Specification for the E-ARK Content Information Type Specification for digital geospatial data records archiving (CITS Geospatial)

A proper front page will be created for the publication occurring after implementation of review comments.

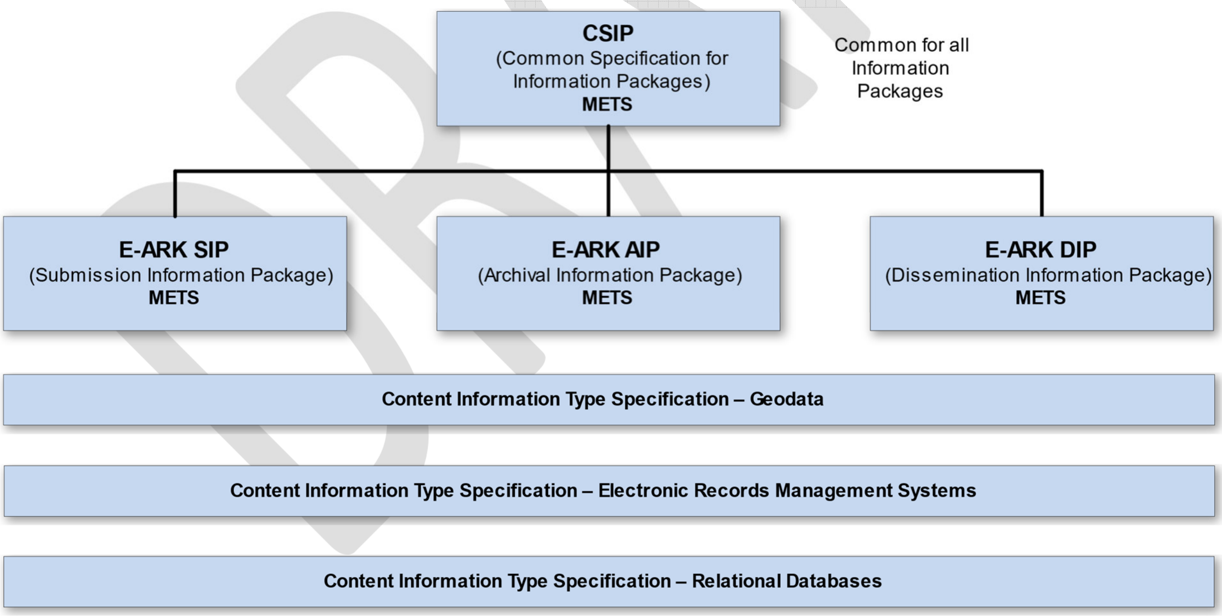
1. Preface

The correct preface will be inserted for the publication occurring after implementation of review comments.

1.1 Aim of the specification

This specification is one of several related specifications. The single most important aim of all of these specifications is the provision of a common set of specifications for packaging digital information for archiving purposes. The specifications are based on common, international standards for transmitting, describing and preserving digital data. They have been produced to help data creators, software developers and digital archives to tackle the challenge of short-, medium- and long-term data management and re-use in a sustainable, authentic, cost-efficient, manageable and interoperable way.

The foundation upon which the specifications are built is the Reference Model for an Open Archival Information System (OAIS) (OAIS Reference model) which has Information Packages as its basis. Familiarity with the core functional entities of OAIS is a prerequisite for understanding the specifications. A visualisation of the current specification network can be seen here:



The E-ARK specification dependency hierarchy

Specification	Aim and Goals
Common Specification for Information Packages	This document introduces the concept of a Common Specification for Information Packages (CSIP). Its three main purposes are to:

	• Establish a common understanding of the requirements which need to be met in order to achieve interoperability of Information Packages. • Establish a common base for the development of more specific Information Package definitions and tools within the digital preservation community. • Propose the details of an XML-based implementation of the requirements using, to the largest possible extent, standards which are widely used in international digital preservation. Ultimately the goal of the Common Specification is to reach a level of interoperability between all Information Packages so that tools implementing the Common Specification can be adopted by institutions without the need for further modifications or adaptations.
E-ARK SIP	The main aims of this specification are to: • Define a general structure for a Submission Information Package format suitable for a wide variety of archival scenarios, e.g. document and image collections, databases or geographical data. • Enhance interoperability between Producers and Archives. • Recommend best practices regarding metadata, content and structure of Submission Information Packages.
E-ARK AIP	The main aims of this specification are to: • Define a generic structure of the AIP format suitable for a wide variety of data types, such as document and image collections, archival records, databases or geographical data. • Recommend a set of metadata related to the structural and the preservation aspects of the AIP as implemented by the reference implementation eArchiving ToolBox (formerly earkweb). • Ensure the format is suitable to store large quantities of data.
E-ARK DIP	The main aims of this specification are to: • Define a generic structure of the DIP format suitable for a wide variety of archival records, such as document and image collections, databases or geographical data. • Recommend a set of metadata related to the structural and access aspects of the DIP.
Content Information Type Specifications	The main aim and goal of a Content Information Type Specification is to: • Define, in technical terms, how data and metadata must be formatted and placed within a CSIP Information Package in

	order to achieve interoperability in exchanging specific Content Information. The number of possible Content Information Type Specifications is unlimited.

1.2 Organisational support

This specification is maintained by the Digital Information LifeCycle Interoperability Standards Board (DILCIS Board). The DILCIS Board (http://dilcis.eu/) was created to enhance and maintain the draft specifications developed in the European Archival Records and Knowledge Preservation Project (E-ARK project) which concluded in January 2017 (http://eark-project.com/). The Board consists of eight members, but there is no restriction on the number of participants in the work. All Board documents and specifications are stored in GitHub (https://github.com/DILCISBoard) while published versions are made available on the Board webpage. Since 2018 the DILCIS Board has been responsible for the core specifications in the Connecting Europe Facility eArchiving Building Block (https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eArchiving).

1.3 Authors

A full list of contributors to this specification, as well as the revision history can be found in Appendix 1.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

GLOSSARY

Term	Description
Archival information package (AIP)	An Information Package, consisting of the Content Information and the associated Preservation Description Information (PDI), which is preserved within an Open Archival Information System (OAIS)
Cardinality	The term describes the number of elements in a set (Archival package). The numbers have the following meanings: (1..1) – In each set, there is exactly 1 such element present (0..1) – The set can contain from 0 to 1 of such elements (1..n) – The set contains at least one element – up to n elements (0..n) – The package can contain such element, but it is not mandatory
Coordinate Reference System (CRS)	CRS is a coordinate-based local, regional or global system used to locate geographical entities. A coordinate reference system defines a specific map projection, as well as transformations between different spatial reference systems. Spatial reference systems are defined by the OGC's Simple feature access using well-known text representation of coordinate reference systems, and support has been implemented by several standards-based geographic information systems.
digital geodata records	Digital geodata records are records containing a spatial graphical component, describing objects in space. They can be created digitally, or digitised from an analogue source (paper maps)
Dissemination Information Package (DIP)	An Information Package, derived from one or more AIPs, and sent by Archives to the Consumer in response to a request to the OAIS.
Format Descriptions	Many of the Format Descriptions given in this document use the descriptions published by the Library of Congress, which gives a comprehensive overview of the format. The information is available at https://www.loc.gov/preservation/digital/formats/intro/intro.shtml
Geodata layer	A Geodata layer is a representation of one or many geospatial datasets within a GIS System. It can contain additional representation information such as visualisation, labelling of the dataset, visibility under certain conditions based on scale, SQL query, etc.
Geoprocessing workflows	Geoprocessing workflows are usually defined as a set of Geoprocessing tasks organised into a process. Geoprocessing tasks are functions of a GIS system used to manipulate, transform or manage geospatial datasets, maps and tables.
Geospatial dataset	Geospatial datasets are sets of records defining a type of object in space. For example, a vector dataset of roads will contain graphical representations of

	roads and a respective table containing attributes describing each object within the data model of the dataset.
Geospatial reference	Geospatial reference is the definition of the coordinate reference system for the geospatial dataset.
GIS	Abbreviation for Geographical Information System, which is a system designed to capture, store, manipulate, analyse, manage, and present spatial or geographic data.
GIS Project	A GIS project is a document that organises geospatial datasets into layers, defines the map representations, then reports and stores information on Geoprocessing workflows.
Information Package	A logical container composed of optional Content Information and optional associated Preservation Description Information. Associated with this Information Package is Packaging Information used to delimit and identify the Content Information and Package Description information used to facilitate searches for the Content Information.
Lossless compression	Lossless compression is a class of data compression algorithms that allows the original data to be perfectly reconstructed from the compressed data. By contrast, lossy compression permits reconstruction only of an approximation of the original data, though usually with improved compression rates.
LZW	Lempel–Ziv–Welch is a universal lossless data compression algorithm created by Abraham Lempel, Jacob Ziv, and Terry Welch.
machine-readable Documentation	A machine-readable document is a document whose content can be readily processed by computers. Such documents are distinguished from machine-readable data by virtue of having sufficient structure to provide the necessary context to support the business processes for which they are created.
Open Archival Information System (OAIS)	An Archive, consisting of an organisation, which may be part of a larger organisation, of people and systems, that has accepted the responsibility to preserve information and make it available for a Designated Community. It meets a set of responsibilities that allow an OAIS archive to be distinguished from other uses of the term ‘Archive’. The term ‘Open’ in OAIS is used to imply that this Recommendation and future related Recommendations and standards are developed in open forums, and it does not imply that access to the archive is unrestricted.
Preservation Description Information (PDI)	The information which is necessary for adequate preservation of the Content Information and which can be categorised as Provenance, Reference, Fixity, Context, and Access Rights Information.
Projected coordinate systems	Geospatial data can have a geographic coordinate system or a projected coordinate system or a geographic projection. Projections are used to define coordinates in distance units as opposed to angular units in geographic coordinate systems.
RDBMS	Relational Database Management System
Representation	A Representation within an archival package contains archival data. If we store the same data in two or more different formats (i.e. Original and a Long-term preservation format) or different types of organisations than we

	organise them within two or more representations within the Representations folder of the Archival Package
Representation Information	The Representation Information must enable or allow the recreation of the significant properties of the original data object. In terms of geospatial data, we need the information required to reconstruct the usage of the records meaningfully. For example, if we want to adequately re-use a GML file, containing only the vector geometry and its accompanying attributes, we need rendering information in the form of symbology definition, labelling logic, the coordinate system and projection, the scales in which it was used and description of meanings of attributes in order to understand the data.
Submission Information Package (SIP)	An Information Package that is delivered by the producer to the OAIS for use in the construction or update of one or more AIPs and/or the associated Descriptive Information.
Well-defined geospatial position	As defined with a coordinate reference system.
Well-documented graphical component	Defined using a standardised or proprietary format for geospatial information, that stores coordinates of objects or their representations in space.

2 Context

2.1 Purpose and scope

The purpose of this document is to describe the basic required content information for preserving digital geospatial records and provides an understanding of the recommended elements needed to archive Geographic Information Systems (GIS) within an Information Package, but the full description is not ready yet, which would enable the re-use of the geospatial records in an environment similar to the original system. The definition of required content information for archiving GIS is not within the scope of this document. However, we provide some basic guidelines regarding which Documentation is needed, but the exact standards and definitions have to be agreed upon or even created by the community.
The document defines which content information is required and what information it needs to contain. It also provides some possible examples and a structure that can be used within the Information Package.

2.2 What are geodata records and Geographic Information Systems?

A Geographic Information System (GIS) is a framework for gathering, managing and analysing data. Outputs from a GIS are a combination of digital **geospatial records (geodata)** and a set of **processes** for transforming basic records into outputs (information products). Outputs can be in the form of maps, lists, new sets of geodata, etc.

In order to archive a GIS in such a way that the outputs from the original system can be reproduced, it is necessary to archive the processes used to manipulate data into outputs.

Geodata are any digital records that describe an object in space using coordinates based on a geographic coordinate system and a set of descriptive elements called attributes. They mostly come in two forms, vector data (points, lines, polygons) and raster data (sets of one or multiple arrays of pixels).

Increasingly, different geospatial formats include geospatially focused datasets or databases that contain primary information about a geographic location. In addition, ancillary and supplemental data that either are included or can be derived using spatial analysis are considered necessary for the full and effective functioning, interpretation and re-use of the data.

