

Earth Observation Payload Data Ground Systems Infrastructure Evolution 2011-2014




LTDP SAFE

EO Collection and EO Product metadata separation Trade-Off

Ref: PDGS-SAFE-GMV-TN-12/0185

Version: 1.0

Date: 18th October 2012

Author	Héctor Hugo Pérez Alonso, GMV	<p>19/10/2012</p> <p>X </p> <hr/> <p>Firmado por: Héctor Hugo Pérez Alonso</p>
Reviewer	Fernando Ibáñez, GMV	<p>19/10/2012</p> <p>X </p> <hr/> <p>Firmado por: Fernando Ibáñez Casado</p>
Approver	Adrián Sanz, GMV	<p>19/10/2012</p> <p>X </p> <hr/> <p>Firmado por: Adrián Sanz Díaz</p>

The work described in this document was performed under ESA Contract.
Responsibility for the contents resides in the author or organisation that prepared it.

Table of Contents

1	Introduction	4
1.1	Purpose.....	4
1.2	Scope	4
1.3	Document Structure.....	4
1.4	Document Status.....	4
1.5	Applicable Documents	4
1.6	Reference Documents.....	5
1.7	Acronyms and Abbreviations	6
1.8	Definitions	6
2	CONTEXT OF COLLECTION METADATA SEPARATION TRADE-OFF.....	7
2.1	Collections and EO Products	7
2.2	Reasons for the separation in SAFE	7
2.3	Steps to follow	8
3	IMPACT IN THE FILE STRUCTURE	9
3.1	Location of ISO 19139 XML Schema file	9
3.2	Location of collection metadata files	10
3.3	Structure and links of Collection SAFE Packages	11
4	IMPACT IN METADATA DEFINITION	13
4.1	OGC EOP metadata	13
4.2	ISO/INSPIRE metadata.....	13
4.3	SAFE Own / QA4EO Metadata	13
5	CONCLUSION	14

List of Tables

Table 1-1: Applicable Documents	5
Table 1-2: Reference Documents.....	6
Table 1-3: Acronyms	6
Table 1-4: Definitions	6

1 Introduction

1.1 Purpose

This document contains the trade-off performed on the separation of the collection metadata and the Earth Observation Product metadata within the SAFE metadata model. The purpose of this document is to analyse the ways to perform this separation in order to be in line with HMA. This document evaluates the alternatives for separating both kind of metadata (EO Product and Collection) and chose the most adequate one for the context of SAFE.

1.2 Scope

The scope of this document is the PDR of the LTDP-SAFE project. This analysis has been requested as results of an action (A08) raised in the PDR-C Review within the LTDP-SAFE project. The results obtained from this analysis will be used as a basis to update the SAFE Core documentation for the PDR meeting.

1.3 Document Structure

This document is structured by the following blocks:

- Introduction
- Context of collection metadata trade-off, describing the concepts and situations at the origin of this analysis
 - Collections and EO products
 - Reasons for the separation in SAFE
 - Steps to follow
- Impact in the file structure
 - Location of ISO 19139 XML Schema
 - Location of collection metadata files
 - Structure and links of Collection SAFE Packages
- Impact in metadata definition
 - OGC EOP Metadata
 - ISO/INSPIRE Metadata
 - SAFE Own Metadata
 - SAFE QA4EO Metadata
- Conclusions

1.4 Document Status

This is the first version of the document issued for open discussion in the SAFE Wiki/Forum web page (<http://wiki.services.eoportal.org/tiki-index.php?page=LTDP+SAFE+Wiki>) before the PDR.

1.5 Applicable Documents

The following table lists the Applicable Documents that have a direct impact on the contents of this document.

