SMURF (Semantically Marked Up Record Format) Specification for SFSB (Simple File-System Based) records

AUTHOR(S)	
Name(s)	Organisation(s)
Tarvo Kärberg	National Archives of Estonia (NAE)
Angela Dappert	DLM Forum Foundation
Andrew Wilson	University of Brighton
Levente Szilágyi	National Archives of Hungary (NAH)
Jože Škofljanec	Slovenian National Archives (SNA)
João Cardoso	Instituto Superior Técnico (IST)
Gregor Završnik	Slovenian National Archives (SNA)

REVIEWER(S)	
Name(s)	Organisation(s)
Kuldar Aas	National Archives of Estonia (NAE)
Alex Thirifays	Danish National Archives (DNA)
Andrew Wilson	University of Brighton

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

REVISION HISTORY AND STATEMENT OF ORIGINALITY

Submitted Revisions History

Revision No.	Date	Authors(s)	Organisation	Description
0.1	22 April 2015	Angela Dappert	DLM	Draft outline.
0.2	28 April 2015	Angela Dappert	DLM	Draft outline slightly updated.
0.3	14 August 2015	Angela Dappert	DLM	Incorporate issues from ERMS meetings.
0.35	16 October 2015	Tarvo Kärberg	NAE	Reorganising, incorporating feedback.
0.4	10 November 2015	Tarvo Kärberg	NAE	Updating the content, incorporating feedback.
0.5	12 November 2015	Andrew Wilson	UPHEC	Updating the content. Adding new information about SIP to AIP transformation.
0.6	16 November 2015	João Cardoso	IST	Updating the content.
0.7	17 November 2015	Levente Szilágyi	NAH	Updating the content. Adding information about metadata tables.
0.71	19 November 2015	Tarvo Kärberg	NAE	Incorporating feedback, cleaning the text, merging the content.
0.8	30 November 2015	Tarvo Kärberg	NAE	Merging the content.
0.9	13 January 2016	Tarvo Kärberg	NAE	Merging the content.
0.91	15 January 2016	Levente Szilágyi	NAH	Updating the content. Adding information about metadata tables.
0.92	19 January 2016	Jože Škofljanec	SNA	Updating the content related to SFSB records.
0.93	21 January 2016	Gregor Zavrsnik	SNA	Updating the content related to geodata.
0.94	22 January 2016	Levente Szilágyi	NAH	Updating the content related to EAD tables.
0.95	26 January 2016	Alex Thirifays	DNA	Quality assurance and proof-reading.
0.96	28 January 2016	Kuldar Aas	NAE	Quality assurance and proof-reading.
0.97	29 January 2016	Andrew Wilson	University of Brighton	Quality assurance and proof-reading.
1.0	29 January 2016	Tarvo Kärberg	NAE	**Final version** (part of D3.3)
1.1	20 July 2016	Tarvo Kärberg	NAE	Incorporating additional feedback from Andrew Wilson and Advisory Board.

1.2	30 September	Tarvo Kärberg	NAE	Incorporating agreements made in the Common Specification work group.
1.3	18 November	Tarvo Kärberg	NAE	Extracting SMURF SFSB as a separate document from the SMURF specification. Rearranging the content.
1.4	27 January 2017	Gregor Zavrsnik	SNA	Adding additional specifications for a SIP containing geodata

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Context

1.1 Purpose and Scope

The purpose of this document is to describe SMURF (semantically marked up record format) specification for SFSB (simple file-system based) records as described below. SFSB records mean records that are contained in simple file-system based folders or files, including those originating from content and data management systems, such as SharePoint. SFSBs exist, for example, in operating systems' native file systems designed to organize files and folders. The data requires manual enrichment with additional descriptive metadata in order to provide context and maintain authenticity. Additional pre-ingest tools (e.g. EAD editor[1]) may be required to acquire or create this metadata. Otherwise these records share the metadata specification and workflows of the general ERMS records, with possibly fewer[2] mandatory elements. For example, one specific sub-type of the SFSB profile is geodata which we will use later in this document to demonstrate the meaning of the SFSB content format.

1.2 Methodology

The specification is based on the definition by project partners of specific use cases and requirements to be implemented by E-ARK tools for simple file-system based records.

The work of gathering use cases and requirements involved discussions and feedback from several stakeholders:

- 6 national archives and 2 records' producers to specify the requirements.
- Archival service providers to develop and test the specifications and tools.

The involvement of so many stakeholders allowed us to learn from actual implementations and refine the specification accordingly.

The entities and metadata elements presented in the SMURF SFSB specification were inherited from the Common IP[3] and the SIP/AIP specifications, extracted from EAD3 standard, and captured from existing implementations in archives.

During the process of developing the specifications, decisions were made about:

- Which metadata elements to support
- How they are to be implemented

How did we make decisions about which metadata elements to adopt in the SMURF SFSB specification? We chose only those that were relevant for the SFSB scenario in order to meet submission use cases and

- in use in all archives in the E-ARK project or
- in use in most archives in the E-ARK project or
- required by national regulation and legislation or, to a lesser extent, required by policy decisions within the national archives and related institutions.

[1] EAD editor, https://github.com/ewg118/eaditor
[2] as SFSB does assume manual creation of metadata and thus includes less mandatory elements
[3] E-ARK Draft Common Specification for Information Packages in the E-ARK project, version 0.13, http://eark-project.com/resources/specificationdocs/50-draftcommonspec-1

Profile-dependent tools need to be used to add functionality to the pre-ingest process so that, for example, missing mandatory metadata can be manually added for the simple file-system based records case if the metadata is not submitted with the files (i.e. a SIP cannot consist only of computer files). The tools are intended to be used by the producer and need to be implemented in a way that is user-friendly for the irregular user.

1.3 Limitations

In this document the following are out of scope:

- The E-ARK AIP format (to be covered by a separate document, the E-ARK AIP specification).[4]
- The E-ARK DIP format (to be covered by a separate document, the E-ARK DIP specification).[5]

To simplify the analysis, the sections of this document do not discuss optimisations with respect to packaging and storage. The data model and metadata element definitions only discuss what information is needed, not how it is packaged, stored and optimised for handling.

2 Layered Data Model

This chapter introduces the structure of the data model. E-ARK uses a layered approach for information package definitions (Figure 1). The Common Specification for Information Packages (SIP, AIP, DIP) forms the outermost layer. The general SIP, AIP and DIP specifications add, respectively, submission, archiving and dissemination information to the Common Specification for Information Packages. These two layers are not part of this document. The third layer of the model represents specific content type specifications, such as this SMURF SFSB specification. Additional layers for business specific specifications and local variant implementations of any specification can be added.

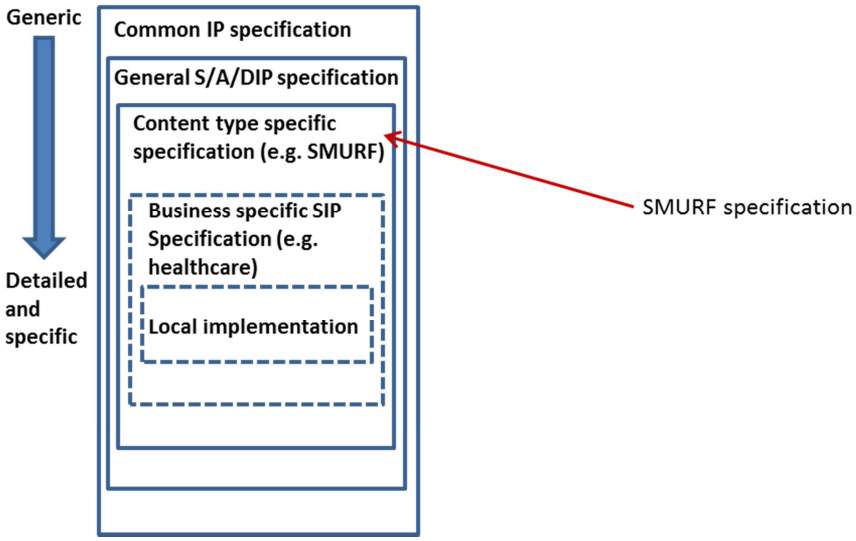

Figure 1: Data Model Structure

[4] E-ARK AIP pilot specification, released January 2016, http://eark-project.com/resources/project-deliverables/53-d43earkaipspec-1/file

[5] E-ARK DIP pilot specification, released April 2016, http://eark-project.com/resources/project-deliverables/61-d53-pilot-dip-specification/file

The SMURF specification omits all information that is specific to a business area (such as social security) or a specific content-type (such as healthcare).[6] However, these specific types of information may be needed by users of the E-ARK SMURF specification. This need is addressed by providing extension mechanisms in our IP specifications so that local (e.g. national) extensions to accommodate local requirements can be added by users.

