BUILDING THE FUTURE OF DIGITAL PRESERVATION IN FRENCH ARCHIVAL SERVICES

Processes, functions and staffing for an effective digital preservation

Marion HUMBERT
Ministry of Culture
Paris
France
marion.humbert@culture.gouv.fr

Stéphanie ROUSSEL
Mintika
Paris
France
stephanie.roussel@mintika.fr

Édouard VASSEUR
Ministry of Armed Forces – Vitam Program
Paris
France
edouard.vasseur@culture.gouv.fr

Abstract –

Studied since the 1980s in French Archival services from a technical point of view, long term digital preservation strategy became a priority a few years ago in the context of Vitam program. The French Archive Interdepartmental Service, in partnership with the Vitam program team and the Ministries of Culture (National Archives), for Europe and Foreign Affairs and of Armed Forces, commissioned a study in 2018 about " The long-term sustainability of digital information ". This collective work offers a modeling of digital preservation processes and an analysis of the skills and profiles needed to implement these processes, not only for the three ministries but for any archival service wishing to diagnose its maturity.

Keywords – Capacity and skills; Preservation planning; Preservation Watch; Reference model; Workflow.

Conference Topics – Designing and Delivering Sustainable Digital Preservation; Building Capacity, Capability and Community.

I. INTRODUCTION

A. Context and objectives

Although many heritage services are now raising the issue of collecting digital records and archives, the subject of the long-term preservation of those data is still poorly understood in its concrete aspects. What perimeter should be given to the word "digital preservation"? Which processes come into play? What skills to involve?

In an attempt to answer these questions, the French Archive Interdepartmental Service (SIAF), in partnership with the Vitam program team and the Ministries of Culture (National Archives), for Europe and Foreign Affairs and of Armed Forces, commissioned a study in 2018 about " The long-term sustainability of digital information ". This study was a complement to the work done in the Digital preservation working group led by the Vitam program [1]. Indeed, thanks to the working group, perimeter of the function “preservation”, tools’ limits and complex

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

DOI: 10.1145/nnnnnnn.nnnnnnn

nature of processes are now better understood and shared by the actors, who defined some software development priorities in order to implement preservation functionalities in the Vitam Software Solution. These monthly meetings facilitated internal implementation and change management to "make digital curation a systematic institutional function" [2].

The study's aim was to assess the workload induced by the implementation of digital preservation using the digital archiving Vitam software in sponsor organizations [3].

B. Methodology

The study was entrusted to two consulting firms specialized in supporting archives and information management services: Serda and Mintika.

It was conducted in 3 phases:

- The definition phase aimed at providing a clear, shared understanding of "digital preservation" in the broader context of information lifecycle management and to provide resources to final users (directory of experts, summary of applicable standards, bibliography)
- The core phase detailed the digital preservation processes required for an effective implementation of the function. It also identified the skills and expertise necessary.
- a third (non-public) chapter provided sponsor organizations with options for change management and long-term support of digital appropriation, human resource estimation and mutualization and outsourcing priorities.

Deliverables of Phases 1 and 2 are generic enough to be used by any archival service, and were not designed solely for the three ministries. Indeed, they are published and publicly available (in French) on "FranceArchives" resource website [4].

C. Perimeter

The study started with an analysis of the perimeter of the "digital preservation" domain and a mapping of concepts related to digital preservation. This immediately reveals that formats, which are usually the first line of inquiry associated with digital preservation, represent but a small part of the information available on the subject.

After discussions between the partners, it was decided, as shown in the figure below, to include in the studied perimeter not only technical and strategical aspects of digital preservation but the starting point of the document life-cycle: acquisition and the aim of preservation: access.

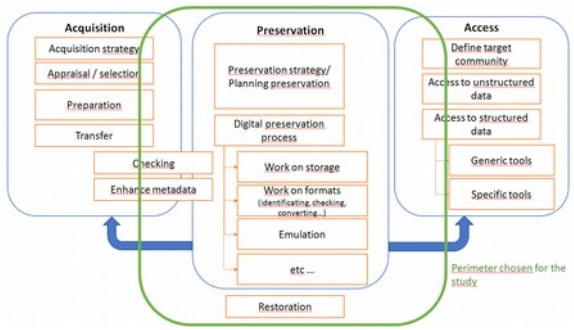

Figure 1: perimeter chosen for the study

II. Modeling processes

A. Why?

Several modelings about digital preservation already exist, especially the ISO OAIS Reference model [5], the proceedings of the international collaborative research project Interpares [6], partially translated in French [7], or the PLANET Model, designed

from OAIS [8] and its projects of implementation in several institutions [9]. These approaches are interesting, but take place in a larger documentation, sometimes not translated in french. Furthermore, the Interpares UML modelling is not the most adapted for large communication when the diffusion of an all-in-one document to professionals and to the hierarchy is one of the main aims of the study.