Acronym	Title	Reference	Issue
[SAFE_PRIMER]	STANDARD ARCHIVE FORMAT FOR EUROPE PRIMER	PGSI-GSEG-EOPG-FS- 010-0001	2
[SAFE_CORE]	STANDARD ARCHIVE FORMAT FOR EUROPE CORE SPECIFICATIONS	PGSI-GSEG-EOPG-FS- 05-0001	2
[SAFE_REC_SPEC]	STANDARD ARCHIVE FORMAT FOR EUROPE RECOMMENDATION FOR SPECIALIZATIONS	PGSI-GSEG-EOPG-FS- 05-0002	2
[PDRC_REP]	SAFE PDR-C Review Report	PDGS-SAFE-GMV-RP- 12/0130	1.0
[SSS]	Software System Specification	SAFE-GMV-SSS-001	1.0

Table 1-: Applicable Documents

1.6 Reference Documents

Acronym	Title	Reference	Issue
[LTDP Guidelines]	European LTDP Common Guidelines	GSCB-LTDP-EOPG-GD-09- 0002 - 30/09/2010	1.1
[OAIS-RM]	Reference Model for an Open Archival Information System (OAIS)	Reference Model for an Open Archival Information System (OAIS) -650.0-B-1- January 2002-Blue Book- Copyright © 2002 Consultative Committee for Space Data Systems (CCSDS)	650.0- B-1
[OGF-DFDL]	OGF-DFDL Standard	http://www.ogf.org/dfdl/	
[SSS]	Software System Specification	SAFE-GMV-SSS-001	N/A
[XFDU]	XML Formatted Data Unit (XFDU) Structure and Construction Rules	XML Formatted Data Unit (XFDU) Structure and Construction Rules -661.0- B-1-September 2008-Blue Book-Copyright © 2008 Consultative Committee for Space Data Systems (CCSDS)	661.0- B-1
[XML_W3C]	XML Schema	http://www.w3.org/XML/ Schema	
[XML-SCHEMA]	XML Schema: Primer Second Edition -W3C Recommendation	XML Schema: Primer Second Edition -W3C Recommendation -October 28, 2004-Version 1.0- Copyright © 2004 World Wide Web Consortium (W3C)	1.0

Acronym	Title	Reference	Issue
[XML-SCHEMA-STRUCT]	XML Schema: Structures Second Edition -W3C Recommendation	XML Schema: Structures Second Edition -W3C Recommendation -October 28, 2004-Version 1.0-Copyright © 2004 World Wide Web Consortium (W3C)	1.0
[XML-SCHEMA-TYPES]	XML Schema: Data Types Second Edition -W3C Recommendation	XML Schema: Data Types Second Edition -W3C Recommendation - October 28, 2004-Version 1.0-Copyright © 2004 World Wide Web Consortium (W3C)	1.0
[MDDEF-TRADE_OFF]	Trade-off Metadata Definition	PDGS-SAFE-GMV-TN-12/0082	1.0
[SIMPL_TRADE_OF F]	Simplification design Trade-off	PDGS-SAFE-GMV-TN-12-0171	1.0

Table 1-: Reference Documents

1.7 Acronyms and Abbreviations

Acronym	Meaning
AIP	Archival Information Package
EO	Earth Observation
EOP	Earth Observation Profile
HMA	Heterogeneous Mission Access
ISO	International Organization for Standardization
LTDP	Long Term Data Preservation
OAIS	Open Archival Information System
OGC	Open Geospatial Consortium
PDGS	Payload Data Ground System
PDR	Preliminary Design Review
PDR-C	Preliminary Design Review – Core
QA4EO	Quality For EO
SAFE	Satellite Archive Format for Europe
SAFE 2.0 Draft	SAFE 2.0 presented at PDR-C (Externalisation trade-off)
XML	eXtensible Markup Language
XSD	XML Schema Definition

Table 1-: Acronyms

1.8 Definitions

Concept	Description
EO Metadata	XML Metadata file

Table 1-: Definitions

2 CONTEXT OF COLLECTION METADATA SEPARATION TRADE-OFF

The following sections describe the concepts and situations which have originated the need of performing the collection metadata separation trade-off.

2.1 Collections and EO Products

An EO Product is an existent set of EO Data, identifiable by some values such as spatial and/or temporal extent and/or a specific band, etc., which is located physically within a larger dataset called EO Collection.

An EO Collection describes a set of related EO Products. Thus EO Collection metadata are those metadata identified for a specific set of related EO products

When defining the structure of metadata for EO Data, the HMA project decided to separate EO Collection Metadata and EO Product Metadata.

