Guideline for CITS Geospatial, Appendix 2

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

Date: 13.12.2024

Version: 1.1.0

Guideline for CITS Geospatial, Appendix 2

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

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.

Co-funded by the European Union

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

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

Name	Long-Term preservation format Profile for Geospatial Raster data using TIFF baseline 6
Category	Example of a raster profile for the CITS Geospatial guideline
Maturity level	Published
Version	1.1.0
Status	Serves as an example of a profile of the CITS Geospatial Guideline
Approval date	Not applicable
Issue date	2024-12-13
Replaces	None
Prerequisite	None
Attachments	No sample Information Package illustrates the use of this profile yet
Languages	English
Publisher / distributor	DILCIS Board, https://dilcis.eu/

Summary

This document specifies the Long-Term preservation format Profile for Geospatial Raster data using TIFF baseline 6 and serves as Appendix 2 for the CITS Geospatial specification guideline**.** It is a normative description of a standard for the long-term preservation of geospatial raster data.

The specification is based on standards, including TIFF baseline 6.

The aim of employing internationally recognised standards is to ensure the long-term preservation of, and access to, the widely used geospatial raster data, as well as the easy exchange of raster data independent of proprietary raster formats.

Version history

Relationship of the present version to previous versions:

Long-Term preservation format Profile for Geospatial Raster data using TIFF baseline 6	First draft of the profile

Table of Contents

1 Introduction

1.1 Status

This document was approved by the DILCIS Board.

1.2 Area of application

1.2.1 Addressees / target group

This is a technical document for IT specialists involved in the long-term archiving of geospatial raster data. This specification covers the long-term preservation of georeferenced images and grid files. All significant properties of other raster formats like satellite imagery and oblique images are not covered by this specification.

1.2.2 Background

This is a standard for the long-term archiving of geospatial raster datasets based on these existing standards for imagery, geometry and information about Coordinate and Reference System (CRS): TIFF baseline 6, ESRI World file, Well Known Text file format (WKT) and GML 3.2.1.

Long-term archiving is the preservation, normally without a time limit, of the information stored in the data files while retaining the bit stream and the ability to interpret and display the data in a way that is human-readable and comprehensible.

If the structure and content of any geospatial raster data format are translated into the proposed preservation formats, like TIFF with additional world files and projection files, it will subsequently be possible to access and exchange the data in the TIFF files at any time, even when the original raster format software is no longer available or can no longer be run. This has been achieved by the use of suitable standards for raster data and additional files with geospatial information that are widely supported internationally.

The recommended solution for the preservation of geospatial raster data consists of four files:

1. A TIFF file for storing the raster object
2. A TFW file for storing the location of the raster object
3. A PRJ file for storing the description of the Coordinate and Reference System used (CRS)
4. A GML file for storing any tiling schema or spatial coverages

Significant properties are those aspects of the digital object which must be preserved over time for the digital object to remain accessible and meaningful. Table 1 describes relations between significant properties and the recommended long-term preservation solution specified in this profile.

XML in table 1 refers to the metadata standard “INSPIRE Metadata Implementing Rules” or the METS standard and GML to the Long-Term preservation format Profile for Geospatial Vector data using GML 3.2.1 (Appendix 1).

Table 1 Coverage of significant properties in the profile

Category	Significant property	TIFF	TFW	PRJ	GML	XML
Content	Compression					
	Spatial resolution					
	Pixel value type					
	Bit depth					
	Bands					
	Raster dimension					
	Raster object location					
	Geospatial resolution					
	Coordinate reference system			Full description	Reference	INSPIRE Reference
	Spatial raster coverage			Full description	Reference	INSPIRE Reference
Context	Geospatial metadata					INSPIRE
	Spatial accuracy					INSPIRE
	Temporal accuracy					INSPIRE
	Lineage information					INSPIRE
	Acquisition information					INSPIRE
	Processing information					INSPIRE
	Source data					INSPIRE
	Contextual information					METS
Structure	Spatial coverage file					
	Tiling index					

1.2.3 Distinctions

It should be noted that this profile is only the long-term storage format for a specific type of geodata: geospatial raster data in the form of georeferenced raster images and grid files. It is assumed that geospatial raster data preserved using this profile is archived as part of an Information Package (CSIP) together with other documents (e.g. metadata files, vector data, SIARD file, documentation, etc.). Requirements for a geospatial Information Package is specified in the CITS Geospatial specification. This Long-Term preservation format Profile for

Geospatial Raster data using TIFF baseline 6 provides additional requirements to the CITS Geospatial specification.

