European Holocaust Research Infrastructure
H2020-INFRAIA-2019-1
GA no. 871111

Deliverable 12.3

Report on long-term preservation of Holocaust digital objects: localised assessments and consultancy

Rebecca Dillmeier

USHMM

Start: April 2023 [M32]
Due: August 2023 [M36]
Actual: January 2024 [M41]

EHRI is funded by the European Union

Document Information

Project URL	www.ehri-project.eu/
Document URL	www.ehri-project.eu/deliverables-ehri-3-2020-2024
Deliverable	D12.3 Report on long-term preservation of Holocaust digital objects: localised assessments and consultancy
Work Package	WP12
Lead Beneficiary	14 - USHMM
Relevant Milestones	MS3
Dissemination level	Public
Contact Person	Rebecca Dillmeier (rdillmeier@ushmm.org)
Abstract (for dissemination)	This report contains a summary of a pilot mentoring project focused on digital preservation for EHRI partner organisations. It includes identification of organisational needs related to digital preservation, and potential implications for EHRI as an ERIC.
Management Summary	N/A

Table of Contents

1. Introduction **5**
2. Why is Digital Preservation Important? **5**
3. Ensuring Long Term Preservation **6**
4. WP12 Plans to Address Digital Preservation **7**
5. Project Plan and Evolution **8**
Description of Organisations Who Participated 8
Meeting Topics 9
NDSA Levels of Preservation 9
Outcomes of Discussions 10
Case Study on the NDSA LoP 11
6. Potential for EHRI **11**
6. Appendix: Survey Questions **12**

Glossary

DPC: Digital Preservation Coalition
EHRI: European Holocaust Research Infrastructure
ERIC: European Research Infrastructure Consortium
IT: Information Technology
LOC: United States Library of Congress
NDSA: National Digital Stewardship Alliance
USHMM: United States Holocaust Memorial Museum

Introduction

Digital access to historical collections, including digital copies of primary sources to Holocaust records, research data sets, and online exhibits and websites, has taken off in the last decade and will only increase into the future. A complex system of hardware and software, and the people who design, implement and manage them, work to ensure access to these materials.

Cultural Heritage organisations often embark on “digital” or “digitalisation” projects without plans for the continued management and maintenance of the data that is produced. Practitioners in these organisations then find themselves learning how to work with digital collections in real time, while also advocating for the funding, policies, infrastructure and personnel required to do this work. Creating “digital” products can sometimes seem like an easy solution to problems of increasing access, but maintaining the integrity and accessibility of digital data across time is a complex endeavour that requires appropriate financial, technical and organisational planning.

One of the focuses of Work Package 12 is to develop a better understanding of the current knowledge base relating to the long term preservation of the digital assets of EHRI-3 and partner organisations who often find themselves in the business of collecting, creating and managing increasing amounts of digital data. A small group of practitioners from variously sized partner organisations gathered to discuss the factors and conditions affecting long-term viability of Holocaust data objects. Although digital tools and techniques have been used heavily for quite some time, the acceleration of use of these approaches to meet demands placed on Holocaust institutions may be an indication that the "digital turn" is still underway, and although the rewards may be obvious, the risks of loss of access to digital data may not be fully appreciated.

Prior to delving further into this topic, we with to note of this work would have been possible without the active participation of a number of people from several institutions who contributed time, ideas and feedback under conditions of confidentiality to allow for frank and open discussions

Why is digital preservation important?

Vast amounts of digital information are produced every day, and much of it is of long-term value for future research. Without developing systems, policies, and the appropriate skill sets in staff to ensure long-term access to this information, it is at risk of loss. Organisations like museums and archives often set out to create data they want to save for the long-term, but just as often staff at these organisations find themselves managing data that develops over time as technology changes. Developing the appropriate skills and knowledge is essential for their ability to ensure the long-term accessibility of the data they find themselves responsible for. There is no such thing as benign neglect when it comes to managing digital materials - without constant monitoring the data will disappear or become unusable.

