

# Case study: the University of Glasgow's digital preservation journey 2017-2019

This case study documents the University of Glasgow's digital preservation journey during 2017 and 2018. The University recognized that action was required to ensure the long-term preservation of key corporate records and archival material. Staff from the University's Digital Preservation Working Group were therefore tasked with identifying the University's priorities and requirements for preserving its key records, with the aim of producing recommendations for a preservation programme. Knowledge and skills were enhanced by participating in a national digital preservation pilot project and learning from practitioners through workshops and information exchange. The case study shares our reflections on the questions which emerged about metadata, workflows and integrating systems. A key priority will be to engage the support of key decision makers within the University, as it was emphasized repeatedly that successful digital preservation depends as much on resources and organizational strategy as it does on technology. Two of the authors have a particular interest in terminology and we share our work to examine digital preservation's confusing and obscure vocabulary. We conclude that transforming digital preservation into standard practice within organizations can best be achieved through continued collaboration within the digital preservation community.

## Keywords

Digital preservation; terminology; metadata; record keeping; RDSS; CASRAI



ALISON SPENCE

Research Information  
Administrator: digital  
preservation  
University of Glasgow



VALERIE McCUTCHEON

Research Information  
Manager  
University of Glasgow



MATT MAHON

Research Information  
Officer  
University of Glasgow

## Introduction

The University of Glasgow has a history of providing research data management (RDM) support, with a fuller service established in 2012. We provide training and advice to help researchers draw up robust, costed RDM plans. In addition, as part of our suite of repositories that encompass publications, theses and impact, we maintain a data registry and repository to store and publish details about data sets created by researchers.

However, limited work has been undertaken to actively preserve other classes of digital records, e.g. corporate records and digital archives deposited with Glasgow University Archive Service. University staff recognized that this issue needed to be addressed to safeguard long-term access to key records, as they underpin business continuity and legal compliance.

2 A group of staff from Information Services, in particular from Research Information Management, were therefore tasked with investigating digital preservation approaches and solutions. The decision was taken to participate in a digital preservation service pilot project which had been set up by Jisc, a membership organization providing digital solutions and advice for UK education and research,<sup>1</sup> as this represented an opportunity to learn more about implementing digital preservation while contributing to wider sector development. Project team members also investigated what good practice for digital preservation comprises, talked to people already carrying out digital preservation activities to learn from their practical experience, and explored current practice within the University for managing records.

## Jisc Open Research Hub

Jisc launched a Research Data Shared Service (RDSS) pilot project in January 2016. The project's intended outcome was to develop a hosted, subscription-based, management service offering a data repository, digital preservation functions and integrated management and reporting, principally for preserving research data, but also open to other types of records. At the time of writing, three years later, Jisc is testing the first version of its new service, Open Research Hub,<sup>2</sup> with the pilot group before a full launch later in 2019. More information about the pilot can be found on Jisc's RDSS project webpage.<sup>3</sup>

Pilot participants were asked to test two digital preservation tools, Archivematica<sup>4</sup> and Preservica,<sup>5</sup> investigate integrations with existing software in their institution, and share feedback with Jisc and other participants to help define service requirements and shape the development of the final product.

As the University of Glasgow already had an established RDM service, we joined the Jisc shared service pilot project in August 2017 to investigate the preservation of digital corporate and archival records. The University's Digital Preservation Policy<sup>6</sup> and draft list of digital preservation requirements<sup>7</sup> provided a framework for our activities. A new member of staff was appointed in January 2018 to help deliver the pilot project.

We tested the preservation tools with a variety of document types and shared our findings with the other pilot institutions via the project discussion board. We discovered that the two preservation tools are quite different in the way they work and the user requirements they seek to meet. Archivematica is primarily a preservation tool. It carries out preservation actions on uploaded material and creates archival-standard packages ready for storage and ongoing management in a separate system of the user's choice. The user can view details of each step in the process and create customized workflows. Preservica provides a full digital management system, offering ongoing storage and management of preserved material, granulated user access and reporting tools in addition to its preservation functions. Working with these two tools prompted us to consider what systems and tools the University already has for managing digital records, how we would transfer records securely between existing document management systems and a preservation tool, and, in particular, the level of automation or staff intervention which would be desirable. Consequently, we have been able to incorporate these issues into our recommendations for a digital preservation programme for our institution.

