

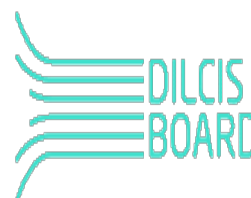
Guideline for CITS Geospatial

Guideline for the E-ARK Content Information Type Specification for digital geospatial data records archiving



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 is maintained by Digital Information LifeCycle Interoperability Standards Board and is licensed under CC BY 4.0



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:

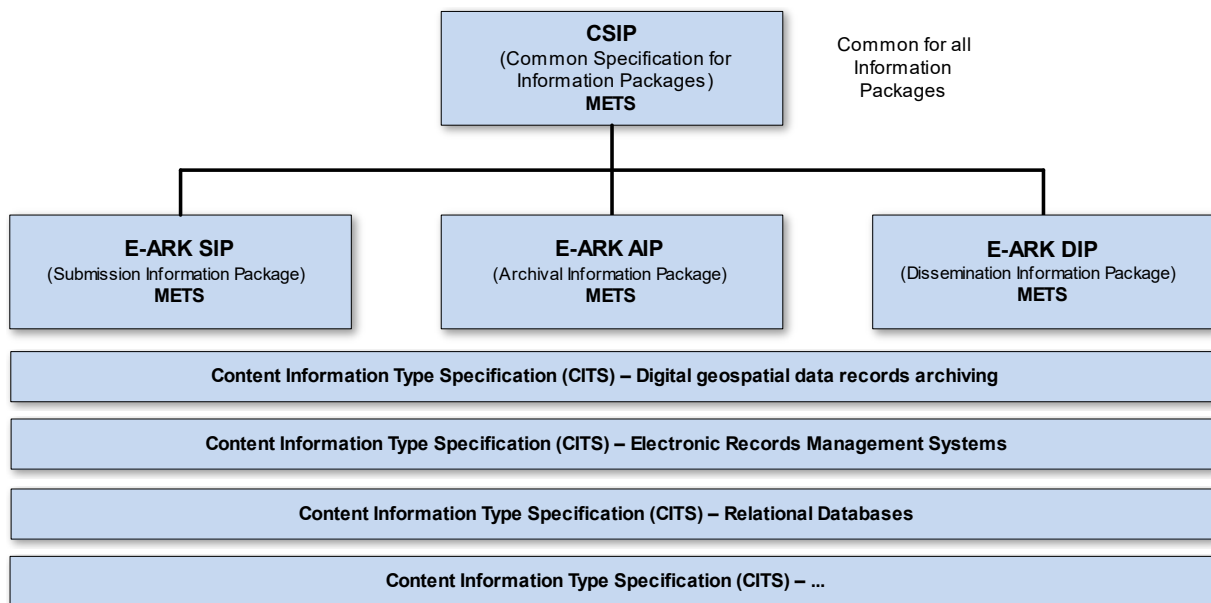


Figure I: Diagram showing E-ARK specification dependency hierarchy. Note that the image only shows a selection of the published CITS and isn't an exhaustive list.

Specification	Aim and Goals
Common Specification for Information Packages	<p>This document introduces the concept of a Common Specification for Information Packages (CSIP). Its three main purposes are to:</p> <ul style="list-style-type: none"> • 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.

Specification	Aim and Goals
	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	<p>The main aims of this specification are to:</p> <ul style="list-style-type: none"> • 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	<p>The main aims of this specification are to:</p> <ul style="list-style-type: none"> • 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	<p>The main aims of this specification are to:</p> <ul style="list-style-type: none"> • 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	<p>The main aim and goal of a Content Information Type Specification is to:</p> <ul style="list-style-type: none"> • 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. <p>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/).</p>

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://eak-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 CONTENT

1	Preface	3
1	Context	9
1.1	Purpose	9
1.2	Scope	9
1.3	Structure of the document	9
2	Introduction to Geospatial data and approach to its preservation	10
2.1	Geospatial geodata	10
2.1.1	Basic Terms	10
2.1.2	Geospatial vector data	10
2.1.3	Geospatial raster data	12
2.1.4	Standards for Geospatial Data	12
2.1.4.1	CITS Geospatial alignment with ISO 19165-1:2018	13
2.2	Significant properties of geospatial data	15
2.3	Recommended reading list	16
3	Rationale for requirements in CITS Geospatial	17
3.1	Folder structure requirements	18
	GEOSTR1 rationale	19
	GEOSTR2 rationale	19
	GEOSTR3 rationale	20
	GEOSTR4 rationale	20
	GEOSTR5 rationale	21
	GEOSTR6 rationale	21
	Folder structure examples explained	21
3.1.1	Geo IP containing one vector representation	21
3.1.1.1	Content	22
3.1.1.2	Context	22
3.1.1.3	Structure	22
3.1.1.4	Rendering	22
3.1.1.5	Behaviour	23
3.1.1.6	Package level elements	23
3.1.2	Geo IP containing multiple vector representations	23
3.1.3	Geo IP containing one representation of multiple raster datasets	25

3.2	Rationales in METS Requirements	26
	GEO_1 Rationale	26
	Rationales in 3.2.2 Package METS requirements	27
	GEO_2 Rationale	27
	GEO_3 Rationale	27
	GEO_4 Rationale	28
	GEO_5 Rationale	28
	GEO_6 Rationale	29
	GEO_7 Rationale	30
	Rationales in 3.2.3 Representation METS requirements	31
	GEO_8 Rationale	31
	GEO_9 Rationale	31
	GEO_10 Rationale	32
3.3	Rationales in data requirements	33
	Rationales in 3.3.1 Geodata general requirements	33
	GEO_11 Rationale	33
	GEO_12 Rationale	33
	GEO_13 Rationale	34
	GEO_14 Rationale	35
	GEO_15 Rationale	35
	GEO_16 Rationale	36
	GEO_17 Rationale	38
	Rationales in 3.3.2. Vector Geodata – requirements	38
	GEO_18 Rationale	38
	GEO_19 Rationale	39
	GEO_20 Rationale	39
	Rationales in 3.3.3 Raster requirements	40
	GEO_21 Rationale	40
	GEO_22 Rationale	41
	GEO_23 Rationale	41
	Rationales in 3.3.5. Long Term Preservation Format Profiles	43
3.4	Rationales in Documentation requirements	50
	Rationales in 3.4 Documentation folders requirements	50
	GEO_24 Rationale	50

GEO_25 Rationale	51
Rationales in 3.4.1 Structure of geospatial records	52
GEO_26 Rationale	52
GEO_27 Rationale	53
GEO_28 Rationale	53
GEO_29 Rationale	55
GEO_30 Rationale	56
Rationales in 3.4.2 Rendering and visualisation	58
GEO_31 Rationale	59
GEO_32 Rationale	60
GEO_33 Rationale	61
GEO_34 Rationale	63
Rationales in 3.4.3 Software and algorithms	64
GEO_35 Rationale	64
GEO_36 Rationale	65
GEO_37 Rationale	65
Rationales in 3.4.4 Coordinate reference system information – requirements	66
GEO_38 Rationale	66
GEO_39 Rationale	68
Rationales in 3.4.5 Other - Contextual Documentation requirements	70
GEO_40 Rationale	71
GEO_41 Rationale	72
3.5 Rationales in metadata requirements	73
Rationales in 3.5 Geospatial Metadata requirements	73
GEO_42 Rationale	94
GEO_43 Rationale	94
4. Examples and tools	96
o Examples	96
o Tools	96
5. Postface	98
Appendix 1: Long-Term preservation format Profile for Geospatial Vector data using GML 3.2.1	100
Appendix 2: Long-Term preservation format Profile for Geospatial Raster data using TIFF baseline 6	101
Appendix 3: Mapping of INSPIRE metadata descriptions to archival descriptive metadata standards	102

1 Context

1.1 Purpose

The purpose of this guideline is to further explain and describe the “Specification for the E-ARK Content Information Type Specification for digital geospatial data records archiving” (also called CITS Geospatial in short). The goal is that as many people as possible will be able to understand the specification and, therefore, to also preserve geodata. The guideline is an evolving document, and more concepts and standards will be explained following the needs of the users of the specification.

1.2 Scope

This guideline provides further information and insights on preserving basic to moderately complex Geospatial records used within GIS, mostly in vector and raster formats. Geospatial domain is developing and many new formats are being used from LIDAR and point clouds to live-streamed data and distributed cloud-based formats. The CITS Geospatial specification is written in a very generic manner, so most of the proposed rules are also applicable to other data types. More examples of standards for newer formats will be added in the future. **This guideline is not describing the usage of CITS Geospatial for GIS and Web services. That topic is addressed in another Guideline.**

The CITS Geospatial specification builds on work done in the E-ARK projects.

1.3 Structure of the document

Section 2 contains an introductory section describing the concept of geodata in general and 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 geospatial data and GIS.

Section 3 provides a rationale for each of the requirements found in the CITS Geospatial specification. This is meant to provide a better basis for understanding the reasons behind the requirements. This section is primarily meant for technicians and developers of the specification, and it is a prerequisite that the reader has knowledge about geospatial data, formats for vector and raster data, the structure of geospatial data in GIS, GIS rendering functionality and the Common Specification for Information Package and the SIP, AIP and DIP specifications.

Section 4 contains an overview of available example packages and tools related to the CITS Geospatial specification as a means to get your hands dirty and take action in developing the field of geodata preservation.

2 Introduction to Geospatial data and approach to its preservation

This section contains an introductory section describing the concept of geospatial data in general and the digital preservation of them. 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.

2.1 Geospatial geodata

The aim of geospatial data is to represent real-world phenomena in a more generalised copy. Digital geospatial data virtually represents real-world objects and phenomena and are sometimes considered digital twins¹. The purpose of digital twins is to conduct operations in a virtual system, that gives us information about the object and its relation to other objects.

2.1.1 Basic Terms

A **Feature** is an abstraction of a real-world phenomenon. An example would be a digital line, representing a section of a road or river – an **instance** of that **Feature type**. We can have Feature types – used to represent different rivers, streams, etc.

An object in the real world can be represented as a feature with its **geometry** represented as a vector (lines represented by multiple sets of coordinates) or as a raster object (a set of pixels with the same values within a larger image). In geospatial data, this geometry always has coordinates based on a Geographical Coordinate Reference System.

A feature can also have its characteristics, which are described as **Feature Attributes**. A river represented as a vector line feature can have an attribute type “Name” that holds a text-based value “Danube” or an attribute “Length”, that holds a numeric value “2850,00”.

A collection of Features containing the same Feature Attribute structure are called a **Feature Dataset**.

A **Geospatial record** represents one or many objects in space and can be composed of one or many Feature Datasets.

2.1.2 Geospatial vector data

Geospatial vector Feature datasets can have different types of geometries (point, line, polygon, multipoint, multiline, multipolygon, and different types of curves and surfaces). Types may vary, based on standards and implementations. Examples:

- Points or multipoints: Shipwrecks, trees, mountain tops, GPS tracks, waterholes, cities, etc.
- Lines: Roads, rivers, power lines, isohypse, etc.
- Polygons: lakes, rivers (smaller scales), road surfaces (smaller scales), Country borders, etc.

¹ Digital Twin (https://en.wikipedia.org/wiki/Digital_twin)

- Mult polygon: A country with islands (multiple polygons or points representing one Feature Instance with one set of Feature Attributes)

In Figure 1, we can see that a Polygon vector **Feature** represents the border of Austria and has three different **Feature Attributes** that describe it.

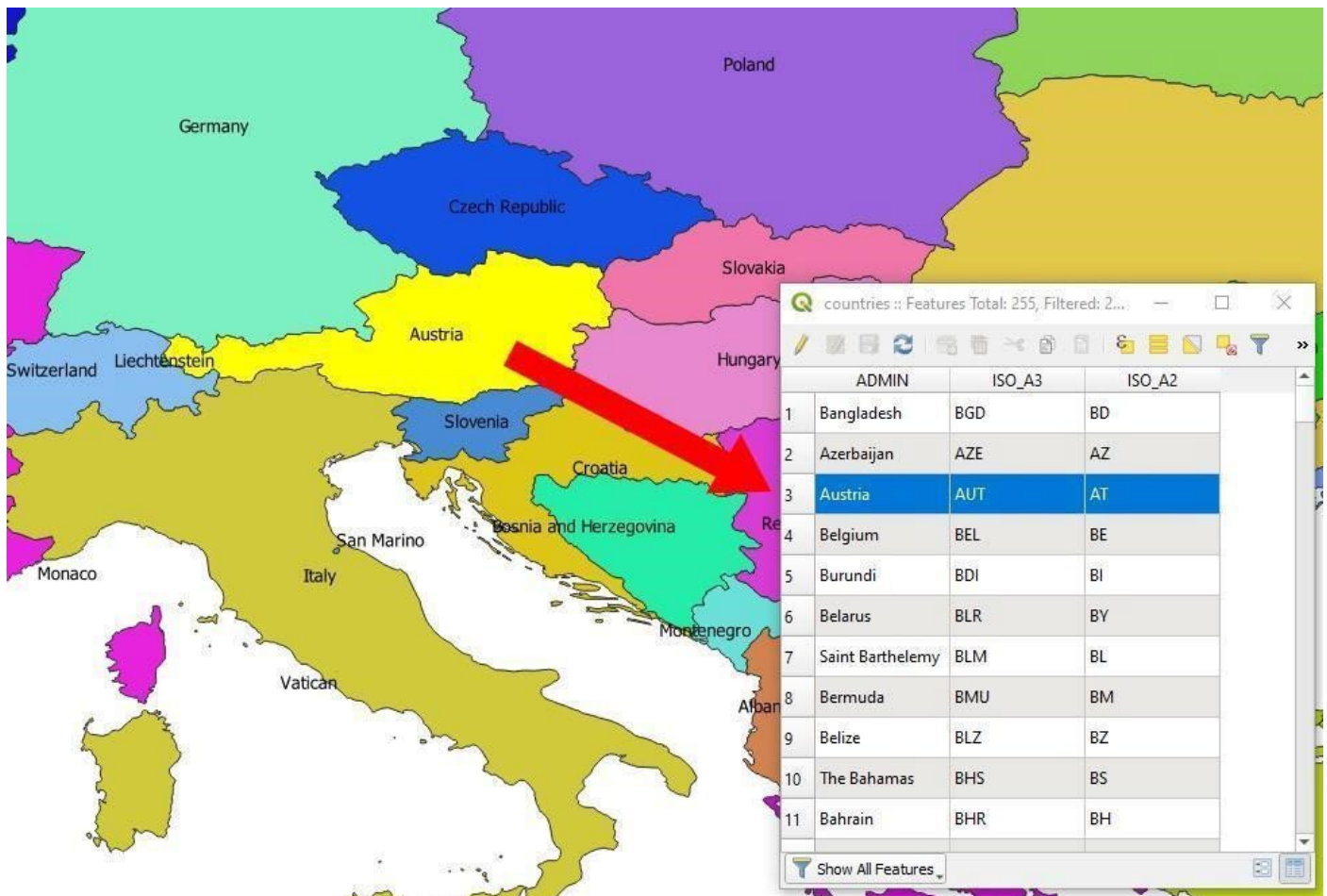


Figure 1 – EU Countries – An example of a geospatial vector dataset with Feature attributes

The explanation of Feature Attribute types and their meanings in the EU Countries dataset is commonly described in a **Feature Catalogue**.

In order to ensure the long-term preservation of geospatial data, it is necessary to ensure that there is a well-documented graphical component, well-defined descriptive attributes and a geographic coordinate system. Also, proprietary and undocumented geodata formats must be converted into a long-term preservation format that is well described and preserves all the significant properties of the original geodata format. Even if some formats are a de facto standard today, they may become unreadable in the distant future.

2.1.3 Geospatial raster data

All spatial raster data are a type of raster images, but not all raster images are spatial raster data. Both grid data and geospatial raster images are spatial raster data.

A raster image is commonly used to represent continuous features over large areas. For example, scanned maps, aerial images, satellite images, and grid files, that represent different values over the pixel area (Elevation, population density, gravimetry measurements, etc.)

A raster dataset is composed of a rectangular array of grid of pixels, each of which represents a value. One pixel represents the smallest unit of information. In the case of ordinary raster images, this value of each pixel usually represents a colour.

Image

A geospatial raster image, like a satellite image, scanned map or an orthophoto, also has pixels that represent colours which can then be rendered as a coloured image of the surface of the earth for human viewing. But it also has an orientation in space and can thus be placed correctly on the surface of the earth or in a 3D space according to a coordinate and reference system representing the earth or this 3D space.

Grid data

In the case of raster grids, the pixels of the image, also called raster cells, can also represent measurements of distances, areas, volumes or heights, derived calculations, classifications or any other units for the area in space that the pixel represents. A digital elevation model is grid data where each pixel represents a height in the terrain or surface of an area.

2.1.4 Standards for Geospatial Data

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

OGC² (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 CITS Geospatial are geospatial formats GML and GeoTIFF, and others³.

ISO TC/211⁴ (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***

² Open GIS Consortium (<https://www.ogc.org/>)

³ List of OGC Standards (<https://www.ogc.org/docs/is>)

⁴ ISO TC/211 (<https://committee.iso.org/home/tc211>)

2.1.4.1 CITS Geospatial alignment with ISO 19165-1:2018

CITS Geospatial is a technical specification that inherits its structure from the CSIP and therefore inherits its folder structure, metadata requirements, and principles (like representations). Basically, it is a Geospatial extension to the CITS Package, whose structure is adopted to serve different types of digital records.

ISO 19165-1 was first released in 2018, and its conception was designed from a geospatial data-centric approach, including preservation concepts from the OAIS standard.

Therefore, there are some conceptual differences between CITS Geospatial and ISO 19165-1:2018. For instance, the Geospatial Information model of the ISO standard includes preservation metadata elements (GP_PreservationMetadata) that are included as a part of the general ISO 19165-1 Metadata model. The elements are aligned with the OAIS Standard. However, the solution is merged with other metadata elements.

CITS Geospatial inherits the approach from CSIP, where separate METS and PREMIS profiles are used to describe fixity and provenance information. CITS Geospatial also describes the structure, context, and rendering content in greater detail than ISO 19165-1.

During the development of the 3.0 version of CITS Geospatial, we aimed to include as many of the geospatial data elements, its properties and auxiliary information and documentation as proposed in the ISO 19165-1. Due to the nature of the CSIP structure, a full complete adoption of the ISO 19165-1 is impossible. However, CITS Geospatial does support the inclusion of ISO 19165-1:2018 based metadata as descriptive metadata (requirement GEO_42).

ISO 19165-1 also describes the packaging mechanism of geospatial data using the Open Packaging Convention. This mechanism could also be adopted within the CITS Geospatial by creating a Long Term Preservation Format Profile that is compliant with the criteria as described in this document.

In conclusion, CITS Geospatial supports most of the content elements as described in ISO 19165-1:2018. It also supports the storage of the Metadata content. However, the package does not use the preservation elements in the same way. Since both structures are written to be machine-readable, a migration specification could be developed to support automated adoption.

2.1.5 Geospatial records in linked data environment

Linked data is one of the upcoming technologies, that offers greater connectivity between otherwise siloed data stores, bringing greater value to the data environment. At its core, Linked Data is about creating a web of interconnected data. It achieves this using:

- **URIs** (Uniform Resource Identifiers): Unique labels to identify abstract or physical resources (such as people, places, datasets, or concepts) unambiguously.

- **RDF (Resource Description Framework)**⁵: A standardized model for representing data as "triples." A triple consists of a subject (the thing being described), a predicate (the property or relationship), and an object (the value of the property).
- **Semantic Web Technologies**: Standardized languages like RDF, SPARQL (a query language for RDF), and OWL⁶ (an ontology language) that allow for reasoning and inference over linked data.
- **Ontologies**: Provide formal structure for preserving knowledge and relationships between different types of entities and can serve as a structure to help AI solutions, like LLMs (Large language models) return better results.

2.1.5.1 *Linked Geospatial Data: Adding the Spatial Dimension*

Linked geospatial data takes the principles of Linked Data and applies them specifically to geospatial information. Here's how it works:

- **Spatial Vocabularies**: Standardized vocabularies like GeoSPARQL⁷ provide ways to represent spatial relationships (e.g., "contains," "overlaps," "is near") within RDF triples.
- **Geospatial URIs**: Geographic entities (a city, a river, a building) are assigned unique URIs for disambiguation.
- **Geospatial Ontologies**: Ontologies provide the formal structure and definitions of geospatial concepts and their relationships. They can also represent the structure of data and can contain some logic

The benefits of Linked Geospatial Data are mostly seen in:

- **Enhanced Interoperability**: By breaking down data silos, linked geospatial data promotes seamless exchange and integration of data from different sources.
- **Discoverability**: Linked datasets are inherently discoverable on the Web, facilitating data reuse and the discovery of new relationships between datasets.
- **Semantic Enrichment**: Ontologies add context and meaning to geospatial data, enabling more powerful analysis and reasoning.
- **Cross-Domain Insights**: Connecting geospatial data with other Linked Datasets (e.g., demographics, weather) enables insights that cut across traditional disciplinary boundaries.

2.1.5.2 *Long-term preservation aspects of linked geospatial data*

From the preservation standpoint, this presents a challenge to current preservation paradigms. At its core, the archival preservation package should be a self-descriptive stand-alone informational unit. However, linked data has a different nature as it relies primarily on connections between different repositories.

In its current version, this guideline only addresses the question of where certain types of elements from the linked data environments can be stored in the geospatial archival package. In the future, we aim to provide further guidance on dealing with the intricacies of linked data, however, this is topic affects all preservation domains and is not specific to geospatial.

⁵ RDF – Resource description Framework - <https://www.w3.org/RDF/>

⁶ OWL – Web Ontology Language - <https://www.w3.org/OWL/>

⁷ GeoSPARQL - <https://www.ogc.org/standard/geosparql/>

2.2 Significant properties of geospatial data

The fundamental challenge of digital preservation is to preserve the accessibility and authenticity of digital objects over time and domains and across changing technical environments.

Significant properties are those aspects of the digital object which must be preserved over time for the digital object to remain accessible and meaningful. An institution with curatorial responsibility for digital objects cannot assert or demonstrate the continued authenticity of those objects over time or across transformation processes unless it can identify, measure, and declare the specific properties on which that authenticity depends. Nor can it undertake the preservation actions required to maintain access to those objects unless it can characterise their current technical representations with sufficient detail.

The InSPECT Framework⁸ report provides a method on how to decide on whether a property of a digital object is significant.

The significant properties of spatial vector and raster data are listed below using the following categories:

- **Content** – Information contained within the Information Object. For example, pixel values and location information (coordinates, orientation, pixel size), feature geometry, related feature attributes. This is usually referred to as data.
- **Context** – Any information that describes the environment in which the content was created, or that affects its intended meaning. For example, creator name, date of creation, spatial accuracy, lineage, source data, sensor information, etc. This is usually referred to as descriptive documentation or metadata.
- **Structure** – Information that describes the extrinsic or intrinsic relationship between two or more types of content, as required to reconstruct the performance. For example, the connection between the vector datasets and their joined tables, the relation between the image file and the world file or how feature datasets are organised within a GIS Project or a Web service mash-up. Structure can be described in the data itself or in an external documentation like a feature catalogue, a GIS project configuration file or metadata related to the data.
- **Rendering** – Any information that contributes to the recreation of the performance of the Information Object. For example, a colour map of pixel values, Styled Description layer for web services, documentation from a cartographic project etc. This is usually referred to as rendering documentation.
- **Behaviour** – Properties that indicate the method in which content interacts with other stimuli. For example, zoom feature, rendering algorithms, analysis functionalities, common transformation processes, documentation of original system functionality, user manuals, training materials, videos of system usage, etc. This is usually referred to as documentation or metadata.

These categories are used as folder names within the Information Package in which such documentation is stored.

⁸ Significant properties described in the INSPECT Framework document (<https://significantproperties.kdl.kcl.ac.uk/>)

2.3 Recommended reading list

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

Introduction to geodata and GIS

- GIS File formats: (https://en.wikipedia.org/wiki/GIS_file_formats)
- Geographical Information Systems (https://en.wikipedia.org/wiki/Geographic_information_system)
- Shashi Shekhar, Hui Xiong, Xun Zhou. Encyclopedia of GIS, Second Edition. Springer 2017, Print ISBN 978-3-319-17886-8, <https://doi.org/10.1007/978-3-319-17885-1>

Preservation of geodata

- Descriptions of geospatial formats and their suitability for long-term preservation <https://www.loc.gov/preservation/digital/formats/content/gis.shtml>
- More documentation on the preservation of geospatial data <http://geopreservation.org/>
- Description of Significant properties concept. <https://significantproperties.kdl.kcl.ac.uk/>

3 Rationale for requirements in CITS Geospatial

This 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, all the CITS Geospatial document requirements are listed and explained in greater detail. Additional explanation contains a more extensive *description*, an *example* and a *rationale* for why the specific requirement stands are 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 CITS_Geodata compliant. 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 for a better reference.

3.1 Folder structure requirements

CITS Geospatial text:

The CITS Geospatial information structure inherited its package structure from the E-ARK Common Specification for Information Packages (CSIP) (blue elements), and it is an extension of it (green elements).

A visualisation of a valid CITS Geospatial Information Package is illustrated in Figure 2. This Figure shows an example of a simple valid Information Package with one representation of a single vector dataset in a GML file format.

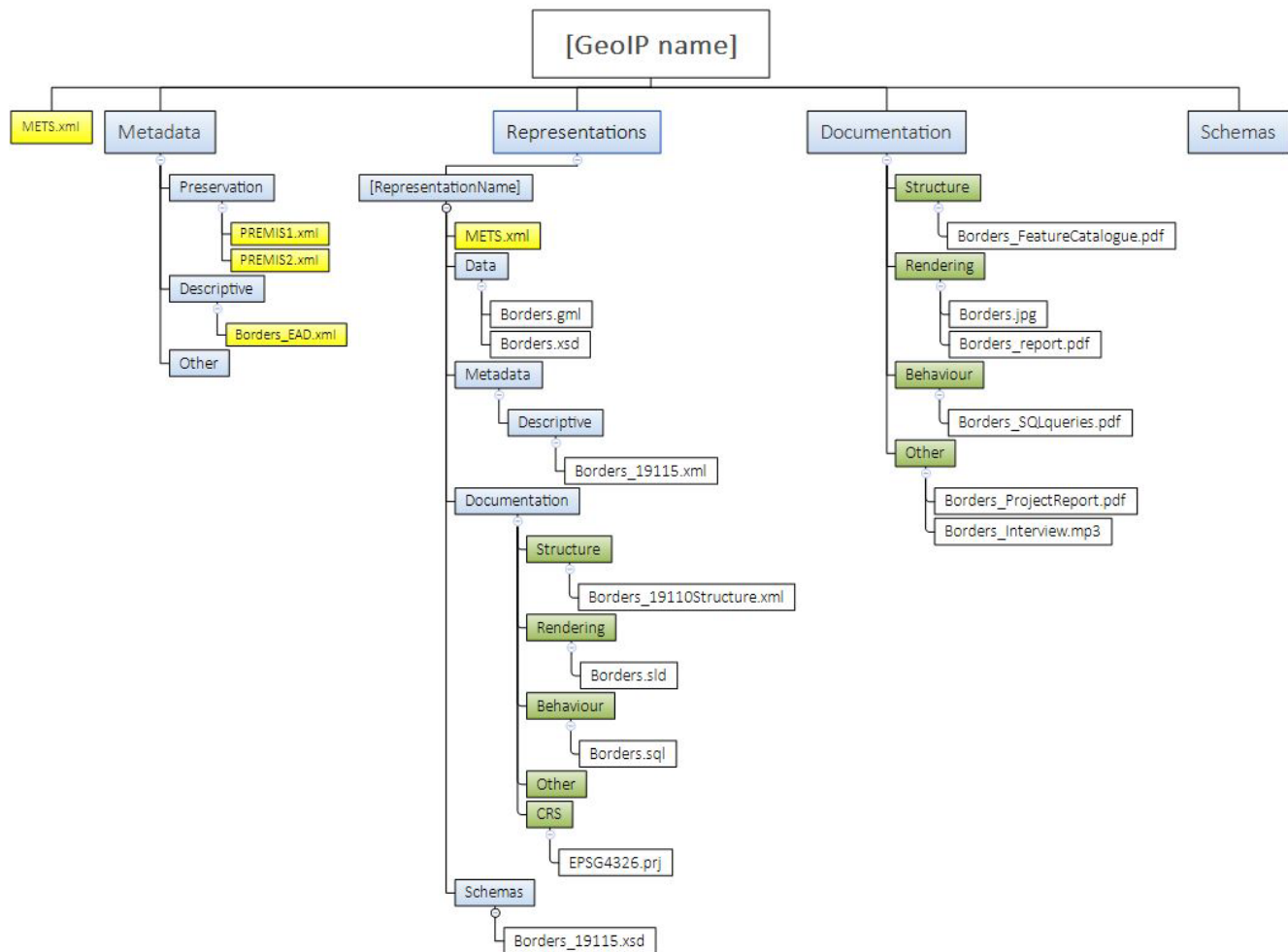


Figure 2 - Example Information Package folder structure

The folder structure in CSIP described in section <https://earkcsip.dilcis.eu/#folderstructureofthecsip> is extended with the following geo specific requirements on the folder structure:

In Figure 2 above from the CITS Geospatial specification, we can see a rather simple example of an Information Package, organised based on CSIP rules and extended with CITS for Geospatial folders. The predefined folders of the Information Package specified in the CSIP rules are in blue boxes, and the extended folders as introduced in the CSIP Geospatial are in green boxes. The predefined package Metadata files in the CSIP rules are in yellow boxes, and the example content of the Geospatial Information Package are in white boxes. Figure 2 is further explained in sections 3.2.

The purpose of data organisation into specific folders is to have known place holders for automated access of the standardised machine-readable content for faster reuse in the future. This is not the only possible approach, but it is in line with the philosophy of the CSIP logic, where we organise content within the Information Package according to the folder structure.

Other possible approaches could use a document pointing out where this type of data and documentation are within the package and how they connect to each other. There are different technologies that support this, like descriptions within METS files, OPC (Open Packaging Convention) files as mentioned within ISO 19165-1 standards, RDF and others. However, the whole idea is to have a commonly accepted convention that is interoperable among the EU member states and others interested in a common digital market.

GEOSTR1 rationale

Requirement:

GEOSTR1: XML schema documents for any structured descriptive geospatial metadata within a package **MUST** be placed in a sub-folder called `schemas` within the Information Package root folder and/or the representation folder. This requirement is an extension of CSIPSTR15.

Description:

An XML schema document is used for validation of structured XML files. XML schema documents must be placed in schemas folders in the Information package. If different XML schema definitions are used for validation of different structured descriptive geospatial metadata files in different representations, the XML schema document should be placed in a schema folder at representation level. If the same XML schema definition can be used for validation of all structured descriptive geospatial metadata files in the Information Package, it should be placed in a schema folder at the package level.

Example:

See Figure 2 above from the CITS Geospatial specification where the file `Borders_19115.xsd` (an XML schema definition file for validation of the `Border_19115.xml` file) is placed in a `representation/schemas` folder.

Rationale:

This requirement ensures that an XML schema definition file is easy to find when a structured descriptive geospatial metadata file must be validated against it.

GEOSTR2 rationale

Requirement:

GEOSTR2: A documentation folder on package or representation level **SHOULD** include a subfolder named `structure`. This requirement is an extension of CSIPSTR16.

Description:

A folder named `structure` should exist in the Information Package. This folder contains documentation about the extrinsic or intrinsic relationship between two or more types of content in the geospatial record, either in the original GIS or in the Information Package. This can be information about the connection between the vector datasets and their joined tables, the relation between an image file and a world file or how feature datasets are organised within a GIS Project or a Web service mash-up.

The `structure` folder can be placed at the package level if it contains a general description of the structure relevant for all geospatial data in the Information Package or at the representation level if the structure information only covers a specific representation in the Information Package.

Example:

See the folder `structure` in the path

`representations/[RepresentationName]/documentation/structure` in figure 2 above from the CITS Geospatial specification. In this example, the `structure` folder stores the file `Borders_19110Structure.xml`, which is a Feature Catalogue in a standardised xml-based file compliant with ISO 19110, describing the geospatial data in the `Borders.gml` file.

Rationale:

This requirement ensures that documentation of the structure of the geospatial dataset can be easily found by the user. Information about the structure of the geospatial record is required to reconstruct the performance.

GEOSTR3 rationale

Requirement:

GEOSTR3: A documentation folder on package or representation level **SHOULD** include a subfolder named `rendering`. This requirement is an extension of CSIPSTR16.

Description:

A folder named `rendering` should exist in the Information Package. This folder contains rendering information like a colour map of pixel values, Styled Description layer for web services, documentation from a cartographic project etc.

The `rendering` folder can be placed at the package level if it contains rendering information relevant for all geospatial data in the Information Package or at the representation level if the rendering information only covers a specific representation in the Information Package.

Example:

See the folder `rendering` in path

`representations/[RepresentationName]/documentation/rendering` in figure 2 above from the CITS Geospatial specification.

Rationale:

This requirement ensures that documentation about the rendering and visualisation of the geospatial dataset can be easily found by the user. Information about the rendering of the geospatial record is required to recreate the performance of the Information Object.

GEOSTR4 rationale

Requirement:

GEOSTR4: A documentation folder on package or representation level **SHOULD** include a subfolder named `behaviour`. This requirement is an extension of CSIPSTR16.

Description:

A folder named `behaviour` should exist in the Information Package. This folder contains documentation about methods in which content interacts with other stimuli (For example, the zoom feature, rendering algorithms, analysis functionalities, common transformation processes, documentation of original system functionality, user manuals, training materials, videos of system usage, etc).

Example:

See the folder `behaviour` in the path `representations/[RepresentationName]/documentation/behaviour` in figure 2 above from the CITS Geospatial specification.

Rationale:

This requirement ensures that documentation about behaviour originally executed on the geospatial dataset can be easily found by the user. Information about behaviour is required to recreate the original usage of the geospatial dataset.

GEOSTR5 rationale

Requirement:

GEOSTR5: A documentation folder on package or representation level **SHOULD** include a subfolder named `CRS`. This requirement is an extension of CSIPSTR16.

Description:

A folder named `CRS` should exist in the Information Package. This folder contains full descriptions of the Coordinate Reference System used in the archived geospatial dataset in the Information package.

Example:

See the folder `CRS` in the path `representations/[RepresentationName]/documentation/CRS` in figure 2 above from the CITS Geospatial specification.

Rationale:

This requirement ensures that documentation about CRS can be easily found by the user for usage or migration of the geospatial collection. Information about CRS used in geospatial records 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.

GEOSTR6 rationale

Requirement:

GEOSTR6: A documentation folder on package or representation level **SHOULD** include a subfolder named `other`. This requirement is an extension of CSIPSTR16.

Description:

A folder named `other` should exist in the Information Package. This folder contains other contextual information about the geospatial records, see section 3.1.2.

Example:

See the folder `other` in the path `representations/[RepresentationName]/documentation/other` in figure 2 above from the CITS Geospatial specification.

Rationale:

This requirement ensures that other contextual documentation can be easily found by the user. See also section Rationales in 3.4.5 Other - Contextual Documentation requirements.

Folder structure examples explained

3.1.1 Geo IP containing one vector representation

Below Figure 2 above is explained further.

3.1.1.1 Content

All the actual **Content** is placed in the `data` folder of the representation in figure 2 above. In this example, the `Borders.gml` file is considered as the **Content** (the Information contained within the Information Object) and thus placed in this folder. The accompanying `Borders.xsd` file is also considered as **Content**, even though it is actually a schema of the xml-based gml file, but it is in this case considered as part of the gml file.

3.1.1.2 Context

The `metadata` folder at the representation level stores the *descriptive* Geospatial metadata about the representation in a standardised machine-readable xml-file, compliant with geospatial metadata standards as described in the chapter 3.5 of this document. Thus, this file is considered as **Context** (Any information that describes the content or environment in which the content was created or that affects its intended meaning). The aim of this placement is to facilitate the automation of access to geospatial metadata.

Any other, non-standardised, documentation in digital form, containing **Context** information could be placed in the `documentation/other` folder at the package level like files `Borders_ProjectReport.pdf` and `Borders_interview.mp3`. If the information Package contains more representations, and there is documentation specific to a specific representation, then it could be placed within the `documentation/other` folder at the representation level (empty in Figure 2.).

The folder `Representations/[RepresentationName]Documentation/CRS` is a possible placeholder for full documentation of the used Coordinate Reference Systems in cases where the content files, in our case `Borders.gml`, do not contain a full definition of the CRS. In the case of GML, it only points to an external CRS repository. More on this topic in chapter 3.4.4.

3.1.1.3 Structure

In Figure 2, the folder `representation/[RepresentationName]/documentation/structure`, we see the file `Borders_19110Structure.xml`, which represents a Feature Catalogue, defining the `Borders.gml` and is a standardised xml-based file compliant with ISO 19110.

On the other hand, there is a folder `documentation/Structure`, containing a file `Borders_FeatureCatalogue.pdf` which contains a classic document describing the Feature Catalogue, however, this document is more descriptive and probably less suitable for automated machine-readable applications.

Generally, we aim to store the standardised machine-readable files within the representation. If the content is not available in such form, then it is generally stored within the package level `documentation` folder. Unless it is representation specific, then we recommend storing it within the representation.

3.1.1.4 Rendering

In Figure 2., the folder `representation/[RepresentationName]/documentation/rendering`, we see the file `Borders.sld`, which represents a standardised machine-readable file, that could be automatically used when rendering the `Borders.gml` within a web service.

On the other hand, there is a folder `documentation/rendering`, containing files `Borders.jpg` and `Borders_report.pdf`. These files also show how the data was originally rendered in the initial system, and the report represents an example of a derived information product to help us recreate adequate information products in potential future systems.

3.1.1.5 Behaviour

In Figure 2., the folder `representation/[RepresentationName]/documentation/behaviour`, we see the file `Borders.sql`, which represents a standardised machine-readable file that could be automatically used when trying to replicate the behaviour of the initial system to reproduce the information products like, the before mentioned `Borders_report.pdf`.

On the other hand, there is a folder `documentation/behaviour` that can contain additional extensive documentation on the initial system, its design, architecture, the use of the `Borders.sql` code and expected outputs.

3.1.1.6 Package level elements

This Information package also holds the two mandatory **METS.xml** files, one at package level (see Rationales in 3.2.2 Package METS requirements) and one at representation level (see Rationales in 3.2.2 Representation METS requirements).

Package level descriptive metadata (**EAD.xml**) and *preservation* metadata (**PREMIS.xml**) describing the Information Package are placed in the `metadata` folder at the package level. This is compliant with CSIP.

Generally, all Schema files (**.xsd**) for validating the different xml files in the Information Package are located in the `schemas` folder at the package level (empty on Figure 2.). However, there could be exceptions. If there are schemas specific to a representation (for example, one representation contains GML version 3.1.1 and the other version 3.2.1.), we can add a `Schemas` folder within a representation.

3.1.2 Geo IP containing multiple vector representations

In the example in Figure 3 below, a Geospatial Information Package (Geo IP) 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 interpret both representations correctly is put in the package level `documentation` folder. It is also possible to include a logical link to point to additional documentation being stored in a different Information Package (in case of a more extensive time series of the same data or similar records but from different organisational units).

