Guideline for CITS Geospatial specification with GIS

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

Date: 13.12.2024
Version: 1.1.0

Guideline for CITS Geospatial specification with GIS

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

The European Commission eArchiving procurement recognizes the E-ARK specifications as the eArchiving specifications which are funded under the eArchiving Common Services Platform Agreement No. LC-01905904-CNECT/LUX/2021/OP/0077.

This specification is published, supported, and developed by the Digital Information LifeCycle Interoperability Standards (DILCIS) Board under the auspices of the DLM Forum.

This specification was previously developed with the support of the European Union:

E-ARK Grant No: 620998 CIP-ICT-PSP.2013.2.5
E-ARK4ALL Agreement No. LC-00921441 CEF-TC-2018-15
E-ARK3 Agreement No. LC-01390244 CEF-TC-2019-3

1 Preface

1.1 Aim of the specification

This document is one of several related specifications which aim to provide a common set of usage descriptions of international standards for packaging digital information for archiving purposes. These specifications are based on common, international standards for transmitting, describing and preserving digital data. They also utilise the Reference Model for an Open Archival Information System (OAIS), which has Information Packages as its foundation. Familiarity with the core functional entities of OAIS is a prerequisite for understanding the specifications.

The specifications are designed to help data creators, software developers, and digital archives to tackle the challenge of short-, medium- and long-term data management and reuse in a sustainable, authentic, cost-efficient, manageable and interoperable way. A visualisation of the current specification network can be seen here:

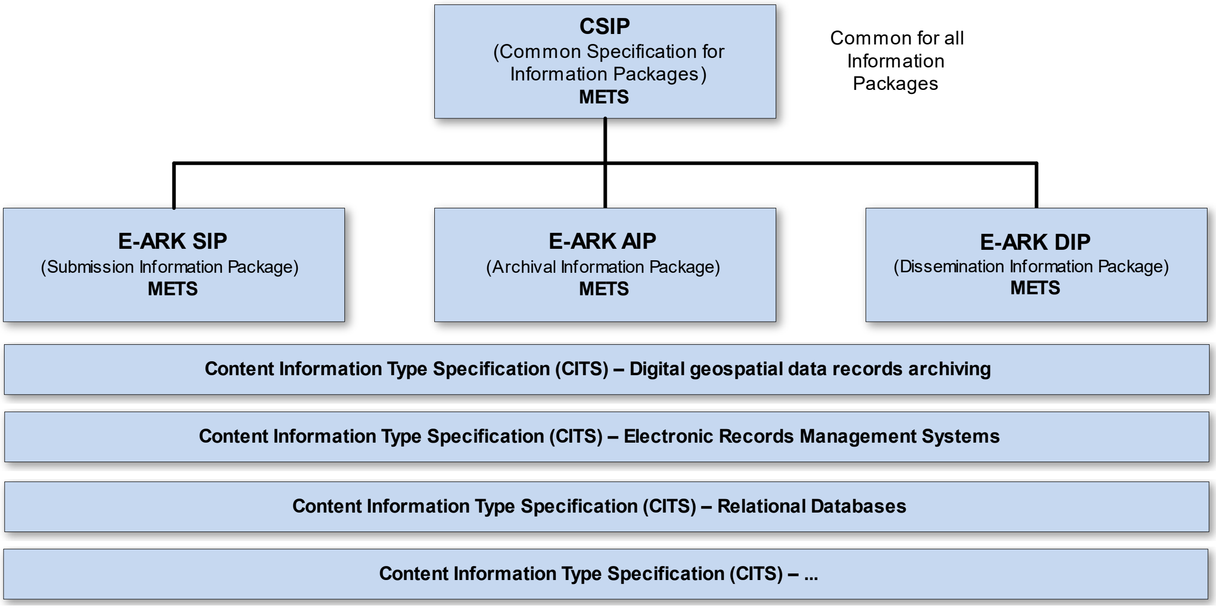

Figure I: Diagram showing E-ARK specification dependency hierarchy. Note that the image only shows a selection of the published CITS and isn't an exhaustive list.

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

Specification	Aim and Goals
	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 eArchiving Reference Implementation (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. For a list of existing Content Information Type Specifications see the DILCIS Board webpage (DILCIS Board, http://dilcis.eu/).

1.2 Organisational support

This specification is maintained by the Digital Information LifeCycle Interoperability Standards Board (DILCIS Board, http://dilcis.eu/). The role of the DILCIS Board is to enhance and maintain the draft specifications developed in the European Archival Records and Knowledge Preservation Project (E-ARK project, http://eark-project.com/), which concluded in January 2017. The Board consists of eight members, but no restriction is placed on the number of participants taking part 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. The DILCIS Board have been responsible for providing the core specifications to the Connecting Europe Facility eArchiving Building Block https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eArchiving/.

1.3 Authors & Revision History

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

TABLE OF CONTENTS

Table of Figures

1 Context

1.1 Purpose

The purpose of this document is to extend the "Guideline for the Specification for the E-ARK Content Information Type Specification for digital geospatial data records archiving (CITS Geospatial)" with content describing preservation of selected elements from Geographical Information Systems (GIS). This document aims to extend the scope of preservation beyond the geospatial data records themselves and focus more on GIS elements defining geospatial information products.

The purpose of preserving digital records from digital record management systems is to ensure access to its digital information products to any potential users in the future. Geospatial digital products are derived using the functionality of a GIS system using the underlying geospatial data records. Hence, the mere preservation of its records does not ensure the preservation of the digital information products produced by the system.

A common approach in digital preservation is also the simulation of existing systems in a virtual environment. With the rising data-driven economy and its use of big data and Artificial Intelligence tools, we also need to think of future users that are not directly human but rely on automated access. Therefore, we aim to define the elements of the system in a way that would enable its reuse by automated means.

Due to the increased complexity of some systems and their proprietary nature, a perfect system description is not always possible. However, we should aim to cover the most significant number of cases possible.

To achieve this in a Geographical Information System, we need to preserve more than just the geospatial records that they are based on. To reproduce digital products in the forms of different types of maps (digital maps for printing, web services, web applications, web mashups, etc.), lists, tables, or new geospatial records in the form of vector, raster, or any other kind of geospatial record, we also need to preserve the recipe for the digital product. The elements of the recipe consist of the context in which data was created and used, different types of rendering rules and configurations, a definition of how data was structured and of course the behaviour, the logic on how it was manipulated to produce the digital information product.

1.2 Scope

This guideline aims to describe the approach of preserving the elements defining structure, rendering, context and behaviour in Geographical Information Systems. However, it is far from complete and will be developed in the future, based on inputs from the geospatial and digital preservation communities.

Since the Geospatial domain is getting more and more complex and interwoven with other systems, this document aims to cover the approaches of preserving the technologies from the oldest to the newest, since the old ones are in the greatest danger of being lost.

The aim of preservation of Geographical Information Systems is in preserving the ability to reproduce an information product in the future, that is based on geospatial data. The level of replication depends on the purpose of the information product, which is usually determined in the appraisal process together with an expert from the archival and preservation domain and the subject matter expert from the data producer.

This Guideline builds on the CITS Geospatial specification. These specifications build on work done in the E-ARK projects.

1.3 Structure of the document

Section 2 contains an introductory section describing GIS and its variations in general and the approaches to its digital preservation. It also includes a recommended reading list for further interest in the topic. This section is meant for colleagues who are new to the field of GIS.

Section 3 provides a rationale and examples of the requirements found in the CITS Geospatial specification, which are GIS specific. This chapter aims to provide a better basis for understanding the requirements when preserving GIS elements. This section is primarily meant for technicians and developers of local specifications. The prerequisite is that the reader knows about:

- Common Structure of Information Packages documents (CSIP) the SIP, AIP and DIP specifications
- CITS Geospatial specification and Guideline for CITS Geospatial specification
- geospatial data and its structures
- basic GIS functionalities (data structuring and modelling, rendering, analysis, data processing functions, etc.)
- geospatial web services (various types and functionalities)

2 Introduction to Geographic Information Systems and approach to its preservation

This section contains an introductory section describing the basics of GIS, its variations and its output types in general. In the second part, there are some approaches to support the digital preservation of systems. In the end, the document also includes a recommended reading list for further interest in the topic. This section is intended for those who are new to the field of GIS.

2.1 Introduction to Geographical Information Systems

A Geographic Information System (GIS)[1] is a conceptualised framework that provides the ability to capture and analyse spatial and geographic data. GIS applications (or GIS apps) are computer-based tools that allow the user to create interactive queries (user-created searches), store and edit spatial and non-spatial data, analyse spatial information output, and visually share the results of these operations by presenting them as maps or other information products (lists, schemas, 3d views, etc.).

Figure 1 - Information products produced from geospatial data using GIS

To create an information product, we need to **structure** spatial and non-spatial data and **render** it using GIS functionality and settings. We can also use the **behaviour** of a GIS, to create, edit, transform, analyse, or convert geospatial data.

[1] Source - https://en.wikipedia.org/wiki/Geographic_information_system

2.2 Types of GIS Systems

Depending on the age and complexity, there are different types of GIS Systems that we can expect to encounter when preserving them.

