



— **Geo-SIP and Geo-Dossier specification**

Package for digital geodata submitted to the Swiss Federal Archives

Date: 5 December 2016

Version: 1.0

Version	Date	Description, comment	Name or role
1.0	05/12/2016	Version for SFA board and swisstopo board	Project team Ellipse
	12/12/2016	Approval by SFA board and swisstopo board	
	22/02/2017	Approval by GCG	

Table of contents

1	INTRODUCTION.....	4
1.1	Digital archiving of geodata	4
1.2	Aim and purpose	4
2	BASIC PRINCIPLES	4
2.1	Package structure.....	4
2.2	Submission type and subtype.....	5
2.3	Limits	5
2.4	Normalisation of file and path names.....	5
3	GEO-SIP	5
3.1	Structure.....	5
3.2	Metadata	5
3.2.1	Submission level.....	5
3.2.2	Classification system level	6
3.2.3	Classification system item level.....	6
3.2.4	Dossier level.....	6
3.2.5	Subdossier level	6
3.2.6	File level.....	6
4	GEO-DOSSIER.....	6
4.1	Principles.....	7
4.2	Structure of a Geo-Dossier	8
4.2.1	General information on the structure of a Geo-Dossier	9
4.2.2	Structure elements and rules.....	10
4.2.2.1	Geo-Dossier folder.....	10
4.2.2.2	Standard folders	10
4.2.2.3	Representation folders	11
4.2.2.4	Grouping folders	12
4.2.2.5	File	12
4.2.2.6	Readme.txt	13
4.2.2.7	Geospatial metadata extract	13
4.2.2.8	Preview	14
4.2.3	Examples.....	14
5	GLOSSARY	17

1 Introduction

1.1 Digital archiving of geodata

The Archiving Act (ArchA)¹ and the Geoinformation Act (GeolA)² stipulate that geodata under federal legislation must be archived. A solution is to be sought that works for the whole of Switzerland, that enables archiving of geodata from the Confederation, cantons and communities and that is based on international standards to the greatest degree possible. The following solution is restricted to the file-based archiving of geodata.

1.2 Aim and purpose

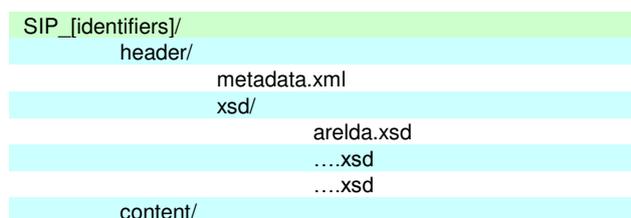
Based on the (general) specification Submission Information Package (SIP)³, this specification has the following aims:

- This specification contains the SFA's requirements⁴ for delivering digital geodata to the SFA. The specification determines how digital submissions of geodata to the SFA must be shaped and thereby serves on the one hand as a specification to be used internally by the submitting authorities and on the other hand as a tool that enables them to communicate with their service providers and software developers.
- This specification supplements the Specification Submission Information Package (SIP) with requirements that must be met when implementing interfaces for creating and submitting digital geodata.
- The following solution is restricted to the archiving of transfer files and is based on the archival submission interface eCH-0160. The solution must be checked once the standard ISO 19165 on the archiving of geodata is available.

2 Basic principles

2.1 Package structure

Regardless of their submission type or their status in the lifecycle (SIP/AIP/DIP), all information packages have the same package structure. They are based on a file/folder structure. Requirements exist in relation to structure and naming conventions (content and structure of names) of the individual folders and files. Even if the submitting authority does not have to manage the structure of a Geo-SIP, because this is automatically generated from a geo-dossier by the package handler, it is shown below what the top folder of a SIP (and therefore also Geo-SIP) looks like:



¹ SR 152.1

² SR 510.62

³ [Specification Submission Information Package](#) (in German)

⁴ See <https://www.bar.admin.ch/bar/de/home/archivierung/ablieferung/digitale-unterlagen.html> (in German)

2.2 Submission type and subtype

The SIP specification defines two submission types: FILES and GEVER. For Geo-SIP, a new subtype *FILES-SIP with geodata* is defined for the type FILES, in a similar manner to *FILES-SIP with integrated documentation*. This addition is purely for organisational purposes and does not require any adjustment of the SIP specification.

