General SIP Specification

AUTHOR(S)	
Name(s)	Organisation(s)
Tarvo Kärberg	National Archives of Estonia (NAE)
Anders Bo Nielsen	Danish National Archives (DNA)
Björn Skog	ES Solutions (ESS)
Gregor Zavrsnik	Slovenian National Archives (SNA)
Hélder Silva	KEEP Solutions (KEEPS)
Karin Bredenberg	National Archives of Sweden / ES Solutions (ESS)
Kathrine Hougaard Edsen Johansen	Danish National Archives (DNA)
Levente Szilágyi	National Archives of Hungary (NAH)
Phillip Mike Tømmerholt	Danish National Archives (DNA)

REVIEWER(S)	
Name(s)	Organisation(s)
Kuldar Aas	National Archives of Estonia (NAE)
Sven Schlarb	Austrian Institute of Technology (AIT)
David Anderson	University of Brighton
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	20.10.2014	Tarvo Kärberg	NAE	First draft.
0.2	13.11.2014	Tarvo Kärberg	NAE	Updating content.
0.3	02.12.2014	Tarvo Kärberg	NAE	Updating content.
0.4	17.01.2015	Tarvo Kärberg	NAE	Updating content.
0.5	21.01.2015	Karin Bredenberg	ESS	Updating content.
0.6	23.01.2015	Anders Bo Nielsen	DNA	Updating content.
0.7	23.01.2015	Kathrine Hougaard Edsen	DNA	Updating content.
0.71	26.01.2015	Björn Skog	ESS	Updating content.
0.72	27.01.2015	Hélder Silva	KEEPS	Updating content.
0.8	27.01.2015	Angela Dappert	DLM/UPHEC	Quality assurance and proof-reading.
0.9	29.01.2017	Kuldar Aas	NAE	Quality assurance and proof-reading.
0.91	30.01.2015	David Anderson	UPHEC	Quality assurance and proof-reading.
1.0	30.01.2015	Tarvo Kärberg	NAE	**Final version (D3.2).**
0.1	11.05.2015	Karin Bredenberg	ESS/NAS	Updating content.
0.2	30.06.2015	Tarvo Kärberg	NAE	Updating content.
0.3	27.07.2015	Tarvo Kärberg	NAE	Updating content.
0.4	23.10.2015	Tarvo Kärberg	NAE	Updating content, synchronising with the SMURF profile.
0.41	17.11.2015	Tarvo Kärberg	NAE	Integrating the feedback.
0.42	07.12.2015	Tarvo Kärberg	NAE	Updating content.
0.5	12.01.2016	Tarvo Kärberg	NAE	Updating content, synchronising with the Common Specification.
0.6	15.01.2016	Anders Bo Nielsen	DNA	Updating content.
0.61	15.01.2016	Gregor Zavrsnik	SNA	Updating content.
0.62	18.01.2016	Tarvo Kärberg	NAE	Updating content.
0.63	20.01.2016	Phillip Mike Tømmerholt	DNA	Updating content.
0.64	25.01.2016	Phillip Mike Tømmerholt	DNA	Updating content.
0.7	26.01.2016	Sven Schlarb	AIT	Quality assurance and proof-reading.
0.8	27.01.2016	Kuldar Aas	NAE	Quality assurance and proof-reading.
0.9	29.01.2016	Andrew Wilson and David Anderson	University of Brighton	Quality assurance and proof-reading.
1.0	29.01.2016	Tarvo Kärberg	NAE	**Final version** (general part of D3.3)
1.1	14.07.2016	Tarvo Kärberg	NAE	Incorporating agreements made in the Common Specification work group.
1.2	12.12.2016	Tarvo Kärberg	NAE	Incorporating agreements made in the Common Specification work group.
1.3	13.01.2017	Tarvo Kärberg	NAE	Small updates.
1.4	31.01.2017	Tarvo Kärberg	NAE	**Finalising the specification.**

Statement of originality:
This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

EXECUTIVE SUMMARY

According to the Open Archival Information System Reference Model (OAIS) every submission of information to an archive by a producer occurs as one or more discrete transmissions of submission information packages.[1] Unfortunately there is currently no central SIP format which would cover all national and business needs as identified in the E-ARK Report on Available Best Practices.[2] The E-ARK project acknowledged this problem and developed a solution in the form of the SIP format which is described in this document.

The first outcome of this work was Deliverable 3.2: E-ARK SIP Draft Specification.[3] This gives an overview of the structure and main metadata elements for the SIP and provides initial input for the technical implementations of pre-ingest and ingest tools. It was followed by Deliverable 3.3 which extends the previous one by providing a revised version of the D3.2 content, adding more details relevant for tool development and implementation, and describing specific profiles for the transfer of relational databases, electronic records management systems (ERMS) and simple file system based records (SFSB).

The target group for this document are records creators, archival institutions and software providers creating or updating their SIP format specifications. This document is also important for electronic records management systems (ERMS) providers as it presents a standardised profile for exporting records and metadata out of their systems.

This document provides an overview of:

- The general structure for Submission Information Packages.
  This chapter explains how records creators should construct/structure their SIPs in order to meet the requirements of the SIP specification and achieve interoperability by following the common rules for all information packages (SIPs, AIPs, DIPs) as described in the Common Specification for Information Packages[4].
- General SIP metadata.
  This chapter provides a detailed overview of metadata sections and the metadata elements in these sections. The tables with all metadata elements could possibly be of interest to technical stakeholders who wish to implement the SIP.
- Content Information Type Specifications.
  This section introduces profiles for SMURF (Semantically Marked Up Records Format) and relational databases. The profiles themselves are separate documents.

[1] Reference Model for an Open Archival Information System (OAIS), 2012, public.ccsds.org/publications/archive/650x0m2.pdf
[2] Deliverable D3.1 E-ARK Report on Available Best Practices, 2014, http://eark-project.com/resources/project-deliverables/6-d31-e-ark-report-on-available-best-practices
[3] Deliverable 3.2 E-ARK SIP Draft Specification, 2015, http://eark-project.com/resources/project-deliverables/17-d32-e-ark-sip-draft-specification
[4] Common Specification for Information Packages, version 0.17, http://eark-project.com/resources/specificationdocs/67-e-ark-draft-common-specification-ver-017

- The submission agreement.
  This chapter provides an overview of submission agreement usages and recommended metadata elements.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENT

The authors of this deliverable would like to thank all national archives, tool developers, the Advisory Board of the E-ARK project and other stakeholders who provided valuable knowledge about their submission information packages and feedback to E-ARK deliverables.

A special gratitude goes to the National Archives of Sweden whose FGS (Förvaltningsgemensam Specifikation) structure influenced the first version of the SIP METS profile development significantly.

The authors would also like to express their gratitude to the team behind the Common Specification for Information Packages document for their enormous effort in agreeing common principles for submission, archival and dissemination packages.

1. INTRODUCTION

1.1. Scope and purpose

This document is a core / general SIP specification which is guided by the following hierarchical model (see Figure 1):

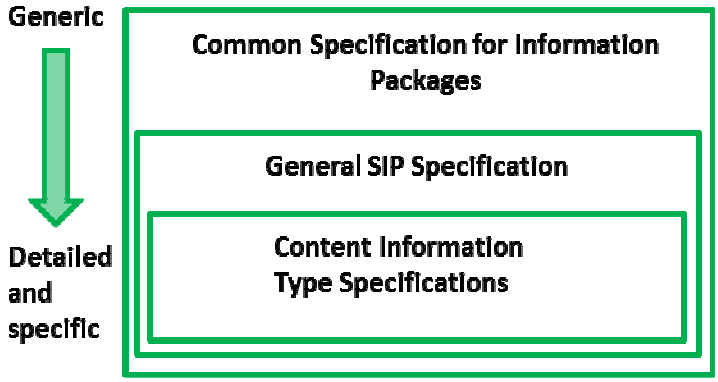

Figure 1: Relations between specifications

- **Common Specification for Information Packages (CSIP)** identifies and standardises the common aspects of information packages (SIP/AIP/DIP) which are equally relevant and implemented by any of the functional entities of the overall digital preservation process (i.e. pre-ingest, ingest, long-term preservation and access). CSIP is a separate document. Therefore, the current specification does not aim largely repeating the information presented there – only the information that is absolutely necessary to understand the SIP specification will be mentioned here.
- General SIP Specification. This is the current document which describes the SIP package structure and minimum set of required metadata for SIP delivery to the archive.
- Content Information Type Specifications are content-dependent specifications which include detailed information on how content, metadata, and documentation for specific content types (for example ERMS or relational databases) can to be handled within the SIP. At the moment, there are 3 such specifications:
    - SIARD 2.0 for relational databases[5]
    - SMURF ERMS for electronic records management systems[6]
    - SMURF SFSB for simple file system based records[7]

[5] The SIARD 2.0 specification for relational databases can be found at http://eark-project.com/resources/specificationdocs/32-specification-for-siard-format-v20
[6] The SMURF profile for ERMS can be found https://github.com/DLMArchivalStandardsBoard/SMURF/tree/master/spec.

1.2. Related work

This document is based on or influenced by the following documents and best practices:

- **Deliverable D3.1** E-ARK Report on Available Best Practices, 2014, http://eark-project.com/resources/project-deliverables/6-d31-e-ark-report-on-available-best-practices

  D3.1 was one of the inputs to the deliverable D3.2 and the D3.2 to the D3.3.

- **Deliverable D2.1** General pilot model and use case definition, 2014, http://eark-project.com/resources/project-deliverables/5-d21-e-ark-general-pilot-model-and-use-case-definition.

  We have developed the SIP specification to support the workflows defined in the general model.

- **FGS package structure**, 2013, https://riksarkivet.se/Media/pdf-filer/Projekt/FGS_Earkiv_Paket.pdf
  This specification was one of the main inputs for the first draft SIP specification. The newest version (https://riksarkivet.se/Media/pdf-filer/doi-t/FGS_Paketstruktur_RAFGS1V1.pdf) was also investigated in the SIP definition process.

- **Reference Model for an Open Archival Information System** (OAIS), 2012, public.ccsds.org/publications/archive/650x0m2.pdf

  We have used the same terminology as introduced in the OAIS model and also the same division of information package types: Submission Information Package (SIP), Archival Information Package (AIP), Dissemination Information Package (DIP).

- **Producer-Archive Interface Methodology Abstract Standard** (PAIMAS), 2004, public.ccsds.org/publications/archive/651x0m1.pdf

  We have looked at the four phases (Preliminary, Formal Definition, Transfer, Validation) of PAIMAS, their aims and expected results and decided to support the phases as far as possible with the current specification. Furthermore, the requirements for the submission agreement were influenced by the PAIMAS standard.

- **Producer-Archive Interface Specification** (PAIS) – CCSDS, 2014, public.ccsds.org/publications/archive/651x1b1.pdf

  We have investigated the structure of a SIP presented in PAIS, but as the implementation of this specification is far from comprehensive (only few prototypes exist), we decided to rely more on the best practices introduced in the best practice report.

[7] The SMURF profile for SFSB can be found at https://github.com/DLMArchivalStandardsBoard/SMURF/tree/master/spec.

- **e-SENS** (Electronic Simple European Networked Services) project, http://www.esens.eu/

  We have investigated the e-Delivery and e-Documents related work in e-SENS and made sure that our work is neither duplicating the work done there nor producing any conflicts between deliverables.