Every level inherits metadata entities and elements from the higher levels. In order to increase adoption we will develop flexible schemas that will allow for extension points where the schema in each layer can be extended to accommodate additional information on the next specific layer until, finally, the local implementation can add specific entities or metadata elements to satisfy very specific local needs. Extension points can be implemented using a built-in extension mechanism via:

- Embedding foreign extension schemas (as supported by METS and PREMIS). These support both increasing the granularity of existing metadata elements by using more detailed data structures as well as adding new types of metadata.
- Single extra metadata elements (as supported by using <odd> metadata elements in EAD3) without the need to define foreign extension schemas.

The structure allows the addition of more detailed requirements for metadata entities e.g. by:

- Increasing the granularity of metadata elements by using more detailed data structures, or
- Adding local controlled vocabularies.

For consistency purposes we reuse design principles between layers as much as possible.

3 Metadata and Mapping

Metadata for a SIP can be obtained in several ways that are not mutually exclusive: automatically from the source system; extracted from the content; added manually during submission agreement or ingest. Ideally metadata should be created or captured as close to its source as possible where it can be most easily or exclusively obtained.

The balance of manual vs automated creation of metadata, as well as the origin of metadata (producer vs archive) varies to a great extent due to different best-practices and legal environments at the local level. However, in most cases the full SIP metadata is a mixture of metadata created manually and in automated fashion, from both the archive and the producer.

The work group has agreed upon the following principles regarding metadata standards in the E-ARK project:

- We allow for use of multiple metadata standards in a single SIP.
- We allow for the duplication of metadata in different standards if it eases information management or if it keeps related metadata elements together (e.g. all metadata about computer files should be recorded by using only one standard), but we keep duplication to a minimum.

[6] Although we do not describe specific content types in this deliverable, we still present geodata as an example of more specific content-oriented SFSB records type.

- We use EAD standard for descriptive metadata in the IP. Search in the archive can be based on these elements (plus additional ones in the catalogue).
- Archival metadata regarding classification, aggregation, disposal schedules and disposal holds will be represented using EAD for SMURF SFSB.
- We use EAD for descriptive metadata in the SIP for file system-based records.
- Different archive solutions can expose different subsets of the SIP metadata elements for search in the catalogue (i.e. the SIP format does not care/prescribe which metadata elements are going to be exposed for searching).
- All SIP metadata is to be treated as metadata rather than as content so that it can be made accessible to search in the archive. This means that it is submitted in the "metadata" folder rather than the "content" folder (i.e. we define only binary files as data, everything else is metadata).

NB. It is important to note that not all the metadata is a part of the current specification. Only the use of the metadata related to SMURF SFSB specification will be explained in the following metadata sections, but some of the metadata is a part of the general SIP specification and the Common Specification for Information Packages (i.e. METS metadata) as explained on page 7.

3.1 Building Aggregations for SFSB

An *aggregation* is any accumulation of record entities at a level above the record object (e.g. folder, series, fonds, etc., cf. Figure 2[7]).

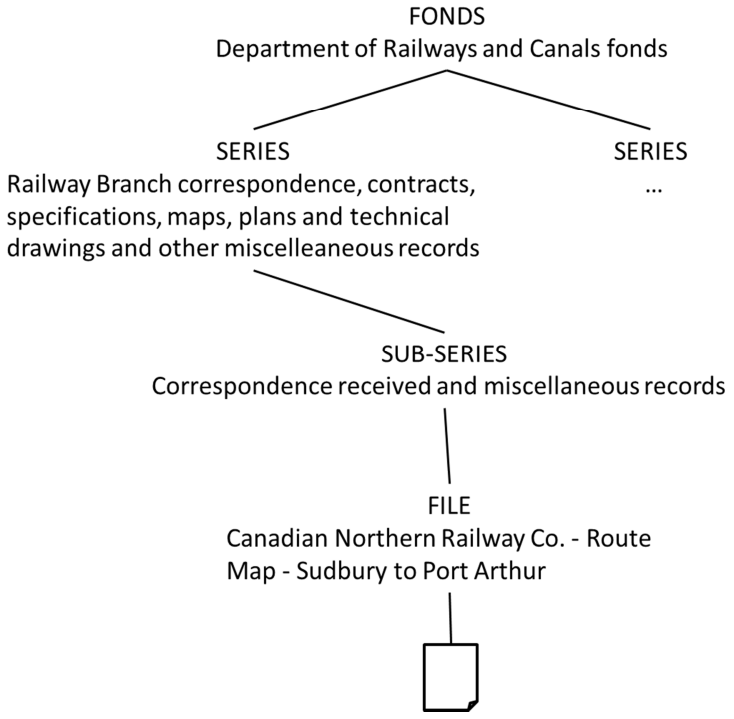

Figure 2: Aggregation

Aggregation is a way of structuring records to place them into the context of their creation and use.

[7] The example is based on ISAD(G) General International Standard Archival Description, http://www.icacds.org.uk/eng/ISAD(G).pdf.

Many of the aspects of the submissions from producers are governed by law or existing constraints. Producers may, for example, not be able to submit complete aggregation information with a record, or may not be able to choose whether they submit a single record or a whole series. In this case we may not specify a mandatory requirement for implementing tools in one specific way or specify rigid metadata structures that are mandatory.

It is not possible to anticipate all possible data dependencies between the producer and archive representations and our data model, metadata profile, workflows and requirements therefore cannot implement all possibilities. We are trying to capture the most common ones found in our stakeholder analysis. Instead of specifying a catch-all solution we provide guidelines for the most critical issues and solve some select cases in our pilot implementations. This provides a good starting point to anyone wanting to implement an aggregation creation for SFSB.

Aggregation structures for SFSB are to be represented using EAD3 metadata. EAD3 uses aggregation values as the "level" attribute on the elements <archdesc> and <c>, to specify the aggregation level at which description belongs (Example 1).

Example 1:

```
<archdesc level="fonds">
        …
        <dsc>
                <c level="series">
                        …
                        <c level="file">
                                Records and computer files
                        </c>
                </c>
        </dsc>
</archdesc>
```

Records are represented as <c level="item"> in EAD3. The relations between records and aggregations can be presented in a way as seen from the Example 2.

Example 2:

```
<archdesc level="fonds">
        <dsc>
        …
        <c level="file">
        …
                <c level="item">
                        <did>
                                …
                        </did>
                        <fileplan>
                                …
                        </fileplan>
                        …
                </c>
        </c>
        </dsc>
</archdesc>
```

As seen above, the record (<c level="item">) is hierarchically related with the aggregation (<c level="file">).

The exact names of aggregation levels depend on the agreements between data producers and archives. EAD3 has defined a set of values (class, collection, file, fonds, item, otherlevel, recordgrp, series, subfonds, subgrp, subseries) for that purpose, but it allows using other values as well if they are defined as "otherlevel" (Example 3).

Example 3:

```
<archdesc level="collection">
        …
        <dsc>
                <c level="series">
                        …
                        <c otherlevel="case"> <!--A new aggregation level-->
                                Records and computer files
                        </c>
                </c>
        </dsc>
</archdesc>
```

As the values of the localtype attribute are not prefixed then it is up to the implementers what values they exactly need to specify and support in their local tools. It may even be the case that this metadata is not used in an archive's EAD implementation, although E-ARK does not recommend such a policy.

3.2 Records and Content

The following section discusses the logic of incorporating SFSB records into a SIP and the minimum metadata needed for describing SFSB records.

