The ADMIRe Project
A Data Management Infrastructure for Research: successes, outputs and challenges

Caroline Williams, Director of Libraries and Research & Learning Resources
Drivers for RDM

1. **Data sharing** – amongst collaborators and more widely (the Open agenda)

2. **Risk management** – security and compliance

3. **Data preservation** – for reuse, longevity of evidence, and avoidance of loss

4. **The challenge of the data deluge** – growth in compute power; technical storage and retrieval requirements

5. **Compliance** - demonstrating compliance with funder and other policy requirements
ADMIRe aimed to …

Develop an infrastructure to support the research data lifecycle, acknowledging & responding to differing practices across disciplines. Steered by a research data management policy endorsed at the highest level.

In order to …

1. Improve research data management capability
2. Extend opportunities for data reuse
3. “Open up” research data
**Project Outputs**

1. **RDM requirements & analysis**
   - RDM survey
   - Faculty use cases
   - Technical requirements catalogue

2. **Technical infrastructure**
   - System model & technical architecture

3. **Support infrastructure**
   - Website
   - Training
   - Awareness raising

http://www.nottingham.ac.uk/researchdata/
RDM Survey Results: Types of data

Types of research data created or worked with:

- Documents: 16%
- Spreadsheets: 14%
- Websites: 7%
- Notebooks: 7%
- Databases: 6%
- Questionnaires: 6%
- Audio/video tapes: 7%
- Photos/films: 7%
- Slides/specimens: 7%
- Digital Objects: 7%
- Raw data: 6%
- Models/Algorithms: 6%
- Contents of an application: 5%
- Other: 3%
- 1%
RDM Survey Results: Data storage

Places where the research data is stored

- Hard disk drive of campus computer: 18%
- Hard disk drive of off-campus computer: 12%
- Hard disk drive of laptop/netbook: 6%
- External hard drive: 10%
- USB/Flash drive: 14%
- CD/DVD: 2%
- Email client/server: 4%
- Floppy disk: 2%
- VHS/Video cassette: 2%
- Audio cassette tape: 2%
- Photographs: 2%
- Slides: 2%
- Microfiche: 2%
- On paper: 2%
- Other: 2%
Research data management training

1. Developing a research data management plan
2. Storing data
3. Creating metadata for data
4. Documenting data (details of methodology, equipment used, details of physical specimens etc.)
Areas where help is required

- Greater data storage capacity: 18%
- A UoN repository to publish data: 17%
- Data management support when writing a research proposal: 14%
- A Research Data Management website for guidance and support: 12%
- Support regarding sensitive data: 10%
- Support to publish data to external subject repositories: 9%
- Help with analysing data: 6%
- Help to make better use of final data sets (e.g., create website to showcase data): 5%
- No help required: 2%
- Also required: 5%
Focus Groups and Use cases

Data types and typical ways of working

Arts & Social Sciences

• Large spectrum of research data types each with its own dynamics: quantitative datasets; text-based data; multimedia data sets; qualitative research data in multiple media, including physical artefacts and ephemera; and personal research data

• Public data sets are very commonly used

• Need to digitise data that is held in paper form looms large

• Data on individuals is held in some cases and needs to be anonymised

• Sharing outside the university is often essential involving Dropbox, Mendeley and shared websites
Focus Groups and Use cases

Data types and typical ways of working

Engineering

• Many research groups; some inter-disciplinary, some distributed
• Data size varies greatly
• Issue with “passing the baton”
• Collaboration on shared documents
• Generally data is not currently made publicly available and external data sets are used by a number of researchers
• In general all iterations of data sets are retained
Focus Groups and Use cases

Data types and typical ways of working

Medicine & Health Sciences

• Research is typically carried out in research groups and projects outlives individual researchers
• Large files and datasets are often generated (of terabyte proportions), and data is never discarded
• Collecting and using data on individuals is a regular occurrence, all data needs to contain a reference to ethical approval
• Lab books are used as standard
• It is common for data to be deposited in external repositories (national, international)