- **Deliverables D3.2** E-ARK SIP Draft Specification, 2015, http://eark-project.com/resources/project-deliverables/17-d32-e-ark-sip-draft-specification and D3.3 E-ARK SIP Pilot Specification, 2016, http://eark-project.com/resources/project-deliverables/51-d33pilotspec

2. GENERAL STRUCTURE AND DATA MODEL FOR SUBMISSION INFORMATION PACKAGES

The SIP specification follows the general structure which is common for all information packages. The SIP data model describes the package submitted to the archive, which consists of representations (submitted data and metadata) and metadata as seen in Figure 2[8] and mandated/required by the SIP, AIP and DIP formats.

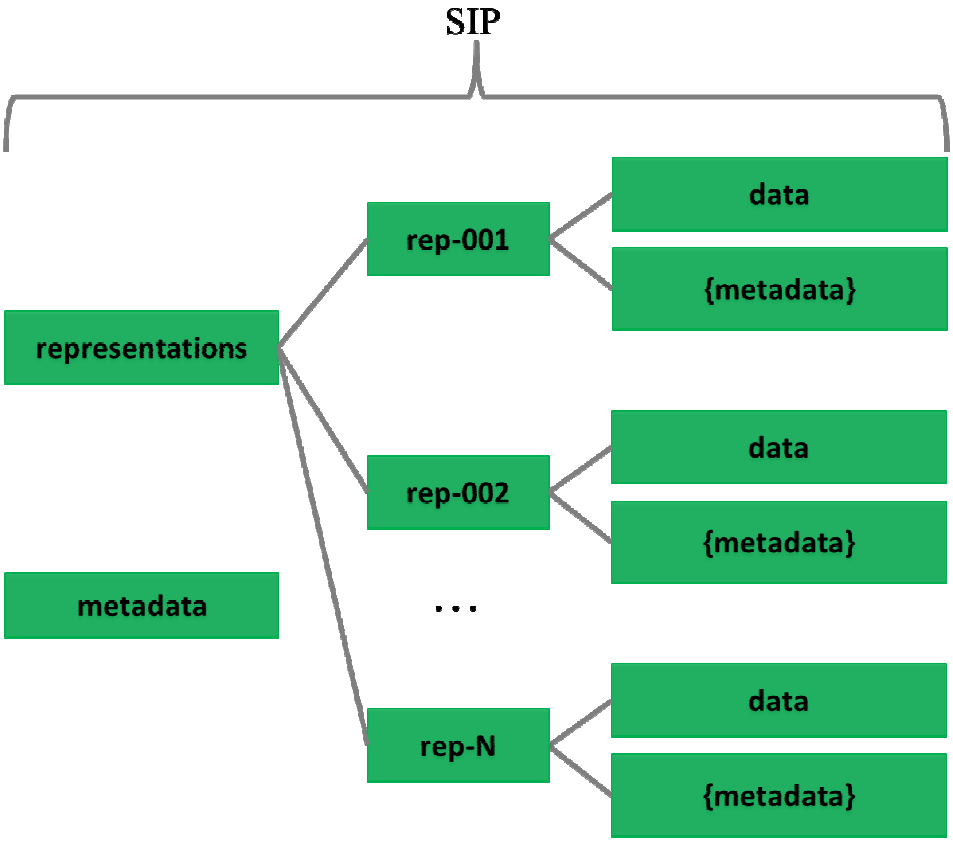

Figure 2: SIP data model

As one SIP can contain more than one representation[9] of the same intellectual entity then it is reasonable to separate different representations (e.g. Rep-001 and Rep-002 under Representations). This requires additional metadata about the SIP. If we store all metadata (even about representations) at the IP level then we do not need to use the Metadata folder at the representation level. In this case, the Metadata directory under representations is considered optional, as are:

[8] This is a conceptual model and does not describe the actual implementation structure.

[9] Digital Object or physical object instantiating or embodying an Intellectual Entity. A Representation that is a Digital Object is the set of stored Files and Structural Metadata needed to provide a complete rendition of the Intellectual Entity. PREMIS Data Dictionary (full document), Version 3.0, 2015, http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf

- Documentation folder – for including additional documents that explain the content or its use (e.g. user manual).
- Schemas folder – for adding schemas for the XML files in the data/metadata directly into the package.

According to the Common Specification for Information Packages, the information package folder must include a mandatory core metadata file called ”METS.xml”, which includes core information needed to identify and describe the structure of the package itself and the rest of its components. Therefore the minimal general SIP structure is as follows (Figure 3). The name of a representation can be freely chosen.

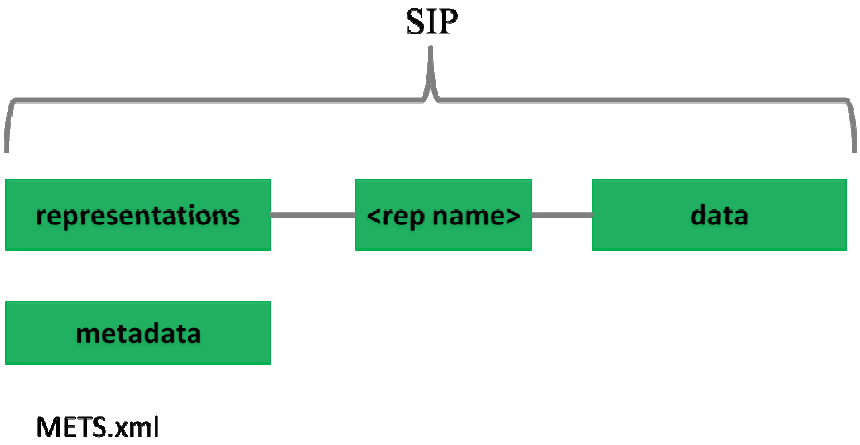

Figure 3: Minimal SIP structure

If needed, a METS.xml file can be present under representations as well to handle scalability issues. This proposed extended IP structure using divided METS files, is introduced in the Common Specification for Information Packages and in deliverable D4.3 E-ARK AIP pilot specification[10] to more easily manage the splitting of large packages using a divided METS structure.

The detailed folder structure of a SIP will also be present and agreed upon in the submission agreement (page 41) by indicating the data model for the submission. Also the details of the internal structure of the data and metadata folders can be further specified in submission agreements.

The metadata model for the SIP will be multi-layered by starting from general common metadata elements and finishing with optional local elements as explained previously.[11]

[10] E-ARK AIP pilot specification, released January 2016, http://eark-project.com/resources/project-deliverables

[11] Please note that the business specific (e.g. healthcare records) or local implementation based metadata is not discussed in this specification. As the specifications can be undertaken at different scales, with different types of data and locations, with their constituent technical components, more detailed or localised specifications may be needed.

3. GENERAL SIP METADATA

The general SIP metadata is based on the METS standard and presented as a profile. METS profiles are intended to describe a class of METS documents in sufficient detail to provide both document authors and programmers with the guidance they need to create and process METS documents conforming to a particular profile.[12]

Creating a METS profile requires a good understanding of the METS Profile components. An overview of these components can be found in the METS online documentation[13] and in Appendix D on page 44 in the D3.2 specification.[14]

There are 5 main sections in this METS profile:

- <metsHdr> - METS header (metadata about the creator, contact persons, etc. of the IP).
- <dmdSec> - descriptive metadata (references to EAD, EAC-CPF, etc.).
- <amdSec> - administrative metadata (information about how files were created and stored, intellectual property rights, etc.).
- <fileSec> - file section, lists all files containing content (may also contain metadata about files).
- <structMap> - structural map, describes the hierarchical structure of the digital object and the whole IP (i.e. object + metadata).

These sections will be described in more detail in sections 3.1 to 3.6. All these sections present the SIP requirements for METS elements in table form according to the following structure:

- Element - The name of the element in plain text used in the accompanying schema for elements or attributes. For more information regarding elements and attributes in XML see WWW Consortium (http://www.w3.org/).
- Definition - Defines the functions of the element. Contains an explanation of the element and some example values.
- Cardinality – Represents the number of occurrences of an element (see below).
    - 0..1 – The element is optional and cannot be repeated.
    - 0..* – The element is optional and can be repeated.
    - 1 – The element is mandatory and can only be stated once.
    - 1..* – The element is mandatory and has one or more occurrences.
- METS - Defines the element in the METS standard used for designing the SIP element. The column uses XML-syntax. [ ] defines where the value is placed.

[12] METS Profiles, http://www.loc.gov/standards/mets/mets-profiles.html
[13] METS Profile Components, http://www.loc.gov/standards/mets/profile_docs/components.html
[14] D3.2 E-ARK SIP Draft Specification, 2015, http://eark-project.com/resources/project-deliverables/17-d32-e-ark-sip-draft-specification

3.1. Root

The root of a METS document can contain a number of optional attributes, namespaces (xmlns:) and schema instance locations (xsi:) of the external standards referenced in the METS record and a number of other elements as seen in Table 1.

Table 1: METS root

Element	Definition	Explanation	Card.	METS
Identity / Content ID	Identification of the package	A code that uniquely identifies the whole SIP and the digital object/objects being submitted. A UUID or GUID should be used to create globally unique identifiers. Example: *"UUID:550e8400-e29b-41d4-a716-446655440004"*	1	<mets: OBJID="[Identity]" >
Description / Package name	Description of the package	Short text describing the package. Example: *"SIP for delivery of personnel information"*	0..1	<mets: LABEL="[Description]">
Content type / General content type	The content type being submitted with this package	The TYPE attribute must be used for identifying the type of the package (genre), for example ERMS, RDBMS, digitised construction plans. However, there is no fixed vocabulary and as such implementers are welcome to use values most suitable for their needs. Example: "*ERMS*"	1	<mets: TYPE="[Content type]"

Element	Definition	Explanation	Card.	METS
Profile[15]	Profile name	Describes the METS-profile being used. The name should contain the version number and the version may be captured in the profile file path. Example: *" http://eark-project.com/resources/METS/v02/METS.xml"*	1	<mets: PROFILE="[Profile]"
Content Information Type Specification	Content type specification used for the content type	An attribute added by this specification. It describes which content information type specification is used for the content. Values of the attribute are fixed in the following vocabulary: 1. SMURFERMS 2. SMURFSFSB 3. SIARD1 4. SIARD2 5. SIARDDK 6. GeoVectorGML 7. GeoRasterGeotiff NB The vocabulary is extensible as additional content information type specifications are developed. Example: "SMURFERMS"	0..1	<mets: CONTENTTYPESPECIFICATION="[Content type specification]"

Example:

```
<mets xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xlink="http://www.w3.org/1999/xlink" xmlns="http://www.loc.gov/METS/"
PROFILE="http://www.dasboard.eu/specifications/sip/v03/METS.xml" TYPE="ERMS" OBJID="5d378f86-
28a1-41d8-a2b9-264b10fbd511" CONTENTTYPESPECIFICATION="SMURFERMS" LABEL="METS file describing the
SIP." xsi:schemaLocation="http://www.loc.gov/METS/ schemas/METS.xsd http://www.w3.org/1999/xlink
schemas/xlink.xsd">
```

[15] This element is not used for representation METS XML files.

3.2. Header

The METS header element <metsHdr> describes metadata about the creator, contact persons, etc. of the submission information package as seen in Figure 4.