'the two preservation tools are quite different in the way they work and the user requirements they seek to meet'

## Key activities and learning

In addition to testing and evaluating the preservation tools, project team members undertook a number of other activities to gain an insight into the University's requirements, understand ways to get a digital preservation programme up and running, and develop insight into good practice.

### Understanding the University's requirements

A number of key questions about aspects of the preservation process emerged for the project team as a result of the Jisc project:

- At what point in a digital object's life cycle do preservation actions need to be applied to ensure that it remains accessible?
- What is the best way to manage metadata?
- How do we integrate preservation activities into existing University systems and workflows?
- How do we ensure the terminology is comprehensible to all users?

As we collated information about current policies and practices for managing current, semi-current and archival digital material within the University, we started asking when preservation actions would need to be undertaken on digital material. Paper records are generally straightforward to manage. Once they are no longer actively used, the records can be placed in a records management store until ready for disposal or transfer to archival storage for permanent preservation. Preservation actions are only usually applied once records have been transferred to the University Archive Service, possibly many years after the records were created. However, it is clear that this model is not an effective template for digital records, which may require active intervention long before they become the responsibility of the Archive Service. In particular, we identified record series, e.g. pension records, which will not be kept permanently but which will be retained for many years. Such records will require active care to ensure that they remain accessible throughout their lifetime.

We also learned that early intervention may be necessary to ensure that the appropriate version of records is retained by departments and that they are accompanied by sufficient metadata to interpret them. For example, in a University team which creates digital films, only compressed web versions were being retained and not the high resolution masters. In addition, staff moved quickly from one project to the next and did not have time to document information about each film. Consequently, when Archive staff were invited to select material for permanent preservation, in some instances they discovered that preservation-quality files were missing or only minimal metadata was available. Staff were able to recall details of more recent projects but key facts about material just three or four years old were uncertain, especially if the creator had left. Even the date of creation was difficult to establish in some cases. We are conscious that this situation could be replicated in other departments. Therefore, liaising with staff to ensure that key records are identified and retained, along with adequate information to support future interpretation and access, will be fundamental to the success of our digital preservation programme.

'liaising with staff to ensure that key records are identified and retained ... will be fundamental to the success'



Stack of analogue film canisters  
National Library of Scotland, used under CC BY 4.0 licence.

How metadata should be managed, stored and accessed has generated a great deal of discussion among team members. We wondered whether digital material should be preserved with a complete copy of the metadata which exists for it, or whether just basic details should be included in the archival package and users referred to the relevant

4

catalogue for full details e.g. the archive catalogue (see Figure 1). As metadata may be updated over time, we feel that it is better to simply preserve basic information with the archival package and maintain the separate catalogue entry as the definitive metadata source. However, as part of our next steps, we would like to talk with other institutions to see how they manage their metadata.

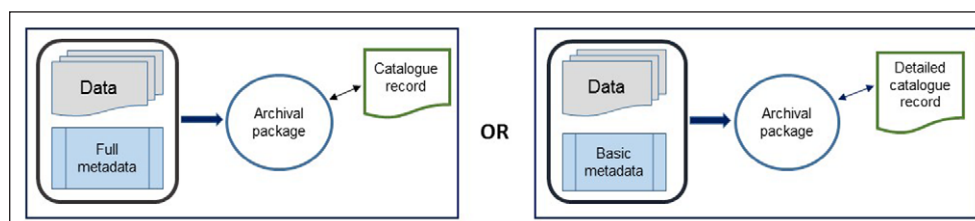


Figure 1. Options for managing and accessing metadata

A particular appeal of the Jisc shared service project was the opportunity to collaborate with others in the sector and learn with them. Regular pilot group meetings and webinars kept us up to date with project developments and more importantly, provided a forum for questions and discussion with other participants. Software user forums were a useful place to ask questions and find out how to make optimal use of the tools. We also carried out some joint testing with colleagues working on the Jisc pilot project at the University of St Andrews. Like us, they are investigating the preservation of corporate records and these, so it was very beneficial to talk through technical issues and organizational questions relating specifically to these record types. We tested several processes using each preservation tool and compared our experiences and outcomes. The staff from St Andrews showed us how to create and upload a tailored processing configuration to specify how these are processed within Archivematica. This was very useful and we have experimented with it subsequently. We agreed that content may need to be encrypted when it is submitted for preservation processing and once it is stored as an archival package, and discussed how this might be achieved. We also explored options for managing user access and thought about who might interact with the preservation tool and in what ways. This collaboration was very productive and has helped us formulate how we might deliver digital preservation and make choices about access, security and workflows.

