Designing and implementing a digital preservation training needs assessment

Findings from the Bodleian Libraries' institutional repository

Sarah Mason
Bodleian Libraries
University of Oxford
Oxford OX2 0EW
sarah.mason@bodleian.ox.ac.uk

Edith Halvarsson
Bodleian Libraries
University of Oxford
Oxford OX2 0EW
edith.halvarsson@bodleian.ox.ac.uk

ABSTRACT

This paper describes the design and pilot of a digital preservation training needs assessment at Bodleian Libraries (University of Oxford). The assessment was designed to establish training needs among staff, with the purpose of creating targeted digital preservation training based on gaps identified. Measurements for the assessment were developed around the DigCurV [7] Framework skill descriptors combined with a literature review of other skills frameworks. The measurements were developed into a set of interview questions, which was piloted on nine members of staff from the Oxford University Research Archive (ORA) in the winter of 2016. Findings from the assessment have influenced the design of training modules which frame digital preservation concepts around workflows in the ORA digital repository. The modules take as their starting point the awareness level, knowledge, practical skills, problem solving approaches and preferred learning style of staff within the ORA team.

KEYWORDS

Auditing, skills, workforce development, training, DigCurV, institutional repository, digital preservation

1 INTRODUCTION

This paper describes the design and trial of a digital preservation staff training needs assessment undertaken at Bodleian Libraries (University of Oxford). In the winter of 2016, the Digital Preservation at Oxford and Cambridge project [1] trialled a set of interview questions on staff members from the Oxford University Research Archive (ORA) [2]. The purpose of the assessment was to record what staff knew about digital preservation, and which transferable skills could be drawn upon to deliver tailored staff training in this area.

Digital Preservation at Oxford and Cambridge (DPOC) is a two-year project funded by the Polonsky Foundation to research digital preservation policy, staffing and technical challenges present at Bodleian Libraries and Cambridge University Library. The project's outcomes and recommendations will form the basis for long-term, sustainable digital preservation programmes at each institution. Each institution has appointed 3 Polonsky Fellows to undertake this work in the following roles: Policy & Planning, Outreach & Training and Technical Specialist.

ORA, the Oxford University Research Archive, is the University of Oxford's institutional repository and maintains the scholarly output of its members. It preserves research publications, journal articles, conference papers, working papers, theses, reports, book sections, unpublished academic work, and more. ORA-data was launched in 2015 as an archival store to help researchers archive, share and cite research data.

Findings from the assessment revealed that awareness of digital preservation risks was good among ORA staff. This knowledge was often acquired through the practical experience of working with digital material in ORA. Interestingly, despite being able to describe core concepts using practical examples, "jargon" digital preservation terms were only recognised by members of staff who had undertaken formal academic training on digital preservation. Having digital preservation terms as part of a staff member's vernacular helps with accessing digital preservation literature and concepts. However, the assessment illustrates that "jargon" is perceived as intimidating and can also create a barrier for entering into discussions in the digital preservation field.

Furthermore, the assessment found that staff felt uncomfortable speaking about digital preservation with depositors due to a lack of up-to-date knowledge of the technical end-to-end workflows used within ORA's underlying submission and repository service. This lack of confidence is a key skill to address; one of the strongest motivators staff expressed for learning about digital preservation was the ability to provide good service and advice to depositors and users. As preservation plans for research outputs are required by many UK funding bodies, staff in ORA are increasingly receiving queries from academics about digital preservation. Based on these findings, DPOC concluded that digital preservation training was better delivered in-house, tailored to Bodleian Libraries' local context, rather than outsourced to external providers. DPOC are currently developing customised training modules around ORA workflows and repository software to contextualise digital preservation within the service. The modules take as their starting point the awareness level, knowledge, practical skills, problem solving approaches and preferred learning style of staff within the ORA team.

2 ASSESSMENT DESIGN

2.1 Reviewing Skills Frameworks

Before designing a set of interview questions for the assessment, the DPOC project needed to identify key skills for staff working

---

in digital preservation. A number of skills frameworks were consulted including the ARA Competency Framework [2], the CILIP Professional Knowledge and Skills Base [6] and the DigCCurr Matrix from the University of North Carolina [11]. However, as there were no available interview question templates to work with, the DPOC team needed to design their own. The DigCurV Framework [7] formed the DPOC project's list of skills for creating interview questions. DigCurV was chosen as it is the most comprehensive framework among those assessed, and provides good granularity as it is tailored around different skill sets for different types of roles (Executive, Manager and Practitioner lenses). Also each skill in the DigCurV Framework has been assigned a unique number, making coding of the literature review and interview transcripts clear and concise.