For both types, HMA specified a different metadata model composed by the following standards:

- EO Products
 - OGC Earth Observation profile for Observations and Measurements (10-157r3)
- EO Collections
 - ISO 19115:2003, Geographic Information – Metadata
 - ISO 19115:2003/Cor 1 2006, Geographic information – Metadata - Corrigendum 1
 - ISO/TS 19139:2007, Geographic information -- Metadata -- XML schema implementation

Hence, HMA provided a clear separation between the metadata for EO Products and the metadata for EO Collections.

2.2 Reasons for the separation in SAFE

During the SRR of the LTDP-SAFE project it was decided to adopt the OGC Earth Observation profile for Observations and Measurements (OGC EOP O&M 10-157r3) as the base metadata model for SAFE. As said before, this model corresponds to the EO Products model specified by HMA.

Additionally, it was agreed to search for any useful ISO 19115 and INSPIRE metadata which were not addressed in the OGC EOP O&M model in order to include them in the SAFE metadata model.

These decisions were applied during the SAFE 2.0 Draft design presented in the PDR-C, with the following consequences:

- a) Some ISO 19115 metadata were added as an extension to the SAFE metadata model
- b) Some INSPIRE metadata were added as an extension to the SAFE metadata model

- c) Some ISO and INSPIRE metadata were considered useful, but were not added to the SAFE metadata model, since it was considered that they could be directly used via the *eop:EarthObservation/om:metadata* link (link to ISO 13139 metadata), which is already present in the OGC EOP metadata model.

Since the ISO 19115 metadata are collection metadata and INSPIRE metadata are basically inherited from ISO (thus collection metadata as well), the proposed SAFE 2.0 draft was mixing EO Products metadata and EO Collection metadata, whose separation was one of the goals of the HMA project.

Additionally, even if the *eop:EarthObservation/om:metadata* link allows to include ISO/TS 19139 metadata to the OGC EOP XML files, the use of this possibility goes against the aforementioned separation, so it has to be avoided.

Therefore, it is needed to redesign the SAFE metadata model structure in order to separate again the EO Product metadata and the EO collection metadata.

2.3 Steps to follow

The steps to follow to change the design of SAFE 2.0 Draft version (presented at the PDR-C) and settle a clear separation between EO Product metadata and EO Collection metadata are listed below:

1. First, extract the ISO and INSPIRE metadata from the SAFE extensions out to an independent schema (not an extension) and change them to follow ISO 19139 (or ISO19115-2 and ISO19139-2).
2. Keep the main XML Metadata file compliant to the OGC EO Profile O&M. This XML file will be located inside the EO Package.
3. Define an extension for the rest of metadata (particular SAFE own metadata, QA4EO metadata, etc.)
4. Add a reference in the EO Product package to the separate EO Collection Metadata file, with either of these options:
 - a. Store the EO Collection Metadata in a collection specific "place" (to be defined in the following sections) and link the SAFE EO Product Metadata stored in the SAFE Package with the EO Collection Metadata.
 - b. Store the EO Collection Metadata and the SAFE EO Product Metadata in the same SAFE package. Here the EO Collection Metadata may be redundantly stored, but handling may be easier.

Note that the meaning of "place" will be resolved in the next section.

3 IMPACT IN THE FILE STRUCTURE

Starting from the previous chapter, the following files have to be added to the model:

- a) The ISO 19139 XML Schema specifying the collection types (ISO/TS 19139:2007, Geographic information -- Metadata -- XML Schema implementation).
- b) An XML file for the collection metadata (compliant to the previous schema).

With these files in the model, the EO Product SAFE Packages belonging to the same Collection will need to point somehow to the same set of EO Collection metadata. The following figure shows an example of three collections with different sets of EO Products.