At its core, digital preservation ensures both the usability of digital data across time as well as its authenticity. When discussing primary historical sources, particularly regarding the Holocaust due to the strength of certain cultural forces intended to deny or minimise its historical truth, it is essential to ensure the integrity of the data across time. Ensuring integrity across time requires not only this byte level monitoring, but also preserving provenance and chain of custody preservation metadata to reinforce authenticity.

Digital preservation at the object level refers to the ability to monitor the integrity at the byte level of a file to ensure that each byte is preserved, allowing for the measurement of the smallest variation over time (such as bit rot) to be spotted and immediately acted upon. Ideally, there are no changes to the bytes of a file, ensuring no change over time. However, that is not guaranteed and such close monitoring allows for practitioners to take action as soon as possible to replace the file with another copy or document the change if no other copies exist.

Users of digital data from archives need to be able to trust the data they are being provided. If we want digital objects to be perceived as an accurate replica of the original and just as trustworthy, it is crucial that the integrity of files is accounted for to maintain a high level of institutional credibility. This refers to both the integrity of the file (its bits and bytes over time) and the integrity of the Institution maintaining that file and the information about it.

As Matthew Addis, the Chief Technology Officer and founder of Arkivum, wrote in a blog post on the DPC website, “But [researchers] can’t reuse data from a research data repository (or any other source for that matter) if [they] can’t trust it. And [they] can’t trust data if they don’t know its authenticity, completeness, correctness, who created it and how they did it.”[1]

Ensuring long term preservation

Two of the largest digital preservation advocacy organisations in the world define digital preservation similarly.[2] They point to the need for policies, processes, technology, and people necessary to successfully carry out digital preservation activities. The Digital Preservation Coalition (DPC)[3], based out of the United Kingdom, and the National Digital Stewardship Alliance (NDSA)[4], based out of the United States, have advocated for the need for digital preservation within the cultural heritage sphere for over a decade. This advocacy has included

[1] https://www.dpconline.org/blog/wdpd/blog-matthew-addis-wdpd

[2] There are other digital preservation advocacy organizations such as Digital Preservation Europe. https://cordis.europa.eu/project/id/034762

[3] The Digital Preservation Coalition is a membership organisation founded in 2002 whose mission is to build a welcoming and inclusive global community working together to bring about a sustainable future for digital assets. They focus on community, advocacy, workforce development, good practice and accountable, sustainable and dynamic governance. https://www.dpconline.org/digipres/what-is-digipres

[4] The NDSA is a consortium of 276 partnering organizations, including universities, professional associations, businesses, government agencies, and nonprofit organizations, all committed to the long-term preservation of digital information. Members work together to preserve access to our digital heritage. NDSA is hosted by the Digital Library Federation.

publicly accessible research findings, support for professional development, and developing a community of practitioners across North America and Europe.

According to the DPC, digital preservation refers to the series of managed activities necessary to ensure continued access to digital materials for as long as necessary. It refers to all of the actions required to maintain access to digital materials beyond the limits of media failure or technological and organisational change. Digital preservation is an active process that requires frequent attention, and that requires a series of actions over time to ensure digital content remains alive, discoverable, accessible and usable.

Digital content faces a unique set of risks, such as:

- Rapid technological change that may leave content unusable or unintelligible as the software that interprets it becomes obsolete.
- Requiring appropriate committed resources to make it possible to store and manage digital content
- Organisational change might leave digital content without a committed custodian.

The NDSA is an international membership organisation that supplies advocacy, expertise and support for the preservation of digital heritage. Launched by the United States Library of Congress (LOC) in 2010, it defines digital preservation as:

> The series of managed activities, policies, strategies and actions to ensure the accurate rendering of digital content for as long as necessary, regardless of the challenges of media failure and technological change.

Ensuring long-term digital preservation is a complex process that requires a network of strategies, policies, technology, and importantly, people with technical skills and dedication to monitoring files. The focus on digital preservation in Work Package 12 is on the knowledge and skills required to monitor the bits and bytes that make up files, including monitoring file and technology obsolescence and file fixity. Fixity ensures through monitoring that a file has not changed over time, ensuring the integrity of the data for future users. Although the discipline of digital preservation also encompasses topics such as preservation of software systems, web and social media archiving, and other areas, the scope of Work Package 12 is limited to preservation of digital files. Each partner institution gathers digital files worthy of long term preservation and the risk of not succeeding is evident.