2.3 Limits

The existing SIP specification limits various technical aspects of a SIP: Package size limit (8 GB), path name length (180 characters), number of files in a folder (5000), number of files in a SIP (1,000,000). However, these are not hard, technical limits, such as e.g. those prescribed by the file system used⁵. These limits have been defined for ease of handling. The package size limit (8 GB) is lifted for Geo-SIP.

2.4 Normalisation of file and path names

Based on the restrictions stipulated by the SIP specification, normalisation must be performed for package creation and archiving if the original files use forbidden characters in their file names or if the path names are too long. In this case, the files and folders are renamed and moved during package creation. The original names and items in the folder hierarchy are recorded in the file metadata.xml so that this information is not lost.

As described below, both the hierarchy of the folders for a package with geodata and the file names of the archivable formats for raster and vector data contain information that disappears from the file and folder structure during normalisation and is instead transferred to metadata.xml. The package handler can undo the normalisation for existing packages so that this information is once again mapped in the file system.

3 Geo-SIP

3.1 Structure

The structure of a Geo-SIP is identical to that of a version 4.0 SIP package (SFA specification, corresponds to eCH-0160, v.1.0). The term Geo-SIP is used below to indicate a SIP V4.0 that contains at least one geo-dossier (see next chapter).

3.2 Metadata

In addition to geospatial metadata, which are saved in the *content* as formalised documentation of the geodata, there are also metadata that are stored in the *header* (see Chapter **Fehler! Verweisquelle konnte nicht gefunden werden. Fehler! Verweisquelle konnte nicht gefunden werden.**). These metadata are, on the one hand, archival metadata that describe a geo-dossier from an archival perspective for indexing and are recorded with the package handler and, on the other hand, metadata that map the structure of the folders and files of the contents and are automatically generated by the package handler.

Requirements and conventions for setting the metadata when creating a Geo-SIP are listed below. The terms used must be standardised in terms of language, either in German or in French.

3.2.1 Submission level

At the submission level, it must be noted that for geodata the *submitting authority* is often not the same as the *records creator*. If the geodata set is in the Federal Spatial Data Infrastructure (FSDI), it is generally also submitted from there. In this case, "Federal Office of Topography" is set as the *submitting authority*.

⁵ The file system used by Microsoft, NTFS, permits folders with up to 4.3 billion files, for example.

Closure period category and *closure period* must be given either at the submission level or the geodata set level.

3.2.2 Classification system level

For the metadata field *Name*, the record creator followed by a colon, a space and the word “Geodata” is entered. Example: “Federal Office of Topography: Geodata”

3.2.3 Classification system item level

If multiple geodata sets are summarised in one Geo-SIP, a classification system item must be created for each geodata set. For official geodata sets, the identifier from the collection of official geodata sets of the Confederation⁶ *is used as the Number* and the corresponding designation according to the collection as the *Title*. For other geodata, continuous numbering is used, beginning with 1. For subsequent submissions, the numbering of the geodata sets must be preserved.

3.2.4 Dossier level

The *Title* corresponds to the name of the dossier folder on the file system (cf. 4.2.2.1-4). The *File reference* is the *Number* of the classification system item. The *Time period* must be specified. If two dossiers have the same *Title*, they must have a different *Time period*. *Closure period category* and *closure period* must be set either at the submission level or the dossier level.

3.2.5 Subdossier level

Within a geo-dossier, folders are mapped on the file system as subdossiers. The names of the folders in the file system are copied over as the *Names* of the subdossiers in the metadata.

3.2.6 File level

One file on the file system corresponds to one file in the metadata. The name of the file in the file system is copied over as the *Name* of the file in the metadata.

4 Geo-Dossier

This chapter describes an ordered folder structure for geodata. Its aims are, on the one hand, to take into account the multi-dimensional characteristics (time, layer, space) of geoinformation and, on the other hand, to make it easy for both humans and machines to interpret the structure. The structure must be flexible enough to incorporate geodata from different producers or suppliers, and it must remain independent of any format. Therefore, it must not make any distinction as to whether raster or vector data are contained, or a combination of the two. It must also be possible to add feature attachments, e.g. text files or multimedia files such as films, to the classic geodata.

The term *Geo-Dossier* is introduced below for this purpose.

The general definition of a dossier is roughly as follows (there is no official definition):

A dossier comprises all documents relating to a specific business matter or originating from a different folder structure (database, file storage). By combining similar business matters or dividing dossiers into subdossiers, this basic structure can be adapted to meet the corresponding needs.