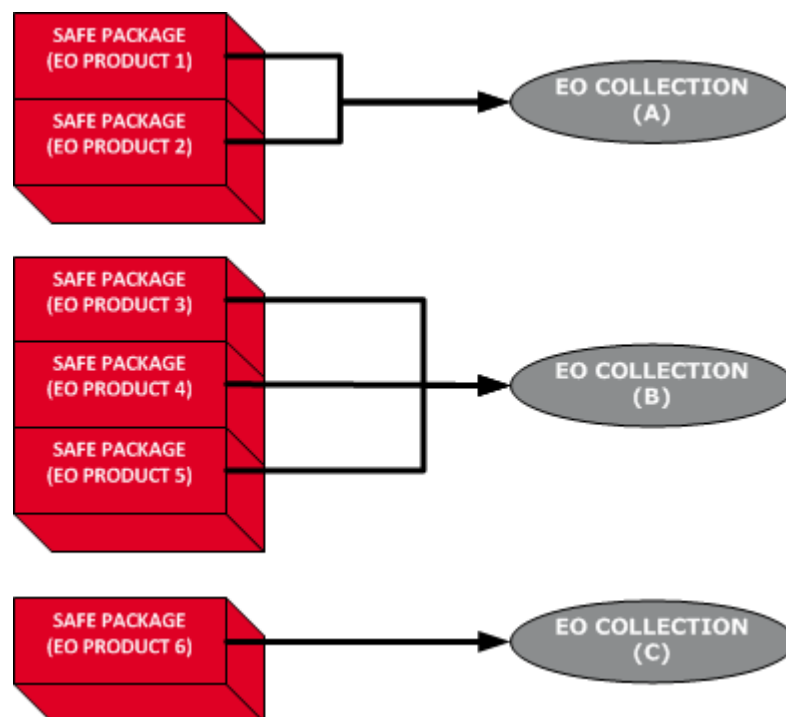


Figure 3-: Scenario of EO Product SAFE Packages and collections

The following subsections explain the main decisions taken to insert the files in the model.

3.1 Location of ISO 19139 XML Schema file

The ISO 19139 XML Schema was already included in the structure, since it is used in the Observations and Measurements main schema (observation.xsd). For the SAFE 2.0 Draft version (presented at the PDR-C) each XML Schema was stored in a specific representation SAFE Package, and this was the case for this one as well.

For the new version, all the XML Schemas have been packed in a few SAFE Packages, in the aim of simplifying the whole structure. However, it seems appropriate to keep the ISO 19139 XML Schema in a specific SAFE Package to clearly separate the collection files from the others. This SAFE Package is

named EO Collection Schema Base Package. For more details about this structure please refer to [SIMPL_TRADE_OFF].

Therefore, as conclusion, *the ISO 19139 XML Schema file will be stored in a specific Representation Package*, and of course will be referenced from the files that need to use it.

3.2 Location of collection metadata files

Two alternatives are envisaged for the storage of the collection metadata files:

1. *Store the EO Collection Metadata in a collection specific "place" and link the SAFE EO Product Metadata with the EO Collection Metadata located at that "place".*

This option avoids including the same collection metadata file in all the SAFE Packages whose EO Products belong to the same collection. On the other hand, it increases the complexity of the solution, since it requires to define the location for the Collection metadata files (probably at the archive), so the same constraints applied for the other externalised files should be applied for the collection metadata files as well:

- These files should be located in the archive, and they should be accessible through any means (API, tools, etc.).
- The preservation of these files should be ensured via a specific package or similar (as for the representation files or the auxiliary files).
 - The EO Collection Schema should be located in an EO Collection Schema Base Package.
 - The EO Collection Metadata from the above XML Schema should be located in an EO Collection Package.
- The links from the SAFE Packages to the EO Collection metadata files could become more complicated.

It is the recommended solution.

2. *Store the EO Collection metadata and the SAFE EO Product Metadata in the same SAFE Package.*

This solution a priori seems quite simpler than the first one, although it implies to include the EO Collection file inside each SAFE Package whose EO Product belongs to the EO Collection.

The main problem of this redundancy could arise when an EO Collection metadata file has to be modified (in particular, the identifier used to make reference to the collection metadata XML file), and it is needed that the rest of EO Products of the collection become aware of the change. This would require a re-transcription of all the SAFE Packages containing EO Products of the collection, in order to replace the obsolete collection metadata file by the new one. This re-transcription is something that has to be avoided by any mean.