Plans to address digital preservation

Work Package 12 gathered, through a combination of self-assessment surveys and group discussions, a better understanding of the factors and conditions affecting long-term availability of Holocaust data objects. This report details the results of the survey and the identification of factors important for digital preservation for EHRI partner organisations.[5]

[5] Questions from survey listed in Appendix

Topics covered include:

- Training and skill sets required
- Relationship between archivists and the information technology (IT) providers
- The National Digital Stewardship Alliance Levels of Digital Preservation
- The challenges of creating one set of guidelines given the wide range of organisations involved in EHRI

Project Evolution

Initially, the leads of the deliverable at USHMM discussed the intent of digital preservation in Work Package 12. USHMM planned to set up a series of mentoring pairs matching organisations based on size, staffing, budget, and levels of maturity in their approach to digital preservation. With only four respondents to the survey, in addition to USHMM, a decision was made to have the small group gather monthly over the course of 4-6 months to discuss tools, resources, and issues to get a better sense of the needs around digital preservation for organisations with Holocaust data. The group quickly identified that it was difficult for a group with such disparate needs to come up with practical solutions to addressing digital preservation at individual organisations, but that we could discuss the kinds of issues faced by differently sized and funded organisations.

We believe the fact that only four organisations responded to the survey highlights the need for developing an awareness of digital preservation for organisations in EHRI. Those who have studied disciplines relating to digital preservation are likely to have been exposed to the very real risks of data loss and know of actual incidents of such loss, and thus will tend to be more attuned to the possibility of losing data. However, institutions may not be positioned to allocate resources in training, staff time, and other costs required to support digital preservation activities.

Description of organisations who participated

The organisations ranged in size from mid-size to very large for both budget and staff size, with a range of digital files under management. Most organisations had a majority of image-based files, and the amount ranged from the several thousands to the tens of millions. Their collections were made up of digitised archival holdings, born digital records like oral histories (including institutional records), data sets, and websites. The smallest of the organizations does not actively collect digital collections, but often finds themselves managing digital copies of family and institutional papers. In the case of family papers, it is often the case that the families want to keep the original, but are willing to donate a digital copy to the repository. The larger organisations had inhouse IT services, while the smallest organisation contracted those services out. Their “preservation storage” is maintained by their IT subcontractor. One of the larger organisations has a digital preservation system backed by a hardware and software ecosystem that automates and reports on many preservation activities. Additionally, some of the organisations were state-funded, while one is a private non-profit. Most, but not all, of the

organisations undertook some level of preservation activities already: managing chain of custody, developing preservation policies, fixity checking, and file audits.

Structure of Meetings

The NDSA has a simple Levels of Preservation framework for practitioners to evaluate the status of their digital preservation practices, and then to further use it to explain to their funders, boards, executives, and IT partners how at risk their digital collections are. Practitioners can use this tool to assess their level of preservation practice and then determine from there what steps they can take to improve. The Levels of Preservation were used to structure the conversations and to identify themes. There are many tools available for this kind of self-evaluation, but we chose the NDSA levels of preservation model because of the low barrier to entry. It is easy to read, understand and explain.

Among participants, there was a clear understanding of why they're collecting, and they are able to record relevant metadata about provenance and other relevant chain of custody metadata, but it was recognized that the difficulties lie in getting the hands on technical experience and developing a collaborative, productive relationship with IT departments.