The SFSB records scenario can occur in a variety of ways. For example, the producer has a shared drive with a set of computer files with (or without) archival descriptions. In order to include the computer files in the SIP we have to define an atomic intellectual entity. In this document the atomic entity for the SFSB profile is a "record", and records can be composed of components. In the SFSB scenario the components are computer files. We will include them in the data directory as required by the E-ARK Draft Common Specification for Information Packages[8] and make sure that we group the computer files into separate representations if needed.[9] For example, we may have the same content of a record represented by two different file formats (DOC and PDF/A) and we may want to include both computer files in the SIP. In that case we have to place them in two separate representation directories as seen in Figure 3.

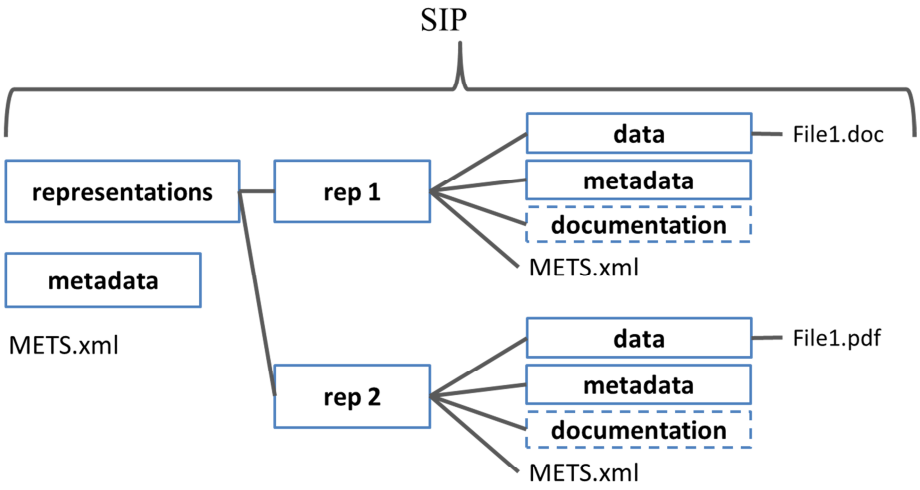

Figure 3: SFSB SIP

The *metadata* folder under representations will have to contain the archival descriptions if they are not already available on the root level. The metadata properties that apply to any SFSB record can be found in Appendices

Appendix I Metadata for SFSB Records. In order to provide search and access capabilities at least a very basic metadata should be recorded in EAD3 (Example 4)

Example 4:

```
<did>
        <!--Unique identifier for a record that is generated automatically by the system.-->
        <unitid localtype="current">ERA.14-4-3-1-1</unitid>
        <!--The identifying name or title of the record.-->
```

[8] E-ARK Draft Common Specification for Information Packages in the E-ARK project, version 0.13, page 23, http://eark-project.com/resources/specificationdocs/50-draftcommonspec-1

[9] This step will be performed manually.

```
        <unittitle>Final report</unittitle>
        <!--System set date and time when the entity was created. Can be represented as
        unitdatestructured or unitdate.-->
        <unitdate datechar="created">2004-05-25T00:00:00</unitdate>
        <!--The unit used to describe the extent of the record (e.g MB, pages, num of
        files/components)-->
        <physdescstructured coverage="whole" physdescstructuredtype="spaceoccupied">
                <quantity>0.4453</quantity>
                <unittype>MB</unittype>
        </physdescstructured>
        <!--<daoset> allows grouping multiple links to born digital records or digital representations of
        the materials being described. If only one link is present then the <daoset> element can be
        replaced by <dao>.-->
        <daoset label="Digital Objects">
                <dao … />
                <dao … />
        </daoset>
</did>
```

The <dao> element is a linking element that uses href to connect to born digital records or digital representations of the record (Example 5).

Example 5:

```
<dao id="dad603af-037b-44d5-8993-5754a42b3962" daotype="borndigital" linktitle="Report"
href="../../representations/rep1/data/Report.doc" />
```

The ID attribute represents a machine-processable unique identifier for the file. The daotype specifies if a file is born digital (borndigital), digitized from physical holdings (derived) or other (otherdaotype). For example, for scanned files we can point out that the files are not “original” (Example 6).

Example 6:

```
<dao id="baa703af-037b-44d5-8993-5754a42b3962" daotype="derived" linktitle="Page10"
href="../../representations/rep1/data/Page10.tif" />
```

In more complex cases (e.g. files migration) it is recommended to use both attributes (daotype, otherdaptype) as using only one attribute may be not enough for describing the file origin. For example, if a file is a result of a file format conversion from DOCX to PDF/A then we recommend using daotype="borndigital" otherdaotype="migrated" because the PDF/A is a born-digital and a migrated file in the same time (Example 7).

Example 7:

```
<dao id="dad603af-037b-44d5-8993-5754a42b3962" daotype="borndigital" linktitle="Report"
href="../../representations/rep1/data/Report.doc" />

<dao id="aad248bf-037b-34d5-7993-5754b42b3971" daotype="borndigital"
otherdaotype="migrated" linktitle="Report" href="../../representations/rep2/data/Report.pdf" />
```

The *documentation* folder is not mandatory, but it allows the inclusion of additional information about the content if needed.

The *METS* files can represent the metadata for a whole SIP package, or for a single representation.

3.2.1 Geodata

A specific example and subcase of the SFSB scenario is the archiving of geodata.

Geodata is a combination of graphical representation of objects in space and their descriptions or attributes. Increasingly, geospatial formats include geospatially focused datasets or databases that contain primary information about a geographic location.

Geodata generally comes in two forms, vector or raster, and can be stored as a set of files or a database. Therefore, it is possible to deal with geodata as a database (and apply the SIARD 2.0 specification) or as SFSB records. In this document we will consider the latter case.

3.2.1.1 Structure of a Geo SIP package

Following the E-ARK Draft Common Specification for Information Packages[10] at least one representation should be put in the representation directory (Figure 4).

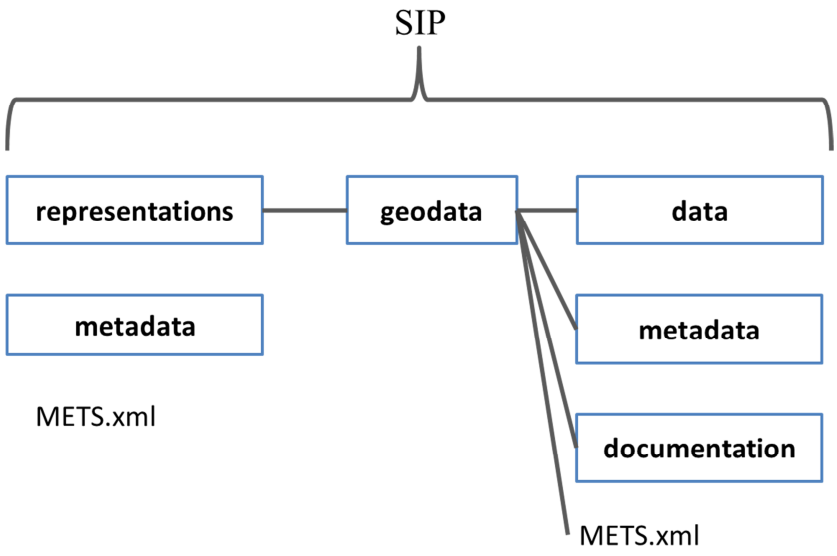

Figure 4: Geodata SIP

In the current case, the IP Package contains only one representation folder (in this case named 'geodata'- although the use of other names is allowed) and the additional documentation is stored within that representation.

Other scenarios are also possible, like multiple representations of the same geodata in the same IP and different possibilities of storing additional documentation:

[10] E-ARK Draft Common Specification for Information Packages in the E-ARK project, version 0.13, page 23, http://eark-project.com/resources/specificationdocs/50-draftcommonspec-1

3.2.1.1.1 GeoSIP containing multiple vector representations and documentation on the root

In this case, a GeoSIP package contains one representation in GML format and an original representation of the same data in ESRI Shapefile format. The representation with the GML file also contains the documentation specific to that representation – export log from shapefile to GML. All other documentation, required to properly interpret both representations is put on the top documentation folder (Figure 5). We could also include a logical link to point to additional documentation being stored in a different SIP (in case of a larger time series of the same data or similar records but from different organizational units).

