Standalone training from the eArchiving Initiative on the

E-ARK Specifications

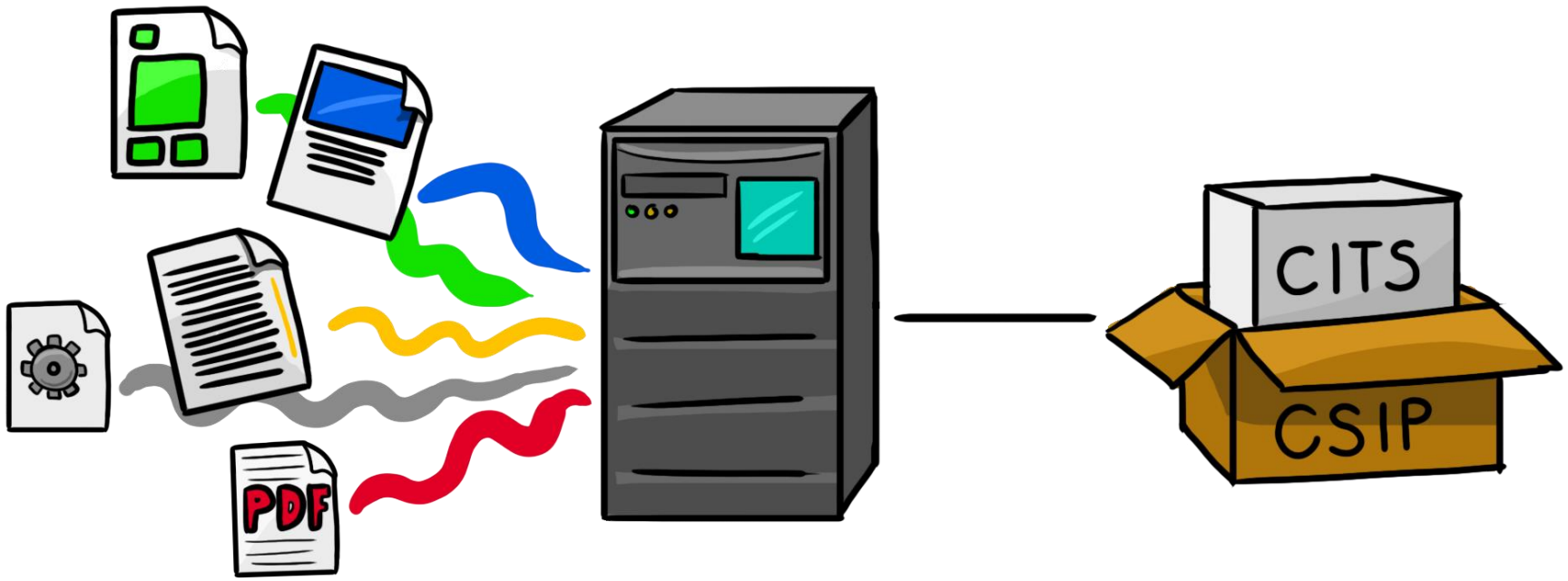
Lesson 3

Introduction to the E-ARK Content Type Specifications

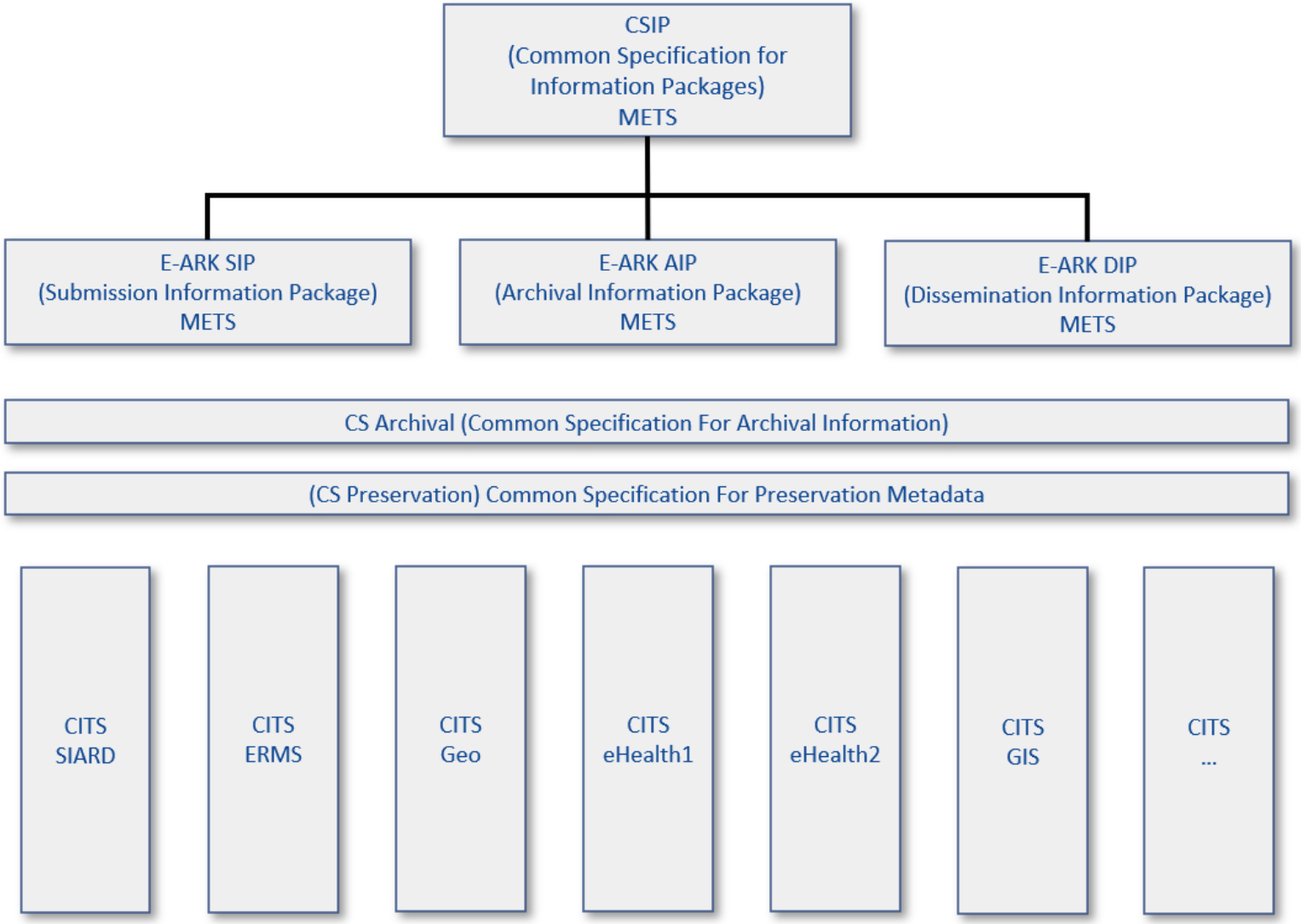
Lesson Structure

- Lesson 1 – Introduction to the E-ARK Specifications
- Lesson 2 – Common Specification and the Information Package Specifications
- **Lesson 3 – Introduction to the E-ARK Content Type Specifications**
- Lesson 4 - 9 – Introduction to Specific Content Types [To be added later]

Content Information Type Specifications (CITS)



E-ARK Content Type Specifications



CS Archival Information

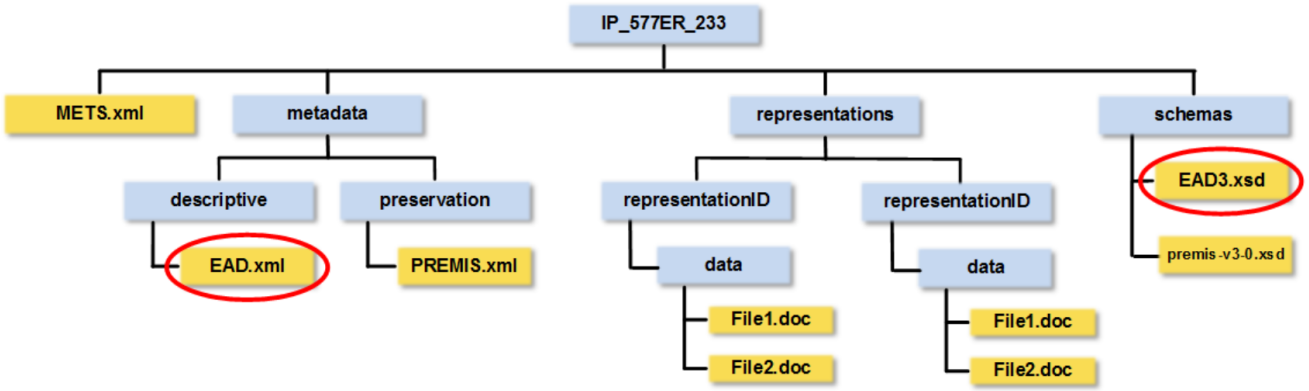
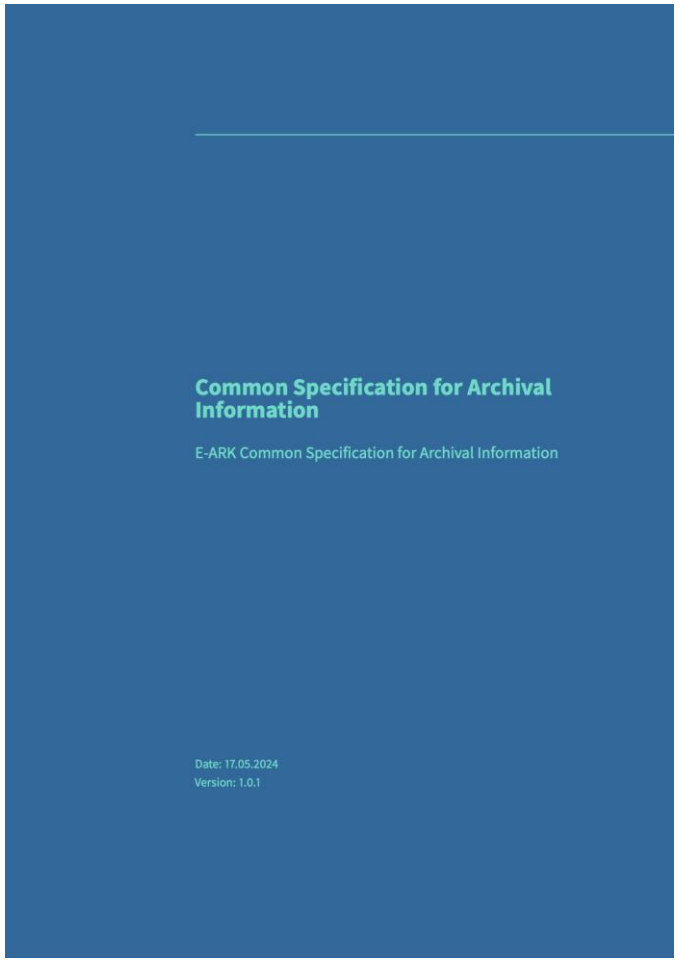


