eArchiving

**Simplify long-term access to information with the CEF
eArchiving building block: a technical solution for developers
that need to migrate, preserve and reuse data.**

The eArchiving Building Block can provide long-term
information assurance. It provides the specifications,
reference software, training and service desk support
for digital archiving, including digital preservation.
This benefits both the design and implementation of
repositories and enables business systems to send
data to those repositories.

Information Package specifications are the foundation
of eArchiving. These describe platform-independent
formats to structure information assets as bulk
data and metadata that remains authentic and
understandable over time. They are thus ideal for:

- migrating information assets between generations
  of technology;

- transferring information assets to repositories;

- long-term managing of information assets in
  repositories;

- permanent access to information assets as
  archived data.

eArchiving also provides sample software for several
business scenarios, which any organisation can
reuse to develop its own institutional archiving and
preservation ecosystem, or to deliver content from
its business systems to external repositories. Using
eArchiving guidelines guarantees that information will
be available and reusable for as long as it is required.

Who can use it?

- Data creators in public and private institutions

- Data repositories such as archival institutions

- Services and solution providers

How it works

eArchiving is based on Information
Package specifications that provide
interoperability across borders, types
of institution and user communities.
These are supported by the Common
Specification for Information
Packages (CS IP), expressed with the
Metadata Encoding & Transmission
Standard (METS), with specialisations
for:

- Submission Information Package
  (SIP), for sending the data to be
  preserved to a repository;

- Archival Information Package (AIP),
  for storing and preserving the data
  in a repository;

- Dissemination Information Package
  (DIP), for accessing data stored in a
  repository.

Common for all Information Packages

CS IP ①

E-ARK SIP ① — E-ARK AIP ① — E-ARK DIP ①

CITS ② — CITS ② — CITS ②
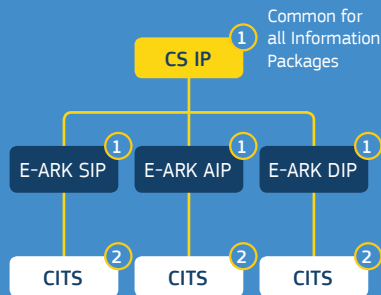
① Expressed with Metadata Encoding and
Transmission Standard (METS)

② Content Information Type Specification (CITS)
e.g. for ERMS, databases, geodata, etc.

Get started with help from CEF

The Connecting Europe Facility (CEF) within the European Commission offers a range of services to support public and private sector entities that want to implement eArchiving, at no cost.

- **Software:** Components to demonstrate how eArchiving technical specifications can be applied in the archival workflow

- **Support:** Our service desk is available to help your team during analysis, design, implementation and post implementation of your project

- **Stakeholder Management:** Access a range of knowledge articles, events, and webinars to support a successful rollout to your stakeholders

- **Developers Community:** Connect with other developers and share your knowledge and experience in building applications

How we comply with European Standards

Developed and tested in the E-ARK project, the eArchiving specifications are based on international standards for describing, transmitting and preserving digital data.

The core standard for eArchiving is the Reference Model for an Open Archival System (OAIS): ISO 14721:2012. Information Packages are written using Metadata Encoding & Transmission Standard (METS) and use other standards for describing the different types of content.

6 reasons to use eArchiving

1 Maturity: re-uses the accumulated knowledge of other organisations facing the same challenges

2 Standardisation: enables information assets to be transmitted, preserved and re-used across borders as well as time

3 Transparency: ensures a high level of confidence among all participants in the information value chain

4 Flexibility: supports scaling of digital archival systems from small to very large

5 Efficiency: accelerates the delivery time of a working digital archive, while controlling costs

6 Risk management: reduces risks in information assurance

Find out more:
ec.europa.eu/cefdigital/eArchiving

The CEF Building blocks and the role of the eArchiving within it

1. The Connected Europe Facility (CEF) and its building blocks

Connected Europe Facility (CEF) is a key EU funding instrument to promote growth, jobs and competitiveness through targeted infrastructure investment at European level. It supports basic and re-usable digital services, known as building blocks, as well as digital services that are more complex. The building blocks, now available at the CEF Digital Single Web Portal (*https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/Building+Blocks*), can be combined with each other and integrated with the more complex services, contributing to the development of a Digital Single Market and a fully connected Europe.

