

Aye Preserve in the West of Scotland, 20th March

Updates

Valerie McCutcheon, [Research Information Management, University of Glasgow](#)

Research Information Management team looks after preservation of research data. However, aware that archive collections and corporate records also require digital preservation and wanted to put in place service to preserve key business and cultural assets.

Got involved with **Jisc's digital preservation shared service project** in 2017, which has been a catalyst for action and helped us build valuable links with a community of preservation practitioners. However, main project work carried out by one part-time member of staff, so still limited resources.

Looking at: corporate records, archives and special collections, research data, theses and audio-visual files. Other document types for future consideration include exam papers, university calendars, email.

[No-one else in the room appears to use document management systems].

Jisc pilot involved defining institutional requirements, identifying current practices for managing files, testing preservation systems with a variety of record types and analysing workflows.

Where are we now?

- Initial understanding of tools.
- Undertaken Jisc project testing of preservation systems.
- Initial understanding of funder requirements.
- Seeking support for next phase.
- Just published case study on our preservation journey <http://doi.org/10.1629/uksg.461>

Maintain blog where we share what we are doing
<https://universityofglasgowdigitalpreservation.wordpress.com/>

Issues for **business case**

- Easier to manage data if it is curated.
- Can digital preservation fit into existing workflows?
- How do we demonstrate value for money? What is the cost of doing nothing? Difficult to assign a value to the risk as we do not know currently how much data we are talking about.

Have proposed a full digital preservation service to senior management team. Currently working to define scale of service and possible activity for next stage.

Barriers to preserving research data

- **Data licensing and sharing:** researchers are unsure whether they have used other people's content appropriately or don't know the best way to licence re-use of their own data, so do not submit their research output for preservation at all.

University of Glasgow is working with Jisc on a project to address licensing. Have produced guides for researchers covering: ownership, making data available, choosing a licence and using research data. See [Dataset licensing blog](#) for discussion of activities.

- **Persistent IDs to discover and cite data:** problematic. Discoverable and durable IDs encourage people to cite data. Researchers are encouraged to choose the most suitable data repository for their research outputs. They can put their metadata in the University of Glasgow repository and add a link to their data. At Glasgow we recommend [Zenodo](#) data repository.
- **Quality:** researchers do not always create good quality metadata to accompany their research outputs. Documenting code can be a particular problem. Long-term, better organisation will produce higher quality metadata, but requires more time, effort and learning now.
- **Researchers' tools:** e.g. paper research notebooks are not easily accessed and searched. University of Glasgow currently working on project with Jisc to explore good practice and options for research notebooks. Information on [Research notebooks blog](#).
- **Funders** do not lay down specific requirements about preservation of outputs. This is non-restrictive but makes deposit/preservation harder to enforce. We have recently carried out a [systematic review of funder requirements](#) for managing research data.

What else are we doing?

- Working with Jisc, Digital Preservation Coalition (DPC), [CASRAI](#) and others.
- Digital preservation terminology: ran workshop at CASRAI Re-connect 2018 conference looking at the [DPC glossary](#); supporting DPC and CASRAI on updating/expanding the glossary.
- Encourage data sharing.

Q & A

Comment: within company, each team manages its records independently, rather than using SharePoint or TEAMS. Asking them to undertake more robust records management procedures would be perceived as a burden, adding to their workload.

Qu. Has there been any advocacy work with funders to get them to encourage preservation?

A. Have spoken to funders but they are not pressing researchers on this. The research councils (UKRI) are in flux at present. There may be more harmonisation among them, which may lead to changes, but there are many other funders. Talk a lot about FAIR (Findable, Accessible, Interoperable, Re-usable) research data practices but no real implementation.

Qu. Are you archiving software to read digital content which requires non-standard software?

A. Jisc is looking into this, but there are licensing issues. It would be useful if legal deposit libraries such as the National Library of Scotland could obtain permission to archive software. NLS response to this: they cannot do this currently but are advocating on the issue.