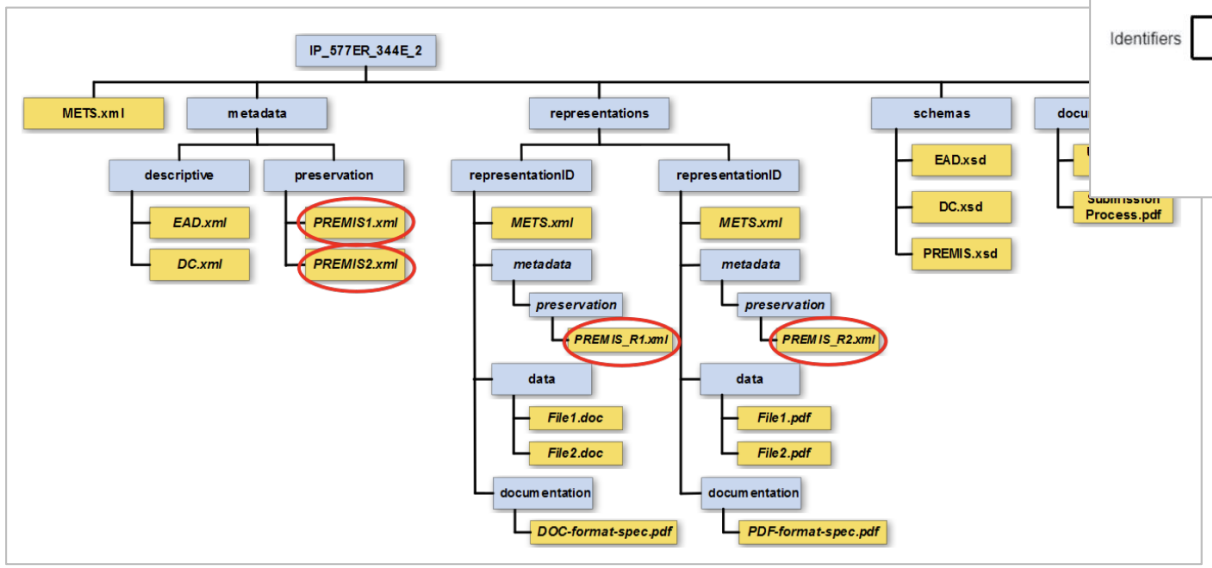
Figure 1: Example of placement of EAD in the IP.

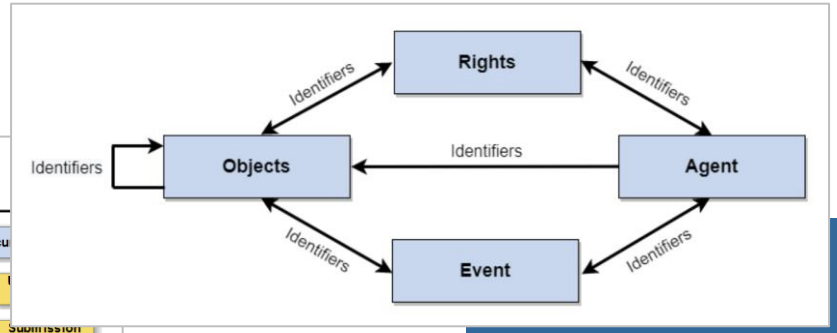


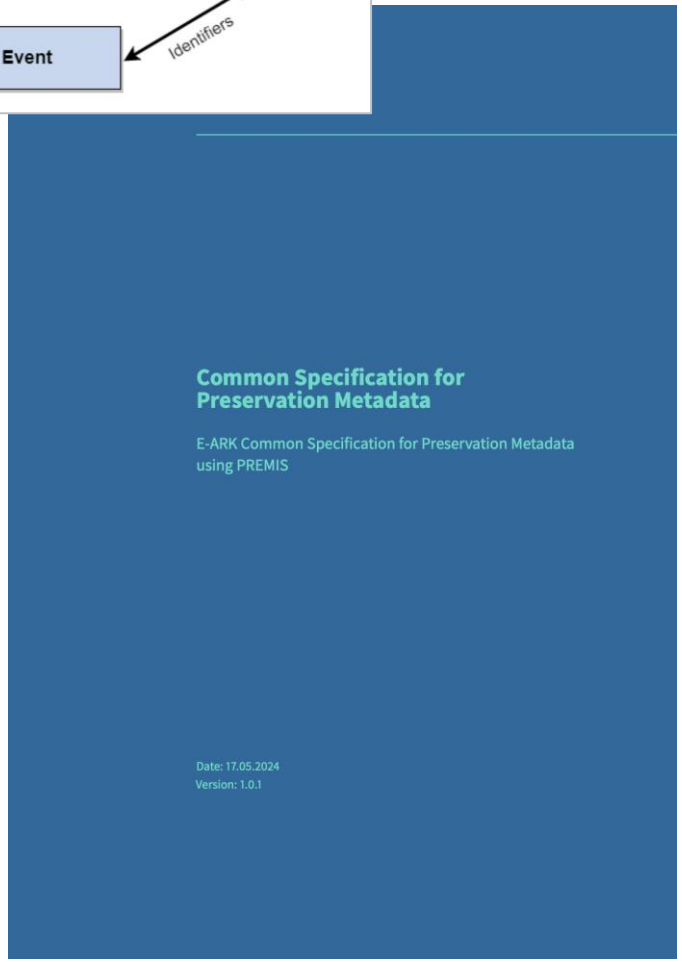
The purpose of this document is to describe the Common Specification (CS) for Archival Information. The term ‘archival information’ includes documents such as the finding aid and the creator information.

The specification is used for both the transfer to archives and for information exchange between different systems that require access to the Archival Information.

CS Preservation Metadata







The purpose of this specification is to maintain the authenticity and integrity of digital objects in digital storage based on the use of the de-facto standard Preservation Metadata: Implementation Strategies (PREMIS) (http://www.loc.gov/standards/premis/).

The document is a simple and basic implementation of PREMIS, and a more elaborate implementation can and should be made in your preservation platform. The specification is designed to be used for the transfer to archives and information exchange between different systems requiring preservation metadata for the digital objects of the transfer.

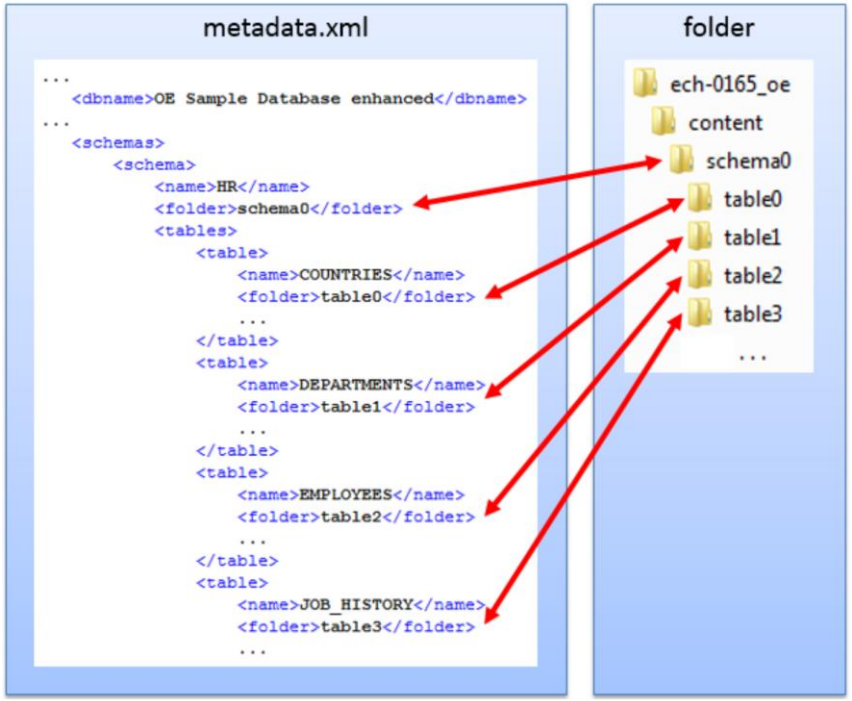
SIARD

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

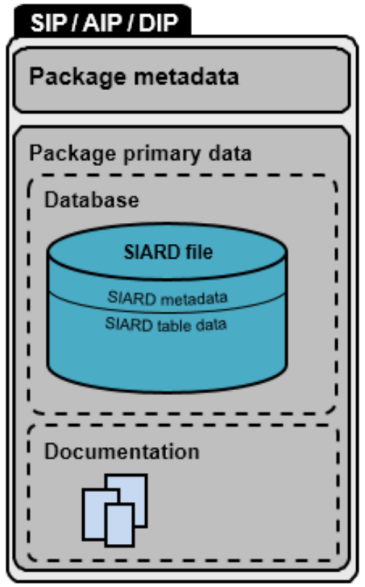
DILCIS BOARD

The term “SIARD” stands for Software Independent Archival of Relational Databases. It is an open file format for the long-term archiving of relational databases in the form of text data based on XML that are packaged in a container file (SIARD archive).