6

<http://www.geo.admin.ch/internet/geoportal/de/home/topics/geobasedata/introductionplan.parsys.9842.downloadList.40450.DownloadFile.tmp/sammlunggeobasisdatensatzedesbundesrechtsde.xlsx> (in German)

The structure of a geo-dossier must be simple and transparent. It is designed first and foremost for readability by humans, with machine readability being a secondary concern. It is not intended to be integrated into a GIS fully automatically.

A geo-dossier prescribes the rules for the structure of its contents. In a similar manner to the requirements listed in the SIP specification under number M_4.8-3 for the structuring of data in a FILES-SIP, provisions are set out here for the structured storing of geodata.

4.1 Principles

- One Geo-SIP can contain one or more geo-dossiers.
- A geodata set is understood as meaning all thematically linked geodata. A geodata set is not redefined for submission to the archive.
- One geo-dossier contains precisely one snapshot⁷ of a geodata set. Exceptions to these principles are historicised data sets of vector data and collections of raster data tiles that cover an area, redundancy-free, within a specific time period.
- A geo-dossier contains either one or several representations of a geodata set.
 - Representations are different versions of a single object, where each version has the same contextual information but in a different form. Geodata sets can be understood as representations of different quality, reference systems or combinations of these.
 - Time stamps are not representations.
- One geodata set can be divided into individual parts, e.g. thematic breakdown (single layers), spatial breakdown (single areas), breakdown according to data type such as vector or raster data. Partial geodata sets such as these are broken down in grouping folders within the geo-dossier.
- The standard structure of a geo-dossier is geared towards the simplest case (single representation of a geodata set without any breakdown). However, the structuring rules allow for more complex structures.
- To reduce the effort involved, it is recommended to create one geo-dossier per snapshot for vector data and to submit everything in a single Geo-SIP. For large raster data sets, however, it is recommended to submit the data for one time period in a geo-dossier and a single Geo-SIP.
- A geo-dossier does not contain any references to external documents. Referenced documents, in particular referenced models and all INTERLRIS imports, must be included in the geo-dossier. The references must then be adapted accordingly so that they point to the local documents.
- At least one preview image must be provided in every geo-dossier.

⁷ Status of an object at a specific point in time. In the context of this document, snapshots refer to the status of a geodata set at a defined point in time.

4.2 Structure of a Geo-Dossier

A geo-dossier consists of folders, with files below these. The structure and naming of the folders and files must comply with the given regulations. The diagram below shows permitted structure elements in an example where the structure element type is specified in the right-hand column. Not all structure elements have to be present.

SIP_[identifiers]/		
header/		
...		
content/		
38.11 VECTOR200 2010 edition/		Geo-Dossier
1_DOC/		Standard folder "Documentation"
Readme.txt		Documentation on the geo-dossier
docu_vector200_2010.pdf		Documentation on the geodata
vector200_2010_gm03.xml		Geospatial metadata extract GM03 as XML
vector200_2010_gm03.pdf		Geospatial metadata extract GM03 as PDF/A
vector200_2010_iso19139.xml		Geospatial metadata extract ISO19139 as XML
vector200_2010_iso19139.pdf		Geospatial metadata extract ISO19139 as PDF/A
2_MODELS/		Standard folder "Models"
vector200.ili		Geodata model
gm03.ili		Geospatial metadata model
gm03.xsd		Schema for geospatial metadata model
imports		Grouping folder for referenced models
INTERLIS.ili		Basic modules for minimum geodata models
CoordSys.ili		Coordinate systems module
3_DATA/		Standard folder "Geodata"
LV03/		Representation folder "LV03"
3_DATA/		Standard folder "Geodata"
buildings/		Grouping folder
vector200_bui.xtf		Geodata
hydrography/		Grouping folder
vector200_hyd.xtf		Geodata
LV95/		Representation folder "LV95"
3_DATA/		Standard folder "Geodata"
buildings/		Grouping folder
vector200_bui.xtf		Geodata
hydrography/		Grouping folder
vector200_hyd.xtf		Geodata
4_GRAPH/		Standard folder "Graphical representation"
PREVIEWS/		Previews grouping folder
preview_bui.tif		Buildings preview layer
preview_hyd.tif		Water preview layer
sld/		Grouping folder
common.xsd		Graphical description
sld.xml		Graphical description
sld.xsd		Graphical description
style.xml		Graphical description
style.xsd		Graphical description
symbol.xml		Graphical description
symbol.xsd		Graphical description
legend.pdf		Legend

Key:

Name
Geo-Dossier
Standard folder
Representation folder
Grouping folder
File

4.2.1 General information on the structure of a Geo-Dossier

As a general rule, a geo-dossier is structured so as to be superordinate and based on individual, independent and autonomous representations of geodata sets that can each be used separately. A geo-dossier includes not only the actual geodata, but also the data models, documentation and geospatial metadata.

