TOGETHER FOREVER,

or How We Created a Common and Collaborative Digital Preservation Service

Johan Kylander
CSC – IT Center for Science
Finland
johan.kylander@csc.fi
https://orcid.org/0000-0002-8084-8233

Heikki Helin
CSC – IT Center for Science
Finland
heikki.helin@csc.fi
https://orcid.org/0000-0003-4002-8203

Kimmo Koivunen
CSC – IT Center for Science
Finland
kimmo.koivunen@csc.fi
https://orcid.org/0000-0002-9614-8370

Juha Lehtonen
CSC – IT Center for Science
Finland
juha.lehtonen@csc.fi
https://orcid.org/0000-0002-9916-5731

Abstract – Collaboration is a key to success in digital preservation. In Finland we have, within the national digital preservation services, established a successful model of collaboration with archives, libraries and museums to preserve our digital cultural heritage. Collaboration is essential through the whole life cycle of digital assets as digital preservation needs to be a dynamic activity in a constantly changing environment. In association with archives, libraries and museums we are able to build capabilities that are needed for digital preservation. We stipulate that we have been able to create a unified and centralized preservation service with a set of common specifications only through a close collaboration with the data curators.

Keywords – collaboration, community building, digital preservation services, common specifications

Conference Topics – Collaboration: a Necessity, an Opportunity or a Luxury?; Building Capacity, Capability and Community

I. INTRODUCTION

”Knowledge begets agony!” The Finnish proverb for ignorance is bliss, but with the process reversed, was an often heard expression cried out in meetings during the early days of planning and sketching the Finnish national digital preservation service. The meetings were actively attended by both the producer of the preservation service as well as various cultural heritage organizations. The expression conveyed both a sense of despair at the staggering task at preserving Finnish digital cultural heritage but also acted as a signal that we are in this together and that all problems are shared problems. This communal spirit that was prevalent during the planning turned out to be vital in successfully creating a functional service.

The national digital preservation services are a result of a long term and profound cooperation between the services’ producer CSC – IT Center for Science (CSC) and the organizations that curate the data. The Digital Preservation Service for Cultural Heritage preserves digital assets from the cultural heritage sectors, represented by archives, libraries and museums, whom are referred to as partner organizations. The national services are currently expanding to include preservation of research data, covering disciplines such as geophysics, astronomy and political sciences. The digital preservation system, the technical solution behind the service, is common for both cultural heritage data and research data.

Together we have managed to create a centralized digital preservation service that employs a shared technological solution for processing different types of data, as well as developing common specifications that define the format and structure of the data ingested to the service. The collaboration is an ongoing effort where we build preservation capabilities for cross-disciplinary and cross-institutional needs.

In this paper we briefly present our services and the forms of our collaborative effort. We present how the services were designed, what are their visible products from the user’s point of view and how the services are managed in cooperation with our partner organizations.

II. NATIONAL PRESERVATION SERVICES

A. Building the Service

Back in 2008, after years of unofficial discussions, the National Digital Library of Finland (NDL) initiative was formed within the remit of the Ministry of Education

16th International Conference on Digital Preservation
iPRES 2019, Amsterdam, The Netherlands.

DOI: 10.1145/nnnnnnn.nnnnnnn

and Culture of Finland. The project aimed at creating a nationally unified structure for contents and services, ensuring the effective and high-quality management, dissemination, and digital preservation of cultural digital information resources. Libraries, archives, and museums from the cultural heritage sector formed the basis for the NDL.

It was decided in the early days of the NDL that a common and shared digital preservation service should be created. This was based on a profound cost-benefit analysis. It was estimated that common infrastructure and services reduces costs, increases system integration, strengthens cooperation, and brings the practices of partner organizations closer together. Besides technical solutions, the collaboration between partner organizations was an essential goal of the NDL.