Geodata has many properties that define its accuracy and usability. These properties are commonly described in metadata and should also be part of the archival package. **Processes** are used to organise and transform Geodata in GISs to derive value from the dataset itself as well as from the combination with other datasets and databases within the system. Hence, to preserve data from a GIS, it is necessary to preserve the relations between geodata and other databases and the logic with which the data was used (Representation Information).

3 Elements of the geodata information package

Four basic groups of elements were identified in order to sufficiently describe digital geospatial records for preservation in an information package.

- **Geodata** (data object) – This contains the actual geodata records (datasets), exported from the existing GIS system. The same geodata can be stored in one or multiple representations[1]. This is explained in greater detail in the Common Specification for Information Packages.

- **Technical Documentation** (Representation Information) – This describes the geodata in a way that enables unequivocal interpretation of data and reconstruction of the original data products. If applicable, technical Documentation should include the: attribute definitions tables, the logical structure of the geodatabase, structure of the GIS project, visualisation information for geodata layers (cartography, labelling, etc.) – a data dictionary. Here we also store the logic of conducting analysis, common queries, common processing workflows etc. The terms are described in the following sections.
  Technical Documentation can vary between representations; therefore, we recommend it is stored in a "documentation" folder within its representation.

- **Context documentation** (knowledge base) – This may contain any other information that provides contextual information for the geodata such as project reports, user manuals, interviews with the data producer, etc. If we only have more than one representation, we can store it within the Representation Documentation folder if it pertains only to one specific representation or in the Documentation folder, that lies on the root of the package if it pertains to all the representations.

- **Geodata specific Metadata** – This includes machine-readable XML files based on geodata metadata standards (ISO 19115:2003[2], EC INSPIRE directive[3]) and any other form of metadata (txt file descriptions, pdf, etc.) is classified as Documentation. Multiple representations of geodata often have their own metadata; therefore, we recommend storing them in the metadata folder within the representation.

[1] As described in the section 4 of the E-ARK CSIP Specification (https://earkcsip.dilcis.eu/pdf/eark-csip.pdf)
[2] https://www.iso.org/standard/26020.html
[3] https://inspire.ec.europa.eu/

The CSIP principles are implemented through a folder structure, as seen in Figure 1.

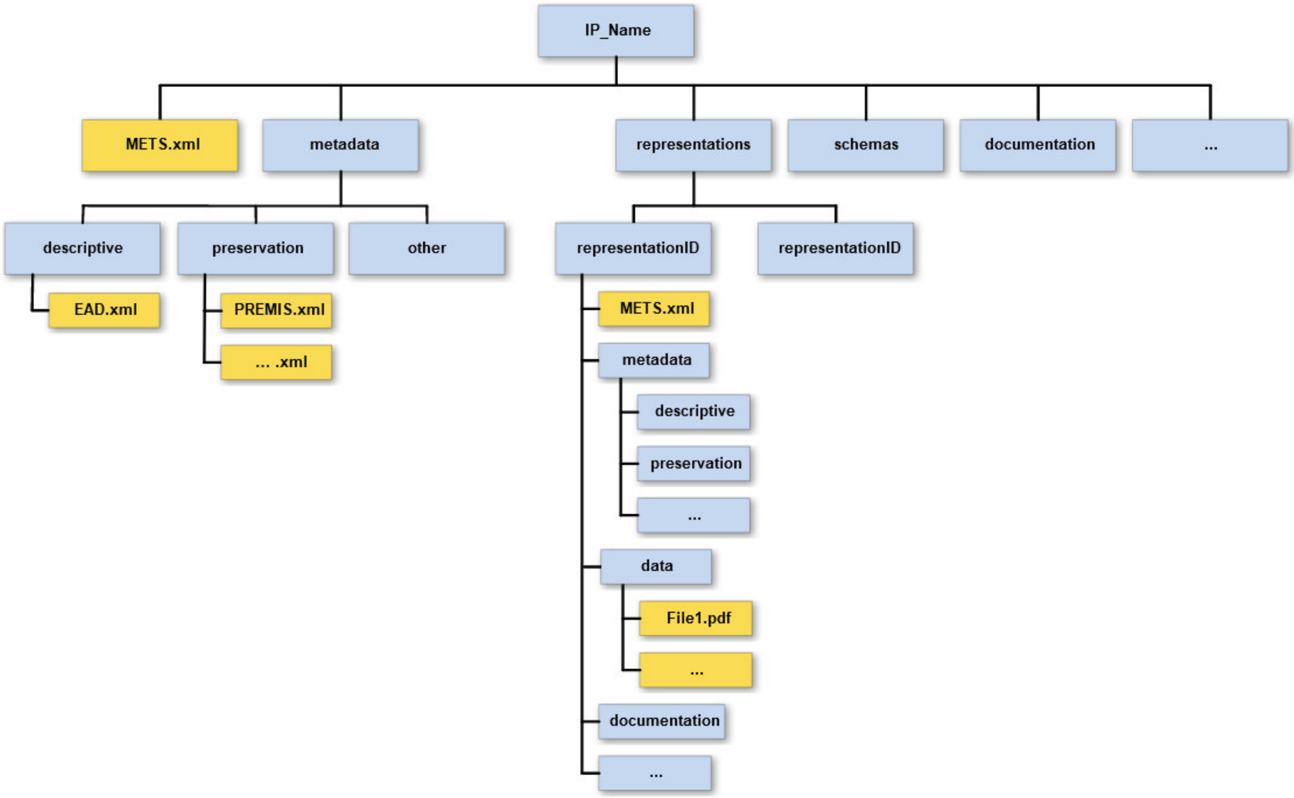


Figure 1: CSIP Information Package folder structure

When the Information Package is used for the transfer of Geospatial information, the foreseen placement of the files is shown in Figure 2. In this figure, white boxes with a number and text denote the section in this document where more information can be found on the type of information. The green boxes denote the type of geospatial documentation and metadata, which is to be included. More information on these is provided in the following subsections.

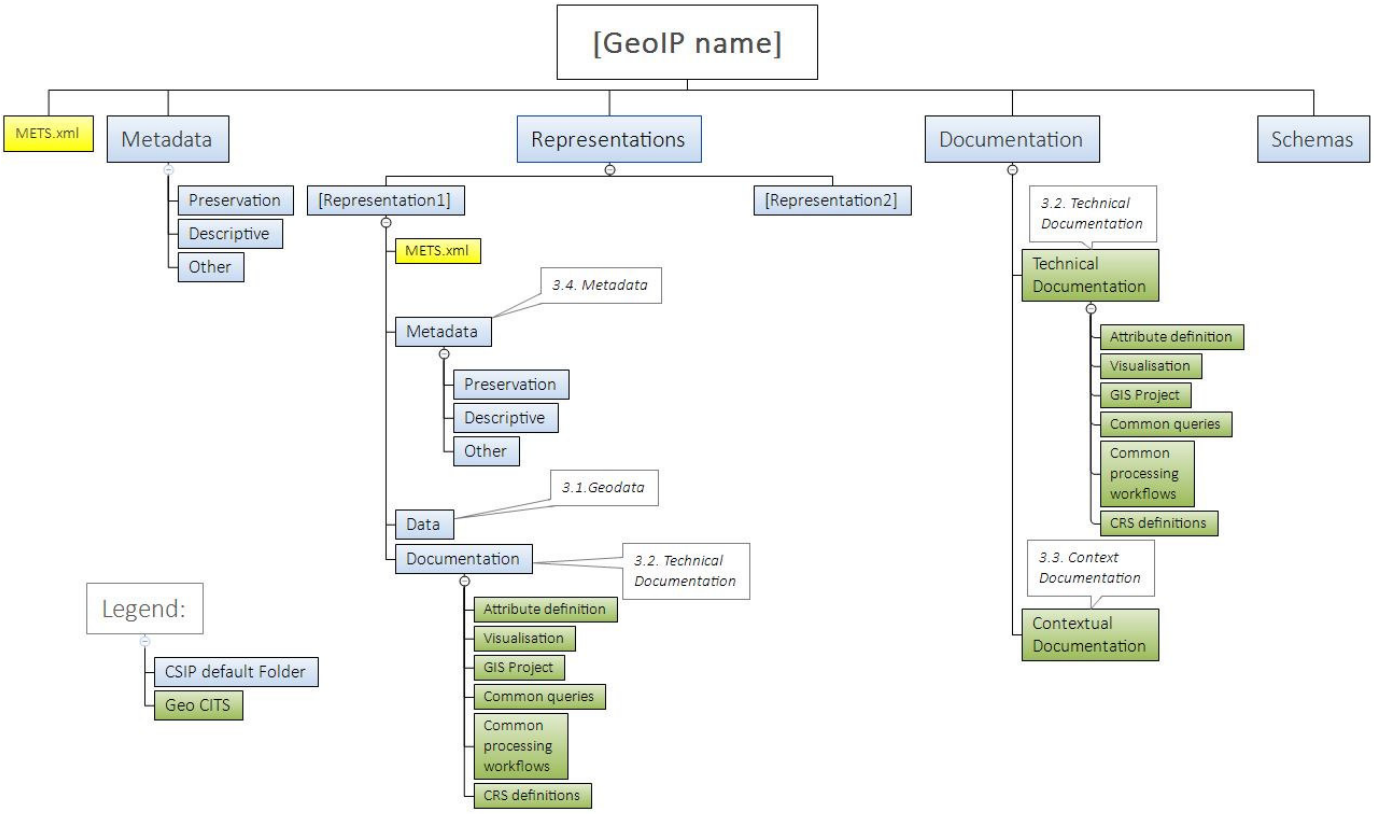


Figure 2: Elements of Geodata Information Packages

3.1 Geodata

Following OAIS[4], geodata presents the data object in the Information Package. Geodata are records representing and describing objects and phenomena in space. Geodata is always composed of the graphical part in ether vector or raster form, which has a defined location in space in the form of geographical coordinates. The descriptive part connects the graphical element with the descriptions in the form of one or more tables, as seen in Figure 3. In the case of geodata in raster format, additional tables may not always be present. When archiving geodata, it is necessary to ensure that there is a well-documented graphical component, well-defined descriptive attributes and a geographic coordinate system.

[4] https://www.iso.org/standard/57284.html

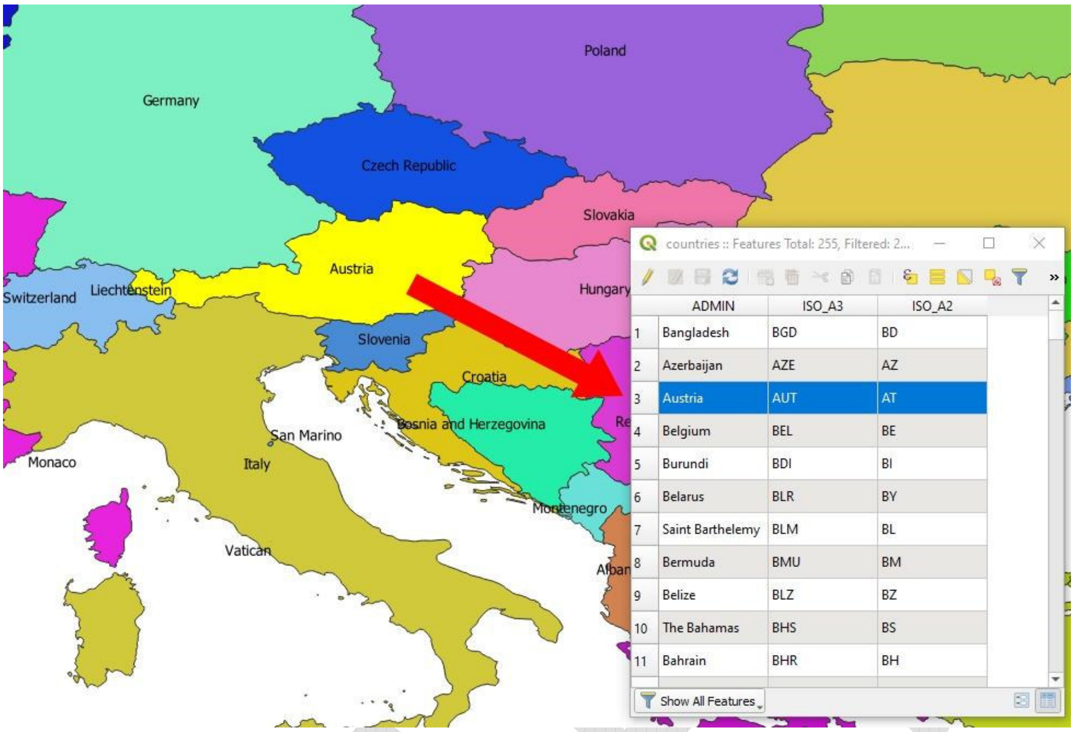


Figure 3: A Vector Graphical component is connected to the descriptive component in a table

In order to ensure the long-term preservation of geodata, proprietary and undocumented formats must be converted into a long-term preservation format that is well described and defines all the necessary elements of the geospatial dataset (unique geospatial reference, usable independent of the system and well-documented). Even if some formats are a *de facto* standard today, they may become unreadable in the distant future.