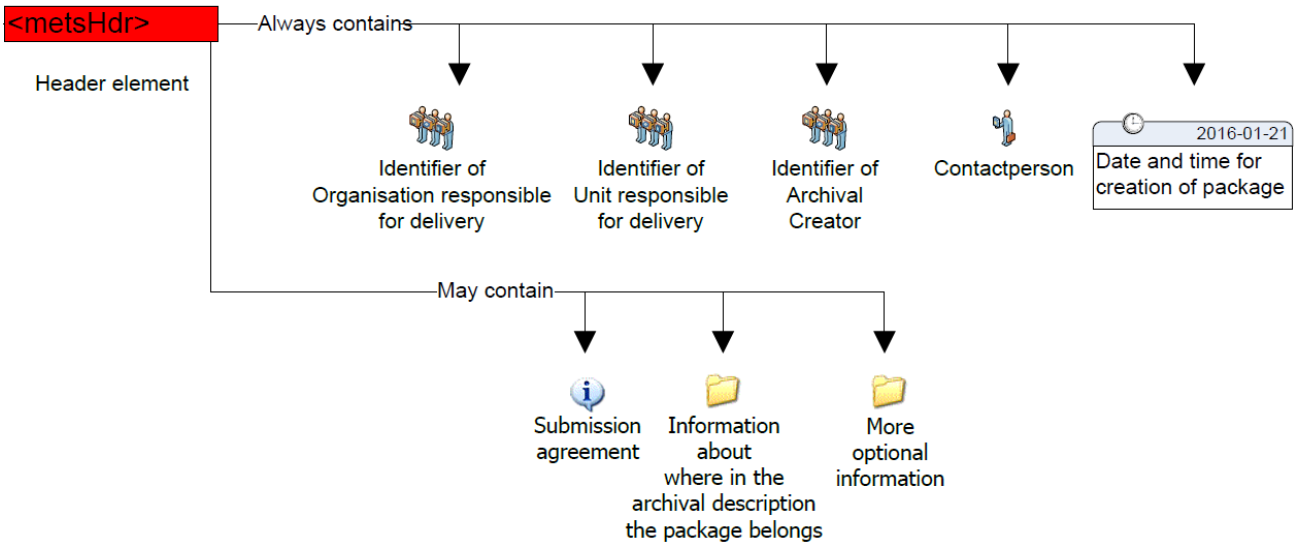

Figure 4: METS header

These are the elements that give information about the submission of the SIP in the METS header element.

Table 2: Metadata about the information package

Element	Definition	Explanation	Card.	METS
Id	Unique ID for the METS header section	Unique ID for the METS header section. Example: *"ac0e8400-e29b-41d4-a716-446655440004"*	0..1	<metsHdr: ID=”[ID value]”>

Element	Definition	Explanation	Card.	METS
Date and time*[16]	Time of creation of package[17]	Date and time for creation of the package must be described according to the XML-standard ("YYYY-MM-DDThh:mm:ssZ").[18] This timestamp states when the whole package and the package file was created.[19] Example: *"2012-04-26T12:45:00+01:00"*	1	<metsHdr: CREATEDATE="[Date and time]">
Status*	Package status	A way of indicating the status of the package for making it easier to know how to handle the package (allows package status specific processing or rendering) if for example an identical package is being submitted replacing the previous one sent. Example: "*NEW*" Example: "*TEST*" Example: "*REPLACEMENT*"	0..1	<metsHdr: RECORDSTATUS="[Status]">
OAIS type of package	SIP, AIP, DIP, AIU, AIC	The type of the IP. Possible values are SIP, AIP, DIP, AIU, AIC. This list of values is managed by the DAS Board and will be updated when required.	1..1	<metsHdr: PACKAGETYPE="[OAIS type of package]">

[16] Header elements marked with an asterisk * represent elements which are required only for the root METS.xml. All other elements can be recorded in the representation METS.xml files.
[17] The Commona Specification for Information Packages also defines the last modification date, but this is not relevant for SIPs.
[18] XML Schema Part 2: Datatypes Second Edition, https://www.w3.org/TR/2004/REC-xmlschema-2-20041028/#isoformats
[19] The package file is the METS-document which describes the content of the whole package.

Element	Definition	Explanation	Card.	METS
DocID*	METS document ID	A unique identifier for the METS document itself. This identifier may be different from the Identity given in the mets-element. The recommendation is to use the file name given to the METS-document. Example: "*SIP20150127.xml*"	0..1	<metsDocumentID>[DocID]

Element	Definition	Explanation	Card.	METS
Submission agreement[20]	Reference to the used submission agreement	Complete reference to the submission agreement for the submission of the package. In this way it provides an alternative identifier for the package due to the requirement that a package can only belong to one submission agreement. Example: *RA 13-2011/5329; 2012-04-12* Example: *http://submissionagreement.kb.se/dnr331-1144-2011/20120711/* **Note:** *It is recommended to use an external schema for better description of a submission agreement.*[21] Example: <external_schema> <name>Submission Agreement (SA)</name> <URL>http://www.dasboard/xmlns/SubmissionAgreement.xsd</URL> <context> The submision agreement specifies the relations between the producer and the archive as it is described in ISO 20652:2006 (Producer-archive interface -- Methodology abstract standard). </context> </external_schema>	0..1	<altrecordID: TYPE=”SUBMISSIONAGREEMENT”>[Submission agreement] **OR PREFERABLY** <mets: <external_schema> <name>Submission Agreement (SA)</name> <URL>http://www.dasboard/xmlns/SubmissionAgreement.xsd</URL> <context> The submision agreement specifies the relations between the producer and the archive as it is described in ISO 20652:2006 (Producer-archive interface -- Methodology abstract standard). </context> </external_schema>

[20] Reference Model for an Open Archival Information System (OAIS), 2012, public.ccsds.org/publications/archive/650x0m2.pdf
[21] For example, the submission agreement developed by Docuteam GmbH http://www.loc.gov/standards/mets/profiles/00000041.xml

Element	Definition	Explanation	Card.	METS
Previous submission agreement	The previous submission agreement(s) the information belongs to in the case the information is recorded.	Reference to the previous submission agreement(s) which the information may have belonged to is recorded if the information is available. Example: *”FM 12-2387/12726, 2007-09-19”* **Note:** *It is recommended to use an external schema for better description of a submission agreement like explained previously.*	0..*	<altrecordID: TYPE=”PREVIOUSSUBMISSIONAGREEMENT”>[Previous submission agreement] **OR PREFERABLY** <mets: <external_schema> <name>Previous Submission Agreement (PSA)</name> <URL>http://www.dasboard/xmlns/SubmissionAgreement.xsd</URL> <context> The submision agreement specifies the relations between the producer and the archive as it is described in ISO 20652:2006 (Producer-archive interface -- Methodology abstract standard).</context> </external_schema>

Element	Definition	Explanation	Card.	METS
Archival reference code	Reference code in the archival description	It is possible to give a reference code indicating where in the archival hierarchy the package shall be placed. Example: *"SE/RA/123456/24/P"*	0..1	<altrecordID: TYPE="REFERENCE CODE">[Archival reference code]
Previous reference code	An earlier used reference code in the archival description	In case where the SIP originates from other institutions maintaining a reference code structure, this element can be used to record these reference codes and therefore support the provenance of the package when a whole archival description is not submitted with the submission. Example: *"SE/FM/123/123.1/123.1.3"*	0..*	<altrecordID: TYPE="PREVIOUSR EFERENCECODE">[Previous reference code]
Archival creator	Name of archival creator	Name of the original creator (organisation) of the data being transferred. Please note that this might be different from the organisation which has been charged with preparing and sending the SIP to the archives. Example: *"The Swedish health agency"*	1[22]	<agent: ROLE="ARCHIVIST " TYPE= "ORGANIZATION"> <name>[Archival creator]

[22] One SIP can contain information form one and only creator.

Element	Definition	Explanation	Card.	METS
Archival creator identification code	A unique identification code for the archival creator	A unique identification code for the archival creator. The code uses a prefix followed by a ”:” Prefix[23] according to vcTypeOfIdentificationCode. Example: *”VAT:SE201345098701”*	0..1	<agent: ROLE=”ARCHIVIST ” TYPE= ”ORGANIZATION”> <note>[Archival creator identification code]
Submitting organisation name[24]	Name of the organisation submitting the package to the archive	Name of the organisation submitting the package to the archive. For stating and extending the information, use of additional agents may be necessary. Example: *”The agency, Personnel*	1	<agent: ROLE=”CREATOR” TYPE= ”ORGANIZATION”> <name>[Submitting organisation name]
Delivering organisation identification code	A unique identification code for the delivering organisation	A unique identification code for the delivering organisation. The code uses a prefix followed by a ”:” Prefix according to vcTypeOfIdentificationCode. Example: *” VAT:SE2098109810-AF87”*	0..1	<agent: ROLE=”CREATOR” TYPE= ”ORGANIZATION”> <note>[Delivering organisation identification code]

[23] All prefixes referred in this table are described in the SIP METS Profile
[24] All similar <agent> elements (the archival creator, delivering organisation, submitting organisation and producing organisation) may not be needed in some implementations of the SIP METS Profile.

Element	Definition	Explanation	Card.	METS
Contact person name	Contact person for the submission	It may be useful to record a contact person for the submission at the time of the submission. This information is only valid during a short time period. Example: *”Sven Svensson”*	0..*	<agent ROLE=”CREATOR” TYPE= ”INDIVIDUAL”> <name>[Contact person name]
Contact person contact information	Contact information for the contact person	Phone number and e-mail for the contact person. This information is only valid during a short time period. Example:[25] *”08-12 34 56”* *“sven.svensson@fm.se”*	0..*	<agent ROLE=”CREATOR” TYPE= ”INDIVIDUAL”> <note>[Contact person contact information]
Software*	The software which has been used to create the package	The metsHdr must include at least one agent describing the software which has been used to create the package.	1	<agent ROLE=”CREATER” TYPE=”OTHER” OTHERTYPE=”SOFTWARE”>
Preservation organisation name	Name of organisation preserving the package	Name of the organisation responsible for preservation of a submitted package. Example: *”National Archives of Hungary”*	1	<agent ROLE= ”PRESERVATION” TYPE= ”ORGANIZATION”> <name>[Preservation organisation name]

[25] As the SIP profile allows for 0..* fields then it is possible to have e-mail, phone, physical address details etc in separate fields. The choice is up to specific implementations.

Element	Definition	Explanation	Card.	METS
Preservation organisation identification code	Identification code of organisation preserving the package	A unique identification code for the organisation responsible for preservation. The code uses a prefix followed by ":" Prefix according to vcTypeOfIdentificationCode. Example: *"ORG:2010340987"*	0..1	<agent ROLE= "PRESERVATION" TYPE= "ORGANIZATION"> <note>[Preservation organisation identification code]

Example:

```
<metsHdr CREATEDATE="2016-01-15T10:40:09" RECORDSTATUS="NEW" PACKAGETYPE="SIP">
  <agent TYPE="ARCHIVIST" ROLE="ORGANIZATION">
    <name>The Hungarian Ministry of Healthcare</name>
    <note>ORG:HU121345098701</note>
  </agent>
  <agent TYPE="CREATOR" ROLE="ORGANIZATION">
    <name> The Hungarian Health Agency</name>
    <note>ORG:HU891345098701</note>
  </agent>
  <agent TYPE="PRESERVATION" ROLE="ORGANIZATION">
    <name> National Archives of Hungary</name>
    <note>ORG:HU2010340987</note>
  </agent>
  <agent TYPE="OTHER" ROLE="CREATOR" OTHERTYPE="SOFTWARE">
    <name>SIP Creator</name>
    <note>VERSION=0.0.2</note>
  </agent>
</metsHdr>
```

3.3. Descriptive metadata

The METS descriptive metadata element <dmdSec> references to archival description metadata (EAD, EAC-CPF, etc.) as seen in Figure 5.

