Sciences. Research groups across the University were also canvassed regarding the format and size of their research data. Gary emphasized that the various faculties must be involved as there is no one set of data standards that can be used to represent them all. “From engineering, you may be getting large datasets, from social sciences, spreadsheets. Other departments are producing figures, code, media and artefacts. For an artist the research output could be an exhibit and although it may not be considered a typical piece of research data, it must still be captured and stored.”

### Awareness of figshare

Whilst canvassing Loughborough research groups it became apparent that some were familiar with *figshare*, and in some cases were already using personal accounts. These users tended to come from disciplines where collaboration was key and they were using *figshare* as a means to attract potential research partners from other related disciplines.

### On a journey of discovery

The Steering Committee worked for over a year to develop and institute policies on what research data they needed to store. They also developed training to meet the requirements of researchers and support staff. Having consulted widely within the institution, their next step was to look at suppliers and systems that could deliver the right solution to support Loughborough’s requirements.

### What IT has to say?

The biggest technical issue facing the IT department was storage space. They looked at the assortment of data that is being produced, where and how it would be stored and what the guarantees around the data would be. Gary explained that the group was looking to understand how the diversity of data could be stored in a cost-effective manner.

“We were very practical from the start, recognising that some data would be too large and too extensive to manage easily and cost effectively. We were looking for a solution that would cater for 90-95% of our research data.” The group were also cognisant of the fact that in some cases, funders do provide repositories for research data and therefore the institution wouldn’t need to also store this data. However, in such cases there was still a need to record where this data was being held. As they were also trying to build a platform to showcase Loughborough’s research, if there was data stored elsewhere, they would still need to be able to link to it.

“From engineering, you may be getting large datasets, from social sciences, spreadsheets. Other departments are producing figures, code, media and artefacts. For an artist the research output could be an exhibit and although it may not be considered a typical piece of research data, it must still be captured and stored.”

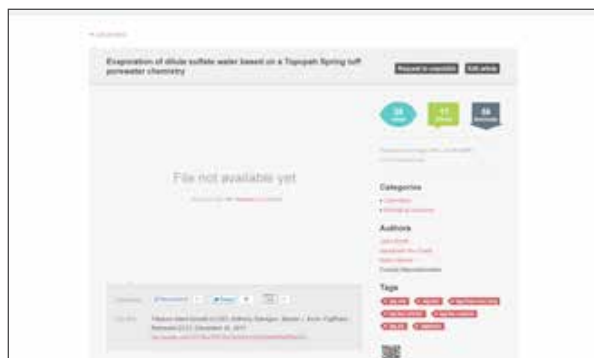


fig. 2: a file held under an author request embargo in *figshare* for institutions

### Shortlisted solutions

The Steering Committee looked at a number of alternative solutions before drawing up a shortlist of three:

- *Arkivum* – the group were very impressed by the data storage and setup. Arkivum was able to provide excellent guarantees on data integrity, but did not include the platform for promotion and discovery to fully meet the institution's requirements
- *DSpace* – open source software that was already in use by Loughborough to host its institutional repository containing research publications. However, there were concerns about how it would handle large datasets and also the need for additional local resources to support the service
- *figshare for institutions* – this provided a ready made platform to showcase Loughborough research. The group were impressed by the feature-rich solution, in particular the visualisation of figures, media and presentations, the automatic generation of DOIs (Digital Object Identifiers) and the potential for greater collaboration that the service offered

Gary stated that during the discussion of the relative merits of each solution it became obvious that no one solution would satisfy all of Loughborough's requirements. Instead the group took the unusual step of challenging two of the suppliers, *Arkivum* and *figshare*, to work together to provide a complete solution for the institution. A challenge that both suppliers readily embraced.