During the next years, the Ministry of Education and Culture, CSC, and partner organizations, in tight co-operation, designed a national digital preservation service, suiting the needs of all cultural heritage sectors. A lot of planning was done in the preservation support group consisting of the aforementioned actors. The preservation support group evaluated new features, devised the common specifications, and gave indispensable input on the release plan for the services. The support group also provided a valuable forum for partner organizations for sharing their experiences on digitization and data management processes, which in turn helped individual organizations in improving their processes.

Several targeted working groups were established to solve particular technical and social challenges. These working groups, consisting of invited experts on certain topics, were tasked with solving specific issues, such as recommending and evaluating file formats for audio-visual data or suggesting technical and/or structural metadata for different types of content. The working groups reported to the preservation support group, thus providing valuable input in creating the service.

The framework for planning the service came from the national NDL project. As one key area of focus was increasing interoperability, foster common practices and integrate systems, a national enterprise architecture for the cultural heritage sector was created. The enterprise architecture describes common high-level principles for managing digital assets, common processes for achieving the goals of the NDL project, and describes centralized services and technological solutions that are to be used. This architecture includes describing processes for improving the long-term digital preservation of cultural heritage by assigning the preservation to the national preservation service. The enterprise architecture also includes the Standard Portfolio. The portfolio is a specification giving an overview of all approved metadata standards and common interfaces that are allowed within the NDL project[1].

The successful collaboration which was conducted on many levels resulted in a set of common national digital preservation specifications and eventually in the national digital preservation service. The first AIP was created on November 2015, roughly seven years after the NDL project started. The volume of content to be preserved in the service is growing steadily. In 2018 the volume nearly doubled with approximately 107 terabytes of new content received for preservation. By the end of 2018, there was a total of over 222 terabytes of content to be preserved in our services.

B. A Formalized Service

The NDL initiative ended in 2017 but the collaboration continues as partner organizations have recognized the importance of cooperation and knowledge sharing. Our goal is to develop and maintain the digital preservation service on a sustainable basis. Currently we have preserved more than a million AIPs and the number of partner organizations preserving their digital content in the national preservation service grows constantly.

The national digital preservation services are owned by the Ministry of Education and Culture and are provided by CSC. The ministry and CSC have signed an agreement in which for example governance of the services is defined. Furthermore, many of our partner organizations have statutory obligations in preserving certain digital assets which increases the long-term scope of our services. Direct preservation costs are funded by the ministry, meaning that for partner organizations the utilization of the service is free of charge.

Partner organizations, under the remit of the Ministry of Education and Culture, may utilize the national digital preservation services after they have agreed with the ministry about borderlines. These include recognition of the collections to be preserved and quota allocated for these collections. After that, the partner organization makes a formal contract with CSC in order to start preservation, as depicted in Fig. 1.

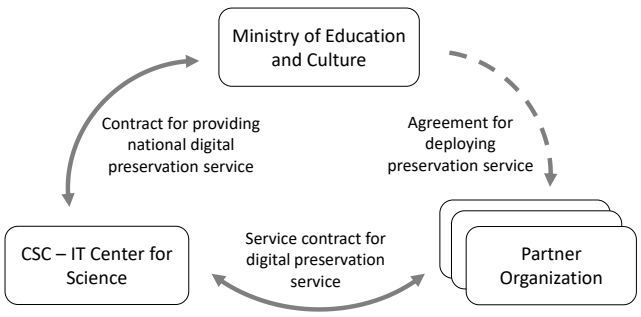

Figure 1. The contract model of the national preservation services.

[1] https://www.digime.fi/en/interoperability/enterprise-architecture/

III. Digital Preservation is Cross-Disciplinary

The foundation for effective collaboration begins with the simple acknowledgment that the data producers and curators are experts in their respective fields and know their own data and data models best. Therefore, they know what data to preserve, what it contains, and what it requires. The service provider, CSC, acts as a facilitator to discuss digital preservation issues together and between organizations. CSC also provides deep know-how about digital preservation and produce preservation tools for partner organizations.

Active collaboration is a necessity in digital preservation for several reasons. In addition to the cost-effectiveness of sharing knowledge and common tools we recognize the notion that digital preservation requires knowledge about a wide area of topics. These include knowledge about the contents of the digital assets, knowledge about how and why the assets have been created and included in a repository, as well as technical expertise on digital preservation methods.