'Software user forums were a useful place to ask questions and find out how to make optimal use of the tools'

### Collaborative working

Many record custodians within universities and the cultural heritage sector are currently seeking to establish what good practice for digital preservation looks like and are addressing similar technical and organizational challenges. It therefore makes sense for us to work together to understand the issues involved, share experiences and identify simple steps which people can implement. Over the past couple of years, project members have undertaken training and met with practitioners in different fields to discuss implementing digital preservation in practice.

Our project team has benefited greatly from our membership of the Digital Preservation Coalition (DPC),<sup>8</sup> a membership organization facilitating knowledge exchange, technological research and engagement. Team members attended various DPC workshops, including 'Getting started with Digital Preservation' in May 2018, a helpful introduction to the organizational risks and priorities we should be considering in addition to technical factors. Practical exercises tackled risk management, digital asset registers and applying the National Digital Stewardship Alliance (NDSA) levels of digital preservation (see Table 1).<sup>9</sup> These helped us to analyse what our organization is already doing and pinpoint some straightforward actions which we could take to improve record preservation, e.g. ensuring copies of essential records are stored in more than one geographical location.

	<b>Level 1 (Protect your data)</b>	<b>Level 2 (Know your data)</b>	<b>Level 3 (Monitor your data)</b>	<b>Level 4 (Repair your data)</b>
Storage and Geographic Location	<ul style="list-style-type: none"> <li>- Two complete copies that are not collocated</li> <li>- For data on heterogeneous media (optical discs, hard drives, etc.) get the content off the medium and into your storage system</li> </ul>	<ul style="list-style-type: none"> <li>- At least three complete copies</li> <li>- At least one copy in a different geographic location</li> <li>- Document your storage system(s) and storage media and what you need to use them</li> </ul>	<ul style="list-style-type: none"> <li>- At least one copy in a geographic location with a different disaster threat</li> <li>- Obsolescence monitoring process for your storage system(s) and media</li> </ul>	<ul style="list-style-type: none"> <li>- At least three copies in geographic locations with different disaster threats</li> <li>- Have a comprehensive plan in place that will keep files and metadata on currently accessible media or systems</li> </ul>
File Fixity and Data Integrity	<ul style="list-style-type: none"> <li>- Check file fixity on ingest if it has been provided with the content</li> <li>- Create fixity info if it was not provided with the content</li> </ul>	<ul style="list-style-type: none"> <li>- Check fixity on all ingests</li> <li>- Use write-blockers when working with original media</li> <li>- Virus-check high risk content</li> </ul>	<ul style="list-style-type: none"> <li>- Check fixity of content at fixed intervals</li> <li>- Maintain logs of fixity info; supply audit on demand</li> <li>- Ability to detect corrupt data</li> <li>- Virus-check all content</li> </ul>	<ul style="list-style-type: none"> <li>- Check fixity of all content in response to specific events or activities</li> <li>- Ability to replace/repair corrupted data</li> <li>- Ensure no one person has write access to all copies</li> </ul>
Information Security	<ul style="list-style-type: none"> <li>- Identify those people authorized to read, write, move and delete individual files</li> <li>- Restrict who has those authorizations to individual files</li> </ul>	<ul style="list-style-type: none"> <li>- Document access restrictions for content</li> </ul>	<ul style="list-style-type: none"> <li>- Maintain logs of who performed what actions on files, including deletions and preservation actions</li> </ul>	<ul style="list-style-type: none"> <li>- Perform audit of logs</li> </ul>
Metadata	<ul style="list-style-type: none"> <li>- Inventory of content and its storage location</li> <li>- Ensure backup and non-collocation of inventory</li> </ul>	<ul style="list-style-type: none"> <li>- Store administrative metadata</li> <li>- Store transformative metadata and log events</li> </ul>	<ul style="list-style-type: none"> <li>- Store standard technical and descriptive metadata</li> </ul>	<ul style="list-style-type: none"> <li>- Store standard preservation metadata</li> </ul>
File Formats	<ul style="list-style-type: none"> <li>- When you can advise on the creation of digital files, encourage use of a limited set of known open formats and codecs</li> </ul>	<ul style="list-style-type: none"> <li>- Inventory of file formats in use</li> </ul>	<ul style="list-style-type: none"> <li>- Monitor file format obsolescence issues</li> </ul>	<ul style="list-style-type: none"> <li>- Perform format migrations, emulation and similar activities as needed</li> </ul>