As a basic principle, a geo-dossier is structured in 4 main groups:

- 1_DOC
- 2_MODELS
- 3_DATA
- 4_GRAPH

Standard folder	Allocated data	Ending	Must/Can
1_DOC	Documentation		
	▪ Readme.txt	.txt	C
	▪ Documentation of the minimum geodata model including object catalogue, UML diagram and registration guidelines	.pdf	C
	▪ Release notes	.pdf	C
	▪ Geospatial metadata (ISO 19139 ⁸ and GM03)	.xml	M
	▪ Geospatial metadata (ISO 19139 and GM03)	.pdf	M
	▪ Other documentation	.pdf, .txt, .xml	C
2_MODELS	Models		
	▪ Minimum geodata model	.ili	C
	▪ Data model for geospatial metadata	.ili	M
	▪ XSD schema: geospatial metadata	.xsd	M
	▪ Imports (INTERLIS)	.ili	M ⁹
3_DATA	Geodata ¹⁰		
	▪ Raster	.tif, .ewf.xml	C
	▪ Vector	.itf, .xtf	C
	▪ Feature attachments ¹¹	.tif, .pdf	C
4_GRAPH	Graphical representation		
	▪ Presentation models	.pdf, .xml, .csv	C
	▪ Legends	.pdf	C
	▪ Preview(s)	.tif	M

⁸ ISO/TS 19139:2007 defines the Geographic MetaData XML (gmd) encoding, an XML schema implementation that has been derived from ISO 19115.

⁹ If INTERLIS models are present and reference external models

¹⁰ At least one type of geodata must be included in the form of a raster or vector. Combinations are also possible, such as raster and vector data. Feature attachments usually occur in combination with vector data.

¹¹ It must also be possible to store "feature attachments", such as detailed descriptions of objects in PDF/A, in a geo-dossier. In future this will also include multimedia files, such as films. Just as different data types – vector data, raster data – exist, feature attachments can also exist as a data type. A vector feature can be linked with one or more feature attachments.

4.2.2 Structure elements and rules

Below is a list of the structure elements shown in the example above and the applicable rules.

4.2.2.1 Geo-Dossier folder

ID	Geo-Dossier folder: Contents	Must/can
4.2.2.1-1	The folder contains all files and folders of a geo-dossier.	M

ID	Geo-Dossier folder: Position	Must/can
4.2.2.1-2	The folder is directly below the folder content/ of the SIP. Example content/38.11 VECTOR200 2010 edition/	M

ID	Geo-Dossier folder: Several geo-dossiers per SIP	Must/can
4.2.2.1-3	One SIP can contain several geo-dossiers.	C

ID	Geo-Dossier folder: Name	Must/can
4.2.2.1-4	The folder name must match the name from geocat.ch with a time reference added. Example Sectoral Plan for Deep Geological Repositories 2012/	M

4.2.2.2 Standard folders

ID	Standard folders: Structure	Must/can
4.2.2.2-1	Directly below the geo-dossier folder must be the four standard folders <ul style="list-style-type: none"> • 1_DOC • 2_MODELS • 3_DATA • 4_GRAPH 	M

ID	Standard folders: Contents	Must/can
4.2.2.2-2	Below the four standard folders 1_DOC, 2_MODELS, 3_DATA und 4_GRAPH, only the following structural elements may be located: <ul style="list-style-type: none"> • Representation folder • Grouping folder • Files in archivable formats 	M

4.2.2.3 Representation folders

ID	Representation folders: Structure	Must/can
4.2.2.3-1	<p><i>Representation folders</i> are mandatory if different representations of a geodata set are mapped. If only one representation is present, there is no representation folder. Representations describe the same information unit in different forms.</p> <p>Example Examples of representations include</p> <ul style="list-style-type: none"> • Quality • Reference system • Combinations of the above 	M

ID	Representation folders: Name	Must/can
4.2.2.3-2	<p>The name of a <i>representation folder</i> can be freely selected in principle, but it must either be meaningful or use common technical terms.</p> <p>Example For the reference frameworks for land surveying: LV03, LV95</p>	M

ID	Representation folders: Position	Must/can
4.2.2.3-3	<p><i>Representation folders</i> may only exist below the standard folder 3_DATA.</p>	M

