

Meet eArchiving workshop report

3-4 December 2019, Borschette Centre, Brussels, Belgium

A Connecting Europe Facility workshop

European Commission

DG CNECT

Unit CNECT/G2 Interactive Technologies, Digital for Culture and Education

Contents

Summary	1
Agenda	2
Workshop outcomes	2
Breakout session	3
Q1. How do you deal with the preservation of your data? What reuse do you do of it?	3
Q2. What kind of data are you dealing with, and what do you do with these opportunities?	4
Q3. What are the potentials and barriers of eArchiving in your domain?	4
Conclusions and lessons learned	6

Summary

“Meet eArchiving” was the first dedicated event organised in Brussels by the European Commission (EC) to enable anyone interested in digital preservation to gain an in-depth understanding of the CEF eArchiving Building Block. The event was hosted by the EC’s Directorate General (DG) CNECT G2 Unit - Interactive Technologies, Digital for Culture and Education, which is in charge of eArchiving; together with the Connecting Europe Facility (CEF) Stakeholder Management Office (SMO) from the EC DG for Information Technologies, DIGIT.

DG CNECT G2 also covers other CEF digital technologies, such as the student eCard¹ and Europeana², dedicated to the reuse of cultural resources and encompassing 3D models, digitised material from museums, art galleries, libraries or archives.

There is a clear need for correctly preserving digital material, and for making it available for reuse while preserving its authenticity, both for future generations and for legal reasons, and guaranteeing interoperability cross-border and through many domains. The importance of reaching out beyond the archives sector implies looking into the many issues of digital longevity in other sectors.

In this context, the Commission decided to hold this workshop with three main objectives:

- (1) raise awareness of digital preservation, and of the potential of eArchiving in the field;
- (2) reach out beyond National Archives, both to other levels of public administration and to other domains; and
- (3) listen to the various users communities to improve policies on these matters.

¹ <https://ec.europa.eu/digital-single-market/en/eu-student-ecard>

² <https://www.europeana.eu/portal/en/about.html>

The workshop lasted two days: the 3 December 2019 included a series of presentations by experts, with each session providing an opportunity for audience questions. The following day, 4 December 2019, had a different format: first an introduction to two domains other than National Archives with a panel discussion; and then some breakout sessions, where the audience was split up into small groups, each with a group discussion facilitator, with the aim to gain a comprehensive understanding of the participants' needs and issues in regard to digital preservation. All sessions were live-streamed via internet, and recorded so they could be viewed after the workshop. The recordings are available on the workshop page.

Agenda

The full agenda, speakers bios and presentations are available on the workshop web page:

<https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/Meet+eArchiving>

Recordings of the streaming can be found here:

<https://webcast.ec.europa.eu/cef-earchiving-expert-group-and-stakeholder-workshops-03-12-19>

<https://webcast.ec.europa.eu/cef-earchiving-expert-group-and-stakeholder-workshops-04-12-19>

An article about the event is in:

<https://ec.europa.eu/cefdigital/wiki/pages/viewpage.action?pageId=171771351>

Workshop outcomes

Some figures:

- Registrations: 186.
- Attendance: 123 people from 33 countries.
- Speakers: 20.
- Web streaming: 538 connections from 28 countries.
- Most attended building block event to date.

Feedback:

- *“Thank you very much for organising this fantastic workshop, which gave great insights into the development of eArchiving.”*
- *“Great event that showcased the increasing interest to archive digital objects and eArchiving could help achieve this goal in a more harmonized, interoperable and cost-efficient way.”*
- *“The eArchiving CEF Building Block could:*
 - o *Provide a clearer description of the actual service offered (...)[, as] no storage/infrastructure is provided.*
 - o *Improve the online identification and location of the E-ARK specs.”*

Breakout session

During the last session of the second day, attendants were divided into nine groups to discuss and answer the following questions. A summarised, not exhaustive transcription of their feedback follows. It has been edited for clarity and grouped into common themes.

Q1. How do you deal with the preservation of your data? What reuse do you do of it?