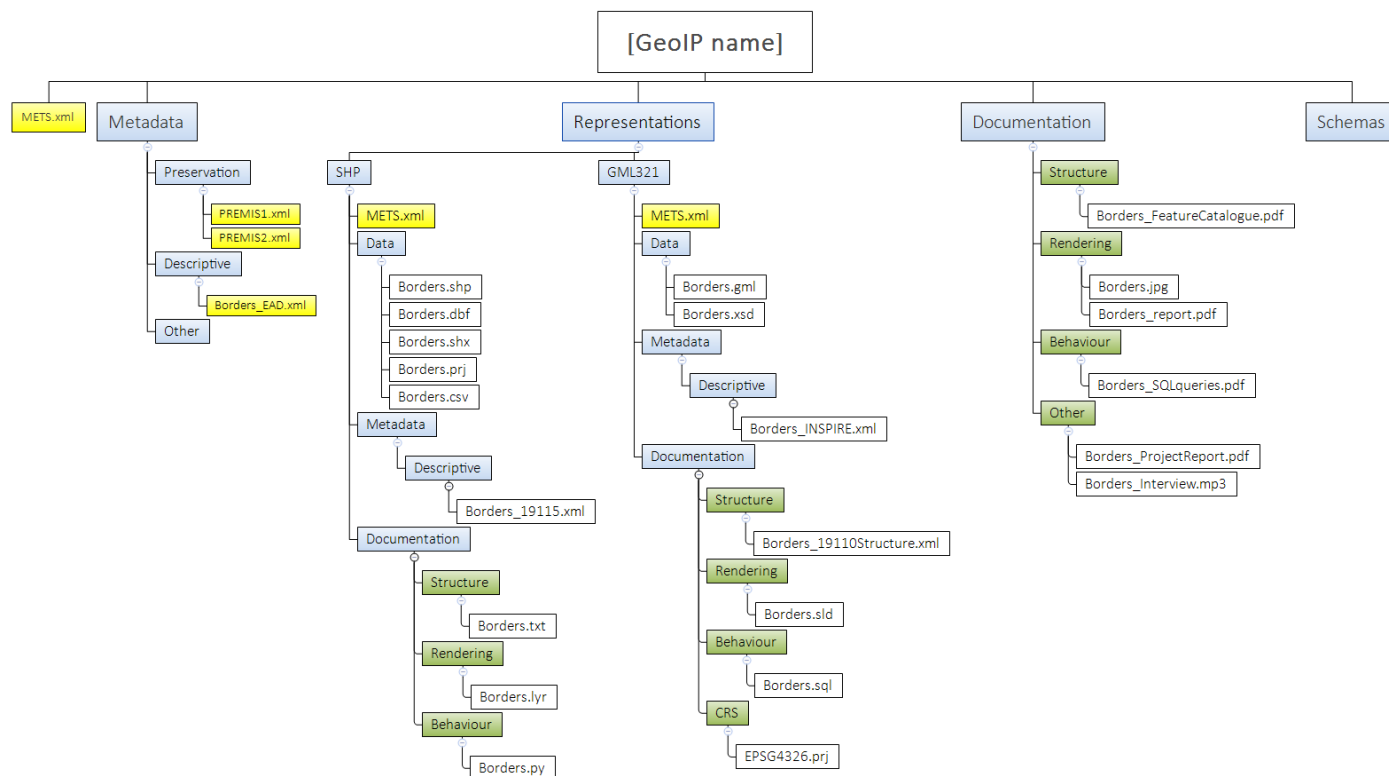


Figure 3 -Geospatial Information package with two representations containing vector data

Key differences between the representations are:

- Geospatial data in original format in representation SHP contains additional information in a separate non-spatial table (Borders.csv)
- Geospatial feature structure is described within a non-standard txt file in representation “SHP”
- Geospatial metadata in the original format is from a local system, compliant with ISO 19115:2003 (Borders_19115_export.xml). In representation GML321, it is updated to contain the INSPIRE compliant metadata.
- Since the GML321 representation references an EPSG register of CRS, we need to add a separate definition of the CRS (ETRS4326.prj) in the technical Documentation (exported from the EPSG registry). See chapter 3.4.4.
- Additional documentation for both representations is stored in the Root Documentation folder.

We can also see that some documentation, not in standardised machine-readable form, was placed under the root documentation folder.

3.1.3 Geo IP containing one representation of multiple raster datasets

In the example in Figure 4, a Geo IP contains one representation of multiple raster images covering an area with an accompanying vector file – containing an index of distribution and positions of the raster images. All additional documentation for the raster datasets is located in the root `Documentation` folder, except for additional CRS information in the `CRS` folder and a Geospatial Metadata xml file in the `Metadata/Descriptive` folder.

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

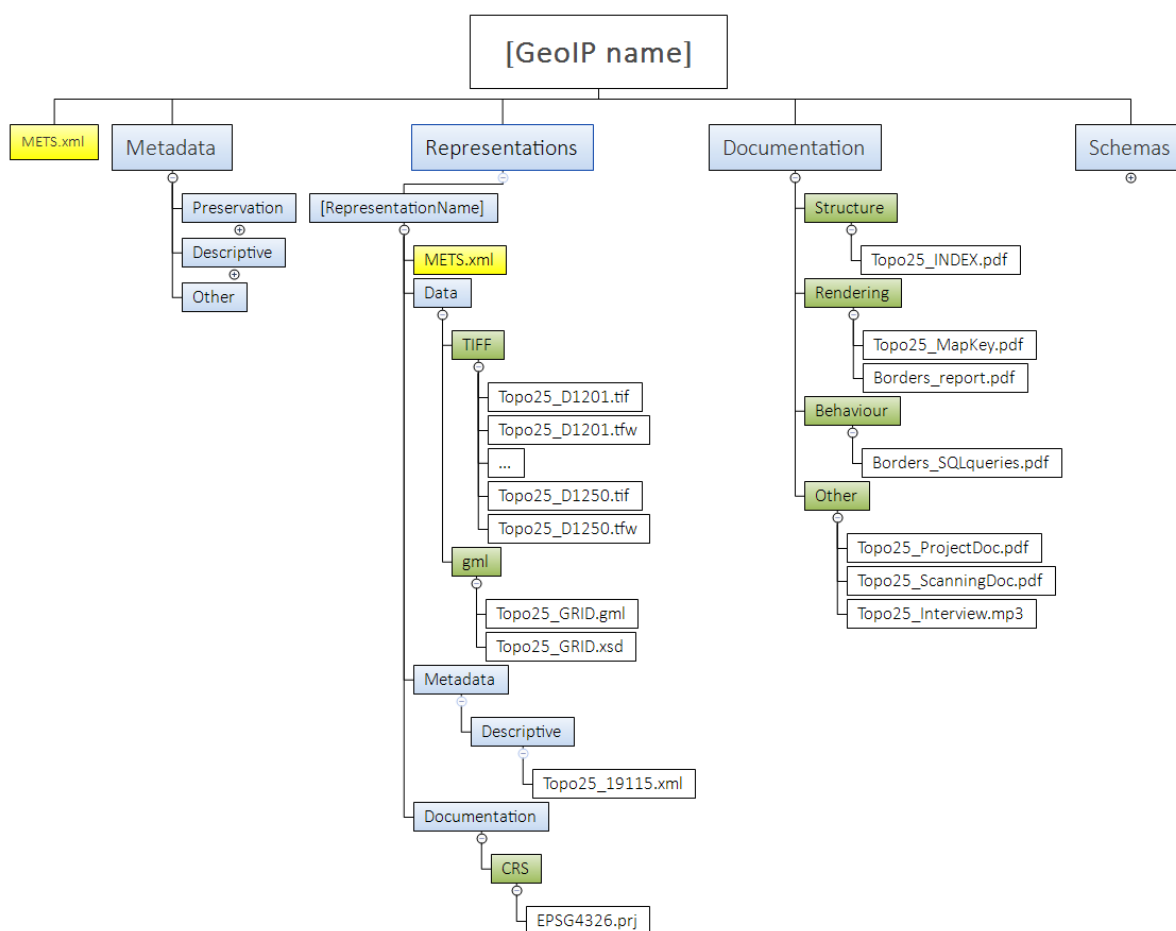


Figure 4 - Folder structure of a Geo IP containing one representation of multiple rasters

In this example, we can see that the Raster and Vector content within the `Data` folder can be organised in various ways. It could represent the structure within the initial system. However, to support more automated validation and future preservation actions it is recommended that geospatial records in different formats are stored in separate folders. In this example, GML files are stored within GML folder, and Raster data is stored in TIFF folder.

3.2 Rationales in METS Requirements

CITS Geospatial text:

A CSIP can consist of zero to many representations, and this is an important feature that needs to be taken into consideration when packing geodata files within CSIPs.

There can easily be different representations of the same geodata records located within one CSIP. For example, one package could consist of:

- one representation with geodata in original format;
- one representation with geodata in a long-term preservation format;
- one representation with geodata in dissemination formats;

There can be several representations of dissemination formats. There can also be many different types of geodata records and databases within the same package.

As for the CITS Geospatial specification, there always needs to be a minimum of one representation and, therefore a minimum of two METS.xml. The Package METS.xml has to be a general METS.xml describing if the package itself is mainly a CITS Geodata package, and then the Representation METS.xml describing what specific main Geodata types (Primarily vector or raster) the representation consists of.

A CITS Geospatial is building upon the general CSIP requirements but do not mention them here. Those requirements should be met before applying the requirements listed below.

GEO_1 Rationale

Requirement:

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

Description:

Geospatial records are placed in the Data folder in a Representation folder in a CSIP. There must be at least one representation of the geospatial data in a representation folder in the IP. Also, the IP must contain only one Package METS.xml file and one Representation METS.xml file in each representation folder.

This first requirement is central for the CITS Geospatial specification since it operates with two central terms: the Package METS.xml and the Representation METS.xml.

The "Package METS.xml" (there is only one) needs to be in the root of the package, and one "Representation METS.xml" needs to exist in the root of each representation within the package.

According to the CSIP specification, it is up to the user to define whether all files in an IP are described in the "Package METS.xml" or whether the user wishes to split it up and let "Representation METS.xml" describe the content within the representations. In the CITS Geospatial specification files in a representation folder are described in a "Representation METS.xml".

Example:

See the example in Figure 2, containing one Package level METS and two Representation level METS files.

Rationale:

An Information package without geospatial content stored within Representation folders doesn't qualify for a geospatial package. It could qualify for a general CSIP Package.

Rationales in 3.2.2 Package METS requirements

GEO_2 Rationale

Requirement:

<p>GEO_2 Ref CSIP2</p>	<p>Type mets/@TYPE</p>	<p>For information packages that primarily contain geospatial data, the value in Package mets/@TYPE MUST be "Geospatial Data" as taken from the CSIP Vocabulary for Content Category.</p>	<p>1..1 MUST</p>
----------------------------	----------------------------	--	----------------------

Description:

This requirement ensures that the IP is compliant with the requirement CSIP2 in the CSIP specification, which states that there **MUST** be a TYPE-attribute with a value taken from the provided vocabulary for Content Category (<https://earkcsip.dilcis.eu/schema/CSIPVocabularyContentCategory.xml>).

"Geospatial Data" is the most relevant value from the Content Category vocabulary found in CSIP.

Example:

TYPE="Geospatial Data"

Rationale:

This information in the "Package METS.xml" can be used in a Finding Aid to group the IP as an IP that primarily contains "Geospatial Data".

GEO_3 Rationale

Requirement:

<p>GEO_3 Ref CSIP4</p>	<p>Content Information Type Specification mets/@csip:CONTENTINFORMATIONTYPE</p>	<p>For information packages that primarily contain geospatial data, the value in Package mets/@csip:CONTENTINFORMATIONTYPE MUST be "citsgeospatial_v3_0" as taken from the CSIP Vocabulary for Detailed Content Type.</p>	<p>1..1 MUST</p>
----------------------------	---	--	----------------------

Description:

This requirement is to make sure that the IP is compliant with the requirement CSIP4 in CSIP, which is a primary way of handling which kind of Content Information Type the package contains. In the case of multiple Content Information Types in an IP the value "MIXED" should be used.

Example:

csip:CONTENTINFORMATIONTYPE=" citsgeospatial_v3_0 "

Rationale:

When the “citsgeospatial_v3_0” value is used, this means that the package can be identified as stated to live up to the CITS Geospatial specification, and therefore be validated based on requirements in this specification.

GEO_4 Rationale

Requirement:

GEO_4 Ref CSIP5	Other Content Information Type Specification mets/@csip:OTHERCONTENTINFORMATIONTYPE	For information packages that primarily contain geospatial data, the Package METS MUST NOT have a mets/@csip:OTHERCONTENTINFORMATIONTYPE	0..0 MUST NOT
--------------------	--	--	----------------------

Description:

The csip:OTHERCONTENTINFORMATIONTYPE-attribute is meant to specify which Content Information Type is used if the csip:CONTENTINFORMATIONTYPE-attribute has the value “OTHER”. It is not meant to exist if there are multiple Content Information Types in an IP. In the case of multiple Content Information Types, then the value “MIXED” should be used. Therefore, for CITS_Geospatial packages, there must not be a csip:OTHERCONTENTINFORMATIONTYPE-attribute in “Package METS.xml”. Note that this is different from the “Representation METS.xml”.

Example:

The `csip:OTHERCONTENTINFORMATIONTYPE` must not be present in the Package METS.xml file.

Rationale:

Because if it is an csip:OTHERCONTENTINFORMATIONTYPE-attribute then it is not a valid “ CITS_Geospatial packages”.

GEO_5 Rationale

Requirement:

GEO_5 Ref CSIP6	METS Profile mets/@PROFILE	For information packages that primarily contain geospatial data, the value in the @PROFILE MUST be "https://citsgeospatial.dilcis.eu/profile/E-ARK-GEOSPATIAL-ROOT.xml "	1..1 MUST
--------------------	-----------------------------------	---	------------------

Description:

This requirement is to make sure that the IP is compliant with the requirement CSIP6 in CSIP.

Each Content Information Type Specification (CITS) has its own METS profile where further requirements are added to the CSIP profile. The “<https://citsgeospatial.dilcis.eu/profile/E-ARK-GEOSPATIAL-ROOT.xml>” is thus an extending profile adding requirements to the CSIP requirements or changing their cardinality by changing optional to mandatory or specifying the number of occurrences of an element. It is not allowed to remove requirements from the CSIP profile since this will make the implementation invalid.

Example:

PROFILE="https://citsgeospatial.dilcis.eu/profile/E-ARK-GEOSPATIAL-ROOT.xml "

Rationale:

The CITS Geospatial METS profile is created for validation purposes. As per 9 April 2021 the profile has not yet been created, but it is planned.

GEO_6 RationaleRequirement:

<p>GEO_6</p> <p>Ref CSIP62</p>	<p>fileSec Representation Content Information Type Specification</p> <p>mets/fileSec/fileGrp[@USE='Representations']/@csip:CONTENTINFORMATIONTYPE</p>	<p>There MUST be a minimum of one mets/fileSec/fileGrp[@USE='Representations']/@csip:CONTENTINFORMATIONTYPE with the value "citsgeospatial_v3_0" as taken from the CSIP Vocabulary for Detailed Content Type that direct to the representation METS.xml in the representation folder containing geospatial data.</p>	<p>1..n</p> <p>MUST</p>
--------------------------------	---	---	-------------------------

Description:

In this requirement, a filegroup (fileGrp) is named "Representations" in the "Package METS.xml" file. It is via the value "Representations" in the fileGroup USE-attribute on the filegroup element that one can mark up that within this filegroup will be a fileSec with a path to one or more METS-files in one or more representations. One METS file per representation.

Thus, this filegroup named "Representations" holds all the paths to the "Representation METS.xml" files in the IP.

Example:

```
<fileGrp USE="Representations" csip:OAISPACKAGE="SIP" csip:CONTENTINFORMATIONTYPE="citsgeospatial_v3_0" ID="ID_Rep1">
```

```
  <file ID="vector_protected_areas_METS.xml" USE="OTHER" MIMETYPE="application/xml"
    CREATED="2015-12-14T14:20:00"
```

```
      CHECKSUM="90c7527e6d4d3c3a6247ceb94b46bcf5" CHECKSUMTYPE="MD5" SIZE="8322">
```

```
        <FLocat LOCTYPE="URL" xlink:href="representations\repl\METS.xml"
          xlink:type="simple"/>
```

```
      </file>
```

```
</fileGrp>
```

GEO_7 Rationale

Requirement:

GEO_7 Ref CSIP105- CSIP112	StructMap METS pointer	For any fileGrp/@csip:CONTENTINFORMATIONTYPE with the value "citsgeospatial_v3_0" there MUST be a corresponding @div- representation in the StructMap-element	1..1 MUST
---	---------------------------	---	------------------

This requirement is to make sure that the IP lives up to the requirement CSIP105 to CSIP112 in CSIP. They are all related to "how to create a structMap-element". The structMap element holds all the internal links to folders and files in the IP. The link/URL is inserted in the xlink:ref element in the example below.

The METS structural map element is the only mandatory element in the METS specification. In the CSIP the <structMap> describes the higher-level structure of all the content in the package and may link to representations in the IP.

CSIP105 states that when a package consists of multiple representations, each described by a representation level METS.xml document, there is a discrete representation div element for each representation.

Example:

```
<structMap TYPE="PHYSICAL" LABEL="CSIP" ID="StructmapID_rep1">
  ...
  <div ID="Structmap_Div_ID_Representations" LABEL="Representations"/>
  <div ID="struct-map-reps-sub-div" LABEL="Representations/rep1">

    <mptr LOCTYPE="URL" xlink:type="simple"
      xlink:href="representations/rep1/METS.xml" xlink:title="ID_Rep1" >
    </mptr>

  </div>
  ...
</structMap>
```

Rationale:

Each representation div references the representation level METS.xml document, documenting the structure of the package and its constituent representations.

Rationales in 3.2.3 Representation METS requirements

Many of the requirements in this section are the same as in section 3.3 - it is important to notice the differences.

GEO_8 Rationale

Requirement:

GEO_8	Type	For representations that primarily contain geospatial data, the value in Package mets/@TYPE MUST be "Geospatial Data" as taken from the CSIP Vocabulary for Content Category.	1..1
Ref CSIP2	mets/@TYPE		MUST

Description:

The same as GEO_2. This requirement ensures that the IP is compliant with the requirement CSIP2 in the CSIP specification, which states that there **MUST** be a TYPE-attribute with a value taken from the provided vocabulary for Content Category (<https://earkcsip.dilcis.eu/schema/CSIPVocabularyContentCategory.xml>).

"Geospatial Data" is the most relevant value from the Content Category vocabulary found in CSIP.

Example:

TYPE="Geospatial Data"

Rationale:

This information in a "Representation METS.xml" can be used in a Finding Aid to group the IP as an IP that primarily contains "Geospatial Data".

GEO_9 Rationale

Requirement:

GEO_9	Content Information Type Specification	For representations that primarily contain geospatial data and that conforms to CITS Geodata the value in Package mets/@csip:CONTENTINFORMATIONTYPE MUST be "citsgeospatial_v3_0" as taken from the CSIP Vocabulary for Detailed Content Type.	1..1
Ref CSIP4	mets/@csip:CONTENTINFORMATIONTYPE		MUST

Description:

The same as SIARD_3⁹. This requirement is to make sure that the IP is compliant with the requirement CSIP4 in CSIP which is a central way of handling which kind of content information type the package contains, in this case it is the representation.

Example:

csip:CONTENTINFORMATIONTYPE=" citsgeospatial_v3_0 "

⁹ See the CITS SIARD specification requirements.

Rationale:

When the “citsgeospatial_v3_0” value is used, this means that the package can be identified as stated to live up to the CITS Geospatial specification, and therefore be validated based on requirements in this specification.

GEO_10 RationaleRequirement:

GEO_10	METS Profile	For information packages that primarily contain geospatial data the value in the @PROFILE MUST be	1..1
Ref CSIP6	mets/@PROFILE	"https://citsgeospatial.dilcis.eu/profile/E-ARK-GEOSPATIAL-REPRESENTATION.xml "	MUST

Description:

This requirement is to make sure that the IP is compliant with the requirement CSIP6 in CSIP.

Each Content Information Type Specification (CITS) has its own METS profile where further requirements are added to the CSIP profile. The “https://citsgeospatial.dilcis.eu/profile/E-ARK-GEOSPATIAL-REPRESENTATION.xml” is thus an extending profile adding requirements to the CSIP requirements or changing their cardinality by changing optional to mandatory or specifying the number of occurrences of an element. It is not allowed to remove requirements from the CSIP profile since this will make the implementation invalid.

Example:

```
PROFILE="https://citsgeospatial.dilcis.eu/profile/E-ARK-GEOSPATIAL-REPRESENTATION.xml "
```

Rationale:

The CITS Geospatial METS profile have been created for validation purposes.

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.

Sections 3.3 – 3.5 of this document do not discuss optimisations with respect to packaging and storage. The requirements for data, metadata and documentation only suggest what information is needed and the appropriate placement of it, not how it is packaged, stored and optimised for automatic handling.

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 MUST 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. However if a geospatial information product is created exclusively by linking to external data sources, an archival package can still be created as long as it contains proper links or references to external data (see also 2.1.5 on linked data).

Example:

A good example is the image "Figure 2." in chapter 3.1. The example illustrates an Information Package with one geospatial record in the GML format in the Representations/[RepresentationName]/data folder.

Rationale:

The purpose of making the Information package compliant with CITS Geospatial is to validate it against the criteria as described in the specification.

GEO_12 Rationale

Requirement:

GEO_12	Subfolders in data representations/[RepresentationName]/data	If there are more geospatial records in a representation, each geospatial file MAY be placed or grouped in subfolders in representations/[RepresentationName]/data	0..n MAY
--------	--	---	-----------------

Description:

This requirement allows the use of subfolders in the data folder. Subfolders inside subfolders are also permitted.

Recommendation:

Use short names or IDs when naming subfolders. If a file path in an IP is longer than 256 characters, this can complicate an IP's validation or migration.

Example:

representations/[RepresentationName]/data/TIFF

See Figure 4 in chapter 3.1.3, where geospatial records are stored within a TIFF or GML folder.

Rationale:

All geospatial files can be placed directly in the root of the data folder, but this can be a mess of files if a representation in the IP contains more than one dataset.

Often it makes sense to group bundles of files. For example, all files belong to a dataset in one folder or all files in the same format in one folder. Grouping files with relations to each other, for example, a GML file and the .xsd schema file validating the GML file, also enables fast identification of files related to other files. Placing each file in a folder with a unique ID can be a way to identify a file uniquely based on the folder ID.

Grouping and identifications of files can also be described in a METS.xml file in the structMap section. If producers organise their data in a folder structure, it is sensible to store the data in the way they are used to finding it. Grouping based on folder structures is more visual and easier for humans to comprehend. This can also be an alternative way (a contingency plan) to navigate the IP's when preservation systems based on the METS.xml file information fails or cannot be used due to lack of GDPR compliance.

GEO_13 Rationale

Requirement:

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

Description:

This requirement ensures 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 Rationales in 3.3.5 Long-term Preservation format profiles). This is a requirement for Submission Information Packages (SIPs) and Archival Information Packages (AIPs), however, a Dissemination Information Package (DIP) can only contain a dissemination format representation.

The Information Package can also contain other representations containing records in original or non-long-term preservation formats see description in GEO_14.

Example:

As proposed in the Long-Term Preservation format Profile for Geospatial Vector data using GML 3.2.1 in Appendix 1, we can have vector geospatial data stored in a GML 3.2.1. format, along with other standardised machine-readable documentation. See Figure 2 in chapter 3.1.

Rationale:

The idea of long-term preservation formats is fundamental in archival practice, as it prevents loss of data in the future. Since the CITS Geospatial is a specification for the long-term preservation of geospatial data, one representation in the IP must 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:

Figure 3 in chapter 3.1.2 shows 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).

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 becomes obsolete.

GEO_15 Rationale

Requirement:

GEO_15 Ref GEO_15	CRS definition	Every geospatial dataset MUST be accompanied with information about its underlying Coordinate Reference System (CRS) in one of two ways: <ul style="list-style-type: none"> • 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/systems (CRS) used in the geospatial file stored is/are 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. 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 in GEO_38.

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 4326¹⁰.

```
<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>
```

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.

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.

GEO_16 Rationale

Requirement:

GEO_16	Geographic location validation	The geographies in the geospatial records SHOULD be located within a fixed bounding box defined in the submission agreement between the producer and the archive according to the expected location and extent of the dataset	0..1 SHOULD
--------	--------------------------------	--	----------------

Description:

This requirement is intended to check against the expected spatial extent of submitted geospatial records. This extent can be limited to the extent of national or administrative unit boundary or determined in the valorisation phase of the Pre-Ingest process.

Example:

¹⁰ World Geodetic System 1984 (https://epsg.org/crs_4326/WGS-84.html?sessionkey=z5jf4xc886)

An archive could decide that all geospatial data from a specific country should be inside a fixed bounding box covering the areas of this country. Note that a bounding box is rectangular and aligns the axis of the CRS. Thus, it covers more than the exact national borders of a country.

A bounding box defined in a GML file:

```
<gml:boundedBy>  
  <gml:Envelope srsName="EPSG:25832" srsDimension="2">  
    <gml:lowerCorner>212481.60 6019669.40</gml:lowerCorner>  
    <gml:upperCorner>961440.75 6510422.51</gml:upperCorner>  
  </gml:Envelope>  
</gml:boundedBy>
```



Figure 5 - Orange coordinates placed outside the expected bounding box due to error-prone migration

Rationale:

This requirement enables the validation of geospatial data migrated to preservation format from another CRS. When a transformation of coordinates from one CRS to another fails or has errors, the migrated coordinates are often placed outside the expected bounding box of the geospatial records, as illustrated in Figure 5.

In the exception where the geospatial records are located internationally or legitimately extend the proposed boundary, a new boundary should be determined by the owner of geospatial records and the archive, proposing the limitations.

GEO_17 Rationale

Requirement:

GEO_17	Metadata	Every geospatial dataset MUST be accompanied by a metadata file that describes the dataset with the basic required information	1..n MUST
--------	----------	---	------------------

Description:

This is a general requirement that ensures that all geospatial records within the Information Package are accompanied by descriptive metadata. If a geospatial information product is created exclusively by linking to external data sources the archival package still needs to include copies of metadata for external data used for the creation of the information project. Metadata can be stored in many different ways:

- With an accompanying standardised xml file (preferably to support automation)
- Using a database of metadata descriptions
- Containing metadata tags within the geospatial file (if the format supports it.)
- Within human-readable documentation (least preferred, but acceptable)

For further reference, see descriptions in GEO_42 and GEO_43

Example:

See examples from GEO_42 and GEO_43

Rationale:

Geospatial records commonly don't contain enough information to be fully self-descriptive. Therefore, it is necessary to follow standards for the description of geospatial records, using metadata that gives us additional context to understand and evaluate them.

Rationales in 3.3.2. Vector Geodata – requirements

CITS Geospatial text:

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

The requirements listed in this subchapter are specific to geospatial records in vector formats, contained within Information Package representations that primarily hold geospatial records and are marked as such in METS. If a representation contains mixed data types, it is still recommended that these validation rules are used for the geospatial vector data types.

GEO_18 Rationale

Requirement:

GEO_18	Valid geospatial vector file	Any geospatial vector datafile in representations/[RepresentationName]/data MUST be a valid	1..n MUST
--------	------------------------------	--	------------------

		vector file compliant with its respective format requirements (must pass the validation with the chosen validator for its format).	
--	--	--	--

Description:

This requirement ensures that the vector file format used for the geospatial vector data is valid.

Example:

A vector file in GML 3.2.1 format should be validated against the schema collection from Open Geospatial Consortium (OGC) to GML version 3.2.1.

Rationale:

To enable automated migration and dissemination of geospatial files in the archive, they must be valid according to format specifications.

GEO_19 Rationale

Requirement:

GEO_19	Feature attribute	Each Vector Feature dataset MUST contain at least one Feature attribute unique to each vector object	1..n MUST
--------	-------------------	---	------------------

Description:

This requirement states that each vector feature, for example, a polygon in a GML file used for outlining borders of countries, must be described by at least one attribute in some kind of feature dataset accompanying the geospatial vector file and identifying each instance of that feature uniquely.

Example:

See the example in Figure 1 where a polygon vector dataset containing borders of countries is set up with multiple vector features (in this case, polygons). Every feature is accompanied by at least one feature attribute. In Figure 1, a single vector feature has three feature attributes (ADMIN, ISO_A3, ISO_A2). The ADMIN feature attribute contains values for "Country name". Each instance of this feature attribute could hold the specific name of the country. And every vector instance has a unique name to differentiate them.

Rationale:

A vector shape (point, line, polygon, multipoint, multiline, multipolygon, etc.) needs to have an attribute to differentiate it from other objects in the dataset and to prevent duplicate entries. A unique feature attribute describes the geometry uniquely and gives it a reference point or a unique key, allowing other GIS functions like the joining of other attributes.

GEO_20 Rationale

Requirement:

GEO_20	Long-Term preservation format Profile for Geospatial Vector data	Geospatial vector data in a long-term preservation representation SHOULD comply with the requirements for the respective Long-Term preservation format Profile for Geospatial Vector data (see chapter 3.3.5)	0..n SHOULD
--------	--	--	--------------------

Description:

This requirement recommends that geospatial vector data be preserved in the preservation format specified in a Long-Term preservation format Profile for Geospatial Vector data (see Rationales in 3.3.5 Long-term Preservation format profiles). The CITS Geospatial specification doesn't directly specify a long-term preservation format. Instead, it allows for Long-term Preservation Format profiles to support various formats for different purposes and local implementations. The proposed example in Appendix 1 is a showcase to support the creation of one's own profile if one intends to use other formats for long-term preservation.

Example: An example of a Long-Term Preservation Format Profile for Geospatial Vector data using GML 3.2.1 is available in Appendix 1 of this document. It proposes criteria for the use of the GML 3.2.1 format for the long-term preservation of geospatial vector data.

Rationale:

Since at least one representation containing data in a long-term preservation format is required for the Information Package to be compliant with CITS Geospatial, this requirement specifies how exactly to do this for the vector geospatial data.

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:

The requirements listed in this subchapter are specific to geospatial records in raster formats, contained within Information Package representations that primarily hold geospatial records and are marked as such in METS. If a representation contains mixed data types, it is still recommended that these validation rules are used for the geospatial raster data types.

GEO_21 Rationale

Requirement:

GEO_21	Valid geospatial raster file	Any geospatial raster datafile in representations/[RepresentationName]/data MUST be a valid raster datafile compliant with its respective format requirements (must pass the validation with the chosen validator for its format).	1..n MUST
--------	------------------------------	---	------------------

Description:

This requirement ensures that the raster file format used for the geospatial raster data is valid.

Example:

A raster file in GeoTIFF could be validated against the OGC GeoTIFF standard.

Rationale:

To enable automated migration and dissemination of geospatial files in the archive they must be valid according to format specifications.

GEO_22 RationaleRequirement:

GEO_22	Long-Term preservation format Profile for Geospatial Raster data	Raster data in the long-term preservation representation SHOULD comply with the requirements for the respective Long-Term preservation format Profile for Geospatial Raster data (see chapter 3.3.5.)	0..n SHOULD
--------	--	--	-----------------------

Description:

This requirement recommends that geospatial raster data are preserved in the preservation format specified in a Long-Term preservation format Profile for Geospatial Raster data (see Rationales in 3.3.5 Long-term Preservation format profiles). The CITS Geospatial specification doesn't directly specify a long-term preservation format. It instead allows for Long-term Preservation Format profiles to support various formats for different purposes and local implementations.

Example: An example of a Long-Term Preservation Format Profile for Geospatial Raster data using TIFF Baseline 6 is available in Appendix 2 of this document, and it proposes criteria for usage of the TIFF format for the preservation of geospatial raster data.

Rationale:

Since at least one representation containing data in a long-term preservation format is required for the Information Package to be compliant with CITS Geospatial, this requirement specifies how to do this for the raster geospatial data.

GEO_23 RationaleRequirement:

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.

Example:

The map below illustrates a tiling index.

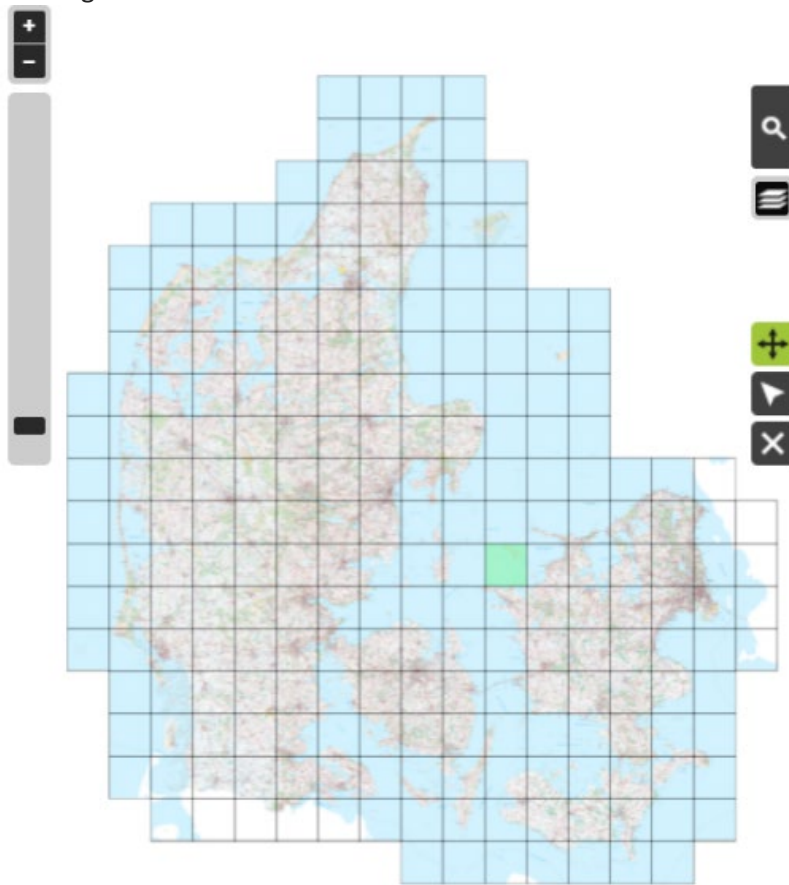


Figure 6 - Example of a tiled vector index for raster geodata

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.

Example:

Two examples of Long Term Preservation format Profiles are listed in these appendixes:

- Appendix 1: Long-Term preservation format Profile for Geospatial Vector data using GML 3.2.1
- Appendix 2: Long-Term preservation format Profile for Geospatial Raster data using TIFF baseline 6

Rationale:

The preservation of both the amount in formats and the formats' significant properties entails significant complexities for an archive's ability to safeguard digital preservation. If data in a digital archival collection is homogeneous and based on a few selected preservation formats, it results in the ability to provide access to archived data in the future and maintain cost-effective preservation. However, homogeneity often occurs at the expense of authenticity, and it is in this area of tension that it makes sense to use a framework with criteria for selecting long-term preservation formats in a structured and documented way.

Description:

An example of a framework for selecting long-term preservation formats using a matrix to score different criteria for each preservation format candidate is described below¹¹. Figure 7 illustrates the use of this framework, accessing a suitable preservation format for georeferenced raster images.

The format that achieves the highest total score is considered to be the most suitable preservation format. Overall assessment results in a recommendation written in note format, which can be taken further as the basis for a subjective management decision to approve a new preservation format.

Arguments for each score in the matrix shown in Figure 7 should be described as illustrated in Figure 8.

¹¹ Format assessment matrix, The Danish National Archives (<https://github.com/the-danish-national-archives/concept-model/tree/main/P2%20Format%20Assessment>)

Criterion	Weight	Format				
		TIFF+TFW	GeoTIFF	JPEG2000	Geopackage (GPKG)	ASCII GRID (*.asc)
Prevalence	2	1	1	1	0	1
Prevalence II	2	1	1	1	0	0
Lifespan	1	1	1	1	0	1
Lifespan II	3	0,5	1	1	1	0
Documentation	2	1	1	1	1	1
Documentation II	1	1	1	1	0	1
Documentation III	1	1	1	0	1	1
Licensing	2	1	1	0	1	1
Structure	1	1	1	1	0	1
Structure II	1	0	0	0	0	1
Significant properties	3	1	1	1	1	0
Dissemination	1	1	1	1	1	1
Searchability	1	0	0	0	0	1
Interoperability	2	1	1	1	1	1
Testing	2	1	1	1	0	0
Testing II	2	1	1	1	1	0
Compression	3	1	1	1	1	1
Storage	1	0	0	1	1	0
Migration	2	1	1	1	1	1
Compatibility	1	1	1	1	1	1
Score		29,5	31	29	23	21
		0,87	0,91	0,85	0,68	0,62

Figure 7 - Example of format assessment of preservation format candidates for georeferenced raster images

Each individual criteria for measuring how persistable a format is as well as the relative importance (weighting) of the criteria in the matrix in figure 7, is described in detail here:

Prevalence

The format is frequently used within its content type. It is a strength for preservation formats to be frequently used across user segments for which the format is created. The more international and the more general the prevalence, the more established the format can be considered to be, which is a strong indicator in relation to preservation suitability. The criterion is weighted 2, as general prevalence is an important indicator for how established a format is and opens up technical support for many years to come.

Prevalence II

The format is frequently used for preservation at cultural heritage institutions. It is a strength for preservation formats to be already in use at other cultural heritage preservation institutions. This occurs because like-minded actors in unison face the same challenges, thus creating opportunities for sparring and resource pooling as well as reducing the risk of technical obsolescence. The criterion is weighted 2, as the prevalence of a format for preservation purposes is a strong basis for cooperation on common issues.

Lifespan

The format has been around for at least ten years. An established format is more robust than a new format that has not been around for a long time, where it can be difficult to assess the direction in which the prevalence will go. The criterion is weighted 1, as lifespan is, to a larger extent, an indication rather than something tangible.

Lifespan II

The format has good future prospects. The criterion is difficult to quantify, but “good future prospects” are based on an estimate on whether the prevalence and software support will increase over the coming years rather than decline. The criterion tells something more about the development of the format rather than just age. The criterion is weighted 3, since whether a format is growing rather than dying is crucial to assessing the preservation suitability of the format.

Documentation

The format is standardised. It is a strength for preservation formats to be technically well-documented. The documentation makes it possible to analyse the functionality of a format and to develop system tools for characterisation, validation, and migration. The criterion is weighted 2, as it is an insurmountable task in terms of resources to disassemble a format for the purpose of developing the system tools for receiving and storing data.

Documentation II

The format standard has not been updated within the last ten years. The more stable a format, the easier it is to integrate it into the preservation planning since the prioritisation of resources for monitoring and to analyse the format decreases. The criterion is weighted 1, as it is only an indication of stability, and the challenge can be solved by adding the necessary resources to monitor and analyse the format more thoroughly.

Documentation III

The documentation of the format is well-described and can be read easily. It is a strength for preservation formats to have well-described and clear documentation written in an understandable language. At the same time, the documentation must not be too long to read. Here we set the limit at 1,000 pages. The criterion is weighted 1, as we do not necessarily have to understand the documentation in all its details.

Licensing

The format is open source. It is a strength for preservation formats to be independent of companies, including the strategic and financial interests of companies. In practice, the strength comes from the fact that working with the formats is cost-effective, the future holds fewer risks, and the possibility that the format will become more prevalent over time is greater. The criterion is weighted 2, as the free use of the format is a sign of health and an important factor behind the creation of an archival academic community around the format.

Structure

The format is self-supporting (not a container format). The criterion means that a format consists solely of its own format. Some formats allow multiple formats to be stored in them, making the format a container for other formats. Examples of container formats are video formats that store images in one format, audio in another format, and any subtitle tracks in a third format. Zip is also a container format that can have all sorts of formats stored in it. The criterion is weighted 1, as container formats only add the implication (but also complexity) that we must approve all the formats of the container as preservation formats.

Structure II

The format can be read as plain text. It is a strength for preservation formats if both machines and people can interpret the data directly. Plain text means any basic text program can open the file, interpret the character set, and render the binary data as readable text. Formats that cannot be read as plain text will not produce readable text after the interpretation of the character set. The criterion is weighted 1, as binary formats are not problematic as long as the technical documentation is saved.

Significant properties

The format supports most of the significant properties within its content type. The investigation of the significant properties of the format is a prerequisite for grading. Here it is ideal if a preservation format has all the properties of the original format so that significant properties are not lost during migration. The criterion is weighted 3, as significant properties are an essential indicator of whether data is preserved as it is created.

Dissemination

The format can be reused without conversion. If it is necessary to convert data from the preservation format in order to make data more usable, it will add an extra work process and risk of loss of authenticity, which would be undesirable. The criterion is weighted 1, as data conversion is recognised to be problematic, but nonetheless, conversion is common practice in the migration strategy, and system-independent formats are often more suitable for long-term preservation (see compatibility below) than proprietary formats that are easy to use today thus causing the need for conversion to more usable formats on dissemination.

Searchability

The format has searchable information. If the format allows for searches within the content, it is easier to index the content and make it available. The example here is a TIFF of a digital document, which exclusively stores an image of the document, which must subsequently be OCR-treated to have searchable text in a separate text or database file. A PDF version of the same digital document will have the text stored as embedded searchable information. The criterion is weighted 1, as the technical possibilities of processing offered by the format have an impact on the ability of an archive to preserve and make data available in a cost-effective and authentic way. However, there are often tools that can mechanically remedy the lack of searchability.

Interoperability

The format is suitable for data exchange. Some formats are more suitable for data exchange than others. This applies, for instance, to mark-up formats such as XML or formats that are widely used for exporting and importing data between IT systems. The criterion is weighted 2, as interoperability makes it easier to work with data in automated processes, and broad system support for data exchange is an indication of robustness.

Testing

The format has tools for identification and characterisation. The verification of format identity makes it possible to ensure that data are transferred to the archive according to format specification and supports subsequent preservation actions such as migration. The purpose is to gain knowledge about whether a file has the format it claims to have through its file extension, as well as gain knowledge about the metadata of the file. It is optimal if a testing tool is already mature for use in a fully developed and open version, where the developers still offer bug fixes and updates for new operating systems. An open version means that the tool can be freely adapted to suit your own business needs. The criterion is weighted 2, as identification and characterisation are important components of the testing process that, in a very basic way, allow for more complex preservation tasks.

Testing II

The format has tools for validation. The validation of the structure and content of files makes it possible to ensure that data is transferred to the archive according to the format specification. Here it is optimal if the tool is already mature for use in a fully developed and open version, where the developers still offer bug fixes and updates for new operating systems. An open version means that the tool can be freely adapted to suit your own business needs. The criterion is weighted 2, since validation is an important component of the testing process that allows for the performance of more complex preservation tasks.