2.2.1 Stand-alone GIS

A stand-alone GIS is the most basic type of GIS that has the following characteristics:

- It was used on a desktop computer or a workstation
- Data was generally stored locally, on a network drive or in a local database
- It was mostly managed by one or a couple of users
- It could contain multiple complex projects that are often not well organised and documented

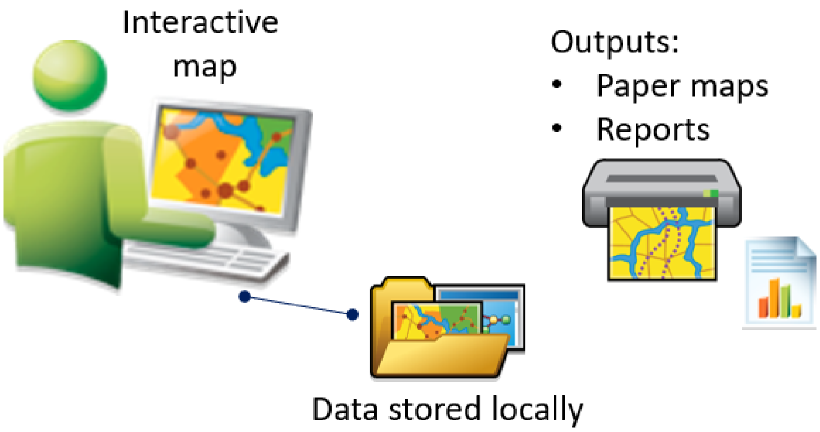

Figure 2 - Stand Alone GIS

2.2.2 Enterprise GIS

Enterprise systems generally serve the whole organisation and are therefore more complex.

- The system comprises one or more GIS servers providing shared access to data which implies possible shared editing and managing workflows.
- Users access it from various environments (desktop clients, Web applications, mobile applications).
- Data is generally stored in an enterprise database; however, we can also expect some local records if subprojects are stored locally. In large organisations, we can encounter multiple departmental databases connected to the central database.
- This type of system is used by a variety of users, from administrators to power users and basic users. A larger organisation is more likely to have developed standard GIS management workflows, which could help when archiving such a system.
- Enterprise systems can provide public access to Web and mobile applications by serving geospatial services. Therefore, it is important, to consider how external users used this data and if this impacts the preservation process.

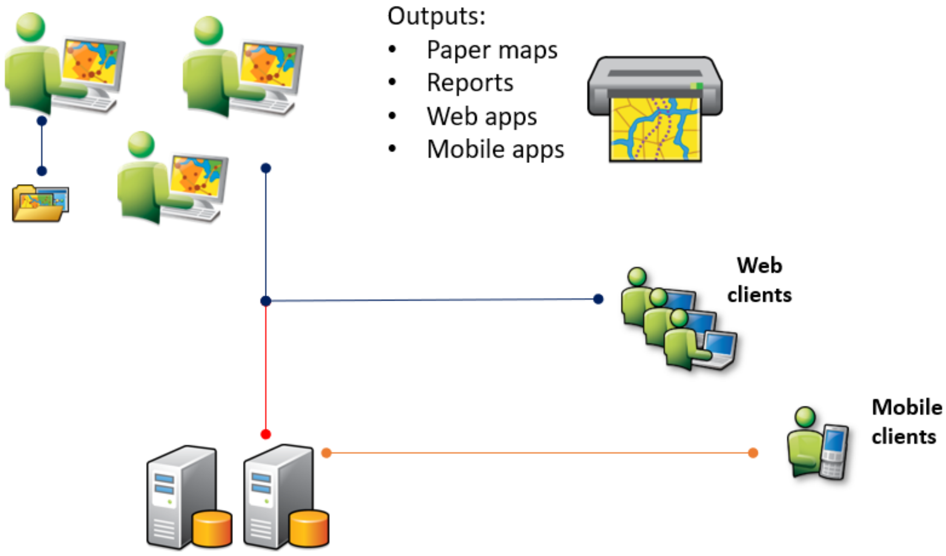

Figure 3 - Enterprise GIS system

As mentioned above, Enterprise GIS can make geospatial records accessible in multiple ways, which defines the format of geospatial records (As a file server, from a relational database or as a geospatial web service).

2.2.2.1 File server access to geospatial records

Geospatial records are stored on a shared file server, which can help manage access and rights to geospatial records. The geospatial records are usually stored as files in typical geospatial formats. From the preservation standpoint, we just preserve the geospatial records and convert them into long-term preservation formats for future use.

Some file formats also behave like databases. Examples are Geopackage[2] (GPKG), which uses SQLite database as a container and ESRI file Geodatabase[3], which uses a folder containing a set of xml and binary files. Some older GIS also uses MS Access as a container for geospatial records. Preservation of these records is more complex and can be approached from the standpoint of preserving relational databases or exporting geospatial records into long-term preservation formats for vector and raster. In that case, an estimation is needed as to which elements defining structure and behaviour should be preserved from the database.

2.2.2.2 Geospatial records from Enterprise Relational Databases

When geospatial datasets are loaded into databases, they can change their format and extent. Data can be split in smaller elements and column datatypes might be different than in original format. The benefit of this approach is to integrate the geospatial records with other tables in an information system.

[2] Geopackage - https://www.geopackage.org/
[3] ESRI File Geodatabase - https://desktop.arcgis.com/en/arcmap/latest/manage-data/administer-file-gdbs/file-geodatabases.htm

In databases, geospatial records are often restructured and stored within the database format for geospatial data to allow better access. For example, a geospatial raster can be stored in a huge number of small tiles for quicker access. Vector data is often stored in proprietary or standardised database-specific formats (most common are Oracle Spatial[4], PostGIS[5], ArcSDE[6], etc.)

Preservation of geospatial records from Enterprise Raster databases needs special consideration since a lot of structure and behavioural elements are stored within the non-spatial tables. A preservation specialist should work together with the GIS specialist to determine which elements need to be preserved.

2.2.2.3 Geospatial Web Services

When Geospatial records are served as a service, they originate as a file-based geospatial record or a database stored record and are transformed to a geospatial web service using a geospatial server.

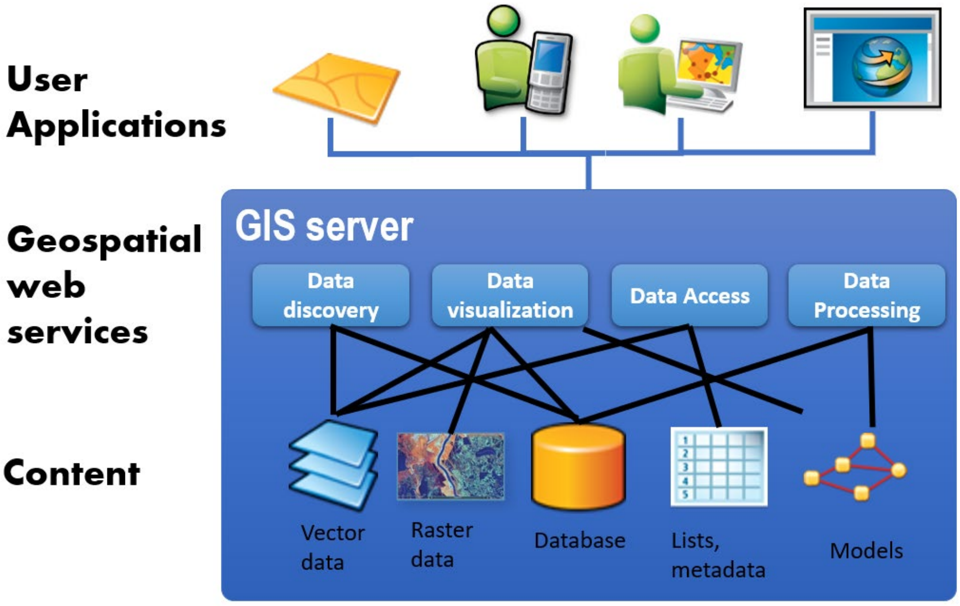

Figure 4 - Geospatial Web Services Types

There are four types of different geospatial web services:

- Catalogue services (serving Metadata catalogues for discovery of data, example solution is GeoNetwork)
- Geospatial data services (Serve data for use in GIS Systems and analysis, examples are WFS[7], WCS[8])

[4] Oracle spatial format - https://www.oracle.com/database/spatial/
[5] PostGIS - https://postgis.net/
[6] ArcSDE - https://en.wikipedia.org/wiki/ArcSDE
[7] https://www.ogc.org/standards/wfs
[8] https://www.ogc.org/standards/wcs

- Data Visualisation services (Merging data into maps, examples are WMS[9], WMTS[10])
- Web processing services (Offer online processing services for GIS analysis, examples are WPS[11] and WCPS[12])

Long term preservation of geospatial web services would differ depending on the service type.