3.1.1 Vector geodata

Vector geodata are sets of coordinates, representing objects in space. Depending on the geometry of the object, such data can be represented in the form of points, lines or polygons. Vector geodata comes in many formats, depending on the system of origin. Generally, it takes the form of tables, where one of the columns contains the graphical information and the other columns represent descriptive information which enables the visualisation, querying and analysis of the data. One format that can be used is the Geography Markup Language (GML).

Geography Markup Language (GML) format

A format that fits the criteria for a long-term preservation format for vector geodata is the Geography Markup Language (GML) format (version 3.2.1./ISO 19136:2007)[5]. This format is well-documented because it is an ISO standard. It contains a definition of the geospatial coordinate system and enables descriptions of both simple elements (point, line, polygon) and complex vector elements (arcs, topology definitions, etc.). Since the GML format was intended as a transfer format, so it is not as commonly used in practice as some *de facto* standards, such as the ESRI shapefile[6]. However, this is less important when considering long-term preservation, where the ability to read it after a long period is critical.

```
<?xml version="1.0" encoding="utf-8" ?>
<ogr:FeatureCollection
     gml:id="aFeatureCollection"
```

[5] https://www.iso.org/standard/32554.html

[6] https://www.loc.gov/preservation/digital/formats/fdd/fdd000280.shtml

```
     xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
     xsi:schemaLocation="http://ogr.maptools.org/ Ljubljana.xsd"
     xmlns:ogr="http://ogr.maptools.org/"
     xmlns:gml="http://www.opengis.net/gml/3.2">
  <gml:boundedBy>
       <gml:Envelope srsName="urn:ogc:def:crs:EPSG::3912">
              <gml:lowerCorner>454632.64 92348.50</gml:lowerCorner>
              <gml:upperCorner>481420.94 111431.81</gml:upperCorner>
       </gml:Envelope>
</gml:boundedBy>
<ogr:featureMember>
    <ogr:Ljubljana gml:id="Ljubljana.0">
      <gml:boundedBy>
              <gml:Envelope srsName="urn:ogc:def:crs:EPSG::3912">
                     <gml:lowerCorner>454632.64 92348.50</gml:lowerCorner>
                     <gml:upperCorner>481420.94 111431.81</gml:upperCorner>
              </gml:Envelope>
         </gml:boundedBy>
      <ogr:geometryProperty>
              <gml:MultiSurface srsName="urn:ogc:def:crs:EPSG::3912" gml:id="Ljubljana.geom.0">
                     <gml:surfaceMember>
                       <gml:Polygon gml:id="Ljubljana.geom.0.0">
                            <gml:exterior>
                                   <gml:LinearRing>
                                          <gml:posList>474745.53 93881.61 474750.62
                                          93867.39 474736.89 93857.48 474728.21
                                          93852.96 474701.70 93853.25 474690.56
                                          ....  ["coordinate  list  truncated  for  presentation
purposes"]
                                          474742.49 93890.11 474745.53 93881.61
                                          </gml:posList>
                                   </gml:LinearRing>
                            </gml:exterior>
                       </gml:Polygon>
                     </gml:surfaceMember>
              </gml:MultiSurface>
         </ogr:geometryProperty>
      <ogr:UNIT>OB</ogr:UNIT>
      <ogr:OB_MID>11027849</ogr:OB_MID>
      <ogr:OB_UIME>Ljubljana</ogr:OB_UIME>
      <ogr:AREA_KM2>275.03</ogr:AREA_KM2>
    </ogr:Ljubljana>
  </ogr:featureMember>
</ogr:FeatureCollection>
```


3.1.2 Raster geodata

Geodata in the form of raster datasets are actually array-oriented data structures, where the information is stored in the value of the pixels (or array cells) that compose the raster. In this way the raster covers a larger area and is mainly used as continuity surfaces that present remote sensing imagery, digital elevation, scanned maps, etc. This often calls for special requirements for raster formats that enables the storage of multiple types of values per location (text, number), or a combination of up to 400 colour bands (hyperspectral imagery) in one image or a multidimensional representation (netCDF7) and others. Another property determining the choice of a long-term preservation format is the image size and the level of lossless compression the format offers.

When choosing a long-term preservation format, it is necessary to ensure that all information that the original format provides is retained and that there is a well-defined geospatial position of the raster and the coordinate system.

Two formats that can be used are described below.

- **TIFF:** A good long-term preservation format for most raster datasets is TIFF[8]; however, in its basic form, it lacks the geospatial position definition. TIFF also supports LZW[9] as lossless compression. A

[7] https://www.loc.gov/preservation/digital/formats/fdd/fdd000330.shtml
[8] TIFF – https://www.loc.gov/preservation/digital/formats/fdd/fdd000022.shtml
[9] LZW Compression – https://www.loc.gov/preservation/digital/formats/fdd/fdd000074.shtml

plain TIFF file should be augmented with additional files describing the geospatial reference, which depend on the GIS Tool (one example would be a TIFF world file (TFW)[10] describing coordinates and PRJ file defining the geographic coordinate reference system and its projection).

- **GeoTIFF:** GeoTIFF[11] is a public domain geospatial metadata standard which allows geo-referencing information to be embedded within a TIFF file. However, there are some uncommon projected coordinate systems which are not supported. In such cases, it is necessary to augment the GeoTIFF with the additional definition of the coordinate system, in a similar manner as is the case with TIFF.

3.1.3 Additional tabular information

Geodata is often a part of a complex data structure, stored in a database along with ordinary tables. To reproduce the information products from a GIS, it is often necessary to store additional tables with the geodata. These tables do not have their own geospatial component. In this case, it is important to store the relationships and logic of the data structure, so that it can be reconstructed in the future. For long-term preservation of additional tabular information (attribute tables, code lists, etc.) along with geodata, formats proposed for RDBMS archiving are used.[12] Two formats that can be used are described below.

- **SIARD:**[13] The Software Independent Archiving of Relational Databases (SIARD) developed by the Swiss Federal Archives (SFA) now maintained in cooperation between the DILCIS Board and SFA. The format is since being further developed by the digital preservation community. It is a normative description of a file format for the long-term preservation of a relational database.
- **CSV:**[14] In simple cases, the CSV format can be used as a text-based format for storing tabular information. It is important that the structure of the table and its nominal code page is defined in the Representation Information part of the Information Package. Although it is a simple format and commonly used, it could be difficult to validate and therefore migrate correctly.

3.1.4 Geospatial reference

In order to properly render the geodata in any future GIS, it is necessary to specify the geospatial reference system. The Coordinate Reference System (CRS) provides information about how to locate geodata objects anywhere on the earth's surface. Elements of the spatial reference system are **projection**, **geodetic datum**, and **unit of measurement.**

The CRS definition for geodata within an archival information package can be accomplished in many ways, including:

- embedded within the data itself (e.g. GML 3.2.1. and GeoTIFF),
- written in the accompanying files (e.g. GML 3.1.1. and earlier, ESRI Shapefile, TIFF, JPEG2000 and GMLJP2[15]),
- recorded in the accompanying Documentation which needs to be recreated for every subsequent dataset.

Ease of use of geodata in a DIP diminishes with the number of steps needed to define the CRS for every geodata layer. If geodata was used using a standard and documented CRS, it can be referenced by linking to a well-documented list of CRSs (an example is a reference code from the International Association of

[10] TFW World file –https://www.loc.gov/preservation/digital/formats/fdd/fdd000287.shtml
[11] GeoTIFF – https://www.loc.gov/preservation/digital/formats/fdd/fdd000279.shtml
[12] https://dilcis.eu/content-types/siard
[13] SIARD Description and examples – https://github.com/DILCISBoard/SIARD
[14] CSV – https://www.loc.gov/preservation/digital/formats/fdd/fdd000323.shtml
[15] http://www.opengeospatial.org/standards/gmljp2

Oil and Gas Producers Geodetic Parameter Registry (EPSG.[16])) or describing it in a well-documented format like the Well Known Text (WKT), as defined within the ISO 19162:2019[17].

Three formats that can be used are described in the following subsections.

CRS as defined within GML 3.2.1

Geo-referencing information in GML is a mandatory part of the file itself, and a reference to it is embedded in the geodata file itself:

```
<gml:boundedBy>
   <gml:Envelope srsName="urn:x-ogc:def:crs:EPSG:4326">
       <gml:lowerCorner>50.23 9.23</gml:lowerCorner>
       <gml:upperCorner>50.31 9.27</gml:upperCorner>
   </gml:Envelope>
</gml:boundedBy>
```

The attribute “srsName” holds the value of the coordinate reference system code, according to EPSG. In this example, the code is 4326[18].

In case of referencing external CRS catalogues, the package should contain a definition of the referenced CRS (with all parameters, needed to recreate it) as a separate technical documentation file.

CRS in ESRI Shapefile (shp)

Although the ESRI shapefile is not the best choice of formats for a long-term preservation format, it could be used as a current DIP format. A SHP needs a <shapefilename>.prj file in order to be properly geo-referenced. A .prj file is a .txt file, containing a definition of the coordinate reference system and all of its elements. Here is an example:

```
PROJCS["NAD_1983_UTM_Zone_10N",GEOGCS["GCS_North_American_1983",
DATUM["D_North_American_1983",SPHEROID["GRS_1980",6378137,298.257222101]],
PRIMEM["Greenwich",0],UNIT["Degree",0.0174532925199433]],
PROJECTION["Transverse_Mercator"],PARAMETER["False_Easting",500000.0],PARAMETER["False_N
orthing",0.0],PARAMETER["Central_Meridian",-123.0],PARAMETER["Scale_Factor",0.9996],
PARAMETER["Latitude_of_Origin",0.0],UNIT["Meter",1.0]]
```

CRS in a TIFF

A basic TIFF raster geodata file does not contain any CRS information. A TIFF file must therefore be accompanied by the “TFW” file that contains its initial coordinates and pixel size and a file that defines the geospatial coordinate system, such as “PRJ”. For example, a D240143.tif file would be accompanied by a D240143.tfw file. That is a .txt file, containing information about the coordinates and the size of the first top-left pixel. Here is an example:

```
0.42333
0.0
0.0
-0.42333
394250.00
```

[16] EPSG – Geodetic Parameter Registry http://www.epsg.org
[17] https://www.iso.org/standard/76496.html
[18] World Geodetic System 1984 (https://epsg.org/crs_4326/WGS-84.html?sessionkey=z5jf4xc886

3.1.5 Other geodata types

Geodata can be presented in different formats and as a part of different data structures (such as geometric networks, coverages, structures combining raster and vector data, etc.), as web services, automated maps, etc.

This document focuses on the fundamental structures of geodata, which are used as basic input components for complex structures. In order to replicate the complex data structures or services in the future, their organisational logic and the way they are used within applications needs to be documented.

3.2 Technical Documentation

Geodata is rarely in a form that is sufficiently self-explanatory to be used and properly interpreted by itself. Consequently, additional information is required in order to enable the user to properly understand, interpret and use geodata. This section describes the required technical Documentation for geodata datasets (where it is applicable). Ideally, a machine-readable format of Documentation is preferred; however, any other form of documenting the system is welcome.

