Time to Change

Effects and Implications of Digital Preservation in an Organizational Context by Michelle Lindlar[1]

TIB

Abstract

Digital preservation is not a pure technical task – the first paragraph of the OAIS, the probably most cited work in the context of digital preservation, clearly says so: An OAIS is an "Archive, consisting of an organization, which may be part of a larger organization, of people and systems that has accepted the responsibility to preserve information and make it available for a Designated Community" (CCSDS, 2012, p. 1-13). But what prerequisites does an institution have to fulfill for "digital preservation readiness"? And what effect does digital preservation have on the organization itself? How does the task digital preservation fare in juxtaposition with the necessary resources "organization", "people" and "systems"?

Digital preservation within an organizational context requires different layers of change: During the implementation process the necessary resources have to be allocated and aligned. Since digital preservation is a cross-sectional task, drawing from the expertise of different stakeholders, this may require organizational change in form of new team structures. Furthermore, digital preservation is a rather fluid task, which requires constant monitoring and adaptability. Both change layers – the initial one as well as the ongoing one – have direct effects and implications on an organization.

Keywords: digital preservation, organization, change processes

Introduction

Digital preservation is often mistaken for a problem of technology. While the rapid technological change we have been facing since the rise of the personal computer and the growth of data that inevitably came with it is certainly the main risk of the preservation process, digital preservation is by no means solely a problem of technology. A frequently used definition of digital preservation is that it is not a promise of safeguarding information for five, ten or fifty years – but rather a promise to develop strategies which meet the constant change imposed on digital objects and their surroundings through the fast-paced development of new technologies (Schwens, 2004).

Memory institutions do have long-standing experience in the conservation of analogue information artifacts. But how are they faring in the preservation of digital objects? The question of organizational requirements and organizational change in the context of digital preservation is certainly not a new one. In 1996 one of the first high level research activities – the CPA/RLG (Commission on Preservation and Access / Research Library Group) "Task Force on the Archiving of Digital Information" – described the connection between risk mitigation and organizational changes: "Compounding the technical challenges of migrating digital information is the problem of managing the process in a legal and organizational environment that is in flux as it moves to accommodate rapidly changing digital technologies" (RLG/CPA, 1996, p. 6).

Digital preservation within an organization may be seen as a pyramid (see figure 1), where the organization type (i.e., national library, archive, research institution, etc.) and the mandate form the basis for all activities. Preservation activities are realized using a combination of three resources: people, hardware, and software. The resources are tied together in processes and workflows. And lastly, policies form a guideline framework for all activities.

But what are the prerequisites in the single building blocks when it comes to digital preservation? What changes will they face? This paper will briefly highlight needed prerequisites for organizational digital preservation readiness as well as needed ongoing change processes for three areas: preservation strategy, staff, and system choice. While the prerequisites will pick up on common misperceptions and highlight initial change processes, the ongoing changes will show what is needed in the ongoing maintenance of sustainable digital preservation processes.
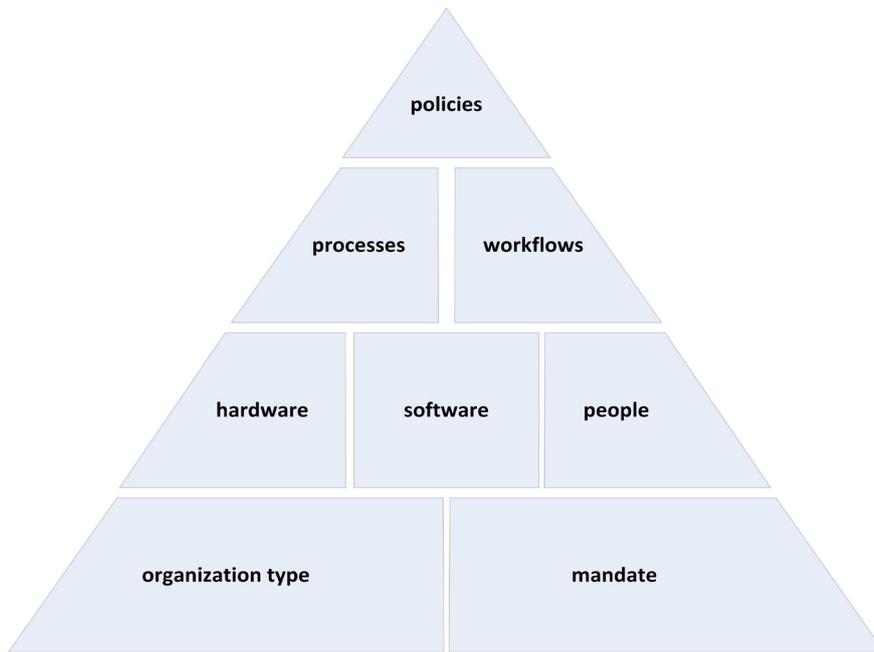
Figure 1: The digital preservation pyramid