- **Catalogue services** – preservation of source content – in this case, geospatial metadata. Rendering and behaviour are mostly dependent on the applications using the services, so they can be stored as additional documentation.
- **Geospatial data services**. – preservation of source content – in this case, geospatial data. Rendering and behaviour are mostly dependent on the applications using the services, so they can be stored as additional documentation.
- **Data Visualisation services** – preservation of source content with structure and rendering settings. In cases of pre-cached tiled services, a representative sample of tiles can be preserved or the whole cache as a temporary dissemination representation.
- **Web Processing services** – preservation of source content with structure and behaviour logic.

9 https://www.ogc.org/standards/wms
10 https://www.ogc.org/standards/wmts
11 https://www.ogc.org/standards/wps
12 https://www.ogc.org/standards/wcps

2.2.3 Distributed GIS

Distributed GIS [13]refers to GI Systems that do not have all the system components in the same physical location. This could be the processing, the database, the rendering, or the user interface. It represents a special case of distributed computing, with examples of distributed systems including web-based GIS and Mobile GIS. Distribution of resources provides corporate and enterprise-based models for GIS (involving multiple databases, different computers undertaking spatial analysis and a diverse ecosystem of often spatially enabled client devices). Distributed GIS permits a shared services model, including data fusion (or mashups) based on Open Geospatial Consortium (OGC) or proprietary web services.

Such a system cannot be practically preserved as a whole, as there are data producers or distributors for multiple different organisations. The approach to preserving it would require a comprehensive record of its data sources and its owners to enable the reuse of information products based on it. There are possible examples where organisations that have no geospatial data of their own issue official records based on external geospatial data. Therefore, in such an organisation, we can only preserve the structure of data references in an application producing geospatial information products.

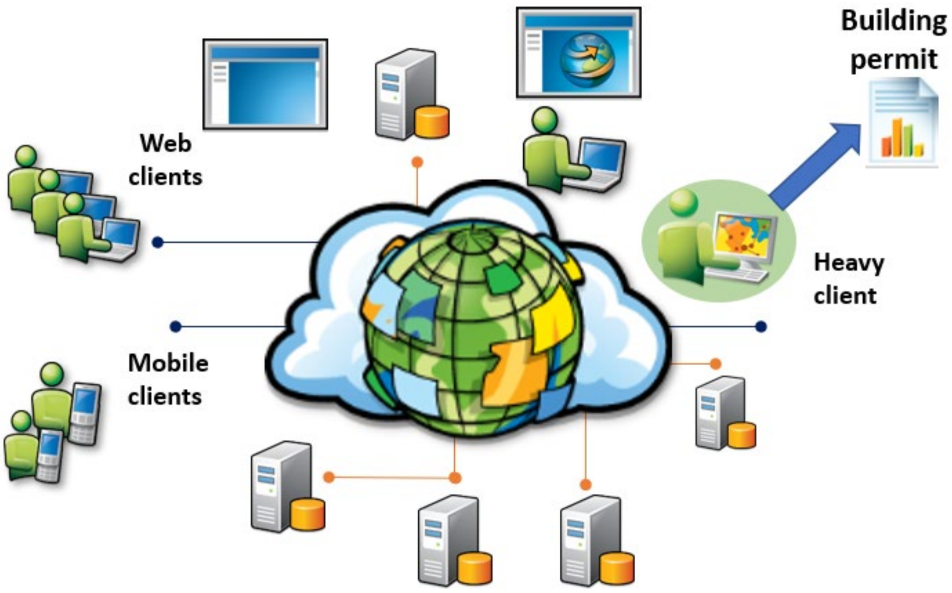

Figure 5 - Distributed GIS

An example of such an official record is a building permit, where in some countries, the organisation issuing it needs to access data from multiple, separately managed, official GIS Databases to produce this information product.

[13] Distributed GIS definition - https://en.wikipedia.org/wiki/Distributed_GIS

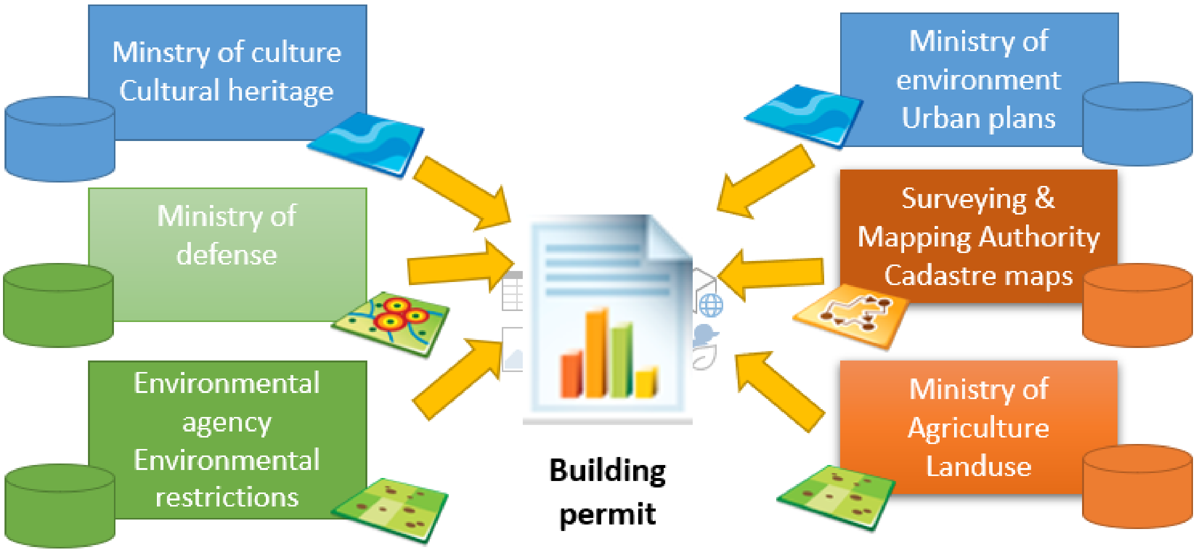

Figure 6 - Example of an information product based on Distributed GIS System

2.2.4 Standards for Geospatial Data

There are many standards covering Geospatial Information, coming from different standardisation bodies:

OGC[14] (Open GIS Consortium) is an international community that is committed to improving access to geospatial or location information. The organisation represents over 500 businesses, government agencies, research organisations, and universities united with a desire to make location information FAIR – Findable, Accessible, Interoperable, and Reusable. The community creates free, publicly available geospatial standards that enable new technologies. Some of the most commonly referenced OGC standards in the CITS Geospatial specification are geospatial formats GML and GeoTIFF, and others[15].

ISO TC/211[16] **(ISO Technical Committee 211)**
This committee is dedicated to standardisation in the field of digital geographic information. ISO Standards referring to Geographic Information are within the ISO 191XX family. The CITS Geospatial mostly refers to standards concerning geospatial metadata (ISO 19115-1;19115-2;19110;19157 and others). The most relevant ISO standard for this specification is, of course the ***ISO 19165-1:2018 Geographic Information – Preservation of digital data and metadata – Fundamentals***

[14] Open GIS Consortium (https://www.ogc.org/)
[15] List of OGC Standards (https://www.ogc.org/docs/is)
[16] ISO TC/211 (https://committee.iso.org/home/tc211)

2.3 GIS Preservation strategies

It used to be true that the data went to the archives to die. The preservation process of old systems is often tedious and costs a lot of unavailable resources from both sides (the archives and the data producers). And the result might end up being useless if proper considerations regarding preservation were not taken ahead of time. In the digital world, systems, formats, hardware and knowledge changes much faster than in the paper world, so when preserving digital records, we also need to preserve data in shorter timeframes or even at its creation. Data is considered a commodity; digital data preservation is turning the linear process into cyclical. Thus, it makes sense that we include a preservation step in the data life cycle management of a GIS.

2.3.1 Preservation of old/discontinued systems

Currently, the greatest danger of losing geospatial records lies in backed up old and discontinued systems. Organisations that have those systems often do not have enough knowledge or awareness to preserve the geospatial data or the systems themselves. Additionally, the archivists and digital data preservation experts often do not have enough knowledge to adequately assess key elements in GIS.

2.3.1.1 Guidelines for Archivists

The key first steps for any national or local archive are:

- **Cooperate with a Geospatial expert:** The inclusion of a geospatial expert in the assessment process is crucial for a proper understanding of the real value of geospatial records and the key elements of the system you need to preserve. They are more likely to anticipate possible needs for the reuse of the archived data.
- **Start with the base maps**: If the archive has just started with the accessions of geospatial records, the best organisation is to start with the base maps (Topographic maps, Aerial imagery, base datasets), commonly produced by a National Mapping agency. Base maps are present in most of the GIS projects in all other organisations. By preserving them upfront and having a “Fond” to reference, we will decrease the duplication of the same data in future accessions.
- **Base your strategy on standards:** It is common to start developing solutions from scratch when we do not have any. However, in the geospatial community, there are many existing standards and open source tools that can be used to establish a preservation platform. When designing your own strategies and guidelines, examine the existing standards for distribution and preservation.
- **Gather documentation with the producer/records manager:** Archivists can benefit by using interviews as a data gathering technique. Written documentation can be augmented using audio interviews with existing GIS users and managers, as they offer a faster gathering approach and save time to both parties. It also offers two-way communication, where the producers learn more about the preservation process and can therefore help the archive in the future.