Note that, unlike other faculties, Dropbox is not used for sharing as it is regarded as insufficiently secure.
Focus Groups and Use cases

Data types and typical ways of working

Science

• *Research groups* will often cross department and institution boundaries, involving commercial companies who have commissioned the research
• No data sets above 5TB currently
• *Lab books are* an essential part of workflows
• *Sharing outside the University* – often needed, but an appropriate way to swap large data sets internationally has yet to be established
# RDM Focus Group Results: Priorities

<table>
<thead>
<tr>
<th>Data Set</th>
<th>RELEVANCE &gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Metadata</strong></td>
<td></td>
</tr>
<tr>
<td>Description of assets, such as Title, Author</td>
<td>High: 13, Med: 10, Low: 1</td>
</tr>
<tr>
<td><strong>Paradata</strong></td>
<td></td>
</tr>
<tr>
<td>Use of assets, such as Activity, Actor, Context, Date, Volume</td>
<td>High: 5, Med: 7, Low: 9, Zero: 3</td>
</tr>
<tr>
<td><strong>Identity</strong></td>
<td></td>
</tr>
<tr>
<td>Allocation of a unique digital identity to each asset (URI, DOI)</td>
<td>High: 9, Med: 10, Low: 4, Zero: 1</td>
</tr>
<tr>
<td><strong>Files</strong></td>
<td></td>
</tr>
<tr>
<td>The digital objects themselves or related assets</td>
<td>High: 18, Med: 3, Low: 2, Zero: 1</td>
</tr>
<tr>
<td><strong>Stuff</strong></td>
<td></td>
</tr>
<tr>
<td>Real world artefacts that need to be referenced</td>
<td>High: 8, Med: 6, Low: 8, Zero: 4</td>
</tr>
<tr>
<td><strong>Vocabularies</strong></td>
<td></td>
</tr>
<tr>
<td>Standardised terms used in metadata and paradata</td>
<td>High: 6, Med: 12, Low: 3, Zero: 2</td>
</tr>
<tr>
<td><strong>Licensing</strong></td>
<td></td>
</tr>
<tr>
<td>Explicit licensing as open data (e.g. Creative Commons)</td>
<td>High: 2, Med: 7, Low: 12, Zero: 1</td>
</tr>
<tr>
<td><strong>Copyright</strong></td>
<td></td>
</tr>
<tr>
<td>Necessary statements</td>
<td>High: 4, Med: 9, Low: 8, Zero: 3</td>
</tr>
<tr>
<td><strong>Links</strong></td>
<td></td>
</tr>
<tr>
<td>Links to internal and external systems (ePrints, CRIS, RC)</td>
<td>High: 5, Med: 8, Low: 9, Zero: 1</td>
</tr>
</tbody>
</table>
RDM Technical Infrastructure: Active Data
RDM Technical Infrastructure: Data Archive
Managing research data is becoming increasingly important. We provide our researchers with training, support and guidance in line with our Research Data Management policy and the findings from our research data management survey.

Equally important is the sharing of data with other researchers and for the greater public good. This site offers guidance for researchers, whilst also showcasing how our data is used to create world-leading research.

What is research data?  
Research data lifecycle  
Data management plans

Creating data  
Organising data  
Sharing & archiving data

Creating data showcase  
Research data showcase

Training and support  
Contact us

- Email us for more information
- Research Data Management flyer
In order to …
Improve research data management capability
Extend opportunities for data reuse
“Open up” research data

<table>
<thead>
<tr>
<th>Successes</th>
<th>Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Requirements and use cases</td>
<td>1. Who owns managing research data?</td>
</tr>
<tr>
<td>2. Raising awareness</td>
<td>2. Awareness and engagement</td>
</tr>
<tr>
<td>3. Working across Professional Service Departments</td>
<td>3. Complexity and scale</td>
</tr>
<tr>
<td>4. Learning from others e.g. Jisc projects and DCC</td>
<td>4. Developing a sustainable service model</td>
</tr>
<tr>
<td>5. Conceptualisation of the technical space</td>
<td></td>
</tr>
</tbody>
</table>
Questions, reflections and comments?