Preservation Strategy

A preservation strategy formulates the "why and how" of digital preservation and should be the starting point of preservation activity within an institution. Comments one may come across when first talking about digital preservation in an institution may include: "Digital preservation will be solved by technology in a few years" – "We already have all our objects on spinning disks, so they are preserved" – "We've chosen PDF and therefore will never have to revisit our choice of file format" or even "Let's print it all out! Paper will last longer than anything digital".

A preservation strategy shall describe an institution's vision of maintaining accessibility and understandability of digital content and outline ways to achieve this vision. It will lead to one or several refined preservation policiesß. When formulating a strategy three main factor groups need to be considered: technological factors, organizational factors, and usability factors (see figure 2). Furthermore, the strategy needs to address the multifacetedness of a digital object and take into account the bit layer, i.e. the "ones and zeros", the logical layer, i.e. the file format, and the semantic layer, i.e. the contextual understandability, alike.

In current discourse, preservation strategy is often used in the context of a chosen preservation action, i.e., migration, emulation and normalization (see Strodl et al., 2007; Van der Hoeven, 2005). In choosing a preservation action strategy, the implications of the aforementioned technological, organizational and usability factors certainly also need careful consideration. Furthermore, the type of content may limit the choice of preservation action – for complex archiving objects

such as computer software, migration or normalization are certainly not suitable actions. However, preservation strategy which is only centered around a preservation action is limited to the question of logical preservation. Bit preservation and semantic preservation should be, as mentioned before, addressed in an overarching organizational preservation strategy.

Prerequisite

A preservation strategy can only be formulated based on a thorough understanding of the organization itself. Going back to the preservation pyramid (figure 1), the mandate plays a major role in the preservation intent. In regards to an archival mandate, different regulations may exist on a (sub-)collection basis, meaning that there might be a clear preservation responsibility only for a small part of the holdings and no preservation right for the rest. The mandate may also be tied to funding available for the task of digital preservation. Legal restrictions fall into the category of mandate – or lack thereof.

The role of the organization is closely tied to the consumer or to the designated community which the organization serves. University libraries, for instance, will target a different community than a highly specialized research institute or a company library. Expectations of the organization's stakeholders – data producers as well as data consumers – need to be taken into consideration. Lastly, the holdings need to be analyzed in regards to content type and technological factors such as data carriers, data formats and complexity.

Figure 2: Input categories of a preservation strategy

Based on a thorough organizational analysis a preservation vision will be formulated, which will be put into the context of organizational resources and capabilities in order to formulate a strategy. Establishing a preservation strategy and first workflows based on the strategy is therefore the initial change process. While digital preservation efforts have so far been mainly driven by large cultural heritage institutions such as libraries and archives at a national level, more and more smaller organizations are now realizing the importance of digital preservation (Strodl et al., 2011). Their initial change process often starts with a single sub-collection for which an urgent preservation need exists, e.g., research institutions needing to guarantee the long-term availability of a research project's data.

Ongoing change
The three main influence factors into a preservation strategy as shown in figure 2 are technology, community/users and the organization itself. As ongoing change in technology is the root risk to be mitigated in digital preservation practice, the task of "technology watch" is an established one in the preservation context. The same holds true for "community watch" which is the monitoring process targeting data producers and consumers. The intention of the producer and the intended (re-)use of data by the consumer need to be understood to fulfil the task of maintaining the information for the designated community, as defined in the OAIS reference model (CCSDS, 2012).

The third category also needs to be monitored – the organization itself. Changes in mandate, organizational structure or in resources such as funding, or the skill set of staff will significantly influence the preservation strategy. A continuous organizational watch as well as continuous internal and external lobbying for the preservation cause need to be included.