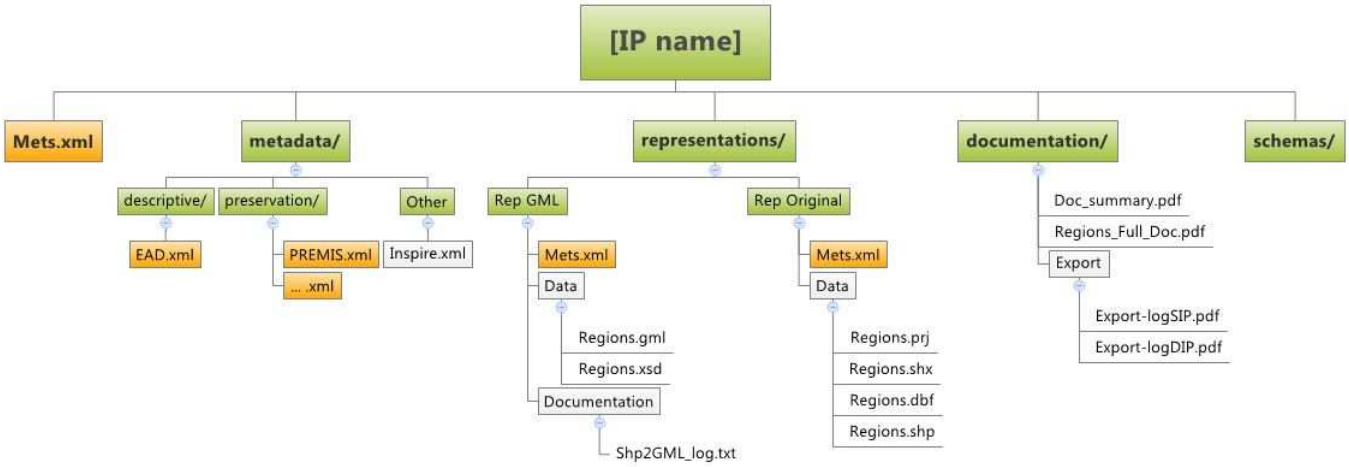

Figure 5: Folder structure of the GeoSIP containing multiple vector representations and documentation on the top

3.2.1.1.2 GeoSIP containing one representation of multiple raster datasets and documentation on the top

In this case, the IP package contains one representation of multiple raster images covering an area with an accompanying vector file – containing positions of the raster images (Figure 6). Documentation for the raster datasets is located in the top Documentation folder. In a case of a large volume of data we could split the data into multiple SIPs and record the organization of the split within the documentation folder.

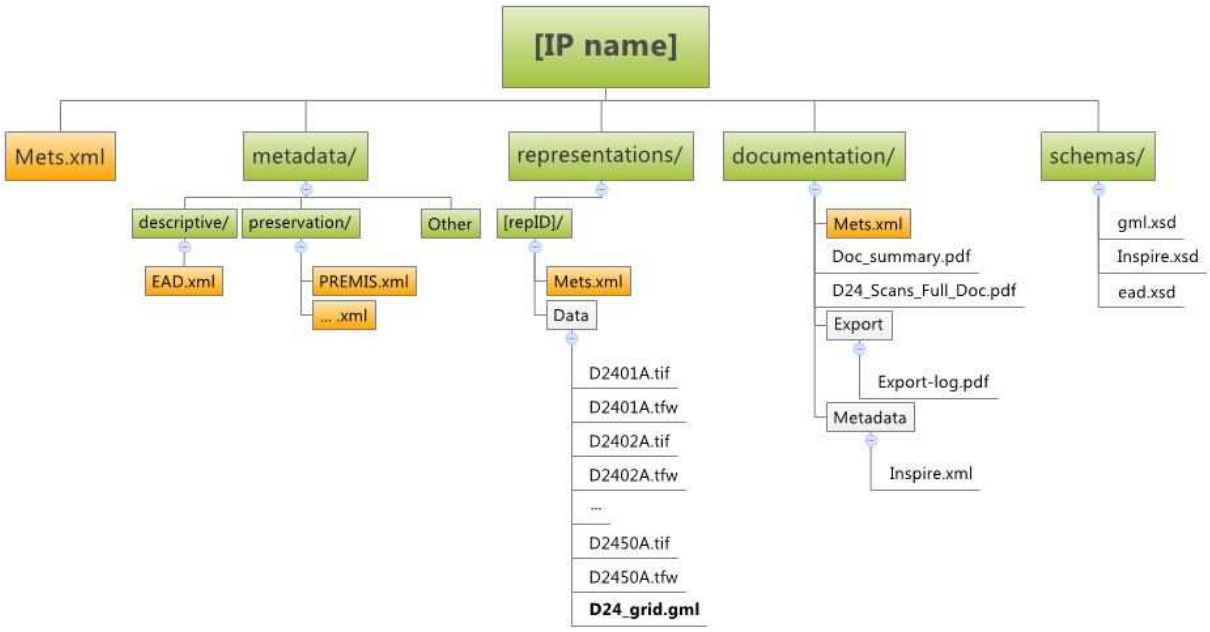

Figure 6: Folder structure of the GeoSIP containing one representation of multiple rasters and documentation on the top

3.2.1.2 Structuring geodata in the Data folder

The Data folder contains the geodata itself and all information that is needed to properly render the information (attributes, georefferencing information and visualization information). Geodata can stored in different formats, however every GeoSIP requires at least one representation which contains geodata in the required archival format – GML for vector data and GeoTIFF (or TIFF with external georefference file).

Graphical information:

Graphical information defines formats of geodata (vector or raster format). Vector data can be stored in different formats, such as SHP11, KML12, DXF13, GML14, etc. The GML format as defined by the ISO19136:2007 standard was chosen as the long term preservation format.

It is also possible to store a representation of the data in the original format if that format is still commonly used and well documented as is the case with the ESRI15 Shapefile (SHP) format. It is widely accepted and supported by most GIS software today.

Geodata in a raster format is stored in a "geo-enabled" raster file. Some common formats for raster geodata are GeoTIFF16, JPEG200017, BIM18, GRID19, ASCII GRID20, TIFF21, etc.

[11] http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf
[12] http://www.opengeospatial.org/standards/kml
[13] http://www.autodesk.com/techpubs/autocad/acad2000/dxf
[14] http://www.opengeospatial.org/standards/gml
[15] http://www.esri.com
[16] http://www.remotesensing.org/geotiff/spec/geotiffhome.html
[17] http://jpeg.org/jpeg2000
[18] http://www.buildingsmart.org
[19] ESRI Grid Format, http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=About_the_ESRI_Grid_format (Informal specification published by ESRI)

For long term preservation format we chose GeoTIFF or ordinary TIFF with a GML bounding box or an external xml file, containing georefferencing information like in the case of Swiss Archives.[22] Both GeoTIFF and GML bounding box need to contain the mandatory georeferencing information.

Attribute information

Attribute information can already be contained within the vector dataset itself. However there are cases when the accompanying attribute information can be contained within two or more tables. In that case, additional attribute tables need to be exported to archival format using the CSV format or the SIARD 2.0.

Relationships between the geodataset and accompanying tables needs to be documented in the "Table relationship" part of the accompanying documentation – described later in the chapter about Geodata documentation.

Attribute information in an ESRI Shapefile is stored in the mandatory *.dbf23 file. Attribute definitions are required in the documentation for interpretation purposes.

Attribute information in a Raster File can be stored in the value of the pixel itself. Attribute definitions are required if pixel value is not a grayscale value of an image.

Georeferencing information

Coordinate Reference System (CRS) information: CRS tells us how to locate objects in the geodata on the earth's surface. Elements of the spatial reference system are projection, geodetic datum, and unit of measure. All the elements can be defined by an EPSG code.[24]

Georeferencing information in GML is a mandatory part of the file itself and it is embedded in the geodata file itself.