ID	Representation folders: Contents	Must/can
4.2.2.3-4	<p>Under <i>Representation folders</i> are those of the four standard folders that are not empty:</p> <ul style="list-style-type: none"> • 1_DOC • 2_MODELS • 3_DATA • 4_GRAPH 	M

ID	Representation folders: Nesting	Must/can
4.2.2.3-5	<p><i>Representation folders</i> can be nested in any fashion.</p> <p>Recommendation It is recommended to only make use of this option where strictly necessary. Ideally, only one single hierarchy level should be used for representation folders.</p>	C

4.2.2.4 Grouping folders

ID	Grouping folders: Structure	Must/can
4.2.2.4-1	<p><i>Grouping folders</i> are optional and can be used to structure associated data. This option should only be used where it is practical to do so. A grouping folder must be created for one or more previews (name: PREVIEWS).</p> <p>Grouping folders can also be used to break a geodata set down into different parts in accordance with a criterion. Here, the parts always contain different information and form the whole geodata set when put together. The names of the grouping folders are mandatory structure elements of a geodata set.</p> <p>Example Examples of criteria for subdivision include</p> <ul style="list-style-type: none"> • Thematic breakdown: single layers • Spatial breakdown: single areas • Breakdown according to data types such as vector data, raster data or feature attachments 	C

ID	Grouping folders: Name	Must/can
4.2.2.4-2	<p>Grouping folders can be named freely, but the name must reflect the group name in the case of grouping of individual files or the subdivision criteria for geodata sets and e.g. specify the layer. The folder PREVIEWS is excluded from this rule; its name is always fixed.</p>	M

ID	Grouping folders: No empty folders	Must/can
4.2.2.4-3	Every grouping folder contains at least one file.	M

ID	Grouping folders: Position	Must/can
4.2.2.4-4	<p>The grouping folders are always below the standard folders</p> <ul style="list-style-type: none"> • 1_DOC • 2_MODELS • 3_DATA • 4_GRAPH 	M

4.2.2.5 File

ID	File: Format	Must/can
4.2.2.5-1	All files must be stored ¹² in archivable formats ¹³ .	M

¹² The list of archivable formats of the Swiss Federal Archives is not currently up to date. In addition to the formats currently included in the list, the following are also permitted (and will be added to the next version of the list): [TIFF+EWF.XML](#) for raster data, INTERLIS-ILI, INTERLIS2-XTF, INTERLIS1-ITF, XML and, under certain conditions, Esri Shape.

¹³[List of archivable formats of the Swiss Federal Archives](#)

ID	File: Name	Must/can
4.2.2.5-2	The file name before the ending can be freely selected. Readme.txt and geospatial metadata extract files are special cases. The name "Readme.txt" is fixed. Geospatial metadata extracts are named "GM03" or "ISO19139".	M

ID	File: Position	Must/can
4.2.2.5-3	Files are located either <ul style="list-style-type: none"> • directly below the standard folders 1_DOC, 2_MODELS, 3_DATA and 4_GRAPH or • directly below a <i>grouping folder</i> • but never directly below a <i>representation folder</i> 	M

4.2.2.6 Readme.txt

ID	Readme.txt: Purpose	Must/can
4.2.2.6-1	Readme.txt is intended to provide a minimum level of help at the start. Readme.txt refers to the documents that will rapidly provide the user with an overview of the geodata and explains the structure of the geo-dossier. Readme.txt refers to the linking of individual elements in a geo-dossier. For linking e.g. a vector element to a feature attachment, paths from the geo-dossier cannot be used. Instead, the parts of path names that are to be used as original names must be taken from the data model, along with the parts that must be adapted to the IT environment in question by way of mapping.	C

ID	Readme.txt: Use	Must/can
4.2.2.6-2	Readme.txt is mandatory if the structure elements <i>representation folder</i> or <i>grouping folder</i> are used.	M

ID	Readme.txt: Name	Must/can
4.2.2.6-3	The name "Readme.txt" is fixed.	M

ID	Readme.txt: Position	Must/can
4.2.2.6-4	Readme.txt is located under the folder 1_DOC at the highest level.	M

4.2.2.7 Geospatial metadata extract

ID	Geospatial metadata extract: Contents	Must/can
4.2.2.7-1	The metadata set associated with the geodata set from geocat.ch appears twice, once modelled according to "GM03" and once according to "ISO19139".	M