In this document, technical Documentation and general contextual Documentation of geodata are differentiated. The purpose of the technical documentation is to provide Representation Information that enables the re-use of geodata in a way that is similar enough to its use in the initial system. Technical Documentation can be stored with the rest of the Documentation, if it contains all the required information and pertains to all the representations.

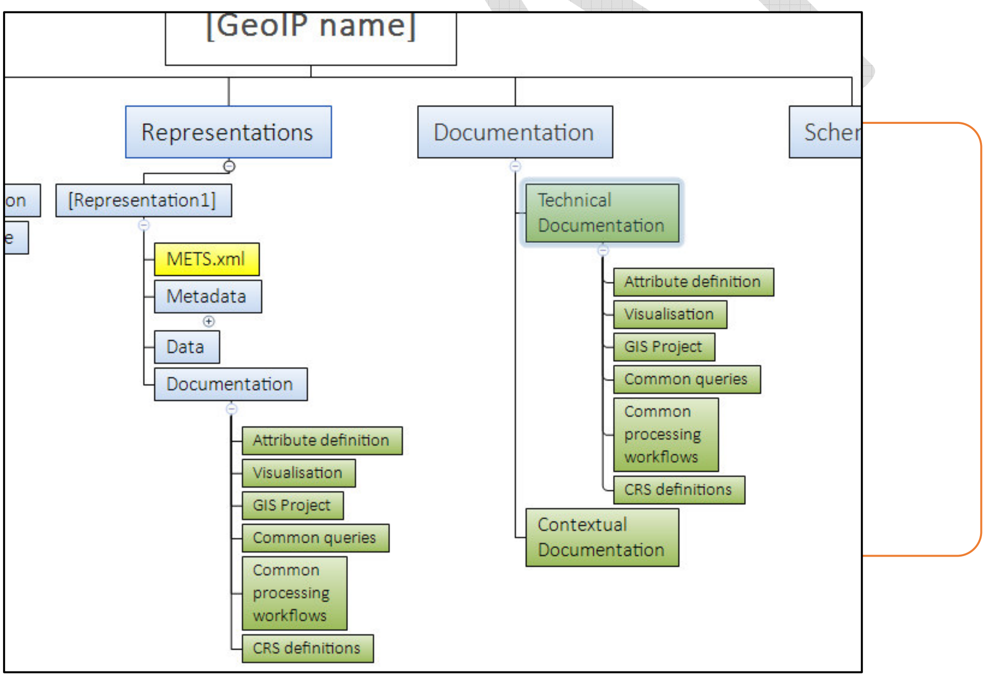


Figure 4: Possible locations of Technical documentation

In ***Fel! Hittar inte referenskälla.***, we can see that it can be stored in two locations in the GeoIP – within a Representation, or under the Root Documentation folder. All technical documentation that is stored in a standardised machine-readable format should be stored within the representation. If the same Documentation is valid for multiple representations, it should be stored in the Documentation folder on the root level. All non-machine readable, descriptive technical Documentation is to be stored in the Technical documentation folder within the Documentation folder on the root level, except if it is representation specific.

Such documentation spans from basic information (how to cartographically render a simple layer or how to set up a basic query using a simple combination of datasets) up to detailed documentation of the initial system that would allow the reconstruction of a technical environment in the future that could produce similar information products as the initial system. When creating a GeoIP, it is required that the producer ensures that the following subsections describing the different types of technical documentation (if applicable) is part of the transfer.

3.2.1 Attribute definition

Attribute definition describes the value types and their meaning for each column in the table of a geodata dataset or an accompanying table. This can be a stand-alone file, such as an XSD schema or a part of a broader Data Dictionary of Feature Catalogue file or document.

Two formats that can be used are described below.

Attribute definition in GML with XSD schema

The attribute definition of table elements in a GML is expressed in an accompanied XSD schema. That is why an XSD schema should be present for every GML file in the Information Package if this format is used. It also provides means for GML validation, and at the same time, it documents the dataset structure.

```
<?xml version="1.0" encoding="utf-8"?>
<schema
   xmlns="http://www.w3.org/2001/XMLSchema"
   xmlns:gml="http://www.opengis.net/gml"
   xmlns:xlink="http://www.w3.org/1999/xlink"
   xmlns:sa="http://www.sa.dk/gml/sa"
   targetNamespace="http://www.sa.dk/gml/sa" elementFormDefault="qualified">
   <import namespace="http://www.opengis.net/gml"
   schemaLocation="http://schemas.opengis.net/gml/3.1.1/base/gml.xsd"/>
   <element name="GEOMETRI" type="sa:GEOMETRIType" substitutionGroup="gml:_Feature"/>
   <annotation>
       <documentation>Geodata in the gml-file are points, locating local archives in Denmark. These
points are linked to SIARD database data in the information package via the local feature/attribute
"arkivID" in the gml-file that references coloumn "arkivID" in table "Paragraf_Syv_Arkiver" in the
SIARD format (database) in the information package.</documentation>
   </annotation>
       <complexType name="GEOMETRIType">
       <complexContent>
          <extension base="gml:AbstractFeatureType">
              <sequence>
                  <element name="arkivID" type="string" minOccurs="1" maxOccurs="1">
                     <annotation>
                      <documentation>Unique identification of the archive.Unik arkividentifikator
som defineres fortløbende. This local feature "arkivID" in the gml file references coloumn "arkivID"
in table "Paragraf_Syv_Arkiver" in the SIARD format (database) in the information package. Thus
this local feature "arkivID" links the geography to other attributes in other parts of the
information package</documentation>
                     </annotation>
                  </element>
                 <element name="arkivNavn" type="string" nillable="false">
                     <annotation>
                           <documentation>Name of the local archive.</documentation>
                     </annotation>
                  </element>
                 <element name="kommune" type="string" nillable="false">
                     <annotation>
                           <documentation>Name of the municipial area that the local archive
belongs to.</documentation>
                     </annotation>
                  </element>
                 <element name="kommuneNummer" type="integer" nillable="false">
                     <annotation>
```

```
                    <documentation>Number og the municipial area.</documentation>
                </annotation>
            </element>
            <element name="adresse" nillable="false">
                <annotation>
                    <documentation>Address of the local archive.</documentation>
                </annotation>
            </element>
            <element ref="gml:location" minOccurs="1">
                <annotation>
                    <documentation>Point locating the local archive.</documentation>
                </annotation>
            </element>
        </sequence>
      </extension>
    </complexContent>
  </complexType>
</schema>
```

Attribute definition in a textual form as a part of the data dictionary

The format of this description can be an ordinary txt, html or pdf document. However, a machine-readable format would be more useful for future use within the data-driven economy. In this example, we have a txt file describing a "Street Layer". In the bottom part, we see a tab-delimited table describing the Attribute names, Data Type, Length, Alias Name and Description.

```
Name           Street
ShapeType      Polyline
FeatureType    Simple
AliasName      Public Streets
HasM           false
HasZ           false
HasAttachment  false

Description    The centerline representation of public roadways or streets. These are a subset of
               the road segments found in the RoadCenterline feature class and only those actively
               managed by a local government public works organisation. The information typically
               is derived from a RoadCenterline layer maintained for address assignment and is a
               framework for the management of streets, roads and work activities associated with
               such.
```

Field	DataType	Length	AliasName	Description
SEGMENTID	String	20	Segment Identifier	The segment identification number
FULLNAME	String	255	Full Road Name	The full name of the road including any prefix and/or suffix
RECLENGTH	Double	8	Record Length	The record length of the roadway segment
LASTUPDATE	Date	8	Last Update Date	The date the feature was last updated in the maintenance database
LASTEDITOR	String	50	Last Editor	The user who performed the last update

3.2.2 Visualisation information

Data visualisation provides an illustration and representation of spatial data. The catalogue of cartographic symbols is a collection of agreed cartographic symbols, which are used via visualisation of spatial data sets to display objects in space. As shown in *Figure 6,* Cartographic symbols are shown in the legend, which explains their meaning.

Figure 5: Example of a Map Key (Legend)

For certain geodata, the visualisation is already done by the producer in the form of (geo-located) raster images or paper maps. In these cases, it is reasonable to archive that kind of visualisation. For each spatial data set, it is possible to produce any number of different visualisations with the appropriate software. It is proposed that:

- Every dataset is described with at least a screenshot image of the geodata dataset shown to its full extent, to enable easy discovery and identification in the archival catalogue.
- If a cartographic key exists, it should be documented in a way that it can be satisfactorily reproduced in a future system.
- If geodata was used to produce complex maps, the logic is preserved in such a way that a similar representation is possible in the future.
- If a visualisation was created using well-documented machine-readable files, they should be preserved.

One format that can be used is S.

SLD files

SLD[19] is an OGC[20] (Open Geospatial Consortium) standard for symbology and is the OGC Styled Layer Description XML format (SLD files). If the producer cannot provide the archive with SLD files, these can be recreated from the description which is provided in the Documentation in an open-source GIS application like QGIS[21]. Raster files can have a colour map associated with the pixel value. The SLD standard is used for rendering geodata in OGC web services and therefore could be used as an appropriate input for an easier DIP creation in the future.

[19] SLD – Styled Layer Description: https://portal.opengeospatial.org/files/?artifact_id=22364
[20] http://www.opengeospatial.org/
[21] QGIS – https://qgis.org/

```xml
<StyledLayerDescriptor xmlns="http://www.opengis.net/sld"
       xmlns:ogc="http://www.opengis.net/ogc"
       xmlns:xlink="http://www.w3.org/1999/xlink"
       xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
       version="1.0.0"
       xsi:schemaLocation="http://www.opengis.net/sld StyledLayerDescriptor.xsd">
   <NamedLayer>
       <Name>Simple Point</Name>
       <UserStyle>
           <Title>SLD Cook Book: Simple Point</Title>
           <FeatureTypeStyle>
               <Rule>
                   <PointSymbolizer>
                       <Graphic>
                           <Mark>
                               <WellKnownName>circle</WellKnownName>
                               <Fill>
                                   <CssParameter name="fill">#FF0000</CssParameter>
                               </Fill>
                           </Mark>
                           <Size>6</Size>
                       </Graphic>
                   </PointSymbolizer>
               </Rule>
           </FeatureTypeStyle>
       </UserStyle>
   </NamedLayer>
</StyledLayerDescriptor>
```

3.2.3 Table relationships

If the IP contains two or more geospatial datasets or additional tables, and if it is applicable, the technical Documentation should describe the relationships between tables in a database or within a GIS project, in order to enable the reconstruction of queries and provide a greater understanding of the usage of complex data structures (example Topologies, Geometric networks, etc.). This can be described in the .xsd file (attribute definition), that accompanies each .gml file or .siard file.

3.2.4 GIS Project (logical structure of datasets)

An Information product based on geospatial data is often a result of When a larger GIS project (containing many geodata datasets and additional tables) is documented, in order to be able to reproduce an information product (a web service) or repeat a workflow, the following should be documented:

- Logical structure of geodata layers and tables in a GIS Project.
- Geodata layer properties (definition query, scale-dependent display, visualisation parameters, etc.).
- Labelling (which layers were used for labelling, scale-dependent display of labels, text rendering properties, etc.).

Two formats that can be used are described below.

- **QGS file – GIS Project definition file on QGIS:** The QGIS application stores its Project structure of geodata layers and its rendering properties (Projection, symbology, labelling, etc.) in a .qgs file, which has an XML structure. So, it can be opened with any text reader, and layer hierarchy and its parameters can be read from the file. However, if used for archiving, additional Documentation on the .qgs file structure should be stored in the Information Package or available in the archive.
- **OWS Context[22]:** This is An Open Format Linking Geospatial Web Services and Information to enable seamless transfer of resources (local data or links to web services) and their organisation and visualisation between different platforms. Therefore, it makes it an excellent format for preserving existing distributed GIS systems.

3.2.4.1 Feature catalogue

The feature catalogue represents a logical structure of attributes. It provides an understanding of the meaning, use and structure of the spatial data and provides a unified classification of spatial data in feature types (classes). Feature types are distinguished by their attributes (properties), by importance and by the relationships between them. ISO 19110:2005 describes this in greater detail.[23]
An example of its usage is within strictly defined digital maps like in the Marine S-57 or S-100 products. There the structure of datasets, that construct a map is precisely defined by a Feature dataset, that specifies Features, its Attributes, Enumerants and Bindings.