Meeting Topics

Meeting 1: Introductions, discussion of issues, introduction to NDSA Levels of Preservation
Meeting 2: Discussion of first two rows of the NDSA Levels of Preservation
Meeting 3: Discussion of last three rows of NDSA Levels of Preservation, tools like DROID and JHOVE
Meeting 4: Discussion about the themes previously identified and how they might apply to various different sizes and types of partner organisations which have many differing needs

NDSA Levels of Preservation[6]

Functional Area	Level			
	Level 1 (Know your content)	**Level 2 (Protect your content)**	**Level 3 (Monitor your content)**	**Level 4 (Sustain your content)**
Storage	Have two complete copies in separate locations Document all storage media where content is stored Put content into stable storage	Have three complete copies with at least one copy in a separate geographic location Document storage and storage media indicating the resources and dependencies they require to function	Have at least one copy in a geographic location with a different disaster threat than the other copies Have at least one copy on a different storage media type Track the obsolescence of storage and media	Have at least three copies in geographic locations, each with a different disaster threat Maximize storage diversification to avoid single points of failure Have a plan and execute actions to address obsolescence of storage hardware, software, and media
Integrity	Verify integrity information if it has been provided with the content Generate integrity information if not provided with the content Virus check all content; isolate content for quarantine as needed	Verify integrity information when moving or copying content Use write-blockers when working with original media Back up integrity information and store copy in a separate location from the content	Verify integrity information of content at fixed intervals Document integrity information verification processes and outcomes Perform audit of integrity information on demand	Verify integrity information in response to specific events or activities Replace or repair corrupted content as necessary
Control	Determine the human and software agents that should be authorized to read, write, move, and delete content	Document the human and software agents authorized to read, write, move, and delete content and apply these	Maintain logs and identify the human and software agents that performed actions on content	Perform periodic review of actions/access logs
Metadata	Create inventory of content, also documenting current storage locations Backup inventory and store at least one copy separately from content	Store enough metadata to know what the content is (this might include some combination of administrative, technical, descriptive, preservation, and structural)	Determine what metadata standards to apply Find and fill gaps in your metadata to meet those standards	Record preservation actions associated with content and when those actions occur Implement metadata standards chosen
Content	Document file formats and other essential content characteristics including how and when these were identified	Verify file formats and other essential content characteristics Build relationships with content creators to encourage sustainable file choices	Monitor for obsolescence, and changes in technologies on which content is dependent	Perform migrations, normalizations, emulation, and similar activities that ensure content can be accessed

Discussions brought to light the need for archivists and other practitioners to be able to gain practical skills to be able to do this work not often taught in archival programs, the need for strategies for communicating about digital preservation needs clearly with executives and IT partners, and strategies for advocating for appropriate funding for:

- Software and hardware needed to implement a robust digital preservation program
- Support for practitioners to continuously develop the technical skills needed to perform this work
- Adequate staff resources to perform this work
- The need to advocate for developing a cooperative relationship with IT services

Outcomes of pilot discussions

It quickly became clear that the initial idea of trying to partner organisations of similar size and budget turned out to be unworkable. The differences in resources, access to specialised technical expertise, and funding models made it next to impossible to share ideas for providing solutions across organisations. Several themes became apparent through the conversations regardless of size, budget, or organisational model:

[6] https://ndsa.org/publications/levels-of-digital-preservation/

- Organisations of all sizes find themselves with digital data in need of preservation
- Organisations embark on “digital” projects without plans for long-term sustainability for the outputs of those projects
- Practitioners find themselves advocating for and educating those who provide funding for work within an organisation (whether that is external funders, Boards of Directors, or Finance/IT departments within larger organisations)
- Practitioners often find themselves performing digital preservation activities and advocacy without it being a formal part of their job
- Practitioners need funding and opportunities for practical hands-on training opportunities to learn how to use and implement digital preservation tools
- Developing close cooperative relationships with IT services is crucial and complicated regardless of organisational model

While there were many similarities in the challenges being faced by organisations, identifying common solutions was difficult due to the varied nature of an organisation’s funding model, leadership, size of collection, and sophistication of the underlying technical infrastructure to support digital data.

Case study on the effectiveness of the NDSA Levels of Preservation
One of the smaller organisations with a growing collection of digitised items in their care, used the NDSA Levels of preservation to demonstrate to the organisation’s Board of Trustees the need to devote resources to digital preservation to ensure the long-term legacy of the organisation’s digitization projects. This certainly made an impression on the Board, although as yet efforts towards the creation and implementation of a digital preservation policy remain in their infancy. A major challenge to progress is that digital preservation is not part of the job description of the colleague responsible for digital matters, and they must take this on as an additional task. Nevertheless, as result, there is now recognition within the organisation of the need to initiate a digital preservation programme.