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

The attribute "`srsName`" holds the value of the coordinate reference system code, in this example the code is 4326.

Georeferencing information in ESRI Shapefile (shp)

ESRI shapefile needs a <shapefilename>.prj file in order to be properly georeferenced. A prj file is a txt file, containing a definition of the coordinate reference system and all of its elements.

[20] http://desktop.arcgis.com/de/desktop/latest/manage-data/raster-and-images/esri-ascii-raster-format.htm
[21] http://partners.adobe.com/public/developer/tiff
[22] https://www.bar.admin.ch/dam/bar/en/dokumente/kundeninformation/Merkblatt%20TIFF%20EWF%20XML.pdf.download.pdf/Merkblatt%20TIFF%20EWF%20XML.pdf
[23] http://www.dbase.com/Knowledgebase/INT/db7_file_fmt.htm
[24] http://www.epsg.org

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

Georeferencing information in GeoTIFF

GeoTIFF contains the CRS information in its header. A GeoTIFF raster needs to be validated for the existence of CRS metadata within its header. Otherwise CRS needs to be added as for an ordinary TIFF format.

Georeferencing information in ordinary TIFF with GML bounding box

If geodata comes in an ordinary TIFF files it should be accompanied by an additional "world" file. For example a D240143.tif file would be accompanied by a D240143.tfw file. That is a txt file, containing information about the coordinates and size of the first top left pixel.

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

CRS information is then provided within the GML vector file that represents the location of the raster file in space and accompanies an ordinary TIFF.

Visualization information

Many GIS software applications enable export of the symbol definitions to a special file, however they may be in a proprietary XML structure (like the *.qml in QGIS, *.tab in MapInfo…) or even in a binary format (*.lyr in ArcGIS). The only OGC standard for symbols is the OGC Styled Layer Description XML format (sld files).

3.2.1.3 Metadata folder

Archival metadata for the GeoSIP are the same as for all other SFSB data. We only identified some specifics for the descriptive metadata (EAD3) based on existing metadata standards for Geodata (ISO 19115) and its adoption by the European Directive INSPIRE. If the producer has the INSPIRE metadata or the ISO 19115 based metadata, they should include them in the additional documentation, not in this folder.

When translating the 22 mandatory metadata for INSPIRE into EAD3, there were 2 elements that we needed to add as <odd> elements, since they were no appropriate mappings in the EAD3 structure.

This elements are:

- **Spatial resolution –** defining the positional accuracy of the geo dataset, which is derived from the acquisition method or the map resolution.
- **Lineage –** This 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.

Example: In the example below this is the description of lineage and spatial accuracy for an administrative units spatial dataset.

```
<odd>
  <head>Lineage</head>
    <p>This dataset was created by manualy digitisting the original cadastre maps in various
       scales (1:1000, 1:1440 and 1:2880). The whole dataset was created by merging individual
       units, that were digitised by different authors, since different municipalities had
       different service providers. The base datasets were created in the period from 1984 to
       1994 and were subsequently added to create the final dataset. Boundaries were
       generalised for a faster display.</p>
</odd>
<odd>
 <head>SpatialResolution</head>
   <p>5000</p>
</odd>
```

E-ARK project has identified the 22 mandatory contextual metadata elements, based on the EC Directive INSPIRE[25] which have been mapped to EAD3 elements (see

Appendix II Metadata for Specific Geodata Content Type) in order to enable automatic capture of descriptive metadata from the INSPIRE description of the spatial data sets.

3.2.1.4 Documentation folder

Since geodata is rarely in a form that is self-explanatory enough to be used by itself, we need additional information in order to enable the user to properly understand, interpret and use geodata. Therefore a checklist of mandatory and optional elements of additional documentation is used in the ingest process.

The required documentation for geodata can be within a GeoSIP or it can be linked from a diffrerent SIP in the same Series or Fond.

When creating a SIP, the producer needs to supply the following elements of documentation:

[25] The Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE),, http://inspire.ec.europa.eu/index.cfm/pageid/3

Information	Description	Cardinality
Documentation checklist	The documentation checklist is a document that needs to be present in the root of the documentation folder and which provides an index of locations of the required and optional information within the documentation folder. For instance, if we need to provide the *Feature catalogue, Logical Structure* and *Attribute* definitions, they can all exist within one document, or they can be in separate files. The documentation checklist serves as an index on geodata.	1..1
Export Section	If archived geodata was not stored in the archival format, any transformation into another format needs to be documented. This part of the documentation should contain detailed information about the export and transformation process in any part of the archival package lifecycle.	0..n
Feature Catalogue	The feature catalogue represents a logical structure of attributes. It provides a better understanding of the meaning, use and structure of the spatial data and provides a unified classification of spatial data in feature types (classes). Feature types are distinguished by their attributes (properties), by importance and by the relations between them.	0..n
Visualization	Data visualization provides an illustration and representation of spatial data. The catalogue of cartographic symbols is a collection of agreed cartographic symbols, which are used during the process of visualizing spatial data sets to display objects in space. Cartographic symbols are shown in the legend, which explains their meaning.	0..n
Logical structure of layers	The logical structure of layers shows the organization of the data layers at the level of logical tables contained in the database or in a connection of unstructured objects and their attributes, organized in a GIS application.	0..n
Attribute definition	Contains descriptions of the attributes for each data layer and the code values for each attribute (if not described in the **feature catalogue**).	1..n
Table relations	In the case of a complex system of tables, the documentation directory should contain diagrams of the relationships between tables in a database or within a GIS project, in order to enable the reconstruction of queries and provide greater understanding of the usage of tables.	0..n
Metadata	Besides the standard archival metadata like METS, PREMIS and EAD, geodata is often accompanied by geodata specific metadata. We expect this data to be in a standardized structure based on standards such as EN ISO 19115, or the EC INSPIRE directive. This data could in the future be accessed by the GIS tools available at the time. Therefore, we expect an Inspire.xml file or a number of them named after their geodata counterparts. If metadata is available in any other self-explainable form – like a text file or a pdf document, it should be referenced under this section.	0..n
Common queries	A list and a description of the most common queries provide additional information about how the data set was used in the production environment from the end user's point of view. The main goal is to enable the re-creation of the original functionality of tools, which enables users to get information in the form of common queries and common reports, as they were present in the original production environment.	1..n
Other contextual documentation	This chapter combines all the documents (links) to the relevant documentation describing the lineage and provenance of the spatial data set. The list of documentation includes: user manuals, related practices in EU and worldwide, methodological rules, scientific articles, publications, etc.	0..n

Table XX - Documentation elements in a geo DIP

The folder “Documentation” stores information which is not already included in the “Metadata” or “Data” folders. It will help archivists and end-users to understand the data-set in a wider social context and will provide a better understanding of the meaning, use and structure of the spatial data. The goal is to provide the end-user with all the necessary information for a proper understanding and interpretation of the geodata data set and to ensure provenance and authenticity.

3.3 The general logic of the mapping and conversion

The EAD file has 3 main inputs (Figure 5):

- Archival descriptions. Contains main archival descriptions (including metadata about aggregations and classification).
- Content. Contains links to computer files as <dao> elements.
- Additional metadata. Specific information that does not fit into the EAD3 standard elements can be saved as <odd> elements or localtype attributes in EAD3 elements (Example 8).