[22] OWS Context homepage – https://www.owscontext.org/
[23] https://www.iso.org/standard/39965.html

3.2.5 Common queries and geoprocessing workflows

In the initial GIS, there is often a requirement to run certain database queries or geo-specific processes (geoprocessing workflows), to produce information products. Common information products from a GIS are maps, lists, charts, new geodata derived from existing data, web services, etc. To reproduce this type of GIS information product in the future, documentation of the queries and geoprocessing workflows is needed. Ideally, these queries should be described using machine-readable formats. However, any other form of Documentation is welcome.

3.2.6 Documenting Geospatial references in the archive

If the CRS of the geodata transferred to the archive is described by referencing to a well-documented list (such as the EPSG database) and is therefore dependent upon the long-term existence of that database, the archives should consider storing a local definition of the referenced CRS. This definition should preferably use one type of the standardised machine-readable formats of the CRS definition such as WKT2 (ISO 19162:2019), or a GML-based description.

Two ways that can be used are described in the following subsections.

GML Definition of the CRS referenced to external EPSG library

This is an example for the CRS with the EPSG code 3794.

GML - Slovene national GRID

```
<gml:ProjectedCRS gml:id="ogrcrs68">
  <gml:srsName>Slovenia 1996 / Slovene National Grid</gml:srsName>
  <gml:srsID>
    <gml:name gml:codeSpace="urn:ogc:def:crs:EPSG::">3794</gml:name>
  </gml:srsID>
  <gml:baseCRS>
    <gml:GeographicCRS gml:id="ogrcrs69">
      <gml:srsName>Slovenia 1996</gml:srsName>
      <gml:srsID>
        <gml:name gml:codeSpace="urn:ogc:def:crs:EPSG::">4765</gml:name>
      </gml:srsID>
      <gml:usesEllipsoidalCS>
        <gml:EllipsoidalCS gml:id="ogrcrs70">
          <gml:csName>ellipsoidal</gml:csName>
          <gml:csID>
            <gml:name gml:codeSpace="urn:ogc:def:cs:EPSG::">6402</gml:name>
          </gml:csID>
          <gml:usesAxis>
            <gml:CoordinateSystemAxis gml:id="ogrcrs71" gml:uom="urn:ogc:def:uom:EPSG::9102">
              <gml:name>Geodetic latitude</gml:name>
              <gml:axisID>
                <gml:name gml:codeSpace="urn:ogc:def:axis:EPSG::">9901</gml:name>
              </gml:axisID>
              <gml:axisAbbrev>Lat</gml:axisAbbrev>
              <gml:axisDirection>north</gml:axisDirection>
            </gml:CoordinateSystemAxis>
          </gml:usesAxis>
          <gml:usesAxis>
            <gml:CoordinateSystemAxis gml:id="ogrcrs72" gml:uom="urn:ogc:def:uom:EPSG::9102">
              <gml:name>Geodetic longitude</gml:name>
              <gml:axisID>
                <gml:name gml:codeSpace="urn:ogc:def:axis:EPSG::">9902</gml:name>
              </gml:axisID>
              <gml:axisAbbrev>Lon</gml:axisAbbrev>
              <gml:axisDirection>east</gml:axisDirection>
            </gml:CoordinateSystemAxis>
          </gml:usesAxis>
        </gml:EllipsoidalCS>
      </gml:usesEllipsoidalCS>
      <gml:usesGeodeticDatum>
        <gml:GeodeticDatum gml:id="ogrcrs73">
          <gml:datumName>Slovenia_Geodetic_Datum_1996</gml:datumName>
```

```
          <gml:datumID>
            <gml:name gml:codeSpace="urn:ogc:def:datum:EPSG::">6765</gml:name>
          </gml:datumID>
          <gml:usesPrimeMeridian>
            <gml:PrimeMeridian gml:id="ogrcrs74">
              <gml:meridianName>Greenwich</gml:meridianName>
              <gml:meridianID>
                <gml:name gml:codeSpace="urn:ogc:def:meridian:EPSG::">8901</gml:name>
              </gml:meridianID>
              <gml:greenwichLongitude>
                <gml:angle gml:uom="urn:ogc:def:uom:EPSG::9102">0</gml:angle>
              </gml:greenwichLongitude>
            </gml:PrimeMeridian>
          </gml:usesPrimeMeridian>
          <gml:usesEllipsoid>
            <gml:Ellipsoid gml:id="ogrcrs75">
              <gml:ellipsoidName>GRS 1980</gml:ellipsoidName>
              <gml:ellipsoidID>
                <gml:name gml:codeSpace="urn:ogc:def:ellipsoid:EPSG::">7019</gml:name>
              </gml:ellipsoidID>
              <gml:semiMajorAxis gml:uom="urn:ogc:def:uom:EPSG::9001">6378137</gml:semiMajorAxis>
              <gml:secondDefiningParameter>
                <gml:inverseFlattening
gml:uom="urn:ogc:def:uom:EPSG::9201">298.257222101</gml:inverseFlattening>
              </gml:secondDefiningParameter>
            </gml:Ellipsoid>
          </gml:usesEllipsoid>
        </gml:GeodeticDatum>
      </gml:usesGeodeticDatum>
    </gml:GeographicCRS>
  </gml:baseCRS>
  <gml:definedByConversion>
    <gml:Conversion gml:id="ogrcrs76">
      <gml:usesMethod xlink:href="urn:ogc:def:method:EPSG::9807"/>
      <gml:usesParameterValue>
        <gml:value gml:uom="urn:ogc:def:uom:EPSG::9102">0</gml:value>
        <gml:valueOfParameter xlink:href="urn:ogc:def:parameter:EPSG::8801"/>
      </gml:usesParameterValue>
      <gml:usesParameterValue>
        <gml:value gml:uom="urn:ogc:def:uom:EPSG::9102">15</gml:value>
        <gml:valueOfParameter xlink:href="urn:ogc:def:parameter:EPSG::8802"/>
      </gml:usesParameterValue>
      <gml:usesParameterValue>
        <gml:value gml:uom="urn:ogc:def:uom:EPSG::9001">0.9999</gml:value>
        <gml:valueOfParameter xlink:href="urn:ogc:def:parameter:EPSG::8805"/>
      </gml:usesParameterValue>
      <gml:usesParameterValue>
        <gml:value gml:uom="urn:ogc:def:uom:EPSG::9001">500000</gml:value>
        <gml:valueOfParameter xlink:href="urn:ogc:def:parameter:EPSG::8806"/>
      </gml:usesParameterValue>
      <gml:usesParameterValue>
        <gml:value gml:uom="urn:ogc:def:uom:EPSG::9001">-5000000</gml:value>
        <gml:valueOfParameter xlink:href="urn:ogc:def:parameter:EPSG::8807"/>
      </gml:usesParameterValue>
    </gml:Conversion>
  </gml:definedByConversion>
  <gml:usesCartesianCS>
    <gml:CartesianCS gml:id="ogrcrs77">
      <gml:csName>Cartesian</gml:csName>
      <gml:csID>
        <gml:name gml:codeSpace="urn:ogc:def:cs:EPSG::">4400</gml:name>
      </gml:csID>
      <gml:usesAxis>
        <gml:CoordinateSystemAxis gml:id="ogrcrs78" gml:uom="urn:ogc:def:uom:EPSG::9001">
          <gml:name>Easting</gml:name>
          <gml:axisID>
            <gml:name gml:codeSpace="urn:ogc:def:axis:EPSG::">9906</gml:name>
          </gml:axisID>
          <gml:axisAbbrev>E</gml:axisAbbrev>
          <gml:axisDirection>east</gml:axisDirection>
        </gml:CoordinateSystemAxis>
      </gml:usesAxis>
      <gml:usesAxis>
        <gml:CoordinateSystemAxis gml:id="ogrcrs79" gml:uom="urn:ogc:def:uom:EPSG::9001">
          <gml:name>Northing</gml:name>
          <gml:axisID>
            <gml:name gml:codeSpace="urn:ogc:def:axis:EPSG::">9907</gml:name>
```

```
          </gml:axisID>
          <gml:axisAbbrev>N</gml:axisAbbrev>
          <gml:axisDirection>north</gml:axisDirection>
        </gml:CoordinateSystemAxis>
      </gml:usesAxis>
    </gml:CartesianCS>
  </gml:usesCartesianCS>
</gml:ProjectedCRS>
```

WKT2 (ISO 19162:2019) Definition of an undefined CRS

This is an example of a CRS definition based on WKT2. Although it references some elements within EPSG, there is no need for a definition of those references, because they are properly described in the text itself.

WKT2 (ISO 19162:2019) – Georeference description of coordinate systems

```
PROJCRS["Slovenia 1996 / Slovene National Grid",
  BASEGEOGCRS["Slovenia 1996",
    DATUM["Slovenia Geodetic Datum 1996",
      ELLIPSOID["GRS 1980",6378137,298.257222101,
        LENGTHUNIT["metre",1,ID["EPSG",9001]],
        ID["EPSG",7019]],
      ID["EPSG",6765]],
    PRIMEM["Greenwich",0,
      ANGLEUNIT["degree",0.0174532925199433,ID["EPSG",9102]],
      ID["EPSG",8901]],
    ID["EPSG",4765]],
  CONVERSION["Slovene National Grid",
    METHOD["Transverse Mercator",
      ID["EPSG",9807]],
    PARAMETER["Latitude of natural origin",0,
      ANGLEUNIT["degree",0.0174532925199433,ID["EPSG",9102]]],
    PARAMETER["Longitude of natural origin",15,
      ANGLEUNIT["degree",0.0174532925199433,ID["EPSG",9102]]],
    PARAMETER["Scale factor at natural origin",0.9999,
      SCALEUNIT["unity",1,ID["EPSG",9201]]],
    PARAMETER["False easting",500000,
      LENGTHUNIT["metre",1,ID["EPSG",9001]]],
    PARAMETER["False northing",-5000000,
      LENGTHUNIT["metre",1,ID["EPSG",9001]]],
    ID["EPSG",19845]],
  CS[Cartesian,2,
    ID["EPSG",4400]],
  AXIS["Easting (E)",east,
    ORDER[1]],
  AXIS["Northing (N)",north,
    ORDER[2]],
  LENGTHUNIT["metre",1,ID["EPSG",9001]],
  USAGE[SCOPE[“Engineering survey, topographic mapping.”],
  AREA[“Slovenia - onshore and offshore.”],
  BBOX[45.42,13.38,46.88,16.61]],
ID["EPSG",3794]]
```

3.3 General (contextual) documentation

This part of the IP describes all remaining information about the geodata. Included here are links to relevant Documentation describing the lineage and provenance of the spatial data set. The Documentation consists of user manuals, related practices in the EU and worldwide, methodological rules, scientific articles, publications, etc.

3.4 Geospatial Metadata folder

When Geospatial data is described with a machine-readable .xml file within a GIS, and the XML schema is based on a standardised metadata format, we can use this content to harvest descriptive elements into archival metadata (EAD, ISAD(G), etc.). That is why we need to define a unique location for such files so that an automatic harvesting process knows where to look. A proposal for translation tables between the INSPIRE metadata schema and ISAD(G) and for the EAD3 schemas is presented in section 4.

In general, GIS systems offer the possibility to describe the geospatial dataset with additional technical and descriptive metadata. Depending on the system, the metadata structure could be proprietary and/or the metadata could adhere to local or global standards. Technical metadata can be derived from the dataset (geometry type, number of records, etc.). However, the descriptive metadata must be entered by the data creator or its manager. And this is the part that we should consider as part of the Information Package. In reality, we might encounter systems where geodata is not accompanied by adequate metadata.

Depending on the age of the GIS system, we might encounter different descriptive information or metadata, for each individual dataset:

- None or very limited (Metadata can be derived from separate Documentation).
- Metadata is available in separate files that are not in a machine-readable format (.pdf, unstructured .txt, etc.).
- Metadata is available in machine-readable format (.xml or other).