Compression

The format is uncompressed or has lossless compression. Compression can cause the quality of data to be reduced so that they do not have the same accuracy as an original format may have had. However, compression may be allowed if the compression is lossless. The criterion is weighted 3, as lossless compression is a crucial factor in ensuring authenticity.

Storage

The format occupies less storage space than the average for its content type. Storage is the capacity in the physical media that the preservation formats occupy. Storage is a relative factor that is affected by technological development, which over time offers greater capacity at the same price, but the criterion's significance is also affected by the budgetary framework of the preservation institution. The criterion is weighted 1 because storage space does not have the same economic significance as in the past.

Migration

The format can be migrated with an acceptable loss of significant properties to another format. Migration here is defined as converting data from one original format to another format within the same content type. Next, the availability of data migration tools is a prerequisite for pursuing a migration strategy, and the better the tools the market can offer, the stronger the format's preservation suitability. In an ideal scenario, migration should result in the retention of all significant properties, the same intellectual content, and the same visual representation. The criterion is weighted 2, as the ability to migrate data to other formats indicates robustness.

Compatibility

The format is compatible with multiple operating systems and applications. In practice, this means that the format is both system and software independent without locking it to specific operating systems or programs, and several competing programs must exist in the market for rendering the data of the format. The criterion is weighted 1, as system independence have an impact on the possibility of pursuing a migration strategy.

GeoTIFF	
CRITERIA	Description
Prevalence	GeoTIFF is probably the most commonly used raster format with embedded georeferencing metadata. The reason for broad acceptance of GeoTIFF is that it provides more convenient and more complete georeferencing information than TIFF+TFW+PRJ combination. GeoTIFF is more convenient because location information and CRS information is provided within tags in the header of TIFF file and additional files TFW and PRJ are not needed.
Prevalence II	TIFF image file format without geo metadata tags is frequently used for preservation of digital images. It is on a list of preferable digital formats in many preservation institutions like Library of Congress and The National Archives in UK.
Lifespan	GeoTIFF image format has been in usage for more than ten years. The first version of GeoTIFF was published in 1995. The latest revisited version of GeoTIFF (1.0) was published in 2000.
Lifespan II	GeoTIFF image format has good prospects in the preservation of georeferenced digital images due to extensive software support. The new standard of GeoTIFF (OGC GeoTIFF 1.1) from 2019 formalises existing GeoTIFF specification version 1.0 and update it to added data in EPSG Geodetic Parameter values.
Documentation	GeoTIFF image format version 1.0 is well standardised and well documented as it is OGC GeoTIFF 1.1 version.
Documentation II	The latest version of GeoTIFF image format (v 1.1) is from year 2019 and is less than 10 years old but it is backward compatible with revision 1.0 from 1995.
Documentation III	GeoTIFF version 1.1 image format OGC documentation is substantial and detailed as well as it is latest documentation for version 1.0 from 2000.
Licensing	Portion of GeoTIFF 1.0 are copyrighted by authors Mike Ruth and Niles Ritter. Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct or commercial advantage, and this copyright notice appears. There are no concerns about patents concerning GeoTIFF use of TIFF tags. No concerns about patents for underlying TIFF 6.0
Structure	GeoTIFF is a self-supporting image format in the same way as TIFF image format is.
Structure II	GeoTIFF image file format is not a plain text format nor its data in the header. The context of GeoTIFF image file cannot be interpreted directly.
Significant properties	<p>GeoTIFF raster format supports storing imagery and is regularly used to store grid data.</p> <p>GeoTIFF image format supports the same significant properties as the aforementioned TIFF+TFW+PRJ combination. The difference is that all the location information is stored within TIFF tags.</p> <p>GeoTIFF is essentially a TIFF 6.0 compliant image (raster) file with additional tags where georeferencing information is stored. In all other properties, the GeoTIFF image format is same as TIFF. The difference between TIFF and GeoTIFF format is in GeoTIFF capability to store additional georeferencing information within 6 GeoKey extended tags. Additional GeoKeys provides most significant information regarding georeferencing. Despite providing extended information about projection, datums and ellipsoid GeoTIFF GeoKeys are not capable of storing information about datum transformation parameters that are needed in custom defined local coordinate systems. As TIFF image format GeoTIFF also doesn't provide raster spatial coverage.</p> <p>A limited number of contextual and structural properties could be described inside the TIFF metadata (tag) structure. So other Context and Structural properties need to be added with external metadata or documentation.</p>
Dissemination	GeoTIFF image format can be directly reused as it is without conversion.
Searchability	GeoTIFF image file format is binary image format, and its image content is not searchable. GeoTIFF image file format is capable of storing image metadata information into Baseline fields like image resolution, dimension, creation date, bits per sample, compression, orientation, author, copyright information,... With software that is capable of reading TIFF Baseline and Extended field information is possible to search TIFF image metadata. Additionally, to those TIFF 6.0 based metadata, GeoTIFF is capable of storing georeferencing information into so-called GeoKeys. GeoKeys could be read with any application that support GeoTIFF image format.
Interoperability	GeoTIFF image format provides a high level of interoperability. GeoTIFF image format is supported by a plethora of operating systems and programs. GeoTIFF images could be transferred between different operating and file systems without conversion.
Testing	<p>For identification and characterisation of GeoTIFF images several programs that are capable of identifying and characterised TIFF image exist. Most capable of them are:</p> <ul style="list-style-type: none"> - DROID - JHOVE validation tool with TIFF module

	<ul style="list-style-type: none"> - DPF Manager - Checkit_tiff - ImageMagick - ExifTool
Testing II	<p>For validation of TIFF image files several programs that are capable verify TIFF image exist. Most capable of them are:</p> <ul style="list-style-type: none"> - JHOVE validation tool with TIFF module and - DPF Manager - Checkit_tiff
Compression	<p>Compression of GeoTIFF image is an option, and GeoTIFF image format support several kinds of compression (5):</p> <ul style="list-style-type: none"> - No compression - Lossless compression (CCITT FAX3, FAX4, LZW, RLE, DEFLATE) - Lossy compression (JPEG)
Storage	<p>Non-compressed GeoTIFF image file format is probably the most used image format in image preservation, and so it is occupying average amount of storage space while GeoTIFF image with JPEG compression file size is smaller.</p>
Migration	<p>GeoTIFF image format is widely used and standardised image format supported by a large number of migration software. Migration of non-compressed GeoTIFF image should be simple with potential small loss of image metadata. With validation after migration, the loss of image metadata could be minimised.</p>
Compatibility	<p>GeoTIFF format is compatible with many operating systems and file systems. In all widely used operating systems (MS Windows, Linux, Apple macOS, BSD) are programs and tools that support reading, writing and visualisation of GeoTIFF image format.</p>
Overall Assessment [0,91]	<p>The overall score of this format is 31 points out of 34, which gives it a weight of 0,91% of all points. Out of all formats used in the geospatial domain, GeoTIFF is very suitable for long term preservation. It supports almost all raster geodata use cases. In comparison to the TIFF+TFW+PRJ format, it provides a more compact solution. For preservation purposes, the metadata tags within the GeoTIFF should be specified.</p>

Figure 8 - Example of descriptions of scores in column "GeoTIFF" in figure 7

3.4 Rationales in Documentation requirements

Rationales in 3.4 Documentation folders requirements

CITS Geospatial text:

Geospatial records are rarely in a form that is sufficiently self-explanatory to be used and interpreted adequately by itself. Consequently, additional information describing context, structure, rendering and behaviour is required to enable the user to understand, interpret and reuse preserved geodata properly. This chapter describes the requirements for Documentation for geospatial datasets (where it is applicable). Ideally, a standardised machine-readable format is preferred. However, any other form of documenting the System is welcome. Standardised machine-readable formats should be placed within the representation. Other documentation should be placed within the package level Documentation folder.

GEO_24 Rationale

Requirement:

GEO_24	Package level documentation	Documentation covering all representations in the Information package SHOULD be placed in documentation/ on package level	0..n SHOULD
--------	-----------------------------	--	-----------------------

Description:

This requirement states where the package-level documentation is to be placed.

Example:

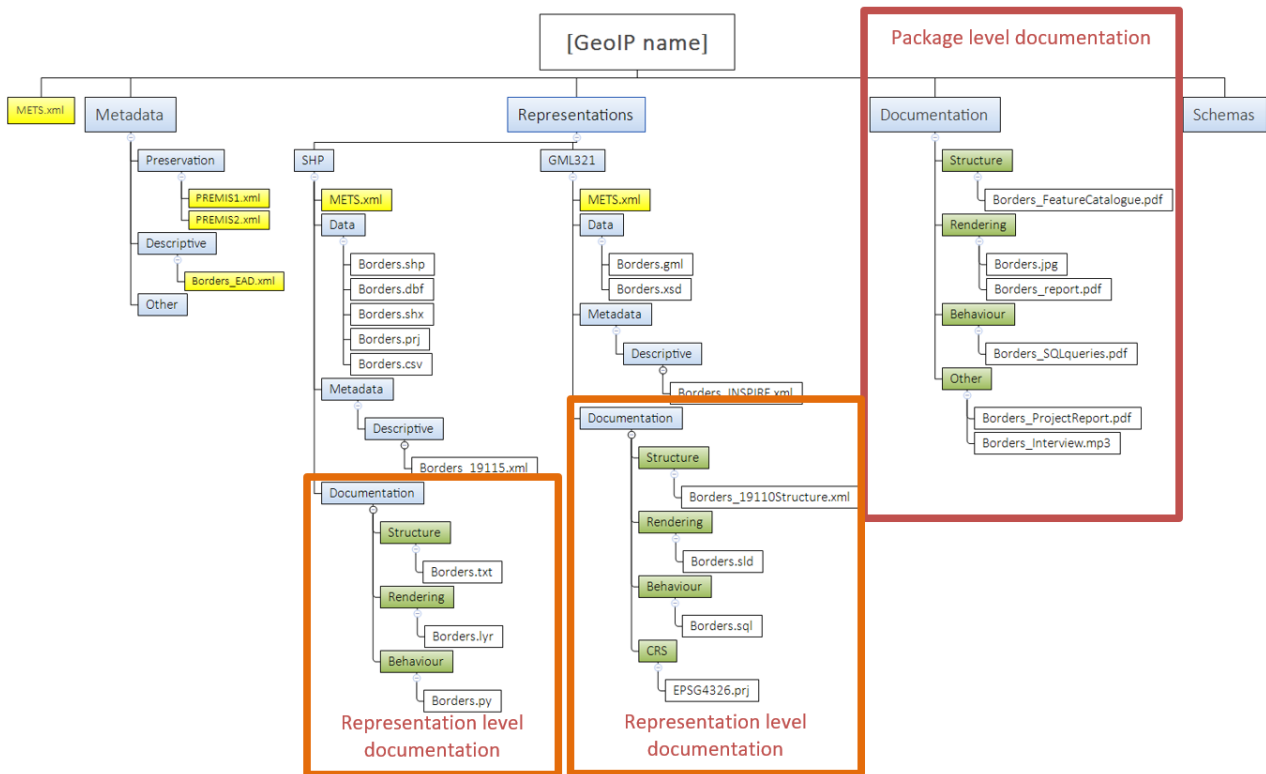


Figure 9 - Locations of Documentation folders in archival package.

Rationale:

Package-level documentation that covers information about all representations is placed on a higher level.

GEO_25 RationaleRequirement:

GEO_25	Representation level documentation	Technical documentation specific to one representation SHOULD be placed in <code>representations/[RepresentationName]/documentation</code>	0..n SHOULD
--------	------------------------------------	---	------------------------------

Description:

This requirement states where the representation-level documentation is to be placed.

Example:

See image in the GEO_24.

Rationale:

Sometimes documentation is specific to the content within the Representation. So, when we need to be more specific, we place the documentation to the Representation level. It is also intended for the standardised machine-readable content to support automated access and record dissemination.

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_26 Rationale

Requirement:

GEO_26	Feature Catalogue documentation	A document containing definitions and descriptions of feature types and feature attribute values SHOULD be provided for all geospatial records in the Information Package	0..n SHOULD
--------	---------------------------------	---	--------------------

Description:

This general requirement states that some kind of feature catalogue must accompany geospatial data in the Information Package. A Geospatial Feature catalogue contains definitions and descriptions of the feature types, feature attributes and feature relationships (feature inheritances and feature associations) occurring in one or more sets of geographic data, together with any feature operations that can be applied. [SOURCE: ISO 19101-1:2014, 4.1.13]. ISO 19110:2016 describes Feature Catalogues in greater detail.¹²

Feature Catalogues can also come in many forms:

- Standardised machine-readable xml file, based on ISO or OGC standards
- A proprietary well documented machine-readable feature catalogue
- A descriptive document that lists and explains all the definitions, descriptions and relations between features.

A Standardised machine-readable xml file is preferred, as it would support future access automation and be stored on the representation level. However, if that form is not available, a descriptive document needs to be available in the package-level documentation.

Example:

See example in the requirement GEO_28

Rationale:

A feature catalogue describes the structure of geospatial datasets and their relationships with other data. This information is essential for future reuse of the data.

¹²ISO 19110:2016 Geographic information — Methodology for feature cataloguing (<https://www.iso.org/standard/57303.html>)

GEO_27 Rationale

Requirement:

GEO_27 ISO 19110 ISO 19115-3	Standardised machine-readable Feature Catalogue	A standardised machine-readable feature catalogue SHOULD be provided in the Information Package	0..n SHOULD
GEO_27a Ref. GEO_27	Placement of Standardised machine-readable Feature Catalogue	If a standardised machine-readable feature catalogue exists it SHOULD be placed in representations/[RepresentationName]/documentation/structure	0..n SHOULD

Description:

This requirement provides an option to fulfil the requirement GEO_26 providing a standardised machine-readable feature catalogue in a subfolder named structure in the `documentation` folder at the representation level of the Information Package. Ideally, it should at least be available in the long-term preservation representation. Additional, **proprietary** machine-readable feature catalogue files should be placed in additional representations containing non-long-term preservation formats.

Taxonomies, Ontologies and/or vocabularies, that contain information relevant to Feature catalogues and that were used with geospatial data in standardized formats (OWL, SKOS, RDF...) can also be added to this folder.

Example:

An example would be an XML file based on ISO 19110:2016 or earlier.

Rationale:

A standardised machine-readable feature catalogue is commonly well-documented and gives us the possibility of future dissemination automation.

GEO_28 Rationale

Requirement:

GEO_28 Ref. GEO_27	Documentation containing Feature Catalogue Descriptions	Documentation, describing elements of a feature catalogue, not compliant with GEO_27 (a non-standardised machine-readable feature catalogue) SHOULD be provided in one of the Documentation folders of the Information Package	0..n SHOULD
--------------------------	---	---	--------------------

Description:

This requirement provides an option to fulfil the requirement GEO_26. When a standardised machine-readable feature catalogue is not available or as an addition to it. A Feature Catalogue documentation should contain a collection of metadata that provides the semantics and the structure of the objects stored in geospatial record(s). A feature catalogue should contain the names and definitions of feature types, names and definitions of their properties which include feature attributes, geometry (shapes and specifications, datum, map projection, etc.), temporal aspect (dimensions and specifications, datum, units, resolutions, etc.), operations, and roles, descriptions of attribute values and domains, relationships, constraints,...

It is recommended that the document describing all representations is placed in a subfolder named structure in the documentation folder at the package level or representation level of the Information Package, depending on the context. Other placements are also possible, depending on the context (to better reflect the producers organisational structure or if a document contains documentation on multiple technical elements, etc.)

Example:

Below an example of a non-standardised machine-readable feature catalogue.

Spatial Object Type: Duct

Duct	
Title:	duct
Definition:	A utility link or link sequence used to protect and guide cable and pipes via an encasing construction.
Description:	<p>A Duct (or Conduit, or Duct-bank, or Wireway) is a linear object which belongs to the structural network. It is the outermost casing. A Duct may contain Pipe(s), Cable(s) or other Duct(s).</p> <p>Duct is a concrete feature class that contains information about the position and characteristics of ducts as seen from a manhole, vault, or a cross-section of a trench and duct.</p>
Subtype of:	UtilityLinkSet
Type:	Spatial Object Type
Attribute:	
Name:	ductWidth
Title:	duct width
Definition:	The width of the duct.
Description:	The measurement of the object – in this case, the duct - from side to side.
Voidable:	true
Multiplicity:	1
Value type:	Length
Association role:	
Name:	ducts
Definition:	A single duct or set of ducts that constitute the inner-duct.
Voidable:	true
Multiplicity:	0..*
Value type:	Duct (spatial object type)
Association role:	
Name:	cables
Definition:	A duct may contain one or more cables.
Voidable:	true
Multiplicity:	0..*
Value type:	Cable (spatial object type)
Association role:	
Name:	pipes
Definition:	The set of pipes that constitute the duct bank.
Voidable:	true

Multiplicity:	0..*
Value type:	Pipe (spatial object type)

More examples can be found on Feature Catalogue “INSPIRE Application Schemas.”¹³

Rationale:

Written documentation containing the feature catalogue is a less favourable option since automated access is not supported. However, it would still document geospatial records well enough for manual interpretation.

GEO_29 Rationale

Requirement:

GEO_29	Logical model	A document describing relationships between multiple geospatial entities or geospatial and non-spatial records SHOULD be provided in the Information Package	0..n SHOULD
GEO_29a Ref GEO_29	Placement of logical model	If a document describing the logical model exists, it SHOULD be provided in a documentation/structure folder	0..1 SHOULD
GEO_29b Ref GEO_29	Placement of machine-readable logical model	If a standardised machine-readable version of a document describing the logical model exists, it SHOULD be provided in representations/[RepresentationName]/documentation/structure	0..1 SHOULD

Description:

This requirement is applicable in cases when the Information Package contains multiple geospatial datasets with defined relationships. A logical model is a visual representation of the Feature Catalogue and can be presented as a UML Diagram or in a standardised machine-readable format that can be imported into a database.

Ontologies, Taxonomies and/or vocabularies, that contain information describing the logical model and were used with geospatial in standardized formats (OWL, SKOS, RDF..) can also be added to this folder

¹³ Feature Catalogue “INSPIRE Application Schemas” (<https://inspire.ec.europa.eu/data-model/approved/r4618-ir/fc/>)

Example:

Below is an example of a logical model describing the relationship between multiple entities in a geospatial record_

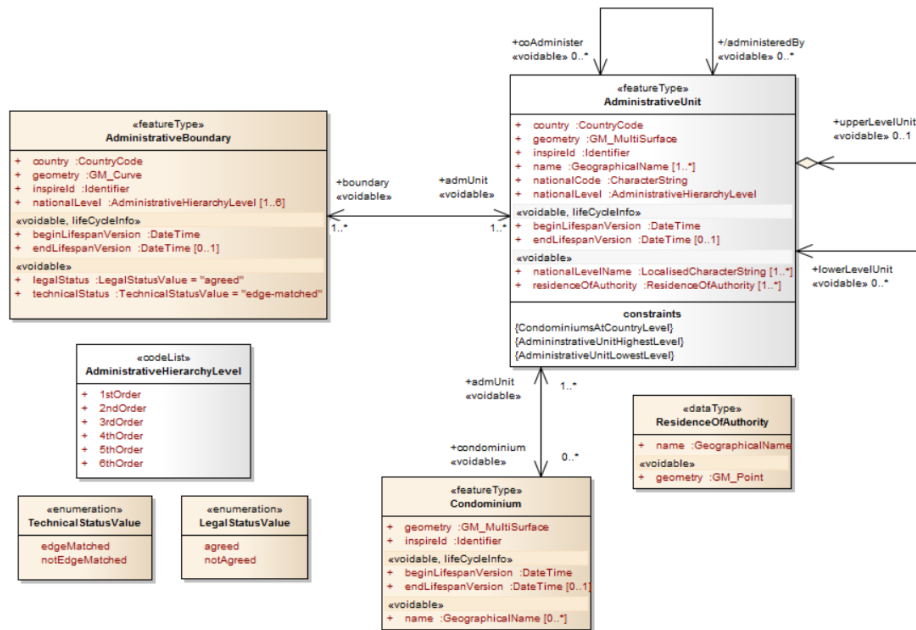


Figure 10 - example of a logical model describing the relationship between multiple entities in a geospatial record

More examples can be found in the INSPIRE Knowledge Base > Data Specifications >Data Models¹⁴.

An example of a standardised machine-readable format for logical models is the XML Metadata Interchange (XMI)¹⁵. This format is defined by the OMG and based on XML and is an open standard file format that enables the interchange of model information between models and tools. Other possible formats are formatted CSV, RDF and others.

Rationale:

This structure of one or many geospatial records in a dataset should be described in a machine-readable way to enable automated validation and use of the files.

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.

¹⁴ INSPIRE Knowledge Base > Data Specifications >Data Models (<https://inspire.ec.europa.eu/data-model/approved/r4618-ir/fc/>)

¹⁵ <https://www.omg.org/spec/XMI/2.5.1/About-XMI/>

Example:

The OGC OWS Context standard can be used for storing information on the configuration and rendering styles of used data and its references. See more on this in the Guideline for GIS.

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.

Rendering and visualisation documentation provides a way to replicate the rendering of geospatial records in future GIS. The catalogue of cartographic symbols is a collection of agreed cartographic symbols, which are used via the visualisation of geospatial datasets to display objects in space. As shown in Figure 11, Cartographic symbols are shown in the legend, which explains their meaning.

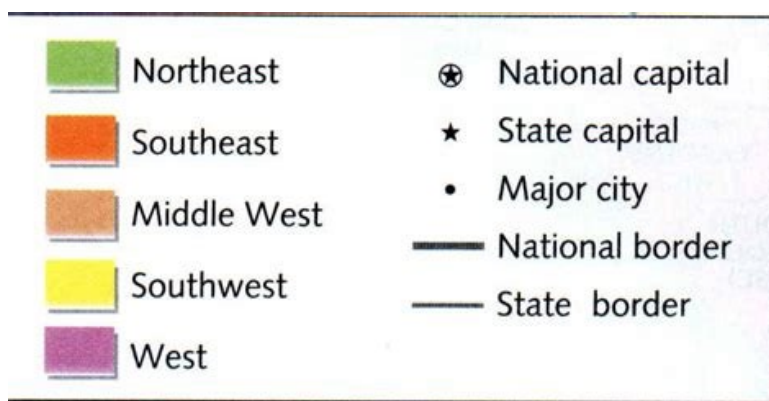


Figure 11 - Legend with cartographic symbols

For specific geospatial data, the visualisation is already embedded into the product by the producer in the form of (geo-located) raster images or scanned paper maps. In these cases, it is reasonable to archive that kind of visualisation. For each geospatial dataset, 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 standardised machine-readable files, they should be preserved.

GEO_31 Rationale

Requirement:

GEO_31	Geospatial dataset visualisation	An image displaying the overall view or a representative section of any geospatial dataset SHOULD be provided in the Information Package and placed in a documentation/rendering folder	0..n SHOULD
--------	----------------------------------	--	-----------------------

Description:

This requirement states that there should be an image representing a visual overview of a geospatial dataset. The image can depict the full extent of the dataset or a representative section, whichever is more informative. Images are to be placed into the `documentation/rendering` folder on the Information Package level. If there are differences between geospatial records within various representations, the images can also be placed within the `representations/[RepresentationName]/documentation/rendering` folder.

If the same record is split into multiple datasets that hold the same type of content, one image is enough. For example, if we have elevation iso lines for different locations, we can use one image to represent them all, even if the data is stored in multiple datasets. If producers use finding aids for their geospatial records, they usually already have the images available.

Example:

The image in Figure 12 shows an example of the usage of images representing geospatial records within a Geospatial Metadata catalogue.

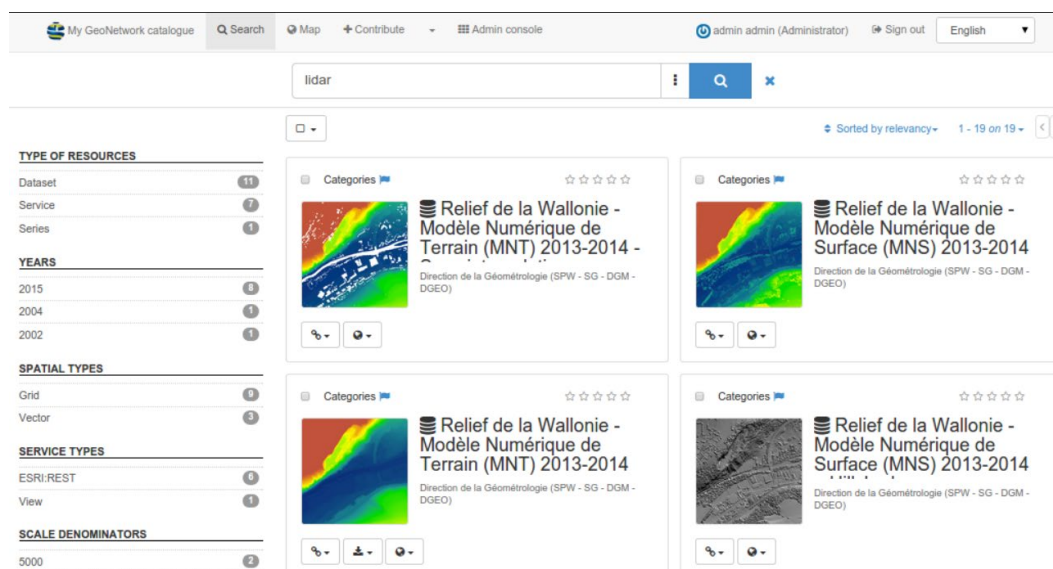


Figure 12 - Usage of images in Geospatial Metadata Catalogue (www.geonetwork-opensource.org)

Rationale:

The purpose of this requirement is to provide an image, that helps us find the geospatial dataset quicker and make it more accessible in the finding aid. The benefit is that images can be used in archival finding aids like archival catalogues or Geospatial Metadata Catalogues without rendering in GIS systems. This gives future users of the preserved geospatial data an easy way to quickly identify the content of the Information Package.

GEO_32 Rationale

Requirement:

GEO_32	Visualisation documentation	A document describing visualisation rules and configurations SHOULD be provided in the Information Package	0..n SHOULD
GEO_32a Ref GEO_32	Placement of visualisation documentation	If a document describing visualisation rules and configurations exists, it SHOULD be provided in a documentation/rendering folder	0..1 SHOULD

Description:

This requirement applies to datasets where geospatial data is rendered using cartographic means depending on the data format.

Standardised machine-readable files should be preferably placed within the `documentation/rendering` folder on the representation level (in the representation containing Long-term preservation formats). This rendering information is often dependent on the system (Desktop tool, Web Application, Mobile device, etc.) used to render the geospatial data. An example of rendering configuration information in QGIS is shown in Figure 13. In QGIS, for example, such rendering information can be stored in a machine-readable *.qml file. All other rendering and visualisation documentation is to be in the `documentation/rendering` folder on the package level of the Information Package package level. If there are differences between geospatial records within various representations, the images or configuration files can also be placed within the `representations/[RepresentationName]/documentation/rendering` folder.

If there is no standardised machine-readable files, then human-readable documentation should be available that contains information about visualisation rules and configurations of geospatial data forming an Information product.

Some of the common visualisation and rendering information used in most GIS tools are listed below.

- **Legend** of a map (e.g. symbology for roads, railways, vegetation etc.)
- **Layer symbology definition** (type of symbol and colouring, classification rules, visibility scale, labelling, definition query)
- **Scale** of map (e.g. 1:50.000)
- **Image value representation type**. How are raster pixel values represented (RGB, greyscale, RGB-NIR, CMYK, NDVI, custom colour map)?
- **Colour map**. How is the colour map documented? How are raster pixel values represented (continuously, classified)? Are any additional classification tables present for the colour map?
- **Rendering algorithm**. Is an algorithm used to render the values (Histogram stretch, Gamma stretch, Statistics based display, Unique values)
- **Raster Pyramids**. Are raster pyramids calculated to display the raster object, and if so, with what parameters.

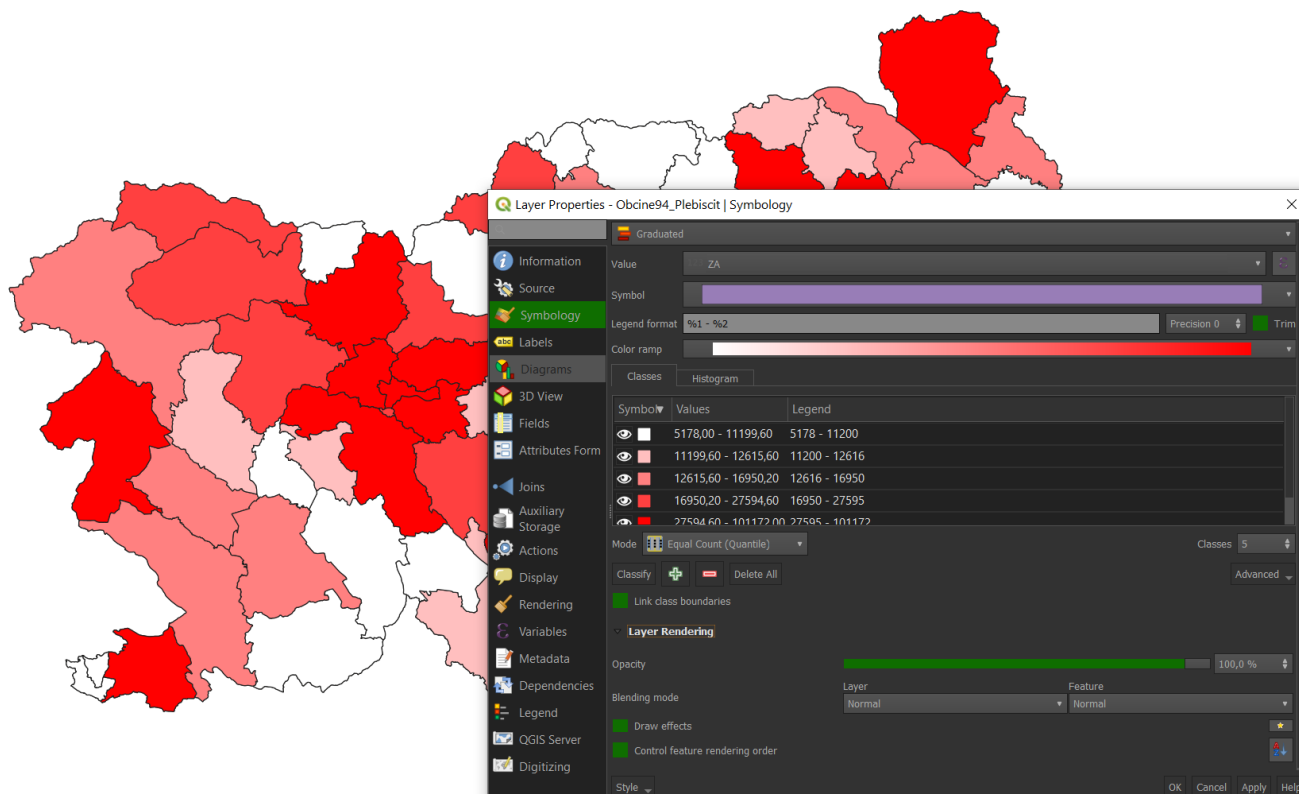


Figure 13 - Vector layer visualization and rendering configuration in QGIS

Examples:

Rendering and Visualisation information can be documented using **standardised machine-readable files**, human-readable documentation defining the cartographic keys and rendering criteria or by examples of information products (maps, lists, screenshots or videos from the original system).

Rationale:

To properly understand, interpret and reproduce information products based on geospatial data, we need to know what methodology was used to render the data. If the rendering tool's functionality is fully documented, an exact reproduction of the user experience can be recreated. However, since this is often not the case, Archivists then decide which elements need to be preserved.

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 an open well-documented machine-readable format to support the automatisisation of dissemination. According to the archiving policies of an organisation, adequate formats need to be selected. For instance, if an organisation is preserving its data for a medium length, like 10-20 years, it can take a more liberal approach. However national archives usually have the strictest requirements. If the producer cannot provide the archive with an open well-documented symbology configuration, it can be recreated from the description provided in the *Documentation* in an open-source GIS application like QGIS¹⁶ or even using a LLM like ChatGPT v.4¹⁷.

Example:

An example of Standardised machine-readable formats for the rendering of geospatial records are SLD¹⁸ or KML¹⁹ files that also have some of that capability. It must be noted, however, that the SLD standard is rather limited. Therefore, many cartographic styling details cannot be adequately represented in SLD, as opposed to, for example, QML²⁰ (from QGIS) and LYR²¹ (ESRI layer file) are much more powerful.

SLD files example: SLD (Styled Layer Description) is an XML-based OGC (Open Geospatial Consortium) standard for symbology of web services. 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 input for easier DIP creation in the future. An example of an SLD file is shown below.

```
<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>
```

Rationale:

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

¹⁶ QGIS (<https://qgis.org/>)

¹⁷ OpenAI ChatGPT- <https://chat.openai.com/>

¹⁸ SLD – Styled Layer Description (https://portal.opengeospatial.org/files/?artifact_id=22364)

¹⁹ KML Standard (<http://www.opengeospatial.org/>)

²⁰ QGIS QML Style file (https://docs.qgis.org/3.34/en/docs/user_manual/appendices/qgis_file_formats.html#qml-the-qgis-style-file-format)

²¹ esri LYR style file (<https://www.loc.gov/preservation/digital/formats/fdd/fdd000626.shtml>)

GEO_34 Rationale

Requirement:

GEO_34	Information product examples	Information product examples based on geospatial record or records example SHOULD be provided in the Information Package and placed in the package level in the Information Package	0..n SHOULD
GEO_34a Ref GEO_34	Placement of information product examples	If information product examples exist, they SHOULD be provided in the Information Package in a documentation/rendering folder	0..1 SHOULD

Description:

This requirement recommends that examples of information products originally derived from the preserved geospatial records in the Information Package are provided in a documentation folder in the Information Package.

Example:

Geographic Information Systems support the creation of different types of Information Products:

- A pop-up window within the application that provides an information
- A digital map or a series of maps that can be printed on physical media or to a digital file (PDF, eps, various image formats)
- A set of lists or reports that are a result of a spatial query that have no or minor graphical components. (i.e. House numbers connected to a water line in a particular area.)
- A flythrough video of a 3D landscape
- A new geospatial record based on use of geospatial algorithms (A new area based on a Buffer function, the terrain elevation model could be a derivate from a LIDAR point cloud, etc.). However, we would not recommend to store these types of information products in the documentation, but rather as additional data within the Data folder or even as a separate Information Package.

Rationale:

To enable reproduction of information products derived from preserved geospatial records in the Information Package, which are similar to the information products data creators used to produce in their original systems.

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	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 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, 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 SHOULD 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_37 Ref GEO_37a	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, 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 any type of source code (scripts, SQL queries, GeoSPARQL queries, Python code, C++ code, etc.) that were used to produce an information product using 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 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. EPSG²² 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.

Example:

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

²² EPSG (<https://epsg.org/home.html>)

```

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]]

```

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 MUST 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 14 - Standardised Machine-readable transformation between MGI 1901 CRS and Slovenia 1996 CRS in WKT2 format

GeoRepository
Login / Register

[Home](#)
[EPSG Home](#)
[EPSG Dataset](#)
[Support Documentation](#)
[About Us](#)
[Contact](#)
[IOGP Geomatics](#)

Coordinate Transformation: MGI 1901 to Slovenia 1996 (1)
WKT GML Print



EPSG
GEODETTIC PARAMETER DATASET
Managed by IOGP's Geomatics Committee



EPSG Dataset : v10.018

Text Search i

Please [login](#) or [register](#) to include deprecated (invalid) items, search remarks and export results.

point Slovenia

Clear all search

Transformation Details [SUPERSEDED]

NAME:	MGI 1901 to Slovenia 1996 (1)
CODE:	3916
VERSION:	GuRS-Svn
VARIANT:	1
ACCURACY (M):	1
USAGE:	

Usage Details

SCOPE:	Transformation of coordinates at 1m level of accuracy.
EXTENT:	Slovenia - onshore

SOURCE CRS:	MGI 1901
TARGET CRS:	Slovenia 1996
TRANSFORMATION METHOD:	Coordinate Frame rotation (geog2D domain)
METHOD IS REVERSIBLE:	YES

TRANSFORMATION PARAMETERS:	Parameter	Value	Reversible	Unit
	X-axis translation	409.545	YES	metre
	Y-axis translation	72.164	YES	metre
	Z-axis translation	486.872	YES	metre
	X-axis rotation	-3.085957	YES	arc-second
	Y-axis rotation	-5.46911	YES	arc-second
	Z-axis rotation	11.020289	YES	arc-second
	Scale difference	17.919665	YES	parts per million

META DATA

REMARKS:	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).
INFORMATION SOURCE:	Geodetska uprava Republike Slovenije, http://www.gu.gov.si
DATA SOURCE:	EPSG
REVISION DATE:	29 August 2018

Figure 15 - A human document describing CRS transformation from MGI 1901 to Slovenia 1996 CRS

Rationale:

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

Rationales in 3.4.5 Other - Contextual Documentation requirements

CITS Geospatial text:

This part of the IP describes all remaining, more general information about the geospatial record. Included here are links to relevant Documentation describing data creation methodology and the spatial data set's provenance. The Documentation could consist of interviews, legal origin documentation, related practices in the EU and worldwide, methodological rules, scientific articles, related publications, etc.

Description:

Besides the documentation described in previous chapters (from 3.4.1 to 3.4.4.), which mainly address elements describing the structure, rendering and behaviour, we also require contextual documentation. Contextual documentation is often in a non-machine-readable format like pdf or TIFF and recently also in audio and video formats.

Examples:

Detailed descriptions of lineage information

- *Acquisition information* describing methods and tools used for creating the geospatial records, like camera used, flight path, scanner used, digitisation accuracy, GPS...
- *Processing information* describing algorithms used and processing performed to produce the data and descriptions of environment procedures such as software and parameters by which the algorithm is applied to generate the data from the source data
- *Source data* information describing the original source data a submitted geospatial dataset (product) is derived from such as data format, CRS and/or storage
- Spatial accuracy of the geospatial records
- Temporal accuracy of the geospatial records

Detailed descriptions of the use of the system used to produce, manage and/or view the geospatial records

- User manuals and other documentation and screenshots describing the system-user dialogue in the original system used for production and/or use of the geospatial records
- Videos or screen captures of the system as seen from the user's point of view
- Interviews with producers and/or users of the geospatial records
- Documentation of finding tool and screenshots of the metadata search and presentation related to the geospatial records

Detailed descriptions of the structure in the information Package of the geospatial records (if this is not documented otherwise using naming, grouping of files in folders or METS.xml structMap elements)

- Relation in the IP between a data file and the metadata file describing
- Relation in the IP between a raster object and a world file
- Relation in the IP between a schema file and the gml file that it validates
- Relation in the IP between raster objects and tiling indexes
- Relation in the IP between a geospatial coverage file and the geospatial objects it covers
- Relation in the IP between geospatial records and additional attributes in another file like a feature catalogue or database file

Detailed descriptions of the broader historical context of the geospatial records

- Provenance
- Description of the purpose of geospatial dataset
- Description of administrative use of the geospatial dataset
- Methodology used

- Regulations and legal context
- Scientific articles about the geospatial records
- Publications

Detailed descriptions of preservation actions

- Documentation of migration of geospatial records to preservation format/Information package
- Migration or dissemination information can be documented in a PREMIS file and placed in a metadata preservation folder.

Rationale:

Contextual documentation of the geospatial data in the Information package is important to be able to understand and use the archived data correctly.

GEO_40 Rationale

Requirement:

GEO_40	Package level contextual documentation	Contextual documentation covering all representations in the Information package SHOULD be placed in documentation/other on package level	0..n SHOULD
--------	--	--	--------------------

Description:

It is recommended that contextual documentation relevant for data in several representations in the Information Package is placed at the package level. However, if the documentation only covers two out of three representations in the IP, a copy of the documentation in each of the two representations at the representation level is a better option.