The primary objective of this modeling was to raise awareness of the magnitude of the task throughout the network of archival public services. Indeed, while more and more electronic archiving systems are being deployed in France, the issue of digital preservation is still poorly understood by most professionals. This work is therefore intended to be educational. It is also possible to imagine future documents that will accompany this modeling to facilitate its appropriation and diffusion. Network heads thus have an important role to play in developing future digital preservation doctrines for all archiving services. It is indeed obvious that not everyone will have the means to deploy such varied skills and bear such high costs as have been highlighted in the modeling.

B. How?

The modelling was developed on the basis of this available professional literature (in particular the OAIS model and works of Interpares) as well as information gathered during interviews conducted with the Ministries of Culture (National Archives), for Europe and Foreign Affairs and of Armed Forces and a number of partners operating major electronic archiving systems (the National French Library, the National Computer Center of Higher Education and the National Audiovisual Institute). This double perspective – a conceptual approach confronted with a very operational approach – is the basis of the work.

The in-depth review of OAIS and Interpares enabled us to identify the key points of a preservation system by specifying for every process the entry and exit points. We identified these key points in the form of post-its that we were then able to arrange at will to correspond to the concrete reality of the projects and partners in the study. This synthesis work was the first step for the modelling.

Finally, OAIS and Interpares were also sources of thinking on how to model and approach the subject of preservation. It was out of the question to redo what had already been done, but we would rather propose, on the basis of this conceptual work, a very operational approach to the implementation of preservation, in the particular context of French archival practice.

Consistency with these different works, in particular with OAIS, is a guarantee of the interoperability of the model. We can thus imagine that this modeling of the digital preservation function could be appropriate for other contexts than the French context.

C. The use of the BPMN modelling language

The language used for modelling is BPMN (Business Process Model and Notation), which provides "a business process model and notation to describe an organization's value chains and business activities in the form of a standardized graphical representation" [10]. This model has the advantage of being standardized (ISO/IEC 19510). It guarantees a

certain stability and interoperability, which seemed important in our context.

BPMN is also widely documented, making it easy for all users to understand and master it. Finally, BPMN is a common language in the developer community, which eases the accustomization to a community unfamiliar with purely archival issues.

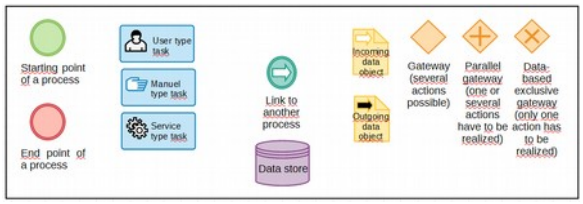

Figure 2 BPMN caption of the modelling

D. *The final result*

The final modelling details 19 processes, split in the following categories:

- ingest: 2 processes
- access: 1 process
- storage: 1 process
- administration: 5 processes
- preservation planning: 6 processes
- management & governance: 4 processes

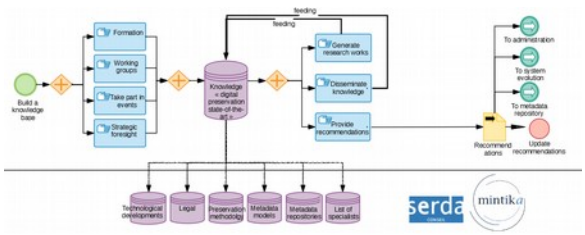

Figure 3 Extract from the modelling of the processes included in "Preservation planning"

By detailing the processes we are able to identify which ones require the more resources: appraisal, ingest, format tests, strategic foresight, international events, participation in development of collaborative tools addressing specific elements in the French environment, i.e. the Lambert conformal conic projection, initially missing from the GeoTIFF specifications.