2.2 Refining DigCurV Skill Descriptors

There are 110 skill descriptors in DigCurV [7], far too many to condense into a concise set of interview questions or online questionnaire. The volume of skill descriptors in DigCurV posed an issue during the assessment design, as addressing each one would take several hours and become too disruptive for staff. DPOC therefore reviewed DigCurV [7] and the earlier frameworks, which produced a reduced list of 71 skills for managers and practitioners. Skills were removed if they applied solely to the Executive lens, which was deemed out of scope for the training needs assessment. Skill descriptors were combined where the DPOC project assessed that there was overlap between them or if a skill was too unclear and lacked clarification. Since the DigCurV [7] Framework lacks clear definitions for all 110 descriptors, the DPOC team created a glossary for the remaining skill descriptors.

A literature review [1, 3–5, 10, 11, 13, 16, 19] produced a final list of 63 of the most commonly associated digital preservation skills. A further review of Bodleian Libraries' training programmes, policies and job descriptions, narrowed the list to a further 43 skills for use in the training needs assessment. This list, combined with the literature review, yielded a shortlist of the top 20 skills for digital preservation that focused mainly on:

- technical skills (KIA1.15, KIA5.1, KIA5.2, KIA5.4. MQA3.12),
- metadata (KIA4.5, KIA4.6),
- communications skills (PQ2.1),
- domain and digital preservation knowledge (KIA1.1, KIA2.5),
- preservation planning (KIA1.16, KIA3.4),
- the designated community, access and searching needs (KIA3.6, KIA4.1, KIA4.2) and
- legal requirements (PC1.1, PC1.2, PC2.1, PC2.4) [7].

Using this list of descriptors, the DPOC project created a set of interview questions which can be delivered in a manageable 1–1.5 hour session.

2.3 Methodology

The DPOC team considered a number of qualitative and quantitative methods for the assessment, and settled on running semi-structured interviews. The use of semi-structured interviews allowed participants to explore alternative areas of interest, aspects which cannot be adequately captured in a structured interview or online questionnaire [18, p. 57] [8, p. 29]. Interviews were run with one member

of the ORA team at a time, as this provided participants with confidentiality not available in a focus group setting. Another strength of one-on-one interviews was that the interviewers received input equally from each participant [18, p.171]. The flexible nature of semi-structured interviews also allowed the interviewers to clarify problematic phrasing and vague terminology identified by participants during the interview. The feedback resulted in a second version of the assessment with revised questions; samples of the question revisions are in Table 1. It was trialled in a second round of interviews in April and May 2017; results from these interviews are currently being analysed.

Two different sets of interview questions were designed and used in the pilot interviews, depending on whether or not a participant was considered to be a manager or practitioner. This differentiation was based on role responsibilities, not the participant's job title [12]. Later, a manager/practitioner combined set of interview questions was added; together, these formed the training needs assessment that would be trialled on the ORA team.

Table 1: Interview question revisions

Original question	Revised question
What is your understanding of a digital file format?	What digital file formats do you work with in your role (for both collection materials and administrative tasks)?
Using your understanding of a digital file format, what formats do you work with in your role for your collection materials and administrative tasks?	Could you explain how you would handle a situation where you received a digital file without a file extension?
From your professional awareness or experience, can you describe to me an example of a complicated information rights issues for a digital deposit?	Think about a time you or one of your colleagues dealt with a complicated rights issues for a digital deposit, could you explain to me how it was managed?
Thinking about your digital collections, are you aware of them having any formal accreditation, certification or being subject to compliance audits?	Thinking about the repository you work with, are you aware if it has any of the following: certifications, or accreditations?

3 PILOT SAMPLE

The DPOC project interviewed 9 members of the ORA team using the first version of the interview questions. The interview sample included:

- 2 managers,
- 5 metadata assistants,
- 1 senior metadata assistant, and
- 1 curator of research data.

The metadata assistants in ORA have a broad range of professional and academic backgrounds. Their time working with the service ranged from 5 years to a few months—although some have previously worked for Bodleian Libraries in other roles. Notably, one of the metadata assistants had completed a Masters level module on digital curation, and was able to respond in more detail around digital preservation concepts and terminology.