2.3.2 Preservation in new and existing GIS systems

A proactive approach is demonstrated here to yield better results and limit the amount of time needed for the process of preservation and give us better results.

If we include a preservation step in the digital data lifecycle process, as shown in Figure 7 below, we can automate most of the data and documentation gathering and conversion processes.

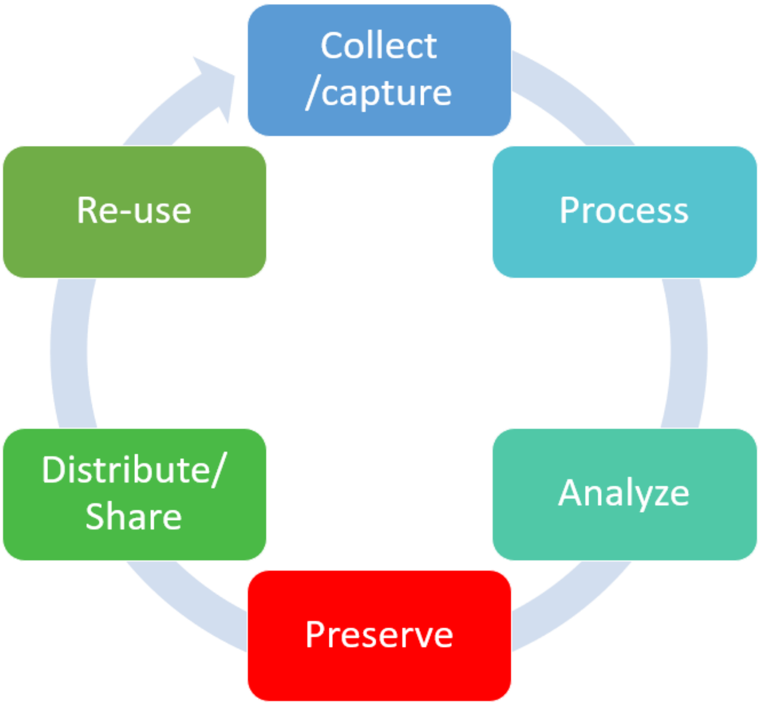

Figure 7 - Preservation in data life cycle process

2.3.2.1 Guidelines for Geospatial data producers and record managers

To achieve the proactive preservation of geospatial records and systems perform the following:

- Include preservation requirements when ordering an upgrade or a new GIS system. Let the solution provider provide as much of the documentation required for preservation as possible. This documentation should be included in the preservation step.
- Create a preservation copy of your data before sharing or publishing it in your distribution system
- Submit geospatial records to archives every five years or anticipate additional preservation steps (change of formats)
- Enforce population of metadata to ensure context documentation and support later data discovery

2.4 Recommended reading list

This section provides a recommended reading list for those interested in the preservation of geodata from GIS.

Introduction to GIS

- History of GIS (https://gisgeography.com/history-of-gis/)
- Wikipedia - Geographic information system (https://en.wikipedia.org/wiki/Geographic_information_system)
- Geospatial formats - https://gisgeography.com/gis-formats/

Preservation of geodata

- Descriptions of geospatial formats and their suitability for long term preservation https://www.loc.gov/preservation/digital/formats/content/gis.shtml
- Description of Significant properties concept. https://significantproperties.kdl.kcl.ac.uk/
- The book: Preservation in Digital Cartography: https://link.springer.com/book/10.1007/978-3-642-12733-5
- Other available guidelines: https://github.com/DILCISBoard/CITS-Geospatial/wiki/Related-guidelines
-

3 GIS Specific Rationale for requirements in CITS Geospatial specification

This section builds on the ***CITS Geospatial specification*** and extends the ***Guidelines for CITS Geospatial specification*** for the use cases, specific to GIS. Therefore, it focuses mostly on the structure, rendering, behaviour and context elements mentioned in the CITS Geospatial specification.

The section is primarily meant for technicians and developers of the specification. It is a prerequisite that the reader has knowledge about the CITS Geospatial specification, the Common Specification for Information Package, and the SIP specifications.

In this section, only the requirements from the CITS Geospatial specification that are GIS specific (complex GIS projects, web services, etc.) are listed and explained in greater detail. Additional explanation contains a more extensive *description*, an *example* and a *rationale* for why the specific GIS requirement is given. The intention is to provide a reasonable basis for understanding the reasons behind the requirements. It also aims to help with the validation of any information package that strives to be compliant with the CITS Geospatial specification. The requirements are isolated in boxes like this:

Requirement:

GEO_1		There **MUST** be a minimum of one representation and, therefore a minimum of one Package METS.xml and a minimum of one Representation METS.xml in a CITS Geospatial compliant package.	1..1 MUST

The requirements are numbered in the same way as the sections in the CITS Geospatial specification. The accompanying text from the CITS Geospatial specification is also repeated (in boxes) for a better reference.

3.3 Rationales in data requirements

Rationales in 3.3.1 Geodata general requirements

CITS Geospatial text:

> **3.3 Data Folder (Geospatial data)**
> This chapter states the requirements for the content data object or objects that form the geospatial record contained in the Information package.

GEO_11 Rationale

Requirement:

GEO_11	Minimum one file in a geospatial format	If the value in mets/@csip: CONTENTINFORMATIONTYPE is "citsgeospatial_v3_0", then there **SHOULD** exist at least one file in a geospatial format in representations/[RepresentationName]/data	0..n SHOULD

Description:
This requirement states that there SHOULD be at least one file in a geospatial format in a data folder in the IP to be a valid CITS Geospatial package.

Example:
See figure 8. The example illustrates an Information Package with one geospatial record in the GML format in the `Representations/[RepresentationName]/data` folder.

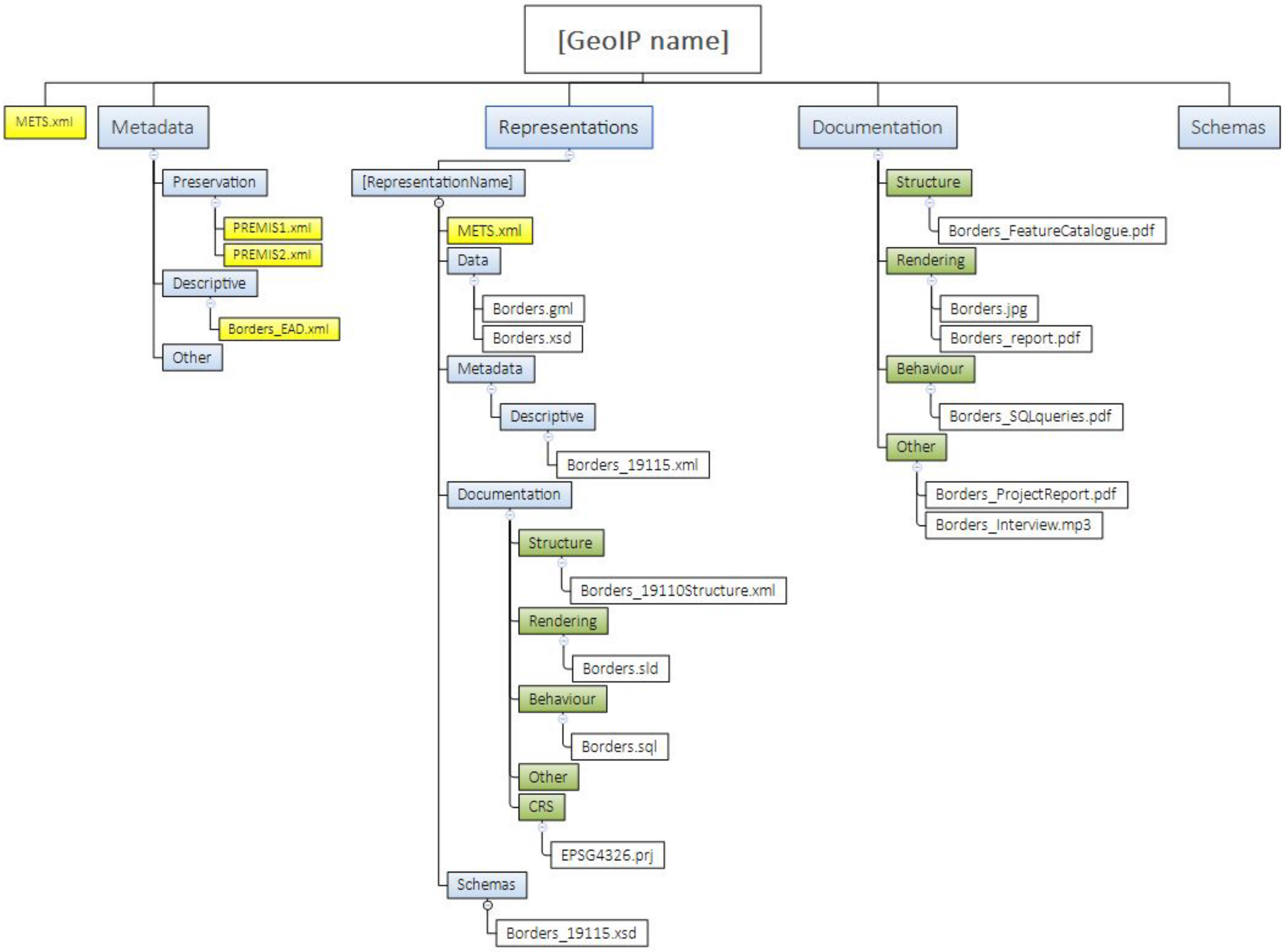

Figure 8 - Example of an Information Package folder structure

Rationale:
This requirement is based on the fact that more and more data is distributed. Thereofore, some organizations do not maintain their own geospatial records but instead use GIS applications that rely solely on web services from external sources. This requirement ensures the preservation of the behavior, structure, and rendering examples without involving any data, essentially creating an Information Package with no data content.