The theory behind the collaboration rests on defining different levels of digital preservation that require different actions and know-how. As depicted in Fig. 2, digital preservation can be divided into three levels [1]. These are, starting from the bottom, bit-level preservation, logical preservation, and semantic preservation. The arrows to the left indicate from which angle the producer of the preservation service, from the

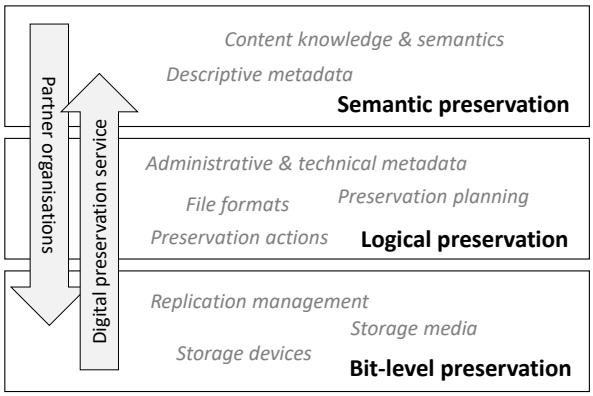

Figure 2. Levels of preservation

bottom going up, and the partner organization, from the top going down, approach digital preservation. It implicates that certain areas, like bit-level preservation, requires more technical know-how and its implementation is (mostly) the responsibility of the service producer. The semantic preservation requires deep found knowledge about the data and why it is to be preserved. This level is naturally (mostly) the responsibility of the curators.

This division is something that we have implemented in our service as the technological solution for the bit-level preservation is mostly managed by the service's provider. Reports on the bit-level preservation, such as hard disk failures or other statistical information, is regularly shared with our partner organizations. It is however in the areas of logical and semantic preservation that our collaboration with partner organizations really bears fruit. At the logical level, various tasks are shared between the service provider and partner organization, depending on the needs of the organization. The organization does not, for example, have to care about details of preservation actions if it does not have enough competence or resources to do so. On the other hand, the semantic level concentrates on semantics of assets to be preserved. These may vary a lot between disciplines. The service provider cannot have enough competence about various semantic details and therefore partner organizations are mainly responsible for the issues in this layer, possibly with help from the service provider.

IV. Collaboration equals Common Requirements

A. Common Specifications

The most visible product of the collaboration, uniting organizations from the diverse cultural heritage sector, is a set of published specifications common to all digital assets ingested into the services. The specifications were created to enable us to build a service that can receive, process, and preserve digital assets from the whole cultural heritage sector. They also define the interfaces of the national preservation services in order to aid partner organizations to integrate their production systems with the preservation services. The specifications have been developed and approved in close collaboration with all partner organizations. They are updated annually undergoing a yearly review process. Up-to-date specifications are available at our website[2].

Because the quantity of data ingested into the services is considerable, a fully automated processing of data during the ingest phase is essential. This, in turn, requires that all metadata in the submission information packages (SIP) ingested to the preservation services are machine readable, thoroughly defined, has to follow a standard and can be automatically validated. The automated processing requires both administrative, technical, structural, as well as descriptive metadata to be included as a part of the SIP in a controlled way.

As an answer to this, we have created the specification *Metadata Requirements and Preparing Content for Digital Preservation* that describes the required metadata and the mandatory structure of the information package. Work on the specification started during the planning phase of the preservation service as a part of the NDL project. Input and approval from partner organizations has been constant already from the beginning. The allowed metadata standards are defined in the Standard Portfolio. Together with the national METS profile, the Standard Portfolio acts as the framework for our specifications, as shown in Fig. 3.