This process-based approach was planned to be complemented with a skills and profiles thinking. In a different context, the subject has already been studied in France by the National French Library [11].

III. Designing digital preservation functions and evaluating ideal staffing

A. *Why?*

In France, the training in heritage preservation doesn't include a major in long-term digital preservation, neither in archives, nor in libraries or museums. Consequently, the skills needed for digital preservation are not precisely defined, or not defined at all, in the tools and repositories of professional associations [12] or administrations' human resources offices [13].

These elements of expertise had to be detailed, particularly those related to information watch operations, critical for the decision process. The technical and archival skills necessary to operate an electronic document and record management system were an important part of the reflection. Part of these skills can be found among computer scientists while others are more specific and imply thorough archival knowledge [14].

The 2018 french modelling will be completed by exploiting and translating international resources about formation in digital preservation, like the SCAPE project [15] and the DigCurV Framework [16], published in 2013 and used by the Oxford University to develop continuous training in patrimonial institutions [17]. In addition to providing surveys and methods of evaluation for active

professionals, these tools will enhance our reflection about the initial formation.

The aims with regard to formation are both increasing the number of digital preservation professionals and having a better understanding of our needs, especially in order to share skills and knowledge with smaller institutions.

B. How?

As mentioned beforehand several times, the objective of our model was to allow partners to project themselves into a future organization. To define profiles and facilitate the appropriation of the model by the study sponsors, it seemed important to use existing human resources categories, even though these frameworks will have to evolve over time to take into account new needs. These profiles were then compared with the references in the professional literature – in particular the NDSA report [18] – and mapping done whenever possible..

C. Identified functions

Profiles for each process (archivist, developer, system administrator, etc.) have been detailed in the form of a diagram and table, in order to facilitate in a given context the identification of the skills required to implement a strategy for the sustainability of digital data.

The published report focused on the profiles specifically related to digital preservation: format management before and after payment into the electronic archiving system, evaluation, conversion, standby. Resources required for the treatment and the analysis of the data themselves or the constitution of the chains of processing when it comes to releasing data from applications were considered out of scope.

Profiles are based on a list of skills needed to implement digital preservation, including archivists (taking in account the balance between archival and technical skills [19]), formats specialists, data restorers, developers and quality specialists.

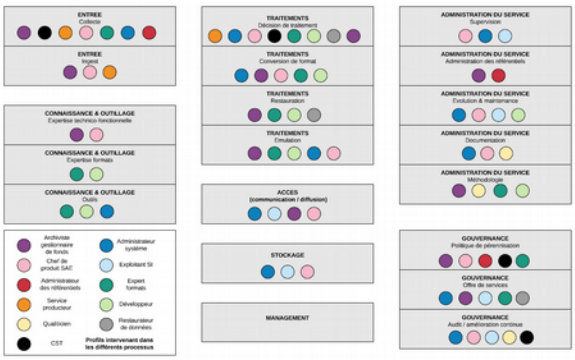

Figure 4 Profiles required for each process

Function	Description	Function in NDSA
Administrateur de référentiel	In charge of management of metadata repositories	Metadata Analyst
Administrateur système	In charge of the upkeeping, configuration, and reliable operation of computer systems	System administrator
Archiviste gestionnaire de fonds	In charge of the respect of archival science principles in the digital environment	Cataloguer / Archives & special collections curator / Digital Archivist / Collection needs analyst / Electronic record archivist
Chef de produit SAE	In charge of administration and supervision of the electronic document and record management system	Administrator / Digital Archivist

Contrôle scientifique et technique	Audit by the archival Authorities	
Développeur	In charge of software development process	Software Developer
Expert format	In charge of developing and maintaining knowledge of archived formats	Format specialist
Exploitant du système d'information	In charge of software maintenance, infrastructures and cybersecurity	
Qualiticien	In charge of quality management system	Workflow specialist
Restaurateur de données	In charge of developing technical solutions to fix damaged digital documents	*Data restorer (not in NDSA report)*
Service producteur	Organization who created the documents and may have expertise on some formats	*Archive-producing department (not in NDSA report)*

Table 1 Correlation between functions identified by NDSA report and our modeling.

The legal-related functions have been considered out of the strict perimeter of digital preservation in the study, and therefore excluded.