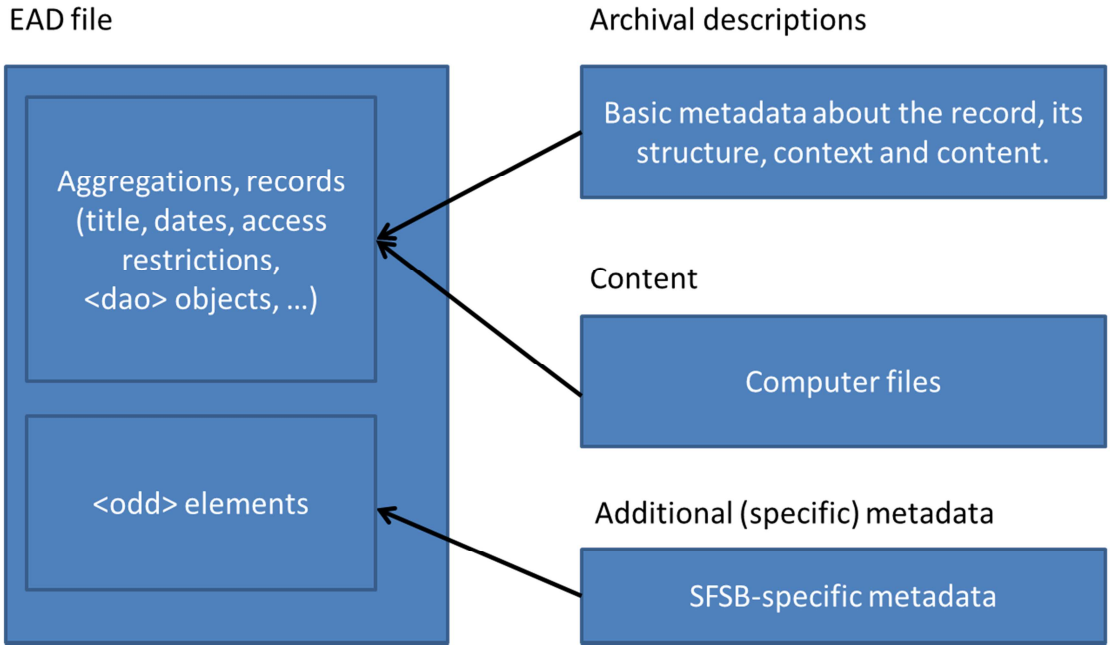

Figure 7: Inputs to the EAD file

Example 8:

```
<c level="item">
<did>
…
<!--System set date and time when the entity was created. “Created” is user-defined value.-->
<unitdate datechar="created">2004-05-25T00:00:00</unitdate>
…
<!--Other Descriptive Data. “Keyword” attribute is user-defined.-->
<odd localtype="Keyword">
        <list>
                <item>keyword A</item>
                <item>keyword B</item>
                <item>keyword C</item>
        </list>
</odd>
```

4 Glossary

Aggregation	Aggregations of records are accumulations of related record entities that, when combined, may exist at a level above that of a single record. Aggregations of records may reflect relationships such as shared characteristics or attributes, or the existence of sequential relationships between related records. [MoReq 2010, v 1.1]
AIP	Archival Information Package
Component	A part of a record that represents a discrete item of content. For completeness, a record, including all its components and their content, must be managed atomically.
DIP	Dissemination Information Package
EAD	Encoded Archival Description. A non-proprietary de facto standard for the encoding of finding aids for use in a networked (online) environment based on ISAD(G). Finding aids are inventories, indexes, or guides that are created by archival and manuscript repositories to provide information about specific collections. While the finding aids may vary somewhat in style, their common purpose is to provide detailed description of the content and intellectual organization of collections of archival materials. EAD allows the standardization of collection information in finding aids within and across repositories. http://www.loc.gov/ead/eadabout.html, accessed on 19 November 2015
EAC-CPF	Encoded Archival Context - Corporate bodies, Persons, and Families (EAC-CPF). A non-proprietary de facto standard for encoding the names of creators of archival materials and related information. EAC-CPF is based on ISAAR(CPF). See http://eac.staatsbibliothek-berlin.de/ [accessed on 6 August 2015]
IP	Information Package
Record	Any 'information created, received and maintained as evidence and information by an organisation or person, in pursuance of legal obligations or in the transaction of business (ISO 15489‑1:2001, 3.15)'. In MoReq2010®, a record may be further characterised as follows. • It has an extensible set of metadata that describe it. • It has one or more components that represent its content. • It is classified with a business classification. • It has a disposal schedule that describes explicitly if, how and when it will be disposed of or destroyed. • It belongs to an aggregation of records.

	• Access to it is controlled and limited to authorised users. • Its destruction may be prevented by a disposal hold. • It may be exported to another MCRS while retaining all of the characteristics listed above. [MoReq 2010, v 1.1]
SFSB	Simple file-system based records: records that exist in simple file-system based folders or files, including those originating from content and data management systems, such as SharePoint, that are not based on true file systems. They require the submission of computer files or folders from the file producers rather than from an ERMS. They require manual enrichment with additional descriptive metadata.
SIP	Submission Information Package
SMURF	**S**emantically **m**arked **u**p **r**ecords **f**ormat SMURF = ERMS (Electronic Records Management Systems) + SFSB (simple file-system based records)

5 Appendices

5.1 Appendix I Metadata for SFSB Records

The following table is based on the EAD3 standard. Further information on EAD elements can be found in TagLibrary-VersionEAD3.pdf available at http://www2.archivists.org/sites/all/files/TagLibrary-VersionEAD3.pdf

Table 1: Metadata for SFSB records

Element name	EAD element	Description and usage	Proposed cardinality	ISAD(g) 2.ed
ID of the Unit	<unitid>	**<unitid>** may contain any alpha-numeric text string that serves as a unique reference point or control number for the described material, such as a lot number, an accession number, a classification number, or an entry number in a bibliography or catalog. <unitid> is primarily a logical designation, which sometimes indirectly provides location information, as in the case of a classification number.	1..1	ISAD(G) 3.1.1
Title of the Unit	<unittitle>	**<unittitle>** is for recording the title statement, either formal or supplied, of the described materials. The title statement may consist of a word or phrase. <unittitle> is used at both the highest unit or <archdesc> level (e.g., collection, record group, or fonds) and at all the subordinate <c> levels (e.g., subseries, files, items, or other intervening stages within a hierarchical description).	1..1	ISAD(G) 3.1.2
Date of the Unit, Structured Date of the Unit	<unitdate>, <unitdatestructured>	**<unitdate>** is for indicating the date or dates the described materials were created, issued, copyrighted, broadcast, etc. <unitdate> may be in the form of text or numbers, and may consist of a single date, a date range, or a combination of single dates and date ranges; **<unitdatestructured>** provides a machine-processable statement of the date or dates the materials described were created, issued, copyrighted, broadcast, etc. <unitdatestructured> must contain one of the following child elements: <datesingle>, <daterange>, or <dateset>. <unitdatestructured> may contain only one child, therefore <dateset> must be used in situations where complex date information needs to be conveyed and requires at least two child elements. A date set may	1..*	ISAD(G) 3.1.3

Element name	EAD element	Description and usage	Proposed cardinality	ISAD(g) 2.ed
		combine two or more <datesingle> and <daterange> elements.		
Scope and Content	<scopecontent>	<scopecontent> contains a narrative statement that summarizes the range and topical coverage of the materials. It provides the researcher with the information necessary to evaluate the potential relevance of the materials being described. <scopecontent> may include information about the form and arrangement of the materials; dates covered by the materials; significant organizations, individuals, events, places, and subjects represented in the materials; and functions and activities that generated the materials being described. It may also identify strengths of or gaps in the materials.	0..1	ISAD(G) 3.3.1
Accruals	<accruals>	Used to indicate anticipated additions to the material already held by the repository. May indicate quantity and frequency of the accruals. **<accruals>** may also be used to indicate no additions are expected.	0..*	ISAD(G) 3.3.3
Conditions Governing Access	<accessrestrict>;	Record in **<accessrestrict>** information about the availability of the described materials, whether due to the nature of the information in the materials being described, the physical condition of the materials, or the location of the materials. Examples include restrictions imposed by the donor, legal statute, repository, or other agency, as well as the need to make an appointment with repository staff. May also indicate that the materials are not restricted;	1..*	ISAD(G) 3.4.1
Conditions Governing Use	<userestrict>	Use **<userestrict>** for information about any limitations, regulations, or special procedures imposed by a repository, donor, legal statute, or other agency. These conditions may be related to reproduction, publication, or quotation of the described materials after access to the materials has been granted. <userestrict> may also be used to indicate the absence of restrictions, such as when intellectual property rights have been dedicated to the public.	1..*	ISAD(G) 3.4.2
Language of the Material	<langmaterial>	**<langmaterial>** records information about languages and scripts represented in the materials being described. <langmaterial> must contain one or more <language> or <languageset> elements, but cannot contain text.	0..*	ISAD(G) 3.4.3
Related Material	<relatedmaterial>	**<relatedmaterial>** is used to identify associated materials in the same repository or elsewhere. These	0..*	ISAD(G) 3.5.3