Example:

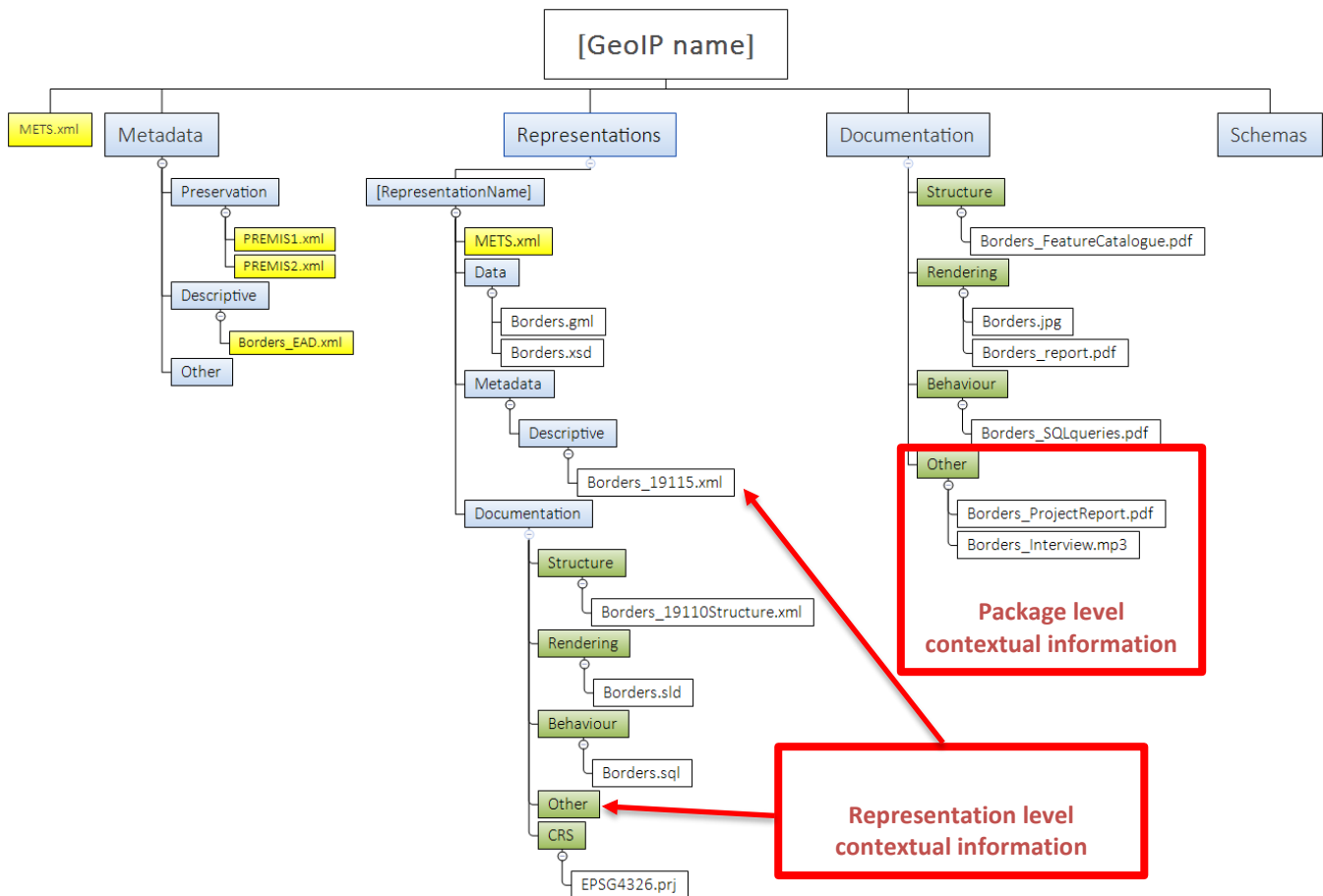


Figure 16 - Location of Package and Representation level Contextual documentation

Rationale:

The purpose of correct placement of contextual documentation in the IP is to enable easy identification of relevant documentation when using data from a representation.

GEO_41 Rationale

Requirement:

GEO_41	Representation level contextual documentation	Contextual documentation covering only content within a particular representation SHOULD be placed in representations/[RepresentationName]/documentation/other	0..n SHOULD
--------	---	---	--------------------

Description:

It is recommended that contextual documentation relevant for only one representation in the Information Package is placed at the representation level in that particular representation. Context can also be described in standardised machine-readable Geospatial metadata that can be placed within the Metadata/descriptive folder on the representation level. Geospatial metadata is described in greater detail in chapter 3.5.

Example: See Example in GEO_40.

Rationale: The purpose of correct placement of contextual documentation in the IP is to enable easy identification of relevant documentation when using data from a representation.

3.5 Rationales in metadata requirements

Rationales in 3.5 Geospatial Metadata requirements

CITS Geospatial text:

Geospatial data in the IP is documented using a form of geospatial metadata, which contains common descriptions of the data as well as descriptions specific to the geospatial domain (accuracy, lineage, scale, measurement units, CRS info, etc.). In original systems, geospatial metadata can be stored in different ways (databases, standardised xml files, common documentation, etc.).

Description:

Digital Geospatial data can be described in different ways. A standardised and structured approach for digital geospatial records is most commonly documented using geospatial metadata, which contains common descriptions of the data as well as descriptions specific to the geospatial domain (accuracy, methodology of creation, reference to source data, scale, measurement units, CRS info, etc.). In original systems, geospatial metadata can be stored in different ways (databases, standardised xml files, txt based readme files, or even only within typical contextual documentation, etc.).

For CITS Geospatial we recommend, that the representation containing the long-term preservation formats holds geospatial metadata in a standardised machine-readable format. The most common standards used for geospatial metadata are:

- **ISO 19115:2003** Geographic information — Metadata
- **INSPIRE Metadata** Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119
- And the latest **ISO 19115-1:2014** Geographic information — Metadata — Part 1: Fundamentals
- **Data Catalogue Vocabulary (DCAT)** and its geospatial version GeoDCAT²³ metadata

Metadata can also be extended using additional ISO Standards like these:

- ISO 19165-1:2018 Geographic information -Preservation of digital data and metadata -Part 1: Fundamentals
- ISO 19115-2:2019 Geographic information — Metadata — Part 2: Extensions for acquisition and processing
- ISO 19157:2013 Geographic information — Data quality

Every metadata standard prescribes the mandatory metadata elements, which are in some cases limited only to essential Dublin Core elements. However, for long-term preservation purposes, we strongly suggest using the set of mandatory elements as described within the INSPIRE Metadata Implementing rules.

Example:

Table 1 below covers proposed mandatory elements required in the INSPIRE Metadata Implementation rules (see INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119²⁴). The INSPIRE metadata rules are also based on ISO 15836 (Dublin Core). Information that cannot be covered by these elements should be described in the accompanying documentation.

Some of the technical guidance recommendations and requirements from the INSPIRE Metadata Implementing Rules are added as descriptions in table 1 below. Obligation states whether the element is mandatory (M), optional (O) or conditional (C). The table also compares the obligation of an element in INSPIRE with the obligation of the element in

²³W3C DCAT – Geospatial dataset property - https://www.w3.org/TR/vocab-dcat-2/#Property:dataset_spatial

²⁴ INSPIRE Metadata Implementing Rules (<https://inspire.ec.europa.eu/documents/inspire-metadata-implementing-rules-technical-guidelines-based-en-iso-19115-and-en-iso-1>)

ISO 19115-1. Each obligation of an INSPIRE element is compared to the obligation of the similar ISO 19115 element in column 2 in table 1.

See the standards INSPIRE Implementing Rules for Metadata and EN ISO 19115-1 for further guidance on how to create an xml file with geospatial metadata that conforms to these rules. The INSPIRE Implementing Rules for Metadata also has examples of XML encoding.

Appendix 3 provide mapping schemas between INSPIRE metadata elements and the archival descriptive metadata standards EAD3 and ISAD(G).

Table 1 - Proposed mandatory metadata elements for an INSPIRE metadata file

No	ISO 19115 Core Element name	INSPIRE Element name	Description of the content of the element	Obligation/ occurrence
1	Dataset title (M)	Resource title (M)	<p><i>Name by which the cited resource is known</i></p> <p>The <i>Resource title</i> is the name of the data set and has to be concise and to the point. It should not contain unexplained acronyms or abbreviations. It is recommended a maximum length of 250 characters and keeping the similarity with the original title of the resource, in the sense of the ‘official naming’.</p> <p>Properties in the xml-file (ISO/TS 19139 path): <code>identificationInfo[1]/*/citation/*/*title</code></p> <p>Elements in the ISO 19115-1 (2014) standard: <code>MD_Metadata.identificationInfo > MD_DataIdentification.citation > CI_Citation.title)</code></p>	M/1
2	Abstract describing the dataset (M)	Resource abstract (M)	<p><i>Brief narrative summary of the content of the resource(s)</i></p> <p>The <i>Resource abstract</i> is a brief narrative summary of the content of the data set. The abstract provides a clear and concise statement that enables the user to understand the content of the data. Do not use unexplained acronyms or abbreviations.</p> <p>Properties in the xml file (ISO/TS 19139 path): <code>identificationInfo[1]/*/abstract</code></p>	M/1

			<p>Elements in the ISO 19115-1 (2014) standard: <code>MD_Metadata.identificationInfo > MD_DataIdentification.abstract</code></p>	
3	- (O)	Resource type	<p><i>Scope to which metadata applies.</i></p> <p>The <i>Resource type</i> is a code identifying the type of resource, e.g., dataset (see MD_ScopeCode), which the metadata describes. 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.</p> <p>Properties in the xml file (ISO/TS 19139 path): <code>hierarchyLevel > Data type:MD_ScopeCode</code></p> <p>The values of MD_ScopeCode in the scope of the INSPIRE Directive are:</p> <ul style="list-style-type: none"> - dataset for spatial datasets; - series for spatial dataset series; - service for spatial data services <p>Elements in the ISO 19115-1 (2014) standard: <code>MD_Metadata/metadataScope/ MD_MetadataScope/resource- Scope</code></p>	M/1
4	Resource identifier (O)	Unique resource identifier (M)	<p><i>Value uniquely identifying an object within a namespace</i></p> <p>Properties in the xml file (ISO/TS 19139 path): <code>identificationInfo[1]/*/citation/ */identifier</code></p> <p>Elements in the ISO 19115-1 (2014) standard: <code>MD_Metadata.identificationInfo>MD _DataIdentification.citation > CI_Citation.identifier>MD_Identifier</code></p>	M/1..n
5	Dataset language (M)	Resource language (C)	<i>Language(s) used within the datasets</i>	M/1..n

			<p>The <i>Resource language</i> is a three-letter code for the language taken from the vocabulary LanguageCode (ISO/TS 19139) based on alpha-3 codes of ISO 639-2. Use only three-letter codes from in ISO 639-2/B (bibliographic codes).</p> <p>The list of all the codes is defined at http://www.loc.gov/standards/iso639-2/ Regional languages also are included in this list.</p> <p>Properties in the xml file (ISO/TS 19139 path): <code>identificationInfo[1]/*/language</code></p> <p>Elements in the ISO 19115-1 (2014) standard: <code>MD_Metadata.identificationInfo> MD_DataIdentification.defaultLocale > PT_Locale</code></p>	
6	Dataset topic category (M)	Topic category (M)	<p><i>Main theme(s) of the dataset.</i></p> <p>The <i>Topic category</i> describes the topic of the dataset and is a selection of the 20 elements in the MD_TopicCategory enumeration MD_TopicCategoryCode . The list below is from B.3.30 MD_TopicCategoryCode << Enumeration>> in the ISO 19115-1 (2014) standard, and the listed examples are not exhaustive:</p> <ul style="list-style-type: none"> ● farming ● biota ● boundaries ● climatologyMeteorologyAtmosphere ● economy ● elevation ● environment ● geoscientificInformation ● health ● imageryBaseMapsEarthCover ● intelligenceMilitary ● inlandWaters ● location ● oceans ● planningCadastre ● society ● structure ● transportation 	M/1..n

			<ul style="list-style-type: none"> ● utilitiesCommunication ● extraTerrestrial ● disaster <p>This topic information is a high-level classification scheme to assist in the grouping and topic-based search of available geospatial data resources.</p> <p>Correct categorisation is very important to help users to search and find the resources they are looking for.</p> <p>Properties in the xml file (ISO/TS 19139 path): <code>identificationInfo[1]/*/topicCategory</code></p> <p>Elements in the ISO 19115-1 (2014) standard: <code>MD_Metadata.identificationInfo > MD_DataIdentification.topicCategory > MD_TopicCategoryCode</code></p>	
7	Keywords (O)	Keyword > Keyword value (M)	<p><i>Commonly used word(s) or formalised word(s) or phrase(s) used to describe the subject</i></p> <p>The INSPIRE Metadata Regulation 1205/2008/EC mandate the presence of at least one keyword to describe the dataset.</p> <p>An INSPIRE Keyword is defined by:</p> <ul style="list-style-type: none"> ● a keyword value (see 2.4.1), which in ISO standard is referred to as “Keyword”; ● an optional originating controlled vocabulary (see 2.4.2), which in ISO standard is referred to as “Thesaurus”. ● It is possible to add as many keywords as relevant to the resource. <p>If only one keyword is used, then for spatial dataset or spatial dataset series, the keyword:</p> <ul style="list-style-type: none"> ● shall describe the relevant INSPIRE Spatial Data Theme (as defined in Annex I, II and III of the INSPIRE Directive) ● shall be expressed in the language of the metadata for the 34 INSPIRE Spatial Data Themes (please use the terms in each of the official languages in which the INSPIRE Directive has been translated) or neutral language values such as a URI. 	M/1..n

			<p>The keyword value is a commonly used word, formalised word or phrase used to describe the subject. While the topic category is too coarse for detailed queries, keywords help to narrow a full-text search, and they allow for structured keyword search</p> <p>Properties in the xml file (ISO/TS 19139 path): <code>identificationInfo[1]/*/descriptiveKeywords/*/keyword</code></p> <p>Elements in the ISO 19115-1 (2014) standard: <code>MD_Metadata.identificationInfo</code> <code>>MD_DataIdentification></code> <code>descriptiveKeywords</code> <code>>MD_Keywords</code></p>	
8	Keywords (O)	Keyword > Originating controlled vocabulary (C)	<p><i>Name of the formally registered thesaurus or a similar authoritative source of keywords</i></p> <p>Mandatory if the keyword value originates from a controlled vocabulary (Conditional).</p> <p>If the keyword value originates from a controlled vocabulary (thesaurus, ontology), for example, GEMET - Concepts, the citation of the originating controlled vocabulary shall be provided.</p> <p>The thesaurusName identification shall include at least the title and a reference date (date of publication, date of last revision or of creation) of the originating controlled vocabulary.</p> <p>Properties in the xml file (ISO/TS 19139 path): <code>identificationInfo[1]/*/descriptiveKeywords/*/thesaurusName</code></p> <p>Elements in the ISO 19115-1 (2014) standard:</p> <p>The controlled vocabulary is described through the thesaurusName property of the instance of descriptiveKeywords to which the keyword pertains:</p> <p>Example descriptiveKeywords: (MD_Keywords) keyword: BOUNDARIES Administrative</p>	C/0..n

			<p>keyword: INDUSTRY Mining Exploration keyword: MINERALS Exploration thesaurusName: (CI_Citation) title: ANZLIC Search Words date: (CI_Date) date: 1996-07 dateType: (CI_DateTypeCode) publication</p>	
9	Geographic location of the dataset (C)	Geographic bounding box (M)	<p><i>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)</i> <i>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)</i></p> <p>This is the extent of the resource (e.g. dataset) in the geographic space, given as a bounding box. Defining the coordinates of a rectangle representing the resource area on a map allows the discovery by geographical area.</p> <p>The bounding box shall be as small as possible.</p> <p>The bounding box shall be expressed in decimal degrees with a precision of at least two decimals. The coordinates of the bounding box are expressed in any geodetic coordinate reference system with a Greenwich Prime Meridian.</p> <p>Properties in the xml file (ISO/TS 19139 path): identificationInfo[1]/*/extent/*/geographicElement/*/westBoundLongitude</p> <p>identificationInfo[1]/*/extent/*/geographicElement/*/eastBoundLongitude</p> <p>identificationInfo[1]/*/extent/*/geographicElement/*/southBoundLatitude</p>	M/1..n

			<p>identificationInfo[1]/*/extent/*/geographicElement/*/northBoundLatitude</p> <p>Elements in the ISO 19115-1 (2014) standard: EX_GeographicBoundingBox westBoundLongitude eastBoundLongitude southBoundLatitude northBoundLatitude</p>	
10	Additional extent information for the dataset (vertical and temporal) (O)	Temporal extent (M)	<p><i>The time period covered by the content of the dataset</i></p> <p>The temporal extent defines the time period covered by the content of the resource. This time period may be expressed as:</p> <ul style="list-style-type: none"> ● an individual date ● an interval of dates (starting date and ending date) ● a mix of individual dates and intervals of dates <p>The INSPIRE Metadata Regulation 1205/2008/EC requires at least one temporal reference chosen from one of these four categories:</p> <ul style="list-style-type: none"> ● temporal extent ● date of publication ● date of last revision ● date of creation <p>To be compliant with ISO 19115 it is necessary to use at least one among the date of publication, date of last revision, or the date of creation.</p> <p>The default reference system shall be the Gregorian calendar, with dates expressed in accordance with ISO 8601 (yyyy-mm-dd where yyyy is the year, mm is the month and dd is the day). Example: From 2008-01-01T11:45:30 to 2008-12-31T09:10:00.</p> <p>Properties in the xml file (ISO/TS 19139 path): identificationInfo[1]/*/extent/*/temporalElement/*/extent</p> <p>Elements in the ISO 19115-1 (2014) standard:</p>	M/1..n

			<pre>MD_Metadata.identificationInfo > MD_Identification.extent > EX_ Extent > EX_TemporalExtent or EX_VerticalExtent</pre>	
11	<p>Additional extent information for the dataset (vertical and temporal) (O)</p>	<p>Date of publication</p>	<p><i>The reference date for the cited resource - publication</i></p> <p>This is the date of publication of the resource when available or the date of entry into force. There may be more than one date of publication. The date of publication differs from the temporal extent. For example, a dataset might have been published in March 2009 (2009-03-15), but the covered information was collected over the year 2008 (temporal extent from 2008-01-01 to 2008-12-31).</p> <p>The default reference system shall be the Gregorian calendar, with dates expressed in accordance with ISO 8601 (yyyy-mm-dd where yyyy is the year, mm is the month and dd is the day). Example: 2009-03-15T11:45:30 or 2009-03-15.</p> <p>Properties in the xml file (ISO/TS 19139 path): <code>identificationInfo[1]/*/citation/*/date[./*/dateType/*/text()='publication']/*/date</code></p> <p>Elements in the ISO 19115-1 (2014) standard: ... <code><gmd:identificationInfo></code> <code><gmd:MD_DataIdentification></code> <code><gmd:citation></code> <code><gmd:CI_Citation></code> ... <code><gmd:date></code> <code><gmd:CI_Date></code> <code><gmd:date></code> <code><gco>Date>2009-03-15</gco>Date></code> <code></gmd:date></code> <code><gmd:dateType></code> <code><gmd:CI_DateTypeCode</code> <code>codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codelist/ML_gmxCodelists.xml#CI_DateTypeCod</code></p>	O/0..n

			<pre>e" codeListValue="publication">publication</g md:CI_DateTypeCode> </gmd:dateType> </gmd:CI_Date> </gmd:date> ... </gmd:CI_Citation> </gmd:citation> ... </gmd:MD_DataIdentification> ... </gmd:identificationInfo></pre>	
<p>12</p>	<p>Additional extent information for the dataset (vertical and temporal) (O)</p>	<p>Date of last revision</p>	<p><i>The reference date for the cited resource - revision</i></p> <p>This date describes when the resource was last revised, if the resource has been revised. The date of revision differs from the temporal extent. For example, a dataset might have been revised in April 2009 (2009-04-15), but the covered information was collected over the year 2008 (temporal extent from 2008-01-01 to 2008-12-31). Example: 2009-04-15 2009-04-15T11:15:00.</p> <p>Properties in the xml file (ISO/TS 19139 path): identificationInfo[1]/*/citation/*/date[./*/dateType/*/text()='revision']/*/date</p> <p>Elements in the ISO 19115-1 (2014) standard:</p> <pre>... <gmd:identificationInfo> <gmd:MD_DataIdentification> <gmd:citation> <gmd:CI_Citation> ... <gmd:date> <gmd:CI_Date> <gmd:date> <gco:DateTime>2009-04-15T11:15:00</gco:DateTime> </gmd:date> <gmd:dateType> <gmd:CI_DateTypeCode</pre>	<p>O/0..n</p>

			<pre>codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codelist/ML_gmxCodelists.xml#CI_DateTypeCode" codeListValue="revision">revision</gmd:CI_DateTypeCode> </gmd:dateType> </gmd:CI_Date> </gmd:date> ... </gmd:CI_Citation> </gmd:citation> ... </gmd:MD_DataIdentification> </gmd:identificationInfo></pre>	
13	Additional extent information for the dataset (vertical and temporal) (O)	Date of creation	<p><i>The reference date for the cited resource - creation</i></p> <p>This date describes when the resource was created. The date of creation differs from the temporal extent. For example, a dataset might have been created in February 2009 (2009-02-15), but the covered information was collected over the year 2008 (temporal extent from 2008-01-01 to 2008-12-31). Example: 2009-02-15 2009-02-15T11:15:00.</p> <p>Properties in the xml file (ISO/TS 19139 path): <code>identificationInfo[1]/*citation/*/date[./*/dateType/*/text()='creation']/*date</code></p> <p>Elements in the ISO 19115-1 (2014) standard:</p> <pre>... <gmd:identificationInfo> <gmd:MD_DataIdentification> <gmd:citation> <gmd:CI_Citation> ... <gmd:date> <gmd:CI_Date> <gmd:date> <gco>Date>2009-02-15</gco>Date> </gmd:date> <gmd:dateType> <gmd:CI_DateTypeCode codeList="http://standards.iso.org/ittf/PubliclyAv</pre>	O/0..n

			<pre> ailableStandards/ISO_19139_Schemas/resources/ codelist/ML_gmxCodelists.xml#CI_DateTypeCod e" codeListValue="creation">creation</gmd:CI_ DateTypeCode> </gmd:dateType> </gmd:CI_Date> </gmd:date> ... </gmd:CI_Citation> </gmd:citation> ... </gmd:MD_DataIdentification> ... </gmd:identificationInfo> </pre>	
14	Lineage (O)	Lineage (M)	<p><i>General explanation of the data producer’s knowledge about the lineage of a dataset.</i></p> <p>According to the INSPIRE Implementing Rules for Metadata, Lineage is “a statement on process history and/or overall quality of the spatial data set. Where appropriate, it may include a statement whether the data set has been validated or quality assured, whether it is the official version (if multiple versions exist), and whether it has legal validity. The value domain of this element is free text.”</p> <p>The process history may be described by information on the source data used and the main transformation steps that took place in creating the current data set (series).</p> <p>The use of acronyms should be avoided. If used, their meaning should be explained.</p> <p>This information can also reference other documents that cover this description in greater detail.</p> <p>Properties in the xml file (ISO/TS 19139 path): dataQualityInfo/*/lineage/*/state ment</p> <p>Elements in the ISO 19115-1 (2014) standard:</p>	M/1

			<pre>MD_Metadata >resourceLineage> LI_Lineage</pre>	
15	Spatial resolution of the dataset (O)	Spatial resolution (C)	<ul style="list-style-type: none"> • <i>Equivalent scale: level of detail expressed as the scale denominator of a comparable hardcopy map or chart</i> • <i>Distance: ground sample distance</i> <p>Spatial resolution refers to the level of detail of the data set. It shall be expressed as a set of zero to many resolution distances (typically for gridded data and imagery-derived products) or equivalent scales (typically for maps or map-derived products).</p> <p>An <i>equivalent scale</i> is generally expressed as a positive integer value expressing the scale denominator. Example: 50000 (e.g. 1:50000 scale map).</p> <p>A resolution <i>distance</i> shall be expressed as a numerical value associated with a unit of length. Example: 0.25 (degrees).</p> <p>Each spatial resolution is either an equivalent scale OR a ground sample distance. Each spatialResolution element must contain either an equivalent scale or a distance but not both.</p> <p>Properties in the xml file (ISO/TS 19139 path): <code>identificationInfo[1]/*/spatialResolution/*/equivalentScale/*/denominator</code> (equivalent scale)</p> <p><code>identificationInfo[1]/*/spatialResolution/*/distance</code> (distance)</p> <p>Elements in the ISO 19115-1 (2014) standard: <pre>MD_Metadata.identificationInfo > MD_Identification.spatialResolution > MD_Resolution.equivalentScale MD_Resolution.distance,</pre> </p>	C/0..n Mandatory if an equivalent scale or a resolution distance can be specified.

			<p>MD_Resolution.vertical, or MD_Resolution.angularDistance, or MD_Resolution.levelOfDetail</p>	
-		<p>Conformity > Degree (M)</p>	<p><i>Indication of the conformance result</i></p> <p>This is the degree of conformity of the resource to the implementing rules adopted under Article 7(1) of INSPIRE Directive 2007/2/EC or other specification.</p> <p>The values are:</p> <ul style="list-style-type: none"> ● true (if conformant) ● false (if not conformant) ● null (with nilReason = “unknown” if not evaluated) <p>Properties in the xml file (ISO/TS 19139 path): dataQualityInfo/*/report/*/result */pass</p> <p>Elements in the ISO 19115-1 (2014) standard: I cannot find the elements in ISO 19115-1:2014</p> <p>Elements in the ISO 19115-1 (2003) standard</p> <pre><gmd:result> <gmd:DQ_ConformanceResult> <gmd:specification> ... <!-- See below --> </gmd:specification> <!-- gmd:explanation is mandated by ISO 19115. A default value is proposed --> <gmd:explanation> <gco:CharacterString>See the referenced specification</gco:CharacterStrin g> </gmd:explanation> <!-- the value is false instead of true if not conformant --> <gmd:pass> <gco:Boolean>>true</gco:Boolean> </gmd:pass> </gmd:DQ_ConformanceResult> </gmd:result></pre>	M/1

<p>16</p>	<p>-</p>	<p>Conformity > Specification (M)</p>	<p><i>Citation of the product specification or user requirement against which data is being evaluated</i></p> <p>This is a citation of the implementing rules adopted under Article 7(1) of INSPIRE Directive 2007/2/EC or other specification to which a particular resource conforms.</p> <p>The following properties are expected:</p> <ul style="list-style-type: none"> ● Title (characterString and free text) ● Reference date (CI_Date): <ul style="list-style-type: none"> ○ dataType: creation, publication or revision date, e.g. publication ○ date: an effective date, e.g. 2010-12-08 <p>Properties in the xml file (ISO/TS 19139 path): dataQualityInfo/*/report/*/result */specification</p> <p>Elements in the ISO 19115-1 (2014) standard: I cannot find the elements in ISO 19115-1:20114</p> <p>Elements in the ISO 19115-1 (2003) standard</p> <pre> <gmd:dataQualityInfo> <gmd:DQ_DataQuality> ... <gmd:report> <gmd:DQ_DomainConsistency> <gmd:result> <gmd:DQ_ConformanceResult> <gmd:specification> <gmd:CI_Citation> <gmd:title> <gco:CharacterString>COMMISSION REGULATION (EU) No 1089/2010 of 23 November 2010 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards interoperability of spatial data sets and services</gco:CharacterString> </gmd:title> <gmd:date> <gmd:CI_Date> <gmd:date> <gco>Date>2010-12-08</gco>Date> </gmd:date> </gmd:CI_Date> </gmd:result> </gmd:DQ_ConformanceResult> </gmd:result> </gmd:DQ_DomainConsistency> </gmd:report> </gmd:DQ_DataQuality> </gmd:dataQualityInfo> </pre>	<p>M/1</p>
-----------	----------	---	---	------------

			<pre> </gmd:date> <gmd:dateType> <gmd:CI_DateTypeCode codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codelist/ML_gmxCodeLists.xml#CI_DateTypeCode" codeListValue="publication">publication</gmd:CI_DateTypeCode> </gmd:dateType> </gmd:CI_Date> </gmd:date> </gmd:CI_Citation> </gmd:specification> ... </gmd:DQ_DataQuality> </gmd:dataQualityInfo> </pre>	
<p>17</p>	<p>Constraints on resource access and use (O)</p>	<p>Limitations on public access (M)</p>	<p><i>Access constraints applied to assure the protection of privacy or intellectual property, and any special restrictions or limitations on obtaining the resource</i></p> <p>This metadata element shall provide information on the limitations and the reasons for them. If there are no limitations on public access, use the free text available in MD_LegalConstraints. otherConstraints to enter "No Limitations" in the language used for the metadata.</p> <p>Limitations on public access shall be represented by at least one of these metadata elements:</p> <ul style="list-style-type: none"> - MD_LegalConstraints. accessConstraints - MD_LegalConstraints. otherConstraints - MD_SecurityConstraints. classification <p>Access constraints: Access constraints applied to assure the protection of privacy or intellectual property and any special restrictions or limitations on obtaining the resource. Value is strictly limited to the value defined in codelist MD_RestrictionCode <<CodeList>> in ISO 19115-1:2014:</p> <ul style="list-style-type: none"> ● copyright ● patent ● patentPending 	<p>M/1..n</p>

			<ul style="list-style-type: none"> ● trademark ● licence ● intellectualPropertyRights ● restricted ● otherRestrictions ● unrestricted ● licenceUnrestricted ● licenceEndUser ● licenceDistributor ● private ● statutory ● confidential ● sensitiveButUnclassified ● in-confidence <p>Other constraints: Other restrictions and legal prerequisites for accessing and using the resource or metadata. Value is free text. Example: No limitations.</p> <p>Classification: Name of the handling restrictions on the resource. Value from codelist B.3.13 MD_ClassificationCode <<CodeList>> in ISO 19115-1:2014:</p> <ul style="list-style-type: none"> ● unclassified ● restricted ● confidential ● secret ● topSecret ● sensitiveButUnclassified ● forOfficialUseOnly ● protected ● limitedDistribution <p>Properties in the xml file (ISO/TS 19139 path):</p> <pre>identificationInfo[1]/*/resourceConstraints/*/accessConstraints</pre> <pre>identificationInfo[1]/*/resourceConstraints/*/otherConstraints</pre> <pre>identificationInfo[1]/*/resourceConstraints/*/classification</pre> <p>Elements in the ISO 19115-1 (2014) standard:</p> <pre>MD_Metadata.identificationInfo>MD_DataIdentification>MD_Constraints>MD_LegalConstraints and/or MD_SecurityConstraints</pre>	
--	--	--	--	--

<p>18</p>	<p>Constraints on resource access and use (O)</p>	<p>Conditions applying to access and use</p>	<p><i>Restrictions on the access and use of a resource or metadata</i></p> <p>The value is free text.</p> <p>If no conditions apply to the access and use of the resource, ‘no conditions apply’ shall be used. If conditions are unknown, ‘conditions unknown’ shall be used.</p> <p>Properties in the xml file (ISO/TS 19139 path): identificationInfo[1]/*/resourceConstraints/*/useLimitation</p> <p>Elements in the ISO 19115-1 (2014) standard: MD_Metadata.identificationInfo>MD_DataIdentification>MD_Constraints.useLimitations</p>	<p>M/1..n</p>
	<p>Resource point of contact (O)</p>	<p>Responsible organisation > Responsible party (M)</p>	<p><i>Identification of, and means of communication with, person(s) and organisation(s) associated with the resource(s)</i></p> <p>This is the description of the organisation responsible for the establishment, management, maintenance or distribution of the resource.</p> <p>This description shall include: name of the organisation and contact email address. The name of the organisation should be given in full, without abbreviations. It is recommended to use institutional email instead of personal emails.</p> <p>The following properties are expected:</p> <ul style="list-style-type: none"> ● organisationName (characterString and free text) ● contactInfo (CI_Contact): <ul style="list-style-type: none"> ○ address <ul style="list-style-type: none"> ■ electronicMailAddress [1..*] (characterString) <p>Properties in the xml file (ISO/TS 19139 path): identificationInfo[1]/*/pointOfContact</p> <p>Elements in the ISO 19115-1 (2014) standard: MD_Metadata.identificationInfo > MD_DataIdentification.pointOf-</p>	<p>M/1</p>

			Contact > CI_Responsibility	
	Resource point of contact (O)	Responsible organisation > Responsible party role (M)	<p><i>Function performed by the responsible party</i></p> <p>This is the role of the responsible organisation.</p> <p>Value from the codelist B.3.5 CI_RoleCode in ISO 19115-1:2014:</p> <ul style="list-style-type: none"> ● resourceProvider ● custodian ● owner ● user ● distributor ● originator ● pointOfContact ● principallInvestigator ● processor ● publisher ● author ● sponsor ● coAuthor ● collaborator ● editor ● mediator ● rightsHolder ● contributor ● funder ● stakeholder <p>There is a direct mapping between the responsible party roles defined in Part D 6 of the INSPIRE Metadata Regulation 1205/2008/EC and the values of the CI_RoleCode codelist of ISO 19115.</p> <p>Properties in the xml file (ISO/TS 19139 path): <code>identificationInfo[1]/*/pointOfContact/*/role</code></p> <p>Elements in the ISO 19115-1 (2014) standard: <code>MD_Metadata.identificationInfo> MD_DataIdentification.pointOfContact>CI_Responsibility>CI_RoleCode</code></p>	M/1

	<p>Metadata point of contact (M)</p>	<p>Metadata point of contact (M)</p>	<p><i>Party responsible for the metadata information</i></p> <p>This description shall include: name of the organisation and contact email address.</p> <p>The name of the organisation should be given in full, without abbreviations. It is recommended to use institutional email instead of personal emails.</p> <p>The following properties are expected:</p> <ul style="list-style-type: none"> ● organisationName (characterString and free text) ● contactInfo (CI_Contact): <ul style="list-style-type: none"> ○ address: <ul style="list-style-type: none"> ■ electronicMailAddress [1..*] (characterString) <p>The role of the responsible party serving as a metadata point of contact is out of scope of the INSPIRE Metadata Regulation 1205/2008/EC, but this property is mandated by ISO 19115. The default value is <code>pointOfContact</code>.</p> <p>Properties in the xml file (ISO/TS 19139 path): <code>MD_Metadata/*/contact</code></p> <p>Elements in the ISO 19115-1 (2014) standard: <code>MD_Metadata.contact>CI_Responsibility</code></p>	<p>M/1..n</p>
<p>19</p>	<p>Metadata date stamp (M)</p>	<p>Metadata date (M)</p>	<p><i>Date that the metadata was created</i></p> <p>The date which specifies when the metadata record was created.</p> <p>The default reference system shall be the Gregorian calendar, with dates expressed in accordance with ISO 8601 (yyyy-mm-dd where yyyy is the year, mm is the month and dd is the day). Example: 2012-02-20.</p> <p>Properties in the xml file (ISO/TS 19139 path): <code>dateStamp</code></p> <p>Elements in the ISO 19115-1 (2014) standard: <code>MD_Metadata.dateInfo</code></p>	<p>M/1</p>

20	Metadata language (O)	Metadata language (M)	<p><i>Language used for documenting metadata</i></p> <p>This is the language in which the metadata elements are expressed.</p> <p>The <i>Resource language</i> is a three-letter code for the language taken from the vocabulary LanguageCode (ISO/TS 19139) based on alpha-3 codes of ISO 639-2. Use only three-letter codes from in ISO 639-2/B (bibliographic codes).</p> <p>The list of all the codes is defined at http://www.loc.gov/standards/iso639-2/ Regional languages also are included in this list.</p> <p>Properties in the xml file (ISO/TS 19139 path): language</p> <p>Elements in the ISO 19115-1 (2014) standard: MD_Metadata/defaultLocale: PT_Locale</p>	M/1
----	------------------------------	------------------------------	---	-----

GEO_42 Rationale

Requirement:

GEO_42 Ref GEO_17	Standardised machine-readable geospatial metadata	Descriptive geospatial metadata in the long-term preservation format representation of the Information Package SHOULD be provided in the form of standardised machine-readable format compliant with geospatial metadata standards	0..n SHOULD
GEO_42a Ref GEO_42	Placement of standardised machine-readable geospatial metadata	If a standardised descriptive geospatial metadata file exists it MUST be provided in representations/[RepresentationName]/metadata/descriptive	Conditional 1..1 MUST
GEO_42b Ref GEO_42 and GEOSTR1	XML schema definition for geospatial metadata	If a standardised descriptive geospatial metadata file exists it MUST be accompanied by an XML schema definition placed in a sub-folder called /schemas within the Information Package root folder or the representation folder	Conditional 1..n MUST

Description:

A minimum of contextual information about each geospatial dataset in the Information Package must be provided in a machine-readable metadata format compliant with a geospatial metadata standard. This metadata file should be placed in a `metadata/descriptive` folder on the representation level in the Information Package.

Example:

See an example of an INSPIRE file in the example Information package. [This is not yet available.]

Rationale:

Some metadata is specific and essential to describing geospatial datasets and may not be present in an archival metadata file (EAD or Dublin Core) used to describe the content of the whole Information Package at the package level. This kind of metadata is usually called *discovery* or *descriptive* metadata and is often used in metadata catalogues and Finding Aids enabling automated search, discovery and identification of the geospatial records. An XML schema definition enables validation of a standardised geospatial metadata file (GEO_42b).

GEO_43 Rationale

Requirement:

GEO_43 Ref. GEO_17	Non-standardised machine-readable	A copy of Geospatial metadata in non-long-term preservation representations MAY be stored in its original form as databases or documentation. However, if this data is stored in a long-term preservation representation, the formats need to comply with the	0..n MAY
-----------------------	-----------------------------------	--	-------------

	Geospatial metadata	archival guidelines (stored in approved long-term preservation formats).	
--	---------------------	--	--

Description:

This requirement addresses proprietary or local/national metadata repositories or documents based on legacy formats or standards that are not aligned with typical worldwide xml-based formats for geospatial metadata.

Example:

An example would be:

- A local database that contains metadata descriptions of geospatial records
- Unstructured text-based documents or files describing the context of geospatial records
- Any other legacy documentation that contains metadata elements describing geospatial records.

Rationale:

Standardised machine-readable geospatial metadata is essential for supporting quick and clear automated identification of the object we need in a vast collection of preserved data. Standard-based geospatial metadata enables us to use existing Geospatial metadata catalogues and validators and enable harvesting of compatible metadata elements to archival catalogues.

4. Examples and tools

This section contains an overview of available example packages and tools as a means to “get your hands dirty” and take action in developing the field of geospatial data preservation.

○ Examples

Examples are often the best teacher. In this section, we will guide the reader to examples of valid CITS Geospatial packages. These can in general, be found at the GitHub site for this specification:

<https://github.com/DILCISBoard/CITS-Geospatial> .

Examples will be described in the table below [By publication of version 1.1 of this guideline the examples are under development]:

Link	Name and Description	Number of representations	Proprietary format
[Link to the example]	[The example name and a short description of the example]	[An integer describing the number of representations present in the example]	[A description of the proprietary formats used in the example]
[Link to the example]	[The example name and a short description of the example]	[An integer describing the number of representations present in the example]	[A description of the proprietary formats used in the example]

It is a plan that more examples will follow. If you have a good example, please let us know via the “Issues”-function in GitHub portal <https://github.com/DILCISBoard/CITS-Geospatial/issues> .

○ Tools

There are several open-source and commercial tools to support geospatial data preservation workflows, management and its reuse.

Name	Description	Link
QGIS	This is a fully functional open-source desktop GIS system. It supports transformations of coordinate systems, geospatial data formats. It includes GDAL and PROJ functionalities and can be used to reconstruct information products based on geospatial data. QGIS can be used to consume INSPIRE compliant view and download services. The QGIS plugin ‘GML	www.qgis.org/

	Application Schema Toolbox' makes use of the GMLAS driver for GDAL and aims to develop tools to manipulate complex features in a GIS desktop application.	
GDAL/OGR	GDAL is an open-source C++ translator library for more than 200 raster and vector geospatial data formats. The libraries can support the transformation from original to Long term preservation formats and back to many dissemination formats. It could also be used for the automation of preservation actions.	https://gdal.org/
PROJ	PROJ is a generic coordinate transformation software that transforms geospatial coordinates from one coordinate reference system (CRS) to another. This includes cartographic projections as well as geodetic transformations.	https://proj.org/
Geonetwork	GeoNetwork is a metadata catalogue application to manage spatially referenced resources. It provides powerful metadata editing and search functions as well as an interactive web map viewer. It is currently used in numerous Spatial Data Infrastructure initiatives across the world. It is an excellent tool to be used together with archival catalogues.	https://geonetwork-opensource.org/
FME	A commercial tool for migration of various geospatial and non-geospatial formats. It supports graphical workflow design and automation.	https://www.safe.com/
Hale>>Studio	An opensource tool for transforming data into INSPIRE compliant datasets. It contains presets for all INSPIRE schemas, Integration of the INSPIRE codelist registry, full support for all INSPIRE default and alternative encodings and more.	https://wetransform.to/halestudio/
OSGeo	The OsGeo-Live environment includes Open Source software that is needed to implement these different INSPIRE components. Furthermore, several software tools have been developed within EU-funded projects and are suitable for reuse.	https://wiki.osgeo.org/wiki/INSPIRE_tools_inventory
OGC API		

If you know of other good examples of freely available tools - please let us know via the "Issues"-function in GitHub portal [Add URL].

5. Postface

AUTHOR(S)	
Name(s)	Organisation(s)
Ann-Kristin Egeland	The Danish National Archives
Anita Graser	Austrian Institute of Technology
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
1.0.0	2021-06-11	Ann Kristin Egeland, Gregor Završnik	DNA Geoarh	Creation of version for public review
1.0.0	2021-08-31	Various	Various	Publication of version 1.0.0

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.