If the structure and content of a relational database are translated into the SIARD format, it will subsequently be possible to access and exchange the data in the database at any time, even when the original database software is no longer available or can no longer be run.



1.2.3 Distinctions



It should be noted that the SIARD format is only the long-term storage format for a specific type of digital documents (relational databases) and is therefore designed entirely independently of package structures such as the SIP (Submission Information Package), AIP (Archival Information Package) and DIP (Dissemination Information Package) in the OAIS model[3].

It is assumed that a database in SIARD format is archived as part of such an information package together with other documents (externalized large object files, translation maps for external file names, database documentation, business documents relevant to the understanding of the database, etc.).

Just as an XML-based Word or e-mail file contains an internal file structure consisting of metadata, primary data and various auxiliary data, an archived relational database in SIARD format contains its own metadata describing the document more precisely in addition to the actual table data – regardless of the metadata catalogue that an archive records in its OAIS packages.

European Commission

CITS SIARD

2.2 Layered data model

This section introduces the data model structure, which is based on a layered approach for information package definitions (Figure 1). The Common Specification for Information Packages (CSIP) forms the outermost layer. The general SIP, AIP and DIP specifications add submission, archiving and dissemination information to the CSIP specification. The third layer of the model represents specific content information type specifications, such as this CITS SIARD specification. Additional layers for business-specific specifications and local variant implementations of any specification can be added.

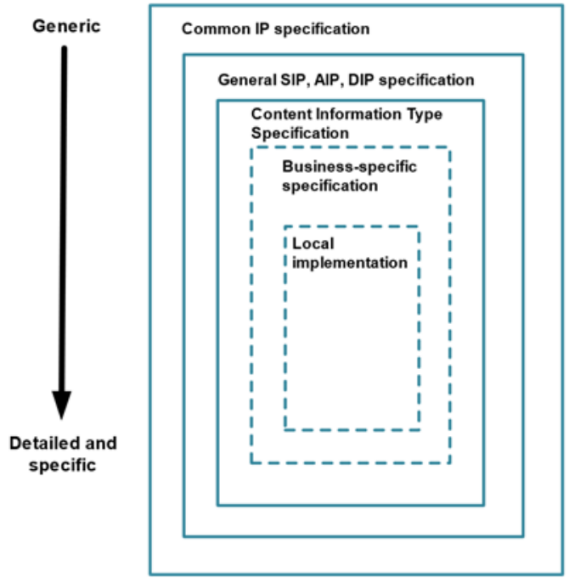
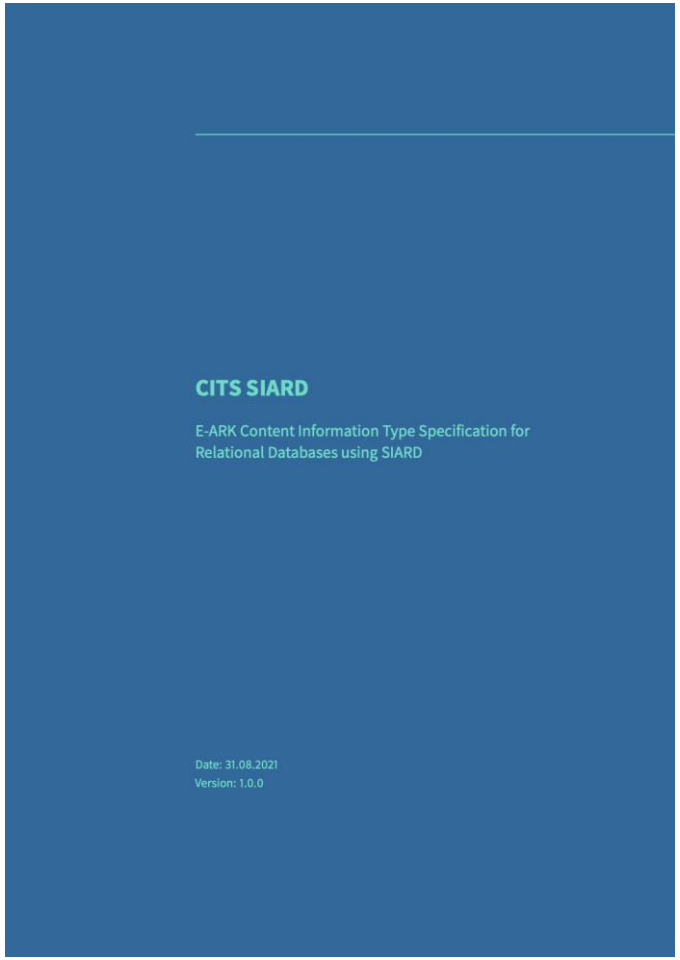


Figure 1: Data Model Structure

Every level in the data model structure inherits metadata entities and elements from the higher levels.

Folder Structure of Northwind Sample Database

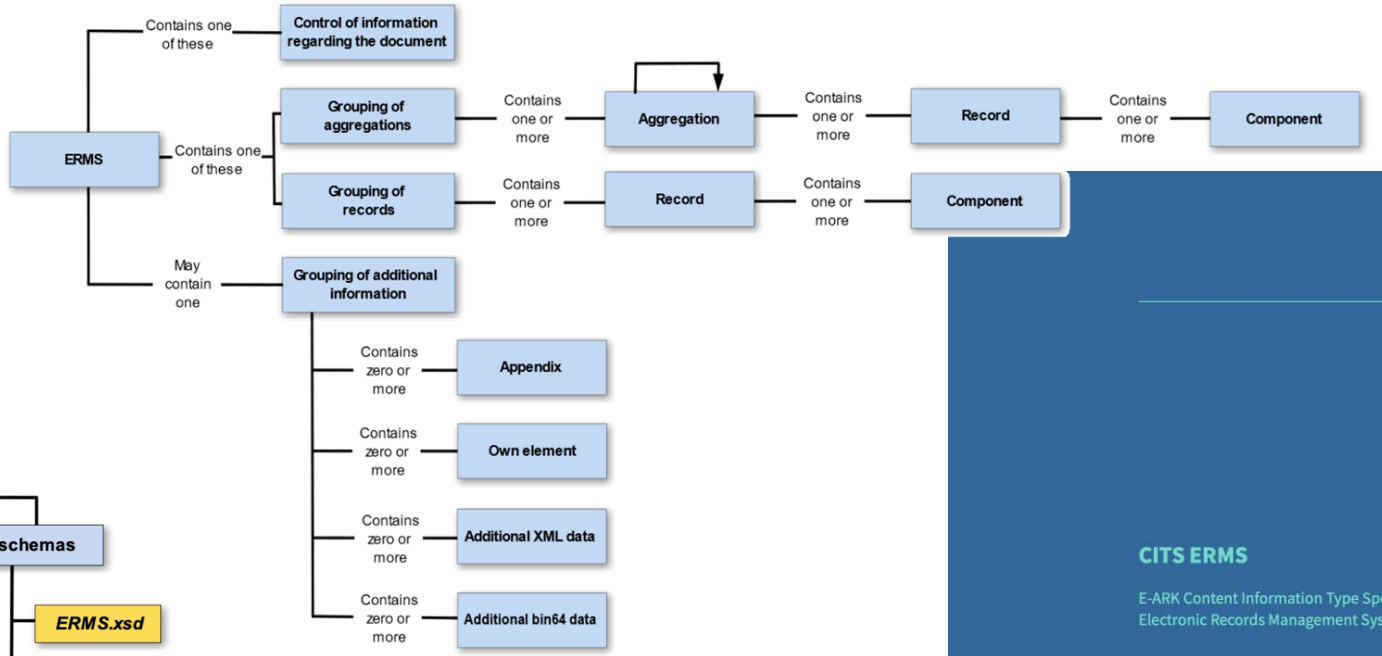
- IP.AVID.RA.18006
  - representations
    - AVID.SA.18006.rep0
      - documentation
      - metadata
        - preservation
        - other
        - descriptive
      - data
        - Northwind_lobs
          - s0_t2_c4
            - seg0
              - t2_c4_r1.bin
              - t2_c4_r2.bin
              - t2_c4_r3.bin
              - t2_c4_r4.bin
              - t2_c4_r5.bin
              - t2_c4_r6.bin
              - t2_c4_r7.bin
              - t2_c4_r8.bin
          - s0_t2_c8
            - seg0
              - t2_c8_r3.bin
          - s0_t11_c6
            - seg0
              - t11_c6_r7.bin
        - northwind.siard
      - METS.xml
  - documentation
    - database_diagrams
      - Original_Northwind ER diagram.png
      - Archived_Northwind ER diagram.png
    - schemas
      - ead3.xsd
      - premis.xsd
      - METS_xlink.xsd
      - xlink.xsd
      - mets.xsd
    - metadata
      - other
      - preservation
        - PREMIS.xml
      - descriptive
        - EAD.xml
  - METS.xml



European Commission

CITS ERMS

The ERMS XML-schema contains the high-level entities seen in Figure 3. As shown, it is possible to export just one record or to export an aggregation.