At the time of this trade-off, it is unknown whether this scenario (changing the existing EO collection metadata files) could occur several times or not but it has been taken into account in order to ensure that this possibility doesn't impact in the proposed solution.

Therefore, although at the PDR-C it was stated that the design should be as simple as possible (see action PDR-C_A02), we have to adopt the first solution for its safeness against possible re-transcriptions of products, in spite of the increase of the complexity. Furthermore, it is the solution used in the [SIMPL_TRADE_OFF].

As conclusion, *the collection metadata files will be stored in a new kind of SAFE Package (this is the “place” that was mentioned before), and the policy of linking the EO Product metadata with the EO Collection metadata is described in the following section.*

3.3 Structure and links of Collection SAFE Packages

The following figure describes the structure of SAFE Packages related to the Collections and their links. This figure only depicts the XML files, since our focus is the metadata and schemas. Hence the binaries and data products of the EO Product SAFE Package are not displayed in the picture for simplicity.

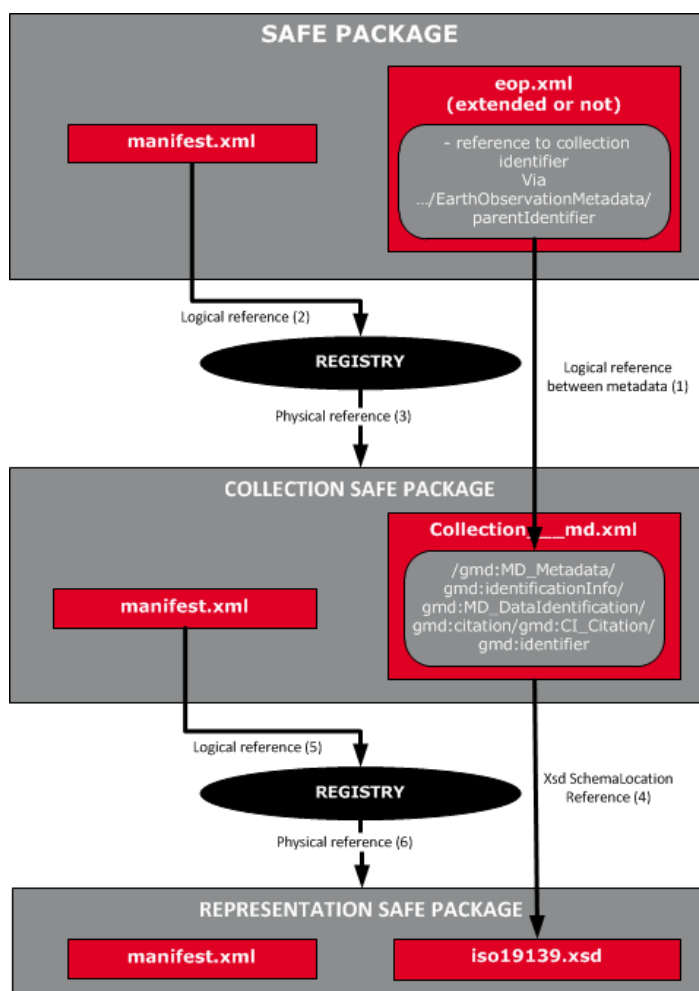


Figure 3--: Structure of SAFE Packages and XML files for collection metadata

At the top of the figure there is an Earth Observation SAFE Package (EO Product Package), with the manifest (`manifest.xml`) and the OGC EO Metadata file (`eop.xml`). This package belongs to a specific Earth Observation collection, so it shall be linked to the corresponding EO Collection SAFE Package. Two kinds of references are foreseen:

1. **Logical reference inside the metadata.** This reference is in charge of establishing the semantic link between the EO Product XML Metadata file and the EO Collection Metadata file. It is defined by a simple identifier/code. It is defined as follows:

- The following ISO 19139 metadata element shall be present inside all the XML Collection metadata files, and it shall contain the identifier of the ISO 19139 metadata set:

/gmd:MD_Metadata/gmd:identificationInfo/gmd:MD_DataIdentification/gmd:citation/gmd:CI_Citation/gmd:identifier

- The following OGC EOP metadata element shall contain the corresponding ISO 19139 metadata set identifier:

/EarthObservation/eop:EarthObservationMetadata/parentIdentifier

2. **Reference through the Registry to the Collection SAFE Package.** This is the link that defines the physical location of the SAFE EO Collection Package. This will be the common way of making reference to external packages in the new SAFE 2.0. The Registry acts as a mapping between logical and physical references (2 and 3 in the figure). For more information about this kind of links refer to [SIMPL_TRADE_OFF].