GEO_13 Rationale

Requirement:

GEO_13	Long term preservation format representation	The Information Package **SHOULD** contain at least one representation of a geospatial record in a long-term preservation format, as defined by the Archive or in the Long-term Preservation Format Profile (See chapter 3.3.5.)	0..n SHOULD

Description:
This requirement aims to ensure the long-term preservation of the data in the IP. It is up to the archive to determine which geospatial formats are considered as long-term preservation formats. However, the CITS

Geospatial specification recommends Long-term Preservation Formats Profiles for vector and raster data, which can also be used (see section "Rationales in 3.3.5 Long-term Preservation format profiles" below).

Example:
As proposed in the Long-Term Preservation format Profile for Geospatial Vector data using GML 3.2.1 in Appendix 1 of the Guidelines for CITS Geospatial specification, we can have vector geospatial data stored in a GML 3.2.1. format, along with other standardised machine-readable documentation.
In cases of web services, we would export source geospatial records to the proposed Long-Term format.

Rationale:
Since the CITS Geospatial specification is a specification for the long-term preservation of geospatial data, one representation in the IP should hold the geospatial data in a long-term preservation format.

GEO_14 Rationale

Requirement:

GEO_14	Original format representation	The Information Package **MAY** contain a separate representation of the same data, containing geospatial data in its original format	0..1 MAY

Description:
This requirement allows an additional representation in the IP with the geospatial data in the original format.

Example:
See Figure 9 of an IP with two representations. One representation contains a Long-Term Preservation vector data format (`GML321`), and the other contains a representation of the original format in an ESRI shapefile format (`SHP`).
In the case of tiled web services, we could have a representation containing the cached tiles from that service as dissemination representation for quick generation of web services, or they can be regenerated on demand.

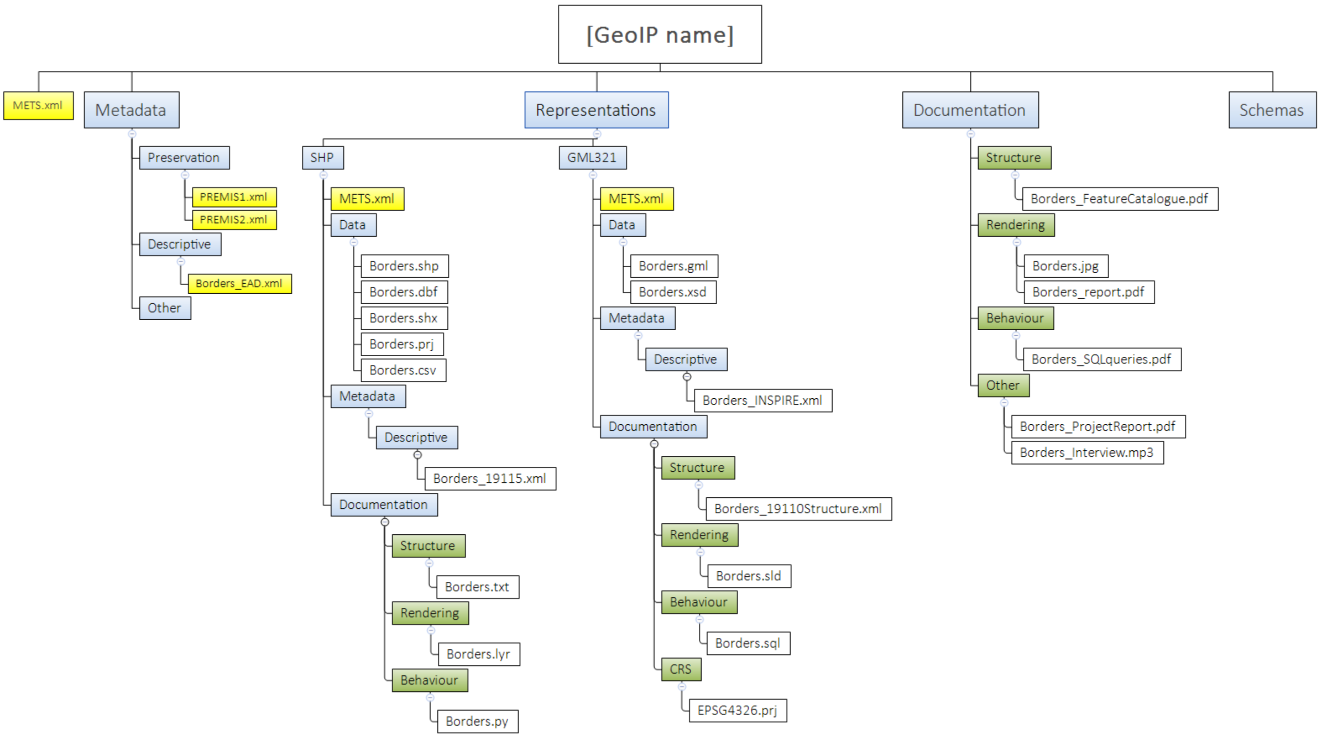

Figure 9 - Example of an Information Package with two representations

Rationale:
Original formats are often richer and easier to use than the preservation format and suitable for dissemination in the short term. However, it does not ensure the long-term preservation of the data. Geospatial data in original format can also be used for validation on submission mitigating loss of data and significant properties during migration to preservation format. The idea is that the users could use this representation until the original formats become obsolete.

In the case of cached tiled web services, the preservation of original tiles could save time in reconstructing the web service, however, it generally requires a lot more space.

GEO_15 Rationale

Requirement:

GEO_15 **Ref. Geo_11**	CRS definition	Every geospatial dataset **MUST** be accompanied with information about its underlying Coordinate Reference System (CRS) in one of two ways: • Full description of the CRS together with the archived data (within the geospatial file itself or in an accompanying file) • The geospatial file contains a reference to a CRS registry	Conditional 1..1 MUST

Description:
This requirement ensures that information about the Coordinate Reference System (CRS) used in the geospatial file stored is provided in a data folder. Information about the CRS can be documented as a complete description of the CRS (both geodetic datum, projection definition with its parameters, units of measurement, etc.) inside the geospatial file or as a reference in the geospatial file to an EPSG code referencing a well-known CRS registry - the EPSG database. If there are multiple versions of a CRS it is also important to include a timestamp or the valid version of the CRS reference. Often coordinate systems with the same EPSG code change over time. See also requirement GEO_38 if a reference to a CRS registry is used .

Example:
See an example of a full description of a CRS in the WKT Format in Figure 11 and a GML version, containing the Timestamp in the "RevisionDate" tag. Examples are available in the in GEO_38 part below.

Reference to an external CRS registry (EPSG code) in a GML-file

Georeferencing information in GML is a mandatory part of the file itself, and a reference to CRS is embedded in the geodata file itself. In the example below the attribute "srsName" holds the value of the coordinate reference system code, according to EPSG. In this example, the code is 4326[17].

```
<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>
```

[17] World Geodetic System 1984 (https://epsg.org/crs_4326/WGS-84.html)

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.

Reference to an external CRS registry (EPSG code) in a WMS Service:
Web services can serve data in different coordinate systems that are defined in the following entities relevant to the WMS:
– the <BoundingBox> element in the service metadata (7.2.4.6.8)[18];
– the CRS parameter in the GetMap request (7.3.3.5);
– the CRS parameter in the map requests part of the GetFeatureInfo request (7.4.3.3).

Rationale:
The Coordinate Reference System (CRS) provides information about how to locate geodata objects anywhere on the earth's surface. For a GIS to display the content of a geospatial file correctly on the earth's surface, the Coordinate Reference System (CRS) corresponding to the coordinates in the geospatial file must be specified in the geospatial file itself or within an accompanying file within the Information package.

Rationales in 3.3.3 Raster requirements

CITS Geospatial text:

Additional to the Geodata general requirements, the following requirements are intended for all raster geodata in the Information package:

GEO_23 Rationale

Requirement:

GEO_23	Tiling index file	If raster objects are organised using an external tiling index file this tiling index **MAY** be placed in representations/[RepresentationName]/data	0..n MAY

Description:
This requirement proposes the inclusion of an external tiling index dataset or document describing the organisation of a record that is comprised of a large number of raster objects. When a raster geodata record is very large, for example, an Orthophoto Campaign covering a whole country, it needs to be composed out of several thousand raster data objects organised in a tiling index.