Geodata-specific metadata commonly contains descriptions that are specific to geodata (CRS info, bounding coordinates, scale, etc.) and would not be easily added to archival contextual metadata formats like ISAD(G) or EAD without extending the schema. So, we recommend that we store this geospatial metadata as the content of the data. It can be used to harvest elements from geospatial metadata, into archival descriptive metadata of the Information Package (GeoIP), during the ingest process.

We recommend that the producer provides metadata based on existing metadata standards for Geodata (ISO 19115:2003 or later)[24] or its adoption by the European Directive INSPIRE.

According to the mandatory elements, that are used in describing datasets by the INSPIRE directive, the following metadata elements must present within the descriptive metadata structure for each geospatial dataset, which is present in the data folder.

Table 1: Proposed metadata elements

Nr	Inspire name	Inspire definition
1	**Resource title**	Name by which the cited resource is known
2	**Resource abstract**	Brief narrative summary of the content of the resource(s)
3	**Resource type**	Scope to which metadata applies. This is the type of resource being described by the metadata, and it is filled in with a value from a classification of the resource-based on its scope. The choice of Resource Type will be probably the first decision made by the user, and it will define the metadata elements that should be filled. (Example: dataset)

[24] https://www.iso.org/standard/26020.html

Nr	Inspire name	Inspire definition
4	**Unique resource identifier**	Value uniquely identifying an object within a namespace
5	**Resource language**	Language(s) used within the datasets
6	**Topic category**	Main theme(s) of the dataset. (Example: Hydrography, administrative areas, transportation, etc.)
7	**Keyword value**	Commonly used word(s) or formalised word(s) or phrase(s) used to describe the subject
8	**Originating controlled vocabulary**	Name of the formally registered thesaurus or a similar authoritative source of keywords
9	**Geographic bounding box**	Western-most coordinate of the limit of the dataset extent, expressed in longitude in decimal degrees (positive east). Eastern-most coordinate of the limit of the dataset extent, expressed in longitude in decimal degrees (positive east) Northern-most coordinate of the limit of the dataset extent, expressed in latitude in decimal degrees (positive north) Southern-most coordinate of the limit of the dataset extent, expressed in latitude in decimal degrees (positive south)
10	**Temporal extent**	Time period covered by the content of the dataset
11	**Date of publication**	Reference date for the cited resource - publication
12	**Date of last revision**	Reference date for the cited resource - revision
13	**Date of creation**	Reference date for the cited resource - creation
14	**Lineage**	General explanation of the data producer's knowledge about the lineage of a dataset. This information can also reference other documents, that cover this description in greater detail explaining more about how and from what source datasets were created, the methodology of the process, etc.
15	**Spatial resolution**	• Equivalent scale: level of detail expressed as the scale denominator of a comparable hardcopy map or chart • Distance: ground sample distance
16	**Specification**	Citation of the product specification or user requirement against which data is being evaluated
17	**Limitations on public access [and use]**	Access constraints applied to assure the protection of privacy or intellectual property, and any special restrictions or limitations on obtaining the resource Limitations on public access: Access constraints – Example: Other Restrictions (limitation not listed)

Nr	Inspire name	Inspire definition
		Other constraints – Example: No limitations Classification – Example: unclassified
18	**Conditions applying to access and use**	Restrictions on the access and use of a resource or metadata
19	**Metadata date**	Date that the metadata was created
20	**Metadata language**	Language used for documenting metadata

To link metadata files with the individual geospatial dataset (file), the name of the XML file should be the same name as the geospatial dataset (In case of GML). However, if the metadata file references more than one file (in case of several raster datasets constituting a single map or imagery dataset), the name can be more general.

Examples of INSPIRE Metadata can be found on the INSPIRE Geoportal.[25]

[25] https://inspire-geoportal.ec.europa.eu/overview.html?view=thematicEuOverview&theme=none

4 Example structures of a Geodata Information Package (GeoIP)

This section presents examples of Information Package structures containing geodata archival package elements. Examples follow the top-level structure of a Common Specification for Information Packages[26] (CSIP), containing at least one representation, that is contained within the Representations directory.

The CSIP Structure also allows for multiple representations of the same geodata in the same. In Figure 7, we can see other folders in green, that represent possible folders containing geodata record specific content. Geodata itself and the accompanying geodata specific metadata is put within existing CSIP folders.

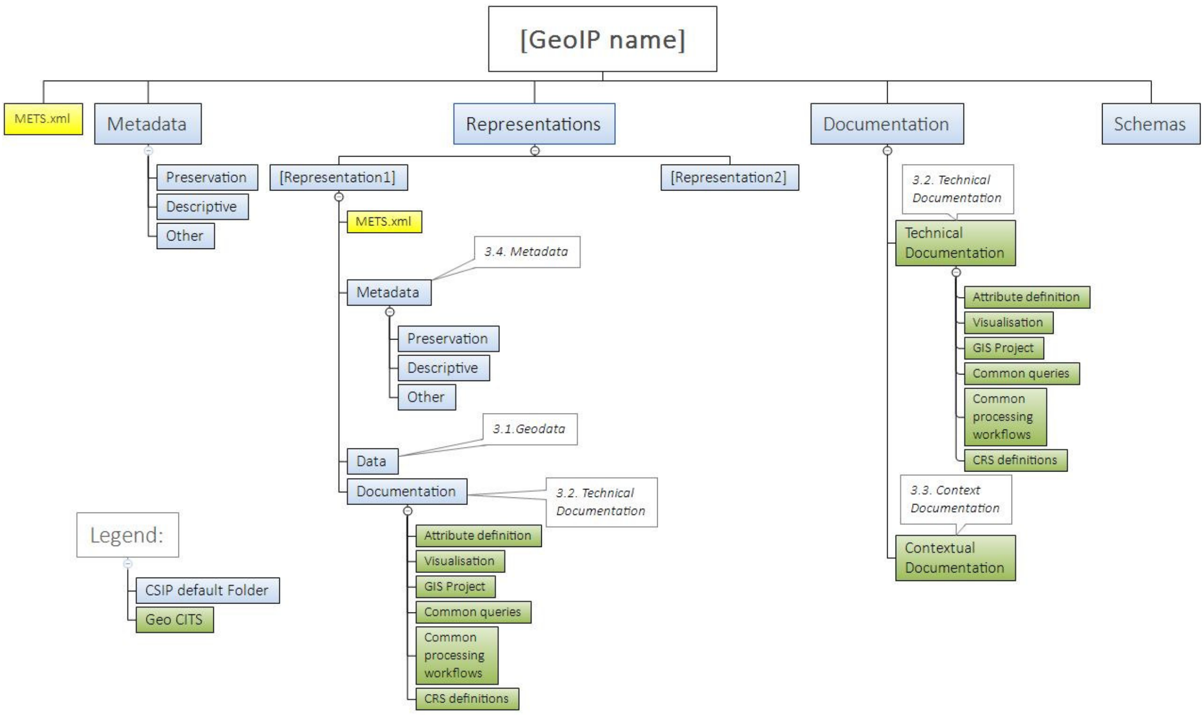


Figure 6: Common CSIP structure and additional GeoIP specific folders

It is possible to add other folders within the prescribed substructure following the principles of CSIP.

26 https://dilcis.eu/specifications/common-specification

4.1 GeoIP containing one Geodata set

Figure 7: Basic Geodata Information Package shows a basic structure of the information package, containing folders, as defined by the CSIP, in blue and the elements in green are described within this specification.

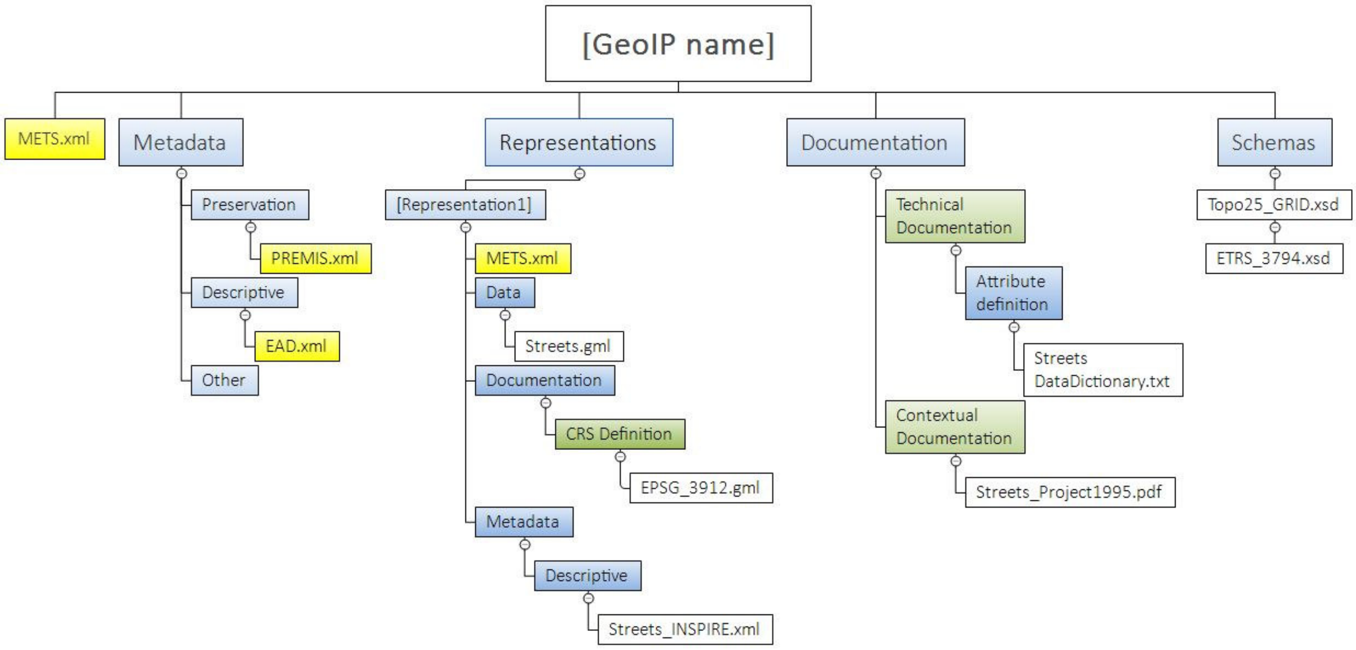


Figure 7: Basic Geodata Information Package

The Data folder (Data) contains only one representation of vector geodata in GML format (Streets.gml), which could be used as a long-term preservation format.

Technical Documentation is put in two different locations, as described at the beginning of section 3.2.:

- The machine-readable representation specific documentation is located in the “Documentation” folder within the representation (since the GML file points to an external CRS database, technical Documentation contains an element defining the CRS Definition.).
- The descriptive txt file defining the streets dataset is located in the root “Documentation” folder within the “Technical Documentation \ Attribute definition”.

Context Documentation is another folder within the root “Documentation” folder, that contains a .pdf document, describing additional content about the Streets Project from 1995.

4.2 GeoSIP containing multiple vector representations

In this case, a GeoSIP package contains an original representation of geodata in ESRI Shapefile format, and one representation in GML format, as a long-term preservation format. All other Documentation required to properly interpret both representations is put in the documentation folder (Figure 9) on the root level.