Just as an XML-based Word or e-mail file contains an internal file structure consisting of metadata, primary data and various auxiliary data, geospatial raster data preserved using this profile contains its own metadata describing the data more precisely in addition to the actual data – regardless of the metadata catalogue that an archive records in its OAIS packages (e.g. EAD metadata).

2 Structure of the document

2.1 Structure of chapters

Each chapter in this specification is constructed according to the same pattern. After a brief introduction, the requirements are listed in a table.

ID	Description of requirement	M/O
contains the ID of the requirement	contains the text of the requirement	stipulates whether mandatory or optional

A requirement is frequently further explained using recommendations, notes and examples, each of which is specifically indicated as such.

ID	Description of requirement	M/O
A_3.1-1	Text of requirement **Example** Text of example **Note** Text of note **Recommendation** The text of recommendations is in italics.	M

2.2 ID for requirements

The requirements are unambiguously identifiable using an ID.

ID
G_2.1-1

This ID is constructed according to the following pattern:

G_	Letter +	=	identifies main chapters
	G_	=	General requirements
	P_	=	Requirements for package structure
	D_	=	Requirements for data
	M_	=	Requirements for metadata

2.1-1 The number begins with the number of the chapter (which groups together requirements on the same topic), and the number after the dash is consecutive, thus designating all the requirements in the chapter.

2.3 The distinction between mandatory and optional requirements

Each requirement is either mandatory or optional. This is indicated by a letter:

Abbreviation	Meaning
M	Mandatory requirement This requirement must be met.
O	Optional requirement This requirement should be met. It simplifies handling and constitutes best practice.

2.4 Notation of folders, files and folder structures

The following symbols and parameters are used for the notation of folders, files, etc.

Symbol	Meaning
`/`	Folder
`header/`	A folder with the name “header”
`xy.txt`	File (with file extension “txt”)
`dir1/`	Example folders (in red)
`abc.pdf`	Example files (in red)
…	Placeholder for files or folders that are not relevant to the explanation
`<xx>`	Placeholder for any desired string of characters

3 General requirements

3.1 Use of standards

To ensure that the contents of georeferenced images and grid files remain interpretable over a long period, the Long-Term preservation format Profile for Geospatial Raster data using TIFF baseline 6 is essentially based on standards.
The profile is a combination of several file types to fulfil the preservation of all significant properties of georeferenced raster images and grid files.

ID	Description of requirement	M/O
G_3.1-1 Ref GEO_21	A raster object (image) of a geospatial raster dataset is stored as an image file in accordance with one of the following standards: • TIFF version 6.0 baseline, Adobe, June 3, 1992[1] **Recommendation** *Other formats like JPEG-2000 is recommended to preserve large raster images*	M
G_3.1-2	Location information stored externally to the raster object is stored in world file format in accordance with ESRI World File standard[2]	M
G_3.1-3	Full description of the Coordinate Reference System (CRS) stored externally to the raster object is stored in a projection file in a Well Known Text file format (WKT) in accordance with one of the two standards • ISO 19162:2019 Geographic information - Well-known text representation of coordinate reference systems • Geographic information — Well-known text representation of coordinate reference systems, version 2.0.6, OGC, June 2019[3]	M
G_3.1-4	A tiling index or spatial coverage file is stored in GML 3.2.1 and in accordance with the Long-Term preservation format Profile for Geospatial Vector data using GML 3.2.1 (Appendix 1).	M
G_3.1-5	A spatial coverage file for georeferenced images is stored in GML 3.2.1 and in accordance with OGC Coverage Implementation, Version 1.1, OGC, September 15, 2017[4]	O

[1] TIFF, Revision 6.0. Adobe, June 3, 1992.
https://www.loc.gov/preservation/digital/formats/fdd/fdd000022.shtml
2 ESRI World file format, Esri technical support, June 16, 2020.
http://webhelp.esri.com/arcims/9.3/General/topics/author_world_files.htm
[3] Geographic information — Well-known text representation of coordinate reference systems: https://www.ogc.org/standards/wkt-crs version 2.0.6, OGC, June 2019
[4] OGC Coverage Implementation Version: 1.1, OGC, September 15, 2017 https://docs.opengeospatial.org/is/09-146r6/09-146r6.html#5