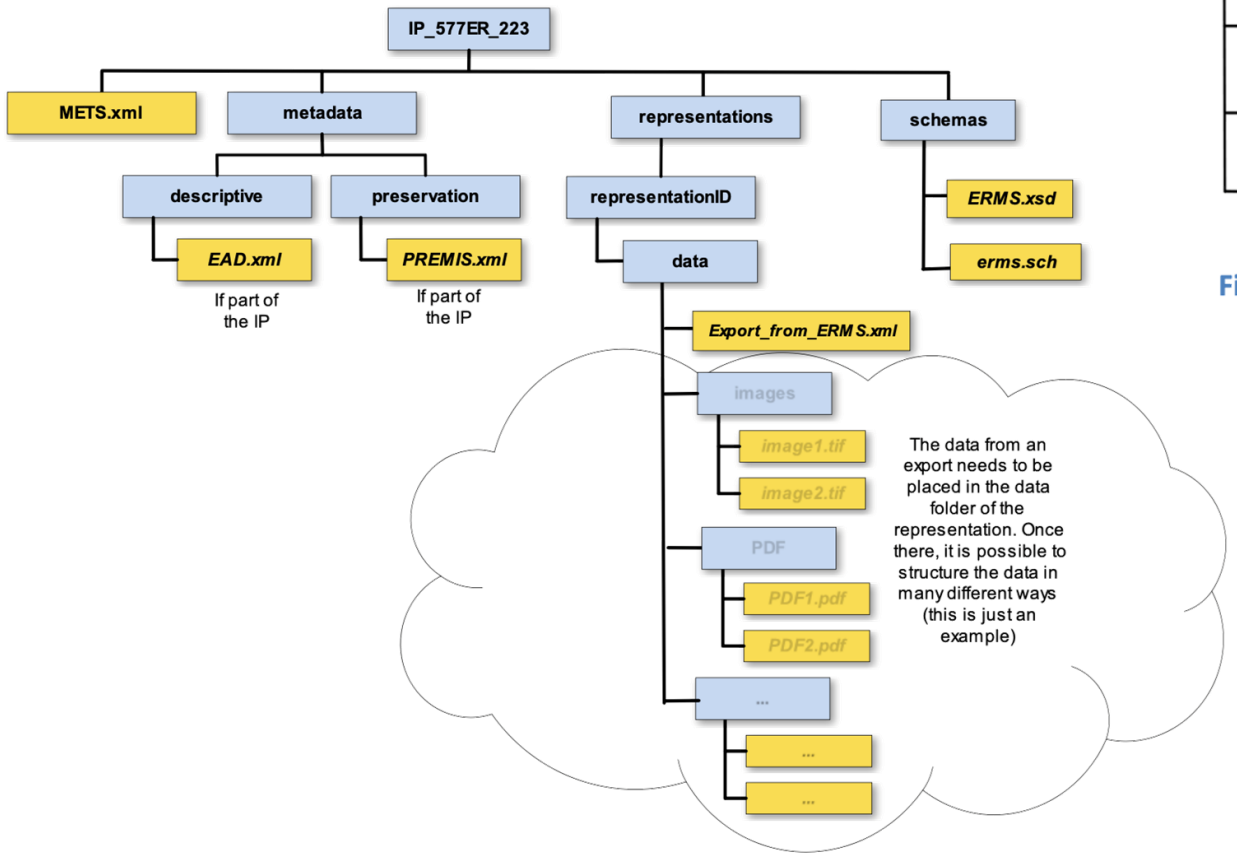
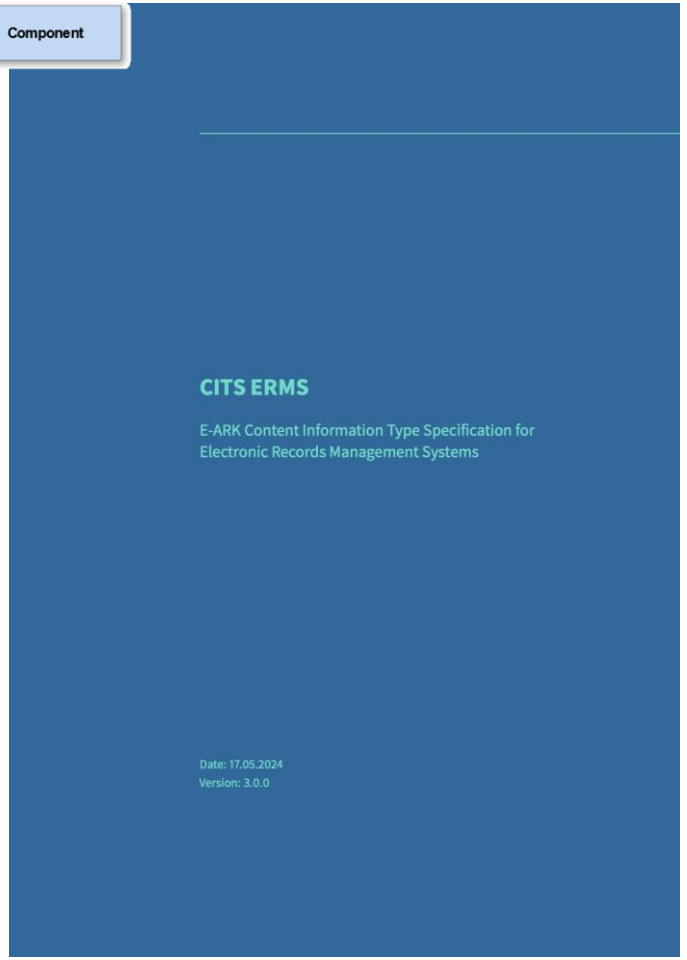
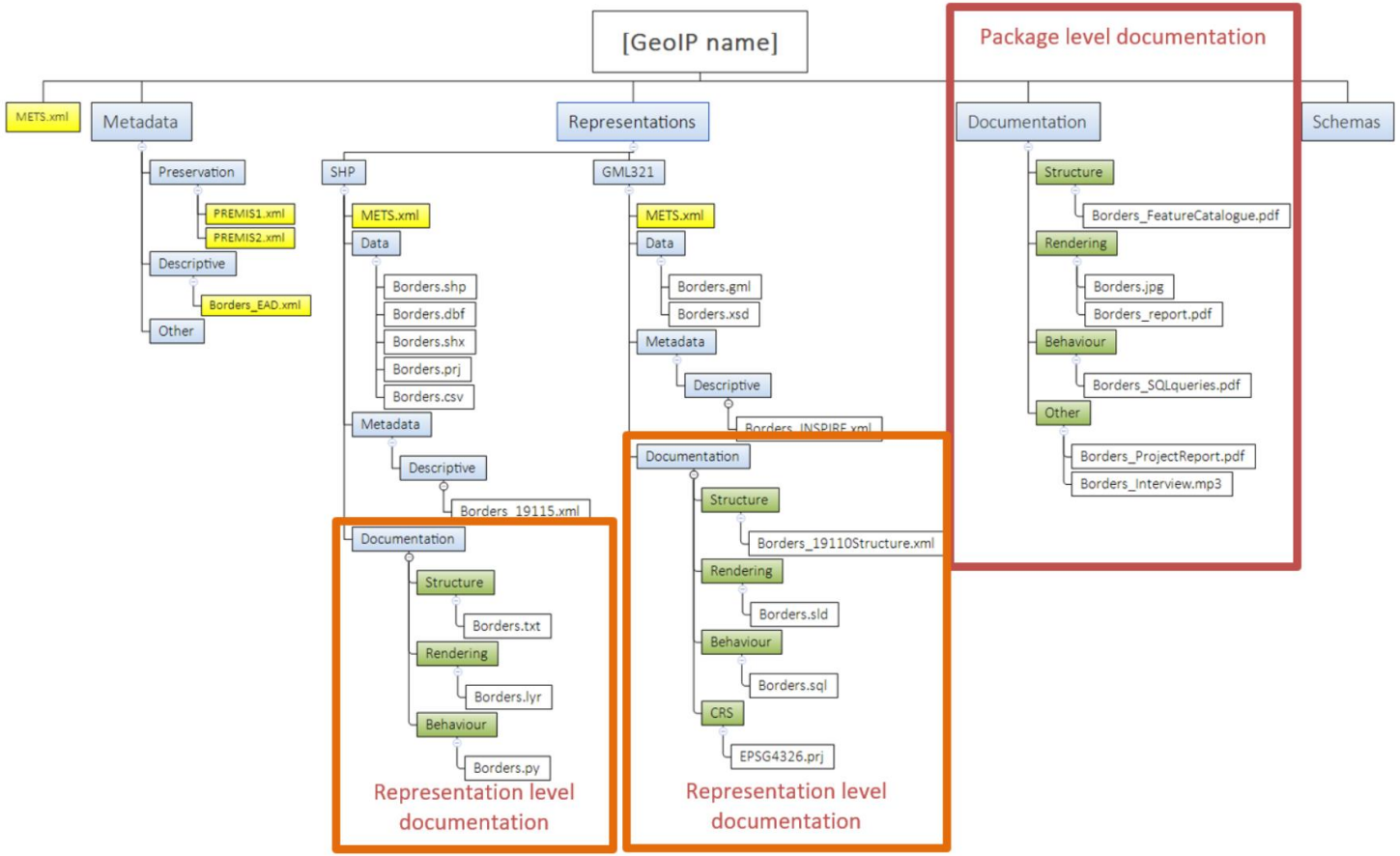