In cases of a Tiled Web Map Service, the tiling schema is defined in the OpenGIS Web Map Tile Service Implementation Standard[19].

Example:
The map in Figure 10 illustrates a tiling index.

[18] Open GIS Web Map Service (WMS) Implementation Specification - https://portal.ogc.org/files/?artifact_id=14416
[19] OpenGIS Web Map Tile Service Implementation Standard - https://portal.ogc.org/files/?artifact_id=35326

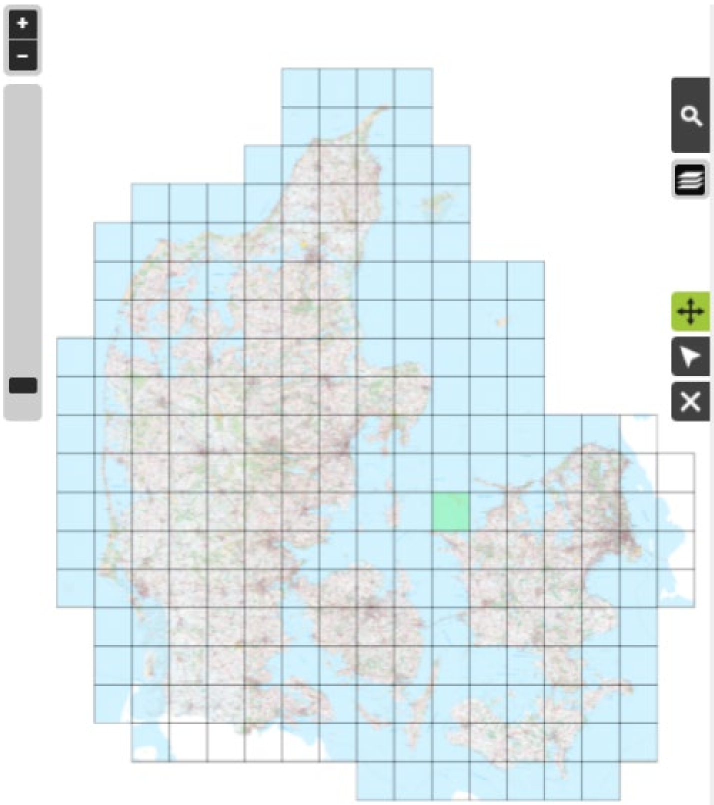

Figure 10 - Example of a tiling index

Rationale:
A tiling index file is often used as a finding aid for access and dissemination of large raster datasets with several raster data objects. It could therefore be considered a part of this record. The tiling index file enables easy visual identification of a relevant raster data object (raster files). A tiling index can also be recreated based on the information in the raster data objects in the Information Package.

Rationales in 3.3.5. Long Term Preservation Format Profiles

CITS Geospatial text:

> A Long Term Preservation format Profile contains a set of one or more base or subsets of base standards, and, where applicable, the identification of chosen clauses, classes, options, and parameters of those base standards, that are necessary for geospatial records to comply with the Archival guidelines for the selection of long-term preservation formats.
>
> A Long Term Preservation format Profile would specify a proposed format for long term specification, its justification according to Archival guidelines (to ensure long-term preservation and reuse), a list of required auxiliary files and documentation and validation criteria to ensure structural and content suitability.

In cases of preserving Geospatial records from a GIS system, the same Long-Term Preservation Format profiles are used. Additional Profiles will be developed in the future.

3.4 Rationales in Documentation requirements

Rationales in 3.4.1 Structure of geospatial records

CITS Geospatial text:

Structure of geospatial records describes the extrinsic or intrinsic relationships between two or more type of content, as required to reconstruct the performance of one or more geospatial records within the information package.

GEO_30 Rationale

Requirement:

GEO_30	GIS Project structure	A document describing the structure of geospatial records in the GIS System **MAY** be provided in the Information Package. A standardised machine-readable version is preferred.	0..n MAY

Description:
This requirement recommends the preservation of information about the structure of the geospatial records in the original GIS System used for creating information products based on the digital records, like a permit, a segment of a cadastral map etc.

Example:
The OGC OWS Context standard can be used for storing information on the configuration and rendering styles of data used and its references. An OWS Context (OWC) document consists of a number of key elements. These key elements are shown in figure 11.

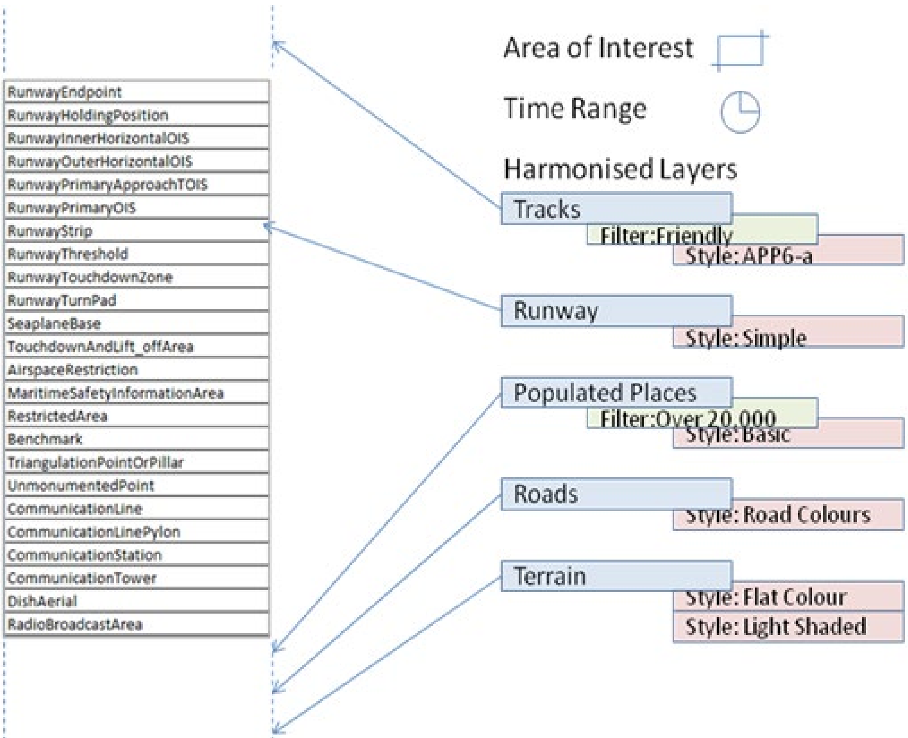

Figure 11 - OWS Context key elements (Source: www.owscontext.org)

The overall information includes a name, abstract and creation date as well as a range of high-level metadata. It also includes an Area of Interest and a Time Range of interest (all optional). When using OWC for visualisation, the Area of Interest is typically the area displayed on the screen when the context document is loaded. Similarly, the Time Range of interest is the time range that any time slider in the application will be set to. The core of the OWC document is an ordered list of resources. Again, for visualisation purposes, the application should load the resources in the list such that the first resource is at the top (i.e., reverse draw order). However, it is recommended, that all data also includes a timestamp of its usage, since the data provided by the referenced services can be changing as well.

Each layer in a context document is known as a 'Resource' which in the Atom encoding is mapped to an 'Entry' XML element. A resource can be realised in a number of different ways, and so an OWC document allows various options to be specified. These are known as offerings. The intention is that these are, as far as is possible by the format used, equivalent and no priority is assigned to their order in the standard. They are intended to be alternatives that the client can use to allow it to visualise or use the resource. This structure is shown in Figure 12.

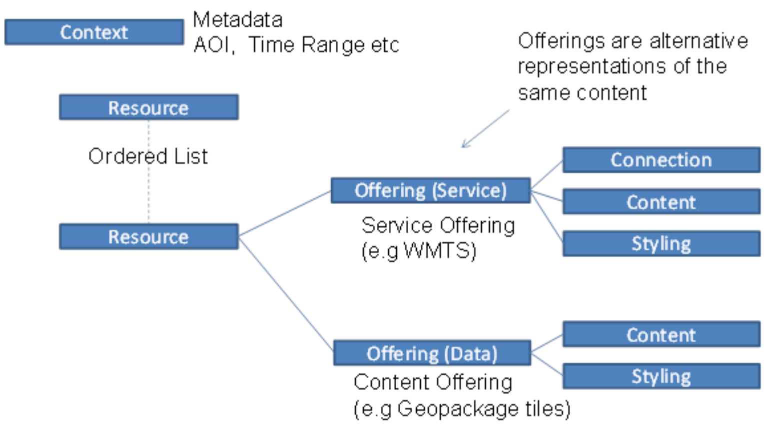

Figure 12 - Structure of an OWS Context document

Rationale:
This information is needed to be able to reproduce an original digital information product based on the preserved digital records in the Information Package. This can be done by replicating it based on the preserved data and the methods by which they were used to the level needed for the digital product to be reused for the same basic purpose in the future (issuing a permit, a segment of a cadastral map, etc.). An alternative approach is to preserve or emulate original GIS systems in virtual environments.

