Digital Preservation Case Study
digital-preservation@glasgow.ac.uk
In 2017

- Senior Management Group
- Information Governance Group
- Information Policy and Strategy Committee
- Research Planning and Strategy Committee
- University Librarian
- Digital Preservation Steering Committee
  - Research Information Management Team
  - Archives and Special Collections
  - IT Services
  - Data Protection and Freedom of Information Office
Our Records

- Extensive traditional collections
- Research data
- Administrative records
- 'Estimate' 2% require digital preservation
2017 What we would like to achieve

- Digital preservation service for key business and cultural assets
- Robust systems and processes that address our essential requirements
- Good practice exchange with others - minimise the cost, maximise the benefit to the tax payer
Current Digital Preservation Project with Jisc 2017-2019

- A **catalyst** to address digital preservation within our institution.
- Not alone - whole **community** of practitioners - collaborate and share experiences

https://universityofglasgowdigitalpreservation.wordpress.com/
2019 What material are we talking about?

Jisc Open Research Hub (RDSS) pilot testing:

Business records
- Stored and managed in business systems (Documentum)

Archives/special collections
- Stored in EMu (Content Management System)

Research data
- Stored in Enlighten (EPrints repository with Arkivum storage)

Theses
Stored in Enlighten (EPrints repository)

Anyone know what EMu means?
## Considering preserving these kinds of files:

<table>
<thead>
<tr>
<th>Collection</th>
<th>Use</th>
<th>Challenge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exam Papers</td>
<td>Revision</td>
<td></td>
</tr>
<tr>
<td>University Calendar</td>
<td>Legal, Corporate Memory</td>
<td>Old versions of word</td>
</tr>
<tr>
<td>Committee Papers</td>
<td>Legal, Corporate Memory</td>
<td>Old versions of word</td>
</tr>
<tr>
<td>Emails e.g. Principals</td>
<td>Corporate Memory</td>
<td>Selection and Appraisal Data Protection</td>
</tr>
<tr>
<td>Staff Records</td>
<td>Corporate Memory,</td>
<td>Selection and Appraisal Data Protection</td>
</tr>
<tr>
<td>Compliance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student Information</td>
<td>Corporate Memory</td>
<td>Selection and Appraisal Data Protection</td>
</tr>
<tr>
<td>Research Data</td>
<td>Compliance, Research Use</td>
<td></td>
</tr>
<tr>
<td>Estates and Buildings</td>
<td></td>
<td>Selection and Appraisal</td>
</tr>
<tr>
<td>Plans and Reports</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Jisc Pilot

Requirements gathering

Records 'selection'

Summary existing systems and processes

Identify lead contacts

Testing
2019 Where are we now?

- Initial understanding of digital preservation tools (Jisc project testing)
- Initial understanding of funder requirements
- Seeking support for next phase
Spence, A., McCutcheon, V., Mahon, M. 2019. 'Case Study: The University of Glasgow's digital preservation journey', Insights: The UKSG Journal (Accepted) http://eprints.gla.ac.uk/179930

Blog: https://universityofglasgowdigitalpreservation.wordpress.com/

(includes link to requirements document)

Policy: www.gla.ac.uk/myglasgow/it/regulationscommitteesandpolicies/digitalpreservation
What might be included in a business case?

Quality of data

Curated data is easier to manage and assess
Can DP activity fit into existing workflows eg. accession of business records into archive?

How can we demonstrate value for money?

What is the cost of not doing?
What is ‘value’ from funder perspective (especially for research data)?
DP Service planning (1)

Different potential levels of service based on different workflows:
DP Service planning (2)

Different possible levels of service based on different workflows:

FULL AND INTEGRATED PRESERVATION
We proposed the latter – central preservation and review.

- £AMOUNT allocated in first year

- Work to define the scale of the service and possible activity

- Develop/revise business case for future service
Barriers to Preservation

Despite our attempts to encourage planning

Many barriers

Not necessarily technical….
Barriers to Preservation - Licencing and Sharing

Jisc

Guides - with wider application...

Ownership of Rights Research Data

Making Research Data Available

Choosing a Licence Research Data

Using Research Data

https://datasetlicencing.wordpress.com/outputs/
Barriers to Preservation – Licencing and Sharing

https://www.jisc.ac.uk/open-research-hub
Barriers to Preservation – Licencing and Sharing

Licence

Licences define how your work can be re-used by others. If you are unsure of the appropriate licence for your work, contact your supervisor.

What licence should this dataset be published under? *

Creative Commons

- Creative Commons Attribution 4.0 International (CC-BY-4.0) Learn more (opens in a new window)
- Creative Commons Attribution Non Commercial 4.0 International (CC-BY-NC-4.0) Learn more (opens in a new window)
- Creative Commons Attribution Non Commercial No Derivatives 4.0 International

Guide: Choosing a licence for research data

https://www.jisc.ac.uk/open-research-hub
Barriers to Preservation – Persistent Identifiers

To make it easy to find the data

To encourage users to cite your data

Zenodo Digital Object Identifier (DOI) recommend if need to cite and not using University of Glasgow repository.

Only supports items in GitHub

Not everyone uses / wants to use GitHub
Barriers to Preservation - Quality

- Some staff and students produce code ad hoc
- May lack knowledge about good documentation
- Long term - better organisation will result in better code
- But more time, effort and learning now
Barriers to Preservation – Tools

https://researchnotebooks.wordpress.com/
Barriers to Preservation – Guidance from Funders

Wellcome Trust - Data Management Plan (DMP) information

- McNair, Kara
- Detailed information on data management requirements for Wellcome Trust funding applications.

<table>
<thead>
<tr>
<th>Wellcome Trust</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data management plan required?</td>
<td>Yes</td>
</tr>
<tr>
<td>RDM policy Link</td>
<td><a href="https://wellcome.ac.uk/funding/managing-grant/policy-data-software-materials-management-and-sharing">https://wellcome.ac.uk/funding/managing-grant/policy-data-software-materials-management-and-sharing</a></td>
</tr>
<tr>
<td>Definition of data</td>
<td>Datasets generated by research. Original software, new materials generated (antibodies, cell lines, reagents) - this is particular for output management.</td>
</tr>
<tr>
<td>Length of data management plan</td>
<td>700 words</td>
</tr>
</tbody>
</table>

https://rdmfunderrequirements.wordpress.com/
What else are we doing?

More work with Jisc, DPC, CASRAI and others……

Terminology

Encouraging sharing of data

Help with specific formats and topics
Thank you

digitalpreservation@glasgow.ac.uk