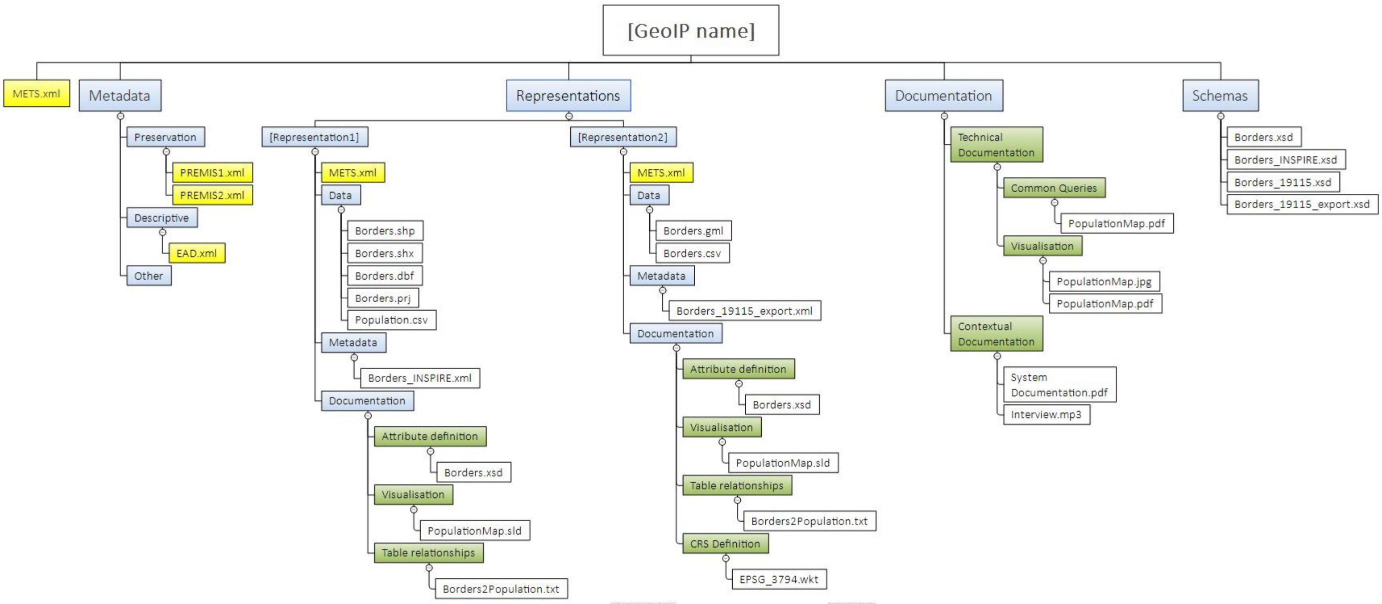


Figure 8: Geo Information package with two representations containing vector data

Key differences between the representations are:

- Geodata-specific metadata file (Borders_19115_export.xml) is updated to contain the information about conversion from the original format.
- Since the GML representation references an EPSG register of CRS, we need to add a separate definition of the CRS (ETRS_3794.gml) in the technical Documentation (exported from the EPSG registry). See example 3.2.7.1.
- Context documentation for both representations is stored in the Root Documentation folder.

We can also see that some Technical documentation, was put under the root “Documentation”. In addition, all xsd schemas, that belong to the XML-based files, are located within the “Schemas” folder.

4.3 GeoSIP containing one representation of multiple raster datasets

In this case, the IP contains one representation of multiple raster images covering an area with an accompanying vector file – containing positions of the raster images (Figure 10). Technical and Contextual Documentation for the raster datasets is located in the root "Documentation" folder. Except for the machine-readable CRS Definition and INSPIRE metadata which is placed close to its representation.

In the case of a large volume of data, we could split the data into multiple IPs described by its own METS document and also record the organisation of the split by modifying the accompanying GML file, to represent the amount of data within the IP.

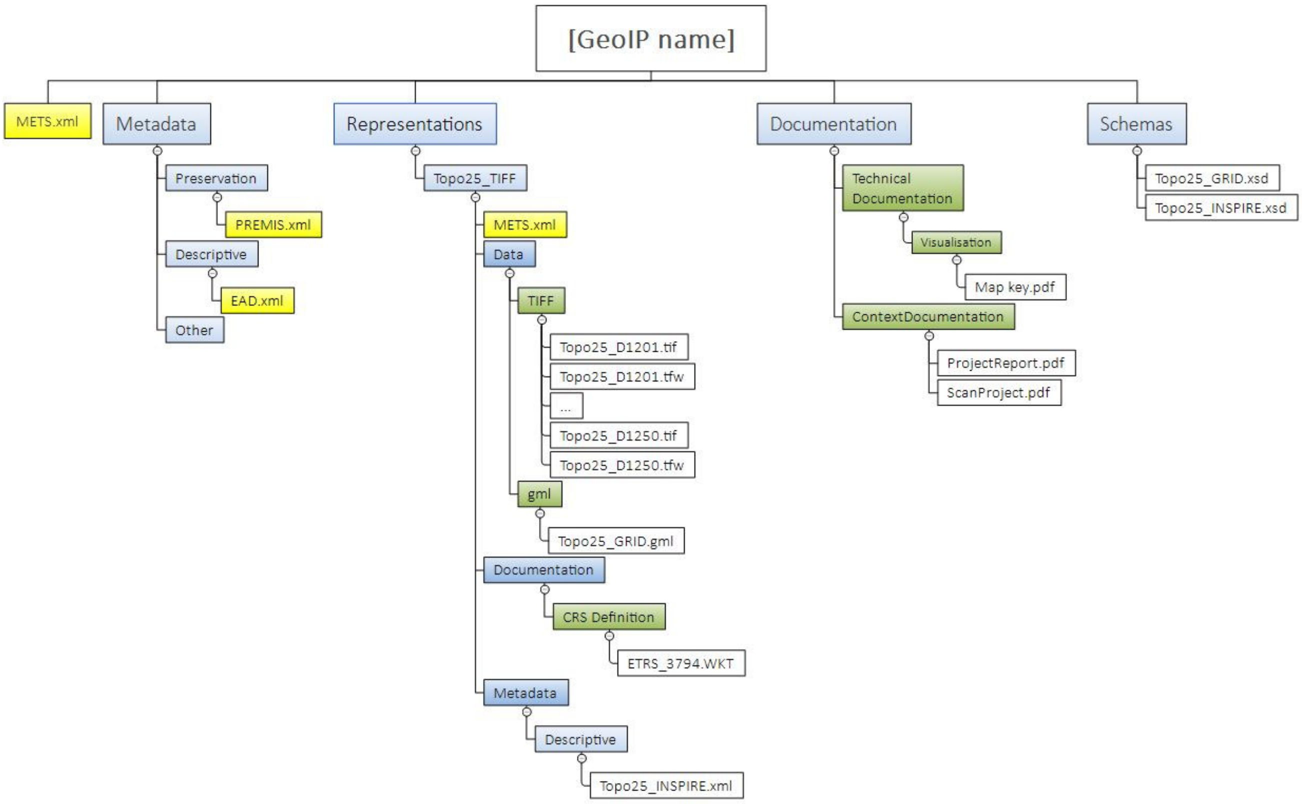


Figure 9: Folder structure of the GeoSIP containing one representation of multiple rasters

4.4 Proposed cardinality for elements in a GEO IP

In this section, we describe the expected number of occurrences of files within the different geo sections for an Information Package containing Geospatial Content.

Geodata elements

Table 2 – Occurrences of geo files in an Information Package

Information	Cardinality
Graphical information	1..n
Data attributes	1..n
Geo referencing information	1..n
Visualisation information	0..n
GIS Metadata[27]	0..n
Graphical information	1..n
Data attributes	0..n
Geo-referencing information	1..n
Visualisation information	0..n
Attribute definition	1..n
Object catalogue	0..n
Relationship to geodata	1..n

GIS System elements

If the Information Package contains a more complex GIS export, the following list of occurrences of files within additional geo sections should be added to the previous list.

Table 3: Added occurrences when it is a GIS System

Information	Cardinality
List of elements in a project	1..n
Object relations (geodata layers)	1..n
Geo-referencing transformations	1..n
Data layer properties	0..n
Labelling	0..n
Map visualisation	0..n
Common queries	1..n
Geoprocessing workflows	0..n
Common reports	0..n

[27] GIS Metadata refers to the information covered in section 3.4 of this document.

5 Metadata translations

5.1 Proposed translation schema for the INSPIRE metadata descriptions for geospatial resources in ISAD(G)

The following table displays the identified counterparts of the required INSPIRE metadata elements used in the ISAD(G) structure. Initial elements are based on the INSPIRE Metadata Implementing Rules: Technical Guidelines, based on EN ISO 19115 version 1.3. and INSPIRE Metadata Implementation Rules at:

http://inspire.jrc.ec.europa.eu/documents/Metadata/MD_IR_and_ISO_20131029.pdf.

Further information on ISAD-G elements can be found at:
https://www.ica.org/en/isadg-general-international-standard-archival-description-second-edition

Table 4: Proposed translation schema the INSPIRE metadata descriptions for geospatial resources in ISAD(G)

INSPIRE el. Nr.	INSPIRE el. Name	Explanation	Metadata data type	Proposed Cardinality	ISDG code	Comments
2.2.1	Resource title	Name by which the cited resource is known	text	1..1	3.1.2 Title	
2.2.2	Resource abstract	Brief narrative summary of the content of the resource(s)	text	0..1	3.3.1 Scope and content	
2.2.3	Resource type	Scope to which metadata applies	CodeList	0..1	3.1.5 Extent and medium of the unit of description	CodeList (see annex B.5.25 of ISO 19115)
2.2.5	Unique resource identifier	Value uniquely identifying an object within a namespace	text	0..1	3.1.1 Reference code	
2.2.6	Coupled resource	Provides information about the datasets that the service operates on	URI	0..n	3.5.3 Related units of description	
2.2.7	Resource language	Language(s) used within the datasets	CodeList	0..n	3.4.3 Language/scripts of material	LanguageCode (ISO/TS 19139)
2.3.1	Topic category (INSPIRE specific)	Main theme(s) of the dataset	CodeList	1..1	No ISAD-G element exists	List of values. See Part D 2 of the INSPIRE Metadata Regulation 1205/2008/EC) ISO19115:B.5.27 MD_TopicCategoryCode

INSPIRE el. Nr.	INSPIRE el. Name	Explanation	Metadata data type	Proposed Cardinality	ISDG code	Comments
2.3.2	Spatial data service type	A service type name from a registry of services	CodeList	0..1	No ISAD-G element exists	List of values. See section 1.3.1 in INSPIRE Metadata Implementing Rules
2.4.1	Keyword value	Commonly used word(s) or formalised word(s) or phrase(s) used to describe the subject	text	0..n	No ISAD-G element exists	descriptors
2.6.1	Temporal extent	Time period covered by the content of the dataset	Date	1..n	3.1.3 Date(s)	
2.6.2	Date of publication	Reference date for the cited resource – publication	Date	0..1	3.1.3 Date(s)	
2.6.3	Date of last revision	Reference date for the cited resource – revision	Date	0..1	3.1.3 Date(s)	
2.6.4	Date of creation	Reference date for the cited resource – creation	Date	0..1	3.1.3 Date(s)	
2.7.1	Lineage	General explanation of the data producer's knowledge about the lineage of a dataset	text	0..1	/	This element is Geodata specific, so we propose that searching using this criterion is done by using Geospatial metadata catalogues and not Archival catalogues or that data is added into "Scope and content" element (3.3.1. ISAD(G))
2.7.2	Spatial resolution	• Equivalent scale: level of detail expressed as the scale denominator of a comparable hardcopy map or chart • Distance: ground sample distance	text	0..1	/	Same as with Lineage
2.9.1	Limitations on public access [and use]	Access constraints applied to assure the protection of privacy or intellectual property, and any special restrictions or limitations on obtaining the resource	CodeList	1..1	3.4.1 Conditions governing access	
2.9.2	Conditions applying to access and use	Restrictions on the access and use of a resource or metadata	text	1..1	**3.4.1 Conditions governing access;** 3.4.2 Conditions governing reproduction; 3.4.4 Physical characteristics and technical requirements	

5.2 Proposed translation between INSPIRE metadata elements and EAD3 metadata elements

The following table displays the identified counterparts of the mandatory INSPIRE metadata elements in the EAD3 structure. Initial elements are based on the INSPIRE Metadata Implementing Rules.: Technical Guidelines, based on EN ISO 19115 version 1.3. and INSPIRE Metadata Implementation Rules at: http://inspire.jrc.ec.europa.eu/documents/Metadata/MD_IR_and_ISO_20131029.pdf.