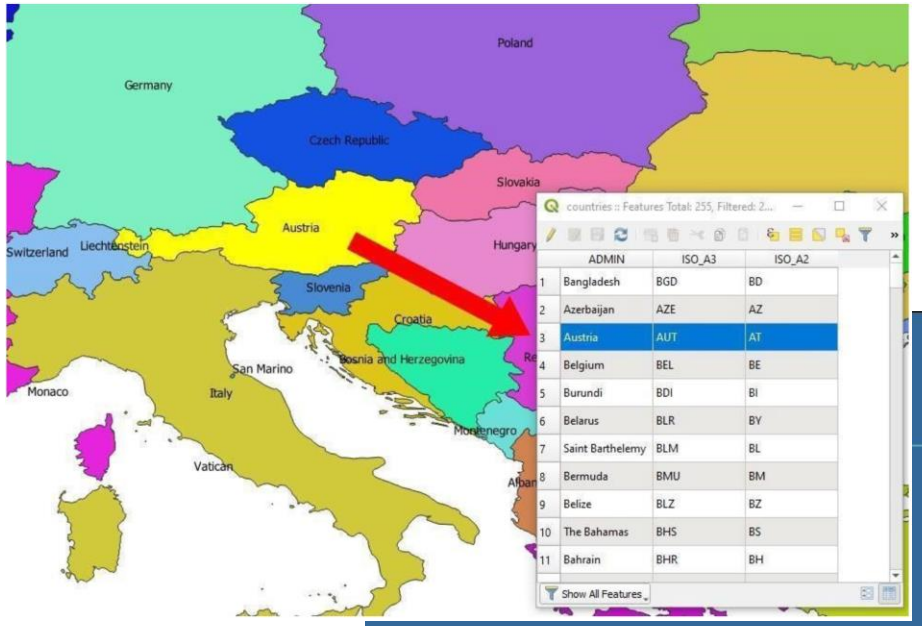
Figure 3: Components of the ERMS XML-format



European Commission

CITS Geodata





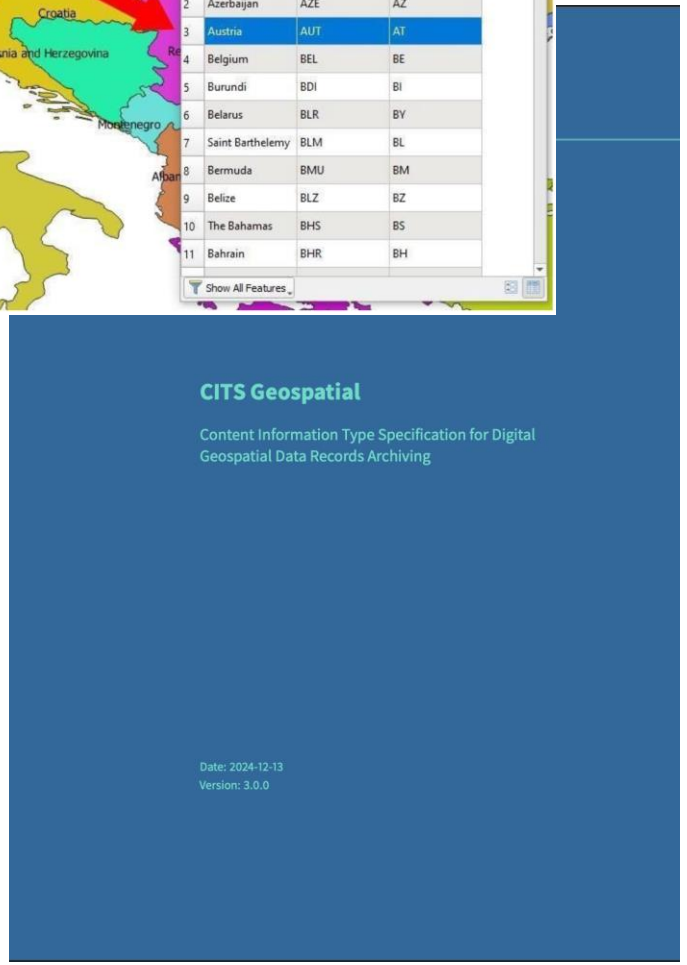


Figure 9 - Locations of Documentation folders in archival package.

CITS eHealth1

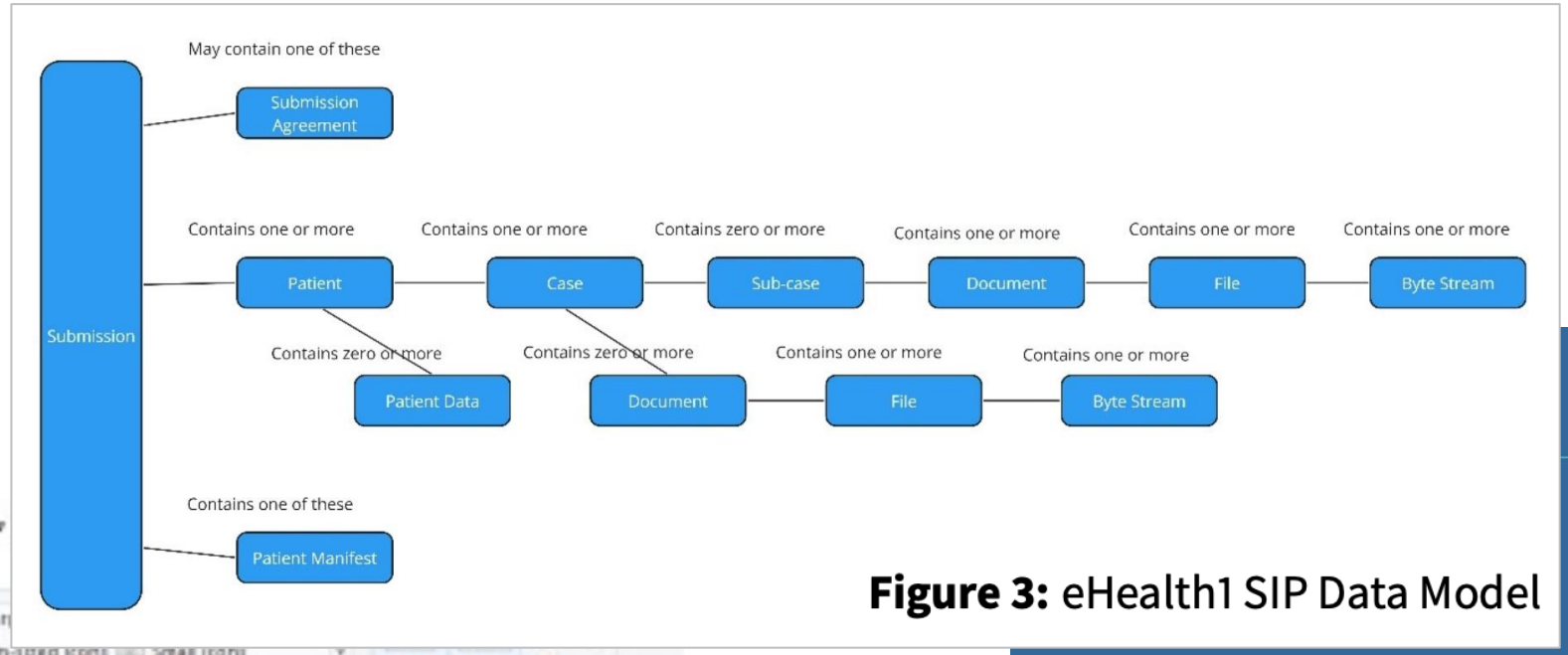
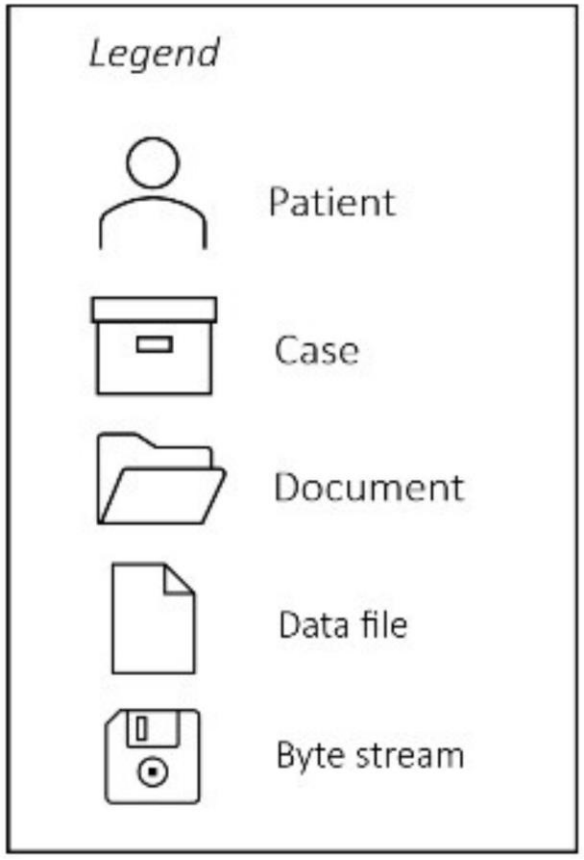
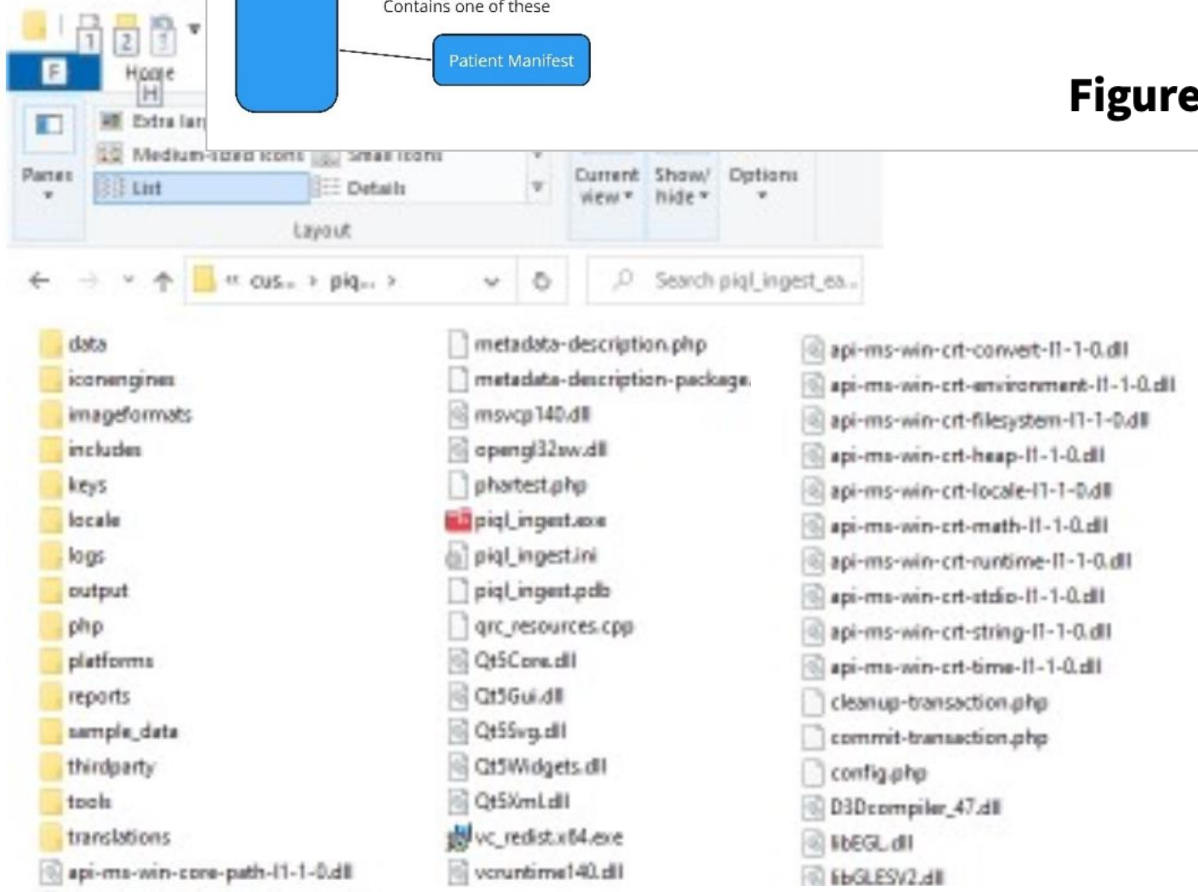
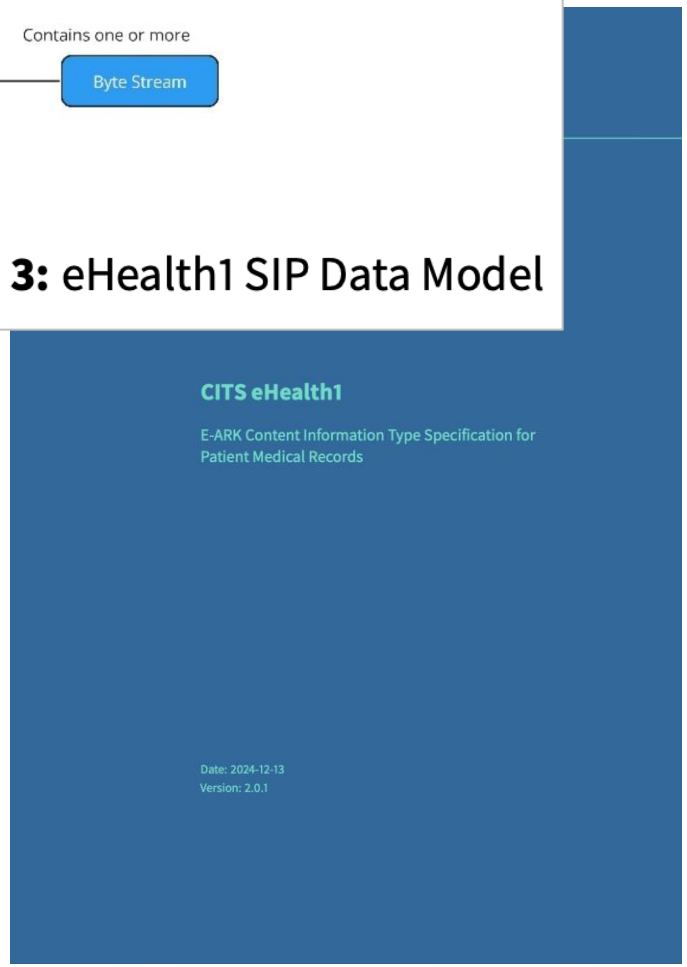