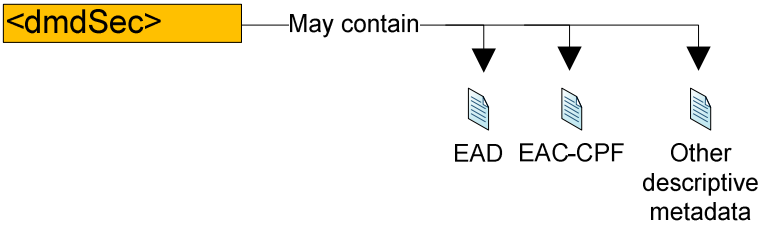

Figure 5: METS descriptive metadata

Archival information can be included in the METS package. Usually, for the archival institutions this information is delivered in EAD and EAC-CPF formats.
To include EAD and EAC-CPF in a METS profile the use of <dmdSec> is to be preferred according to the METS implementation guide referenced above. The complete rules for all elements and attributes in the <dmdSec> are stated in the profile, the specific elements used when referencing and embedding is shown below.
Other metadata standards for description and administrative purposes can be used and referencing them must adhere to the <amdSec> and <dmdSec> rules stated in the profile.

Table 3: EAD metadata

Element	Definition	Explanation	Card.	METS
EAD metadata file	Metadata file in EAD format referenced in the METS document	Metadata file in EAD format when it is referenced in the METS document. Also an unique identifier (@ID), creation date (@CREATED) of the metadata in this section and the locator type (@LOCTYPE) used in the xlink:href has to be recorded.	0..*	<dmdSec> <mdRef: ID="IDc9abe6db-84eb-4af3-9d45-ca235a959313" MDTYPE=”EAD” xlink:href=”file://[26] [EAD-metadata file]” xlink:type=”simple”>

[26] The coice of using ”///” or ”//” depends on implementations. For example, a Windows file path could be file://C:/somefile.txt and file:///somefile.txt would be the corresponding Unix path in case of absolute paths.

Element	Definition	Explanation	Card.	METS
EAC-CPF metadata file	Metadata file in EAC-CPF format referenced in the METS document	Metadata file in EAC-CPF format when it is referenced in the METS document. Also an unique identifier (@ID), creation date (@CREATED) of the metadata in this section and the locator type (@LOCTYPE) used in the xlink:href has to be recorded.	0..*	<dmdSec> <mdRef: ID="IDa2abe6db-84eb-2af3-9d45-ca235a959312" MDTYPE= ”EAC-CPF” xlink:href=”file://[EAC-CPF metadata file]” xlink:type=”simple”>

Example:

```
<dmdSec ID="ca14390a-6732-45ba-a466-3c32b7bbacd3">
     <mdRef ID="a9769e81-900a-466c-aee2-86249642532b" LOCTYPE="URL" MDTYPE="EAD"
MDTYPEVERSION="3" xlink:type="simple" xlink:href="file://metadata/descriptive/ead.xml"
MIMETYPE="text/xml" SIZE="4700" CREATED="2017-01-31T10:22:02.1202592+02:00"
CHECKSUM="ee673e7c9f787a02f4ce30c553e781ac065487c3c9c16ac6381b51e2059f8cf2" CHECKSUMTYPE="SHA-
256" />
</dmdSec>
<dmdSec ID="c9adf006-2383-4e8a-830c-506e47db2188">
     <mdRef ID="b9520eb0-4b1b-4c4f-8c9d-d0677eb2618a" LOCTYPE="URL" MDTYPE="EAC-CPF"
MDTYPEVERSION="2010" xlink:type="simple" xlink:href="file://metadata/descriptive/eaccpf.xml"
MIMETYPE="text/xml" SIZE="1787" CREATED="2017-01-31T10:22:02.1202592+02:00"
CHECKSUM="1eefac646c8e42d57d7fa9ede829e7b780e89bda7bf6ab5374535664118f3fcf" CHECKSUMTYPE="SHA-
256" />
</dmdSec>
```

3.4. Administrative metadata

The METS administrative metadata element <amdSec> references to technical and preservation metadata as seen in Figure 6.

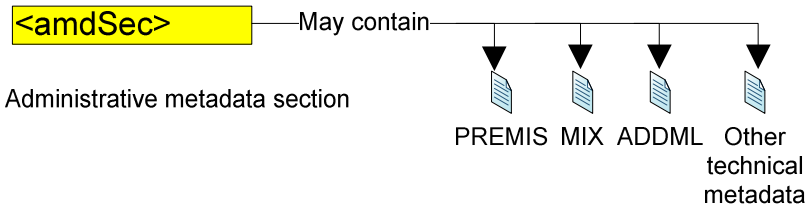

Figure 6: METS administrative metadata

Preservation metadata can be included[27] in the METS package. It is recommended that PREMIS[28] is used for preservation metadata. For further reading:

- More information about PREMIS can be found at: http://www.loc.gov/standards/premis/ .
- A guide on using PREMIS with METS may be found at: http://www.loc.gov/standards/premis/guidelines-premismets.pdf.
- Decisions made during the use of PREMIS can be recorded using this document: http://www.loc.gov/standards/premis/premis_mets_checklist.pdf

The guide on using PREMIS with METS (referenced above) recommends using the <amdSec> in order to reference PREMIS metadata. The complete rules for all elements and attributes in the <amdSec> are stated in the profile, the specific elements used when referencing are shown below. However, please note that preservation metadata varies for different content types and therefore best practice guidelines should be applied as required.

Table 4: PREMIS metadata

Element	Definition	Explanation	Card.	METS
PREMIS metadata file	Metadata file in PREMIS format referenced in the METS document	Metadata file in PREMIS format when it is referenced in the METS document. Its needs to be stated in the submission agreement if referencing or embedding are used.	0..*	<amdSec> <digiProvMD> <mdRef: MDTYPE= ”PREMIS” xlink:href=”file://[

[27] The Common Specification recommends avoiding embedding preservation or any other type of metadata directly in to the METS file.
[28] We expect that PREMIS will be created automatically by the SIP or AIP Creation Tools.

Element	Definition	Explanation	Card.	METS
				PREMIS metadata file]” xlink:type=”simple ”>

Example:

```
<amdSec ID="ID1a57e479-20e2-4e99-868b-88d0f816d109">
    <digiprovMD ID="ID41d8bb3c-f7c1-4254-aa9f-825009314fb0">
      <mdRef MIMETYPE="text/xml" xlink:href="file://metadata/preservation/premis1.xml"
LOCTYPE="URL" CREATED="2015-11-18T15:50:14"
CHECKSUM="8aa278038dbad54bbf142e7d72b493e2598a94946ea1304dc82a79c6b4bac3d5" xlink:type="simple"
ID="ID58ecdae0-b6af-4ad9-abf1-f6c2971f253a" MDTYPE="OTHER" CHECKSUMTYPE="SHA-256"/>
    </digiprovMD>
</amdSec>
```

3.5. Files

The METS file section element <fileSec> lists all files containing content (may also contain metadata about files) as seen in Figure 7.

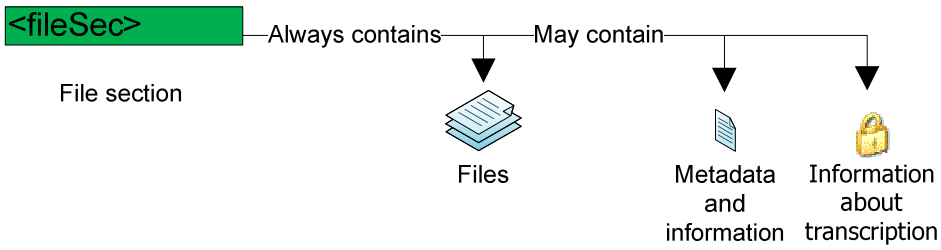

Figure 7: METS files

All files found in the submission package should be referenced once and only once in the METS-document describing the submission. The elements and attributes are the same regardless of the content type submitted.
When describing the content and documentation files in METS they are placed in the fileSec element in one or more fileGrp elements. The fileGrp element can be used for grouping files together in different ways. In this profile we do not group files in different groups, we only use one mandatory fileGrp. Use of more fileGrp’s must be decided in every implementation and described in a METS profile.

Table 5: Files metadata

Element	Definition	Explanation	Card.	METS
Identification of the file	Identification of the file object	A code that uniquely identifies the file inside the METS-document for referencing in the structMap element. Suggested use is a prefix "ID" directly followed by an UUID or GUID or own local identification code. ID follows the rules of the xml attribute XML:ID. Example: *"ID550e8400-e29b-41d4-a716-4466554400bg"*	1	<file ID="[Identification of the file]">
File location	Name of the file	Name of the file and the path to locate it in the package. The file name must use the prefix file://. The attribute LOCTYPE is mandatory to describe how to find the file and uses a value list present in METS. Example: *"file://personnelexport.xml"*	1	<file <flocat: LOCTYPE="URL" xlink:href="file://[File name]" xlink:type="simple">
Date and time	Timestamp for the file[29]	The timestamp seen on the file and used for validating the file. In most cases this is the last modification date. Described using xml type datetime rules. Example: *"2012-04-20T13:30:00+01:00"*	1	<file CREATED="[Date and time]">

[29] It is the timestamp recorded inside the file (i.e. information we can read with JHove and similar tools).

Element	Definition	Explanation	Card.	METS
MIME type[30]	Simplest way of describing a file type	The simplest way of describing a file type. Example: *"text/xml"*	1	<file MIMETYPE="[MIME type]">
File format name[31]	Name of file format	Sometimes a more detailed name needs to be given to the file format when use of PREMIS has not been agreed upon in the submission agreement. Example: *"Extensible Markup Language"* Example: *"PDF/A"* Example: *"ISO/IEC 26300:2006"*	0..1	<file ext:FILEFORMATNAME="[File format name]"
File format version	Version of file format	The version of the file format and use of PREMIS has not been agreed upon in the submission agreement. Example: *"1.0"*	0..1	<file ext:FILEFORMATVERSION="[File format version]"

[30] Media Types, 2015, https://www.iana.org/assignments/media-types/media-types.xhtml
[31] This and all following file elements can be recorded in the SIP by using extension schemas (as shown here), but also by using PREMIS or some other metadata standard.