Table 1. National Digital Stewardship Alliance levels of digital preservation, Version 1 NDSA, 201, used under CC BY 4.0 licence

Meanwhile, the DPC’s workshop on migrating data between systems (July 2018) offered a valuable opportunity to learn from information professionals who have been managing and preserving digital records for some time. As before, contributors emphasized that the key factors are organizational and policy related, rather than technical. All agreed that cleaning and enhancing metadata takes up a considerable amount of time, so this needs to be built into project timescales. Speakers recommended planning and testing exit strategies throughout the life of your preservation system, not just when you decide to remove your data from it. Our group enjoyed an animated discussion about ways to verify that *all* data has been migrated correctly from one system to another and is intact: it was most useful to hear how attendees

‘key factors are organizational and policy related, rather than technical’

6 had addressed this. Learning about good practice now, before we set up any digital preservation tools and workflows, should help us to put together a robust preservation system with exit strategies built in.

We also collaborated with DPC colleagues to organize our own digital preservation events, in March 2017 and, on International Digital Preservation Day (30 November 2017), 'Aye Preserve: Digital Preservation in the West of Scotland'.<sup>10</sup> This brought together digital practitioners from different backgrounds to network, share best practice, discuss challenges and update themselves on recent developments and training. It was a great opportunity for us to benefit particularly from the experiences of practitioners in the business sector, where digital preservation is already routine.

Colleagues implementing digital preservation in other universities have also been generous in sharing their learning with us. For example, the University of Westminster's University Records and Archives team has been actively preserving digital archival material since 2016, using a cloud-based, managed digital preservation service. Staff showed us the automated preservation workflows which they have set up for specific types of material and it was very helpful to see preservation processes running in a live environment. The University of Westminster Records and Archives team's webinar<sup>11</sup> addressed questions we have about managing relationships between preserved data and metadata in catalogue systems, cataloguing digital material, and ensuring sensitive records are dealt with appropriately. It also raised the interesting question of what is 'good enough'. If two files out of one hundred in an archival package have not migrated successfully from their original file format to a preservation format, do you investigate and then run the whole preservation process again? Or, given time pressures and quantity of material to process, do you have to accept that this is 'good enough'? When we start to process records at scale, then we may need to accept that not all errors can be fixed. These are useful questions for us to contemplate now as we consider what level of preservation service we will be able to deliver. The Westminster University Records and Archives team also suggested that processing digitized records first of all, before tackling born-digital material, acted as a helpful first step to test processes and gain confidence, so this is something we will bear in mind.

'we may need to accept that not all errors can be fixed'

The Research Data team at The University of Strathclyde also took time to show us how they preserve their research data. We were able to watch how they transfer the data from their data management system into the preservation tool and create a workflow for preserving digital research data. Again, it was useful to discuss the choices they have made about workflows and tools.

Overall, learning from colleagues' practical experience and seeing both simple first steps and established preservation programmes in action has been very constructive. They have helped us to put together recommendations for establishing a digital preservation programme, given us a solid understanding of first principles and reassured us that we will not 'get it wrong'.

### Tackling terminology barriers

Technical terms within any field can act as a barrier to effective communication and a shared understanding. The field of digital preservation is full of sector-specific terminology and the project team has had to familiarize themselves with many new terms in order to understand the processes and literature. However, digital preservation terminology can be confusing or unclear, and the same term can mean different things to different people. Some terms, e.g. 'digital curation', are so problematic that people avoid using them.<sup>12</sup> There are a few glossaries online, such as the one produced by the DPC,<sup>13</sup> but their contents differ and there is no fully comprehensive resource which people can refer to, so clarifying exactly what a term means is not always straightforward.

'Some terms, e.g. 'digital curation', are so problematic that people avoid using them'

The University project team therefore decided to run a digital preservation terminology workshop at the CASRAI (Consortia Advancing Standards in Research Administration Information) Reconnect UK 2018 conference, based on the DPC glossary. CASRAI is

7 an international, not-for-profit organization seeking to develop efficient information requirements for research organizations. Participants were asked to highlight terms that they did not understand and list words which they had expected to find but which were absent. Terms considered ambiguous or unclear included 'authentication' and 'metadata', and attendees discussed possible definitions. Meanwhile, 'graphical user interface', 'bit rot' and 'user experience' were among the terms identified as not currently in the glossary.