Element name	EAD element	Description and usage	Proposed cardinality	ISAD(g) 2.ed
		materials may be related by sphere of activity, or subject matter.		
Description of Subordinate Components	<dsc>	Use **<dsc>** to wrap subordinate components in the archival hierarchy of the materials being described. Although <dsc> may repeat, it is recommended to include only a single <dsc> element. Because it is a wrapper element and not an essential part of archival description, <dsc> may be deprecated in future versions of EAD. Avoiding multiple <dsc> elements within an EAD instance will make future migrations simpler.	0..*	/
Component (Unnumbered)	<c>	As a wrapper for a set of elements, **<c>** provides information about the content, context, and extent of a subordinate body of materials. It is always a child or descendant of <dsc> and often a child and/or parent of another <c>. Each <c> identifies a logical section, or level, of the described materials. The physical filing separations between components need not always coincide with the intellectual separations. For example, a <c> that designates dramatic works might end in the same box in which the next <c> begins with short stories. Also, not every <c> directly corresponds to a folder or other physical entity. Some <c> elements simply represent a logical point in a hierarchical description.	0..1	/
Digital Archival Object	<dao>	<dao> is a linking element that uses @href to connect to born digital records or digital representations of the described materials. Digital representations may include graphic images, audio or video clips, images of text pages, and electronic transcriptions of text. The objects can be selected examples, or digital surrogates of all the materials in a collection, fonds, or an individual file.	1..*	/
Descriptive Identification Note	<didnote>	**<didnote>** can encode textual notes within <did> that are not more appropriately encoded in the other available elements.	0..1	ISAD(G) 3.6.1
Other Descriptive Data	<odd>	**<odd>** may be useful in converting legacy finding aids to the EAD format, by designating as "other" information that does not easily map to a more specific element. <odd> may be used when information about the described materials does not correspond to another element's definition, when the information is heterogeneous enough to make a single classification difficult, and when shifting the information to permit more specific content designation would be too costly or burdensome.	0..*	ISAD(G) 3.6.1

5.2 Appendix II Metadata for Specific Geodata Content Type

The following table is based on EAD3 standard and INSPIRE Metadata Implementing Rules: Technical Guidelines, based on EN ISO 19115 and EN ISO 19119 version 1.3. Further information on EAD elements can be found in TagLibrary-VersionEAD3.pdf available at http://www2.archivists.org/sites/all/files/TagLibrary-VersionEAD3.pdf and INSPIRE Metadata Implementing Rules at http://inspire.jrc.ec.europa.eu/documents/Metadata/MD_IR_and_ISO_20131029.pdf.

Table 2: Metadata for geodata

Element name	EAD element	Description and usage	Proposed cardinality	ISAD(g) 2.ed	INSPIRE el. No.	INSPIRE el. Name	INSPIRE Explanation
Title of the Unit	<unittitle>	**<unittitle>** is for recording the title statement, either formal or supplied, of the described materials. The title statement may consist of a word or phrase. <unittitle> is used at both the highest unit or <archdesc> level (e.g., collection, record group, or fonds) and at all the subordinate <c> levels (e.g., subseries, files, items, or other intervening stages within a hierarchical description).	1..1	ISAD(G) 3.1.2	2.1.1	Resource title	Name by which the cited resource is known.
Scope and Content	<scopecontent>	<scopecontent> contains a narrative statement that summarizes the range and topical coverage of the materials. It provides the researcher with the information necessary to evaluate the potential relevance of the materials being described. <scopecontent> may include information about the form and arrangement of the materials; dates covered by the materials; significant organizations, individuals, events, places, and subjects represented in the materials; and functions and activities that generated the materials being	0..1	ISAD(G) 3.3.1	2.2.2	Resource abstract	Brief narrative summary of the content of the resource(s).

Element name	EAD element	Description and usage	Proposed cardinality	ISAD(g) 2.ed	INSPIRE el. No.	INSPIRE el. Name	INSPIRE Explanation
		described. It may also identify strengths of or gaps in the materials.					
Other Record Identifier	<otherrecordid>	**<otherrecordid>** can be used to record an identifier that is an alternative to the mandatory identifier provided in <recordid>. These might include identifiers from systems that were used to generate the EAD instance or that are no longer current but had some part in the history and maintenance of the EAD instance.	0..1	/	2.2.5	Unique resource identifier	Value uniquely identifying an object within a namespace.
Related Material	<relatedmaterial>	**<relatedmaterial>** is used to identify associated materials in the same repository or elsewhere. These materials may be related by sphere of activity, or subject matter.	0..*	ISAD(G) 3.5.3	2.2.6	Coupled resource	Provides information about the datasets that the service operates on.
Language of the Material	<langmaterial>	**<langmaterial>** records information about languages and scripts represented in the materials being described. <langmaterial> must contain one or more <language> or <languageset> elements, but cannot contain text.	0..*	ISAD(G) 3.4.3	2.2.7	Resource language	Language(s) used within the datasets.
Subject	<subject>	**<subject>** indicates a topic reflected in the described materials.	1..1	/	2.3.1	Topic category	Main theme(s) of the dataset.
Other Descriptive Data	<odd>	Other descriptive metadata	0..1	/	2.3.2	Spatial data service type	A service type name from a registry of services

Element name	EAD element	Description and usage	Proposed cardinality	ISAD(g) 2.ed	INSPIRE el. No.	INSPIRE el. Name	INSPIRE Explanation
Index Entry	<indexentry>	A wrapper element that pairs an index term with zero or more linking elements. Each <indexentry> must contain an access element, such as <corpname>, <persname>, <subject>, etc., or <namegrp> to handle multiple access elements. It may also contain <ref>, <ptr>, or <ptrgrp> to identify and/or provide a link to the relevant position in the finding aid. If desired, use controlled vocabulary terms to facilitate access to information within and across finding aid systems.	0..*	/	2.4.1	Keyword value	Commonly used word(s) or formalised word(s) or phrase(s) used to describe the subject.
Geographic Coordinates	<geographiccoordinates>	Use **<geographiccoordinates>** to express a set of geographic coordinates such as latitude, longitude, and altitude representing a point, line, or area on the surface of the earth.	0..*	/	2.5.1	Geographic bounding box	Western-most coordinate of the limit of the dataset extent, expressed in longitude in decimal degrees (positive east). Eastern-most coordinate of the limit of the dataset extent, expressed in longitude in decimal degrees (positive east) Northern-most coordinate of the limit of the dataset extent, expressed in latitude in decimal degrees

Element name	EAD element	Description and usage	Proposed cardinality	ISAD(g) 2.ed	INSPIRE el. No.	INSPIRE el. Name	INSPIRE Explanation
							(positive north) Southern-most coordinate of the limit of the dataset extent, expressed in latitude in decimal degrees (positive north).
Date of the Unit, Structured Date of the Unit	<unitdate>, <unitdatestructured>	**<unitdate>** is for indicating the date or dates the described materials were created, issued, copyrighted, broadcast, etc. <unitdate> may be in the form of text or numbers, and may consist of a single date, a date range, or a combination of single dates and date ranges; **<unitdatestructured>** provides a machine-processable statement of the date or dates the materials described were created, issued, copyrighted, broadcast, etc. <unitdatestructured> must contain one of the following child elements: <datesingle>, <daterange>, or <dateset>. <unitdatestructured> may contain only one child, therefore <dateset> must be used in situations where complex date information needs to be conveyed and requires at least two child elements. A date set may combine two or more <datesingle> and <daterange> elements.	1..*	ISAD(G) 3.1.3	2.6.1	Temporal extent	Time period covered by the content of the dataset.
Date of the Unit, Structured Date of	<unitdate>,	**<unitdate>** is for indicating the date or dates the described materials were created, issued,	0..1	ISAD(G) 3.1.3	2.6.2	Date of	Reference date for the cited resource –