Element	Definition	Explanation	Card.	METS
Format registry name	Name of the format registry identifying the file format	Name of the format registry identifying the file format when use of PREMIS has not been agreed upon in the submission agreement. According to vocabulary[32] vcTypeOfFormatregistry Example: *" PRONOM""*	0..1	<file ext:FORMATREGISTRY=”[Format registry name]”
Format key	Key of the file format in the registry	Key of the file format in the registry when use of PREMIS has not been agreed upon in the submission agreement. Example: *“fmt/101"*	0..1 1 If Format registry name is used	<file ext:FORMATREGISTRYKEY=”[Format key]”
File size	Size of the file in bytes	Size of the file in bytes. Example: *”8765324”*	1	<file SIZE=”[File size]”>
Function	Identifies the function of the file	Sometimes it is useful to record information on the function of the file. A vocabulary for stating the “Function” is recommended in every implementation and the vocabulary should be stated in the METS profile. Example: *”Submission file”*	0..1	<file: USE=”[Function]”>

[32] All vocabularies referred in this table are described in the SIP METS Profile

Element	Definition	Explanation	Card.	METS
Checksum type	Used algorithm for creating the checksum	Algorithm used for creating the checksum. Values are predefined in METS. The algorithm to use is to be stated in the submission agreement. Example: *”SHA-256”*	1	<file: CHECKSUMTYPE=”[Checksum type]”>
Checksum value	Calculated checksum for the file	Check sum for the file. Example: *”574b69cf71ceb5534c8a2547f5547d”*	1	<file: CHECKSUM=”[Checksum]”>
Transformation algorithm	Transformation algorithm used for transformation	Transformation algorithm used for any file transformation (decryption/encryption). The algorithm to use is to be stated in the submission agreement. The attribute “TRANSFORMATIONTYPE” is to be used to state the transformation type according to predefined set of values. Example: ”DES”	0..1	<file <transformFile TRANSFORM-TYPE=”decryption” TRANSFORM-ALGORITHM= ”[Transformation algorithm]”>
Transformation key	Transformation key for a transformed file	Transformation key for a transformed file. The attribute “TRANSFORMATIONTYPE” states the transformation type according to a predefined set of values. Example: *”574b69cf71ceb5534c8a2547f5547d”*	0..1 1 If Transformation algorithm is used	<file <transformFile TRANSFORMTYPE=”decryption” TRANSFORMKEY= ”[Transformation key]”>

Example of the <fileSec> element (root METS file):

```
<fileSec>
   <fileGrp USE="files root" ID="IDae911aa8-24f0-4bd8-a684-32044b89d687">
      <fileGrp USE="schemas" ID="IDae911aa8-24f0-4bd8-a684-32056b89d789">
         <file MIMETYPE="application/xsd" USE="Schema" CHECKSUMTYPE="SHA-256" CREATED="2015-12-
04T09:59:45" CHECKSUM="41d38f0a204e7dbda2838d93ad8eb5cf6bed92acd9c2f06f497faf47722e990d"
ID="ID04918b96-cf9f-41fa-ab13-3d550aaf94f5" SIZE="6814">
            <FLocat xlink:href="file://schemas/METS.xsd" xlink:type="simple" LOCTYPE="URL"/>
         </file>
      </fileGrp>
      <fileGrp USE="representations" ID="IDae055ba8-24f0-4bd8-a684-32056b89d882">
         <fileGrp USE="representation123" ID="IDbc911aa8-24f0-4bd8-a684-32056b89d789">
            <file MIMETYPE="application/xml" USE="Representation METS" CHECKSUMTYPE="SHA-256"
CREATED="2015-12-04T09:59:45"
CHECKSUM="41d38f0a204e7dbda2838d93ad8eb5cf6bed92acd9c2f06f497faf47722e990d" ID="ID04918b96-cf9f-
41fa-ab13-3d550aaf94f5" SIZE="6814">
               <FLocat xlink:href="file://representations/representation123/METS.xsd"
xlink:type="simple" LOCTYPE="URL"/>
            </file>
         </fileGrp>
      </fileGrp>
      <fileGrp USE="documentation" ID="ID7d136e4c-26fe-40da-85a2-67a42efd6b27">
      ...
      </fileGrp>
   </fileGrp>
</fileSec>
```

Example of the <fileSec> element (representation METS file):

```
<fileSec>
   <fileGrp USE="Files representation representation123" ID="IDae911aa8-24f0-4bd8-a684-
32044b89d687">
      <fileGrp USE="data" ID="IDae911aa8-24f0-4bd8-a684-321556389d687">
         <fileGrp USE="user-defined-data-subfolder" ID="IDae911aa8-24f0-4bd8-a684-32044b89d789">
            <file MIMETYPE="application/pdf" USE="data" CHECKSUMTYPE="SHA-256" CREATED="2015-12-
04T09:59:45" CHECKSUM="41d38f0a204e7dbda2838d93ad8eb5cf6bed92acd9c2f06f497faf47722e990d"
ID="ID04918b96-cf9f-41fa-ab13-3d550aaf94f5" SIZE="6814">
               <FLocat xlink:href="file://data/contentfile.pdf" xlink:type="simple"
LOCTYPE="URL"/>
            </file>
         </fileGrp>
         ...
      </fileGrp>
      <fileGrp USE="documentation" ID="ID7d136e4c-26fe-40da-85a2-67a42efd6b27">
      ...
      </fileGrp>
   </fileGrp>
</fileSec>
```

3.6. Structure

The mandatory METS structural map element <structMap> describes the hierarchical structure for the digital object as seen in Figure 8 and follows completely the requirements set in the Common Specification for Information Packages.

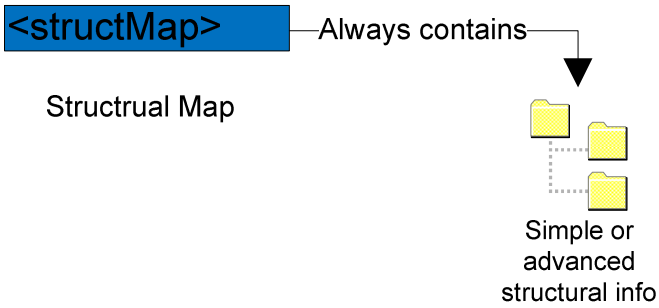

Figure 8: METS structural section

Example:

```
<structMap ID="cb595ef2-46bf-4c0f-b777-ed21d63c5ac1" TYPE="physical" LABEL=" Common Specification
structural map">
   <div ID="ae592e3b-b647-4585-945f-619c5a8b980b" LABEL="9da99df7-2237-48d6-90ef-01d99447c16f">
      <div ID="cbbbb887-e73e-4c35-a3fe-13c85ff47d6f" LABEL="metadata">
         <div ID="fa312f6f-c2b0-4124-89b5-9eec9b8267f9" LABEL="descriptive">
            <fptr FILEID="IDc04f8f55-802e-4646-b5f9-78b8e864e530"/>
            <fptr FILEID="IDa2da0aa8-bf9c-4a79-a83d-2944cb2031ab"/>
         </div>
         <div ID="d90b45ac-f587-42ff-bf20-b69eec1ca2fe" LABEL="preservation">
            <fptr FILEID="IDc2ccef19-802e-4646-b5f9-78b8e864e532"/>
            <fptr FILEID="IDa2da11a8-bf9c-4a79-a83d-2944cbfee654"/>
         </div>
      </div>
      <div ID="ebcb69c0-a719-483a-8991-a7af65fa290c" LABEL="schemas">
         <fptr FILEID="ID845a7a5b-0cfe-43ff-acd9-14f5f0463e28"/>
      </div>
      <div ID="d958361d-a147-41dc-b657-b11a20205cab" LABEL="representations"/>
         <div ID="be0ac8a7-0be9-4222-b307-5306d75b88c1" LABEL="representations/rep1">
            <mptr xlink:href="file://representations/rep1/METS.xml" xlink:type="simple"
LOCTYPE="URL"/>
         </div>
         <div ID="d7fc076e-0287-4105-acf4-1a44d0172c7a" LABEL="representations/rep2">
            <mptr xlink:href="file://representations/rep2/METS.xml" xlink:type="simple"
LOCTYPE="URL"/>
         </div>
   </div>
</structMap>
```

4. CONTENT INFORMATION TYPE SPECIFICATIONS

As discussed above (Chapter 2), an SIP can include content-type specific data and metadata. Types of data files and their structural relationships, and metadata elements vary for different content-types. Metadata is submitted to an archive so that it can support functions in the archive. The metadata created by business systems can be in different structures / formats. The amount and type of available metadata depends very much on the type and owner/developer of the system. As such there are also differences in how much metadata can a specific system or type of system export and in which formats. To deal with these differences there's the possibility of content type profiles which define detailed metadata requirements beyond the Common Specification for Information Packages and general SIP.

This specification does not offer one single structure in which the content-type specific metadata could be stored as a whole. In order to efficiently use the metadata to support archival functions the SIP defines separate SIP METS sections as containers for the various metadata functions, such as the METS header for package management function, the <dmdSec> for EAD metadata standard (i.e. using "<dmdSec> for package discovery) and other descriptive metadata standards, the <amdSec> for preservation (PREMIS), technical and other functions and standards. In order to use the submitted metadata it has to be mapped to and referenced from the SIP METS sections.[33] To do this the content-type specific metadata elements need to be mapped to those containers and implemented in the agreed standards. Therefore, complementary metadata profiles are needed for content types. This document refers to 3 profiles which define how the submitted content-specific metadata should be mapped to the SIP structure:

- The SMURF (semantically marked up record format) for ERMS will contain mappings for ERMS (electronic records management systems) based on MoReq2010 as described in 4.1.
- The SMURF for SFSB (simple file-system based) records as described in 4.2.
- The SIARD 2.0 profile for relational databases as described in a section 4.3.

All SIPs will need to be transformed into AIPs in the archival ingest process. The SIP to AIP conversion is described in the AIP specification.[34]

[33] In the case of descriptive metadata it is even additionally integrated with external systems, such as the catalogue in order to support external access to the archive. One might also want to do that for other metadata (e.g. technical or preservation) in order to ease management of the archive.

[34] E-ARK AIP pilot specification, released January 2016, http://eark-project.com/resources/project-deliverables

4.1. Electronic records management systems (ERMS)

The first case represents ERMS records encapsulated in the SIP.[35] This profile aims to standardise the export of records management systems into a single easy to use model. The basic workflow is described on Figure 9.

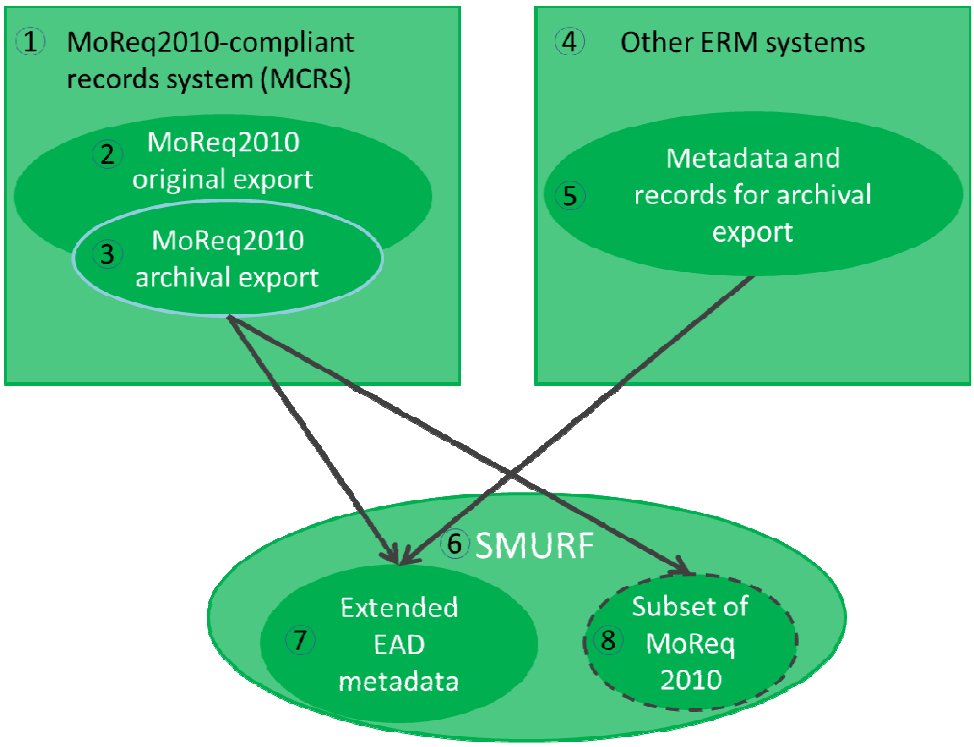

Figure 9: Extraction at pre-ingest

In case of ERMS we distinguish two scenarios – MCRS and non-MCRS (1, 4). The latter is assumed to be able to export metadata and records in a native export format (5),[36] the first supports in addition the specific MoReq2010 export format (2). Further, the export for archival purposes can differ from the original export (3).