Paul Stokes, Jisc

Jisc is a not-for-profit membership organisation dealing with digital services and solutions for Higher Education Institutions (HEI) and others. Operates shared digital infrastructure and services. Launched **Research Data Shared Service (RDSS) project** in 2016 to investigate creating a subscription service for managing and preserving research data.

Drivers for Research Data Shared Service project:

- Research integrity and open research agenda important to universities
- Mandate from funders for open access to research outputs
- GDPR and related legal/access issues
- Need to address risks and costs of managing digital data
- Sustainability.

Requirements for preservation solution:

- Needs to be multi-tenant (one installation serves multiple separate users)
- Suitable for wide range of content types
- Modular and interoperable
- Positive user experience
- Make digital preservation easier for researchers
- Reporting functionality
- Community governance and input.

Just launched [Open Research Hub](#), an interoperable system for managing, preserving and sharing institutional digital material. Current focus is research data but open to all organisations looking for support with digital preservation. APIs allow hub to integrate with a number of other data management/preservation systems and repositories. Agile product, so development is ongoing.

Now looking at preservation of wider range of material e.g.

- Theses
- Audience beyond HEI e.g. Police
- Records management and archives
- Dynamic preservation of websites

The Open Research Hub **data model** has been created as a tool not just for the Jisc project but for wider use by the preservation community. Documentation and [diagram](#) are available on GitHub <https://github.com/JiscRDSS/Canonical-data-model>.

Other supporting activities:

- Research data management toolkit
- Commissioned report on what research data should be preserved: [Neil Beagrie \(2019\) 'What to Keep: a Jisc research data study'](#) .

Q & A

Q. The first question people (should) ask when thinking about buying a system, is how to get data out again if they want to stop using the system. How has Jisc addressed this?

A. While testing various systems, Jisc found getting data in or out of them presented challenges. Open Research Hub project has concentrated on open APIs to facilitate interoperability with other systems.

Q. How much does it cost to use Open Research Hub?

A. Depends on various factors: how much content you are storing, size of organisation, where you are storing your data, which services you want to use, etc. No cost calculator available at present. Jisc happy to work with individual organisations to find best option for them and discuss costings.

Sara Day Thomson, [Digital Preservation Coalition \(DPC\)](#)

Not-for-profit organisation. Working with others to build:

- More responsive and better-informed institutions
- Good practice and standards
- Prepared workforces
- Supportive community
- High quality and sustainable services.

Capacity building

- Commission/write technology watch reports: each focuses on a specific topic. Forthcoming reports include preserving e-mail and preserving software.
- Run briefing days.
- Working with Jisc on the Open Research Hub and digital preservation. Looking at software preservation, assigning values/costs to digital preservation, and the Preservation Action Registry.

Preservation Action Registry

This is a collaborative project to develop a set of protocols and tools to enable preservation actions (procedures such as virus checking or file format identification) to be shared across platforms and repositories. The Preservation Action Registry would enable one organisation to acquire actions from another system and share good practice.

Details about the project can be found on [Open Preservation Foundation website](#) and [Preservation Action Registry project](#) website.

Good practice and standards

- Contribute to the development of good practice and standards e.g. revision of NDSA levels of digital preservation <https://ndsa.org//activities/levels-of-digital-preservation/> .
- Project work with specific clients e.g. Nuclear Decommissioning Authority feeds into the development of policies and guidance for wider community.

Prepared workforces

- Resources on website e.g. handbook (currently being updated), short guides and tools.
- In-person training around country. Events posted on digital-preservation@jiscmail.ac.uk and DPC website.

- Scholarships to attend training and conferences.

Supportive community

- About to publish an Executive Guide on Digital Preservation, which will contain tools and messages which practitioners can draw on to advocate for digital preservation within their organisation.

High quality and sustainable services

- Stream/record number of DPC events to widen access.
- Offer video calls, open resources and informal meetings.
- One-to-one support for members on individual issues.

Q & A

Qu. How does the DPC see its role with regard to skills training on digital literacy and preservation awareness for students and future researchers?

