Institutional case studies - Reading, Sheffield Hallam and ULCC

Robert Darby (Reading), Eddy Verbaan (Sheffield Hallam), Rory McNicholl (ULCC)
Session outline

- Institutional contexts
- Selecting and implementing
- Lessons learned and top tips
- Implementation examples
Sheffield Hallam: large/teaching intensive

Fourth largest university in the UK and the largest modern university (34,718 students)

- Dedicated research staff: 126.12 FTE
- Dedicated teaching staff: 1435.39 FTE
- 24% of all academic staff spend 1 day or more on research (approx 374.76 FTE)

Research grants and contracts total to £8.5m:

- 7% (£0.6m) from RCUK
- 35% (£3m) from EU grants
- 41% (£3.5m) from contractual research with government, local authorities and health authorities

In the top 5 of modern universities in REF 2014; research was particularly strong in Art and Design, Architecture, Education, Sport and History

Clear ambition to improve our ranking: a new and ambitious PVCR, preparations for the next REF are in full swing

“make contributions that have benefits in regional, national and international spheres ... [we are] sought after by partners in business, industry, the professions, government and charities ... developing practical solutions and adding value to their activities.”
RDM at Sheffield Hallam

- Key driver: compliance with EPSRC requirements
- RIC working Group (IS&T, LIS, RIO, academics) drew up the EPSRC Roadmap
- Library Resources was appointed business owner because of synergies with publications repository
  - Business case for 2-year fixed term 0.5 FTE Research Data Manager was approved in July 2013 and recruited in March 2014
  - Library restructure from January 2016 includes new Library Research Support team, but no dedicated role for RDM
- A new PVC Research and Innovation (arrived January 2014) has been a driving force
- Systems:
  - ePrints publications repository, [SHURA](#)
    - externally hosted at Southampton
  - Our RMS is Converis
    - externally hosted, currently only the pre-award module is implemented to replace infoEd, may be a future contender for institutional data catalogue
  - Research Store for ‘live’ data was established in December 2013
Data at Sheffield Hallam

- DAF inspired survey December 2014
  - Data is varied, often with complex hierarchical relationships
  - Respondents expected to produce on average 632GB of data for currently running projects
  - 6% anticipate to produce large datasets (>1TB) especially in engineering, health and sport research
  - Data preservation is erratic and involves own storage solutions: only 7% had deposited in a repository before and 12% was planning to do so at the end of their current project
  - Reluctance to share data because of confidentiality (54%), ethical issues (37%) or commercial value (23%)
- Follow-up survey June 2015
  - Respondents anticipated that 41.4% should be archived
  - Archival needs by 2017 for n=20: 5.5TB
  - Estimated maximum total need: approx 36TB
The project

- Work led by RIC Working Group, since March 2014 a Steering Group, but delivered by Research Data Manager in close collaboration with all stakeholders in a University-wide role

- Project objectives:
  - establish a catalogue and archive for digital data
  - provide training and guidance to researchers to develop best practice
  - develop policy and governance frameworks and appropriate processes

- Staff budget is owned by LIS, repository budget is owned by IS&T

- Professional services stakeholders have different contributions
  - RIO focuses on pre-award and compliance (DMP)
  - IS&T focus on short-term and long-term storage, limited resources
  - LIS focuses on repository and awareness raising
### Project timeline

<table>
<thead>
<tr>
<th>Period</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>May-October 2014</td>
<td>Policy that re-iterates RCUK and EPSRC requirements. The Records Retention Schedule was updated to reflect EPSRC requirements for all research data</td>
</tr>
<tr>
<td>September-December 2014</td>
<td>Initial requirements gathering (DAF, interviews)</td>
</tr>
<tr>
<td>November 2014</td>
<td>Repository requirements specification with MOSCOW prioritisation</td>
</tr>
<tr>
<td>September-November 2014</td>
<td>Exploration of solutions and choice of supplier (in-house developed document management system for metadata with in-house data store, ePrints, figshare)</td>
</tr>
<tr>
<td>December-April 2015</td>
<td>Agreement on work and procurement</td>
</tr>
<tr>
<td>May-September 2015</td>
<td>SHURDA is being built</td>
</tr>
</tbody>
</table>
Project status