The SMURF ERMS profile (6) defines a set of Extended EAD metadata (7)[37] which are created during the pre-ingest phase. In some cases it may be not possible to map all relevant original elements to a set of

[35] The scope of this chapter is to give short introductions; more details are available in a separate document SMURF (semantically marked up record format) for ERMS.
[36] The metadata extracted from a non-MCRS system should be mapped and transformed into the SMURF format by using external mechanisms (i.e. XSL transformation) or by updating the export format to support the SMURF profile.
[37] The EAD extraction will be created automatically by a MCRS.

Extended EAD metadata, therefore some MoReq 2010 elements (8) are allowed[38] for guaranteeing that all required elements are included in the SIP.

The SMURF extraction **should** be complemented with more general information about the information package and **could** be complemented with PREMIS, EAC-CPF metadata as well (Figure 10).

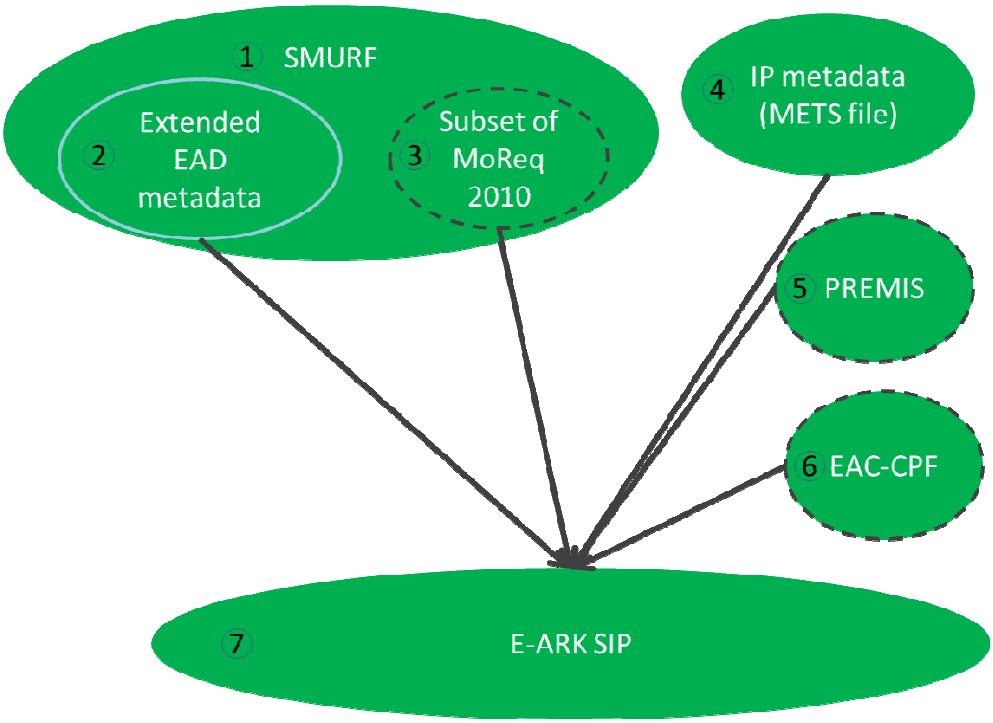

Figure 10: Creation at Pre-Ingest

The SMURF profile (1) includes MoReq2010 metadata that has been mapped to EAD (2) and some additional elements required by archives.[39] The structural metadata for the submission information package (represented as a METS file) will be added (4) during the SIP preparation process. If possible the EAC-CPF metadata (6) should be created and SIP creation events logged as PREMIS metadata (5). The full SIP will consist of items 1, 4 and optionally (5), (6).

[38] We do not recommend using MoReq2010 elements in the SMURF profile and therefore only the mapping from MoReq2010 elements to EAD will be provided.
[39] As referred earlier, we do not recommend using the original MoReq 2010 elements (3) in the SIP.

4.2. Simple file system based records (SFSB)

The second case represents an encapsulation of computer files into the SIP. It is based on an assumption that the files can be described in an extended EAD format[40] (Figure 11).

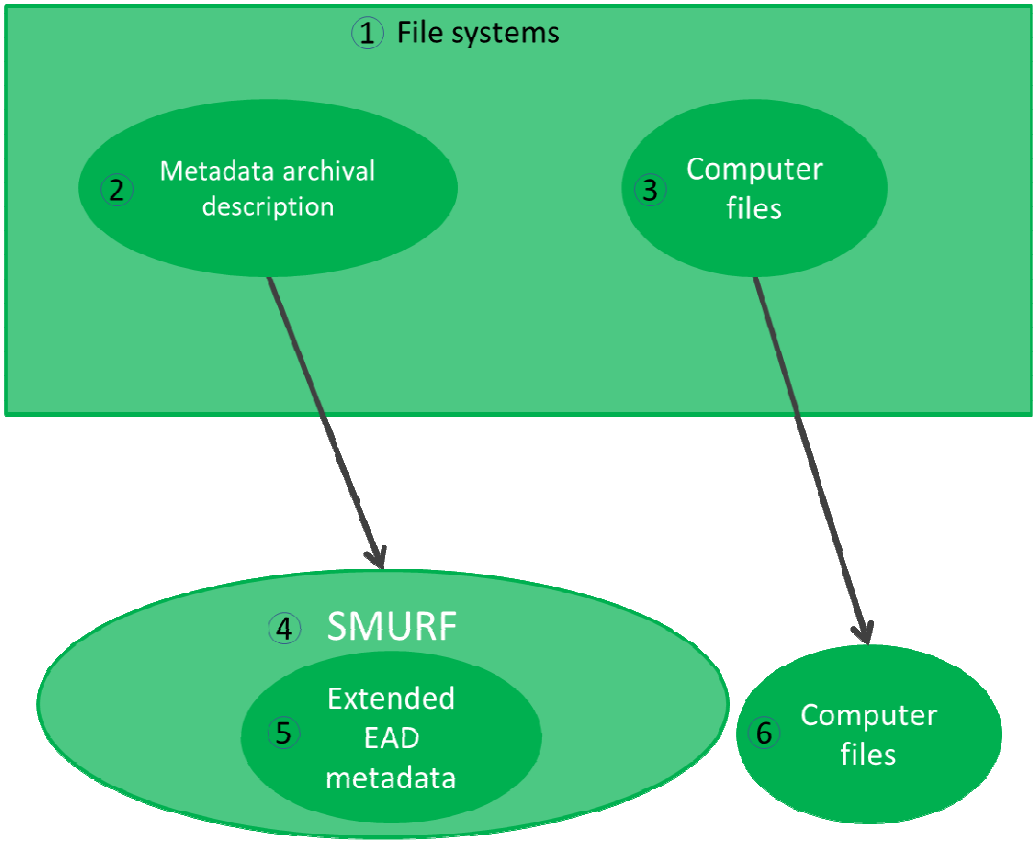

Figure 11: SFSB metadata and computer files

The blocks in the diagram refer to the following. Computer files reside in some file system (e.g. shared drives, 3). The metadata (2) about the files needed for the long time preservation may or may not exist. If the metadata exists then it has to be transformed into the EAD metadata (5). If the metadata does not exist then it has to be created and included in the SIP.

The SMURF metadata **should** be complemented with more general information about the information package and **could** be complemented with PREMIS, EAC-CPF metadata as well to build a full SIP (Figure 12).

[40] It is assumed that already available EAD creator tools (e.g. EAD editor at https://github.com/ewg118/eaditor) can be used.

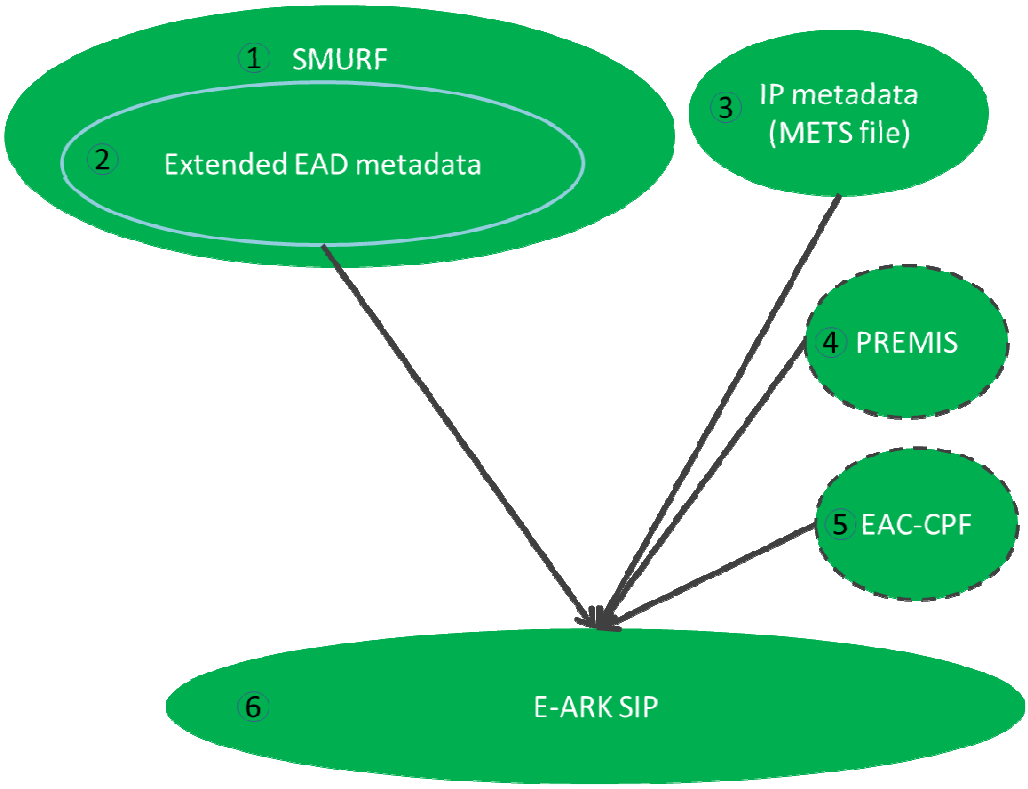

Figure 12: SFSB SIP

The blocks in the diagram refer to the following:

1. The SMURF profile for SFSB records.
2. Archival descriptions following the EAD extended schema for SFSB records.
3. Structural metadata for the submission information package (represented as METS file).
4. If possible then SIP creation events should be logged as PREMIS metadata.
5. If possible then EAC-CPF metadata should be created during the SIP creation process.
6. The SIP consists of items 1, 3 and optionally (4), (5).