Element name	EAD element	Description and usage	Proposed cardinality	ISAD(g) 2.ed	INSPIRE el. No.	INSPIRE el. Name	INSPIRE Explanation
the Unit	<unitdatestructured>	copyrighted, broadcast, etc. <unitdate> may be in the form of text or numbers, and may consist of a single date, a date range, or a combination of single dates and date ranges; **<unitdatestructured>** provides a machine-processable statement of the date or dates the materials described were created, issued, copyrighted, broadcast, etc. <unitdatestructured> must contain one of the following child elements: <datesingle>, <daterange>, or <dateset>. <unitdatestructured> may contain only one child, therefore <dateset> must be used in situations where complex date information needs to be conveyed and requires at least two child elements. A date set may combine two or more <datesingle> and <daterange> elements.				publication	publication.
Date of the Unit, Structured Date of the Unit	<unitdate>, <unitdatestructured>	**<unitdate>** is for indicating the date or dates the described materials were created, issued, copyrighted, broadcast, etc. <unitdate> may be in the form of text or numbers, and may consist of a single date, a date range, or a combination of single dates and date ranges; **<unitdatestructured>** provides a machine-processable statement of the date or dates the materials described were created, issued, copyrighted, broadcast, etc. <unitdatestructured> must contain one of the following child elements:	0..1	ISAD(G) 3.1.3	2.6.3	Date of last revision	Reference date for the cited resource – revision.

Element name	EAD element	Description and usage	Proposed cardinality	ISAD(g) 2.ed	INSPIRE el. No.	INSPIRE el. Name	INSPIRE Explanation
		<datesingle>, <daterange>, or <dateset>. <unitdatestructured> may contain only one child, therefore <dateset> must be used in situations where complex date information needs to be conveyed and requires at least two child elements. A date set may combine two or more <datesingle> and <daterange> elements.					
Date of the Unit, Structured Date of the Unit	<unitdate>, <unitdatestructured>	**<unitdate>** is for indicating the date or dates the described materials were created, issued, copyrighted, broadcast, etc. <unitdate> may be in the form of text or numbers, and may consist of a single date, a date range, or a combination of single dates and date ranges; **<unitdatestructured>** provides a machine-processable statement of the date or dates the materials described were created, issued, copyrighted, broadcast, etc. <unitdatestructured> must contain one of the following child elements: <datesingle>, <daterange>, or <dateset>. <unitdatestructured> may contain only one child, therefore <dateset> must be used in situations where complex date information needs to be conveyed and requires at least two child elements. A date set may combine two or more <datesingle> and <daterange> elements.	0..1	ISAD(G) 3.1.3	2.6.4	Date of creation	Reference date for the cited resource – creation.
Other Descriptive	<odd>		0..1	/	2.7.2	Spatial	• Equivalent scale: level of detail

Element name	EAD element	Description and usage	Proposed cardinality	ISAD(g) 2.ed	INSPIRE el. No.	INSPIRE el. Name	INSPIRE Explanation
Data						resolution	expressed as the scale denominator of a comparable hardcopy map or chart • Distance: ground sample distance
Conditions Governing Access	<accessrestrict>;	Record in **<accessrestrict>** information about the availability of the described materials, whether due to the nature of the information in the materials being described, the physical condition of the materials, or the location of the materials. Examples include restrictions imposed by the donor, legal statute, repository, or other agency, as well as the need to make an appointment with repository staff. May also indicate that the materials are not restricted;	1..1	ISAD(G) 3.4.1	2.9.1	Limitations on public access [and use]	Access constraints applied to assure the protection of privacy or intellectual property, and any special restrictions or limitations on obtaining the resource. Limitations on public access: - Access constraints - Example: otherRestrictions (limitation not listed). - Other constraints - Example: No limitations - Classification - Example: unclassified

Element name	EAD element	Description and usage	Proposed cardinality	ISAD(g) 2.ed	INSPIRE el. No.	INSPIRE el. Name	INSPIRE Explanation
	<accessrestrict>;	Record in **<accessrestrict>** information about the availability of the described materials, whether due to the nature of the information in the materials being described, the physical condition of the materials, or the location of the materials. Examples include restrictions imposed by the donor, legal statute, repository, or other agency, as well as the need to make an appointment with repository staff. May also indicate that the materials are not restricted;	1..1	ISAD(G) 3.4.1	2.9.2	Conditions applying to access and use	Restrictions on the access and use of a resource or metadata.
Conditions Governing Use	<userestrict>	Use **<userestrict>** for information about any limitations, regulations, or special procedures imposed by a repository, donor, legal statute, or other agency. These conditions may be related to reproduction, publication, or quotation of the described materials after access to the materials has been granted. <userestrict> may also be used to indicate the absence of restrictions, such as when intellectual property rights have been dedicated to the public.	1..1	ISAD(G) 3.4.1; ISAD(G) 3.4.2; ISAD(G) 3.4.4	2.9.2	Conditions applying to access and use	Restrictions on the access and use of a resource or metadata.
ID of the Unit	<unitid>	**<unitid>** may contain any alpha-numeric text string that serves as a unique reference point or control number for the described material, such as a lot number, an accession number, a classification number, or an entry number in a bibliography or catalog. <unitid> is primarily a logical designation, which sometimes indirectly provides location	1..1	ISAD(G) 3.1.1			

Element name	EAD element	Description and usage	Proposed cardinality	ISAD(g) 2.ed	INSPIRE el. No.	INSPIRE el. Name	INSPIRE Explanation
		information, as in the case of a classification number.					
Accruals	<accruals>	Used to indicate anticipated additions to the material already held by the repository. May indicate quantity and frequency of the accruals. **<accruals>** may also be used to indicate no additions are expected.	0..*	ISAD(G) 3.3.3			
Description of Subordinate Components	<dsc>	Use **<dsc>** to wrap subordinate components in the archival hierarchy of the materials being described. Although <dsc> may repeat, it is recommended to include only a single <dsc> element. Because it is a wrapper element and not an essential part of archival description, <dsc> may be deprecated in future versions of EAD. Avoiding multiple <dsc> elements within an EAD instance will make future migrations simpler.	0..*	/			
Component (Unnumbered)	<c>	As a wrapper for a set of elements, **<c>** provides information about the content, context, and extent of a subordinate body of materials. It is always a child or descendant of <dsc> and often a child and/or parent of another <c>. Each <c> identifies a logical section, or level, of the described materials. The physical filing separations between components need not always coincide with the intellectual separations. For example, a <c> that designates dramatic works might end in the same	0..1	/			

Element name	EAD element	Description and usage	Proposed cardinality	ISAD(g) 2.ed	INSPIRE el. No.	INSPIRE el. Name	INSPIRE Explanation
		box in which the next <c> begins with short stories. Also, not every <c> directly corresponds to a folder or other physical entity. Some <c> elements simply represent a logical point in a hierarchical description.					
Digital Archival Object	<dao>	<dao> is a linking element that uses @href to connect to born digital records or digital representations of the described materials. Digital representations may include graphic images, audio or video clips, images of text pages, and electronic transcriptions of text. The objects can be selected examples, or digital surrogates of all the materials in a collection, fonds, or an individual file.	1..*	/			
Descriptive Identification Note	<didnote>	**<didnote>** can encode textual notes within <did> that are not more appropriately encoded in the other available elements.	0..1	ISAD(G) 3.6.1			
Other Descriptive Data	<odd>	**<odd>** may be useful in converting legacy finding aids to the EAD format, by designating as "other" information that does not easily map to a more specific element. <odd> may be used when information about the described materials does not correspond to another element's definition, when the information is heterogeneous enough to make a single classification difficult, and when	0..*	ISAD(G) 3.6.1			

Element name	EAD element	Description and usage	Proposed cardinality	ISAD(g) 2.ed	INSPIRE el. No.	INSPIRE el. Name	INSPIRE Explanation
		shifting the information to permit more specific content designation would be too costly or burdensome.					