The interviews were completed over a two month period with each interview averaging between 1–1.5 hours. The same two members of the DPOC team took part in each interview to ensure that questions were delivered consistently across the sample [17, p.1193]. This method allowed one person to focus on accurate note taking, negating the need for tape recording, which the Oxford DPOC team felt would be too cumbersome to transcribe and had the potential to alter interviewees' responses [15, p.169].

4 FINDINGS FROM THE PILOT INTERVIEWS

4.1 Core DigCurV Skills

The assessments identified a number of core skill strengths as identified by the selected DigCurV [7] skill descriptors. Not surprisingly, due to the type of material ORA staff work with, all staff had a good understanding of rights management, as well as strong communication and metadata editing skills. Staff also displayed good problem solving skills when required to learn new technology. Two members of staff had experience of using UNIX command lines and scripts.

4.2 Digital Preservation Risks and Concepts

Staff were able to give a number of examples of digital preservation risks, drawing on experience of their role reviewing digital file submissions for ORA. The most common risk areas described were hardware obsolescence (55% of respondents), file format 'obsolescence' and software obsolescence (55% of respondents). One staff member also mentioned storage failure and low quality metadata as risks to digital material. Although staff were able to describe these risks, only two recognised the digital preservation terms and concepts listed in one of the questions (the list contained references to the OAIS model and concepts, normalization and fixity). In several cases, the interviewers were able to frame what these concepts referred to by pointing to examples already supplied by staff earlier on in the interview. This led the DPOC team to start considering the possibility of using these work experiences and the ORA workflow as an entry to digital preservation concepts.

4.3 Digital Preservation and the ORA Service

While many preservation risks were understood by staff, there was much uncertainty about how risks can be mitigated. A reoccurring theme throughout the assessment was that staff did not know, or had out-dated knowledge about technical workflows and preservation practices associated with ORA, including the role and implementation of Bodleian Libraries' Fedora instances and Hydra. Opinions were split among the 9 staff interviewed regarding whether or not ORA carries out digital preservation activities: 4 members of staff were unsure about the answer, 2 thought that ORA is doing basic bit-level preservation and 3 thought the service currently does not do any preservation activities (see Figure 1).

Q: Do you think we already carry out activities to preserve the digital collections you work with? (n=9)
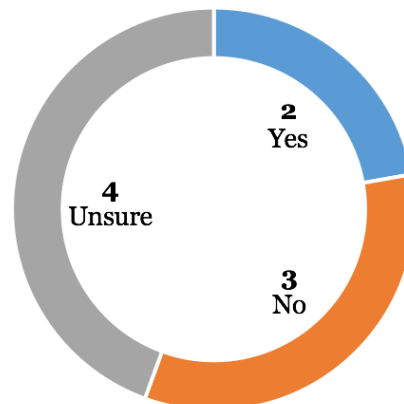
Figure 1: Excerpt from assessment: digital preservation activities (all interviewees, managers & practitioners)

Not knowing what current preservation actions the ORA service uses to mitigate risks impacted on how comfortable staff were about explaining these concepts to depositors and users. 6 out of the 7 practitioners (one was a manager/practitioner) were not comfortable discussing digital preservation with depositors—with one commenting that "it would be nice to be able to answer those questions" and another commenting that "I do not know enough about the service and its future [to do so]". One commented that "I use the words 'archiving' and 'storage' with depositors, because I'm not sure we do [digital preservation]."

4.4 Manager Buy-in

Both managers showed a clear understanding of digital preservation, associated risks and techniques to mitigate those risks. This finding was encouraging, as it demonstrated managerial commitment to digital preservation. One manager expressed awareness that confidence with digital preservation was lacking in the ORA team, and believed that training would ensure staff could "answer queries and understand the wider environments which they work in." This support for digital preservation training included the manager's willingness to release staff to attend any available training sessions and provide additional feedback on session content.

4.5 Practitioner Problem Solving

The practitioner set of questions delved into problem solving approaches. When metadata assistants were asked how they would handle a scenario of reviewing a digital file with an unknown or missing file extension, their responses showed that most were comfortable experimenting with digital files and doing online self-directed research. The questions around problem solving approaches also revealed that 5 out of 6 metadata assistants relied heavily on

the knowledge base built up within their local team to address scenarios that required them to learn a new task or program. Several staff explained that if they received problematic or unidentified digital files, they would initially ask their immediate team for advice to see if anyone had previous experience of that file format.