Comments and answers could be analysed from the following points of view:

a) Legal and policy framework:

- Institutional regulations or national laws dealing with digitisation and digital archiving processes are in place in many cases. At the same time: there are gaps and lack of detail in current laws and regulations (e.g. on file formats, metadata, transfer and archiving). In the middle, one can find hybrid preservation systems: digital objects together with paper archives.
- Specific or formal Digital Preservation plans covering policies, a mission statement, legal obstacles, data selection procedures carried out with data producers, data retention periods, preservation workflows and preservation metadata.
- Different levels of digital preservation depending on the type of data:
 - o public data should be preserved for a very long time;
 - o private data, such as economic data (e.g. invoices) should be preserved for 10-12 years, depending on the country;
 - o citizen-related data, such as health data, should be preserved up to 75 years; similarly for private or public pensions;

b) Technical infrastructure and resources:

- While some answers mentioned digital repositories in use, none mentioned the cloud in general though.
- Both open-source and proprietary solutions are being used for digital preservation.
- Public agencies that should maintain their own intermediate digital archives, before transferring the data to the national archives for permanent preservation. Such organisations need to have the requisite technical skills, competences and resources.
- Some mentioned issues regarding medium or long-term preservation of Electronic Records Management Systems. Also, the chosen (or mandatory) metadata profile used can have a big impact on Digital Preservation.

c) Questions on:

- Difference between archiving and creating back-ups.
- Guidance on what tools to use for which digital preservation action, e.g. using DBPTK for database Digital Preservation, etc.

d) Concerning the reuse:

- The need for all types of metadata is very important.
- It is also necessary to have an accurate search interface.
- Reuse is happening more and more separately, through intermediary systems (i.e. the organisation maintaining the archive provides data to other systems which serve the actual end users).
- Data access depends on its sensitivity:

- *Closed data*: kept for legal, compliance and/or economic reasons, but only for internal reuse. For example, invoices/financial data or national digital treasures.
 - *Restricted access*: data can be reused externally but only by authorised users to avoid misuse, for example health or work-related (i.e. for pensions).
 - *Open or public access*: data is not restricted and is publicly available for historic or other research and reuse, sometimes after a digitisation project.
- Transparency is vital for public records.

The use of the Archives Portal Europe (APE) was mentioned.

Q2. What kind of data are you dealing with, and what do you do with these opportunities?

The kinds of data mentioned can be classified into the following categories:

- Administrative and legal records, both historical and digitally-born:
 - Historical records should be digitised then made available via APE, for example for people interested in European history.
 - Transactional data can be reused for analysis (mentioned for tax purposes.)
 - Commercial data could be used for contracts (mentioned for the banking sector.)
 - GDPR should naturally be implemented.
 - Health-care data, patient information.
- Social Media/Internet:
 - Web pages.
 - Tweets and other social media.
 - Issue: how should we archive linked content (e.g. a social network), and how should its original form be presented?
- Databases.
- eInvoice: data are in XML format and comes in very high volumes.
- Maps.
- Public data for Cultural Heritage.

General opportunities mentioned for data are:

- Restricted data can be anonymised and used for statistical analysis.
- Common standards for different sectors would be useful at the EU level.

The big amount of databases provide possibilities for quantitative research.

Finally, some issues regarding data were mentioned:

- The lack of a common legal framework is an issue.
- Structured vs. unstructured data in terms of analysis and potential: it is hard to do much with unstructured data.
- How to separate digitised vs born-digital data, as these are to be preserved and reused differently.
- Digitally signed vs unsigned documents need to be handled differently; especially, appropriate metadata about the signature for the document to remain trustworthy.
- Lots of file formats; archives can be overwhelmed.

Q3. What are the potentials and barriers of eArchiving in your domain?

a) Regulation

- Lack of digital archiving regulation is blocking the establishment of cross-border available digital archiving services. For example, the CEF eID building block is established based on the eIDAS legislation, which allows for sufficient compliance and scaling.

- Rules and regulations at the EU level would make it much easier for Member States to adopt CEF eArchiving.
- Lack of political willingness is probably the main barrier which translates into the lack of support and funding for digital archiving;
- The digitisation of processes and data is (legally) not supported enough or even allowed in some countries or areas.

b) Standardisation

- Harmonization at the EU level, which is of real benefit to research agencies, enabling them to get ahead, thus creating lots of potential.
- Standards (a) facilitate interoperability; (b) allow other organisations to expend less effort; (c) improve portability.
- The CEF eArchiving interoperability provides the potential to share infrastructures and therefore lower costs, and makes possible to exchange data between different *digital safes*.
- CEF eArchiving allows for evidence to improve authenticity.
- CEF eArchiving helps to avoid vendor lock-in. We can potentially get better quality, cheaper and a larger variety of systems.
- The European Archives Group has established a workgroup to deal with the topic of “Archiving by Design” to bring archival issues (data and metadata quality control, transfer, preservation, long-term use) closer to the systems where data is created. The group is led by the National Archives of Norway and the Netherlands.