Appendix 1: Long-Term preservation format Profile for Geospatial Vector data using GML 3.2.1

See separate Appendix 1

Appendix 2: Long-Term preservation format Profile for Geospatial Raster data using TIFF baseline 6

See separate Appendix 2

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:

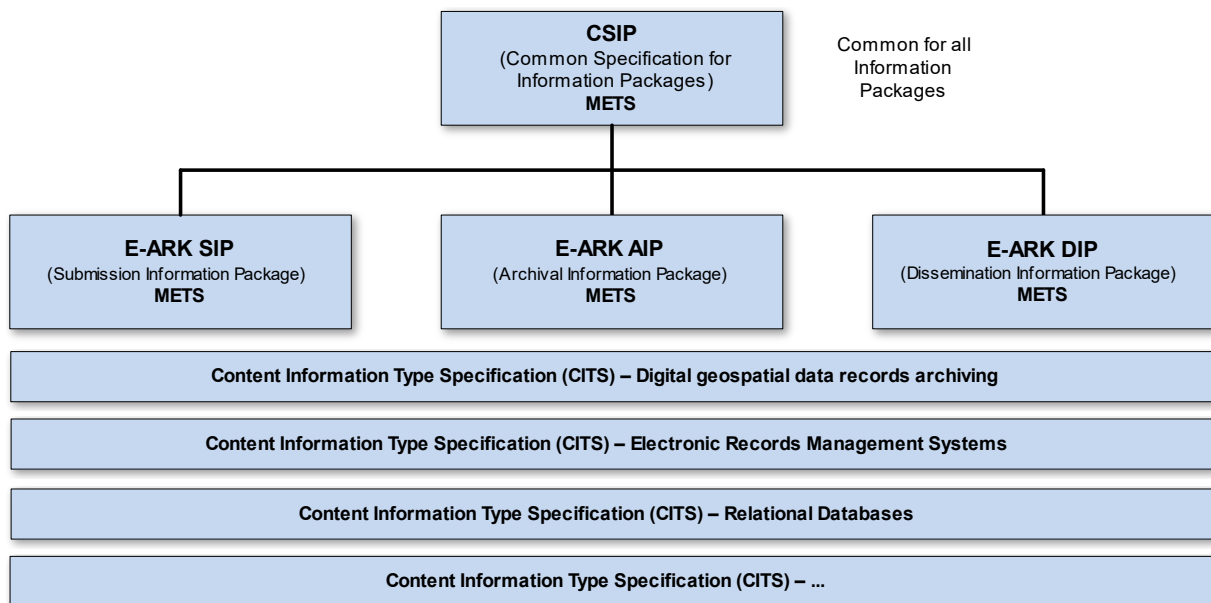


Figure I: Diagram showing E-ARK specification dependency hierarchy. Note that the image only shows a selection of the published CITS and isn't an exhaustive list.

Specification	Aim and Goals
Common Specification for Information Packages	<p>This document introduces the concept of a Common Specification for Information Packages (CSIP). Its three main purposes are to:</p> <ul style="list-style-type: none"> • 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.

Specification	Aim and Goals
	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	<p>The main aims of this specification are to:</p> <ul style="list-style-type: none"> • 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	<p>The main aims of this specification are to:</p> <ul style="list-style-type: none"> • 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	<p>The main aims of this specification are to:</p> <ul style="list-style-type: none"> • 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	<p>The main aim and goal of a Content Information Type Specification is to:</p> <ul style="list-style-type: none"> • 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. <p>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/).</p>

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://eak-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 CONTENT

1	Preface	3
1	Context	9
1.1	Purpose	9
1.2	Scope	9
1.3	Structure of the document	9
2	Introduction to Geospatial data and approach to its preservation	10
2.1	Geospatial geodata	10
2.1.1	Basic Terms	10
2.1.2	Geospatial vector data	10
2.1.3	Geospatial raster data	12
2.1.4	Standards for Geospatial Data	12
2.1.4.1	CITS Geospatial alignment with ISO 19165-1:2018	13
2.2	Significant properties of geospatial data	15
2.3	Recommended reading list	16
3	Rationale for requirements in CITS Geospatial	17
3.1	Folder structure requirements	18
	GEOSTR1 rationale	19
	GEOSTR2 rationale	19
	GEOSTR3 rationale	20
	GEOSTR4 rationale	20
	GEOSTR5 rationale	21
	GEOSTR6 rationale	21
	Folder structure examples explained	21
3.1.1	Geo IP containing one vector representation	21
3.1.1.1	Content	22
3.1.1.2	Context	22
3.1.1.3	Structure	22
3.1.1.4	Rendering	22
3.1.1.5	Behaviour	23
3.1.1.6	Package level elements	23
3.1.2	Geo IP containing multiple vector representations	23
3.1.3	Geo IP containing one representation of multiple raster datasets	25

3.2	Rationales in METS Requirements	26
	GEO_1 Rationale	26
	Rationales in 3.2.2 Package METS requirements	27
	GEO_2 Rationale	27
	GEO_3 Rationale	27
	GEO_4 Rationale	28
	GEO_5 Rationale	28
	GEO_6 Rationale	29
	GEO_7 Rationale	30
	Rationales in 3.2.3 Representation METS requirements	31
	GEO_8 Rationale	31
	GEO_9 Rationale	31
	GEO_10 Rationale	32
3.3	Rationales in data requirements	33
	Rationales in 3.3.1 Geodata general requirements	33
	GEO_11 Rationale	33
	GEO_12 Rationale	33
	GEO_13 Rationale	34
	GEO_14 Rationale	35
	GEO_15 Rationale	35
	GEO_16 Rationale	36
	GEO_17 Rationale	38
	Rationales in 3.3.2. Vector Geodata – requirements	38
	GEO_18 Rationale	38
	GEO_19 Rationale	39
	GEO_20 Rationale	39
	Rationales in 3.3.3 Raster requirements	40
	GEO_21 Rationale	40
	GEO_22 Rationale	41
	GEO_23 Rationale	41
	Rationales in 3.3.5. Long Term Preservation Format Profiles	43
3.4	Rationales in Documentation requirements	50
	Rationales in 3.4 Documentation folders requirements	50
	GEO_24 Rationale	50

GEO_25 Rationale	51
Rationales in 3.4.1 Structure of geospatial records	52
GEO_26 Rationale	52
GEO_27 Rationale	53
GEO_28 Rationale	53
GEO_29 Rationale	55
GEO_30 Rationale	56
Rationales in 3.4.2 Rendering and visualisation	58
GEO_31 Rationale	59
GEO_32 Rationale	60
GEO_33 Rationale	61
GEO_34 Rationale	63
Rationales in 3.4.3 Software and algorithms	64
GEO_35 Rationale	64
GEO_36 Rationale	65
GEO_37 Rationale	65
Rationales in 3.4.4 Coordinate reference system information – requirements	66
GEO_38 Rationale	66
GEO_39 Rationale	68
Rationales in 3.4.5 Other - Contextual Documentation requirements	70
GEO_40 Rationale	71
GEO_41 Rationale	72
3.5 Rationales in metadata requirements	73
Rationales in 3.5 Geospatial Metadata requirements	73
GEO_42 Rationale	94
GEO_43 Rationale	94
4. Examples and tools	96
o Examples	96
o Tools	96
5. Postface	98
Appendix 1: Long-Term preservation format Profile for Geospatial Vector data using GML 3.2.1	100
Appendix 2: Long-Term preservation format Profile for Geospatial Raster data using TIFF baseline 6	101
Appendix 3: Mapping of INSPIRE metadata descriptions to archival descriptive metadata standards	102

1 Context

1.1 Purpose

The purpose of this guideline is to further explain and describe the “Specification for the E-ARK Content Information Type Specification for digital geospatial data records archiving” (also called CITS Geospatial in short). The goal is that as many people as possible will be able to understand the specification and, therefore, to also preserve geodata. The guideline is an evolving document, and more concepts and standards will be explained following the needs of the users of the specification.

1.2 Scope

This guideline provides further information and insights on preserving basic to moderately complex Geospatial records used within GIS, mostly in vector and raster formats. Geospatial domain is developing and many new formats are being used from LIDAR and point clouds to live-streamed data and distributed cloud-based formats. The CITS Geospatial specification is written in a very generic manner, so most of the proposed rules are also applicable to other data types. More examples of standards for newer formats will be added in the future. **This guideline is not describing the usage of CITS Geospatial for GIS and Web services. That topic is addressed in another Guideline.**

The CITS Geospatial specification builds on work done in the E-ARK projects.

1.3 Structure of the document

Section 2 contains an introductory section describing the concept of geodata in general and 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 geospatial data and GIS.

Section 3 provides a rationale for each of the requirements found in the CITS Geospatial specification. This is meant to provide a better basis for understanding the reasons behind the requirements. This section is primarily meant for technicians and developers of the specification, and it is a prerequisite that the reader has knowledge about geospatial data, formats for vector and raster data, the structure of geospatial data in GIS, GIS rendering functionality and the Common Specification for Information Package and the SIP, AIP and DIP specifications.

Section 4 contains an overview of available example packages and tools related to the CITS Geospatial specification as a means to get your hands dirty and take action in developing the field of geodata preservation.

2 Introduction to Geospatial data and approach to its preservation

This section contains an introductory section describing the concept of geospatial data in general and the digital preservation of them. 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.

2.1 Geospatial geodata

The aim of geospatial data is to represent real-world phenomena in a more generalised copy. Digital geospatial data virtually represents real-world objects and phenomena and are sometimes considered digital twins¹. The purpose of digital twins is to conduct operations in a virtual system, that gives us information about the object and its relation to other objects.

2.1.1 Basic Terms

A **Feature** is an abstraction of a real-world phenomenon. An example would be a digital line, representing a section of a road or river – an **instance** of that **Feature type**. We can have Feature types – used to represent different rivers, streams, etc.

An object in the real world can be represented as a feature with its **geometry** represented as a vector (lines represented by multiple sets of coordinates) or as a raster object (a set of pixels with the same values within a larger image). In geospatial data, this geometry always has coordinates based on a Geographical Coordinate Reference System.

A feature can also have its characteristics, which are described as **Feature Attributes**. A river represented as a vector line feature can have an attribute type “Name” that holds a text-based value “Danube” or an attribute “Length”, that holds a numeric value “2850,00”.

A collection of Features containing the same Feature Attribute structure are called a **Feature Dataset**.

A **Geospatial record** represents one or many objects in space and can be composed of one or many Feature Datasets.

2.1.2 Geospatial vector data

Geospatial vector Feature datasets can have different types of geometries (point, line, polygon, multipoint, multiline, multipolygon, and different types of curves and surfaces). Types may vary, based on standards and implementations. Examples:

- Points or multipoints: Shipwrecks, trees, mountain tops, GPS tracks, waterholes, cities, etc.
- Lines: Roads, rivers, power lines, isohypse, etc.
- Polygons: lakes, rivers (smaller scales), road surfaces (smaller scales), Country borders, etc.

¹ Digital Twin (https://en.wikipedia.org/wiki/Digital_twin)

- Mult polygon: A country with islands (multiple polygons or points representing one Feature Instance with one set of Feature Attributes)

In Figure 1, we can see that a Polygon vector **Feature** represents the border of Austria and has three different **Feature Attributes** that describe it.

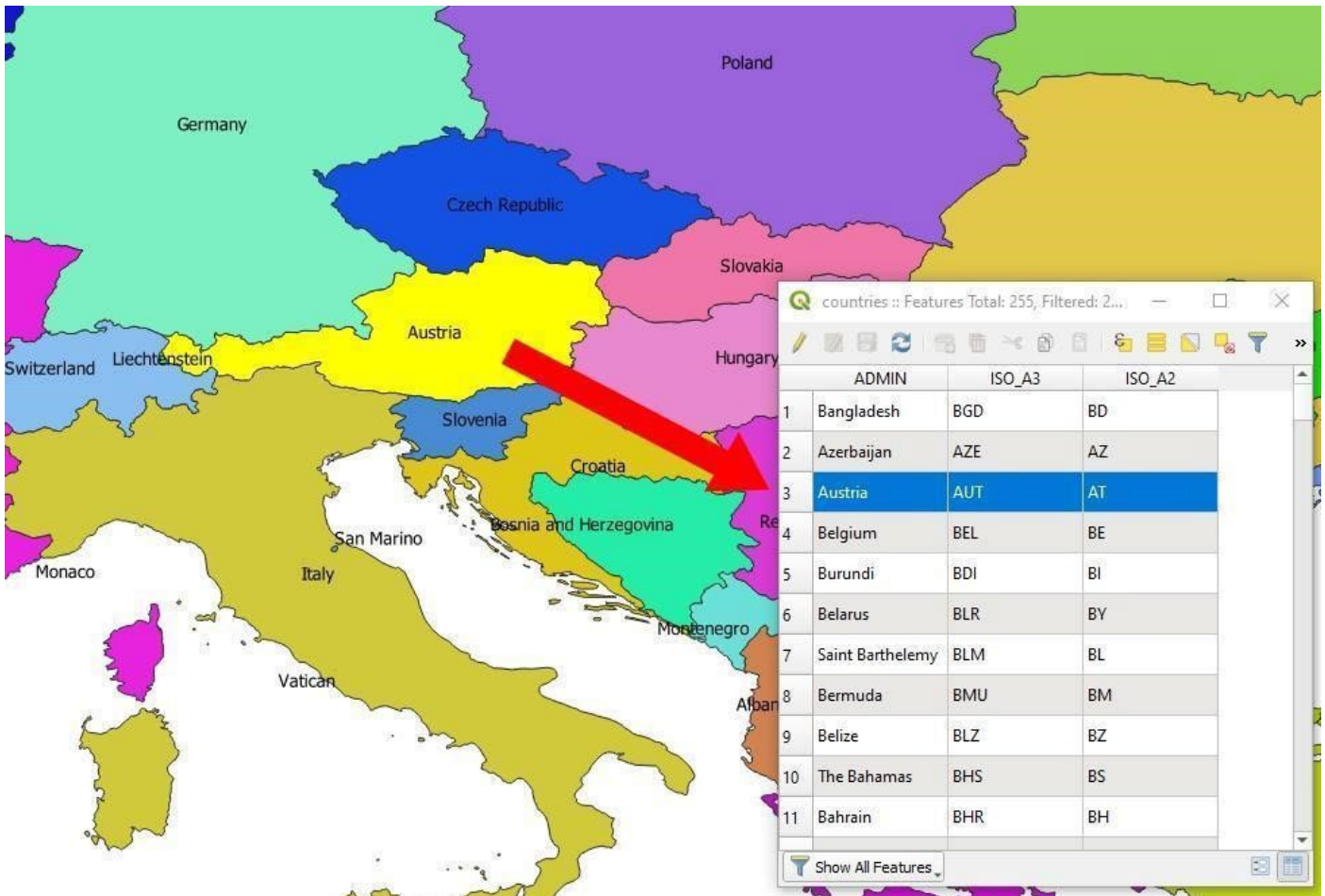


Figure 1 – EU Countries – An example of a geospatial vector dataset with Feature attributes

The explanation of Feature Attribute types and their meanings in the EU Countries dataset is commonly described in a **Feature Catalogue**.

In order to ensure the long-term preservation of geospatial data, it is necessary to ensure that there is a well-documented graphical component, well-defined descriptive attributes and a geographic coordinate system. Also, proprietary and undocumented geodata formats must be converted into a long-term preservation format that is well described and preserves all the significant properties of the original geodata format. Even if some formats are a de facto standard today, they may become unreadable in the distant future.

2.1.3 Geospatial raster data

All spatial raster data are a type of raster images, but not all raster images are spatial raster data. Both grid data and geospatial raster images are spatial raster data.

A raster image is commonly used to represent continuous features over large areas. For example, scanned maps, aerial images, satellite images, and grid files, that represent different values over the pixel area (Elevation, population density, gravimetry measurements, etc.)

A raster dataset is composed of a rectangular array of grid of pixels, each of which represents a value. One pixel represents the smallest unit of information. In the case of ordinary raster images, this value of each pixel usually represents a colour.

Image

A geospatial raster image, like a satellite image, scanned map or an orthophoto, also has pixels that represent colours which can then be rendered as a coloured image of the surface of the earth for human viewing. But it also has an orientation in space and can thus be placed correctly on the surface of the earth or in a 3D space according to a coordinate and reference system representing the earth or this 3D space.

Grid data

In the case of raster grids, the pixels of the image, also called raster cells, can also represent measurements of distances, areas, volumes or heights, derived calculations, classifications or any other units for the area in space that the pixel represents. A digital elevation model is grid data where each pixel represents a height in the terrain or surface of an area.

2.1.4 Standards for Geospatial Data

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

OGC² (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 CITS Geospatial are geospatial formats GML and GeoTIFF, and others³.

ISO TC/211⁴ (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***

² Open GIS Consortium (<https://www.ogc.org/>)

³ List of OGC Standards (<https://www.ogc.org/docs/is>)

⁴ ISO TC/211 (<https://committee.iso.org/home/tc211>)

2.1.4.1 CITS Geospatial alignment with ISO 19165-1:2018

CITS Geospatial is a technical specification that inherits its structure from the CSIP and therefore inherits its folder structure, metadata requirements, and principles (like representations). Basically, it is a Geospatial extension to the CITS Package, whose structure is adopted to serve different types of digital records.

ISO 19165-1 was first released in 2018, and its conception was designed from a geospatial data-centric approach, including preservation concepts from the OAIS standard.

Therefore, there are some conceptual differences between CITS Geospatial and ISO 19165-1:2018. For instance, the Geospatial Information model of the ISO standard includes preservation metadata elements (GP_PreservationMetadata) that are included as a part of the general ISO 19165-1 Metadata model. The elements are aligned with the OAIS Standard. However, the solution is merged with other metadata elements.

CITS Geospatial inherits the approach from CSIP, where separate METS and PREMIS profiles are used to describe fixity and provenance information. CITS Geospatial also describes the structure, context, and rendering content in greater detail than ISO 19165-1.

During the development of the 3.0 version of CITS Geospatial, we aimed to include as many of the geospatial data elements, its properties and auxiliary information and documentation as proposed in the ISO 19165-1. Due to the nature of the CSIP structure, a full complete adoption of the ISO 19165-1 is impossible. However, CITS Geospatial does support the inclusion of ISO 19165-1:2018 based metadata as descriptive metadata (requirement GEO_42).

ISO 19165-1 also describes the packaging mechanism of geospatial data using the Open Packaging Convention. This mechanism could also be adopted within the CITS Geospatial by creating a Long Term Preservation Format Profile that is compliant with the criteria as described in this document.

In conclusion, CITS Geospatial supports most of the content elements as described in ISO 19165-1:2018. It also supports the storage of the Metadata content. However, the package does not use the preservation elements in the same way. Since both structures are written to be machine-readable, a migration specification could be developed to support automated adoption.

2.1.5 Geospatial records in linked data environment

Linked data is one of the upcoming technologies, that offers greater connectivity between otherwise siloed data stores, bringing greater value to the data environment. At its core, Linked Data is about creating a web of interconnected data. It achieves this using:

- **URIs** (Uniform Resource Identifiers): Unique labels to identify abstract or physical resources (such as people, places, datasets, or concepts) unambiguously.

- **RDF (Resource Description Framework)**⁵: A standardized model for representing data as "triples." A triple consists of a subject (the thing being described), a predicate (the property or relationship), and an object (the value of the property).
- **Semantic Web Technologies**: Standardized languages like RDF, SPARQL (a query language for RDF), and OWL⁶ (an ontology language) that allow for reasoning and inference over linked data.
- **Ontologies**: Provide formal structure for preserving knowledge and relationships between different types of entities and can serve as a structure to help AI solutions, like LLMs (Large language models) return better results.

2.1.5.1 *Linked Geospatial Data: Adding the Spatial Dimension*

Linked geospatial data takes the principles of Linked Data and applies them specifically to geospatial information. Here's how it works:

- **Spatial Vocabularies**: Standardized vocabularies like GeoSPARQL⁷ provide ways to represent spatial relationships (e.g., "contains," "overlaps," "is near") within RDF triples.
- **Geospatial URIs**: Geographic entities (a city, a river, a building) are assigned unique URIs for disambiguation.
- **Geospatial Ontologies**: Ontologies provide the formal structure and definitions of geospatial concepts and their relationships. They can also represent the structure of data and can contain some logic

The benefits of Linked Geospatial Data are mostly seen in:

- **Enhanced Interoperability**: By breaking down data silos, linked geospatial data promotes seamless exchange and integration of data from different sources.
- **Discoverability**: Linked datasets are inherently discoverable on the Web, facilitating data reuse and the discovery of new relationships between datasets.
- **Semantic Enrichment**: Ontologies add context and meaning to geospatial data, enabling more powerful analysis and reasoning.
- **Cross-Domain Insights**: Connecting geospatial data with other Linked Datasets (e.g., demographics, weather) enables insights that cut across traditional disciplinary boundaries.

2.1.5.2 *Long-term preservation aspects of linked geospatial data*

From the preservation standpoint, this presents a challenge to current preservation paradigms. At its core, the archival preservation package should be a self-descriptive stand-alone informational unit. However, linked data has a different nature as it relies primarily on connections between different repositories.

In its current version, this guideline only addresses the question of where certain types of elements from the linked data environments can be stored in the geospatial archival package. In the future, we aim to provide further guidance on dealing with the intricacies of linked data, however, this is topic affects all preservation domains and is not specific to geospatial.

⁵ RDF – Resource description Framework - <https://www.w3.org/RDF/>

⁶ OWL – Web Ontology Language - <https://www.w3.org/OWL/>

⁷ GeoSPARQL - <https://www.ogc.org/standard/geosparql/>

2.2 Significant properties of geospatial data

The fundamental challenge of digital preservation is to preserve the accessibility and authenticity of digital objects over time and domains and across changing technical environments.

Significant properties are those aspects of the digital object which must be preserved over time for the digital object to remain accessible and meaningful. An institution with curatorial responsibility for digital objects cannot assert or demonstrate the continued authenticity of those objects over time or across transformation processes unless it can identify, measure, and declare the specific properties on which that authenticity depends. Nor can it undertake the preservation actions required to maintain access to those objects unless it can characterise their current technical representations with sufficient detail.

The InSPECT Framework⁸ report provides a method on how to decide on whether a property of a digital object is significant.

The significant properties of spatial vector and raster data are listed below using the following categories:

- **Content** – Information contained within the Information Object. For example, pixel values and location information (coordinates, orientation, pixel size), feature geometry, related feature attributes. This is usually referred to as data.
- **Context** – Any information that describes the environment in which the content was created, or that affects its intended meaning. For example, creator name, date of creation, spatial accuracy, lineage, source data, sensor information, etc. This is usually referred to as descriptive documentation or metadata.
- **Structure** – Information that describes the extrinsic or intrinsic relationship between two or more types of content, as required to reconstruct the performance. For example, the connection between the vector datasets and their joined tables, the relation between the image file and the world file or how feature datasets are organised within a GIS Project or a Web service mash-up. Structure can be described in the data itself or in an external documentation like a feature catalogue, a GIS project configuration file or metadata related to the data.
- **Rendering** – Any information that contributes to the recreation of the performance of the Information Object. For example, a colour map of pixel values, Styled Description layer for web services, documentation from a cartographic project etc. This is usually referred to as rendering documentation.
- **Behaviour** – Properties that indicate the method in which content interacts with other stimuli. For example, zoom feature, rendering algorithms, analysis functionalities, common transformation processes, documentation of original system functionality, user manuals, training materials, videos of system usage, etc. This is usually referred to as documentation or metadata.

These categories are used as folder names within the Information Package in which such documentation is stored.

⁸ Significant properties described in the INSPECT Framework document (<https://significantproperties.kdl.kcl.ac.uk/>)

2.3 Recommended reading list

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

Introduction to geodata and GIS

- GIS File formats: (https://en.wikipedia.org/wiki/GIS_file_formats)
- Geographical Information Systems (https://en.wikipedia.org/wiki/Geographic_information_system)
- Shashi Shekhar, Hui Xiong, Xun Zhou. Encyclopedia of GIS, Second Edition. Springer 2017, Print ISBN 978-3-319-17886-8, <https://doi.org/10.1007/978-3-319-17885-1>

Preservation of geodata

- Descriptions of geospatial formats and their suitability for long-term preservation <https://www.loc.gov/preservation/digital/formats/content/gis.shtml>
- More documentation on the preservation of geospatial data <http://geopreservation.org/>
- Description of Significant properties concept. <https://significantproperties.kdl.kcl.ac.uk/>

3 Rationale for requirements in CITS Geospatial

This 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, all the CITS Geospatial document requirements are listed and explained in greater detail. Additional explanation contains a more extensive *description*, an *example* and a *rationale* for why the specific requirement stands are 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 CITS_Geodata compliant. 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 for a better reference.

3.1 Folder structure requirements

CITS Geospatial text:

The CITS Geospatial information structure inherited its package structure from the E-ARK Common Specification for Information Packages (CSIP) (blue elements), and it is an extension of it (green elements).

A visualisation of a valid CITS Geospatial Information Package is illustrated in Figure 2. This Figure shows an example of a simple valid Information Package with one representation of a single vector dataset in a GML file format.

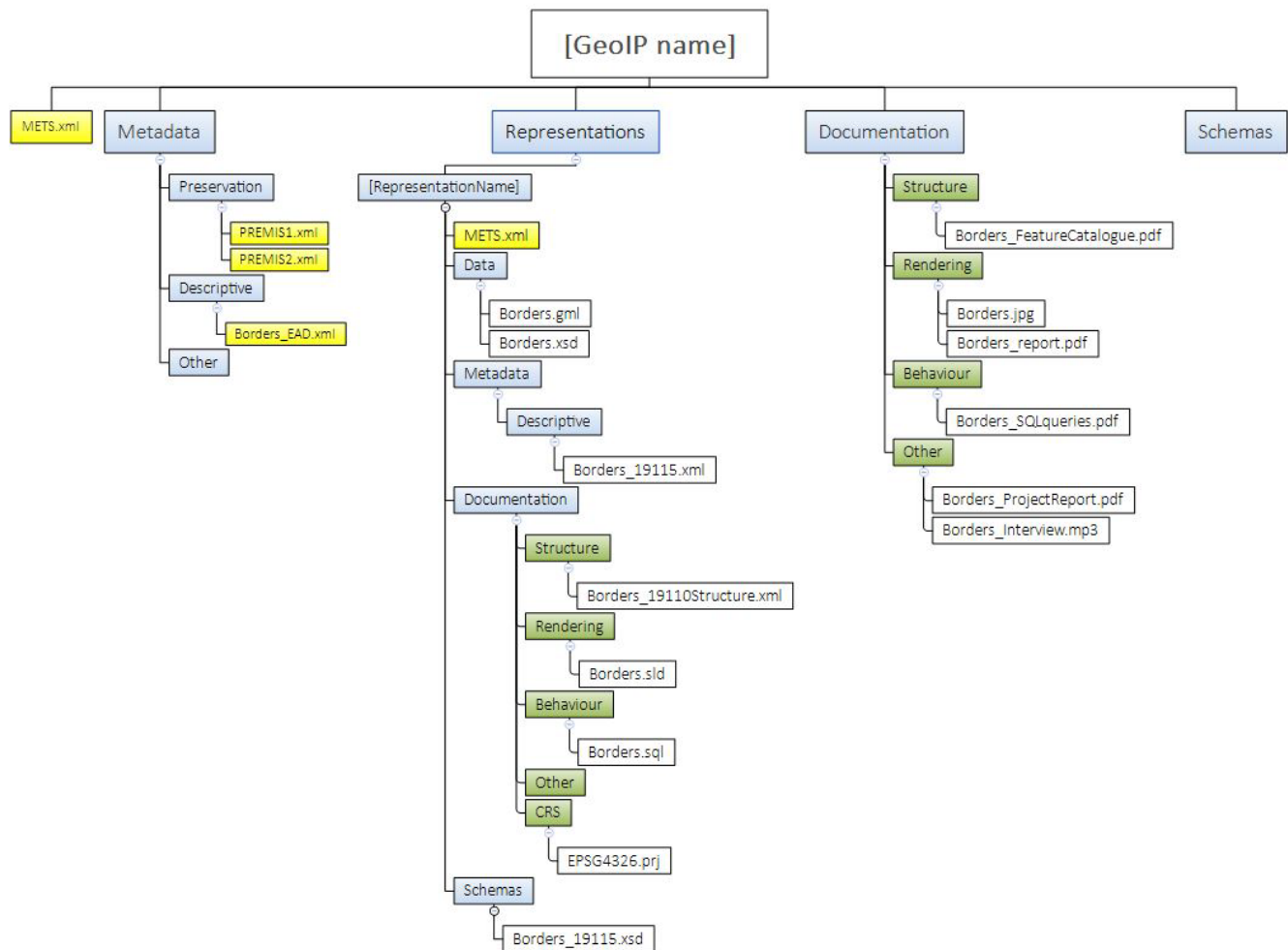


Figure 2 - Example Information Package folder structure

The folder structure in CSIP described in section <https://earkcsip.dilcis.eu/#folderstructureofthecsip> is extended with the following geo specific requirements on the folder structure:

In Figure 2 above from the CITS Geospatial specification, we can see a rather simple example of an Information Package, organised based on CSIP rules and extended with CITS for Geospatial folders. The predefined folders of the Information Package specified in the CSIP rules are in **blue boxes**, and the extended folders as introduced in the CSIP Geospatial are in **green boxes**. The predefined package Metadata files in the CSIP rules are in **yellow boxes**, and the example content of the Geospatial Information Package are in white boxes. Figure 2 is further explained in sections 3.2.

The purpose of data organisation into specific folders is to have known place holders for automated access of the standardised machine-readable content for faster reuse in the future. This is not the only possible approach, but it is in line with the philosophy of the CSIP logic, where we organise content within the Information Package according to the folder structure.

Other possible approaches could use a document pointing out where this type of data and documentation are within the package and how they connect to each other. There are different technologies that support this, like descriptions within METS files, OPC (Open Packaging Convention) files as mentioned within ISO 19165-1 standards, RDF and others. However, the whole idea is to have a commonly accepted convention that is interoperable among the EU member states and others interested in a common digital market.

GEOSTR1 rationale

Requirement:

GEOSTR1: XML schema documents for any structured descriptive geospatial metadata within a package **MUST** be placed in a sub-folder called `schemas` within the Information Package root folder and/or the representation folder. This requirement is an extension of CSIPSTR15.

Description:

An XML schema document is used for validation of structured XML files. XML schema documents must be placed in schemas folders in the Information package. If different XML schema definitions are used for validation of different structured descriptive geospatial metadata files in different representations, the XML schema document should be placed in a schema folder at representation level. If the same XML schema definition can be used for validation of all structured descriptive geospatial metadata files in the Information Package, it should be placed in a schema folder at the package level.

Example:

See Figure 2 above from the CITS Geospatial specification where the file `Borders_19115.xsd` (an XML schema definition file for validation of the `Border_19115.xml` file) is placed in a `representation/schemas` folder.

Rationale:

This requirement ensures that an XML schema definition file is easy to find when a structured descriptive geospatial metadata file must be validated against it.

GEOSTR2 rationale

Requirement:

GEOSTR2: A documentation folder on package or representation level **SHOULD** include a subfolder named `structure`. This requirement is an extension of CSIPSTR16.

Description:

A folder named `structure` should exist in the Information Package. This folder contains documentation about the extrinsic or intrinsic relationship between two or more types of content in the geospatial record, either in the original GIS or in the Information Package. This can be information about the connection between the vector datasets and their joined tables, the relation between an image file and a world file or how feature datasets are organised within a GIS Project or a Web service mash-up.

The `structure` folder can be placed at the package level if it contains a general description of the structure relevant for all geospatial data in the Information Package or at the representation level if the structure information only covers a specific representation in the Information Package.

Example:

See the folder `structure` in the path

`representations/[RepresentationName]/documentation/structure` in figure 2 above from the CITS Geospatial specification. In this example, the `structure` folder stores the file `Borders_19110Structure.xml`, which is a Feature Catalogue in a standardised xml-based file compliant with ISO 19110, describing the geospatial data in the `Borders.gml` file.

Rationale:

This requirement ensures that documentation of the structure of the geospatial dataset can be easily found by the user. Information about the structure of the geospatial record is required to reconstruct the performance.

GEOSTR3 rationale

Requirement:

GEOSTR3: A documentation folder on package or representation level **SHOULD** include a subfolder named `rendering`. This requirement is an extension of CSIPSTR16.

Description:

A folder named `rendering` should exist in the Information Package. This folder contains rendering information like a colour map of pixel values, Styled Description layer for web services, documentation from a cartographic project etc.

The `rendering` folder can be placed at the package level if it contains rendering information relevant for all geospatial data in the Information Package or at the representation level if the rendering information only covers a specific representation in the Information Package.

Example:

See the folder `rendering` in path

`representations/[RepresentationName]/documentation/rendering` in figure 2 above from the CITS Geospatial specification.

Rationale:

This requirement ensures that documentation about the rendering and visualisation of the geospatial dataset can be easily found by the user. Information about the rendering of the geospatial record is required to recreate the performance of the Information Object.

GEOSTR4 rationale

Requirement:

GEOSTR4: A documentation folder on package or representation level **SHOULD** include a subfolder named `behaviour`. This requirement is an extension of CSIPSTR16.

Description:

A folder named `behaviour` should exist in the Information Package. This folder contains documentation about methods in which content interacts with other stimuli (For example, the zoom feature, rendering algorithms, analysis functionalities, common transformation processes, documentation of original system functionality, user manuals, training materials, videos of system usage, etc).

Example:

See the folder `behaviour` in the path `representations/[RepresentationName]/documentation/behaviour` in figure 2 above from the CITS Geospatial specification.

Rationale:

This requirement ensures that documentation about behaviour originally executed on the geospatial dataset can be easily found by the user. Information about behaviour is required to recreate the original usage of the geospatial dataset.

GEOSTR5 rationale

Requirement:

GEOSTR5: A documentation folder on package or representation level **SHOULD** include a subfolder named `CRS`. This requirement is an extension of CSIPSTR16.

Description:

A folder named `CRS` should exist in the Information Package. This folder contains full descriptions of the Coordinate Reference System used in the archived geospatial dataset in the Information package.

Example:

See the folder `CRS` in the path `representations/[RepresentationName]/documentation/CRS` in figure 2 above from the CITS Geospatial specification.

Rationale:

This requirement ensures that documentation about CRS can be easily found by the user for usage or migration of the geospatial collection. Information about CRS used in geospatial records 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.

GEOSTR6 rationale

Requirement:

GEOSTR6: A documentation folder on package or representation level **SHOULD** include a subfolder named `other`. This requirement is an extension of CSIPSTR16.

Description:

A folder named `other` should exist in the Information Package. This folder contains other contextual information about the geospatial records, see section 3.1.2.

Example:

See the folder `other` in the path `representations/[RepresentationName]/documentation/other` in figure 2 above from the CITS Geospatial specification.

Rationale:

This requirement ensures that other contextual documentation can be easily found by the user. See also section Rationales in 3.4.5 Other - Contextual Documentation requirements.

Folder structure examples explained

3.1.1 Geo IP containing one vector representation

Below Figure 2 above is explained further.

3.1.1.1 Content

All the actual **Content** is placed in the `data` folder of the representation in figure 2 above. In this example, the `Borders.gml` file is considered as the **Content** (the Information contained within the Information Object) and thus placed in this folder. The accompanying `Borders.xsd` file is also considered as **Content**, even though it is actually a schema of the xml-based gml file, but it is in this case considered as part of the gml file.

3.1.1.2 Context

The `metadata` folder at the representation level stores the *descriptive* Geospatial metadata about the representation in a standardised machine-readable xml-file, compliant with geospatial metadata standards as described in the chapter 3.5 of this document. Thus, this file is considered as **Context** (Any information that describes the content or environment in which the content was created or that affects its intended meaning). The aim of this placement is to facilitate the automation of access to geospatial metadata.

Any other, non-standardised, documentation in digital form, containing **Context** information could be placed in the `documentation/other` folder at the package level like files `Borders_ProjectReport.pdf` and `Borders_interview.mp3`. If the information Package contains more representations, and there is documentation specific to a specific representation, then it could be placed within the `documentation/other` folder at the representation level (empty in Figure 2.).

The folder `Representations/[RepresentationName]Documentation/CRS` is a possible placeholder for full documentation of the used Coordinate Reference Systems in cases where the content files, in our case `Borders.gml`, do not contain a full definition of the CRS. In the case of GML, it only points to an external CRS repository. More on this topic in chapter 3.4.4.

3.1.1.3 Structure

In Figure 2, the folder `representation/[RepresentationName]/documentation/structure`, we see the file `Borders_19110Structure.xml`, which represents a Feature Catalogue, defining the `Borders.gml` and is a standardised xml-based file compliant with ISO 19110.

On the other hand, there is a folder `documentation/Structure`, containing a file `Borders_FeatureCatalogue.pdf` which contains a classic document describing the Feature Catalogue, however, this document is more descriptive and probably less suitable for automated machine-readable applications.

Generally, we aim to store the standardised machine-readable files within the representation. If the content is not available in such form, then it is generally stored within the package level `documentation` folder. Unless it is representation specific, then we recommend storing it within the representation.

3.1.1.4 Rendering

In Figure 2., the folder `representation/[RepresentationName]/documentation/rendering`, we see the file `Borders.sld`, which represents a standardised machine-readable file, that could be automatically used when rendering the `Borders.gml` within a web service.

On the other hand, there is a folder `documentation/rendering`, containing files `Borders.jpg` and `Borders_report.pdf`. These files also show how the data was originally rendered in the initial system, and the report represents an example of a derived information product to help us recreate adequate information products in potential future systems.

3.1.1.5 Behaviour

In Figure 2., the folder `representation/[RepresentationName]/documentation/behaviour`, we see the file `Borders.sql`, which represents a standardised machine-readable file that could be automatically used when trying to replicate the behaviour of the initial system to reproduce the information products like, the before mentioned `Borders_report.pdf`.

On the other hand, there is a folder `documentation/behaviour` that can contain additional extensive documentation on the initial system, its design, architecture, the use of the `Borders.sql` code and expected outputs.

3.1.1.6 Package level elements

This Information package also holds the two mandatory **METS.xml** files, one at package level (see Rationales in 3.2.2 Package METS requirements) and one at representation level (see Rationales in 3.2.2 Representation METS requirements).

Package level descriptive metadata (**EAD.xml**) and *preservation* metadata (**PREMIS.xml**) describing the Information Package are placed in the `metadata` folder at the package level. This is compliant with CSIP.

Generally, all Schema files (**.xsd**) for validating the different xml files in the Information Package are located in the `schemas` folder at the package level (empty on Figure 2.). However, there could be exceptions. If there are schemas specific to a representation (for example, one representation contains GML version 3.1.1 and the other version 3.2.1.), we can add a `Schemas` folder within a representation.

3.1.2 Geo IP containing multiple vector representations

In the example in Figure 3 below, a Geospatial Information Package (Geo IP) 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 interpret both representations correctly is put in the package level `documentation` folder. It is also possible to include a logical link to point to additional documentation being stored in a different Information Package (in case of a more extensive time series of the same data or similar records but from different organisational units).