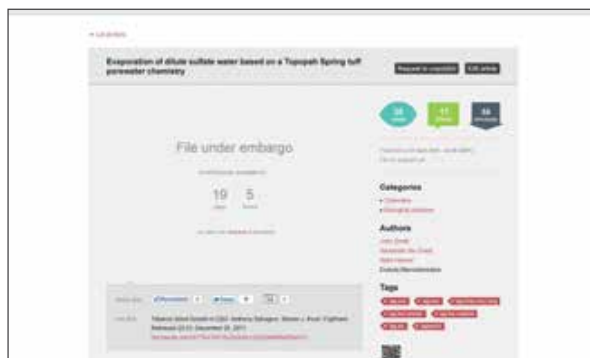


fig. 3: a file held under a time-sensitive embargo in *figshare* for institutions

## Moving forward

With the decision made and contracts signed, the Steering Group was reconstituted as a Project Implementation Group chaired by the institution's Chief Operating Officer. Gary says that it was at this point that he formally became part of the project as one of two Project Managers assigned to it from the IT department.

Work is now underway provide Loughborough with the research data management solution it requires. *Arkivum* have already installed a physical appliance at Loughborough to provide efficient uploading of content and have exposed access to their APIs (Application Programming Interface) to *figshare*. Meanwhile *figshare* have been busy creating the landing page for Loughborough's research data.

One key development that Loughborough has requested within *figshare for institutions* is the ability for a department to approve the publication of any data. This is particularly important where funding for the research is from a commercial source which may insist upon prior agreement before data is released.

The next big step is to pilot the service with Loughborough academics, which will be starting shortly. According to Gary, they hope to get a representative group of academics drawn from varying disciplines and with different research outputs.

### Working in partnership

Communication is a vital aspect of the project. There are ongoing meetings between the project partners, *Arkivum*, *figshare* and Loughborough University that occur approximately every two months. There are also regular fortnightly video calls between projects leads for the three partners to monitor progress and keep each other informed of ongoing developments.

### Many parts working in harmony

Authentication to *figshare for institutions* will be via single sign-on using Loughborough's existing IdP (Identity Provider) and profiles for research active staff provided from the institution's HR system.

*Symplectic Elements* is already well established at Loughborough for delivery of publication information, and future development work is proposed to change the way that *Elements* integrates with *figshare*. "We intend to have a single interface of research publications and research data". This is scheduled to happen in the second phase of the project roll out.

### What will success look like?

This is a key question for the institution. Success could be measured through improved citation counts for academics, greater prominence for the institution in national and international research-based league tables, increased collaboration with external researchers and further success in obtaining research funding.

We will follow and report on the progress that Gary and his colleagues at Loughborough make with this ground-breaking roll-out later in 2015.

### Researcher involvement is critical

Seven research active academics on the Steering Committee itself and numerous research groups canvassed. And with further academics involved in piloting the project, Loughborough will get a wide spread of research data from different fields.

### The moving parts of the selected solution

#### **Arkivum**

Large-scale, long-term, safe digital archiving

#### **figshare for institutions**

Live data – citable, shareable and discoverable

#### **Symplectic Elements**

To manage publications and other research outputs

#### **DSpace**

Pre-prints and Green Open Access publications

### Classes of data to be retained

#### **1 Primary data generated during the original research project.**

- Primary datasets are potentially large, e.g. multiple Terabytes for engineering.
- Access in the future will be infrequent if ever.
- Requirement is to keep the data for compliance with research council funding and for possible future re-use within the institution.
- Open access to primary data is likely to be the exception rather than the rule.
- Primary data will often be restricted, e.g. because it is not in a form ready for open access, or because of the confidentiality constraints from creating data with partners, e.g. industry.

#### **2 Derived/summarized data used to support publications.**

- Derived data supports publications and is made openly accessible along with the publications.
- Derived data might be a 'slice' of the primary data, or it might be a summary form created from the primary data.
- The size of the derived data is a lot smaller.
- Access is likely to be a lot more frequent to the derived data than to the primary data.

## How it all works