Figure 3: eHealth1 SIP Data Model







European Commission

CITS eHealth2

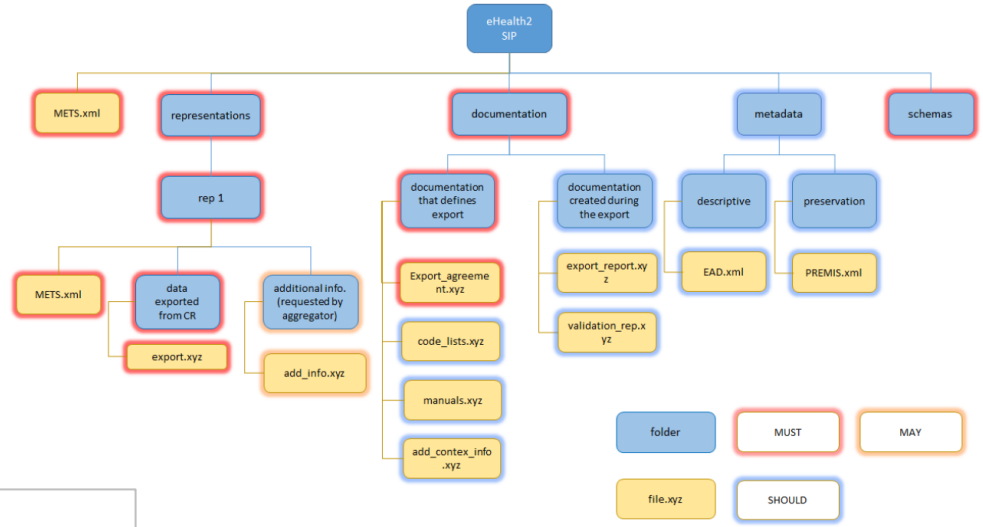
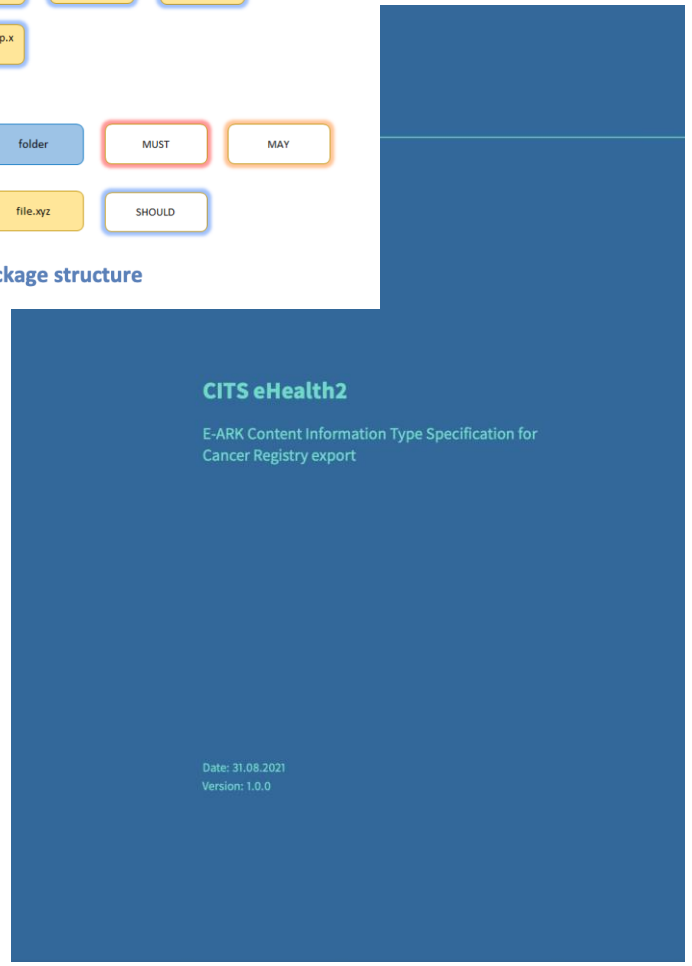