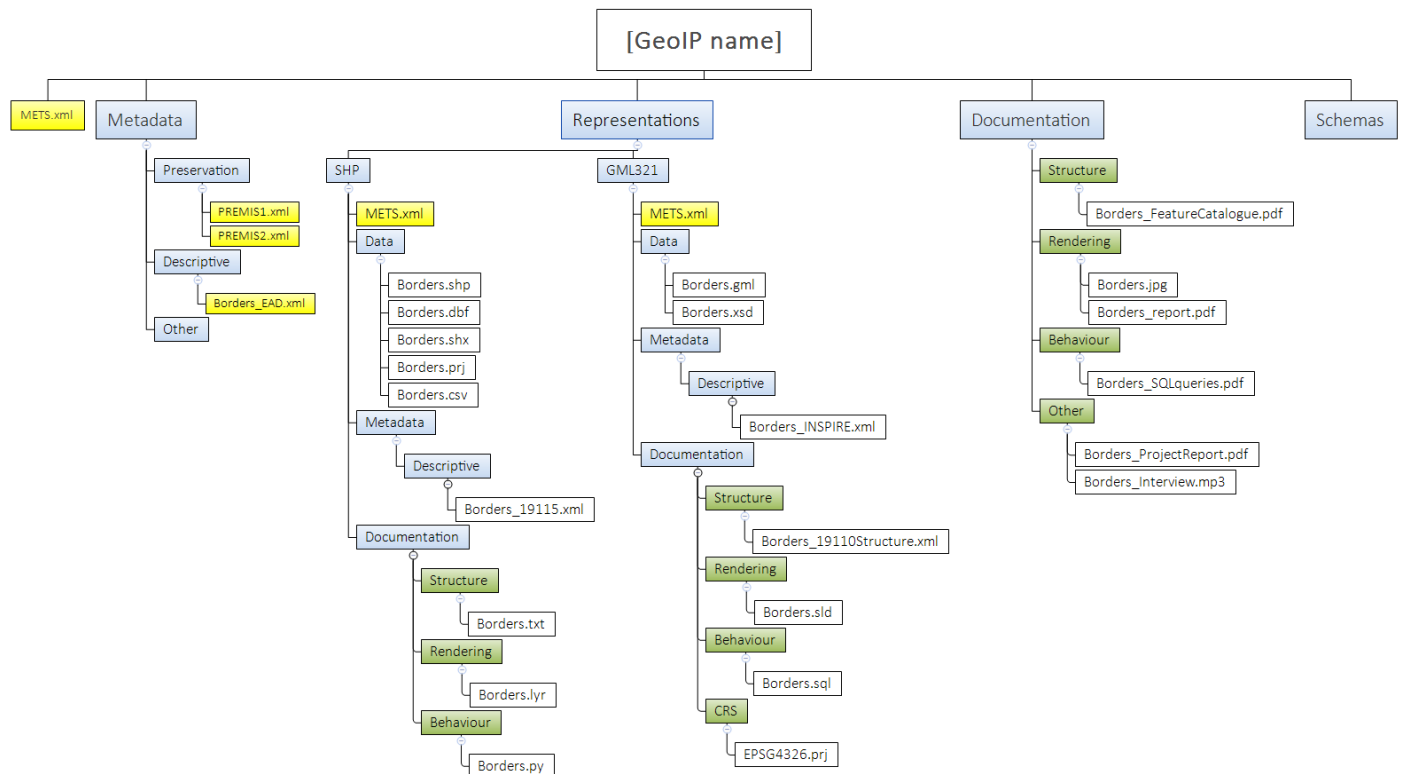


Figure 3 - Geospatial Information package with two representations containing vector data

Key differences between the representations are:

- Geospatial data in original format in representation SHP contains additional information in a separate non-spatial table (Borders.csv)
- Geospatial feature structure is described within a non-standard txt file in representation "SHP"
- Geospatial metadata in the original format is from a local system, compliant with ISO 19115:2003 (Borders_19115_export.xml). In representation GML321, it is updated to contain the INSPIRE compliant metadata.
- Since the GML321 representation references an EPSG register of CRS, we need to add a separate definition of the CRS (ETRS4326.prj) in the technical Documentation (exported from the EPSG registry). See chapter 3.4.4.
- Additional documentation for both representations is stored in the Root Documentation folder.

We can also see that some documentation, not in standardised machine-readable form, was placed under the root documentation folder.

3.1.3 Geo IP containing one representation of multiple raster datasets

In the example in Figure 4, a Geo IP contains one representation of multiple raster images covering an area with an accompanying vector file – containing an index of distribution and positions of the raster images. All additional documentation for the raster datasets is located in the root `Documentation` folder, except for additional CRS information in the `CRS` folder and a Geospatial Metadata xml file in the `Metadata/Descriptive` folder.

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

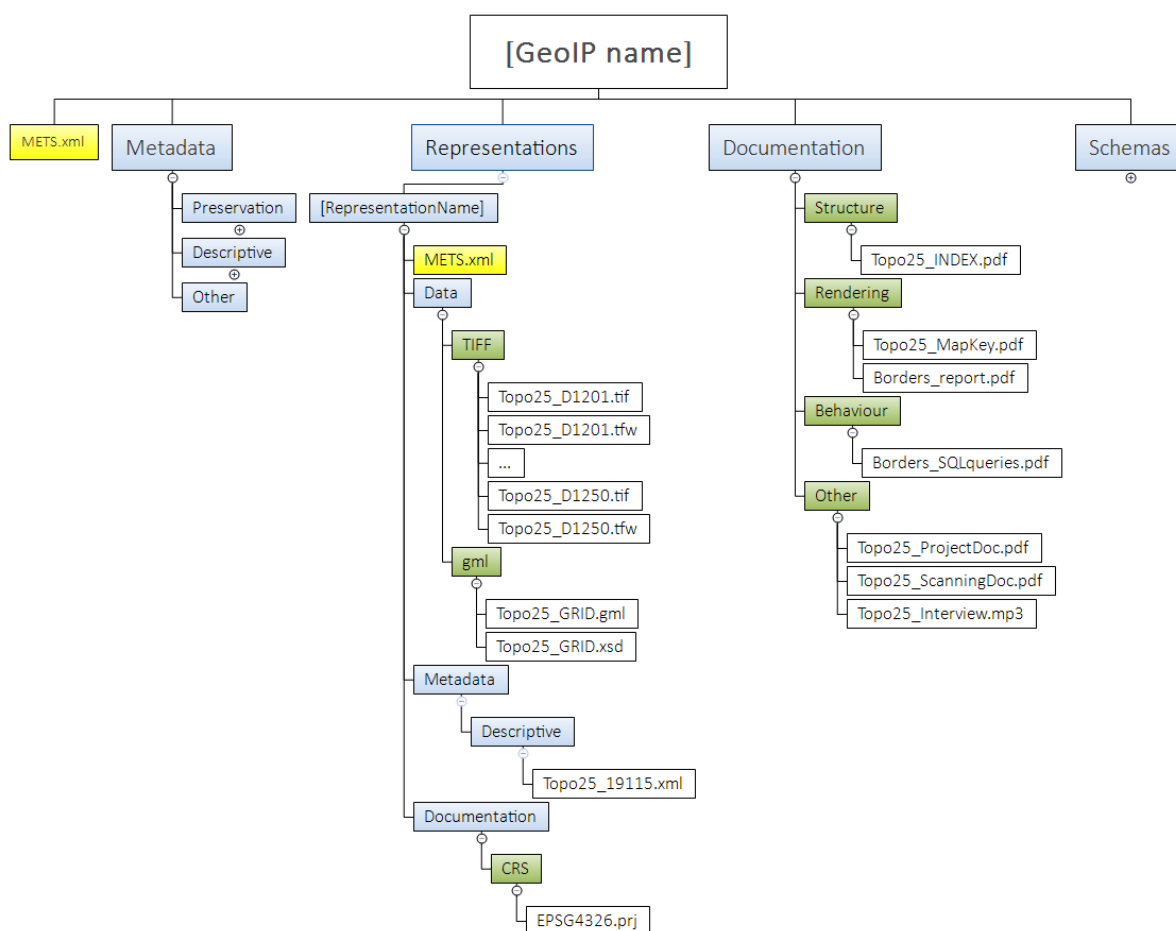


Figure 4 - Folder structure of a Geo IP containing one representation of multiple rasters

In this example, we can see that the Raster and Vector content within the `Data` folder can be organised in various ways. It could represent the structure within the initial system. However, to support more automated validation and future preservation actions it is recommended that geospatial records in different formats are stored in separate folders. In this example, GML files are stored within GML folder, and Raster data is stored in TIFF folder.

3.2 Rationales in METS Requirements

CITS Geospatial text:

A CSIP can consist of zero to many representations, and this is an important feature that needs to be taken into consideration when packing geodata files within CSIPs.

There can easily be different representations of the same geodata records located within one CSIP. For example, one package could consist of:

- one representation with geodata in original format;
- one representation with geodata in a long-term preservation format;
- one representation with geodata in dissemination formats;

There can be several representations of dissemination formats. There can also be many different types of geodata records and databases within the same package.

As for the CITS Geospatial specification, there always needs to be a minimum of one representation and, therefore a minimum of two METS.xml. The Package METS.xml has to be a general METS.xml describing if the package itself is mainly a CITS Geodata package, and then the Representation METS.xml describing what specific main Geodata types (Primarily vector or raster) the representation consists of.

A CITS Geospatial is building upon the general CSIP requirements but do not mention them here. Those requirements should be met before applying the requirements listed below.

GEO_1 Rationale

Requirement:

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

Description:

Geospatial records are placed in the Data folder in a Representation folder in a CSIP. There must be at least one representation of the geospatial data in a representation folder in the IP. Also, the IP must contain only one Package METS.xml file and one Representation METS.xml file in each representation folder.

This first requirement is central for the CITS Geospatial specification since it operates with two central terms: the Package METS.xml and the Representation METS.xml.

The "Package METS.xml" (there is only one) needs to be in the root of the package, and one "Representation METS.xml" needs to exist in the root of each representation within the package.

According to the CSIP specification, it is up to the user to define whether all files in an IP are described in the "Package METS.xml" or whether the user wishes to split it up and let "Representation METS.xml" describe the content within the representations. In the CITS Geospatial specification files in a representation folder are described in a "Representation METS.xml".

Example:

See the example in Figure 2, containing one Package level METS and two Representation level METS files.

Rationale:

An Information package without geospatial content stored within Representation folders doesn't qualify for a geospatial package. It could qualify for a general CSIP Package.

Rationales in 3.2.2 Package METS requirements

GEO_2 Rationale

Requirement:

GEO_2 Ref CSIP2	Type mets/@TYPE	For information packages that primarily contain geospatial data, the value in Package mets/@TYPE MUST be "Geospatial Data" as taken from the CSIP Vocabulary for Content Category.	1..1 MUST
--------------------	--------------------	---	------------------

Description:

This requirement ensures that the IP is compliant with the requirement CSIP2 in the CSIP specification, which states that there **MUST** be a TYPE-attribute with a value taken from the provided vocabulary for Content Category (<https://earkcsip.dilcis.eu/schema/CSIPVocabularyContentCategory.xml>).

"Geospatial Data" is the most relevant value from the Content Category vocabulary found in CSIP.

Example:

TYPE="Geospatial Data"

Rationale:

This information in the "Package METS.xml" can be used in a Finding Aid to group the IP as an IP that primarily contains "Geospatial Data".

GEO_3 Rationale

Requirement:

GEO_3 Ref CSIP4	Content Information Type Specification mets/@csip:CONTENTINFORMATIONTYPE	For information packages that primarily contain geospatial data, the value in Package mets/@csip:CONTENTINFORMATIONTYPE MUST be "citsgeospatial_v3_0" as taken from the CSIP Vocabulary for Detailed Content Type.	1..1 MUST
--------------------	---	---	------------------

Description:

This requirement is to make sure that the IP is compliant with the requirement CSIP4 in CSIP, which is a primary way of handling which kind of Content Information Type the package contains. In the case of multiple Content Information Types in an IP the value "MIXED" should be used.

Example:

csip:CONTENTINFORMATIONTYPE=" citsgeospatial_v3_0 "

Rationale:

When the “citsgeospatial_v3_0” value is used, this means that the package can be identified as stated to live up to the CITS Geospatial specification, and therefore be validated based on requirements in this specification.

GEO_4 Rationale

Requirement:

<p>GEO_4 Ref CSIP5</p>	<p>Other Content Information Type Specification mets/@csip:OTHERCONTENTINFORMATIONTYPE</p>	<p>For information packages that primarily contain geospatial data, the Package METS MUST NOT have a mets/@csip:OTHERCONTENTINFORMATIONTYPE</p>	<p>0..0 MUST NOT</p>
----------------------------	---	---	-------------------------------

Description:

The csip:OTHERCONTENTINFORMATIONTYPE-attribute is meant to specify which Content Information Type is used if the csip:CONTENTINFORMATIONTYPE-attribute has the value “OTHER”. It is not meant to exist if there are multiple Content Information Types in an IP. In the case of multiple Content Information Types, then the value “MIXED” should be used. Therefore, for CITS_Geospatial packages, there must not be a csip:OTHERCONTENTINFORMATIONTYPE-attribute in “Package METS.xml”. Note that this is different from the “Representation METS.xml”.

Example:

The `csip:OTHERCONTENTINFORMATIONTYPE` must not be present in the Package METS.xml file.

Rationale:

Because if it is an csip:OTHERCONTENTINFORMATIONTYPE-attribute then it is not a valid “CITS_Geospatial packages”.

GEO_5 Rationale

Requirement:

<p>GEO_5 Ref CSIP6</p>	<p>METS Profile mets/@PROFILE</p>	<p>For information packages that primarily contain geospatial data, the value in the @PROFILE MUST be "https://citsgeospatial.dilcis.eu/profile/E-ARK-GEOSPATIAL-ROOT.xml "</p>	<p>1..1 MUST</p>
----------------------------	--	--	---------------------------

Description:

This requirement is to make sure that the IP is compliant with the requirement CSIP6 in CSIP.

Each Content Information Type Specification (CITS) has its own METS profile where further requirements are added to the CSIP profile. The “<https://citsgeospatial.dilcis.eu/profile/E-ARK-GEOSPATIAL-ROOT.xml>” is thus an extending profile adding requirements to the CSIP requirements or changing their cardinality by changing optional to mandatory or specifying the number of occurrences of an element. It is not allowed to remove requirements from the CSIP profile since this will make the implementation invalid.

Example:

PROFILE="https://citsgeospatial.dilcis.eu/profile/E-ARK-GEOSPATIAL-ROOT.xml "

Rationale:

The CITS Geospatial METS profile is created for validation purposes. As per 9 April 2021 the profile has not yet been created, but it is planned.

GEO_6 RationaleRequirement:

GEO_6 Ref CSIP62	fileSec Representation Content Information Type Specification mets/fileSec/fileGrp[@USE='Representations']/@csip:CONTENTINFORMATIONTYPE	There MUST be a minimum of one mets/fileSec/fileGrp[@USE='Representations']/@csip:CONTENTINFORMATIONTYPE with the value "citsgeospatial_v3_0" as taken from the CSIP Vocabulary for Detailed Content Type that direct to the representation METS.xml in the representation folder containing geospatial data.	1..n MUST
---------------------	--	--	------------------

Description:

In this requirement, a filegroup (fileGrp) is named "Representations" in the "Package METS.xml" file. It is via the value "Representations" in the fileGroup USE-attribute on the filegroup element that one can mark up that within this filegroup will be a fileSec with a path to one or more METS-files in one or more representations. One METS file per representation.

Thus, this filegroup named "Representations" holds all the paths to the "Representation METS.xml" files in the IP.

Example:

```
<fileGrp USE="Representations" csip:OAISPACKAGE="SIP" csip:CONTENTINFORMATIONTYPE="citsgeospatial_v3_0" ID="ID_Rep1">
```

```
  <file ID="vector_protected_areas_METS.xml" USE="OTHER" MIMETYPE="application/xml"
    CREATED="2015-12-14T14:20:00"
```

```
      CHECKSUM="90c7527e6d4d3c3a6247ceb94b46bcf5" CHECKSUMTYPE="MD5" SIZE="8322">
```

```
        <FLocat LOCTYPE="URL" xlink:href="representations\repl\METS.xml"
          xlink:type="simple"/>
```

```
      </file>
```

```
</fileGrp>
```


GEO_7 Rationale

Requirement:

GEO_7 Ref CSIP105- CSIP112	StructMap METS pointer	For any fileGrp/@csip:CONTENTINFORMATIONTYPE with the value "citsgeospatial_v3_0" there MUST be a corresponding @div- representation in the StructMap-element	1..1 MUST
---	---------------------------	---	------------------

This requirement is to make sure that the IP lives up to the requirement CSIP105 to CSIP112 in CSIP. They are all related to "how to create a structMap-element". The structMap element holds all the internal links to folders and files in the IP. The link/URL is inserted in the xlink:ref element in the example below.

The METS structural map element is the only mandatory element in the METS specification. In the CSIP the <structMap> describes the higher-level structure of all the content in the package and may link to representations in the IP.

CSIP105 states that when a package consists of multiple representations, each described by a representation level METS.xml document, there is a discrete representation div element for each representation.

Example:

```
<structMap TYPE="PHYSICAL" LABEL="CSIP" ID="StructmapID_rep1">
  ...
  <div ID="Structmap_Div_ID_Representations" LABEL="Representations"/>
  <div ID="struct-map-reps-sub-div" LABEL="Representations/rep1">

    <mptr LOCTYPE="URL" xlink:type="simple"
      xlink:href="representations/rep1/METS.xml" xlink:title="ID_Rep1" >
    </mptr>

  </div>
  ...
</structMap>
```

Rationale:

Each representation div references the representation level METS.xml document, documenting the structure of the package and its constituent representations.

Rationales in 3.2.3 Representation METS requirements

Many of the requirements in this section are the same as in section 3.3 - it is important to notice the differences.

GEO_8 Rationale

Requirement:

GEO_8	Type	For representations that primarily contain geospatial data, the value in Package mets/@TYPE MUST be "Geospatial Data" as taken from the CSIP Vocabulary for Content Category.	1..1
Ref CSIP2	mets/@TYPE		MUST

Description:

The same as GEO_2. This requirement ensures that the IP is compliant with the requirement CSIP2 in the CSIP specification, which states that there **MUST** be a TYPE-attribute with a value taken from the provided vocabulary for Content Category (<https://earkcsip.dilcis.eu/schema/CSIPVocabularyContentCategory.xml>).

"Geospatial Data" is the most relevant value from the Content Category vocabulary found in CSIP.

Example:

TYPE="Geospatial Data"

Rationale:

This information in a "Representation METS.xml" can be used in a Finding Aid to group the IP as an IP that primarily contains "Geospatial Data".

GEO_9 Rationale

Requirement:

GEO_9	Content Information Type Specification	For representations that primarily contain geospatial data and that conforms to CITS Geodata the value in Package mets/@csip:CONTENTINFORMATIONTYPE MUST be "citsgeospatial_v3_0" as taken from the CSIP Vocabulary for Detailed Content Type.	1..1
Ref CSIP4	mets/@csip:CONTENTINFORMATIONTYPE		MUST

Description:

The same as SIARD_3⁹. This requirement is to make sure that the IP is compliant with the requirement CSIP4 in CSIP which is a central way of handling which kind of content information type the package contains, in this case it is the representation.

Example:

csip:CONTENTINFORMATIONTYPE=" citsgeospatial_v3_0 "

⁹ See the CITS SIARD specification requirements.

Rationale:

When the “citsgeospatial_v3_0” value is used, this means that the package can be identified as stated to live up to the CITS Geospatial specification, and therefore be validated based on requirements in this specification.

GEO_10 RationaleRequirement:

GEO_10	METS Profile	For information packages that primarily contain geospatial data the value in the @PROFILE MUST be	1..1
Ref CSIP6	mets/@PROFILE	"https://citsgeospatial.dilcis.eu/profile/E-ARK-GEOSPATIAL-REPRESENTATION.xml "	MUST

Description:

This requirement is to make sure that the IP is compliant with the requirement CSIP6 in CSIP.

Each Content Information Type Specification (CITS) has its own METS profile where further requirements are added to the CSIP profile. The “https://citsgeospatial.dilcis.eu/profile/E-ARK-GEOSPATIAL-REPRESENTATION.xml” is thus an extending profile adding requirements to the CSIP requirements or changing their cardinality by changing optional to mandatory or specifying the number of occurrences of an element. It is not allowed to remove requirements from the CSIP profile since this will make the implementation invalid.

Example:

```
PROFILE="https://citsgeospatial.dilcis.eu/profile/E-ARK-GEOSPATIAL-REPRESENTATION.xml "
```

Rationale:

The CITS Geospatial METS profile have been created for validation purposes.

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.

Sections 3.3 – 3.5 of this document do not discuss optimisations with respect to packaging and storage. The requirements for data, metadata and documentation only suggest what information is needed and the appropriate placement of it, not how it is packaged, stored and optimised for automatic handling.

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 MUST 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. However if a geospatial information product is created exclusively by linking to external data sources, an archival package can still be created as long as it contains proper links or references to external data (see also 2.1.5 on linked data).

Example:

A good example is the image "Figure 2." in chapter 3.1. The example illustrates an Information Package with one geospatial record in the GML format in the Representations/[RepresentationName]/data folder.

Rationale:

The purpose of making the Information package compliant with CITS Geospatial is to validate it against the criteria as described in the specification.

GEO_12 Rationale

Requirement:

GEO_12	Subfolders in data representations/[RepresentationName]/data	If there are more geospatial records in a representation, each geospatial file MAY be placed or grouped in subfolders in representations/[RepresentationName]/data	0..n MAY
--------	--	---	-----------------

Description:

This requirement allows the use of subfolders in the data folder. Subfolders inside subfolders are also permitted.

Recommendation:

Use short names or IDs when naming subfolders. If a file path in an IP is longer than 256 characters, this can complicate an IP's validation or migration.

Example:

representations/[RepresentationName]/data/TIFF

See Figure 4 in chapter 3.1.3, where geospatial records are stored within a TIFF or GML folder.

Rationale:

All geospatial files can be placed directly in the root of the data folder, but this can be a mess of files if a representation in the IP contains more than one dataset.

Often it makes sense to group bundles of files. For example, all files belong to a dataset in one folder or all files in the same format in one folder. Grouping files with relations to each other, for example, a GML file and the .xsd schema file validating the GML file, also enables fast identification of files related to other files. Placing each file in a folder with a unique ID can be a way to identify a file uniquely based on the folder ID.

Grouping and identifications of files can also be described in a METS.xml file in the structMap section. If producers organise their data in a folder structure, it is sensible to store the data in the way they are used to finding it. Grouping based on folder structures is more visual and easier for humans to comprehend. This can also be an alternative way (a contingency plan) to navigate the IP's when preservation systems based on the METS.xml file information fails or cannot be used due to lack of GDPR compliance.

GEO_13 Rationale

Requirement:

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

Description:

This requirement ensures 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 Rationales in 3.3.5 Long-term Preservation format profiles). This is a requirement for Submission Information Packages (SIPs) and Archival Information Packages (AIPs), however, a Dissemination Information Package (DIP) can only contain a dissemination format representation.

The Information Package can also contain other representations containing records in original or non-long-term preservation formats see description in GEO_14.

Example:

As proposed in the Long-Term Preservation format Profile for Geospatial Vector data using GML 3.2.1 in Appendix 1, we can have vector geospatial data stored in a GML 3.2.1. format, along with other standardised machine-readable documentation. See Figure 2 in chapter 3.1.

Rationale:

The idea of long-term preservation formats is fundamental in archival practice, as it prevents loss of data in the future. Since the CITS Geospatial is a specification for the long-term preservation of geospatial data, one representation in the IP must 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:

Figure 3 in chapter 3.1.2 shows 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).

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 becomes obsolete.

GEO_15 Rationale

Requirement:

GEO_15 Ref GEO_15	CRS definition	Every geospatial dataset MUST be accompanied with information about its underlying Coordinate Reference System (CRS) in one of two ways: <ul style="list-style-type: none"> • 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/systems (CRS) used in the geospatial file stored is/are 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. 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 in GEO_38.

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 4326¹⁰.

```
<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>
```

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.

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.

GEO_16 Rationale

Requirement:

GEO_16	Geographic location validation	The geographies in the geospatial records SHOULD be located within a fixed bounding box defined in the submission agreement between the producer and the archive according to the expected location and extent of the dataset	0..1 SHOULD
--------	--------------------------------	--	----------------

Description:

This requirement is intended to check against the expected spatial extent of submitted geospatial records. This extent can be limited to the extent of national or administrative unit boundary or determined in the valorisation phase of the Pre-Ingest process.

Example:

¹⁰ World Geodetic System 1984 (https://epsg.org/crs_4326/WGS-84.html?sessionkey=z5jf4xc886)

An archive could decide that all geospatial data from a specific country should be inside a fixed bounding box covering the areas of this country. Note that a bounding box is rectangular and aligns the axis of the CRS. Thus, it covers more than the exact national borders of a country.

A bounding box defined in a GML file:

```
<gml:boundedBy>
  <gml:Envelope srsName="EPSG:25832" srsDimension="2">
    <gml:lowerCorner>212481.60 6019669.40</gml:lowerCorner>
    <gml:upperCorner>961440.75 6510422.51</gml:upperCorner>
  </gml:Envelope>
</gml:boundedBy>
```



Figure 5 - Orange coordinates placed outside the expected bounding box due to error-prone migration

Rationale:

This requirement enables the validation of geospatial data migrated to preservation format from another CRS. When a transformation of coordinates from one CRS to another fails or has errors, the migrated coordinates are often placed outside the expected bounding box of the geospatial records, as illustrated in Figure 5.

In the exception where the geospatial records are located internationally or legitimately extend the proposed boundary, a new boundary should be determined by the owner of geospatial records and the archive, proposing the limitations.

GEO_17 Rationale

Requirement:

GEO_17	Metadata	Every geospatial dataset MUST be accompanied by a metadata file that describes the dataset with the basic required information	1..n MUST
--------	----------	---	------------------

Description:

This is a general requirement that ensures that all geospatial records within the Information Package are accompanied by descriptive metadata. If a geospatial information product is created exclusively by linking to external data sources the archival package still needs to include copies of metadata for external data used for the creation of the information project. Metadata can be stored in many different ways:

- With an accompanying standardised xml file (preferably to support automation)
- Using a database of metadata descriptions
- Containing metadata tags within the geospatial file (if the format supports it.)
- Within human-readable documentation (least preferred, but acceptable)

For further reference, see descriptions in GEO_42 and GEO_43

Example:

See examples from GEO_42 and GEO_43

Rationale:

Geospatial records commonly don't contain enough information to be fully self-descriptive. Therefore, it is necessary to follow standards for the description of geospatial records, using metadata that gives us additional context to understand and evaluate them.

Rationales in 3.3.2. Vector Geodata – requirements

CITS Geospatial text:

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

The requirements listed in this subchapter are specific to geospatial records in vector formats, contained within Information Package representations that primarily hold geospatial records and are marked as such in METS. If a representation contains mixed data types, it is still recommended that these validation rules are used for the geospatial vector data types.

GEO_18 Rationale

Requirement:

GEO_18	Valid geospatial vector file	Any geospatial vector datafile in representations/[RepresentationName]/data MUST be a valid	1..n MUST
--------	------------------------------	--	------------------

		vector file compliant with its respective format requirements (must pass the validation with the chosen validator for its format).	
--	--	--	--

Description:

This requirement ensures that the vector file format used for the geospatial vector data is valid.

Example:

A vector file in GML 3.2.1 format should be validated against the schema collection from Open Geospatial Consortium (OGC) to GML version 3.2.1.

Rationale:

To enable automated migration and dissemination of geospatial files in the archive, they must be valid according to format specifications.

GEO_19 Rationale

Requirement:

GEO_19	Feature attribute	Each Vector Feature dataset MUST contain at least one Feature attribute unique to each vector object	1..n MUST
--------	-------------------	---	------------------

Description:

This requirement states that each vector feature, for example, a polygon in a GML file used for outlining borders of countries, must be described by at least one attribute in some kind of feature dataset accompanying the geospatial vector file and identifying each instance of that feature uniquely.

Example:

See the example in Figure 1 where a polygon vector dataset containing borders of countries is set up with multiple vector features (in this case, polygons). Every feature is accompanied by at least one feature attribute. In Figure 1, a single vector feature has three feature attributes (ADMIN, ISO_A3, ISO_A2). The ADMIN feature attribute contains values for "Country name". Each instance of this feature attribute could hold the specific name of the country. And every vector instance has a unique name to differentiate them.

Rationale:

A vector shape (point, line, polygon, multipoint, multiline, multipolygon, etc.) needs to have an attribute to differentiate it from other objects in the dataset and to prevent duplicate entries. A unique feature attribute describes the geometry uniquely and gives it a reference point or a unique key, allowing other GIS functions like the joining of other attributes.

GEO_20 Rationale

Requirement:

GEO_20	Long-Term preservation format Profile for Geospatial Vector data	Geospatial vector data in a long-term preservation representation SHOULD comply with the requirements for the respective Long-Term preservation format Profile for Geospatial Vector data (see chapter 3.3.5)	0..n SHOULD
--------	--	--	--------------------

Description:

This requirement recommends that geospatial vector data be preserved in the preservation format specified in a Long-Term preservation format Profile for Geospatial Vector data (see Rationales in 3.3.5 Long-term Preservation format profiles). The CITS Geospatial specification doesn't directly specify a long-term preservation format. Instead, it allows for Long-term Preservation Format profiles to support various formats for different purposes and local implementations. The proposed example in Appendix 1 is a showcase to support the creation of one's own profile if one intends to use other formats for long-term preservation.

Example: An example of a Long-Term Preservation Format Profile for Geospatial Vector data using GML 3.2.1 is available in Appendix 1 of this document. It proposes criteria for the use of the GML 3.2.1 format for the long-term preservation of geospatial vector data.

Rationale:

Since at least one representation containing data in a long-term preservation format is required for the Information Package to be compliant with CITS Geospatial, this requirement specifies how exactly to do this for the vector geospatial data.

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:

The requirements listed in this subchapter are specific to geospatial records in raster formats, contained within Information Package representations that primarily hold geospatial records and are marked as such in METS. If a representation contains mixed data types, it is still recommended that these validation rules are used for the geospatial raster data types.

GEO_21 Rationale

Requirement:

GEO_21	Valid geospatial raster file	Any geospatial raster datafile in representations/[RepresentationName]/data MUST be a valid raster datafile compliant with its respective format requirements (must pass the validation with the chosen validator for its format).	1..n MUST
--------	------------------------------	---	------------------

Description:

This requirement ensures that the raster file format used for the geospatial raster data is valid.

Example:

A raster file in GeoTIFF could be validated against the OGC GeoTIFF standard.

Rationale:

To enable automated migration and dissemination of geospatial files in the archive they must be valid according to format specifications.

GEO_22 RationaleRequirement:

GEO_22	Long-Term preservation format Profile for Geospatial Raster data	Raster data in the long-term preservation representation SHOULD comply with the requirements for the respective Long-Term preservation format Profile for Geospatial Raster data (see chapter 3.3.5.)	0..n SHOULD
--------	--	--	-----------------------

Description:

This requirement recommends that geospatial raster data are preserved in the preservation format specified in a Long-Term preservation format Profile for Geospatial Raster data (see Rationales in 3.3.5 Long-term Preservation format profiles). The CITS Geospatial specification doesn't directly specify a long-term preservation format. It instead allows for Long-term Preservation Format profiles to support various formats for different purposes and local implementations.

Example: An example of a Long-Term Preservation Format Profile for Geospatial Raster data using TIFF Baseline 6 is available in Appendix 2 of this document, and it proposes criteria for usage of the TIFF format for the preservation of geospatial raster data.

Rationale:

Since at least one representation containing data in a long-term preservation format is required for the Information Package to be compliant with CITS Geospatial, this requirement specifies how to do this for the raster geospatial data.

GEO_23 RationaleRequirement:

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.

Example:

The map below illustrates a tiling index.

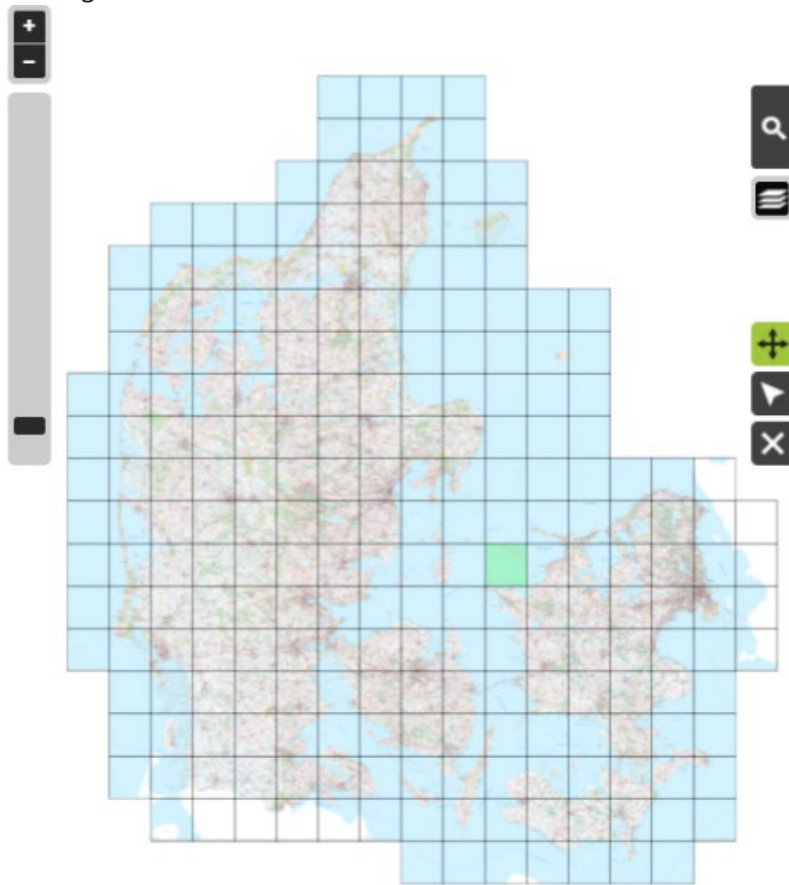


Figure 6 - Example of a tiled vector index for raster geodata

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.

Example:

Two examples of Long Term Preservation format Profiles are listed in these appendixes:

- Appendix 1: Long-Term preservation format Profile for Geospatial Vector data using GML 3.2.1
- Appendix 2: Long-Term preservation format Profile for Geospatial Raster data using TIFF baseline 6

Rationale:

The preservation of both the amount in formats and the formats' significant properties entails significant complexities for an archive's ability to safeguard digital preservation. If data in a digital archival collection is homogeneous and based on a few selected preservation formats, it results in the ability to provide access to archived data in the future and maintain cost-effective preservation. However, homogeneity often occurs at the expense of authenticity, and it is in this area of tension that it makes sense to use a framework with criteria for selecting long-term preservation formats in a structured and documented way.

Description:

An example of a framework for selecting long-term preservation formats using a matrix to score different criteria for each preservation format candidate is described below¹¹. Figure 7 illustrates the use of this framework, accessing a suitable preservation format for georeferenced raster images.

The format that achieves the highest total score is considered to be the most suitable preservation format. Overall assessment results in a recommendation written in note format, which can be taken further as the basis for a subjective management decision to approve a new preservation format.

Arguments for each score in the matrix shown in Figure 7 should be described as illustrated in Figure 8.

¹¹ Format assessment matrix, The Danish National Archives (<https://github.com/the-danish-national-archives/concept-model/tree/main/P2%20Format%20Assessment>)

Criterion	Weight	Format				
		TIFF+TFW	GeoTIFF	JPEG2000	Geopackage (GPKG)	ASCII GRID (*.asc)
Prevalence	2	1	1	1	0	1
Prevalence II	2	1	1	1	0	0
Lifespan	1	1	1	1	0	1
Lifespan II	3	0,5	1	1	1	0
Documentation	2	1	1	1	1	1
Documentation II	1	1	1	1	0	1
Documentation III	1	1	1	0	1	1
Licensing	2	1	1	0	1	1
Structure	1	1	1	1	0	1
Structure II	1	0	0	0	0	1
Significant properties	3	1	1	1	1	0
Dissemination	1	1	1	1	1	1
Searchability	1	0	0	0	0	1
Interoperability	2	1	1	1	1	1
Testing	2	1	1	1	0	0
Testing II	2	1	1	1	1	0
Compression	3	1	1	1	1	1
Storage	1	0	0	1	1	0
Migration	2	1	1	1	1	1
Compatibility	1	1	1	1	1	1
Score		29,5	31	29	23	21
		0,87	0,91	0,85	0,68	0,62

Figure 7 - Example of format assessment of preservation format candidates for georeferenced raster images

Each individual criteria for measuring how persistable a format is as well as the relative importance (weighting) of the criteria in the matrix in figure 7, is described in detail here:

Prevalence

The format is frequently used within its content type. It is a strength for preservation formats to be frequently used across user segments for which the format is created. The more international and the more general the prevalence, the more established the format can be considered to be, which is a strong indicator in relation to preservation suitability. The criterion is weighted 2, as general prevalence is an important indicator for how established a format is and opens up technical support for many years to come.

Prevalence II

The format is frequently used for preservation at cultural heritage institutions. It is a strength for preservation formats to be already in use at other cultural heritage preservation institutions. This occurs because like-minded actors in unison face the same challenges, thus creating opportunities for sparring and resource pooling as well as reducing the risk of technical obsolescence. The criterion is weighted 2, as the prevalence of a format for preservation purposes is a strong basis for cooperation on common issues.

Lifespan

The format has been around for at least ten years. An established format is more robust than a new format that has not been around for a long time, where it can be difficult to assess the direction in which the prevalence will go. The criterion is weighted 1, as lifespan is, to a larger extent, an indication rather than something tangible.

Lifespan II

The format has good future prospects. The criterion is difficult to quantify, but “good future prospects” are based on an estimate on whether the prevalence and software support will increase over the coming years rather than decline. The criterion tells something more about the development of the format rather than just age. The criterion is weighted 3, since whether a format is growing rather than dying is crucial to assessing the preservation suitability of the format.

Documentation

The format is standardised. It is a strength for preservation formats to be technically well-documented. The documentation makes it possible to analyse the functionality of a format and to develop system tools for characterisation, validation, and migration. The criterion is weighted 2, as it is an insurmountable task in terms of resources to disassemble a format for the purpose of developing the system tools for receiving and storing data.

Documentation II

The format standard has not been updated within the last ten years. The more stable a format, the easier it is to integrate it into the preservation planning since the prioritisation of resources for monitoring and to analyse the format decreases. The criterion is weighted 1, as it is only an indication of stability, and the challenge can be solved by adding the necessary resources to monitor and analyse the format more thoroughly.

Documentation III

The documentation of the format is well-described and can be read easily. It is a strength for preservation formats to have well-described and clear documentation written in an understandable language. At the same time, the documentation must not be too long to read. Here we set the limit at 1,000 pages. The criterion is weighted 1, as we do not necessarily have to understand the documentation in all its details.

Licensing

The format is open source. It is a strength for preservation formats to be independent of companies, including the strategic and financial interests of companies. In practice, the strength comes from the fact that working with the formats is cost-effective, the future holds fewer risks, and the possibility that the format will become more prevalent over time is greater. The criterion is weighted 2, as the free use of the format is a sign of health and an important factor behind the creation of an archival academic community around the format.

Structure

The format is self-supporting (not a container format). The criterion means that a format consists solely of its own format. Some formats allow multiple formats to be stored in them, making the format a container for other formats. Examples of container formats are video formats that store images in one format, audio in another format, and any subtitle tracks in a third format. Zip is also a container format that can have all sorts of formats stored in it. The criterion is weighted 1, as container formats only add the implication (but also complexity) that we must approve all the formats of the container as preservation formats.

Structure II

The format can be read as plain text. It is a strength for preservation formats if both machines and people can interpret the data directly. Plain text means any basic text program can open the file, interpret the character set, and render the binary data as readable text. Formats that cannot be read as plain text will not produce readable text after the interpretation of the character set. The criterion is weighted 1, as binary formats are not problematic as long as the technical documentation is saved.

Significant properties

The format supports most of the significant properties within its content type. The investigation of the significant properties of the format is a prerequisite for grading. Here it is ideal if a preservation format has all the properties of the original format so that significant properties are not lost during migration. The criterion is weighted 3, as significant properties are an essential indicator of whether data is preserved as it is created.

Dissemination

The format can be reused without conversion. If it is necessary to convert data from the preservation format in order to make data more usable, it will add an extra work process and risk of loss of authenticity, which would be undesirable. The criterion is weighted 1, as data conversion is recognised to be problematic, but nonetheless, conversion is common practice in the migration strategy, and system-independent formats are often more suitable for long-term preservation (see compatibility below) than proprietary formats that are easy to use today thus causing the need for conversion to more usable formats on dissemination.

Searchability

The format has searchable information. If the format allows for searches within the content, it is easier to index the content and make it available. The example here is a TIFF of a digital document, which exclusively stores an image of the document, which must subsequently be OCR-treated to have searchable text in a separate text or database file. A PDF version of the same digital document will have the text stored as embedded searchable information. The criterion is weighted 1, as the technical possibilities of processing offered by the format have an impact on the ability of an archive to preserve and make data available in a cost-effective and authentic way. However, there are often tools that can mechanically remedy the lack of searchability.

Interoperability

The format is suitable for data exchange. Some formats are more suitable for data exchange than others. This applies, for instance, to mark-up formats such as XML or formats that are widely used for exporting and importing data between IT systems. The criterion is weighted 2, as interoperability makes it easier to work with data in automated processes, and broad system support for data exchange is an indication of robustness.

Testing

The format has tools for identification and characterisation. The verification of format identity makes it possible to ensure that data are transferred to the archive according to format specification and supports subsequent preservation actions such as migration. The purpose is to gain knowledge about whether a file has the format it claims to have through its file extension, as well as gain knowledge about the metadata of the file. It is optimal if a testing tool is already mature for use in a fully developed and open version, where the developers still offer bug fixes and updates for new operating systems. An open version means that the tool can be freely adapted to suit your own business needs. The criterion is weighted 2, as identification and characterisation are important components of the testing process that, in a very basic way, allow for more complex preservation tasks.

Testing II

The format has tools for validation. The validation of the structure and content of files makes it possible to ensure that data is transferred to the archive according to the format specification. Here it is optimal if the tool is already mature for use in a fully developed and open version, where the developers still offer bug fixes and updates for new operating systems. An open version means that the tool can be freely adapted to suit your own business needs. The criterion is weighted 2, since validation is an important component of the testing process that allows for the performance of more complex preservation tasks.