G_3.1-6	A spatial coverage file for georeferenced grid files is stored in GML 3.3 and in accordance with OGC Coverage Implementation Schema - ReferenceableGridCoverage Extension with Corrigendum, Version: 1.0.1, OGC, January 20, 2019[5]	O
G_3.1-7	Geospatial metadata is stored in an XML file compliant with the INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119[6]	M

3.2 Character sets and characters

ID	Description of requirement	M/O
G_3.2-1	Plain text files are encoded in the UTF-8 character set. The use of characters is limited as described in ISO/IEC 10646:2003 Annex D and The Unicode Standard 5.1, Chapter 3.	M

3.3 File format extensions

ID	Description of requirement	M/O
G_3.3-1	TIFF files has the file extension `.tif`	M
G_3.3-3	A world file has the extension `.tfw`	M
G_3.3-5	A projections files has the extension `.prj`	M

[5] OGC Coverage Implementation Schema - ReferenceableGridCoverage Extension with Corrigendum, Version: 1.0.1, OGC, January 20, 2019 https://docs.opengeospatial.org/is/16-083r3/16-083r3.html

[6] INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119, https://inspire.ec.europa.eu/documents/inspire-metadata-implementing-rules-technical-guidelines-based-en-iso-19115-and-en-iso-1

4 Requirements for package structure

This section lists requirements on how to place different files of a geospatial raster dataset in an Information Package that conforms to the Common Specification for Information Packages (CSIP). Each geospatial raster dataset in an Information Package preserved using this profile consists of one or more of the following files:

Data files

- a *TIFF file* storing the raster image or the gridded data
- a *TFW file* with location information of the raster image or grid file
- a *GML file* storing location information or a tiling index, or spatial coverage of raster objects
- a *PRJ file* with a full description of the Coordinate Reference System (CRS) used for location information in a TFW file

Metadata files

- an *XML metadata file* compliant with the "INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119" with descriptive geospatial metadata at the dataset level

Documentation files

- *a text, image, audio or video file* describing the dataset (contextual documentation)