Figure 1: eHealth2 archival information package structure



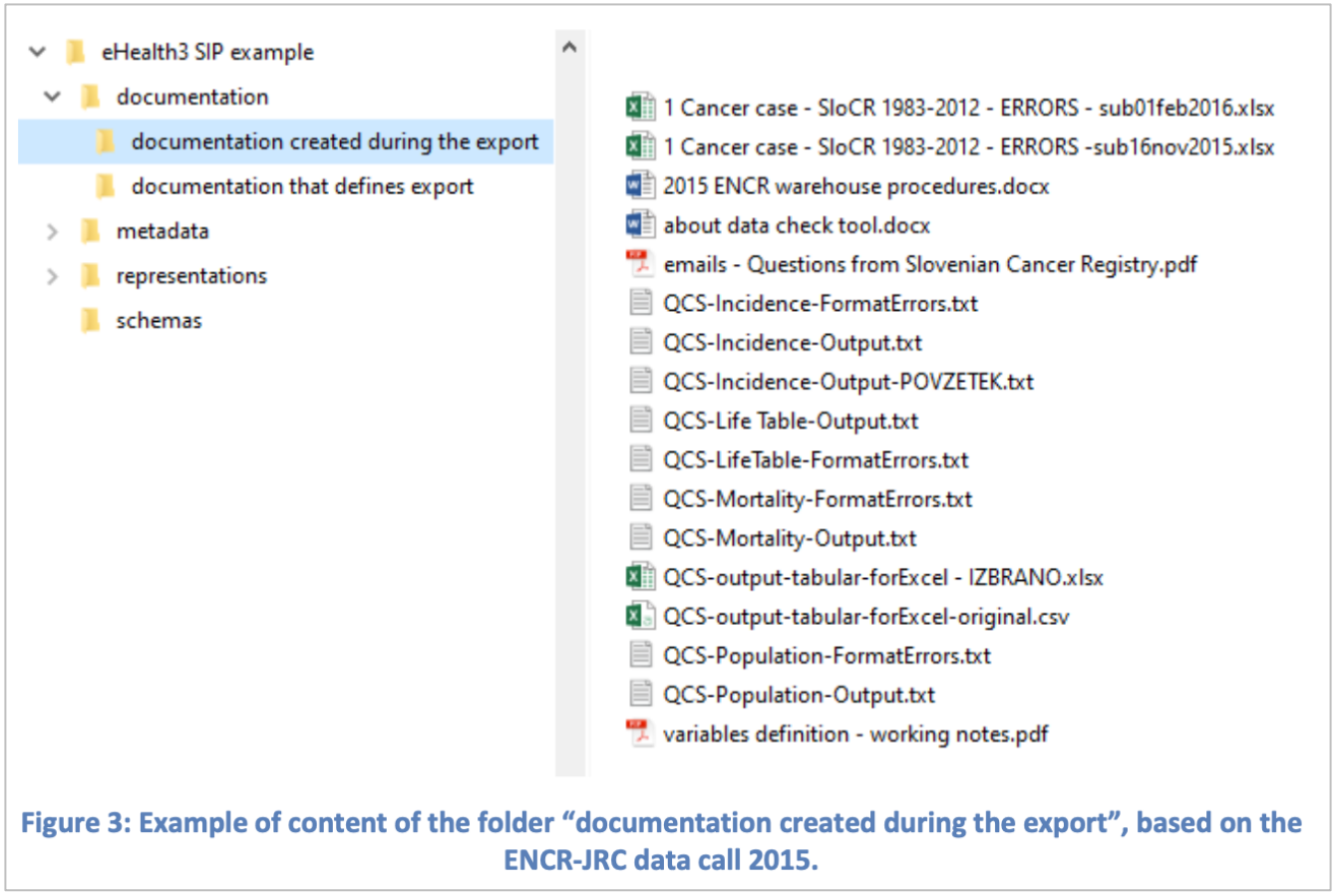


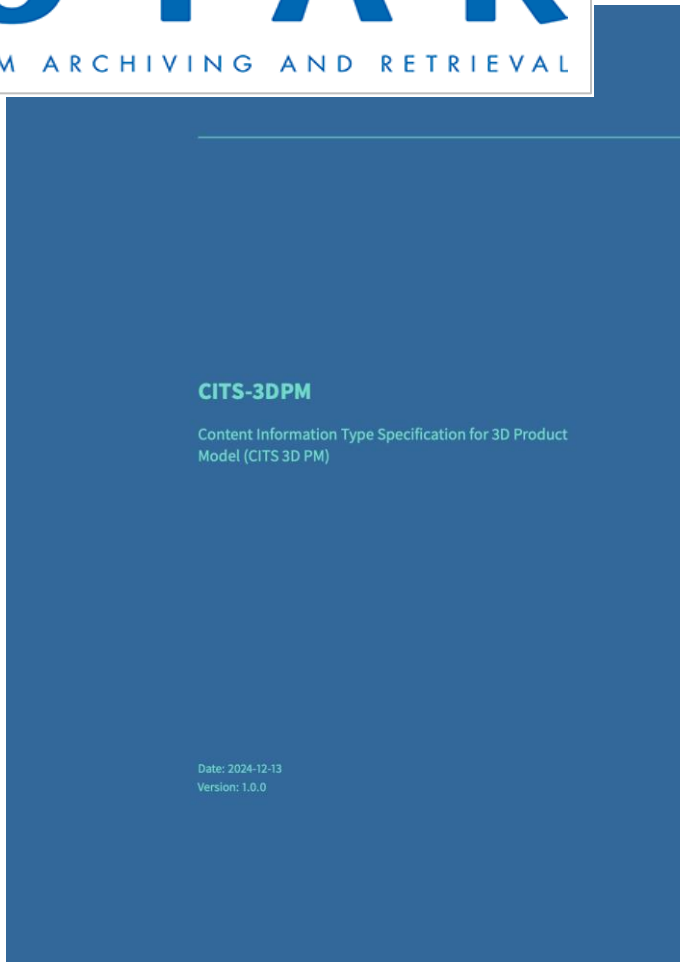
Figure 3: Example of content of the folder “documentation created during the export”, based on the ENCR-JRC data call 2015.

European Commission

CITS 3DPM

LOTAR

LONG TERM ARCHIVING AND RETRIEVAL



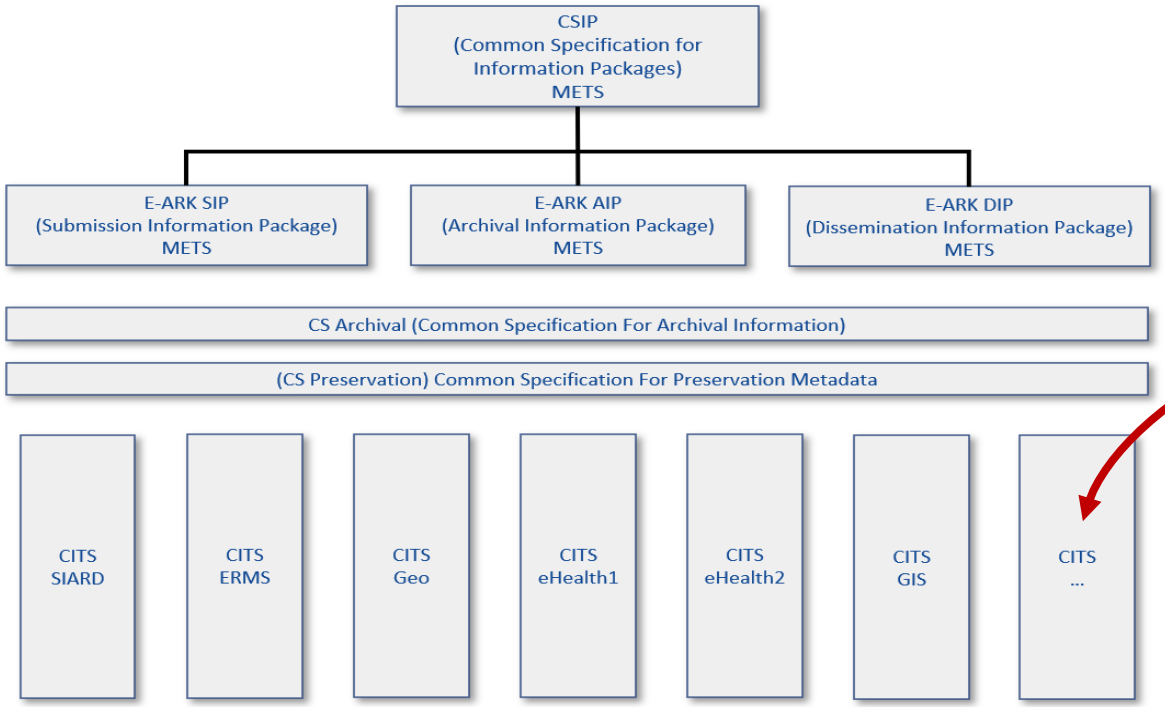
2.3 Use Cases for Archiving of Product Model Data

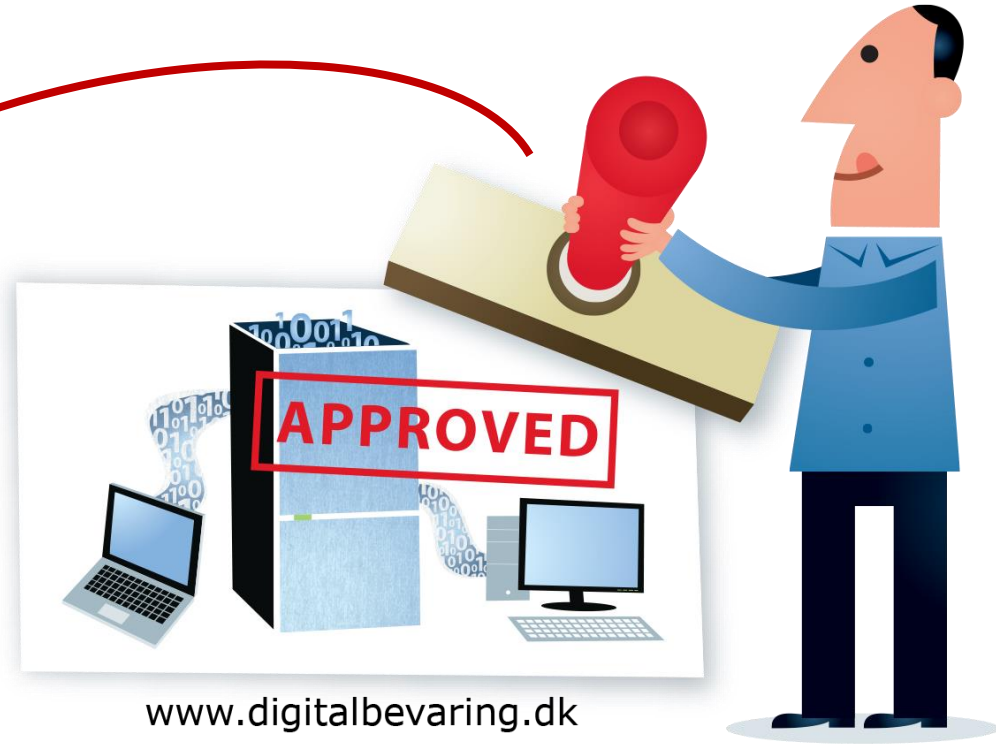
Further detail of the use cases for long-term archiving of data in general and the rationales given in LOTAR specifically for Product Model data are given in the accompanying guideline. In summary, the use cases for archiving of Product Model data are determined to be:

- To enable the submission of 3D Product Model data from engineering departments in an organisation to a centralised or distributed archive, in a common format;
- To store archival 3D Product Model data in a manner that will allow consolidation of archives intra-organisationally or with sources added through mergers or acquisitions;
- To allow dissemination of Product Model archival data within the organisation or to external regulatory bodies preserving both the integrity of the data objects and the information packages.