A. The DPC's role is to provide resources to support trainers (many already exist). Also provides scholarships to enable people to access training and other opportunities e.g. attend conferences.

iPRES Conference 2022

[iPRES](#) is an international conference on digital preservation, held annually in September. The 2022 conference will take place in Glasgow. The DPC is looking for volunteers to get involved and assist with the event: please contact the DPC for more information.

Group Discussion

Issues

1. Advocacy and ongoing funding for digital preservation within organisation

Suggested approaches:

- Tailor message for audience
- Emphasise impact and cost of risks e.g. fire or data loss
- Try to calculate the risk of something happening e.g. 1 in 20 risk of losing half your data. However, it can be difficult to quantify occasional risks.
- Understand costs to help make the case and present digital material as an asset rather than a liability. DPC is trying to speak to accountancy profession to identify methodology to define digital material as an asset: any case studies where organisations have costed data as an asset would be welcome.
- Well-organised and managed data is of more value than disorganised data landfill.

Curation Costs Exchange <http://www.curationexchange.org/> . Useful online tool to help organisations articulate preservation costs.

- Break down preservation activities into short-term and long-term goals and tackle them incrementally.
- Try and obtain initial funding for scoping exercise, then seek further funding to deliver activities identified.

Related issues: lack of awareness of need for digital preservation within organisations, lack of quality metadata, poor quality material submitted for preservation. Need to work with users to improve preparation of content ready for preservation. Time/workload pressures make compliance difficult.

2. Cost and practical issues of **preserving very large deposits** in repository

One organisation received 28 TB deposit of data from researcher. Issues:

- Researcher had not spoken to funder at start of their project about preserving data, so preservation costs not included in budget.
- Did not liaise with university Research Data Management team prior to deposit either.
- Research Data Management team has 'take anything' policy but faced challenge of a) uploading data into the repository and b) ongoing storage costs of £70,000 annually.

Did try and negotiate with researcher e.g. did everything need to be kept, could files e.g. TIFF files be converted into alternative, smaller formats? Problem: researcher had moved onto another project and did not have time to review and amend content. Have ensured that this researcher now budgets for preservation in future projects but need to ensure that situation does not arise with other researchers.

Group comments:

- Jisc is currently looking at issues relating to preserving large datasets
- Deposit agreements are a useful way of defining the researcher's role in preservation and what the archive can do for researchers
- How much research data is re-used? High download statistics do not necessarily indicate high re-use.
- Researchers are not always good at creating metadata and organising data ready for deposit: this is an issue which needs addressing within the HEI sector.
- National Library of Scotland has moved to JPEG 2000 uncompressed images files: much smaller than TIFFs.

Successes

a) University of Strathclyde: implemented **Archivematica**.

Put together a business case a few years ago, used for archives and research data management. Would not have got it for archives on its own, research data management agenda has more influence. Has not solved all problems, but it has made a substantial difference.

Key issue: ensuring secure and complete transfer of content into system.

Archivematica User Group meeting - 16th May at Strathclyde University, see [UK User Group webpage](#) for information and contact details if interested in attending.

b) Several people have had **digital preservation policies** approved.

National Library of Scotland: initial version not very good. Worked with DPC to produce a new version and consulted with many more stakeholders. New policy approved end of January 2019; includes 'going forward' plan.

[Core trust seal certification scheme](#) for preservation repositories. Even if you do not want to apply for certification, the requirements provide a useful checklist of issues to consider when planning for digital preservation within your organisation e.g. access, level of content curation.

The [DPC handbook](#) also contains useful guidance about what to include in a policy and ways of setting it out.

c) Audio-visual file formats

National Library of Scotland moving image archive: production files are huge.

Investigated FFV1 lossless video format to see whether it would be a suitable preservation format. Established that there is no loss of data when original files are converted to FFV1. Compression is impressive – reduces file size to one third of original.

Next steps

Attendees found it helpful to hear what others are doing and to see examples of good practice. Suggestions of ways to continue discussions and share good practice or offers to host another Aye Preserve event would be welcome.