4.6 Practitioners' Preferred Learning Styles

The importance of the local team was also evident when staff were asked about their learning style preferences. The metadata assistants expressed that they preferred small groups for interactive training modules, but that larger groups for lecture-style training was good, provided they were not required to interact with other participants. This often reflected confidence in interacting with a new subject. One participant commented that "I would feel more comfortable having training with my closest team [the metadata assistants]" and another that "I don't like working in a group with people I do not know." However, in terms of general introductory modules, the same participant commented that "I think it is best to open it to all [staff], since then you get a much more varied perspective on things". This suggested that for general awareness topics a large group approach is acceptable, but that ORA-specific or other in-depth workshops should be kept small and be limited to the core practitioners.

5 MODULE DEVELOPMENT

Based on the findings and skills profile of ORA staff, the DPOC project are currently developing training material for a tailored digital preservation module around ORA. Developing the training module for ORA has proven to be an interesting exercise in itself, as the DPOC team members are new to their organisations. Training is therefore being developed alongside a technical repository review of ORA as part of the DPOC project [14]. This has required extensive collaboration with technical support staff around Bodleian Libraries, in order to inform the DPOC team of ORA's workflows and underlying infrastructure. For organisations looking to frame their digital preservation training in a similar fashion, it is evident that getting buy-in and advice from a variety of managerial and service support staff will be necessary to develop meaningful, targeted training modules.

5.1 Tailored Training Module: ORA

The module currently in development covers digital preservation concerns in end-to-end ORA workflows, from submission of research outputs by academics, to long-term storage and management by the library. It also includes explanations of how bit-level preservation is addressed in ORA's current storage infrastructure. DPOC hope that framing digital preservation in a familiar context will help contextualize and introduce digital preservation terminology, and directly impact staff's understanding of their own role within the service.

5.2 Hands-on Workshops

As well as having modules around the ORA workflows, the DPOC team are preparing hands-on workshops for analysing digital files using tools like DROID, JHOVE, VeraPDF and the BitCurator suite. As some staff have command line experience, a mixture of GUI and command line tasks will be run. Although these tools will not be used by reviewers on a day-to-day basis, knowledge of how they work and why they are useful tools will help staff when speaking about digital preservation to depositors and users.

5.3 All-staff Digital Preservation Awareness Module

During the assessment pilot phase, there was an interest in developing digital preservation awareness training that could be delivered to all library staff, not just those currently working with digital collections. Due to this demand, an adapted and brief online questionnaire was later sent out to all members of staff at Bodleian Libraries. The questionnaire enabled the DPOC team to triangulate some of the themes which were raised from qualitative interviews with ORA staff [9]. They used the findings to develop an all-staff digital preservation awareness training module.

This module includes a unit on personal digital archiving, which has been run multiple times. Several other units are currently being developed on digital preservation terms and concepts—these will also cover some basic language from the OAIS reference model. Although the assessment has shown that OAIS terminology is not necessary for identifying digital preservation risks, the DPOC project would like to explore if familiarity with OAIS terminology makes participation in discussions more accessible. These awareness modules will mix lecture-style, interactive and online delivery methods.

5.4 Training Evaluation

The success of the ORA and general digital preservation awareness training will be measured by pre- and post-module evaluation. Where two versions of a similar module or hands-on workshop will be trialled, one metadata assistant will attend and evaluate the two versions, facilitating revisions and improvements to the training materials. This method will allow the DPOC team to ensure an effective training programme is in place at the conclusion of the project.

6 NEXT STEPS

Alongside designing the ORA digital preservation training itself, the DPOC project will also be developing and trialling the next phase of the training needs assessment. The pilot trial of the sets of interview questions on ORA highlighted that there is also a need to design one for software developers working with digital collections. The current practitioner interview questions were found to be unsuitable because its focus does not align with skills required for a software developer. Further research is required to define the ideal skills set for software developers and other technical staff working with collection material. Once done, a specific set of developer interview questions will be developed. When completed, it will be trialled on developers working with ORA services.

The interview questions, online questionnaire and glossary of skill descriptors have been developed into a draft training needs assessment toolkit. This will enable other institutions to audit the skills and training needs of their staff. The authors invite comment and feedback on the draft toolkit, available on the DPOC project website: www.dpoc.ac.uk.

7 CONCLUSION

Findings from the assessment pilot have been invaluable for developing training modules around digital preservation at Bodleian Libraries. While there are a number of outsourced training solutions available, the assessment revealed that training modules benefit from being tailored to Bodleian Libraries' local context. Staff in ORA want and need to be informed about their digital services; they also want to learn about digital preservation in relation to their digital services. While developing training is more labour-intensive, tailored and targeted in-house training will provide staff with the knowledge to speak confidently to users and depositors about their digital service. As the amended set of interview questions is rolled out to other members of staff in Bodleian Libraries, the DPOC project will be able to see whether or not this trend continues across other teams within the institution.