What might this mean for EHRI and its partner organisations moving forward?
EHRI supports developing a network of people and data related to the Holocaust with a focus on a transnational perspective. Its work has brought together practitioners from organisations of all sizes, and many of those organisations find themselves managing an increasing amount of digital data that requires archivists and other practitioners to approach their work with a new set of skills.

Many of these organisations’ digital data and files are at risk due to a lack of institutional understanding of the needs of digital preservation, leading to: staff taking on new responsibilities the organisation hasn’t planned for (budget, training, etc), new requirements for a closer relationship with IT departments that did not previously exist, and often the need for new policy development.

As EHRI transitions into a permanent body, a focus on the long-term sustainability of its own digital footprint will become more important. It is likely that like its partner organisations, the digital data it creates and manages will continue to grow. EHRI has an opportunity to reflect on how it wants to engage with the long-term preservation of its digital data alongside how it will engage with the needs of the partner organisations who make up its network.

Future opportunities for EHRI to further support digital preservation could include:

- Adding digital preservation to “technical” (non-research) fellowship funding opportunities
- Further developing partnerships with the research infrastructures in Europe that provide digital preservation services
- Make space for discussing the relationship and tension between technology and ethics (How does GDPR influence what European organisations can do? Should organisations be looking to commercial services like Amazon/Microsoft?)
- Providing models for evaluating the needs of starting a digital program (costs, technical infrastructure requirements, staffing, etc)

It is not a foregone conclusion that EHRI steps into the space of digital preservation, but it is an area worthy of further exploration in the EHRI Implementation Phase and once it is established as a body in the European Research Infrastructure Consortium. Like all other cultural heritage organisations, Holocaust-focused archives, museums, and libraries are increasingly turning to the creation of digital data to support enhanced access to their collections. As Holocaust denial and distortion continues around the world, ensuring the integrity of our collections and the data about our collections across time remains critical to the work of Holocaust organisations. As EHRI continues to evolve into an ERIC, digital preservation should remain a topic of discussion in the national nodes. National nodes could offer an avenue towards developing the support participating organisations need in order to confidently manage their digital programs, including ensuring long term sustainability is a part of planning for digitisation digital projects.

Appendix 1: Survey Questions

1. Does your organization have an active digitization program?
    a. If yes, please describe what type of material is being digitized and how
2. Are you collecting digital materials?
    a. If yes, please provide some more information below. Are you collecting materials that were born digital or materials that were digitized from analog?
3. How does your organization store digital materials? Please describe the storage space required and what technology is being used.
4. Are you a member of any digital preservation organizations?
    a. If yes, which ones?
5. What types of materials does your organization have that need to be preserved?
    a. Digital files, digitized archival holdings, data sets, websites, software
6. How would you describe the size of your collection?
    a. Small (several hundreds of digital objects)
    b. Medium (several thousands or tens of thousands of digital objects)
    c. Large (hundreds of thousands of digital objects or more)
7. What kinds of digital preservation activities is your organization involved in?
    a. Maintaining digital chain of custody; preservation policy development; redundant copies; fixity checking; file validation; characterization; periodic file auditing; periodic system auditing; none of the above
8. My organization preserves digital materials using:
    a. Open source tools; commercial system; none of the above
9. On a scale of 1-5 (1 less comfortable, 5 very comfortable)
    a. Do you feel like you have a good understanding of what digital preservation means in the context of your organization?
    b. Do you feel you have the appropriate skills, knowledge and resources to perform digital preservation?
    c. Do you feel comfortable educating others about digital preservation?
    d. Do you feel alone in implementing a digital preservation program?
    e. Are you interested in an opportunity to receive mentoring from an organization with experience with digital preservation?
    f. Are you interested in participating as a mentor to share digital preservation concepts and practical applications with another organization?
10. Please select some topics of interest:
    a. Working with IT department; practical tools for doing digital preservation; short and long term budget planning; materials to be digitized; file formats; how to document preservation process and policies
11. Please add anything else you would like us to know about digital collections and preservation at your organization or about what you are most interested in learning or sharing