[2] http://www.digitalpreservation.fi/en/specifications

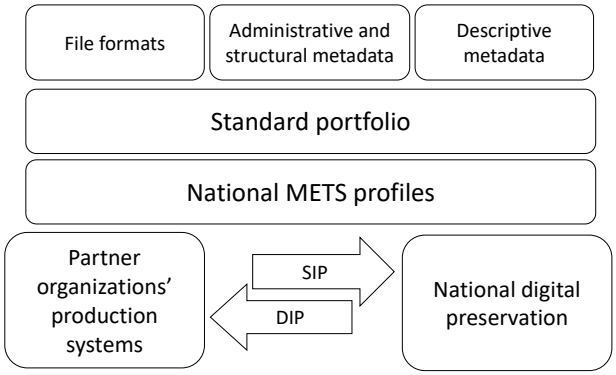

Figure 3.The specifications and standard portfolio govern the information packages.

The metadata specifications are not only functional requirements for our preservation services. They also have a pedagogical value in creating a common understanding of digital preservation metadata and of its quality by steering partner organizations toward implementing the required metadata and using recommended standards. This approach has been very successful, as some smaller organizations did not take these kinds of issues into account before introduction to our specifications. Further, some national organizations, not specifically performing any digital preservation activities, have utilized our specifications in order to improve their daily processes.

It is a common practice in digital preservation to carefully select file formats for preservation. These formats should be technically viable for a long time in the future. In our case the data to be preserved is diverse and includes text, image, audio, moving image, and database files. That is why we and our partner organizations have surveyed what file formats are commonly in use in the cultural heritage sector. We have created a set of approved file formats for preservation and a second set of file formats approved for ingest. The sets are based on an evaluation criteria that for example evaluate whether a file format is open, standardized, technically stable, and preferably backwards compatible with earlier versions. File formats that meet these criteria are approved for preservation. Common formats widely in use but that do not fully meet our evaluation criteria can be approved for ingest, implicitly with the notion that they will be migrated to another file format during the preservation.

The file format sets form the backbone of our specification *File Formats*. Every year the partner organizations can propose new formats if there is a need for it. These are reviewed according to the evaluation criteria and added to the specifications if they pass the criteria and are approved by the preservation collaboration group.

B. Regular Forms of Cooperation

The digital preservation services are administered by a steering group and the development is overseen by a collaboration group. The collaboration group is an open group consisting of partner organizations and others who are interested in the development of the services. The collaboration group discusses the long-term development of the service, approves the specification updates, and gives valuable input to the producers of the preservation services. At the same time it is a forum for discussing general digital preservation and presenting topics ranging from risk management, validation tools, and interfaces. The collaboration group meets at least four times a year.

Annually, in the fall, new demands and suggestions for updating the specifications are collected in the form of a survey sent to interested parties throughout Finland, including all partner organizations. The survey results are processed, resulting in specification updates that are reviewed and approved by the collaboration group. The new features are typically implemented in the beginning of the following year. The new specification versions are published at the same time as the features are implemented.

C. Increasing Capabilities and Knowledge

One cornerstone of our fruitful collaboration is a shared understanding of digital preservation and the goals of the service. Regular meetings and other support activities helps us maintain this shared vision. We arrange workshops and training sessions on different preservation topics several times a year in which participants are sharing knowledge with each other.

An important aspect of collaboration is raising the understanding of digital preservation issues among all participating actors. Its importance is twofold. Firstly, the quality of the digital preservation services improves as the services' producer gains more insight on preservation actions available as the input from the partner organizations increase. Secondly, curating digital assets and processing digital assets and metadata for ingest requires know-how. Therefore, an increased knowledge about digital preservation across the whole field is a necessity for increasing the volume of data ingested into our services.

Sharing knowledge and providing support for partner organizations as well as other institutions in Finland is an important part of the digital preservation services. The services maintain a support email address through which we annually answer about one hundred issues.

Extended support is given when a partner organization joins the service and starts preparing data for ingest. We also provide support in the form of tools which are published on GitHub[3]. These are tools for assisting partner organizations in creating SIPs for ingest and validating their data, for example the pre-ingest tool [2]. We strongly believe that sharing our technological knowledge in the form of a published set

[3] https://github.com/Digital-Preservation-Finland

of tools is an important aspect of both collaboration and in the end of the quality of our service.