The workshop participants considered that it is easier to help develop terms when they are presented in a more layered arrangement. They also agreed that, rather than arranging glossary terms alphabetically, terms grouped together in families e.g. acronyms for organizations, preservation tools/processes, or standards, are more useful, can be defined more effectively and be contextualized. This is the way the CASRAI dictionary works: it defines the root term, then presents context-specific uses (root term with qualifying adjective) as separate entries, rather than trying to make the root definition cover all uses (see Figure 2, for example). A report was compiled summarizing the workshop findings and was shared on the CASRAI forum for further discussion.<sup>14</sup>

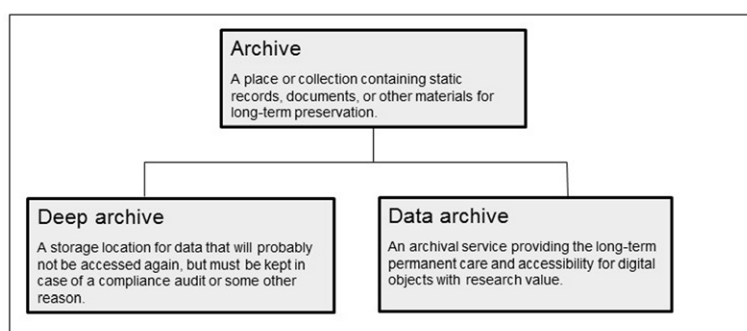


Figure 2. Example of definitions of types of archives in the CASRAI dictionary

Following on from this workshop, CASRAI and the DPC are investigating whether they can work together to enhance and maintain a comprehensive collection of digital preservation terms and we hope to be able to support these efforts.

## Conclusion

Addressing digital preservation can seem a daunting task, with its subject-specific vocabulary and complex technical requirements around metadata, file formats and integration of tools and systems. Managing files is not necessarily a prime concern for staff or stakeholders within the organization, and competing priorities make it difficult to promote digital preservation, secure adequate resources and implement change. However, this period of experimentation and learning for our project team has demonstrated how the collaborative approach of the preservation community is helping to define good practice, identify robust, workable solutions for any scale of organization and develop resources for everyone's benefit.

## Next steps

The University of Glasgow Digital Preservation Working Group has submitted a paper to the Senior Management Team to establish what the University's priorities are for digital preservation in the short and medium term. Meanwhile, Jisc has just launched its new preservation and data repository service, Open Research Hub, developed out of the pilot project. Once we know how the University wants to move forward with digital preservation, we will be able to decide whether Jisc's new service will best fulfil the University's needs or if another approach is required.

While we are waiting for the Senior Management Team's decision, we will use some of the assessment tools we have learned about to evaluate the University's current processes for managing digital material. The results should help us to identify straightforward steps to improve what we do.

8 We also plan to start actively preserving some digital records. Glasgow University Archive Service staff have identified a collection of born-digital films which would benefit from preservation, so we have started work to identify what preservation actions are required, format the metadata and set up a preservation workflow. As well as meeting a real need, this will provide useful experience in developing preservation processes.

Finally, we will continue to organize and co-host 'Aye Preserve' events in the West of Scotland with our colleagues from the DPC and will work with them and CASRAI to clarify preservation terminology. Our digital preservation blog, Digital Preservation @ University of Glasgow,<sup>15</sup> will remain a place to share our journey with the wider community.

'the Jisc pilot has acted as a catalyst to take more decisive action'

Participation in the Jisc pilot has acted as a catalyst to take more decisive action and establish a clearer framework for digital preservation within our institution. Talking to people who are already undertaking digital preservation activities and learning from their expertise and practical experience has given us confidence to make a start.

#### Acknowledgements

We would like to thank those people who have shared their knowledge and experience and assisted us with digital preservation, particularly colleagues at the Digital Preservation Coalition, University of St Andrews, the University of Glasgow, the University of Strathclyde, Westminster University Records and Archives and participants at events.

This work was supported by funding from Jisc award number 4811.