As the CPA/RLG Task Force pointed out, not only technology (and the community) is constantly changing, but also the organizational environment needs to change with it to fulfil the task of preserving the information throughout those external change processes (RLG/CPA, 1996).

Staff – "the who"
As digital preservation is not a mere question of technology, staff is an important factor in any preservation activity. When introducing digital preservation in an organization, various ideas about "the who" in digital preservation exist, such as: "Digital preservation is an IT task" – "The content specialists are responsible for the preservation" – "We have a classical conservation team – it should be their job". Each of those statements has some truth to it, as digital preservation is a cross-sectional task, including technical aspects, content aspects and legal aspects alike.

However, the nature of being a cross-sectional task may complicate the question of where within an organizational structure digital preservation should be positioned. Furthermore digital preservation activities call for new skill sets. While there has been an increase of job advertisements for digital preservation staff over the past 3 years (Kim et al., 2012), dedicated digital preservation staff is still not the norm for all institutions with a long-term stewardship for digital data (DPOE, 2010). Job positions related to digital preservation have various titles – data curation librarian, digital preservation analyst, web archiving engagement and liaison officer – to just name a few. But what requirements do organizations actively seek in preservation

staff according to job descriptions? A recent study of 110 digital curation job descriptions showed that besides degree and job experience requirements, "working in an information technology intensive environment" (58%) and "familiarity with standards and specifications" (55%) were the most sought after qualifications. The fact that 45% of the analyzed job descriptions called for "project management skills" may be taken as an affirmative indicator for the project-status of many digital preservation programs as well as for the high interdisciplinary and cross-sectional work. Only 23% of the job descriptions listed "working knowledge for curation" as a requirement (Kim et al., 2012).

Prerequisite
As mentioned above, the position of digital preservation within an institution needs to be carefully considered and depends on different organizational factors. During the initial phase it is beneficial to position the digital preservation staff as close to the overall management level as possible. This is usually achieved through a first "project" phase of digital preservation and has the advantage of potentially shorter decision making processes.

As a permanent task some organizations position digital preservation within the information technology department – a decision which should be considered carefully, taking into consideration the needed input from other departments for tasks on the logical and semantic preservation layer. It can be assumed that due to the cross-sectional nature of digital preservation, it is a task easier to position within a multi-linear or matrix organizational structure as opposed to a top-down hierarchy organizational structure. One of the first change processes is therefore the formation of a new team or department. Roles and responsibilities for the workflows defined on the basis of the preservation strategy should be defined as early as possible.

Dedicated digital preservation staff should have a sound knowledge of both, information technology and traditional library processes. As an initial change process, staff suited for the intended preservation strategy needs to be allocated (Bähr et al., 2011).

Digital preservation also changes internal and external communication flows. The cross-sectional input needed for digital preservation tasks – e.g., in evaluating format choice options suitable from an organizational, technological and stakeholder point of view – create the need for new internal communication flows. On an external level, communication with data producers and data consumers need to be established to communicate recommendations and to query needs.

Kim et al. point out that dedicated curation staff might also provide reference services or specialized research consultancy to other staff/users (2012).

Ongoing change
As digital holdings to be preserved grow, input of "new" departments and stakeholders may be needed. As part of the organizational watch outlined in the previous chapter on preservation strategy, staff should closely watch changes in the digital holdings of the organization and take action where needed. There may also be a temporal need for special staff resources not foreseen before, e.g., software development resources to develop tools which analyze data or automatically ingest data into a preservation system.

The most important ongoing change when it comes to staff is based on the fluid nature of digital preservation: due to the fast technological and organizational change, digital preservation is a commitment to life-long learning. While recent projects like the DigCurv project have outlined initial and ongoing curricula for digital curators (Molloy, 2013), the digital preservation community itself forms a valuable information source through various blog-platforms, journals, workshops, and webinars (Bähr et al., 2011). Digital preservation staff must be willing to regularly update their knowledge on state-of the art processes and technology and maintain close communications with the digital preservation community and the stakeholder community alike.

System choice
The last factor in this short analysis of change processes connected to preservation building blocks is the choice of preservation system – the "with what" of the preservation process. Again, first off common comments heard in institutions when starting with digital preservation activities: "This software will solve all your digital preservation problems" – "How can software help me – software is the root of all problems" – "Software for digital preservation should be selected based on technical criteria only".