For example by using the **eSignature** public administrations and businesses are able to create and verify electronic signatures across the member states, since they are valid and structured in EU interoperable formats. After we signed the document, we can use the **eDelivery** building block using the network of nodes for a secure and reliable exchange of documents and data both across borders and sectors. In addition, in the end when we need to archive this documents or data we will be able to use a unified interoperable approach provided by the **eArchiving** building block, using its guidelines and tools to ensure the long-term preservation of different types of content.

2. The eArchiving Building block

This is a new building block, whose development started just last year. The aim of eArchiving is to provide the core specifications, software, training and knowledge to help data creators, software developers and digital archives tackle the challenge of short, medium and long-term data management and reuse in a sustainable, authentic, cost-efficient, manageable and interoperable way. The core of eArchiving is formed by Information Package specifications, which describe a common format for storing bulk data. Next to the specifications, eArchiving offers a set of sample software to demonstrate the format in different scenarios and business environments, and consultancy concerning long-term digital preservation risks and their mitigation.

2.1. The Information Package

The Information Package specifications describe a common format for storing bulk data and metadata in a platform-independent, authentic and long-term understandable way. The specifications are ideal for migrating long-term valuable data between generations of information systems, transferring data to dedicated long-term repositories (i.e. digital

archives), or preserving and reusing data over extended (and shorter) periods of time and generations of software systems.
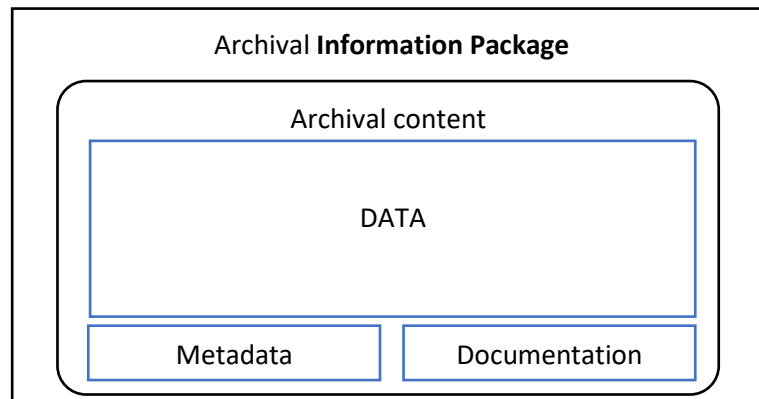
Figure 1 – Structure of the Archival package

On the high level, the Information package contains basic metadata for archival processing and the Archival content is structured and defined by its type. So far the following data types have been specified: Electronic Document Management Systems (ERMS), Relational Databases Management Systems, Geospatial Information (referred as geodata), and other unstructured content. Other content types and subtypes will be added in the future.

3. Geospatial information (GI) archiving within the eArchiving building block

We strive to develop or suggest guidelines for GI archiving as well as tools for exporting, packaging, describing and validating information packages containing GI. We are also developing training for archivists and GI producers in order to promote the use of guidelines and tools.

3.1. The Common specification for the Geodata content type

The Geodata content type refers to digital geospatial data, used as records in GIS systems. The aim of this specification is to specify all the information needed, so that the data can be used in a similar fashion in any future GIS systems. And since we are not archiving applications as such, we need to at least need to store the data, and document the metadata, logic and representation information.

The details can be reviewed on this link:

https://github.com/DILCISBoard/E-ARK-Geodata/blob/master/Specification/CSGeo_v1_1_DRAFT.pdf

4. How can INSPIRE and eArchiving benefit from each other?

The state of geodata content type specification is currently in its basic form and is covering the simplest uses and it is already partially based on the INSPIRE guidelines for metadata.

We are striving to provide guidelines to match the current types and future of GIS systems, that are based on existing standards and guidelines, which are already accepted and in use within the GI community.

So this is where we see the possibilities of cooperation with the existing body of knowledge, being developed within the INSPIRE group.

So we are interested in discovering possible synergies between our two areas of work and how we can benefit from each other (INSPIRE and EARK4ALL) and not duplicate the work where possible.

4.1. We are especially interested in:

- Inroporating existing and future guidelines/directives for "geo-sources" (datasets, services...) into our new guidelines for archiving

- maintenance of our existing guidelines according to changes within the INSPIRE directive

- existing and upcoming tools, that you provide (data and metadata creation and validation, metadata management...)

4.2. How eArchiving can help the INSPIRE community?

In the future, the archives will probably become a vast source of older Geospatial information, that will be used for discovering patterns for future decision-making. So the GI community would benefit if:

- the current GI providers and archives will serve its user in an interoperable way,
- If existing and future systems would add the archiving into the data production process, there by ensuring completeness and lowering future costs.