Further information on EAD elements can be found in the Tag Library for EAD available at: https://www.loc.gov/ead/EAD3taglib/EAD3.html

Table 5: Proposed translation schema of INSPIRE metadata descriptions in EAD3

INSPIRE el. Nr.	INSPIRE element name	Explanation	Proposed Cardinality	EAD3 mapping	EAD description	Comments
2.2.1	Resource title	Name by which the cited resource is known	1..1	<unittitle>	**<unittitle>** is for recording the title statement, either formal or supplied, of the described materials. The title statement may consist of a word or phrase. <unittitle> is used at both the highest unit or <archdesc> level (e.g., collection, record group, or fonds) and at all the subordinate <c> levels (e.g., subseries, files, items, or other intervening stages within a hierarchical description).	
2.2.2	Resource abstract	Brief narrative summary of the content of the resource(s)	0..1	<scopecontent>	**<scopecontent>** contains a narrative statement that summarises the range and topical coverage of the materials. It provides the researcher with the information necessary to evaluate the potential relevance of the materials being described. <scopecontent> may include information about the form and arrangement of the materials; dates covered by the materials; significant organisations, individuals, events, places, and subjects represented in the materials; and functions and activities that generated the materials being described. It may also identify strengths of or gaps in the materials.	

INSPIRE el. Nr.	INSPIRE element name	Explanation	Proposed Cardinality	EAD3 mapping	EAD description	Comments
2.2.3	Resource type	Scope to which metadata applies. This is the type of resource being described by the metadata, and it is filled in with a value from a classification of the resource based on its scope. The choice of Resource Type will be probably the first decision made by the user, and it will define the metadata elements that should be filled. (Example: dataset)	0..1	<physdesc>	**<physdesc>** is a wrapper element for bundling information about the appearance or construction of the described materials, such as their dimensions, a count of their quantity or statement about the space they occupy, and terms describing their genre, form, or function, as well as any other aspects of their appearance, such as colour, substance, style, and technique or method of creation. The information may be presented as plain text, or it may be divided into the <dimension>, <extent>, <genreform>, and <physfacet> sub-elements.	**CodeList (see annexe B.5.25 of ISO 19115)**
2.2.5	Unique resource identifier	Value uniquely identifying an object within a namespace	0..1	<originalsloc>	**<originalsloc>** may be used to provide information about the location, availability, and/or destruction of originals.	If applicable
2.2.6	Coupled resource	Provides information about the datasets that the service operates on	0..n	<relatedmaterial> ; <arrangement>	**<relatedmaterial>** is used to identify associated materials in the same repository or elsewhere. These materials may be related by the sphere of activity or subject matter; Use **<arrangement>** to record the logical or physical groupings within a hierarchical structure and their relationships. This includes how the described materials have been subdivided into smaller units, e.g., record groups into series. May also indicate the filing sequence of the described materials, for example, chronological or alphabetical arrangement.	
2.2.7	Resource language	Language(s) used within the datasets	0..n	<langmaterial>	**<langmaterial>** records information about languages and scripts represented in the materials being described. <langmaterial> must contain one or more <language> or <languageset> elements, but cannot contain text.	LanguageCode (ISO/TS 19139)
2.3.1	Topic category (Specific to INSPIRE)	Main theme(s) of the dataset	1..1	<controlaccess> <subject> <part>	**<controlaccess>** An element that binds together elements containing access headings for the described materials. **<subject>** An element for encoding topics represented in the materials described. **<part>** A required and repeatable child of controlled access elements used to encode one or more parts of an access term.	A separate descriptor. List of values. See Part D 2 of the INSPIRE Metadata Regulation 1205/2008/EC) ISO19115:B.5.27

INSPIRE el. Nr.	INSPIRE element name	Explanation	Proposed Cardinality	EAD3 mapping	EAD description	Comments
						MD_TopicCategoryCode
2.3.2	Spatial data service type	A service type name from a registry of services	0..1	<controlaccess> <subject> <part>	**<controlaccess>** An element that binds together elements containing access headings for the described materials. **<subject>** An element for encoding topics represented in the materials being described. **<part>** A required and repeatable child of controlled access elements used to encode one or more parts of an access term.	A separate descriptor List of values. See section 1.3.1 in INSPIRE Metadata Implementing Rules
2.4.1	Keyword value	Commonly used word(s) or formalised word(s) or phrase(s) used to describe the subject	0..n	<controlaccess> <subject> <part>	**<controlaccess>** An element that binds together elements containing access headings for the described materials. **<subject>** An element for encoding topics represented in the materials being described. **<part>** A required and repeatable child of controlled access elements used to encode one or more parts of an access term.	Each keyword is given in its own <part> element.
2.5.1	Geographic bounding box	Western-most coordinate of the limit of the dataset extent, expressed in longitude in decimal degrees (positive east). Eastern-most coordinate of the limit of the dataset extent, expressed in longitude in decimal degrees (positive east) Northern-most coordinate of the limit of the dataset extent, expressed in latitude in decimal degrees (positive north) Southern-most coordinate of the limit of the dataset extent, expressed in latitude in decimal degrees (positive north).	0..n	<geographiccoordinates>	Use **<geographiccoordinates>** to express a set of geographic coordinates such as latitude, longitude, and altitude representing a point, line, or area on the surface of the earth.	

INSPIRE el. Nr.	INSPIRE element name	Explanation	Proposed Cardinality	EAD3 mapping	EAD description	Comments
2.6.1	Temporal extent	Time period covered by the content of the dataset	1..n	<unitdate>; <unitdatestructured>	**<unitdate>** is for indicating the date or dates the described materials were created, issued, copyrighted, broadcast, etc. <unitdate> may be in the form of text or numbers, and may consist of a single date, a date range, or a combination of single dates and date ranges; **<unitdatestructured>** provides a machine-processable statement of the date or dates the materials described were created, issued, copyrighted, broadcast, etc. <unitdatestructured> must contain one of the following child elements: <datesingle>, <daterange>, or <dateset>. <unitdatestructured> may contain only one child; therefore, <dateset> must be used in situations where complex date information needs to be conveyed and requires at least two child elements. A date set may combine two or more <datesingle> and <daterange> elements.	Use the relevant subelements

INSPIRE el. Nr.	INSPIRE element name	Explanation	Proposed Cardinality	EAD3 mapping	EAD description	Comments
2.6.2	Date of publication	Reference date for the cited resource - publication	0..1	<unitdate>, <unitdatestructured>	**<unitdate>** is for indicating the date or dates the described materials were created, issued, copyrighted, broadcast, etc. <unitdate> may be in the form of text or numbers, and may consist of a single date, a date range, or a combination of single dates and date ranges; **<unitdatestructured>** provides a machine-processable statement of the date or dates the materials described were created, issued, copyrighted, broadcast, etc. <unitdatestructured> must contain one of the following child elements: <datesingle>, <daterange>, or <dateset>. <unitdatestructured> may contain only one child; therefore <dateset> must be used in situations where complex date information needs to be conveyed and requires at least two child elements. A date set may combine two or more <datesingle> and <daterange> elements.	Use the relevant subelements

INSPIRE el. Nr.	INSPIRE element name	Explanation	Proposed Cardinality	EAD3 mapping	EAD description	Comments
2.6.3	Date of last revision	Reference date for the cited resource - revision	0..1	<unitdate>, <unitdatestructured>	**<unitdate>** is for indicating the date or dates the described materials were created, issued, copyrighted, broadcast, etc. <unitdate> may be in the form of text or numbers, and may consist of a single date, a date range, or a combination of single dates and date ranges; **<unitdatestructured>** provides a machine-processable statement of the date or dates the materials described were created, issued, copyrighted, broadcast, etc. <unitdatestructured> must contain one of the following child elements: <datesingle>, <daterange>, or <dateset>. <unitdatestructured> may contain only one child; therefore, <dateset> must be used in situations where complex date information needs to be conveyed and requires at least two child elements. A date set may combine two or more <datesingle> and <daterange> elements.	Use the relevant subelements

INSPIRE el. Nr.	INSPIRE element name	Explanation	Proposed Cardinality	EAD3 mapping	EAD description	Comments
2.6.4	Date of creation	Reference date for the cited resource - creation	0..1	<unitdate>, <unitdatestructured>	**<unitdate>** is for indicating the date or dates the described materials were created, issued, copyrighted, broadcast, etc. <unitdate> may be in the form of text or numbers, and may consist of a single date, a date range, or a combination of single dates and date ranges; **<unitdatestructured>** provides a machine-processable statement of the date or dates the materials described were created, issued, copyrighted, broadcast, etc. <unitdatestructured> must contain one of the following child elements: <datesingle>, <daterange>, or <dateset>. <unitdatestructured> may contain only one child; therefore, <dateset> must be used in situations where complex date information needs to be conveyed and requires at least two child elements. A date set may combine two or more <datesingle> and <daterange> elements.	Use the relevant subelements
2.7.1	Lineage	General explanation of the data producer's knowledge about the lineage of a dataset, covering the history of data creation and methodologies used during the process	0..1	<scopecontent>	It is a separate element in INSPIRE; however, it would best fit as an addition of the **<scopecontent>**. So we recommend appending it to this element.	
2.7.2	Spatial resolution	• Equivalent scale: level of detail expressed as the scale denominator of a comparable hardcopy map or chart • Distance: ground sample distance	0..1	<controlaccess>	It is a separate element, that is often used as a search criterion within Geospatial Metadata repositories. Therefore we suggest adding it as a search term through **<controlaccess>**.	

INSPIRE el. Nr.	INSPIRE element name	Explanation	Proposed Cardinality	EAD3 mapping	EAD description	Comments
2.9.1	Limitations on public access [and use]	Access constraints applied to assure the protection of privacy or intellectual property, and any special restrictions or limitations on obtaining the resource Limitations on public access: Access constraints - Example: otherRestrictions (limitation not listed). Other constraints - Example: No limitations Classification - Example: unclassified	1..1	<accessrestrict> <legalstatus>	Record in **<accessrestrict>** information about the availability of the described materials, whether due to the nature of the information in the materials being described, the physical condition of the materials, or the location of the materials. Examples include restrictions imposed by the donor, legal statute, repository, or other agency, as well as the need to make an appointment with repository staff. May also indicate that the materials are not restricted; Use **<legalstatus>** to identify the status of the material being described as defined by law, for example, the Public Records Act of 1958 in the United Kingdom.	This element has three separate elements that need to be described: • General set of limitations (ISO 19115 B.5.24) • Description of the limitation • Level of confidentiality (ISO 19115 B.5.11)
2.9.2	Conditions applying to access and use	Restrictions on the access and use of a resource or metadata	1..1	<accessrestrict>; <userestrict>	Record in **<accessrestrict>** information about the availability of the described materials, whether due to the nature of the information in the materials being described, the physical condition of the materials, or the location of the materials. Examples include restrictions imposed by the donor, legal statute, repository, or other agency, as well as the need to make an appointment with repository staff. May also indicate that the materials are not restricted; Use **<userestrict>** for information about any limitations, regulations, or special procedures imposed by a repository, donor, legal statute, or other agency. These conditions may be related to reproduction, publication, or quotation of the described materials after access to the materials has been granted. <userestrict> may also be used to indicate the absence of restrictions, such as when intellectual property rights have been dedicated to the public.	

6 Appendix 1

AUTHOR(S)	
Name(s)	Organisation(s)
Gregor Završnik	Geoarh

REVIEWER(S)	
Name(s)	Organisation(s)
Janet Anderson	Highbury IVS

Project co-funded by the European Commission within the ICT Policy Support Programme **Dissemination Level**		
P	**Public**	X
C	**Confidential, only for members of the Consortium and the Commission Services**	

REVISION HISTORY AND STATEMENT OF ORIGINALITY

Submitted Revisions History

Revision No.	Date	Authors(s)	Organisation	Description
0.1	31 October 2018	Gregor Završnik	Geoarh	Draft outline based on SFSB SMURF document.
1.0	20 December 2018	Gregor Završnik	Geoarh	
2.0	31 May 2019	Gregor Završnik	Geoarh	Changes introduced based on received comments from the users.
2.0.2	12 August 2020	Gregor Završnik	Geoarh	Changes introduced based on received comments from the users: – CRS Definition added to technical documentation if referenced externally. – GeoIP schemas aligned with CSIP structure. – Additions to the glossary. – Images changed to improve understanding.
2.0.4.	30 September 2020	Gregor Završnik	Geoarh	Changes introduced based on received comments from the users: – GeoIP schemas aligned with CSIP structure. – Examples added for technical documentation elements.

Statement of originality:
This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.