Of course software is neither at the root of, nor is software itself the answer to all digital preservation problems. Software is, however, an integral part of digital preservation. It is needed for file analysis such as file format identification, technical metadata and validation; for preservation actions like migration or emulation; for accessibility with rendering software and for data management in repositories or digital preservation systems. In all of those tasks software is only as good as the processes and workflows that implement and use it. And – especially in the case of data management in repositories and digital preservation systems – software is only as good as the metadata stored with the object.

In this chapter we are not focusing on singular file analysis or rendering tools, but on larger preservation "systems" where the choice needs to be made between a system off-the-shelf (commercial or open source), a custom built system or an available service.

The system selection should be based on both organizational and technological decision criteria.

Custom built systems certainly have the highest degree of transparency. Quality and modularity of the system can be determined by the organization itself and licensing costs can be kept to a minimum. When extending existing open source repository systems, some form of community support is usually available. This comes at the cost of internal resources for initial integration and development as well as ongoing internal IT staff requirements for support and development, as the system needs to be adopted to meet new needs. Depending on organizational resources, the time from project start to roll-out of the solution can be longer than in the case of off-the shelf systems or available preservation services.

Off-the-shelf systems usually require little to no internal IT resources for development. The resources needed for integration differ from solution to solution. Usually the time from project start to roll out of the system is comparatively short. The continuous development of the solution is usually taken care of by the software provider who may also offer support or service models.

This, of course, may go hand in hand with licensing costs and is tied to a dependency on the software vendor. Also, as off-the-shelf systems try to target a wide customer base, there are usually drawbacks in fulfillment of institutional needs and integration of other existing systems may be tricky.

Available preservation services certainly have the lowest staff cost and no direct hardware and software cost. As there is little to no local technological implementation necessary on the organization's side, the time from project start to roll out of the solution is relatively short. The organization itself usually has no direct control over the data and all preservation actions are based on the decisions of the service provider. Access is usually only possible in pre-defined scenarios, so called "trigger events".

Prerequisite
In order to make a system choice, the overall preservation strategy needs to be in place. As part of the preservation strategy, an analysis of organizational parameters like budget, staff and legal restrictions should have been evaluated. Depending on the strategy a number of prerequisites for the system choice may exist; the mandate, for example, may bind the organization to storage within its own premises, which rules out preservation services. Furthermore, technological parameters like the existing software infrastructure play a role in system architect decisions but also in regards to the availability of interfaces, if existing systems such as catalogue systems or retrieval platforms shall be integrated into the chosen system. If the prerequisites are fulfilled, the initial change process is a system choice and introduction based on the aforementioned criteria

Ongoing change
It has been mentioned in the previous chapters that processes need to be adapted as new materials need to be preserved or new technology becomes available. This of course means a continuous change of the system, e.g., in form of including new workflows or in form of extending the system to suit new technologies. Monitoring the suitability for the digital preservation processes of all parts of the system itself should be an integral part of the aforementioned "technology watch" process. Scaling the system as the holdings grow is another change process which is to be expected.

An at first sight surprising, but very important preparation for a change process is the planning of the exit scenario from the get-go. As part of trustworthy digital preservation an organization needs to know how to completely extract data out of a system and move it to a new one. This scenario needs to be planned for from the start, as it may become necessary to change the system due to different factors, such as vendor problems or new state-of-the art processes.

Conclusion
This paper highlighted organizational prerequisites and ongoing changes needed in a digital preservation process. Three areas were analyzed: preservation strategy, staff and system choice. It was demonstrated that all three sections require initial change processes in an organization when first introducing the task of digital preservation, as well as ongoing change processes in a continuous process.

As it has been pointed out by the RLG/CPA task force, digital preservation is a fluid process and the organization involved in it

needs to adapt its processes alongside the task, to be able to meet the challenge (RLG/CPA, 1996). These changes hold true for all parts of the organization involved in the digital preservation process. A main outcome of this analysis is the importance of "organizational watch" – the re-evaluation of organizational factors in the preservation strategy. While technology watch and community watch are often included in best practice recommendations for digital preservation processes or as requirements in certification procedures, a constant self-analysis in form of an "organizational watch" is usually not mentioned yet.