D. Ideal staffing

The three scenarios identified for each ministry are based on the needs and only take into account the work load directly related to the digital preservation of data.

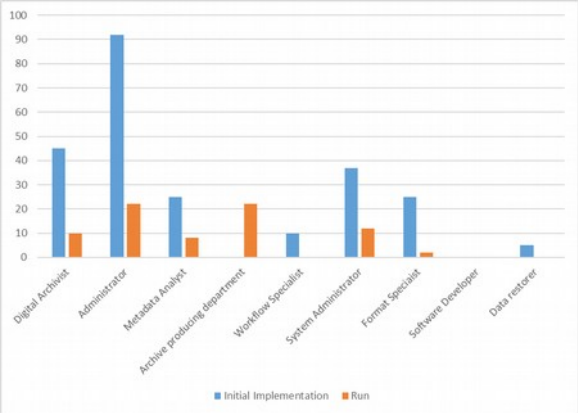

Figure 5 Estimation of the human resources (day/year) necessary to the initial implementation and run of digital preservation in the National Archives.

This projection will evolve as the sponsors will develop their maturity and set up the processes described in the modelling.

IV. Work in progress

A. Advocacy

In order to raise awareness about the subject of long-term digital preservation within sponsor organizations, a presentation of the results of the study will be submitted to the Interdepartmental Committee of French Archives in April 2019. The presentation will point out the possibilities to share resources and expertise between the three ministries, particularly for the tasks related to technology intelligence. The Ministry for Europe and Foreign Affairs presented internally the results of the study, in order to prepare change management as soon as the processes will be ready to be implemented.

These political and strategical aspects will be completed by a more technical and operational working group, created within the national PIN Group [20] (Digital Information Preservation) in order to pursue the tests and exchanges started in the Vitam program

B. Implementation

Each sponsor should now assess its situation and resources with respect to the modelling and skills identified in the study. It offers 25-year estimation in order to help anticipating the resources needed for an efficient strategy of digital preservation. Several others archives in France plan to use the study to perform diagnoses.

For example, the Ministry of Armed Forces will use the results of the study in the following ways:

- Elaboration of a digital preservation governance within the institution.
- Evaluation of the human and budgetary resources necessary for an efficient strategy of digital preservation. A part of this action consist in the identification of the current resources.
- Identification of ten actors in the preservation process and elaboration of an expert profile, who will coordinate all the actors and operations linked to digital preservation.
- Bridging the gap between the “Digital preservation” processes and the processes “Preventive and curative preservation” which already exist, and integration of these processes in the data life cycle (ingest, management, access).
- Update of existing models and internal processes, inclusion of results in ongoing projects.
- Bridging the gap between the traditional archival concepts (appraisal, description, preservation…) and the “Digital Preservation” processes.

Similarly, the National Archives are planning in 2019 to complete their digital preservation strategy thanks to the scenarios of the study. Besides, institutions with particular competences in formats and long-term digital preservation, like the National Computer Center of Higher Education are working in order to enhance the Vitam software solution and will share their developments with all the community of users.

For many institutions, especially the smallest, contributing to research and development works, creating new posts will be difficult, and the preservation processes will have to be implemented with constant resources. That is why the implementation of the modeling has two important issues: firstly, national and international coordination in order to reduce costs and have the widest possible network of expertise and then networking of existing competences.

C. Communication

Even though the study addressed internal needs of the sponsor organizations, a larger diffusion among French archival services has been made through a publication in FranceArchives, the national portal of French Archives. The model and mind-map of the scope, the processes, a bibliography and a summary of the most important standards and norms in the digital preservation area are downloadable.

Furthermore, the results of the study will be presented in April 2019 during the French Archivists Association forum. A panel associating representatives from SIAF, Vitam team, the Ministry of Armed Forces and the National Computer Center of Higher Education will present the deliverables as the foundation of many concrete applications.

V. Conclusion

Carried out at an interdepartmental level, this study took place at an ideal time, just while the Vitam project and its autonomous implementations in each ministries become operational. Without disrupting ongoing projects within each entity, it has enabled stakeholders to consolidate their strategy. They are expected in turn to adjust the practical implementation of the Vitam software solution, to identify new skills to be internalized or outsourced, to identify new profiles to develop and how to include these processes in the pre existing standards and frameworks [17].