CONTRIBUTORS

Sarah Barkla, Bodleian Libraries
Alan Mckechnie, Bodleian Libraries
Jason Partridge, Bodleian Libraries
Thomas Stanbury, Bodleian Libraries
David Tomkins, Bodleian Libraries
Hilary Tricker, Bodleian Libraries
David Watson, Bodleian Libraries

REFERENCES

[1] Suzie Allard, Christopher Lee, Nancy Y Mcgovern, and Alice Bishop. 2016. *The Open Data Imperative*. Technical Report. Council on Library and Information Resources, Washington, DC. http://www.clir.org

[2] Archives and Records Association. 2015. *Competency Framework*. Technical Report. London.

[3] Winston Atkins, Andrea Goethals, Carol Kussmann, Meg Phillips, and Mary Vardigan. 2013. *Staffing for Effective Digital Preservation: an NDSA Report*. Technical Report. NDSA Standards and Practices Working Group, Washington, DC. 26 pages. http://hdl.loc.gov/loc.gdc/lcpub.2013655113.1

[4] K. R. Blumenthal, P. Griesinger, J. Kim, S. Peltzman, and V. Steeves. 2016. What Makes a Digital Steward: Survey Instrument. (2016). https://osf.io/zndwq/https://osf.io/zndwq/

[5] Youngok Choi and Edie Rasmussen. 2009. What qualifications and skills are important for digital librarian positions in academic library? A job advertisement analysis. *The Journal of Academic Librarianship* 35, 5 (2009), 457 – 467.

[6] CILIP. 2014. *Your Professional Knowledge and Skills Base*. Technical Report. 1–20 pages. www.cilip.org.uk/pksb

[7] DigCurV. 2013. DigCurV - Skills in Digital Curation Curriculum Framework. (2013). http://www.digcurv.gla.ac.uk/index.html

[8] R Edwards and J Holland. 2013. *What is qualitative interviewing?* Bloomsbury, London.

[9] T Jick. 1979. Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly* 24, 4 (1979), 602 – 610. DOI:http://dx.doi.org/10.2307/2392366

[10] Jeonghyun Kim, Edward Warga, William E Moen, Warga Kim, and Moen 67. 2013. Competencies Required for Digital Curation: An Analysis of Job Advertisements.

The International Journal of Digital Curation International Journal of Digital Curation 8, 81 (2013), 66–83. DOI:http://dx.doi.org/10.2218/ijdc.v8i1.242

[11] Christopher Lee. 2009. *Matrix of Digital Curation Knowledge and Competencies*. Technical Report. School of Information and Library Science, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina. https://ils.unc.edu/digccurr/digccurr-matrix.html

[12] Library of Congress. 2010. Digital Preservation Outreach Education. (2010). http://www.digitalpreservation.gov/education/educationneeds.html

[13] National Research Council. 2015. *Preparing the Workforce for Digital Curation*. Technical Report. Washington, D.C. DOI:http://dx.doi.org/10.17226/18590

[14] OCLC and CRL. 2007. *Trustworthy Repositories Audit and Certification: Criteria and Checklist*. Technical Report. https://www.crl.edu/sites/default/files/d6/attachments/pages/trac

[15] Zina O'Leary. 2004. *The Essential Guide to Doing Research*. Sage Publications, London.

[16] Carole L. Palmer, Cheryl A. Thompson, Karen S. Baker, and Megan Senseney. 2014. Meeting Data Workforce Needs: Indicators Based on Recent Data Curation Placements. In *iConference 2014 Proceedings*. iSchools. DOI:http://dx.doi.org/10.9776/14133

[17] Michael Quinn Patton. 1999. Enhancing the quality and credibility of qualitative analysis. *Heath Services Research* 34, 5 (1999), 1189 – 1208. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1089059/pdf/hsresearch00022-0112.pdf

[18] Keith F Punch. 2005. *Introduction to Social Research* (2nd ed.). Sage Publications, London.

[19] Bronwen Sprout and Sarah Romkey. 2012. A Persistent Digital Collections Strategy for UBC Library. In *Memory of the World in the Digital Age: Digitization and Preservation Conference*. Vancouver, British Columbia. http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/VC