References

Bähr, T., Lindlar, M. and Vlaeminck, S., 2011. Puzzling over Digital Preservation – Identifying traditional and new skills needed for digital preservation. [pdf] In: IFLA, World Library and Information Congress : 77th IFLA General Conference and Assembly. Puerto Rico, 13-18 August 2011. Available at: <conference.ifla.org/past/2011/217-bahr-en.pdf> [Accessed 12 December 2013]

British Library, 2013. British Library Digital Preservation Strategy 2013-2016. [pdf] Available at: <http://www.bl.uk/aboutus/stratpolprog/collectioncare/discovermore/digitalpreservation/strategy/dpstrategy.html> [Accessed 12 December 2013]

CCSDS, 2012. Reference Model for an Open Archival Information System (OAIS). Recommended Practice. [pdf] Available at: <http://public.ccsds.org/publications/archive/650x0m2.pdf > [Accessed 11 December 2013]

DPOE, 2010. 2010 DPOE Training Needs Assessment Survey. Cross-Tabulated questions. [pdf] Available at: <http://www.digitalpreservation.gov/education/documents/DPOE2010Survey_CrossTabs.pdf> [Accessed 18 December 2013]

Joint Operations Group of Archives New Zealand and National Library of New Zealand, 2011. Digital Preservation Strategy. June 2011. [pdf] Available at: < http://archives.govt.nz/digital-preservation-strategy-pdf> [Accessed 12 December 2013]

Kim, J., Warga, E. and Moen, W., 2012. Digital Curation in the Academic Library Job Market. [pdf] In: ASIS&T, ASIS&T 75th Annual Meeting. Baltimore, MD, USA, 26-30 October2012. Available at: <https://www.asis.org/asist2012/proceedings/Submissions/283.pdf> [Accessed 12 December 2013]

Molloy, Laura et al., 2013. D4.1 Initial curriculum for digital curators. DigCurV project deliverable. [pdf] Available at: <http://www.digcur-education.org/dut/Resources/D4.1-Initial-curriculum-for-digital-curators> [Accessed 12 December 2013]

RLG/CPA, 1996. Report of the Task Force on Archiving of Digital Information. Commissioned by: The Commission on Preservation and Access and the Research Libraries Group. [pdf] Available at: <www.clir.org/pubs/reports/pub63watersgarrett.pdf > [Accessed 12 December 2013]

Schwens, U. and Liegmann, H., 2004. Langzeitarchivierung digitaler Ressourcen. In: R. Kuhlen, T. Seeger and D. Strauch, ed. Handbuch zur Einführung in die Informationswissenschaft und praxis. Munich: Saur, 5th edition.

Strodl, S., Becker, Ch., Neumayer, R. and Rauber, A., 2007. How to Choose a Digital Preservation Strategy: Evaluating a Preservation Planning Procedure. In: ACM/IEEE, Joint Conference on Digital Libraries. Vancouver, British Columbia, Canada, 18-23 June 2007.

Strodl, S., Petrov, P., and Rauber, A., 2011. Research on Digital Preservation within projects co-funded by the European Union in the ICT programme. [pdf] Availabe at: <http://cordis.europa.eu/fp7/ict/telearn-digicult/future-of-the-past_en.pdf> [Accessed 12 December 2013]

Van Der Hoeven, J. and Van Wijngaarden, H., 2005. Modular emulation as a long-term preservation strategy for digital objects. [pdf] In: 5th International Web Archiving Workshop (IWAW05). Vienna, Austria, 22-23 September 2005. Available at: <http://iwaw.europarchive.org/05/hoeven.pdf> [Accessed 12 December 2013]

NOTES

1. Michelle Lindlar works as a technical analyst and as a preservation researcher at the German National Library of Science and Technology (TIB). As part of her work, she is currently involved in the EU FP-7 DURAARK project, where she leads the digital preservation work package. After having worked as a system administrator and IT project manager she joined the field of digital preservation in 2009 – and hasn't regretted it for a minute since. She can be reached by email at: michelle.lindlar@tib.uni-hannover.de

2. Publicly available examples for two rather detailed preservation strategies are the joint digital preservation strategy of Archives New Zealand and the National Library new Zealand (Joint Operations Group, 2011) and the British Library's preservation strategy for the years 2013-2016 (British Library, 2013).