By its concrete and quantified nature, it provides archival services with essential tools for dialogue with decision makers and enables them to consider future joint actions by pooling certain necessary functions, especially in preservation watch.

ACKNOWLEDGMENTS

The SIAF and the Vitam program team would like to thank Pierre FUZEAU and Olivier MOLLET from the firm Serda; Baptiste NICHELE from the firm Mintika; Wilfried PRIEUR; Blandine WAGNER; Sophie RESSEGUIER and Olivier CHEVENIER from the Ministry of Armed Forces; Martine SIN BLIMA-BARRU from the National Archives; Erwann RAMONDENC and Juliet GARNIER from the Ministry for Europe and Foreign Affairs; Eleonore ALQUIER and Valérie CHAUMELLE-SERRUS from the National Audiovisual Institute; Stéphane REECHT, Thomas LEDOUX and Laurent DUPLOUY from the National French Library; Lorène BÉCHARD and Olivier ROUCHON from National Computer Center of Higher Education and Aurélien CONRAUX from the Digital Archives, Standards and Authorities Office (SIAF), for their contributions.

REFERENCES

[1] M. REBOURS, “The Digital Preservation Working Group led by the interdepartmental Digital Archiving Program - Vitam”, *Ipres 2018.*

[2] C. PROM, “Making Digital Curation a Systematic Institutional Function”, *The International Journal of Digital Curation*, p. 139-152, 2011.

[3] E. LABORDE, M. REBOURS, E. VASSEUR, “Vitam Software Solution’s preservation”, *Ipres 2018.*

[4] ISO 14721:2012, Open Archival Information System. Traduced in French in 2017.

[5] “Préserver les données numériques : de quoi parle t-on ?”, *Portail FranceArchives,* 2018. URL: https://francearchives.fr/fr/article/162808225 (accessed 13 March 2019).

[6] Interpares 2 Project, Chain of Preservation Model, 2002-2007. Translated in French in 2010.

[7] C. NOUGARET, F. BANAT-BERGER, “La traduction française d’Interpares 2 : un nouvel outil conceptuel pour l’archivistique”, *La Gazette des Archives,* n°245, p. 67-77, 2017.

[8] B. SIERMAN, P. WHEATLEY, *Evaluation of Preservation Planning within OAIS, based on the Planets Functional Model*, 2010.

[9] P. LUCKER, M. RAS, B. SIERMAN, R. van VEENENDAAL, “Preservation Watch at the National Archives of Netherlands”, *IPres 2018.*

[10] ISO/IEC 19510:2013, Object Management Group Business Process Model and Notation, 2013.

[11] E. BERMÈS, L. FAUDUET, “The Human Face of Digital Preservation: Organizational and Staff Challenges, and Initiatives at the Bibliothèque nationale de France”, *The International Journal of Digital Curation,* p. 226-237, 2011.

[12] Association des archivistes français, *Référentiel métiers, la profession d’archiviste ou les métiers des archives*, 2e édition, Paris, 2016.

[13] Ministère de l’action et des comptes publics, *Répertoire interministériel des métiers de l’État*, 2017.

[14] J. KIM, E. WARGA, W. MOEN, “Competencies Required for Digital Curation: An Analysis of Job Advertisements”, *International Journal of Digital Curation*, p. 66-83, 2013.

[15] SCAPE, 2014, Scalable Preservation Environments (scape-project.eu)

[16] DigCurV, 2013, DigCurV Curriculum Framework (https://www.digcurv.gla.ac.uk)

[17] S. MASON, E. HALVARSSON, “Designing and implementing a digital preservation training needs assessment”, Ipres 2017.

[18] A. WINSTON, et al., *Staffing for digital preservation 2017*, National Digital Stewardship Alliance, p.20-36, 2017.

[19] C. A. LEE, H. TIBBO, “Where’s the Archivist in Digital Curation? Exploring the possibilities through a Matrix of Knowledge and Skills”, *Archivaria*, p. 123-168, 2011.

[20] Groupe de travail “Pérennisation des informations numériques” (http://pin.association-aristote.fr/doku.php/accueil)

[21] C. BECKER, G. ANTUNES, J. BARATEIRO, R. VIEIRA, “A Capability Model for Digital Preservation. Analysing Concerns, Drivers, Constraints, Capabilities and Maturities”, *Ipres 2011*.