Compression

The format is uncompressed or has lossless compression. Compression can cause the quality of data to be reduced so that they do not have the same accuracy as an original format may have had. However, compression may be allowed if the compression is lossless. The criterion is weighted 3, as lossless compression is a crucial factor in ensuring authenticity.

Storage

The format occupies less storage space than the average for its content type. Storage is the capacity in the physical media that the preservation formats occupy. Storage is a relative factor that is affected by technological development, which over time offers greater capacity at the same price, but the criterion's significance is also affected by the budgetary framework of the preservation institution. The criterion is weighted 1 because storage space does not have the same economic significance as in the past.

Migration

The format can be migrated with an acceptable loss of significant properties to another format. Migration here is defined as converting data from one original format to another format within the same content type. Next, the availability of data migration tools is a prerequisite for pursuing a migration strategy, and the better the tools the market can offer, the stronger the format's preservation suitability. In an ideal scenario, migration should result in the retention of all significant properties, the same intellectual content, and the same visual representation. The criterion is weighted 2, as the ability to migrate data to other formats indicates robustness.

Compatibility

The format is compatible with multiple operating systems and applications. In practice, this means that the format is both system and software independent without locking it to specific operating systems or programs, and several competing programs must exist in the market for rendering the data of the format. The criterion is weighted 1, as system independence have an impact on the possibility of pursuing a migration strategy.

GeoTIFF	
CRITERIA	Description
Prevalence	GeoTIFF is probably the most commonly used raster format with embedded georeferencing metadata. The reason for broad acceptance of GeoTIFF is that it provides more convenient and more complete georeferencing information than TIFF+TFW+PRJ combination. GeoTIFF is more convenient because location information and CRS information is provided within tags in the header of TIFF file and additional files TFW and PRJ are not needed.
Prevalence II	TIFF image file format without geo metadata tags is frequently used for preservation of digital images. It is on a list of preferable digital formats in many preservation institutions like Library of Congress and The National Archives in UK.
Lifespan	GeoTIFF image format has been in usage for more than ten years. The first version of GeoTIFF was published in 1995. The latest revisited version of GeoTIFF (1.0) was published in 2000.
Lifespan II	GeoTIFF image format has good prospects in the preservation of georeferenced digital images due to extensive software support. The new standard of GeoTIFF (OGC GeoTIFF 1.1) from 2019 formalises existing GeoTIFF specification version 1.0 and update it to added data in EPSG Geodetic Parameter values.
Documentation	GeoTIFF image format version 1.0 is well standardised and well documented as it is OGC GeoTIFF 1.1 version.
Documentation II	The latest version of GeoTIFF image format (v 1.1) is from year 2019 and is less than 10 years old but it is backward compatible with revision 1.0 from 1995.
Documentation III	GeoTIFF version 1.1 image format OGC documentation is substantial and detailed as well as it is latest documentation for version 1.0 from 2000.
Licensing	Portion of GeoTIFF 1.0 are copyrighted by authors Mike Ruth and Niles Ritter. Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct or commercial advantage, and this copyright notice appears. There are no concerns about patents concerning GeoTIFF use of TIFF tags. No concerns about patents for underlying TIFF 6.0
Structure	GeoTIFF is a self-supporting image format in the same way as TIFF image format is.
Structure II	GeoTIFF image file format is not a plain text format nor its data in the header. The context of GeoTIFF image file cannot be interpreted directly.
Significant properties	<p>GeoTIFF raster format supports storing imagery and is regularly used to store grid data.</p> <p>GeoTIFF image format supports the same significant properties as the aforementioned TIFF+TFW+PRJ combination. The difference is that all the location information is stored within TIFF tags.</p> <p>GeoTIFF is essentially a TIFF 6.0 compliant image (raster) file with additional tags where georeferencing information is stored. In all other properties, the GeoTIFF image format is same as TIFF. The difference between TIFF and GeoTIFF format is in GeoTIFF capability to store additional georeferencing information within 6 GeoKey extended tags. Additional GeoKeys provides most significant information regarding georeferencing. Despite providing extended information about projection, datums and ellipsoid GeoTIFF GeoKeys are not capable of storing information about datum transformation parameters that are needed in custom defined local coordinate systems. As TIFF image format GeoTIFF also doesn't provide raster spatial coverage.</p> <p>A limited number of contextual and structural properties could be described inside the TIFF metadata (tag) structure. So other Context and Structural properties need to be added with external metadata or documentation.</p>
Dissemination	GeoTIFF image format can be directly reused as it is without conversion.
Searchability	GeoTIFF image file format is binary image format, and its image content is not searchable. GeoTIFF image file format is capable of storing image metadata information into Baseline fields like image resolution, dimension, creation date, bits per sample, compression, orientation, author, copyright information,... With software that is capable of reading TIFF Baseline and Extended field information is possible to search TIFF image metadata. Additionally, to those TIFF 6.0 based metadata, GeoTIFF is capable of storing georeferencing information into so-called GeoKeys. GeoKeys could be read with any application that support GeoTIFF image format.
Interoperability	GeoTIFF image format provides a high level of interoperability. GeoTIFF image format is supported by a plethora of operating systems and programs. GeoTIFF images could be transferred between different operating and file systems without conversion.
Testing	<p>For identification and characterisation of GeoTIFF images several programs that are capable of identifying and characterised TIFF image exist. Most capable of them are:</p> <ul style="list-style-type: none"> - DROID - JHOVE validation tool with TIFF module

	<ul style="list-style-type: none"> - DPF Manager - Checkit_tiff - ImageMagick - ExifTool
Testing II	<p>For validation of TIFF image files several programs that are capable verify TIFF image exist. Most capable of them are:</p> <ul style="list-style-type: none"> - JHOVE validation tool with TIFF module and - DPF Manager - Checkit_tiff
Compression	<p>Compression of GeoTIFF image is an option, and GeoTIFF image format support several kinds of compression (5):</p> <ul style="list-style-type: none"> - No compression - Lossless compression (CCITT FAX3, FAX4, LZW, RLE, DEFLATE) - Lossy compression (JPEG)
Storage	<p>Non-compressed GeoTIFF image file format is probably the most used image format in image preservation, and so it is occupying average amount of storage space while GeoTIFF image with JPEG compression file size is smaller.</p>
Migration	<p>GeoTIFF image format is widely used and standardised image format supported by a large number of migration software. Migration of non-compressed GeoTIFF image should be simple with potential small loss of image metadata. With validation after migration, the loss of image metadata could be minimised.</p>
Compatibility	<p>GeoTIFF format is compatible with many operating systems and file systems. In all widely used operating systems (MS Windows, Linux, Apple macOS, BSD) are programs and tools that support reading, writing and visualisation of GeoTIFF image format.</p>
Overall Assessment [0,91]	<p>The overall score of this format is 31 points out of 34, which gives it a weight of 0,91% of all points. Out of all formats used in the geospatial domain, GeoTIFF is very suitable for long term preservation. It supports almost all raster geodata use cases. In comparison to the TIFF+TFW+PRJ format, it provides a more compact solution. For preservation purposes, the metadata tags within the GeoTIFF should be specified.</p>

Figure 8 - Example of descriptions of scores in column "GeoTIFF" in figure 7

3.4 Rationales in Documentation requirements

Rationales in 3.4 Documentation folders requirements

CITS Geospatial text:

Geospatial records are rarely in a form that is sufficiently self-explanatory to be used and interpreted adequately by itself. Consequently, additional information describing context, structure, rendering and behaviour is required to enable the user to understand, interpret and reuse preserved geodata properly. This chapter describes the requirements for Documentation for geospatial datasets (where it is applicable). Ideally, a standardised machine-readable format is preferred. However, any other form of documenting the System is welcome. Standardised machine-readable formats should be placed within the representation. Other documentation should be placed within the package level Documentation folder.

GEO_24 Rationale

Requirement:

GEO_24	Package level documentation	Documentation covering all representations in the Information package SHOULD be placed in documentation/ on package level	0..n SHOULD
--------	-----------------------------	--	-----------------------

Description:

This requirement states where the package-level documentation is to be placed.

Example:

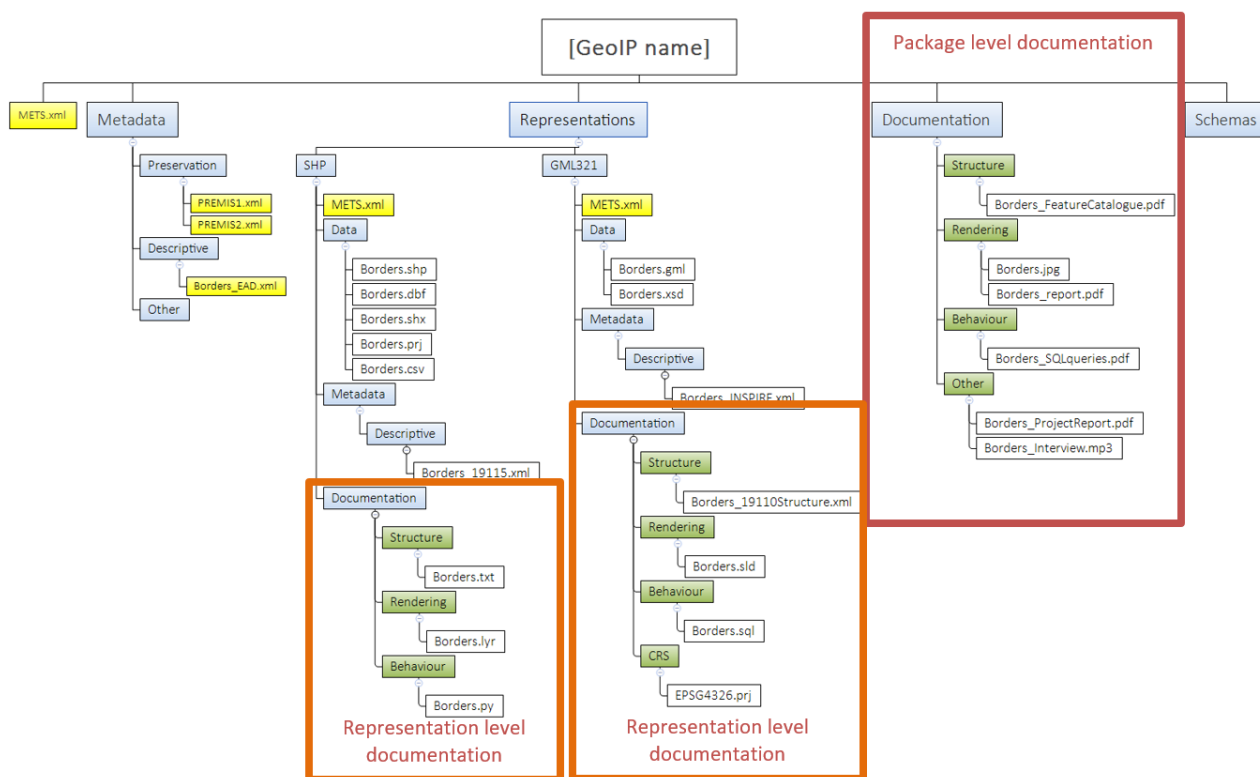


Figure 9 - Locations of Documentation folders in archival package.

Rationale:

Package-level documentation that covers information about all representations is placed on a higher level.

GEO_25 RationaleRequirement:

GEO_25	Representation level documentation	Technical documentation specific to one representation SHOULD be placed in <code>representations/[RepresentationName]/documentation</code>	0..n SHOULD
--------	------------------------------------	---	------------------------------

Description:

This requirement states where the representation-level documentation is to be placed.

Example:

See image in the GEO_24.

Rationale:

Sometimes documentation is specific to the content within the Representation. So, when we need to be more specific, we place the documentation to the Representation level. It is also intended for the standardised machine-readable content to support automated access and record dissemination.

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_26 Rationale

Requirement:

GEO_26	Feature Catalogue documentation	A document containing definitions and descriptions of feature types and feature attribute values SHOULD be provided for all geospatial records in the Information Package	0..n SHOULD
--------	---------------------------------	---	--------------------

Description:

This general requirement states that some kind of feature catalogue must accompany geospatial data in the Information Package. A Geospatial Feature catalogue contains definitions and descriptions of the feature types, feature attributes and feature relationships (feature inheritances and feature associations) occurring in one or more sets of geographic data, together with any feature operations that can be applied. [SOURCE: ISO 19101-1:2014, 4.1.13]. ISO 19110:2016 describes Feature Catalogues in greater detail.¹²

Feature Catalogues can also come in many forms:

- Standardised machine-readable xml file, based on ISO or OGC standards
- A proprietary well documented machine-readable feature catalogue
- A descriptive document that lists and explains all the definitions, descriptions and relations between features.

A Standardised machine-readable xml file is preferred, as it would support future access automation and be stored on the representation level. However, if that form is not available, a descriptive document needs to be available in the package-level documentation.

Example:

See example in the requirement GEO_28

Rationale:

A feature catalogue describes the structure of geospatial datasets and their relationships with other data. This information is essential for future reuse of the data.

¹²ISO 19110:2016 Geographic information — Methodology for feature cataloguing (<https://www.iso.org/standard/57303.html>)

GEO_27 Rationale

Requirement:

GEO_27 ISO 19110 ISO 19115-3	Standardised machine-readable Feature Catalogue	A standardised machine-readable feature catalogue SHOULD be provided in the Information Package	0..n SHOULD
GEO_27a Ref. GEO_27	Placement of Standardised machine-readable Feature Catalogue	If a standardised machine-readable feature catalogue exists it SHOULD be placed in representations/[RepresentationName]/documentation/structure	0..n SHOULD

Description:

This requirement provides an option to fulfil the requirement GEO_26 providing a standardised machine-readable feature catalogue in a subfolder named structure in the `documentation` folder at the representation level of the Information Package. Ideally, it should at least be available in the long-term preservation representation. Additional, **proprietary** machine-readable feature catalogue files should be placed in additional representations containing non-long-term preservation formats.

Taxonomies, Ontologies and/or vocabularies, that contain information relevant to Feature catalogues and that were used with geospatial data in standardized formats (OWL, SKOS, RDF...) can also be added to this folder.

Example:

An example would be an XML file based on ISO 19110:2016 or earlier.

Rationale:

A standardised machine-readable feature catalogue is commonly well-documented and gives us the possibility of future dissemination automation.

GEO_28 Rationale

Requirement:

GEO_28 Ref. GEO_27	Documentation containing Feature Catalogue Descriptions	Documentation, describing elements of a feature catalogue, not compliant with GEO_27 (a non-standardised machine-readable feature catalogue) SHOULD be provided in one of the Documentation folders of the Information Package	0..n SHOULD
--------------------------	---	---	--------------------

Description:

This requirement provides an option to fulfil the requirement GEO_26. When a standardised machine-readable feature catalogue is not available or as an addition to it. A Feature Catalogue documentation should contain a collection of metadata that provides the semantics and the structure of the objects stored in geospatial record(s). A feature catalogue should contain the names and definitions of feature types, names and definitions of their properties which include feature attributes, geometry (shapes and specifications, datum, map projection, etc.), temporal aspect (dimensions and specifications, datum, units, resolutions, etc.), operations, and roles, descriptions of attribute values and domains, relationships, constraints,...

It is recommended that the document describing all representations is placed in a subfolder named structure in the documentation folder at the package level or representation level of the Information Package, depending on the context. Other placements are also possible, depending on the context (to better reflect the producers organisational structure or if a document contains documentation on multiple technical elements, etc.)

Example:

Below an example of a non-standardised machine-readable feature catalogue.

Spatial Object Type: Duct

Duct	
Title:	duct
Definition:	A utility link or link sequence used to protect and guide cable and pipes via an encasing construction.
Description:	<p>A Duct (or Conduit, or Duct-bank, or Wireway) is a linear object which belongs to the structural network. It is the outermost casing. A Duct may contain Pipe(s), Cable(s) or other Duct(s).</p> <p>Duct is a concrete feature class that contains information about the position and characteristics of ducts as seen from a manhole, vault, or a cross-section of a trench and duct.</p>
Subtype of:	UtilityLinkSet
Type:	Spatial Object Type
Attribute:	
Name:	ductWidth
Title:	duct width
Definition:	The width of the duct.
Description:	The measurement of the object – in this case, the duct - from side to side.
Voidable:	true
Multiplicity:	1
Value type:	Length
Association role:	
Name:	ducts
Definition:	A single duct or set of ducts that constitute the inner-duct.
Voidable:	true
Multiplicity:	0..*
Value type:	Duct (spatial object type)
Association role:	
Name:	cables
Definition:	A duct may contain one or more cables.
Voidable:	true
Multiplicity:	0..*
Value type:	Cable (spatial object type)
Association role:	
Name:	pipes
Definition:	The set of pipes that constitute the duct bank.
Voidable:	true

Multiplicity:	0..*
Value type:	Pipe (spatial object type)

More examples can be found on Feature Catalogue “INSPIRE Application Schemas.”¹³

Rationale:

Written documentation containing the feature catalogue is a less favourable option since automated access is not supported. However, it would still document geospatial records well enough for manual interpretation.

GEO_29 Rationale

Requirement:

GEO_29	Logical model	A document describing relationships between multiple geospatial entities or geospatial and non-spatial records SHOULD be provided in the Information Package	0..n SHOULD
GEO_29a Ref GEO_29	Placement of logical model	If a document describing the logical model exists, it SHOULD be provided in a documentation/structure folder	0..1 SHOULD
GEO_29b Ref GEO_29	Placement of machine-readable logical model	If a standardised machine-readable version of a document describing the logical model exists, it SHOULD be provided in representations/[RepresentationName]/documentation/structure	0..1 SHOULD

Description:

This requirement is applicable in cases when the Information Package contains multiple geospatial datasets with defined relationships. A logical model is a visual representation of the Feature Catalogue and can be presented as a UML Diagram or in a standardised machine-readable format that can be imported into a database.

Ontologies, Taxonomies and/or vocabularies, that contain information describing the logical model and were used with geospatial in standardized formats (OWL, SKOS, RDF..) can also be added to this folder

¹³ Feature Catalogue “INSPIRE Application Schemas” (<https://inspire.ec.europa.eu/data-model/approved/r4618-ir/fc/>)

Example:

Below is an example of a logical model describing the relationship between multiple entities in a geospatial record_

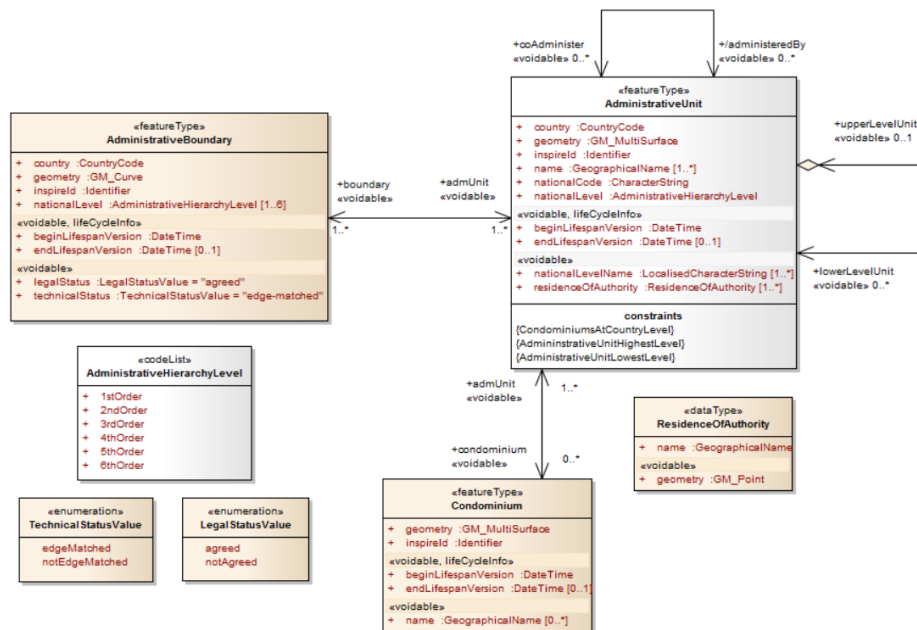


Figure 10 - example of a logical model describing the relationship between multiple entities in a geospatial record

More examples can be found in the INSPIRE Knowledge Base > Data Specifications >Data Models¹⁴.

An example of a standardised machine-readable format for logical models is the XML Metadata Interchange (XMI)¹⁵. This format is defined by the OMG and based on XML and is an open standard file format that enables the interchange of model information between models and tools. Other possible formats are formatted CSV, RDF and others.

Rationale:

This structure of one or many geospatial records in a dataset should be described in a machine-readable way to enable automated validation and use of the files.

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.

¹⁴ INSPIRE Knowledge Base > Data Specifications >Data Models (<https://inspire.ec.europa.eu/data-model/approved/r4618-ir/fc/>)

¹⁵ <https://www.omg.org/spec/XMI/2.5.1/About-XMI/>

Example:

The OGC OWS Context standard can be used for storing information on the configuration and rendering styles of used data and its references. See more on this in the Guideline for GIS.

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.

Rendering and visualisation documentation provides a way to replicate the rendering of geospatial records in future GIS. The catalogue of cartographic symbols is a collection of agreed cartographic symbols, which are used via the visualisation of geospatial datasets to display objects in space. As shown in Figure 11, Cartographic symbols are shown in the legend, which explains their meaning.

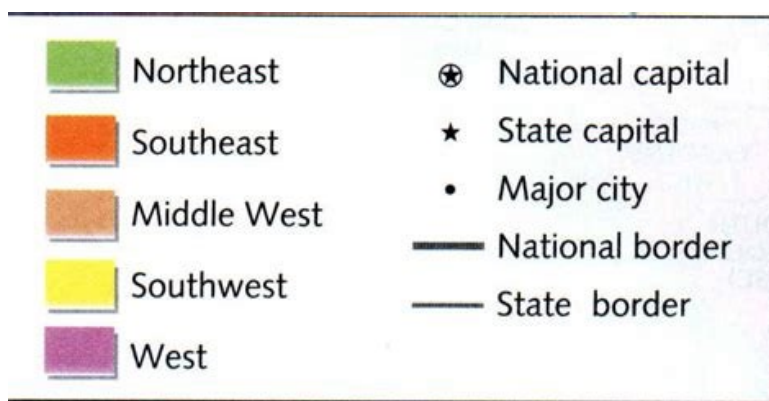


Figure 11 - Legend with cartographic symbols

For specific geospatial data, the visualisation is already embedded into the product by the producer in the form of (geo-located) raster images or scanned paper maps. In these cases, it is reasonable to archive that kind of visualisation. For each geospatial dataset, 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 standardised machine-readable files, they should be preserved.

GEO_31 Rationale

Requirement:

GEO_31	Geospatial dataset visualisation	An image displaying the overall view or a representative section of any geospatial dataset SHOULD be provided in the Information Package and placed in a documentation/rendering folder	0..n SHOULD
--------	----------------------------------	--	-----------------------

Description:

This requirement states that there should be an image representing a visual overview of a geospatial dataset. The image can depict the full extent of the dataset or a representative section, whichever is more informative. Images are to be placed into the `documentation/rendering` folder on the Information Package level. If there are differences between geospatial records within various representations, the images can also be placed within the `representations/[RepresentationName]/documentation/rendering` folder.

If the same record is split into multiple datasets that hold the same type of content, one image is enough. For example, if we have elevation iso lines for different locations, we can use one image to represent them all, even if the data is stored in multiple datasets. If producers use finding aids for their geospatial records, they usually already have the images available.

Example:

The image in Figure 12 shows an example of the usage of images representing geospatial records within a Geospatial Metadata catalogue.

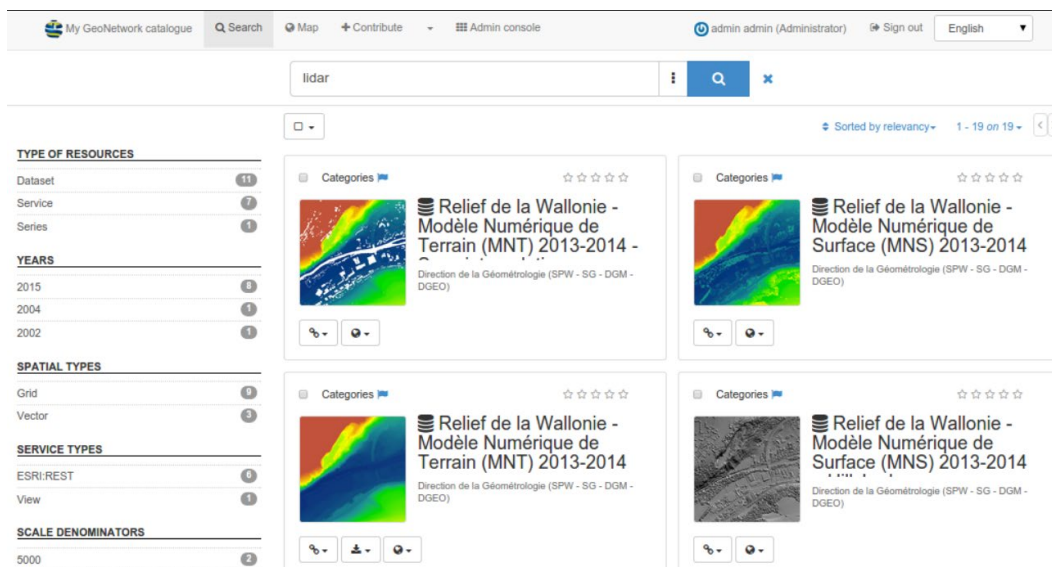


Figure 12 - Usage of images in Geospatial Metadata Catalogue (www.geonetwork-opensource.org)

Rationale:

The purpose of this requirement is to provide an image, that helps us find the geospatial dataset quicker and make it more accessible in the finding aid. The benefit is that images can be used in archival finding aids like archival catalogues or Geospatial Metadata Catalogues without rendering in GIS systems. This gives future users of the preserved geospatial data an easy way to quickly identify the content of the Information Package.

GEO_32 Rationale

Requirement:

GEO_32	Visualisation documentation	A document describing visualisation rules and configurations SHOULD be provided in the Information Package	0..n SHOULD
GEO_32a Ref GEO_32	Placement of visualisation documentation	If a document describing visualisation rules and configurations exists, it SHOULD be provided in a documentation/rendering folder	0..1 SHOULD

Description:

This requirement applies to datasets where geospatial data is rendered using cartographic means depending on the data format.

Standardised machine-readable files should be preferably placed within the `documentation/rendering` folder on the representation level (in the representation containing Long-term preservation formats). This rendering information is often dependent on the system (Desktop tool, Web Application, Mobile device, etc.) used to render the geospatial data. An example of rendering configuration information in QGIS is shown in Figure 13. In QGIS, for example, such rendering information can be stored in a machine-readable *.qml file. All other rendering and visualisation documentation is to be in the `documentation/rendering` folder on the package level of the Information Package package level. If there are differences between geospatial records within various representations, the images or configuration files can also be placed within the `representations/[RepresentationName]/documentation/rendering` folder.

If there is no standardised machine-readable files, then human-readable documentation should be available that contains information about visualisation rules and configurations of geospatial data forming an Information product.

Some of the common visualisation and rendering information used in most GIS tools are listed below.

- **Legend** of a map (e.g. symbology for roads, railways, vegetation etc.)
- **Layer symbology definition** (type of symbol and colouring, classification rules, visibility scale, labelling, definition query)
- **Scale** of map (e.g. 1:50.000)
- **Image value representation type**. How are raster pixel values represented (RGB, greyscale, RGB-NIR, CMYK, NDVI, custom colour map)?
- **Colour map**. How is the colour map documented? How are raster pixel values represented (continuously, classified)? Are any additional classification tables present for the colour map?
- **Rendering algorithm**. Is an algorithm used to render the values (Histogram stretch, Gamma stretch, Statistics based display, Unique values)
- **Raster Pyramids**. Are raster pyramids calculated to display the raster object, and if so, with what parameters.

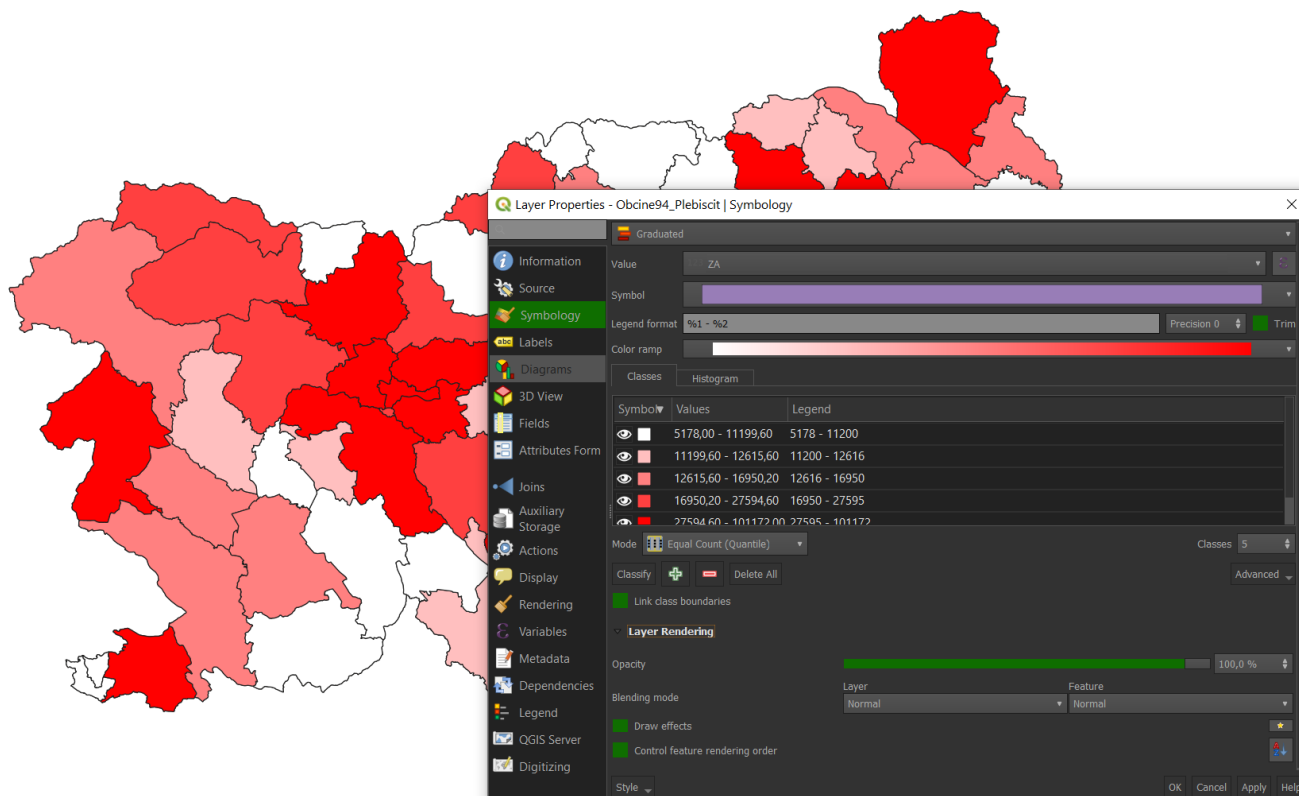


Figure 13 - Vector layer visualization and rendering configuration in QGIS

Examples:

Rendering and Visualisation information can be documented using **standardised machine-readable files**, human-readable documentation defining the cartographic keys and rendering criteria or by examples of information products (maps, lists, screenshots or videos from the original system).

Rationale:

To properly understand, interpret and reproduce information products based on geospatial data, we need to know what methodology was used to render the data. If the rendering tool's functionality is fully documented, an exact reproduction of the user experience can be recreated. However, since this is often not the case, Archivists then decide which elements need to be preserved.

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 an open well-documented machine-readable format to support the automatised dissemination. According to the archiving policies of an organisation, adequate formats need to be selected. For instance, if an organisation is preserving its data for a medium length, like 10-20 years, it can take a more liberal approach. However national archives usually have the strictest requirements. If the producer cannot provide the archive with an open well-documented symbology configuration, it can be recreated from the description provided in the *Documentation* in an open-source GIS application like QGIS¹⁶ or even using a LLM like ChatGPT v.4¹⁷.

Example:

An example of Standardised machine-readable formats for the rendering of geospatial records are SLD¹⁸ or KML¹⁹ files that also have some of that capability. It must be noted, however, that the SLD standard is rather limited. Therefore, many cartographic styling details cannot be adequately represented in SLD, as opposed to, for example, QML²⁰ (from QGIS) and LYR²¹ (ESRI layer file) are much more powerful.

SLD files example: SLD (Styled Layer Description) is an XML-based OGC (Open Geospatial Consortium) standard for symbology of web services. 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 input for easier DIP creation in the future. An example of an SLD file is shown below.

```
<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>
```

Rationale:

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

¹⁶ QGIS (<https://qgis.org/>)

¹⁷ OpenAI ChatGPT- <https://chat.openai.com/>

¹⁸ SLD – Styled Layer Description (https://portal.opengeospatial.org/files/?artifact_id=22364)

¹⁹ KML Standard (<http://www.opengeospatial.org/>)

²⁰ QGIS QML Style file (https://docs.qgis.org/3.34/en/docs/user_manual/appendices/qgis_file_formats.html#qml-the-qgis-style-file-format)

²¹ esri LYR style file (<https://www.loc.gov/preservation/digital/formats/fdd/fdd000626.shtml>)

GEO_34 Rationale

Requirement:

GEO_34	Information product examples	Information product examples based on geospatial record or records example SHOULD be provided in the Information Package and placed in the package level in the Information Package	0..n SHOULD
GEO_34a Ref GEO_34	Placement of information product examples	If information product examples exist, they SHOULD be provided in the Information Package in a documentation/rendering folder	0..1 SHOULD

Description:

This requirement recommends that examples of information products originally derived from the preserved geospatial records in the Information Package are provided in a documentation folder in the Information Package.

Example:

Geographic Information Systems support the creation of different types of Information Products:

- A pop-up window within the application that provides an information
- A digital map or a series of maps that can be printed on physical media or to a digital file (PDF, eps, various image formats)
- A set of lists or reports that are a result of a spatial query that have no or minor graphical components. (i.e. House numbers connected to a water line in a particular area.)
- A flythrough video of a 3D landscape
- A new geospatial record based on use of geospatial algorithms (A new area based on a Buffer function, the terrain elevation model could be a derivate from a LIDAR point cloud, etc.). However, we would not recommend to store these types of information products in the documentation, but rather as additional data within the Data folder or even as a separate Information Package.

Rationale:

To enable reproduction of information products derived from preserved geospatial records in the Information Package, which are similar to the information products data creators used to produce in their original systems.

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	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 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, 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 SHOULD 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_37 Ref GEO_37a	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, 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 any type of source code (scripts, SQL queries, GeoSPARQL queries, Python code, C++ code, etc.) that were used to produce an information product using 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 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. EPSG²² 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.

Example:

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

²² EPSG (<https://epsg.org/home.html>)

```

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]]

```

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 MUST 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 14 - Standardised Machine-readable transformation between MGI 1901 CRS and Slovenia 1996 CRS in WKT2 format

GeoRepository Login / Register

EPSG Home | EPSG Dataset | Support Documentation | About Us | Contact | IOGP Geomatics

Coordinate Transformation: MGI 1901 to Slovenia 1996 (1)

WKT GML Print

Transformation Details [SUPERSEDED]

NAME: MGI 1901 to Slovenia 1996 (1)
 CODE: 3916
 VERSION: GuRS-Svn
 VARIANT: 1
 ACCURACY (M): 1
 USAGE:

Usage Details

SCOPE: Transformation of coordinates at 1m level of accuracy.
 EXTENT: [Slovenia - onshore](#)

SOURCE CRS: [MGI 1901](#)
 TARGET CRS: [Slovenia 1996](#)
 TRANSFORMATION METHOD: [Coordinate Frame rotation \(geog2D domain\)](#)
 METHOD IS REVERSIBLE: YES

TRANSFORMATION PARAMETERS:	Parameter	Value	Reversible	Unit
	X-axis translation	409.545	YES	metre
	Y-axis translation	72.164	YES	metre
	Z-axis translation	486.872	YES	metre
	X-axis rotation	-3.085957	YES	arc-second
	Y-axis rotation	-5.46911	YES	arc-second
	Z-axis rotation	11.020289	YES	arc-second
	Scale difference	17.919665	YES	parts per million

META DATA

REMARKS: 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).
 INFORMATION SOURCE: Geodetska uprava Republike Slovenije, <http://www.gu.gov.si>
 DATA SOURCE: EPSG
 REVISION DATE: 29 August 2018

Figure 15 - A human document describing CRS transformation from MGI 1901 to Slovenia 1996 CRS

Rationale:

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

Rationales in 3.4.5 Other - Contextual Documentation requirements

CITS Geospatial text:

This part of the IP describes all remaining, more general information about the geospatial record. Included here are links to relevant Documentation describing data creation methodology and the spatial data set's provenance. The Documentation could consist of interviews, legal origin documentation, related practices in the EU and worldwide, methodological rules, scientific articles, related publications, etc.

Description:

Besides the documentation described in previous chapters (from 3.4.1 to 3.4.4.), which mainly address elements describing the structure, rendering and behaviour, we also require contextual documentation. Contextual documentation is often in a non-machine-readable format like pdf or TIFF and recently also in audio and video formats.

Examples:

Detailed descriptions of lineage information

- *Acquisition information* describing methods and tools used for creating the geospatial records, like camera used, flight path, scanner used, digitisation accuracy, GPS...
- *Processing information* describing algorithms used and processing performed to produce the data and descriptions of environment procedures such as software and parameters by which the algorithm is applied to generate the data from the source data
- *Source data* information describing the original source data a submitted geospatial dataset (product) is derived from such as data format, CRS and/or storage
- Spatial accuracy of the geospatial records
- Temporal accuracy of the geospatial records

Detailed descriptions of the use of the system used to produce, manage and/or view the geospatial records

- User manuals and other documentation and screenshots describing the system-user dialogue in the original system used for production and/or use of the geospatial records
- Videos or screen captures of the system as seen from the user's point of view
- Interviews with producers and/or users of the geospatial records
- Documentation of finding tool and screenshots of the metadata search and presentation related to the geospatial records

Detailed descriptions of the structure in the information Package of the geospatial records (if this is not documented otherwise using naming, grouping of files in folders or METS.xml structMap elements)

- Relation in the IP between a data file and the metadata file describing
- Relation in the IP between a raster object and a world file
- Relation in the IP between a schema file and the gml file that it validates
- Relation in the IP between raster objects and tiling indexes
- Relation in the IP between a geospatial coverage file and the geospatial objects it covers
- Relation in the IP between geospatial records and additional attributes in another file like a feature catalogue or database file

Detailed descriptions of the broader historical context of the geospatial records

- Provenance
- Description of the purpose of geospatial dataset
- Description of administrative use of the geospatial dataset
- Methodology used

- Regulations and legal context
- Scientific articles about the geospatial records
- Publications

Detailed descriptions of preservation actions

- Documentation of migration of geospatial records to preservation format/Information package
- Migration or dissemination information can be documented in a PREMIS file and placed in a metadata preservation folder.

Rationale:

Contextual documentation of the geospatial data in the Information package is important to be able to understand and use the archived data correctly.

GEO_40 Rationale

Requirement:

GEO_40	Package level contextual documentation	Contextual documentation covering all representations in the Information package SHOULD be placed in documentation/other on package level	0..n SHOULD
--------	--	--	-----------------------

Description:

It is recommended that contextual documentation relevant for data in several representations in the Information Package is placed at the package level. However, if the documentation only covers two out of three representations in the IP, a copy of the documentation in each of the two representations at the representation level is a better option.