4.3. Relational databases

The third case represents a relational database encapsulated in the SIP. This case structure presumes that the database is previously exported in the SIARD 2.0 format (a harmonised format for database archiving based on SIARD, Figure 13).[41]

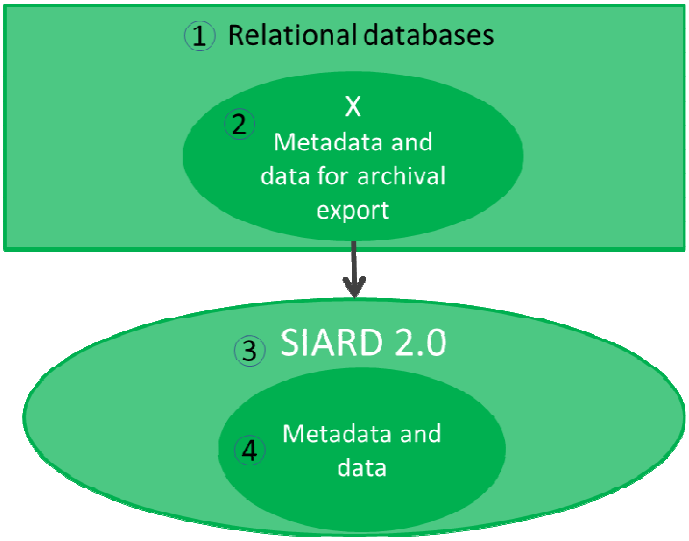

Figure 13: Export to SIARD 2.0

Various relational databases (e.g. Oracle, PostgreSQL, etc.) exist (1). These databases contain the metadata and records in its native format (2) which can be extracted into a standardised format (4) by following SIARD 2.0 (3). The SIARD extraction **should** be complemented with more general information about the information package and **could** be complemented with PREMIS, EAC-CPF, EAD metadata as well (Figure 14).

[41] The SIARD 2.0 specification (http://www.eark-project.com/resources/specificationdocs/32-specification-for-siard-format-v20/file) represents the SIP profile for the relational databases content type.

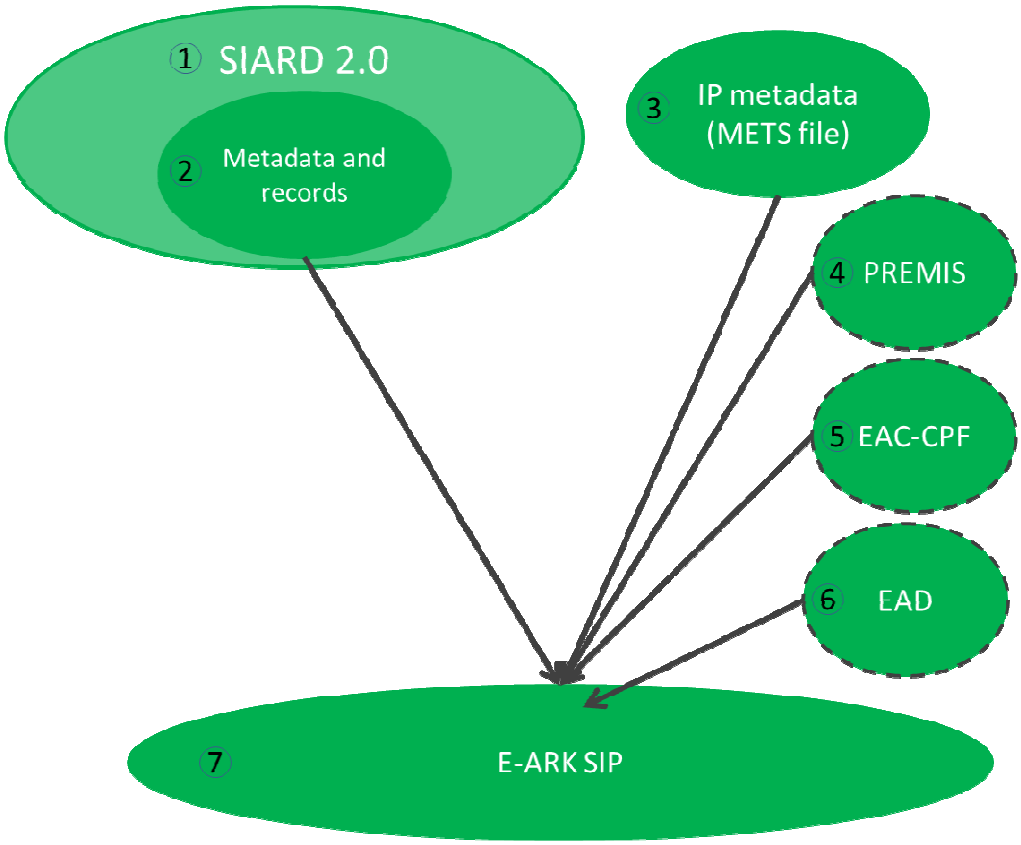

Figure 14: SIARD 2.0 to SIP

BLOBs and CLOBs in relational databases

The Figure 13 and Figure 14 show the most common profile for relational databases with metadata and records. However, in some cases there can be binary data in a relational database which will be exported as external files in SIP creation. This might cause a situation where it is necessary to consult with *"RECOMMENDATION for storing large objects outside the SIARD file",* which is a specific and technical recommendation that is not included in the SIARD 2.0 specification.

Binary data in regard to relational databases is defined as information which is stored in the database as a bit stream following a specific file format. The potentially huge size of binary data within a database can lead to problems in the handling and archival processing of the database. Binary data is mostly referred to as binary large object (BLOB). Similarly large amounts of character data are named CLOB. CLOBs pose a problem due to size more than lack of a proper data type. For the rest of this section CLOBs will be treated as BLOBs.

An example of a relational database with BLOBs could be a database where images are stored.

Databases and the handling of binary data has always been a challenge, regardless of whether the handling was based on:

1. Internal BLOBs - where data is contained in the records.

2. External direct references (path and filename) – where BLOBs are stored as files.
3. External indirect reference (file ID)- where BLOBs are stored as files.
4. Other methods which may exist.

The first method using internal BLOBs is supported in the SIARD 2.0 format, but if a table contains data with BLOBs that are more than 2000 bytes or 2000 characters in size, BLOBs will be produced as separate files and a reference to the location of the individual files stored in the cell content. The SIARD 2.0 format therefore also supports external reference to BLOBs stored as files inside the SIARD table folder structure (i.e. inside the SIARD ZIP package file).

The above scenario will therefore have no consequences regarding the Figure 13 and Figure 14 presented above.

The SIARD 2.0 format, however, also supports methods using external files outside the SIARD table folder structure (i.e. outside the SIARD ZIP package file) but it does not describe in detail how to handle BLOBs if this is the case. It is in this particular scenario that it is advisable to consult the detailed recommendations in *"RECOMMENDATION for storing large objects outside the SIARD file"* document.

When a SIP creation includes BLOBs stored as external files outside the table folder structure this will have influence on the SIP package since in this case there is not only one SIARD-file containing data from the database, but a SIARD-file and one or several other folders containing the external BLOB files.
A diagram for external files outside the SIARD table folder structure is presented in Figure 15:

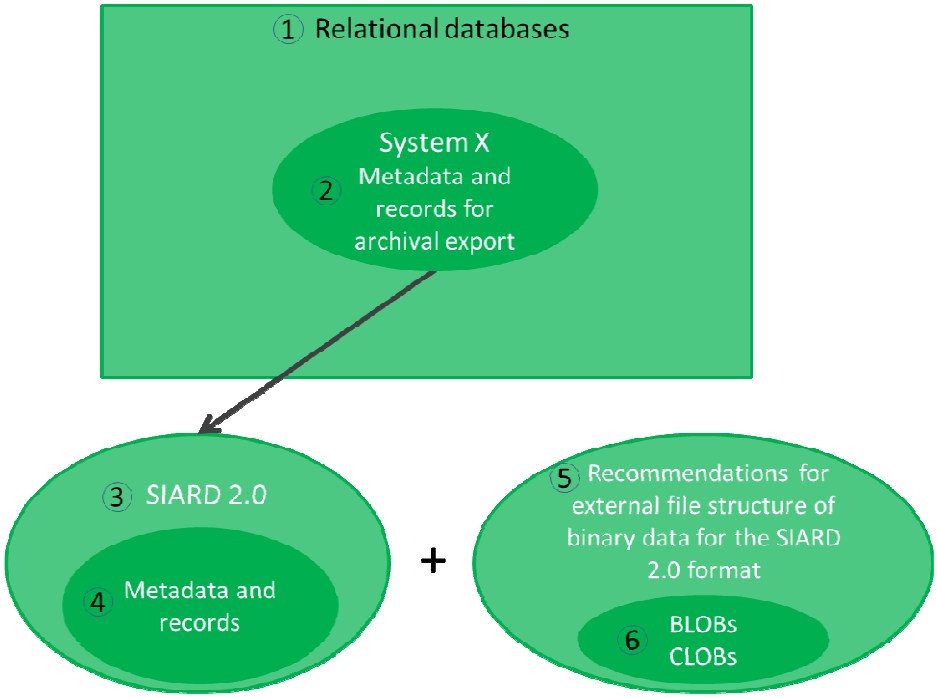

Figure 15: Relational databases with BLOBs/CLOBs stored as external files

1. Various relational databases (e.g. Oracle, PostgreSQL, etc.).
2. The metadata and records in a relational database.
3. The SIARD 2.0 specification.
4. The metadata and records in the SIARD 2.0 format.
5. Recommendations for external file structure of binary data for the SIARD 2.0 format.
6. BLOBs and/or CLOBs stored as external files outside the table structure.

External BLOBs influence on METS file

If there are several data files and folders in the SIP package, this consequently has influence on the IP metadata (METS file). Therefore, *"RECOMMENDATION for segmenting IP using METS"* describes how to represent the files in METS.

Further information can be found in the SIARD2.0 Profile document.[42]

[42] The SIARD 2.0 specification (http://www.eark-project.com/resources/specificationdocs/32-specification-for-siard-format-v20/file).

5. SUBMISSION AGREEMENT

Interaction between the Archive and Producers is often formalized and guided by a Submission Agreement, which establishes specific details of the interaction such as the type of information submitted, the metadata the Producer is expected to provide, the logistics of the actual transfer of custody from the Producer to the archive, and any access restrictions attached to the submitted material.[43] According to the OAIS model the submission agreement is an agreement reached between an Archive and the Producer that specifies a data model, and any other arrangements needed, for the Data Submission Session. This data model identifies format/contents and the logical constructs used by the Producer and how they are represented on each media delivery or in a telecommunication session.[44]

The E-ARK project acknowledged the importance of submission agreements and provided a way for referencing it in a METS.xml regardless of its form.[45] This document does describe a recommended format for a Submission Agreement (Appendix B: Submission Agreement), but of course does not forbid the use of any other Submission Agreement format.

According to the PAIMAS standard the submission agreement should include a complete and precise definition of:[46]

- information to be transferred (e.g., SIP contents, SIP packaging, data models, Designated Community, legal and contractual aspects);
- transfer definition (e.g. specification of the Data Submission Sessions);
- validation definition;
- change management (e.g. conditions for modification of the agreement, for breaking the agreement);
- schedule (submission timetable).

The submission agreement is inspired by the PAIMAS requirements and the submission agreement template provided by the National Oceanic and Atmospheric Administration (NOAA). This document will propose a list of elements which are recommended to be recorded in the submission agreement (8.2).

[43] Lavoie B, The Open Archival Information System (OAIS) Reference Model: Introductory Guide (2nd Edition), 2014, www.dpconline.org/component/docman/doc_download/1359-dpctw14-02
[44] Reference Model for an Open Archival Information, 2012, public.ccsds.org/publications/archive/650x0m2.pdf
[45] A submission agreement can be delivered in a digital (e.g. PDF or XML file) or an analogue way (e.g. paper document).
[46] PAIMAS, 2004, http://public.ccsds.org/publications/archive/651x0m1.pdf

6. SUMMARY

This document described the proposed general structure for submission information packages by explaining how the SIP can be generally constructed by following the common rules developed for all (including archival, dissemination) information packages.

The central section described the metadata sections and elements in these sections.

It also introduced the profiles for the SIP and shows how SIPs can manage various content and metadata. The profiles itself can be found in separate documents.