- SHU Research Data Archive (SHURDA)
- Digital and paper data, with related workflows and policies
- Live on Monday 28 September, but development work continues in order to meet original requirements fully
- Not yet widely publicised - working with pilots in close collaboration with staff and doctoral students
- Free service
- Mediated deposit (Research Store to SHURDA)
- 3 EPSRC funded datasets
  - 1 simple file (Excel spreadsheet)
  - 2 complex datasets (different spectroscopic files with the raw data underpinning figures and tables; raw data organised by chapter)
University of Reading: small(ish)/research-intensive

Scale
- ~13,600 students; ~4,300 postgraduates
- ~1,700 academic staff; ~1,100 research-active
- ~1,150 articles in 2014 (Scopus)

Research awards and income 2013-14
- Income from research grants and contracts: £34m
- New Research Council awards: £11.6m (NERC: £5.3m; EPSRC: £3.7m)
- UK/EU government bodies: £12m; charities: £4.5m; industry: £4.5m
Research present and future

REF 2014

- 48th in UK by research quality overall; 19th by research intensity (adjusted for no. of researchers submitted)
- 78% of submissions internationally excellent (3*/4*); 27% world leading (4*)
- World-leading strengths: Agriculture, Food, Environment (Meteorology), Business and Management, Built environment, Typography

Research Strategy

- Enhance research achievement and status by 2020 - ambitious targets
  - Increase research income
  - Higher ranking in next REF
- New research structure being introduced
  - No longer organised by Faculty/School
  - Five subject-based Research Themes, 37 Research Divisions
RDM at Reading

- Key drivers: RCUK, EPSRC - compliance, management of institutional assets
- Former Dean of Science (left July 2014) a moving force for RDM
  - Business case to University Executive Board in October 2013 to appoint RDM Project Manager and implement research data archive
- RDM Project Manager appointed on 3-year fixed term April 2014
  - Made permanent as Research Data Manager in February 2015
  - Planning to appoint RDM Support Officer mid-2016
- Based in and funded by Research and Enterprise Services (RES)
  - Service development led by RES, but IT and Library have been involved
  - Steep learning curve for IT and Library
- EPrints publications repository, CentAUR
  - Locally-hosted, based in Library, supported by EPrints Services
- No CRIS or plans to procure one
  - CentAUR being developed with EPrints Services to provide CRIS capabilities
## Implementation timeline

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>DAF Survey</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formation of Working Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Requirements specification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluation of solutions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business case submitted for approval</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business case approved</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RDM Policy published</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Repository service procurement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Implementation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DataCite procurement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research Data Archive launched</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Project status

- **University of Reading Research Data Archive** launched 7 September
  - First iteration service in live testing phase
  - Publicised widely - early adopters sought to help test and develop service
  - No charge during test period - business model yet to be decided

- **Full set of policies and procedures** - first versions for reality-testing

- **Self-deposit service but mediated deposit process for time being**
  - Pre-deposit consultation
  - Observed deposit session - assistance from colleagues in RES and Library

- **Slow start - but opportunity to learn and improve**
  - Five datasets so far - relatively small and simple
  - Feedback and issues logged and fed into development plan/reported to ULCC
  - Large dataset deposit use case - development requirement
Constraints and requirements

Constraints

- Budget - small
- 1 May 2015! (EPSRC)
- No internal development resource
- Existing systems

General requirements

- Integrated solution ‘out of the box’ - or as close as possible
  - Catalogue
  - Publication platform
  - Archive
- Installation, support and development outsourced
- Hosted service preferred
- Core service tried and tested
- Existing users/community
- Ability to develop service in response to evolving requirements
EPrints-ReCollect with ULCC-Arkivum

- **Integrated solution with standard components**
  - Repository software (EPrints) with standard data repository plugin ([ReCollect](#))
  - Archive storage in ULCC data centre under sub-contract with Arkivum
  - Tried and tested with [UEL](#) and [LSHTM](#)