In the middle of the figure there is a SAFE EO Collection Package, which is the new kind of package defined in section 3.2. It contains a manifest (manifest.xml) and the EO Collection Metadata file (e.g., collection_md.xml). The latter is the XML file with the metadata which are common to a specific collection, and it shall be compliant with the ISO 19139 XML Schema. This EO Collection ISO 19139 XML Schema should be located in a Representation Information Package (EO Collection Schema Base Package). Again, in this case we have two kinds of references:

1. **XSD schemaLocation reference.** This reference is inside the XML file (it comes from the XML specification), and points to the location of the XML Schema to be used for validations. In SAFE these references will point to relative paths within the own package. For more information about this kind of links refer to [SIMPL_TRADE_OFF].
2. **Reference through the Registry to the Representation SAFE Package.** This is the link that defines the physical location of the SAFE Representation Package (EO Collection Schema Base Package) with the ISO 19139 Schema (defined in section 3.1). It maps again the logical and physical references (5 and 6 in the figure). For more information about this kind of links refer to [SIMPL_TRADE_OFF].

4 IMPACT IN METADATA DEFINITION

4.1 OGC EOP metadata

The only impact in OGC EOP O&M metadata is the use of the */EarthObservation/eop:EarthObservationMetadata/parentIdentifier* to make reference to the collection metadata set. This is explained in the previous section.

4.2 ISO/INSPIRE metadata

In the SAFE 2.0 Draft version (designed for the PDR-C), a subset of ISO and INSPIRE metadata were identified and added to a specific extension of SAFE, trying to address information that was not covered by the OGC EOP O&M metadata and trying to reduce the number of types in the SAFE XML Schemas.

Now, in this trade-off, we have seen that the information in the ISO/INSPIRE extension was in fact collection information, and hence it has to be placed in a specific (and separated) XML file compliant to the ISO 19139 XML Schema.

This situation simplifies the whole picture. The ISO 19139 contains all the metadata types for collections and it has to be used as a base. Therefore all the collection metadata types will be available for their use in a SAFE EO Collection Package and the goals are reached:

- the SAFE format provides all the means to generate SAFE EO Collection Packages according to the needs of the specializations
- the SAFE format provides the means to link the EO SAFE Products with the SAFE EO Collection Packages (see previous section).

4.3 SAFE Own / QA4EO Metadata

These metadata will be placed in a specific extension of SAFE (safe-own). Their structure remains the same, but their definition has to be changed as per the RIDs of the PDR-C (some considerations were provided about the right way of adding extensions, some XML examples were wrong in the Core Specifications, etc).

5 CONCLUSION

The separation of collection metadata and EO Product metadata in SAFE is feasible, and leads to a better convergence with the HMA approach.

It adds some complexity to the solution, because:

- It leads to handle a new type of SAFE Package: the **SAFE EO Collection Packages**, in order to avoid dangerous redundancies that could lead to update all the products upon a re-transcription. The new packages are:
 - SAFE EO Collection Schema Base Package, it contains the ISO 19139 Schema.
 - SAFE EO Collection Package, it contains the ISO 19139 XML.
- it requires to use a particular policy of references between metadata

On the other hand, it simplifies the way to handle the ISO/INSPIRE metadata in the SAFE metadata model, since in SAFE 2.0 they will directly follow the ISO 19139 Schema.

Additionally, thanks to the separation between collection and EO Product metadata, this solution is more flexible, since it allows for including any ISO 19139 collection metadata in the SAFE EO Collection Packages without affecting the EO Product metadata.

Change Record

Issue	Revision	Date	Change Status	Origin
1	Pr1	18th October 2012	Initial version	Héctor Hugo Pérez Alonso GMV

< End of Document >