7. REFERENCES

1. A Checklist for Documenting PREMIS-METS Decisions in a METS Profile, 2010, URL: http://www.loc.gov/standards/premis/premis_mets_checklist.pdf
2. E-ARK Report on Available Best Practices, 2014, URL: http://eark-project.com/resources/project-deliverables/6-d31-e-ark-report-on-available-best-practices
3. e-SENS (Electronic Simple European Networked Services) project, 2015, URL: http://www.esens.eu/
4. Encoded Archival Context for Corporate Bodies, Persons, and Families, 2015, URL: http://eac.staatsbibliothek-berlin.de/
5. FGS packet structure, 2013, URL:https://riksarkivet.se/Media/pdf-filer/Projekt/FGS_Earkiv_Paket.pdf
6. Guidelines for using PREMIS with METS for exchange, Revised September 17, 2008 URL: http://www.loc.gov/standards/premis/guidelines-premismets.pdf.
7. Media Types, 2015, URL: https://www.iana.org/assignments/media-types/media-types.xhtml
8. METS, 2015, URL: http://www.loc.gov/standards/mets/
9. METS Profile Components, 2011, URL: http://www.loc.gov/standards/mets/profile_docs/components.html
10. METS Profiles, 2012, URL: http://www.loc.gov/standards/mets/mets-profiles.html
11. Producer, Submission Agreements: Glossary of Terms, 2015, URL: http://sites.tufts.edu/dca/about-us/research-initiatives/taper-tufts-accessioning-program-for-electronic-records/project-documentation/submission-agreements-glossary-of-terms/
12. Producer-Archive Interface Methodology Abstract Standard (PAIMAS), 2004, URL: public.ccsds.org/publications/archive/651x0m1.pdf
13. Producer-Archive Interface Specification (PAIS) – CCSDS, 2014, URL: public.ccsds.org/publications/archive/651x1b1.pdf
14. Records Creator, Submission Agreements: Glossary of Terms, 2015, URL: http://sites.tufts.edu/dca/about-us/research-initiatives/taper-tufts-accessioning-program-for-electronic-records/project-documentation/submission-agreements-glossary-of-terms/
15. Reference Model for an Open Archival Information System (OAIS), 2012, URL: public.ccsds.org/publications/archive/650x0m2.pdf

8. APPENDICIES

8.1. Appendix A: Quality requirements for a submission information package

Every SIP should follow the requirements set out in the common specification for information packages.

8.1.1. General requirements

Requirement 1.1: *It MUST be possible to include any data or metadata, regardless of its type or format, in a Submission Information Package.*

Requirement 1.2: *A Submission Information Package Specification MUST NOT restrict the means, methods or tools for exchanging it.*

Requirement 1.3: *The Submission Information Package Specification MUST NOT define the scope of data and metadata which constitutes an Information Package.*

Requirement 1.4: *A Submission Information Package SHOULD be highly scalable.*

Requirement 1.5: *A Submission Information Package MUST be machine-readable*

Requirement 1.6: *A Submission Information Package SHOULD be human-readable*

Requirement 1.7: *A Submission Information Package MUST support the preservation method best suited for the data.*

8.1.2. Identification of the Information Package

Requirement 2.1: *The Information Package type (SIP, AIP or DIP) MUST be clearly indicated.*

Requirement 2.2: *The Submission Information Package MUST clearly indicate the Content Information Type(s) of its data and metadata.*

Requirement 2.3: *A Submission Information Package MUST bear an identifier which is unique and persistent in the scope of the repository.*

Requirement 2.4: *A Submission Information Package SHOULD bear an identifier which is globally unique and persistent.*

Requirement 2.5: *All components of a Submission Information Package MUST bear an identifier which is unique and persistent within the repository.*

8.1.3. Structure of the Information Package

Requirement 3.1: *A Submission Information Package MUST be built in such a way that its data and metadata can be logically and physically separated from one another.*

Requirement 3.2: *The structure of the Submission Information Package SHOULD allow for the separation of different types of metadata*

Requirement 3.3: *The structure of the Submission Information Package MUST allow for the separation of data and metadata representations.*

Requirement 3.4: *The structure of a Submission Information Package MUST explicitly define the possibilities for adding additional logical components into the Information Package.*

Requirement 3.5: *A Submission Information Package MUST follow a common conceptual structure regardless of its technical implementation.*

Requirement 3.6: *A Submission Information Package MUST be implemented by one and only one implementation at any point in time.*

8.1.4. Information Package Metadata

Requirement 4.1: *Metadata in a Submission Information Package MUST be based on standards.*

Requirement 4.2: *Metadata in a Submission Information Package MUST allow for unambiguous use.*

Requirement 4.3: *A Submission Information Package MUST NOT restrict the addition of any additional metadata.*

8.2. Appendix B: Submission Agreement

Table 6: Submission agreement

Elements			Explanations
PROJECT INFORMATION			
Project			Elements of a transfer project.
	Project Name		Name of the transfer project (e.g. Transfer I, 2016).
	Project ID		Identification code of the transfer project (e.g. 201601122044).
CHANGE MANAGEMENT			
Version / Revision			Elements for tracking the changes in versions of the submission agreement.
		Release date	The date of the version.
		Change authority	The information about the person who changed the submission agreement (e.g. John Smith (The National Archives of Estonia)).
		Change description	A short textual description of the change.
		Section(s) affected	This element is meant for recording more detailed information about changes.
PRODUCER, ARCHIVE, DESIGNATED COMMUNITY			
Producer Organization			Elements describing the Producer.
	Organization name	Organization name	The official name of the organization
	Main Contact		Elements describing the main contact of the transfer project.
		Address	The address of the main contact.

Elements			Explanations
		Telephone	The telephone number of the main contact.
		E-mail	The e-mail of the main contact.
		Additional Information	Meant for recording any additional information needed to describe the contact.
	Individual Contacts		Elements describing other individual contacts of the organization.
		Name	The full name of the contact person.
		Role	The role of the contact person.
		Telephone	The telephone number of the contact person.
		E-mail	The e-mail of the contact person.
		Additional Information	Meant for recording any additional information needed to describe the contact.
Archive Organization			Elements describing the Archive.
	Organization name		The official name of the organization
	Main Contact		Elements describing the main contact of the transfer project.
		Address	The address of the main contact.
		Telephone	The telephone number of the main contact.
		E-mail	The e-mail of the main contact.
		Additional Information	Meant for recording any additional information needed to describe the contact.
	Individual Contacts		Elements describing other individual contacts of the organization.
		Name	The full name of the contact person.
		Role	The role of the contact person.
		Telephone	The telephone number of the contact person.
		E-mail	The e-mail of the contact person.

Elements			Explanations
		Additional Information	Element for recording any additional information needed to describe the contact.
Designated Community			Elements describing the Designated Community.
	Description		The textual description of the skills and knowledge base of the designated community.
	Individual Contacts		Elements describing the individual contacts of the designated community.
		Name	The full name of the contact person.
		Role	The role of the contact person.
		Telephone	The telephone number of the contact person.
		E-mail	The e-mail of the contact person.
		Additional Information	Meant for recording any additional information needed to describe the contact.
SUBMISSION INFORMATION PACKAGE			
Content and metadata			Elements describing the content and metadata of the submission information package.
	Description		A description of data origination, content and coverage.
	Platform Information		A short description of the source system.
	Representation Information		A description of the means to represent the data content (e.g. referencing to data dictionaries, decoding software etc.).
	Preservation Descriptive Information		A description for keeping data independently understandable (e.g. processing history, platform documentation, algorithm information, technical reports, user manuals, etc.).
	Supplemental Information		Meant for recording any additional information needed to describe the content or metadata of the SIP.

Elements		Explanations
	Access Constraints	A description of access restrictions and other legal or contractual access aspects.
Data Model		Elements describing the agreements for the SIP data model.
	Content Type	A short description of the content type (e.g. ERMS content).
	Reference	A reference to the full agreed data model description.
	Additional Information	A description of any other additional information (e.g. description of the physical folder structure of the SIP) related to the data model.
SUBMISSION SESSION INFORMATION		
Submission Session		Elements describing the agreements for the submission session.
	Submission Method	The description of the submission method (e.g. through a digital interface, a physical transfer).
	Delivery Schedule	A description of a delivery schedule (a submission session may have a routine or a complex schedule).
	Data Submission Inventory	A description of the complete inventory of data objects (and other items) in the submission session.
INGEST		
Submission Reception		Elements describing the agreements for the ingest.
	Validation	A description of procedures for the quality assurance.
	Error Handling	A description of procedures for the error handling.
	Receipt Confirmation	A description of the receipt confirmation.
SUBMISSION RISKS		

Elements			Explanations
Risks			Elements describing the risks and mitigation options of the submission.
	Risk Factor		Meant for listing all risk factors (e.g. the designated community is not properly defined) of the submission.
	Mitigation Option		Meant for listing all mitigation options (e.g. define the designated community together with producers) for the risks.

8.3. Appendix C: Terminology

Archival creator[47]	An organization unit or individual that creates records and/or manages those records during their active use.
Archive*	An Organisation that intends to preserve information for Access and use by a Designated Community.
Delivering organisation	The organisation delivering the package to the archive. For stating and extending the information use of the "Producer organisation name" and "Submitting organisation name" elements is recommended.
ERMS	A type of content management system known as an electronic records management system.
Information Package*	A logical container composed of optional Content Information and optional associated Preservation Description Information. Associated with this Information Package is Packaging Information used to delimit and identify the Content Information and Package Description information used to facilitate searches for the Content Information.
Ingest Functional Entity*	The OAIS functional entity that contains the services and functions that accept Submission Information Packages from Producers, prepares Archival Information Packages for storage, and ensures that Archival Information Packages and their supporting Descriptive Information become established within the OAIS.
OAIS*	The Open Archival Information System is an archive (and a standard: ISO 14721:2003), consisting of an organisation of people and systems that has accepted the responsibility to preserve information and make it available for a Designated Community.
Producing organisation[48]	The organizational unit or individual that has the authority to transfer records to an archive. Usually the producer is also the records creator, the organizational unit or individual that created and managed the records during their active use. This is not always the case, sometimes the producer is different from the records creator. For example: An author dies and her literary executor gains the authority to transfer her papers to an archive. The author is the records creator and the literary executor is the producer. For example: Department X gets reorganized out of existence and Department Y,

[47] Records Creator, Submission Agreements: Glossary of Terms, 2015, http://sites.tufts.edu/dca/about-us/research-initiatives/taper-tufts-accessioning-program-for-electronic-records/project-documentation/submission-agreements-glossary-of-terms/

[48] Producer, Submission Agreements: Glossary of Terms, 2015, http://sites.tufts.edu/dca/about-us/research-initiatives/taper-tufts-accessioning-program-for-electronic-records/project-documentation/submission-agreements-glossary-of-terms/

	which takes over the functional responsibilities of Department X, gains the authority to transfer the records of Department X to the archive. Department X is the records creator and Department Y is the producer. Counter example: The Department of Widget Science transfers some of its own records to the archive. The Department of Widget Science is the records creator and the producer.
Submission Information Package (SIP)*	An Information Package that is delivered by the Producer to the OAIS for use in the construction or update of one or more AIPs and/or the associated Descriptive Information.
Submitting organisation	Name of the organisation submitting the package to the archive. Extends the delivery information since it may be the case that the content of a creator is held by another part of the organisation.

***** Reference Model for an Open Archival Information System (OAIS), 2012, http://public.ccsds.org/publications/archive/650x0m2.