European Commission

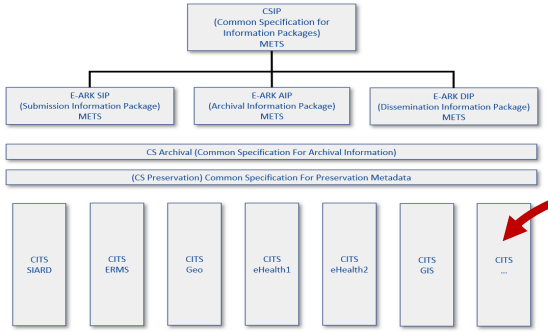
New CITS





www.digitalbevaring.dk

New CITS

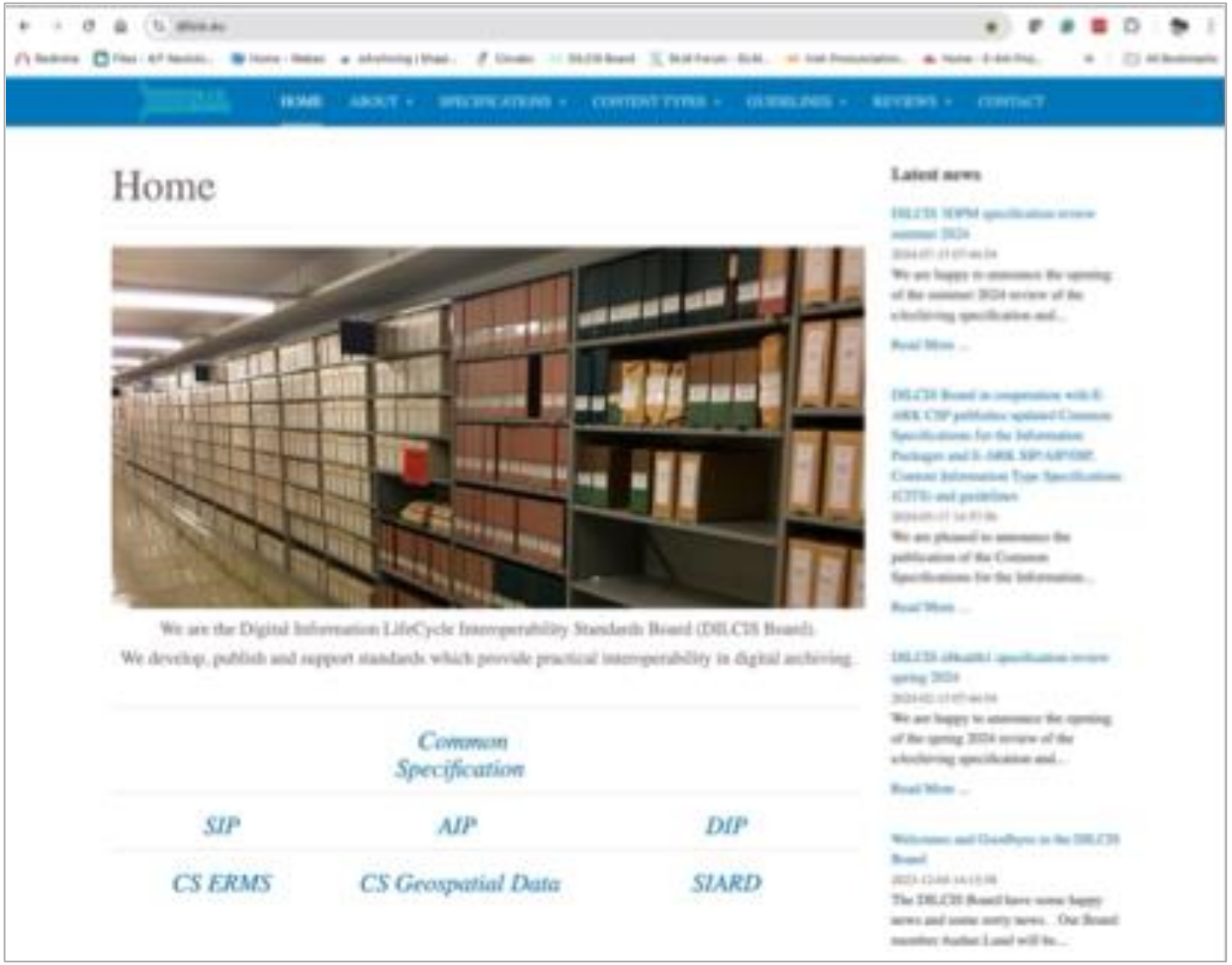


www.digitalbevaring.dk

It is hoped that many Content Information Type Specifications will be developed with the wider community to create new specifications for domains of interest to them. The DILCIS Board aims to work with the community to maintain a list of available specifications following some simple principles:

- The DILCIS Board is responsible for establishing reasonable guidelines and quality requirements for new Content Information Type Specifications, and publishing these on the Board website;
- The Board has the responsibility and mandate to manage a registry of available Content Information Type Specifications which meet the guidelines and quality requirements;
- The Board does NOT take ownership of or have responsibility for maintaining and sustaining any Content Information Type Specifications;
- There shall be no limitations to who is allowed to propose additional Content Information Type Specifications; and
- To ensure the quality of available specifications, the Board validates each proposed specification against the guidelines and quality requirements mentioned above. The validation shall be carried out free of charge and within a reasonable time-frame.

DILCIS Board



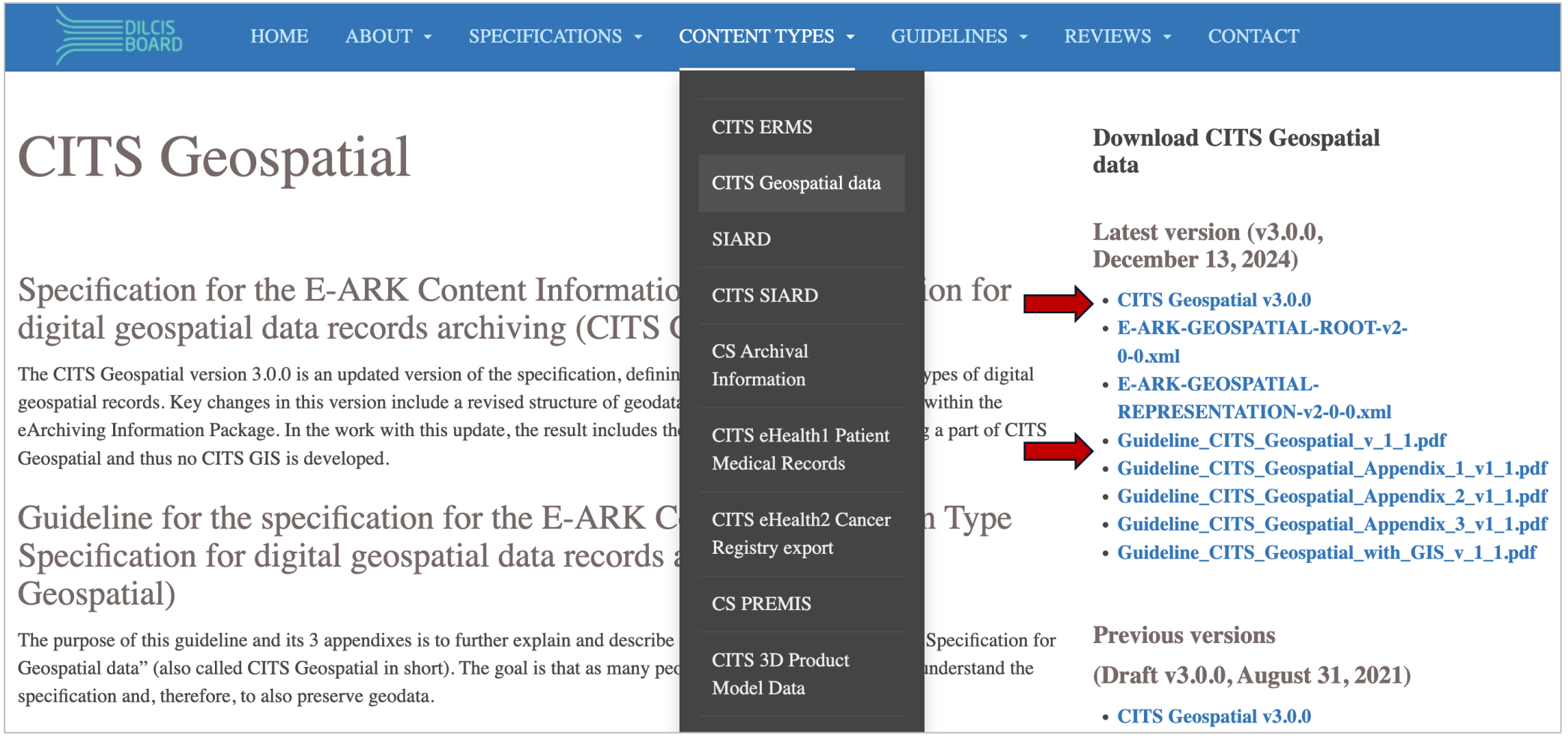
The Digital Information LifeCycle Interoperability Standards Board (DILCIS Board)

The Digital Information LifeCycle Interoperability Standards Board (DILCIS Board) https://dilcis.eu/ is an international group of experts committed to maintain and sustain a set of interoperability specifications that allow for the transfer, long-term preservation, and reuse of digital information regardless of the origin or type of the information.

More specifically, the DILCIS Board maintains specifications initially developed within the E-ARK Project (02.2014 – 01.2017), and which are now the core of the eArchiving Building Block.

European Commission

DILCIS Board



Thank you

Contact

https://www.e-ark-foundation.eu/

support@e-ark-foundation.eu

https://www.linkedin.com/groups/8343650/

https://bsky.app/profile/eu-earchiving.bsky.social

https://www.youtube.com/@e-ark