As we have stressed above that successful digital preservation requires collaboration, we must also look beyond the national boundaries. Examples of somewhat similar approaches to our can be found for example in Germany and the Netherlands. The nestor Network [3] in Germany is doing somewhat similar forms of collaboration in Germany to what we do in Finland. Perhaps the biggest difference however is that they do not provide a concrete preservation service. The same is true for the Dutch Digital Heritage Network [4]. Both of these are successful national examples. Nonetheless, organizations fostering wider international cooperation are needed. Examples, such as the Digital Preservation Coalition (DPC) or Open Preservation Foundation (OPF), are needed. CSC joined the OPF in early 2019. International cooperation is essential in a wider perspective but international cooperation still cannot solve all national challenges.

V. CONCLUSIONS

While developing and maintaining a centralized preservation service for cultural heritage we have encountered several obstacles while trying to solve things in a cross-disciplinary and cross sectoral environment. We have learned that collaboration is the key to success and we consider it the only way to enable organizations from different fields and people with very different backgrounds to build a shared vision for digital preservation.

Even within the scope of cultural heritage, sectors can have different requirements and needs. Even sharing a common language can be challenging. However, with thorough collaboration, mainly in the form of regular meetings with commonly set targets, a consensus can be achieved. We have realized that the self-organizing nature of the early collaboration, tasking the organizations with leading expert groups and contributing heavily to the requirements, was a key factor in achieving a common understanding. From the common understanding grew the foundation for common requirements and ultimately a common service.

Maintaining this collaborative spirit is important throughout the whole life-cycle of preserving digital assets – it is a continuing effort. Addressing issues on the level of logical preservation, such as evaluating file formats and devising material specific preservation planning, requires input from both the data curators and the service provider. These are issues that we continue to process collectively through the preservation collaboration group.

Collaboration in the form of sharing knowledge results in better quality of data and its management. Sharing knowledge can be in the form of seminars, workshops and hands-on support. It also comes in the form of meetings, common specifications, and tools that automate certain tasks, easing the requirements for technological know-how and allowing organizations to spend their resources on curating their data instead.

As our services mature and expand to include more organizations with more limited technological and personnel resources, the importance of tools and even services that aid and automate certain processes for preparing content for ingestion into our services will increase. Ultimately, an increased knowledge on preservation issues in the cultural heritage sector as well as tools that lighten the workload of our partner organizations, enables the amount of data sent to our services to increase and the services role as an important centralized repository for Finnish digital cultural heritage and research data to be realized.

An important insight and lesson has been that submitting data to the national digital preservation services has forced the organization to think about data quality, the quality aspects in the creation of digital data, and proper management of digital resources. In some cases even changes in organizational culture has been a side effect of becoming a partner organization and submitting data to our services. None of this would have been possible without the collaborative process of common meetings, seminars and workshops.

Given all this, it is clear, at least to us, that collaboration for successful digital preservation is essential, and it should be broadened even more. There might be slightly different approaches to collaboration, but still it is very important, no matter in what shape. We are in this together, hopefully forever.

ACKNOWLEDGMENTS

The authors would like to thank all members of the Digital Preservation team at CSC, as well as the national preservation collaboration group for their valuable comments and input during the preparation of national digital preservation services. Especially, we thank the Ministry of Education and Culture of Finland for making this all possible.

REFERENCES

[1] K. Thibodeau, *Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years*, CLIR 2002.

[2] K. Lehtonen, P. Somerkoski, J. Törnroos, M. Vatanen and K. Koivunen, “Modular Pre-Ingest Tool for Diverse Needs of Producers”, in 14th International Conference on Digital Preservation (iPRES 2017), Kyoto, Japan, September, 2017.

[3] S. Schrimpf and Y. Tunnat, “15 years of nestor: German Network of Expertise in Digital Preservation”, in 15th International Conference on Digital Preservation (iPRES 2018), Boston, MA, USA, September, 2018.

[4] Dutch Digital Heritage Network, https://www.netwerkdigitaalerfgoed.nl/en/