Rationales in 3.4.2 Rendering and visualisation

CITS Geospatial text:

Rendering and visualisation documentation represents any information that contributes to the recreation of the performance of the Information Object. Example: Colour map of pixel values in raster datasets, Symbology configuration for vector datasets, Map setup; Web service, etc. To document visualisation, documentation and samples of geospatial information products (maps, lists, charts, new geodata derived from existing data, web services, etc.) from GIS are required.

GEO_33 Rationale

Requirement:

GEO_33	Rendering configuration	A standardised machine-readable rendering configuration for one or more geospatial datasets **MAY** be provided in the Information Package	0..n MAY
GEO_33a Ref GEO_33	Placement of rendering configuration	If a standardised machine-readable rendering configuration for one or more geospatial datasets exists, it **SHOULD** be provided in representations/[RepresentationName]/documentation/rendering	0..n SHOULD

Description:
This requirement recommends that rendering configurations are documented in a standardised machine-readable format to support dissemination automatisation.

Example:
An example of Standardised machine-readable formats for the rendering of geospatial records are SLD[20] files. KML[21] files also have some of that capability:

SLD files example
SLD is an OGC[22] (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 provided in the `Documentation` in an open-source GIS application like QGIS[23]. 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. An example of an SLD file is shown in figure 13.

```
<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>
```

20 SLD – Styled Layer Description (https://portal.opengeospatial.org/files/?artifact_id=22364)
21 KML Standard (http://www.opengeospatial.org/)
22 Open Geospatial Consortium (http://www.opengeospatial.org/)
23 QGIS (https://qgis.org/)

```
                            <Size>6</Size>
                        </Graphic>
                    </PointSymbolizer>
                </Rule>
            </FeatureTypeStyle>
        </UserStyle>
    </NamedLayer>
</StyledLayerDescriptor>
```

Figure 13 - Example of a Styled Layered Description XML file (SLD)

Rationale:

To enable automated and correct dissemination of the preserved geospatial records in the Information Package.

Rationales in 3.4.3 Software and algorithms

CITS Geospatial text:

> To facilitate the reproduction of information products in future GIS, we often need to run specific database queries or geo-specific processes (geoprocessing workflows). However, some information can only be accessed using application functionalities. Therefore the preservation of user manuals and system documentation is also required to preserve the behaviour aspect.

GEO_35 Rationale

Requirement:

GEO_35	GIS system documentation	Documentation regarding the original system, where geospatial records were used, **SHOULD** be provided in the Information Package.	0..n SHOULD
GEO_35a Ref GEO_35	Placement of GIS system documentation	If documentation regarding the original system exists it **SHOULD** be provided in a documentation/behaviour folder	0..n SHOULD

Description:
This requirement recommends the preservation of documentation of the System used to create, manage or manipulate the geospatial data in the Information package. Ideally, the location should be in the `documentation/behaviour` folder on the package or representation level. However, other placements are also possible. The functions of different GIS systems and other applications that use geospatial products varies from system to system. The functions are commonly not specific to a given geospatial dataset or product.

Example:
When collecting systems behaviour information like software and algorithm information, it is recommended to ask users of the geospatial dataset to provide any user documentation available for the system in which the geospatial datasets are being used. This could be any existing manuals, articles on common practices or white papers describing the common methods for use, manipulation, analysis, application and source code, etc., of the geospatial dataset. It is recommended to conduct and record online interviews in which users demonstrate the use of the most common functionalities on a screen and use this video recording as documentation.

Rationale:
The purpose of describing the initial GIS system used to create, manage or manipulate the preserved geospatial data in the Information Package is to be able to recreate a rendering tool and its functionality in the future.

GEO_36 Rationale

Requirement:

GEO_36	Common queries, algorithms	Documentation on the logic of common queries and algorithms used for analysis, transformation, creation and maintenance of geospatial records **SHOULD** be provided in the Information Package	**0..n** **SHOULD**
GEO_36a Ref GEO_36	Placement of common queries, algorithms	If documentation on the logic of common queries and algorithms exists, it **MUST** be provided in a documentation/behaviour folder	0..n SHOULD

Description:
This requirement recommends preserving information on common queries and algorithms used for creation, transformation, analysis or maintenance of the geospatial records in the Information Package in a `documentation/behaviour` folder on the package level.

Example:
Documentation examples could include:

- UML diagrams of Common workflows used
- Well documented algorithms that can be programmed into a new tool in the future.
- User manuals, documenting workflows and algorithms used to create Information products based on geospatial and other records.

Rationale:
This information enables correct reuse of the preserved geospatial records similar to the original use of data at the time the data was created for a specific purpose.

GEO_37 Rationale

Requirement:

GEO_37	Common queries, algorithms machine-readable	Code of queries and algorithms used with the geospatial records in the Information Package **MAY** be provided in the Information Package	0..n MAY
GEO_37a Ref GEO_37	Placement of machine-readable common queries, algorithms	If code of queries and algorithms used with the geospatial records exists, it **SHOULD** be provided in a documentation/behaviour folder	0..n SHOULD

Description:
This requirement recommends that the common queries described in GEO_36 could be documented in a machine-readable format. Suppose algorithms are available in a standardised machine-readable format (like a standard sql). In that case, ideally they should be placed within the `documentation/behaviour` of the long-term preservation or dissemination representation (if applicable) to allow for possible automated access.

Example:
Standardised machine-readable examples would include scripts or sql queries that were used on the geospatial data.

Rationale:
A machine-readable version of common queries and algorithms can enable automated dissemination or guide to correct recreation of queries and algorithms performed on data in the IP.

Rationales in 3.4.4 Coordinate reference system information – requirements

CITS Geospatial text:

> A coordinate Reference System definition is essential for the effective reuse of all geospatial records. When the CRS of the geodata in the Information Package is described by only referencing a well-known external database of CRS definitions (such as the EPSG database), the availability of these definitions is dependent upon the long-term existence of that database. Therefore, a CITS Geospatial Information Package must contain these definitions to be self-descriptive.

GEO_38 Rationale

Requirement:

GEO_38 Ref GEO_15	Standardised machine-readable format CRS definition	If the CRS definition in a geospatial file is documented only by a reference to a CRS registry, a standardised machine-readable format CRS definition compliant with standards for CRS definition **SHOULD** be provided in the Information Package	0..n SHOULD
GEO_38a Ref GEO_38	Placement of standardised machine-readable format CRS definition	If a standardised machine-readable format CRS definition exists, it **SHOULD** be provided in a documentation/CRS folder	0..n SHOULD

Description:
This requirement recommends that if the Coordinate Reference System (CRS) definition in the geospatial data files is only defined as a reference code (e.g., EPSG[24] code) to a CRS registry (see GEO_15), then the CRS should be described fully in a machine-readable format and placed in a documentation/CRS folder at representation level.

[24] EPSG (https://epsg.org/home.html)

Example:
A full description of a CRS can be documented in an accompanying projection file (.prj) in the WKT2 (ISO 19162:2019) format:

```
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]]
```

Figure 14 - Full description of a CRS in a projection file in WKT2 format

Rationale:
Preservation of information on the CRS used in the preserved geospatial records in the IP is essential to be able to display the content of a geospatial file correctly on the surface of the earth in the Coordinate Reference System (CRS) corresponding to the coordinates in the geospatial file. Coordinate Reference Systems also become obsolete and replaced by new ones.

GEO_39 Rationale

Requirement:

GEO_39	CRS transformation parameters	For systems using data in multiple CRS systems, standardised machine-readable transformation parameters between those CRS **MAY** be provided in the Information Package	0..n MAY
GEO_39a Ref GEO_39	Placement of CRS transformation parameters	If standardised machine-readable transformation parameters exist, they **SHOULD** be provided in a documentation/CRS folder	0..n SHOULD

Description:
This requirement applies if multiple geospatial datasets in different coordinate systems are available along with transformation parameters, or if a geospatial dataset has a different CRS from common national or global datasets and is preserved along with transformation parameters. This information could be available in a machine-readable format and provided in a `documentation/CRS` folder on the representation level.