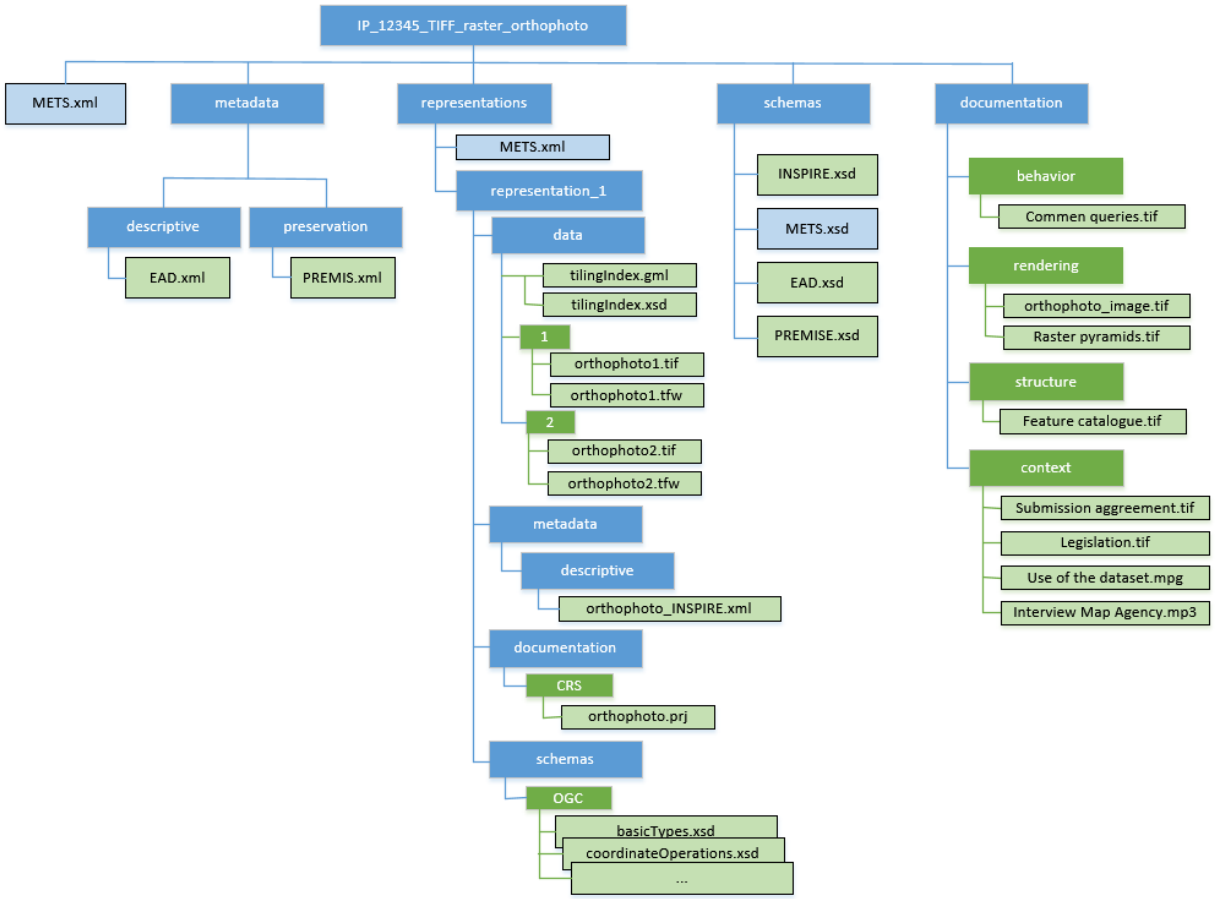

Figure 1: Folder structure of the Long-Term preservation format Profile for Geospatial Raster data using TIFF baseline 6

ID	Description of requirement	M/O
P_4.0-1 Ref GEO_11 and GEO_12 and GEO_13	If the value in mets/@csip: CONTENTINFORMATIONTYPE is "GeoData", then there must exist at least one file with extension `.tif` in `representations/[RepresentationName]/data` or in any of its subfolders	M
P_4.0-2 Ref GEO_12	If there are more raster files in a representation, each raster file could be placed in subfolders in `representations/[RepresentationName]/data` named consecutively with a number starting with 1	O
P_4.0-3 Ref GEO_15 and GEO_38	A full description of the Coordinate Reference System (CRS) for all geospatial records is provided in one of two ways: - in an accompanying projection file (TFW) in `representations/[RepresentationName]/documentation/data` - in an accompanying projection file (TFW) in `representations/[RepresentationName]/documentation/CRS`	M
P_4.0-4 Ref GEO_31	An image displaying the overall view of any geospatial dataset in the IP is provided in a `documentation/rendering` folder	M
P_4.0.5	Location information of a georeferenced raster image or grid file is provided in a plain text file in accordance with TFW (Tiff World File) grouped in the same folder as the raster object it relates to **Note** A TFW file contains the location information in the form of geospatial vector coordinates and the size of the first top left pixel of the raster image, thus enabling the placement of the raster image in a Coordinate Reference System (on a map)	M
P_4.0.6 Ref GEO_23	If raster objects are organised using an external tiling index file, the tiling index is placed in a folder grouping the raster objects it covers **Note** If all raster objects in a representation are organised in the same tiling index, the tiling index file is placed in the root of `representations/[RepresentationName]/data`	M

5 Requirements for data

The category *Content* in table 1, displayed below, covers the data part of the significant properties necessary to preserve. The table illustrates which file format that stores each significant property. A TFW World file or GML file must be accompanied by a PRJ projection file to fully store information about the *Raster object location* and *Geospatial resolution*.

Category	Significant property	TIFF	TFW	PRJ	GML	XML
Content	Compression					
	Spatial resolution					
	Pixel value type					
	Bit depth					
	Bands					
	Raster dimension					
	Raster object location					
	Geospatial resolution					
	Coordinate reference system			Full description	Reference	INSPIRE Reference
	Spatial raster coverage			Full description	Reference	INSPIRE Reference

5.1 TIFF version 6.0 baseline requirements

A georeferenced raster image is composed of a rectangular array or grid of pixels, each of which represents a value. In the case of ordinary raster images, this value usually represents a colour. In other spatial raster datasets like grid files, pixels can also represent measurements, derived calculations, classifications or any other units that the pixel represents.
The TIFF version 6.0 baseline format stores the raster object (image) of the spatial raster dataset. Requirements for a TIFF file are defined below.