#### Abbreviations and Acronyms

A list of the abbreviations and acronyms used in this and other *Insights* articles can be accessed here – click on the URL below and then select the 'full list of industry A&As' link: <http://www.uksg.org/publications#aa>

#### Competing interests

The authors have declared no competing interests.

#### References

1. Jisc:  
<https://www.jisc.ac.uk/> (accessed February 7, 2019).
2. Jisc open research hub:  
<https://www.jisc.ac.uk/open-research-hub> (accessed February 7, 2019).
3. Jisc RDSS project:  
<https://www.jisc.ac.uk/rd/projects/research-data-shared-service> (accessed February 7, 2019).
4. Archivematica is an open source product developed by Artefactual, last modified January 2019:  
<https://www.archivematica.org/en/> (accessed February 7, 2019).
5. Preservica, last modified January 2019:  
<https://preservica.com/> (accessed February 7, 2019).
6. "Digital Preservation policy," University of Glasgow:  
<https://www.gla.ac.uk/myglasgow/staff/governance/> (accessed February 7, 2019).
7. "Digital Preservation requirements," University of Glasgow:  
<http://eprints.gla.ac.uk/174422/> (accessed February 7, 2019).
8. Digital Preservation Coalition, last modified January 2019:  
<https://www.dpconline.org/> (accessed February 7, 2019).
9. National Digital Stewardship Alliance, *Levels of Digital Preservation*, (Washington, DC: *Proceedings of the Archiving [IS&T] Conference*, April 2013):  
<https://ndsa.org/activities/levels-of-digital-preservation/> (accessed February 7, 2019). For more information about the work of the National Digital Stewardship Alliance:  
<https://ndsa.org/> (accessed February 7, 2019).
10. Valerie McCutcheon, "Aye Preserve November 2017," *Digital Preservation @ University of Glasgow* (blog), December 6, 2017:  
<https://universityofglasgowdigitalpreservation.wordpress.com/2017/12/> (accessed February 7, 2019).
11. Rebecca Short and Anna McNally, "Work in Progress: reflections on our first year of digital preservation" (webinar), Westminster University Records and Archives, December 2017:  
<https://www.youtube.com/watch?v=k-SVO6IQE2M> (accessed February 7, 2019).



12. Feedback from workshop at CASRAI Reconnect UK conference 2018.
13. "Glossary," Digital Preservation Coalition:  
<https://www.dpconline.org/handbook/glossary>. Examples of other glossaries are produced by the Digital Curation Centre:  
<http://www.dcc.ac.uk/digital-curation/glossary>, Bristol University Research Data Management Glossary:  
<http://vocab.bris.ac.uk/data/glossary/glossary-1.0.html> and in the CASRAI dictionary:  
[https://dictionary.casrai.org/Main\\_Page](https://dictionary.casrai.org/Main_Page) (all accessed February 7, 2019).
14. Alison Spence, "Workshop notes: vocabularies and definitions in Digital Preservation," (June 2018):  
<http://eprints.gla.ac.uk/175828/> (accessed February 7, 2019). For more information about CASRAI, see their website:  
<https://casrai.org/> (accessed February 7, 2019).
15. *Digital Preservation @ University of Glasgow* (blog):  
<https://universityofglasgowdigitalpreservation.wordpress.com/> (accessed February 7, 2019).

**Article copyright: © 2019 Alison Spence, Valerie McCutcheon and Matt Mahon. This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use and distribution provided the original author and source are credited.**



Corresponding author

Alison Spence

Research Information Administrator: digital preservation

University of Glasgow Library

University of Glasgow

Glasgow G12 8QE, UK

E-mail: [alison.spence@glasgow.ac.uk](mailto:alison.spence@glasgow.ac.uk)

ORCID ID: <http://orcid.org/0000-0002-5483-2342>

Co-authors

Valerie McCutcheon

ORCID ID: <http://orcid.org/0000-0002-0705-4832>

Matt Mahon

ORCID ID: <http://orcid.org/0000-0001-8950-8422>

To cite this article:

Spence A, McCutcheon V and Mahon M, Case study: the University of Glasgow's digital preservation journey 2017-2019, *Insights*, 2019, 32: 10, 1–9; DOI: <https://doi.org/10.1629/uksg.461>

Submitted on 28 January 2019

Accepted on 06 February 2019

Published on 20 March 2019

Published by UKSG in association with Ubiquity Press.