4.2.2.8 Preview

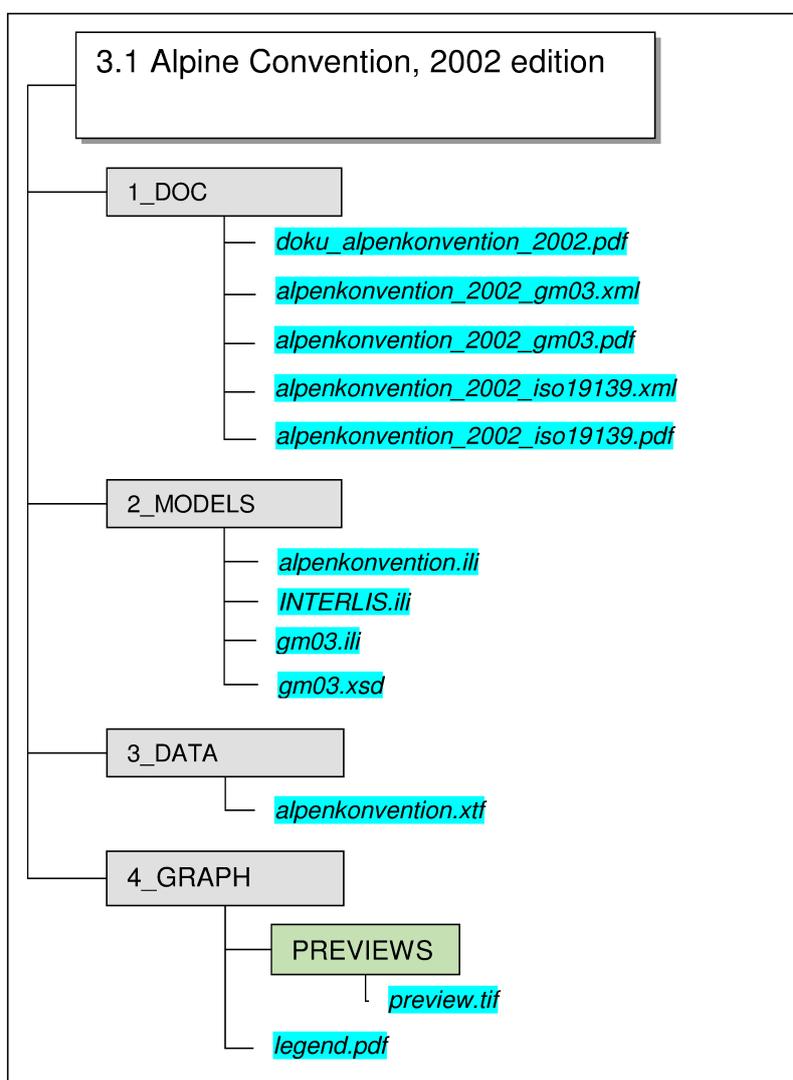
ID	Preview: Mandatory element	Must/can
4.2.2.8-1	Each geo-dossier has at least one preview image.	M

ID	Preview: Position	Must/can
4.2.2.8-2	Preview images are positioned below a grouping folder with the fixed file name PREVIEWS. The folder PREVIEWS is directly below the top folder 4_GRAPH.	M