ID	Description of requirement	M/O
D_5.1-1 Ref G_3.1-1	A TIFF file must be a valid TIFF version 6.0 baseline file	M
D_5.1-2	TIFF file is compressed according to the following compression rules and described in a TIFF Tag: ● Black/white raster objects must be compressed using CCITT/TSS grp. 3, grp. 4, PackBit or LZW ● Raster objects with grey tones or colours must be compressed using PackBit or LZW	O

	• In a TIFF file lossy compressions like JPEG should be avoided. Especially if the exact value of pixels is important.	
D_5.1.3	Pixel value type is described in a TIFF tag, and the domain is either an integer or a floating point (decimal)	M
D_5.1-4	Bit depth is described in a TIFF tag, and TIFF RGB raster objects may only use the following bit depths: 1, 2, 4, 8, 24, 32 and 64 bit	M

5.2 TFW world file requirements

ID	Description of requirement	M/O
D_5.2-1	A TIFF file must be accompanied by an additional ESRI World file (TFW) storing location information of the georeferenced TIFF file	M
D_5.2-2 Ref G_3.1-2	A TFW file must be a valid ESRI World file (TFW) **Note** A TFW file is a Tiff World File and contains the vector coordinates, and the size of the first top left pixel of the raster image enables the placement of the raster image in a Coordinate Reference System (on a map). A tfw file has 6-lines describing the location (x,y), scale (pixel size) and orientation (rotation) of the image. Line 1: pixel size in the x-direction in map units Line 2: rotation about the y-axis. Line 3: rotation about the x-axis. Line 4: pixel size in the y-direction in map units Line 5: x-coordinate of the upper left corner of the image. Line 6: y-coordinate of the upper left corner of the image. As pixels are considered as square lines, 1 and 4 are the same.	M
D_5.5-3	What is important/requirements to enable validation of TFW file? Similar GEO_16 requirement? Validation of content in several TFW files against an expected bounding box?	O

5.3 PRJ projection file requirements

ID	Description of requirement	M/O
D_5.3-1	A raster dataset must be accompanied by one or more additional projection files (PRJ) with full descriptions of all Coordinate Reference Systems (CRS) used in the dataset	M
D_5.3-2 Ref G_3.1-3	A PRJ file must be in a valid Well Known Text file format (WKT)	M
D_5.3-4	A PRJ file should contain all the information necessary to identify the used Coordinate Reference System in the dataset and to transform the described Coordinate Reference System to other Coordinate Reference Systems	O

5.4 GML spatial coverage file requirements

ID	Description of requirement	M/O
D_5.4-1	A raster dataset could be accompanied by one or more additional spatial coverage files **Note** A spatial raster coverage file provides a fast and searchable overview in the form of vector areas of all the raster objects in a dataset with corresponding attributes.	O
D_5.4-2 Ref G_3.1-5	A spatial coverage file for georeferenced images is a valid GML file in accordance with OGC Coverage Implementation version. 1.1 (GMLCOV) **Note** A GML spatial coverage file contains complete raster geo referencing information and, as such, provide information similar to that stored in a TFW world file and PRJ projection file.	M
D_5.4-3 Ref G_3.1-6	A spatial coverage file for georeferenced grid files is a valid OGC Coverage Implementation Schema - ReferenceableGridCoverage Extension with Corrigendum file	M

5.5 GML tiling index requirements

ID	Description of requirement	M/O
D_5.5-1	A raster dataset could be accompanied by one or more additional tiling indexes	O
D_5.5-2 Ref G_3.1-4	A tiling index is compliant with GML 3.2.1 **Note** When a geospatial raster record is large like an Orthophoto Campaign covering a whole country, it is divided into several raster objects and organised in a tiling index with tiles covering a fixed area (e.g. 1x1km).	M

6 Requirements for metadata

ID	Description of requirement	M/O
M_6.0-1 Ref GEO_42 Ref G_3.1-7	Geospatial metadata in the long-term preservation format representation of the Information package must comply with the INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119. Metadata is placed in *Representations/[RepresentationName]/metadata/descriptive*	M