- **Fully hosted, managed and developed service**
  - ULCC well-established provider of hosted services to HE and FE
  - Many years’ expertise hosting and developing EPrints repositories, understanding of RDM
  - Strong customer service focus, supported by SLA
  - Interest in developing solutions/standard functionality of wider potential benefit
  - Value for money

- **Established user community**
  - EPrints publications repositories
  - Various universities using ReCollect to develop data repositories
Lessons learned 1

- EPrints-ReCollect is not quite an ‘out of the box’ solution
  - Needs adapting to local context and requirements
  - Basic installation retains publications repository features and configurations
- It takes work to define features, configuration and functionality to fit requirements
  - Metadata profile: full field set, mandatory/optional fields
  - Reconfiguring publications repository elements (page phrases, admin functionality and workflows, etc.)
  - Deposit and admin workflows
  - Reporting
Lessons learned 2

- Experience commissioning/developing software services is helpful
  - Requirements specification
  - Implementation planning and project management
  - Acceptance testing
  - Development planning

- EPrints ‘previous’ will give you a head start
  - Takes time to find your way around, learn what you can do
  - Documentation not always available/easy to find

- ULCC is a great knowledgebase
  - Will help you acquire the knowledge and skills to take ownership of your Eprints
Top tips!

- Define your functional requirements as precisely as possible
  - Use existing resources, e.g. Leeds, Jisc - but specify to local context

- Take inspiration from existing services
  - UEL, LSHTM, ReShare, Edinburgh
  - Use examples to model your metadata profile, workflows, policies, etc.

- Use your service provider’s expertise and resources
  - ULCC will advise, point you to examples and give you access to a demo repository

- Involve local stakeholders in implementation
  - A working group including Library, IT, and Research Office stakeholders will enable informed development and help spread awareness/understanding of RDM in key service functions
  - IT services will be essential to local integration of your solution and must be involved in planning and liaison with service providers
Implementation examples

● Access/authentication and user profiles
  ○ Authentication methods
  ○ Getting user data into the system

● Deposit workflows
  ○ Different models of how users will interact with the service

● File upload
  ○ Different permissions and processes
  ○ Challenges of large datasets

● Reporting and compliance management
  ○ Reporting on access and use against policy requirements
  ○ Examples: RIOXX, DMAOnline
Over to Rory...
Access, authentication and user profiles

The repository’s database of existing creators
The institutional staff database
External source of researcher profiles

Creator data autocomplete from multiple sources
Deposit Workflows

Five(ish) workflow stages for SHU

Review, amend and deposit

Deposit Item: My Dataset

Please use this page to check the record before depositing. You can use the "edit item" buttons at any point to amend any details.

Location of data file(s) → Upload → Core Metadata → Further details → Deposit

Agreement: I grant the University a licence of use on the terms set out above.

Unspecified fields: Location of Digital Data Files or number of boxes for non-digital data

Upload

Document: Slideshow

Type: Slideshow, Downloadable by: Registered users only. License: Creative Commons Attribution 4.0.

Core Metadata

Email: shurda@shu.ac.uk
Creators: McNeill, Rory
Because document upload is not done by the researcher, they are given the opportunity to record important access metadata at the item level.
Robert: “We have a researcher who would like to deposit in our repository a large dataset, of about 140-150 GB in 2 GB units in its current state… The researcher does not want to have a long flat list of files…”

Rory: “Zip/tar the files and upload via http(s). But your repository has a 10GB upload limit, so we’ll probably need to transfer the files to ULCC and we’ll ingest them into the repository (somehow)”

Robert: “hmmm…”

Rory: “Although…”

[to be continued...]
Reporting and compliance

Using existing statistics to report

Extend RIOXX reporting to report on Data compliance...

Data on last access: When (and who?)
Reporting and compliance
Future developments

- Reporting and compliance
  - Reporting for policy compliance, i.e. usage and date of last access
  - Efficient retention and disposal
- ORCID integration
- Local data deposit linked to remote metadata
  - ULCC and Arkivum working on this
- Upload/download of large datasets
  - ULCC and Arkivum developing prototype solution using ownCloud
- Preserving dataset directory/hierarchical structure
  - The interpretive logic of the dataset
File upload - The large data question