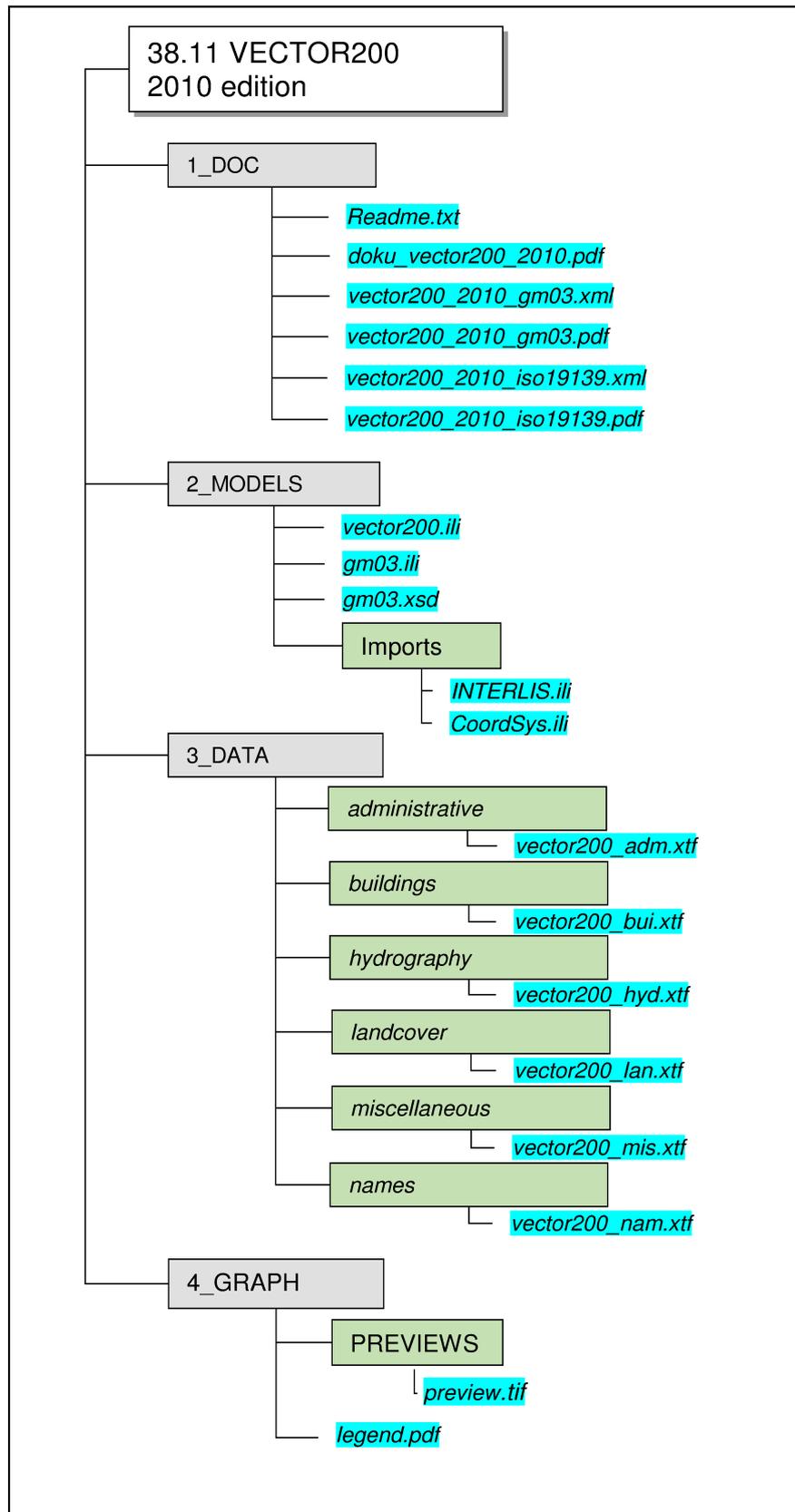
ID	Preview: Name	Must/can
4.2.2.8-3	The file names of the preview images can be freely chosen.	M