Example:

WKT Version 2


```
COORDINATEOPERATION["MGI 1901 to Slovenia 1996 (1)",VERSION["GuRS-Svn"],SOURCECRS[GEOGCRS["MGI 1901",DATUM["MGI
1901",ELLIPSOID["Bessel 1841",6377397.155,299.1528128,ID["EPSG",7004]],ID["EPSG",1031]],CS[ellipsoidal,2,ID["EPSG",6422]],AXIS["Geodetic
latitude (Lat)",north],AXIS["Geodetic longitude
(Lon)",east],ANGLEUNIT["degree",0.0174532925199433,ID["EPSG",9102]],ID["EPSG",3906]]],TARGETCRS[GEOGCRS["Slovenia
1996",DATUM["Slovenia Geodetic Datum 1996",ELLIPSOID["GRS
1980",6378137,298.257222101,ID["EPSG",7019]],ID["EPSG",6765]],CS[ellipsoidal,2,ID["EPSG",6422]],AXIS["Geodetic latitude
(Lat)",north],AXIS["Geodetic longitude
(Lon)",east],ANGLEUNIT["degree",0.0174532925199433,ID["EPSG",9102]],ID["EPSG",4765]]],METHOD["Coordinate Frame rotation (geog2D
domain)",ID["EPSG",9607]],PARAMETER["X-axis translation",409.545,LENGTHUNIT["metre",1,ID["EPSG",9001]]],PARAMETER["Y-axis
translation",72.164,LENGTHUNIT["metre",1,ID["EPSG",9001]]],PARAMETER["Z-axis
translation",486.872,LENGTHUNIT["metre",1,ID["EPSG",9001]]],PARAMETER["X-axis rotation",-3.085957,ANGLEUNIT["arc-
second",0.0000048481368111,ID["EPSG",9104]]],PARAMETER["Y-axis rotation",-5.46911,ANGLEUNIT["arc-
second",0.0000048481368111,ID["EPSG",9104]]],PARAMETER["Z-axis rotation",11.020289,ANGLEUNIT["arc-
second",0.0000048481368111,ID["EPSG",9104]]],PARAMETER["Scale difference",17.919665,SCALEUNIT["parts per
million",0.000001,ID["EPSG",9202]]],OPERATIONACCURACY[1],USAGE[SCOPE["Transformation of coordinates at 1m level of
accuracy."],AREA["Slovenia - onshore."],BBOX[45.42,13.38,46.88,16.61]],ID["EPSG",3916],REMARK["Info source also gives a separate reverse tfm
with slightly different parameter values. Given the tfm accuracy these differences are not significant and this tfm can be considered reversible. May be
taken as approx tfm MGI 1901 to WGS 84 (see code 3917)."]]
```


Figure 15 - Machine-readable CRS transformation between MGI 1901 and Slovenia 1996 in WKT2

GML Version:

```
<?xml version="1.0" encoding="utf-8"?>
<gml:GeodeticCRS xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:epsg="urn:x-ogp:spec:schema-
xsd:EPSG:2.3:dataset" gml:id="epsg-crs-4326" xmlns:gml="http://www.opengis.net/gml/3.2">
  <gml:metaDataProperty>
    <epsg:CommonMetaData>
      <epsg:type>geographic 2d</epsg:type>
      <epsg:informationSource>EPSG. See 3D CRS for original information
source.</epsg:informationSource>
      <epsg:revisionDate>2022-11-29</epsg:revisionDate>
      <epsg:changes>
        <epsg:changeID xlink:href="https://epsg.org/api/v1/Change/2002.151/export?format=gml" />
        <epsg:changeID xlink:href="https://epsg.org/api/v1/Change/2003.37/export?format=gml" />
        <epsg:changeID xlink:href="https://epsg.org/api/v1/Change/2006.81/export?format=gml" />
        <epsg:changeID xlink:href="https://epsg.org/api/v1/Change/2007.079/export?format=gml" />
        <epsg:changeID xlink:href="https://epsg.org/api/v1/Change/2020.027/export?format=gml" />
        <epsg:changeID xlink:href="https://epsg.org/api/v1/Change/2022.086/export?format=gml" />
      </epsg:changes>
      <epsg:show>true</epsg:show>
      <epsg:isDeprecated>false</epsg:isDeprecated>
      <epsg:Usage>
        <epsg:extent xlink:href="https://epsg.org/api/v1/Extent/2830/export?format=gml" />
        <gml:scope>Horizontal component of 3D system.</gml:scope>
      </epsg:Usage>
    </epsg:CommonMetaData>
  </gml:metaDataProperty>
  <gml:metaDataProperty>
    <epsg:CRSMetaData>
      <epsg:projectionConversion
xlink:href="https://epsg.org/api/v1/Conversion/15593/export?format=gml" />
      <epsg:sourceGeographicCRS
xlink:href="https://epsg.org/api/v1/CoordRefSystem/4979/export?format=gml" />
    </epsg:CRSMetaData>
  </gml:metaDataProperty>
  <gml:identifier codeSpace="EPSG">4326</gml:identifier>
  <gml:name>WGS 84</gml:name>
  <gml:scope />
  <gml:ellipsoidalCS xlink:href="https://epsg.org/api/v1/CoordSystem/6422/export?format=gml" />
  <gml:geodeticDatum xlink:href="https://epsg.org/api/v1/Datum/6326/export?format=gml" />
</gml:GeodeticCRS>
```

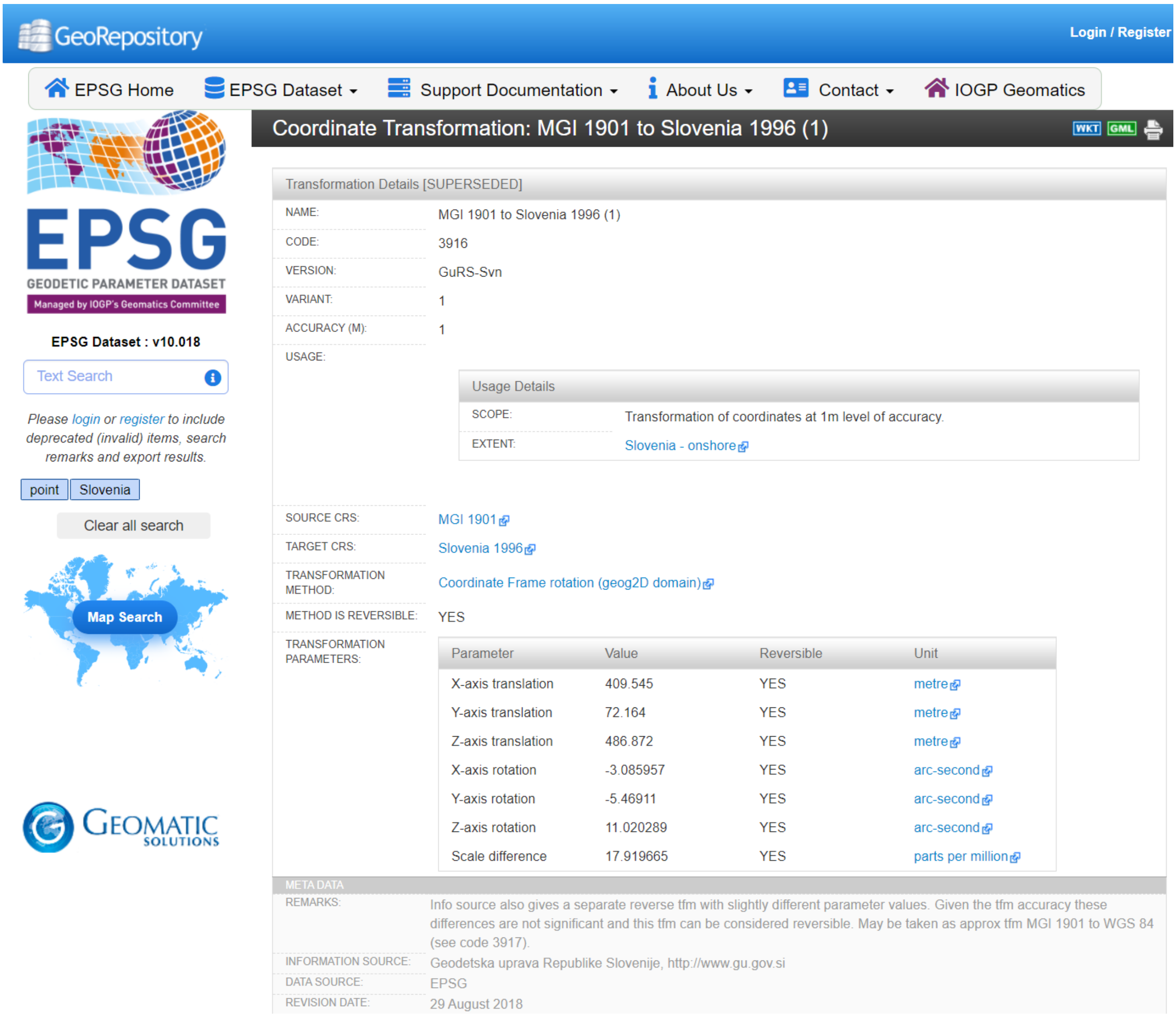

Figure 16 - Human-readable description of CRS transformation from MGI 1901 to Slovenia 1996

<u>Rationale:</u>

Preservation of transformation parameters enables correct and automated transformations between CRS in the future use of the preserved geospatial records.

Postface

AUTHOR(S)	
Name(s)	Organisation(s)
Martin Dew-Hattens	Danish National Archives
Anita Graser	Austrian Institute of Tehcnology
Gregor Završnik	Geoarh

REVIEWER(S)	
Name(s)	Organisation(s)
Jaime Kaminski	Highbury R&D
Karin Bredenberg	Kommunalförbundet Sydarkivera

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
1.0	11.6.2021	Martin Dew-Hattens Gregor Završnik	DNA Geoarh	Creation of guideline
1.0.0	31.8.2021	Various	Various	Publication of version 1.0.0
1.1.0	13.12.2024	Varoius	Varoius	Publication of version 1.1.0

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.