Example:

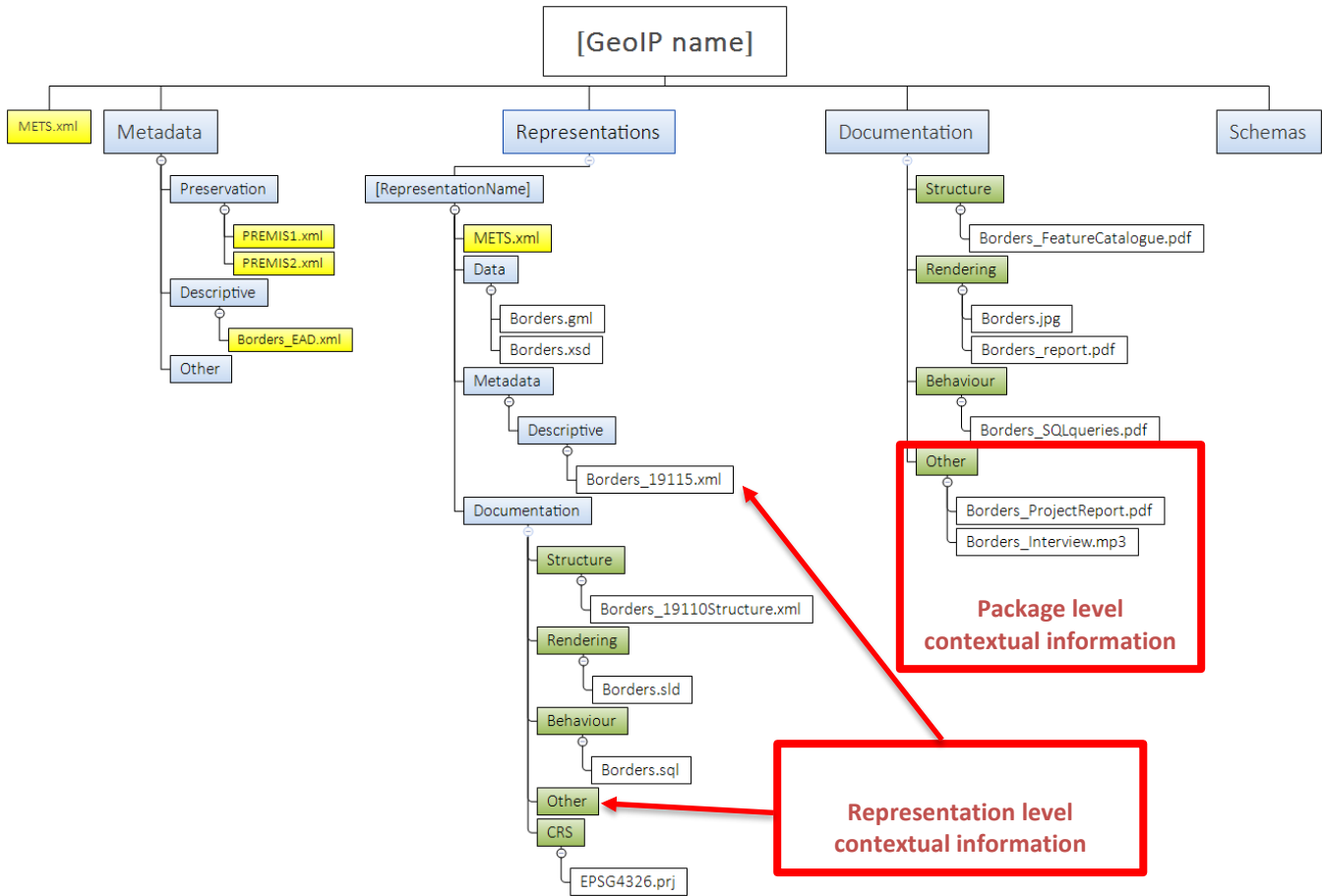


Figure 16 - Location of Package and Representation level Contextual documentation

Rationale:

The purpose of correct placement of contextual documentation in the IP is to enable easy identification of relevant documentation when using data from a representation.

GEO_41 Rationale

Requirement:

GEO_41	Representation level contextual documentation	Contextual documentation covering only content within a particular representation SHOULD be placed in representations/[RepresentationName]/documentation/other	0..n SHOULD
--------	---	---	--------------------

Description:

It is recommended that contextual documentation relevant for only one representation in the Information Package is placed at the representation level in that particular representation. Context can also be described in standardised machine-readable Geospatial metadata that can be placed within the Metadata/descriptive folder on the representation level. Geospatial metadata is described in greater detail in chapter 3.5.

Example: See Example in GEO_40.

Rationale: The purpose of correct placement of contextual documentation in the IP is to enable easy identification of relevant documentation when using data from a representation.

3.5 Rationales in metadata requirements

Rationales in 3.5 Geospatial Metadata requirements

CITS Geospatial text:

Geospatial data in the IP is documented using a form of geospatial metadata, which contains common descriptions of the data as well as descriptions specific to the geospatial domain (accuracy, lineage, scale, measurement units, CRS info, etc.). In original systems, geospatial metadata can be stored in different ways (databases, standardised xml files, common documentation, etc.).

Description:

Digital Geospatial data can be described in different ways. A standardised and structured approach for digital geospatial records is most commonly documented using geospatial metadata, which contains common descriptions of the data as well as descriptions specific to the geospatial domain (accuracy, methodology of creation, reference to source data, scale, measurement units, CRS info, etc.). In original systems, geospatial metadata can be stored in different ways (databases, standardised xml files, txt based readme files, or even only within typical contextual documentation, etc.).

For CITS Geospatial we recommend, that the representation containing the long-term preservation formats holds geospatial metadata in a standardised machine-readable format. The most common standards used for geospatial metadata are:

- **ISO 19115:2003** Geographic information — Metadata
- **INSPIRE Metadata** Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119
- And the latest **ISO 19115-1:2014** Geographic information — Metadata — Part 1: Fundamentals
- **Data Catalogue Vocabulary (DCAT)** and its geospatial version GeoDCAT²³ metadata

Metadata can also be extended using additional ISO Standards like these:

- ISO 19165-1:2018 Geographic information -Preservation of digital data and metadata -Part 1: Fundamentals
- ISO 19115-2:2019 Geographic information — Metadata — Part 2: Extensions for acquisition and processing
- ISO 19157:2013 Geographic information — Data quality

Every metadata standard prescribes the mandatory metadata elements, which are in some cases limited only to essential Dublin Core elements. However, for long-term preservation purposes, we strongly suggest using the set of mandatory elements as described within the INSPIRE Metadata Implementing rules.

Example:

Table 1 below covers proposed mandatory elements required in the INSPIRE Metadata Implementation rules (see INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119²⁴). The INSPIRE metadata rules are also based on ISO 15836 (Dublin Core). Information that cannot be covered by these elements should be described in the accompanying documentation.

Some of the technical guidance recommendations and requirements from the INSPIRE Metadata Implementing Rules are added as descriptions in table 1 below. Obligation states whether the element is mandatory (M), optional (O) or conditional (C). The table also compares the obligation of an element in INSPIRE with the obligation of the element in

²³W3C DCAT – Geospatial dataset property - https://www.w3.org/TR/vocab-dcat-2/#Property:dataset_spatial

²⁴ INSPIRE Metadata Implementing Rules (<https://inspire.ec.europa.eu/documents/inspire-metadata-implementing-rules-technical-guidelines-based-en-iso-19115-and-en-iso-1>)

ISO 19115-1. Each obligation of an INSPIRE element is compared to the obligation of the similar ISO 19115 element in column 2 in table 1.

See the standards INSPIRE Implementing Rules for Metadata and EN ISO 19115-1 for further guidance on how to create an xml file with geospatial metadata that conforms to these rules. The INSPIRE Implementing Rules for Metadata also has examples of XML encoding.

Appendix 3 provide mapping schemas between INSPIRE metadata elements and the archival descriptive metadata standards EAD3 and ISAD(G).

Table 1 - Proposed mandatory metadata elements for an INSPIRE metadata file

No	ISO 19115 Core Element name	INSPIRE Element name	Description of the content of the element	Obligation/ occurrence
1	Dataset title (M)	Resource title (M)	<p><i>Name by which the cited resource is known</i></p> <p>The <i>Resource title</i> is the name of the data set and has to be concise and to the point. It should not contain unexplained acronyms or abbreviations. It is recommended a maximum length of 250 characters and keeping the similarity with the original title of the resource, in the sense of the ‘official naming’.</p> <p>Properties in the xml-file (ISO/TS 19139 path): <code>identificationInfo[1]/*/citation/*/*title</code></p> <p>Elements in the ISO 19115-1 (2014) standard: <code>MD_Metadata.identificationInfo > MD_DataIdentification.citation > CI_Citation.title)</code></p>	M/1
2	Abstract describing the dataset (M)	Resource abstract (M)	<p><i>Brief narrative summary of the content of the resource(s)</i></p> <p>The <i>Resource abstract</i> is a brief narrative summary of the content of the data set. The abstract provides a clear and concise statement that enables the user to understand the content of the data. Do not use unexplained acronyms or abbreviations.</p> <p>Properties in the xml file (ISO/TS 19139 path): <code>identificationInfo[1]/*/abstract</code></p>	M/1

			<p>Elements in the ISO 19115-1 (2014) standard: <code>MD_Metadata.identificationInfo > MD_DataIdentification.abstract</code></p>	
3	- (O)	Resource type	<p><i>Scope to which metadata applies.</i></p> <p>The <i>Resource type</i> is a code identifying the type of resource, e.g., dataset (see MD_ScopeCode), which the metadata describes. 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.</p> <p>Properties in the xml file (ISO/TS 19139 path): <code>hierarchyLevel > Data type:MD_ScopeCode</code></p> <p>The values of MD_ScopeCode in the scope of the INSPIRE Directive are:</p> <ul style="list-style-type: none"> - dataset for spatial datasets; - series for spatial dataset series; - service for spatial data services <p>Elements in the ISO 19115-1 (2014) standard: <code>MD_Metadata/metadataScope/ MD_MetadataScope/resource- Scope</code></p>	M/1
4	Resource identifier (O)	Unique resource identifier (M)	<p><i>Value uniquely identifying an object within a namespace</i></p> <p>Properties in the xml file (ISO/TS 19139 path): <code>identificationInfo[1]/*/citation/ */identifier</code></p> <p>Elements in the ISO 19115-1 (2014) standard: <code>MD_Metadata.identificationInfo>MD _DataIdentification.citation > CI_Citation.identifier>MD_Identifier</code></p>	M/1..n
5	Dataset language (M)	Resource language (C)	<i>Language(s) used within the datasets</i>	M/1..n

			<p>The <i>Resource language</i> is a three-letter code for the language taken from the vocabulary LanguageCode (ISO/TS 19139) based on alpha-3 codes of ISO 639-2. Use only three-letter codes from in ISO 639-2/B (bibliographic codes).</p> <p>The list of all the codes is defined at http://www.loc.gov/standards/iso639-2/ Regional languages also are included in this list.</p> <p>Properties in the xml file (ISO/TS 19139 path): <code>identificationInfo[1]/*/language</code></p> <p>Elements in the ISO 19115-1 (2014) standard: <code>MD_Metadata.identificationInfo></code> <code>MD_DataIdentification.defaultLocale > PT_Locale</code></p>	
6	Dataset topic category (M)	Topic category (M)	<p><i>Main theme(s) of the dataset.</i></p> <p>The <i>Topic category</i> describes the topic of the dataset and is a selection of the 20 elements in the MD_TopicCategory enumeration MD_TopicCategoryCode . The list below is from B.3.30 MD_TopicCategoryCode << Enumeration>> in the ISO 19115-1 (2014) standard, and the listed examples are not exhaustive:</p> <ul style="list-style-type: none"> ● farming ● biota ● boundaries ● climatologyMeteorologyAtmosphere ● economy ● elevation ● environment ● geoscientificInformation ● health ● imageryBaseMapsEarthCover ● intelligenceMilitary ● inlandWaters ● location ● oceans ● planningCadastre ● society ● structure ● transportation 	M/1..n

			<ul style="list-style-type: none"> • utilitiesCommunication • extraTerrestrial • disaster <p>This topic information is a high-level classification scheme to assist in the grouping and topic-based search of available geospatial data resources.</p> <p>Correct categorisation is very important to help users to search and find the resources they are looking for.</p> <p>Properties in the xml file (ISO/TS 19139 path): <code>identificationInfo[1]/*/topicCategory</code></p> <p>Elements in the ISO 19115-1 (2014) standard: <code>MD_Metadata.identificationInfo > MD_DataIdentification.topicCategory > MD_TopicCategoryCode</code></p>	
7	Keywords (O)	Keyword > Keyword value (M)	<p><i>Commonly used word(s) or formalised word(s) or phrase(s) used to describe the subject</i></p> <p>The INSPIRE Metadata Regulation 1205/2008/EC mandate the presence of at least one keyword to describe the dataset.</p> <p>An INSPIRE Keyword is defined by:</p> <ul style="list-style-type: none"> • a keyword value (see 2.4.1), which in ISO standard is referred to as “Keyword”; • an optional originating controlled vocabulary (see 2.4.2), which in ISO standard is referred to as “Thesaurus”. • It is possible to add as many keywords as relevant to the resource. <p>If only one keyword is used, then for spatial dataset or spatial dataset series, the keyword:</p> <ul style="list-style-type: none"> • shall describe the relevant INSPIRE Spatial Data Theme (as defined in Annex I, II and III of the INSPIRE Directive) • shall be expressed in the language of the metadata for the 34 INSPIRE Spatial Data Themes (please use the terms in each of the official languages in which the INSPIRE Directive has been translated) or neutral language values such as a URI. 	M/1..n

			<p>The keyword value is a commonly used word, formalised word or phrase used to describe the subject. While the topic category is too coarse for detailed queries, keywords help to narrow a full-text search, and they allow for structured keyword search</p> <p>Properties in the xml file (ISO/TS 19139 path): <code>identificationInfo[1]/*/descriptiveKeywords/*/keyword</code></p> <p>Elements in the ISO 19115-1 (2014) standard: <code>MD_Metadata.identificationInfo</code> <code>>MD_DataIdentification></code> <code>descriptiveKeywords</code> <code>>MD_Keywords</code></p>	
8	Keywords (O)	Keyword > Originating controlled vocabulary (C)	<p><i>Name of the formally registered thesaurus or a similar authoritative source of keywords</i></p> <p>Mandatory if the keyword value originates from a controlled vocabulary (Conditional).</p> <p>If the keyword value originates from a controlled vocabulary (thesaurus, ontology), for example, GEMET - Concepts, the citation of the originating controlled vocabulary shall be provided.</p> <p>The thesaurusName identification shall include at least the title and a reference date (date of publication, date of last revision or of creation) of the originating controlled vocabulary.</p> <p>Properties in the xml file (ISO/TS 19139 path): <code>identificationInfo[1]/*/descriptiveKeywords/*/thesaurusName</code></p> <p>Elements in the ISO 19115-1 (2014) standard:</p> <p>The controlled vocabulary is described through the thesaurusName property of the instance of descriptiveKeywords to which the keyword pertains:</p> <p>Example descriptiveKeywords: (MD_Keywords) keyword: BOUNDARIES Administrative</p>	C/0..n

			<p>keyword: INDUSTRY Mining Exploration keyword: MINERALS Exploration thesaurusName: (CI_Citation) title: ANZLIC Search Words date: (CI_Date) date: 1996-07 dateType: (CI_DateTypeCode) publication</p>	
9	Geographic location of the dataset (C)	Geographic bounding box (M)	<p><i>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)</i> <i>Northern-most coordinate of the limit of the dataset extent, expressed in latitude in decimal degrees (positive north)</i> <i>Southern-most coordinate of the limit of the dataset extent, expressed in latitude in decimal degrees (positive south)</i></p> <p>This is the extent of the resource (e.g. dataset) in the geographic space, given as a bounding box. Defining the coordinates of a rectangle representing the resource area on a map allows the discovery by geographical area.</p> <p>The bounding box shall be as small as possible.</p> <p>The bounding box shall be expressed in decimal degrees with a precision of at least two decimals. The coordinates of the bounding box are expressed in any geodetic coordinate reference system with a Greenwich Prime Meridian.</p> <p>Properties in the xml file (ISO/TS 19139 path): identificationInfo[1]/*/extent/*/geographicElement/*/westBoundLongitude</p> <p>identificationInfo[1]/*/extent/*/geographicElement/*/eastBoundLongitude</p> <p>identificationInfo[1]/*/extent/*/geographicElement/*/southBoundLatitude</p>	M/1..n

			<p>identificationInfo[1]/*/extent/*/geographicElement/*/northBoundLatitude</p> <p>Elements in the ISO 19115-1 (2014) standard: EX_GeographicBoundingBox westBoundLongitude eastBoundLongitude southBoundLatitude northBoundLatitude</p>	
10	<p>Additional extent information for the dataset (vertical and temporal) (O)</p>	<p>Temporal extent (M)</p>	<p><i>The time period covered by the content of the dataset</i></p> <p>The temporal extent defines the time period covered by the content of the resource. This time period may be expressed as:</p> <ul style="list-style-type: none"> • an individual date • an interval of dates (starting date and ending date) • a mix of individual dates and intervals of dates <p>The INSPIRE Metadata Regulation 1205/2008/EC requires at least one temporal reference chosen from one of these four categories:</p> <ul style="list-style-type: none"> • temporal extent • date of publication • date of last revision • date of creation <p>To be compliant with ISO 19115 it is necessary to use at least one among the date of publication, date of last revision, or the date of creation.</p> <p>The default reference system shall be the Gregorian calendar, with dates expressed in accordance with ISO 8601 (yyyy-mm-dd where yyyy is the year, mm is the month and dd is the day). Example: From 2008-01-01T11:45:30 to 2008-12-31T09:10:00.</p> <p>Properties in the xml file (ISO/TS 19139 path): identificationInfo[1]/*/extent/*/temporalElement/*/extent</p> <p>Elements in the ISO 19115-1 (2014) standard:</p>	M/1..n

			<pre>MD_Metadata.identificationInfo > MD_Identification.extent > EX_ Extent > EX_TemporalExtent or EX_VerticalExtent</pre>	
11	<p>Additional extent information for the dataset (vertical and temporal) (O)</p>	<p>Date of publication</p>	<p><i>The reference date for the cited resource - publication</i></p> <p>This is the date of publication of the resource when available or the date of entry into force. There may be more than one date of publication. The date of publication differs from the temporal extent. For example, a dataset might have been published in March 2009 (2009-03-15), but the covered information was collected over the year 2008 (temporal extent from 2008-01-01 to 2008-12-31).</p> <p>The default reference system shall be the Gregorian calendar, with dates expressed in accordance with ISO 8601 (yyyy-mm-dd where yyyy is the year, mm is the month and dd is the day). Example: 2009-03-15T11:45:30 or 2009-03-15.</p> <p>Properties in the xml file (ISO/TS 19139 path): <code>identificationInfo[1]/*/citation/*/date[./*/dateType/*/text()='publication']/*/date</code></p> <p>Elements in the ISO 19115-1 (2014) standard: ... <code><gmd:identificationInfo></code> <code><gmd:MD_DataIdentification></code> <code><gmd:citation></code> <code><gmd:CI_Citation></code> ... <code><gmd:date></code> <code><gmd:CI_Date></code> <code><gmd:date></code> <code><gco:Date>2009-03-15</gco:Date></code> <code></gmd:date></code> <code><gmd:dateType></code> <code><gmd:CI_DateTypeCode</code> <code>codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codelist/ML_gmxCodeLists.xml#CI_DateTypeCod</code></p>	O/0..n

			<pre>e" codeListValue="publication">publication</g md:CI_DateTypeCode> </gmd:dateType> </gmd:CI_Date> </gmd:date> ... </gmd:CI_Citation> </gmd:citation> ... </gmd:MD_DataIdentification> ... </gmd:identificationInfo></pre>	
12	Additional extent information for the dataset (vertical and temporal) (O)	Date of last revision	<p><i>The reference date for the cited resource - revision</i></p> <p>This date describes when the resource was last revised, if the resource has been revised. The date of revision differs from the temporal extent. For example, a dataset might have been revised in April 2009 (2009-04-15), but the covered information was collected over the year 2008 (temporal extent from 2008-01-01 to 2008-12-31). Example: 2009-04-15 2009-04-15T11:15:00.</p> <p>Properties in the xml file (ISO/TS 19139 path): identificationInfo[1]/*/citation/*/ date[./*/dateType/*/ text()='revision']/*/ date</p> <p>Elements in the ISO 19115-1 (2014) standard:</p> <pre>... <gmd:identificationInfo> <gmd:MD_DataIdentification> <gmd:citation> <gmd:CI_Citation> ... <gmd:date> <gmd:CI_Date> <gmd:date> <gco:DateTime>2009-04-15T11:15:00</gco:DateTime> </gmd:date> <gmd:dateType> <gmd:CI_DateTypeCode</pre>	O/0..n

			<pre>codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codelist/ML_gmxCodeLists.xml#CI_DateTypeCode" codeListValue="revision">revision</gmd:CI_DateTypeCode> </gmd:dateType> </gmd:CI_Date> </gmd:date> ... </gmd:CI_Citation> </gmd:citation> ... </gmd:MD_DataIdentification> </gmd:identificationInfo></pre>	
13	Additional extent information for the dataset (vertical and temporal) (O)	Date of creation	<p><i>The reference date for the cited resource - creation</i></p> <p>This date describes when the resource was created. The date of creation differs from the temporal extent. For example, a dataset might have been created in February 2009 (2009-02-15), but the covered information was collected over the year 2008 (temporal extent from 2008-01-01 to 2008-12-31). Example: 2009-02-15 2009-02-15T11:15:00.</p> <p>Properties in the xml file (ISO/TS 19139 path): <code>identificationInfo[1]/*citation/*/date[./*/dateType/*/text()='creation']/*date</code></p> <p>Elements in the ISO 19115-1 (2014) standard:</p> <pre>... <gmd:identificationInfo> <gmd:MD_DataIdentification> <gmd:citation> <gmd:CI_Citation> ... <gmd:date> <gmd:CI_Date> <gmd:date> <gco>Date>2009-02-15</gco>Date> </gmd:date> <gmd:dateType> <gmd:CI_DateTypeCode codeList="http://standards.iso.org/ittf/PubliclyAv</pre>	O/0..n

			<pre> ailableStandards/ISO_19139_Schemas/resources/ codelist/ML_gmxCodelists.xml#CI_DateTypeCod e" codeListValue="creation">creation</gmd:CI_ DateTypeCode> </gmd:dateType> </gmd:CI_Date> </gmd:date> ... </gmd:CI_Citation> </gmd:citation> ... </gmd:MD_DataIdentification> ... </gmd:identificationInfo> </pre>	
14	Lineage (O)	Lineage (M)	<p><i>General explanation of the data producer’s knowledge about the lineage of a dataset.</i></p> <p>According to the INSPIRE Implementing Rules for Metadata, Lineage is “a statement on process history and/or overall quality of the spatial data set. Where appropriate, it may include a statement whether the data set has been validated or quality assured, whether it is the official version (if multiple versions exist), and whether it has legal validity. The value domain of this element is free text.”</p> <p>The process history may be described by information on the source data used and the main transformation steps that took place in creating the current data set (series).</p> <p>The use of acronyms should be avoided. If used, their meaning should be explained.</p> <p>This information can also reference other documents that cover this description in greater detail.</p> <p>Properties in the xml file (ISO/TS 19139 path): dataQualityInfo/*/lineage/*/state ment</p> <p>Elements in the ISO 19115-1 (2014) standard:</p>	M/1

			<pre>MD_Metadata >resourceLineage> LI_Lineage</pre>	
15	Spatial resolution of the dataset (O)	Spatial resolution (C)	<ul style="list-style-type: none"> • <i>Equivalent scale: level of detail expressed as the scale denominator of a comparable hardcopy map or chart</i> • <i>Distance: ground sample distance</i> <p>Spatial resolution refers to the level of detail of the data set. It shall be expressed as a set of zero to many resolution distances (typically for gridded data and imagery-derived products) or equivalent scales (typically for maps or map-derived products).</p> <p>An <i>equivalent scale</i> is generally expressed as a positive integer value expressing the scale denominator. Example: 50000 (e.g. 1:50000 scale map).</p> <p>A resolution <i>distance</i> shall be expressed as a numerical value associated with a unit of length. Example: 0.25 (degrees).</p> <p>Each spatial resolution is either an equivalent scale OR a ground sample distance. Each spatialResolution element must contain either an equivalent scale or a distance but not both.</p> <p>Properties in the xml file (ISO/TS 19139 path): <code>identificationInfo[1]/*/spatialResolution/*/equivalentScale/*/denominator</code> (equivalent scale)</p> <p><code>identificationInfo[1]/*/spatialResolution/*/distance</code> (distance)</p> <p>Elements in the ISO 19115-1 (2014) standard: <pre>MD_Metadata.identificationInfo > MD_Identification.spatialResolution > MD_Resolution.equivalentScale MD_Resolution.distance,</pre> </p>	C/0..n Mandatory if an equivalent scale or a resolution distance can be specified.

			<p>MD_Resolution.vertical, or MD_Resolution.angularDistance, or MD_Resolution.levelOfDetail</p>	
-		<p>Conformity > Degree (M)</p>	<p><i>Indication of the conformance result</i></p> <p>This is the degree of conformity of the resource to the implementing rules adopted under Article 7(1) of INSPIRE Directive 2007/2/EC or other specification.</p> <p>The values are:</p> <ul style="list-style-type: none"> ● true (if conformant) ● false (if not conformant) ● null (with nilReason = “unknown” if not evaluated) <p>Properties in the xml file (ISO/TS 19139 path): dataQualityInfo/*/report/*/result */pass</p> <p>Elements in the ISO 19115-1 (2014) standard: I cannot find the elements in ISO 19115-1:2014</p> <p>Elements in the ISO 19115-1 (2003) standard</p> <pre><gmd:result> <gmd:DQ_ConformanceResult> <gmd:specification> ... <!-- See below --> </gmd:specification> <!-- gmd:explanation is mandated by ISO 19115. A default value is proposed --> <gmd:explanation> <gco:CharacterString>See the referenced specification</gco:CharacterStrin g> </gmd:explanation> <!-- the value is false instead of true if not conformant --> <gmd:pass> <gco:Boolean>>true</gco:Boolean> </gmd:pass> </gmd:DQ_ConformanceResult> </gmd:result></pre>	M/1

16	-	Conformity > Specification (M)	<p><i>Citation of the product specification or user requirement against which data is being evaluated</i></p> <p>This is a citation of the implementing rules adopted under Article 7(1) of INSPIRE Directive 2007/2/EC or other specification to which a particular resource conforms.</p> <p>The following properties are expected:</p> <ul style="list-style-type: none"> ● Title (characterString and free text) ● Reference date (CI_Date): <ul style="list-style-type: none"> ○ dateType: creation, publication or revision date, e.g. publication ○ date: an effective date, e.g. 2010-12-08 <p>Properties in the xml file (ISO/TS 19139 path): dataQualityInfo/*/report/*/result */specification</p> <p>Elements in the ISO 19115-1 (2014) standard: I cannot find the elements in ISO 19115-1:20114</p> <p>Elements in the ISO 19115-1 (2003) standard</p> <pre> <gmd:dataQualityInfo> <gmd:DQ_DataQuality> ... <gmd:report> <gmd:DQ_DomainConsistency> <gmd:result> <gmd:DQ_ConformanceResult> <gmd:specification> <gmd:CI_Citation> <gmd:title> <gco:CharacterString>COMMISSION REGULATION (EU) No 1089/2010 of 23 November 2010 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards interoperability of spatial data sets and services</gco:CharacterString> </gmd:title> <gmd:date> <gmd:CI_Date> <gmd:date> <gco>Date>2010-12-08</gco>Date> </pre>	M/1
----	---	--	---	-----

			<pre> </gmd:date> <gmd:dateType> <gmd:CI_DateTypeCode codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codelist/ML_gmxCodelists.xml#CI_DateTypeCode" codeListValue="publication">publication</gmd:CI_DateTypeCode> </gmd:dateType> </gmd:CI_Date> </gmd:date> </gmd:CI_Citation> </gmd:specification> ... </gmd:DQ_DataQuality> </gmd:dataQualityInfo> </pre>	
17	<p>Constraints on resource access and use (O)</p>	<p>Limitations on public access (M)</p>	<p><i>Access constraints applied to assure the protection of privacy or intellectual property, and any special restrictions or limitations on obtaining the resource</i></p> <p>This metadata element shall provide information on the limitations and the reasons for them. If there are no limitations on public access, use the free text available in MD_LegalConstraints.otherConstraints to enter "No Limitations" in the language used for the metadata.</p> <p>Limitations on public access shall be represented by at least one of these metadata elements:</p> <ul style="list-style-type: none"> - MD_LegalConstraints.accessConstraints - MD_LegalConstraints.otherConstraints - MD_SecurityConstraints.classification <p>Access constraints: Access constraints applied to assure the protection of privacy or intellectual property and any special restrictions or limitations on obtaining the resource. Value is strictly limited to the value defined in codelist MD_RestrictionCode <<CodeList>> in ISO 19115-1:2014:</p> <ul style="list-style-type: none"> ● copyright ● patent ● patentPending 	M/1..n

			<ul style="list-style-type: none"> ● trademark ● licence ● intellectualPropertyRights ● restricted ● otherRestrictions ● unrestricted ● licenceUnrestricted ● licenceEndUser ● licenceDistributor ● private ● statutory ● confidential ● sensitiveButUnclassified ● in-confidence <p>Other constraints: Other restrictions and legal prerequisites for accessing and using the resource or metadata. Value is free text. Example: No limitations.</p> <p>Classification: Name of the handling restrictions on the resource. Value from codelist B.3.13 MD_ClassificationCode <<CodeList>> in ISO 19115-1:2014:</p> <ul style="list-style-type: none"> ● unclassified ● restricted ● confidential ● secret ● topSecret ● sensitiveButUnclassified ● forOfficialUseOnly ● protected ● limitedDistribution <p>Properties in the xml file (ISO/TS 19139 path):</p> <pre>identificationInfo[1]/*/resourceConstraints/*/accessConstraints</pre> <pre>identificationInfo[1]/*/resourceConstraints/*/otherConstraints</pre> <pre>identificationInfo[1]/*/resourceConstraints/*/classification</pre> <p>Elements in the ISO 19115-1 (2014) standard:</p> <pre>MD_Metadata.identificationInfo>MD_DataIdentification>MD_Constraints>MD_LegalConstraints and/or MD_SecurityConstraints</pre>	
--	--	--	--	--

<p>18</p>	<p>Constraints on resource access and use (O)</p>	<p>Conditions applying to access and use</p>	<p><i>Restrictions on the access and use of a resource or metadata</i></p> <p>The value is free text.</p> <p>If no conditions apply to the access and use of the resource, 'no conditions apply' shall be used. If conditions are unknown, 'conditions unknown' shall be used.</p> <p>Properties in the xml file (ISO/TS 19139 path): identificationInfo[1]/*/resourceConstraints/*/useLimitation</p> <p>Elements in the ISO 19115-1 (2014) standard: MD_Metadata.identificationInfo>MD_DataIdentification>MD_Constraints.useLimitations</p>	<p>M/1..n</p>
	<p>Resource point of contact (O)</p>	<p>Responsible organisation > Responsible party (M)</p>	<p><i>Identification of, and means of communication with, person(s) and organisation(s) associated with the resource(s)</i></p> <p>This is the description of the organisation responsible for the establishment, management, maintenance or distribution of the resource.</p> <p>This description shall include: name of the organisation and contact email address. The name of the organisation should be given in full, without abbreviations. It is recommended to use institutional email instead of personal emails.</p> <p>The following properties are expected:</p> <ul style="list-style-type: none"> ● organisationName (characterString and free text) ● contactInfo (CI_Contact): <ul style="list-style-type: none"> ○ address <ul style="list-style-type: none"> ■ electronicMailAddress [1..*] (characterString) <p>Properties in the xml file (ISO/TS 19139 path): identificationInfo[1]/*/pointOfContact</p> <p>Elements in the ISO 19115-1 (2014) standard: MD_Metadata.identificationInfo > MD_DataIdentification.pointOf-</p>	<p>M/1</p>

			Contact > CI_Responsibility	
	Resource point of contact (O)	Responsible organisation > Responsible party role (M)	<p><i>Function performed by the responsible party</i></p> <p>This is the role of the responsible organisation.</p> <p>Value from the codelist B.3.5 CI_RoleCode in ISO 19115-1:2014:</p> <ul style="list-style-type: none"> ● resourceProvider ● custodian ● owner ● user ● distributor ● originator ● pointOfContact ● principallInvestigator ● processor ● publisher ● author ● sponsor ● coAuthor ● collaborator ● editor ● mediator ● rightsHolder ● contributor ● funder ● stakeholder <p>There is a direct mapping between the responsible party roles defined in Part D 6 of the INSPIRE Metadata Regulation 1205/2008/EC and the values of the CI_RoleCode codelist of ISO 19115.</p> <p>Properties in the xml file (ISO/TS 19139 path): <code>identificationInfo[1]/*/pointOfContact/*/role</code></p> <p>Elements in the ISO 19115-1 (2014) standard: <code>MD_Metadata.identificationInfo> MD_DataIdentification.pointOfContact>CI_Responsibility>CI_RoleCode</code></p>	M/1

	<p>Metadata point of contact (M)</p>	<p>Metadata point of contact (M)</p>	<p><i>Party responsible for the metadata information</i></p> <p>This description shall include: name of the organisation and contact email address.</p> <p>The name of the organisation should be given in full, without abbreviations. It is recommended to use institutional email instead of personal emails.</p> <p>The following properties are expected:</p> <ul style="list-style-type: none"> ● organisationName (characterString and free text) ● contactInfo (CI_Contact): <ul style="list-style-type: none"> ○ address: <ul style="list-style-type: none"> ■ electronicMailAddress [1..*] (characterString) <p>The role of the responsible party serving as a metadata point of contact is out of scope of the INSPIRE Metadata Regulation 1205/2008/EC, but this property is mandated by ISO 19115. The default value is <code>pointOfContact</code>.</p> <p>Properties in the xml file (ISO/TS 19139 path): <code>MD_Metadata/*/contact</code></p> <p>Elements in the ISO 19115-1 (2014) standard: <code>MD_Metadata.contact>CI_Responsibility</code></p>	<p>M/1..n</p>
<p>19</p>	<p>Metadata date stamp (M)</p>	<p>Metadata date (M)</p>	<p><i>Date that the metadata was created</i></p> <p>The date which specifies when the metadata record was created.</p> <p>The default reference system shall be the Gregorian calendar, with dates expressed in accordance with ISO 8601 (yyyy-mm-dd where yyyy is the year, mm is the month and dd is the day). Example: 2012-02-20.</p> <p>Properties in the xml file (ISO/TS 19139 path): <code>dateStamp</code></p> <p>Elements in the ISO 19115-1 (2014) standard: <code>MD_Metadata.dateInfo</code></p>	<p>M/1</p>

<p>20</p>	<p>Metadata language (O)</p>	<p>Metadata language (M)</p>	<p><i>Language used for documenting metadata</i></p> <p>This is the language in which the metadata elements are expressed.</p> <p>The <i>Resource language</i> is a three-letter code for the language taken from the vocabulary LanguageCode (ISO/TS 19139) based on alpha-3 codes of ISO 639-2. Use only three-letter codes from in ISO 639-2/B (bibliographic codes).</p> <p>The list of all the codes is defined at http://www.loc.gov/standards/iso639-2/ Regional languages also are included in this list.</p> <p>Properties in the xml file (ISO/TS 19139 path): language</p> <p>Elements in the ISO 19115-1 (2014) standard: MD_Metadata/defaultLocale: PT_Locale</p>	<p>M/1</p>
-----------	-------------------------------------	-------------------------------------	---	------------

GEO_42 Rationale

Requirement:

GEO_42 Ref GEO_17	Standardised machine-readable geospatial metadata	Descriptive geospatial metadata in the long-term preservation format representation of the Information Package SHOULD be provided in the form of standardised machine-readable format compliant with geospatial metadata standards	0..n SHOULD
GEO_42a Ref GEO_42	Placement of standardised machine-readable geospatial metadata	If a standardised descriptive geospatial metadata file exists it MUST be provided in representations/[RepresentationName]/metadata/descriptive	Conditional 1..1 MUST
GEO_42b Ref GEO_42 and GEOSTR1	XML schema definition for geospatial metadata	If a standardised descriptive geospatial metadata file exists it MUST be accompanied by an XML schema definition placed in a sub-folder called /schemas within the Information Package root folder or the representation folder	Conditional 1..n MUST

Description:

A minimum of contextual information about each geospatial dataset in the Information Package must be provided in a machine-readable metadata format compliant with a geospatial metadata standard. This metadata file should be placed in a `metadata/descriptive` folder on the representation level in the Information Package.

Example:

See an example of an INSPIRE file in the example Information package. [This is not yet available.]

Rationale:

Some metadata is specific and essential to describing geospatial datasets and may not be present in an archival metadata file (EAD or Dublin Core) used to describe the content of the whole Information Package at the package level. This kind of metadata is usually called *discovery* or *descriptive* metadata and is often used in metadata catalogues and Finding Aids enabling automated search, discovery and identification of the geospatial records. An XML schema definition enables validation of a standardised geospatial metadata file (GEO_42b).

GEO_43 Rationale

Requirement:

GEO_43 Ref. GEO_17	Non-standardised machine-readable	A copy of Geospatial metadata in non-long-term preservation representations MAY be stored in its original form as databases or documentation. However, if this data is stored in a long-term preservation representation, the formats need to comply with the	0..n MAY
-----------------------	-----------------------------------	--	-------------

	Geospatial metadata	archival guidelines (stored in approved long-term preservation formats).	
--	---------------------	--	--

Description:

This requirement addresses proprietary or local/national metadata repositories or documents based on legacy formats or standards that are not aligned with typical worldwide xml-based formats for geospatial metadata.

Example:

An example would be:

- A local database that contains metadata descriptions of geospatial records
- Unstructured text-based documents or files describing the context of geospatial records
- Any other legacy documentation that contains metadata elements describing geospatial records.

Rationale:

Standardised machine-readable geospatial metadata is essential for supporting quick and clear automated identification of the object we need in a vast collection of preserved data. Standard-based geospatial metadata enables us to use existing Geospatial metadata catalogues and validators and enable harvesting of compatible metadata elements to archival catalogues.

4. Examples and tools

This section contains an overview of available example packages and tools as a means to “get your hands dirty” and take action in developing the field of geospatial data preservation.

○ Examples

Examples are often the best teacher. In this section, we will guide the reader to examples of valid CITS Geospatial packages. These can in general, be found at the GitHub site for this specification:

<https://github.com/DILCISBoard/CITS-Geospatial> .

Examples will be described in the table below [By publication of version 1.1 of this guideline the examples are under development]:

Link	Name and Description	Number of representations	Proprietary format
[Link to the example]	[The example name and a short description of the example]	[An integer describing the number of representations present in the example]	[A description of the proprietary formats used in the example]
[Link to the example]	[The example name and a short description of the example]	[An integer describing the number of representations present in the example]	[A description of the proprietary formats used in the example]

It is a plan that more examples will follow. If you have a good example, please let us know via the “Issues”-function in GitHub portal <https://github.com/DILCISBoard/CITS-Geospatial/issues> .

○ Tools

There are several open-source and commercial tools to support geospatial data preservation workflows, management and its reuse.

Name	Description	Link
QGIS	This is a fully functional open-source desktop GIS system. It supports transformations of coordinate systems, geospatial data formats. It includes GDAL and PROJ functionalities and can be used to reconstruct information products based on geospatial data. QGIS can be used to consume INSPIRE compliant view and download services. The QGIS plugin ‘GML	www.qgis.org/

	Application Schema Toolbox' makes use of the GMLAS driver for GDAL and aims to develop tools to manipulate complex features in a GIS desktop application.	
GDAL/OGR	GDAL is an open-source C++ translator library for more than 200 raster and vector geospatial data formats. The libraries can support the transformation from original to Long term preservation formats and back to many dissemination formats. It could also be used for the automation of preservation actions.	https://gdal.org/
PROJ	PROJ is a generic coordinate transformation software that transforms geospatial coordinates from one coordinate reference system (CRS) to another. This includes cartographic projections as well as geodetic transformations.	https://proj.org/
Geonetwork	GeoNetwork is a metadata catalogue application to manage spatially referenced resources. It provides powerful metadata editing and search functions as well as an interactive web map viewer. It is currently used in numerous Spatial Data Infrastructure initiatives across the world. It is an excellent tool to be used together with archival catalogues.	https://geonetwork-opensource.org/
FME	A commercial tool for migration of various geospatial and non-geospatial formats. It supports graphical workflow design and automation.	https://www.safe.com/
Hale>>Studio	An opensource tool for transforming data into INSPIRE compliant datasets. It contains presets for all INSPIRE schemas, Integration of the INSPIRE codelist registry, full support for all INSPIRE default and alternative encodings and more.	https://wetransform.to/halestudio/
OSGeo	The OsGeo-Live environment includes Open Source software that is needed to implement these different INSPIRE components. Furthermore, several software tools have been developed within EU-funded projects and are suitable for reuse.	https://wiki.osgeo.org/wiki/INSPIRE_tools_inventory
OGC API		

If you know of other good examples of freely available tools - please let us know via the "Issues"-function in GitHub portal [Add URL].

5. Postface

AUTHOR(S)	
Name(s)	Organisation(s)
Ann-Kristin Egeland	The Danish National Archives
Anita Graser	Austrian Institute of Technology
Gregor Završnik	Geoarh

REVIEWER(S)	
Name(s)	Organisation(s)
Jaime Kaminski	Highbury R&D
Karin Bredenber	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
1.0.0	2021-06-11	Ann Kristin Egeland, Gregor Završnik	DNA Geoarh	Creation of version for public review
1.0.0	2021-08-31	Various	Various	Publication of version 1.0.0
1.1.0	2024-12-13	Various	Various	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.

Appendix 1: Long-Term preservation format Profile for Geospatial Vector data using GML 3.2.1

See separate Appendix 1

Appendix 2: Long-Term preservation format Profile for Geospatial Raster data using TIFF baseline 6

See separate Appendix 2

Appendix 3: Mapping of INSPIRE metadata descriptions to archival descriptive metadata standards

See separate Appendix 3