4.2.3 Examples

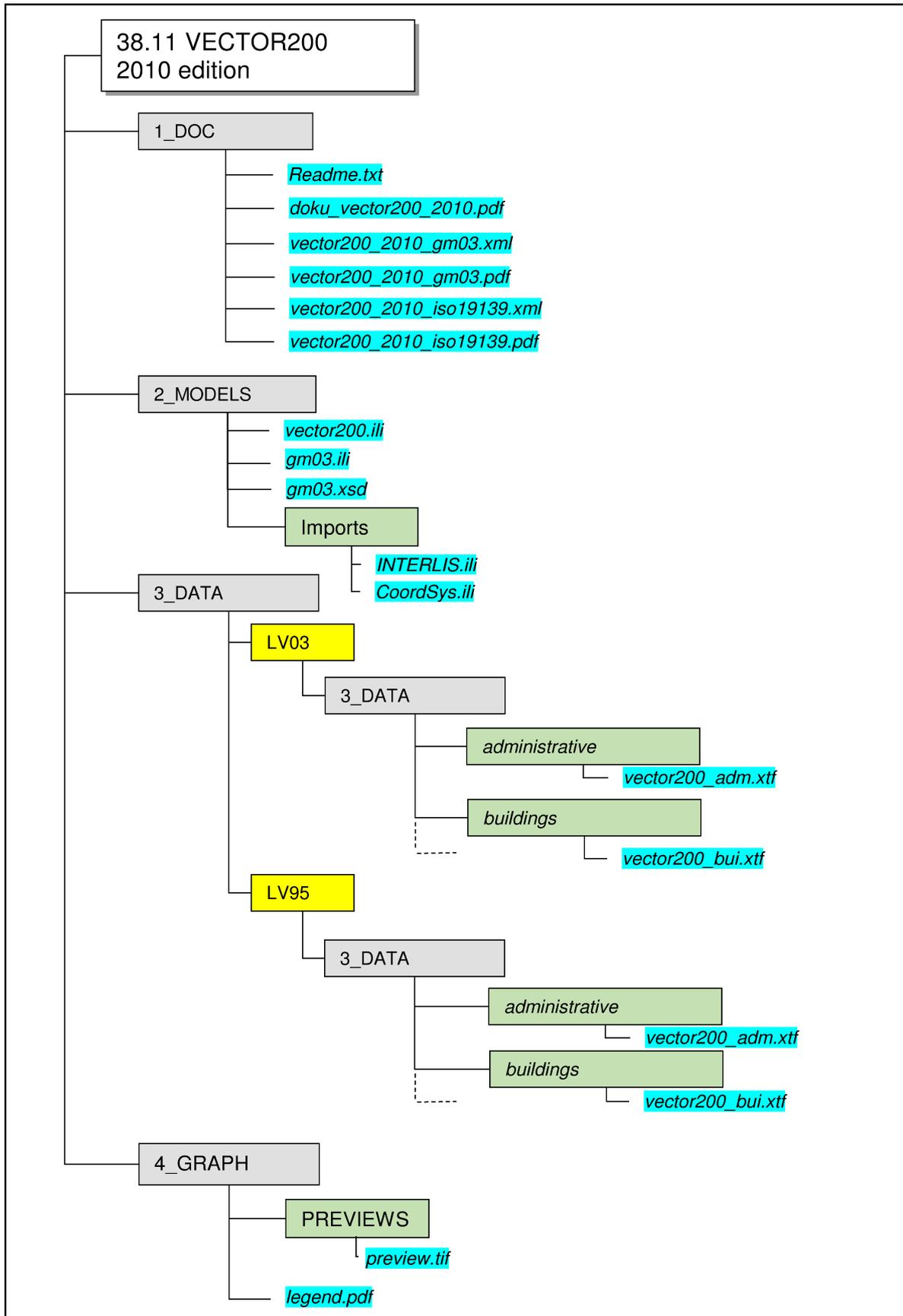
The following example shows the simplest possible structure of a geo-dossier. Only one single representation is used here. In this example, documenting of the structure via "Readme.txt" is not necessary.



In the example below, the option to divide using grouping folders is used in order to map the subdivision of a geodata set by layers and store the INTERLIS imports in a separate folder. However, it would also be permitted to dispense with the grouping folders “Imports”, “administrative”, “buildings” etc. and to store all elements of the type “File” (Units.ili, vector200_adm.xtf, vector200_bui.xtf etc.) directly below 2_MODELS and 3_DATA respectively.



The third example shows a data set that exists in two reference systems. Division using representation folders is mandatory here. Here, too, the division of the layers and imports using grouping folders is shown in the example is optional.



5 Glossary

Term	Description
Submitting authority	Refers to the authority or organisational unit that submits documents to the SFA. It is often (but not necessarily) identical to the records creator.
AIP	Archival Information Package: Archival Information Packages result from SIPs during the process of archiving digital documents. They represent the form of information packages in which digital documents are stored in the digital repository.
FSDI	Federal Spatial Data Infrastructure: The FSDI enables the efficient use and exchange of federal geodata.
File archive	Refers primarily to a quantity of files. Within the scope of digital archiving at the SFA, it is used for submissions containing files that are submitted without a classification system in the sense of records management using a GEVER system. However, the files may well be organised using another management system.
DIP	Dissemination Information Package: A DIP is a container for dossiers that are requested by a user via an ordering procedure.
Dossier	This term refers to all documents relating to a specific business matter. A dossier basically corresponds to a business matter. However, by combining similar business matters or dividing dossiers into subdossiers, this basic structure can be adapted to meet the corresponding needs. The compilation of dossiers is carried out on the basis of the classification system.
Metadata	Metadata can be described as “information about primary data” (data about data), since they have a descriptive nature.
Primary data	Primary data are data that are exclusively created directly by each records creator.
Representation	Representations are different versions of a single object, where each version has the same contextual information but in a different form. Geodata sets can be understood as representations of different quality, reference systems or combinations of these. Different time stamps are not representations.
SIP	Submission Information Package: SIPs are information packages that are submitted to the archive by the creating authorities. They contain digital documents (primary data and metadata).