c) Cost savings, best practices, support

- Cost benefits, saving money by having servers rather than buildings full of papers.
- Potential of having CEF eArchiving as an EC promoted digital preservation system (or even an EC-maintained digital preservation service), which is cheap and possible to be used with limited skills/staff.
- It is easier to compare and develop best practice. However, we need to have good training to achieve this.
- The support, coordination and best practices are very useful.
- Promoting and supporting O/S, good, robust systems.
- CEF eArchiving can be used to gather and compare national or institutional best-practices with CEF eArchiving and each-other.
- CEF eArchiving can be used to be more effective in awareness raising and lobbying (i.e. individual countries or institutions do not have sufficient resources for that).

d) Communication

- There is a lack of general awareness, which is a challenge for the Public Sector in general.
- Companies delivering digital services similar to those of the BBs see their jobs threatened, so try to block their implementation (eSignature etc.); therefore the advantages of adopting the CEF eArchiving services should be promoted.
- In public organisations, there are challenges for decision makers.
- It is difficult to prepare a good tender and specification, to identify the needs and communicate these. Standards are needed to better describe which services an eArchiving system should offer, including a common service description, and a certification description.

- The requirements for a CEF eArchiving system are not well described yet.

e) Funding, resourcing, training

- All groups reported that there is a lack of sufficient funding, staffing, training and expertise.
- Training for archivists, who are not computer scientists. Some archivists have theoretical knowledge (in digital preservation) but do not have the necessary practical skills, so we need more practical, good education to train digital archivists.
- Training is essential to have in different EU languages.
- Archivists still have a “paper mindset” so this needs to change to also encompass the digital mindset.

f) Technical issues. Storage

- Lack of awareness of technical issues.
- CEF eArchiving is currently about the processes and formats in digital preservation. A full solution (needed by institutions) has also to look at storage, replication, security etc.
- Migrating content to a digital archives:
 - a. has to be secure and logged (in order to be able to say that the data is still trustworthy);
 - b. has to come with negotiating the reuse/ownership rights with all the initial owners and the digital repository.
- Varieties across EU for e.g. retention periods. For example, Portugal had a 5-year period for paper invoices, now it is 12 years for digital data.

Conclusions and lessons learned

The European Commission hereby thanks all the participants, in person and remote, for their participation and feedback, which was overwhelmingly positive. The main conclusion is the fact that there is an interest in digital preservation.

Coming back to the objectives laid out before the workshop:

1. Raise awareness of digital preservation challenges, and of the eArchiving building block.

The workshop provided this awareness indeed, both at the business (digital preservation) and the policy (CEF eArchiving) levels. The workshop was the most followed event so far for a CEF building block, and despite some physical space worries, the information reached a lot of people interested in the digital archiving challenges.

There were some shortcomings, though: the EC was requested to provide better descriptions of the CEF eArchiving and the building blocks offer, including necessary requirements, transparency on technical issues, a certification process, rules for tender specifications, and in general, what the advantages for taking CEF eArchiving up are.

2. Reach beyond the National Archives, both to other domains and to other levels of public administration.

The representation of EU member states was extensive, not only geographically but also in terms of different domains: National and regional archives were present, as well as other

public institutions, standardisation organisms, cross-border organisations, software companies, small and medium enterprises, etc.

After some sessions, the attendants asked whether the EU institutions would take over CEF eArchiving, if we were leading by example. Although not extensively and certainly not mandatory, CEF eArchiving is already being used in official and other publications, and has been successfully tested in DG TAXUS. Also the European Commission's Historical Archives were reported willing to analyse the potential of CEF eArchiving.

The participants reported some other issues that prevented them to take CEF eArchiving up. Other than the lack of sufficient information mentioned above, there is also a lack of training for archivists who are not IT experts, and in the different EU languages.

3. *Collect feedback to improve the policy on eArchiving.*

The lack of regulation is mentioned as a barrier to implement cross-border digital archiving services. Since these competences are on the member states, CEF eArchiving is the best answer that the EC may offer at this point. On the other side, eIDAS has EU-level legislation, so there could be a possibility to align their policies.

Furthermore, the lack of funding is to be faced soon; the EC plans to create generic services for CEF eArchiving, that is, the support for implementing the building block.

All the feedback gathered in the sessions' questions and answers, and from the breakout sessions, will help us to improve the digital archiving policy in the CEF eArchiving building block.