

Earth Observation Payload Data Ground Systems Infrastructure Evolution 2011-2014


LTDP SAFE

Data Packaging Formats Trade-off

Ref: PDGS-SAFE-GMV-TN-12/0073

Version: 1.0

Date: 18th May 2012

Author	Adrián Sanz, GMV	18/05/2012
Reviewer		X 
Approver		A. Sanz GMV SAFE Project Manager Firmado por: Adrián Sanz Díaz

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

©GMV 2012; all rights reserved (GMVAD 20763/12 V1/12)
This document shall only be reproduced for the agreed purpose for which it has been supplied.

ESA UNCLASSIFIED - For Official Use

Table of Contents

1	Introduction	4
1.1	Purpose.....	4
1.2	Scope	4
1.3	Document Status	4
1.4	Applicable Documents	4
1.5	Reference Documents.....	4
1.6	Acronyms and Abbreviations	5
2	Data Packaging	6
3	Data Packaging Formats Alternatives	7
3.1	Metadata Encoding and Transmission Standard (METS)	7
3.2	MPEG-21 Digital Item Declaration Language (DIDL)	8
3.3	IMS Content Packaging (IMS-CP)	10
4	Data Packaging Formats Assessment	12
4.1	Standardisation	12
4.2	Open Source Specification.....	12
4.3	Documentation	13
4.4	Flexibility.....	13
4.5	OAIS compliance	14
4.6	Worldwide usage	14
5	Conclusions	15
6	References.....	17

List of Tables

Table 1-1: Applicable Documents	4
Table 1-2: Reference Documents.....	4
Table 1-3: Acronyms	5
Table 4-1: Standardisation assessment	12
Table 4-2: Open Source Specification assessment.....	13
Table 4-3: Documentation assessment.....	13
Table 4-4: Flexibility assessment.....	14
Table 4-5: OAIS Compliance assessment.....	14
Table 4-6: Worldwide usage assessment	14
Table 5-1: Data Packaging format comparison summary table	16

1 Introduction

1.1 Purpose

This document provides the trade-off of the existing packaging standards that can be adopted for the LTDP SAFE project. The purpose of this document is to analyse the existing formats available nowadays to identify if there is a potential benefit on using a different alternative.

1.2 Scope

The aim of this document is to present a trade-off of the available packaging standards identified as candidates for LTDP SAFE project. This information will be used for engineering design of the SAFE format to be assessed during the PDR-C process.

1.3 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-C.

1.4 Applicable Documents

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

Acronym	Title	Reference	Issue
[SRR_REP]	SAFE SRR Review Report	SAFE-GMV-REP-001	1.0
[OAIS]	Reference Model for an Open Archival Information System (OAIS)	CCSDS 650.0-B-1; January 2002	B.1
[XFDU]	XML FORMATTED DATA UNIT (XFDU) STRUCTURE AND CONSTRUCTION RULES	CCSDS 661.0-B-1; September 2008	B.1
[METS]	METS: PRIMER AND REFERENCE MANUAL	Revised 2010-4	1.6
[MPEG-21]	MPEG-21 Overview v.5	ISO/IEC JTC1/SC29/WG11/N5231	v.5
[IMS-CP]	IMS Content Packaging Specification Primer	Public Draft v2.0 March 2007	1.2

Table 1-: Applicable Documents

1.5 Reference Documents

Acronym	Title	Reference	Issue
[SSS]	Software System Specification	SAFE-GMV-SSS-001	1.0

Table 1-: Reference Documents

1.6 Acronyms and Abbreviations

Acronym	Meaning
AIP	Archive Information Packages
CCSDS	Consultative Committee for Space Data Systems
DLF	Digital Library Federation
FOXML	Fedora Object XML
IEC	International Electrotechnical Commission
ISO	International Organization for Standardization
LC	Library of Congress
METS	Metadata Encoding and Transmission Standard
MPEG	Moving Picture Experts Group
NDMSO	Network Development and MARC Standards Office
OAI-ORE	Open Archive Initiative – Object Reuse and Exchange
OAIS	Open Archival Information System
PDGS	Payload Data Ground System
PDR-C	Preliminary Design Review – Core
SAFE	Satellite Archive Format for Europe
SRR	Software Requirements Review
SSS	Software System Specification
W3C	World Wide Web Consortium
XFDU	XML Formatted Data Unit
XML	eXtensible Markup Language
XSD	XML Schema Definition

Table 1--: Acronyms

2 Data Packaging

OAIS Reference Model defines the **information object** concept as combination of:

- Data Object (either physical or digital)
- Representation Information (allowing the full interpretation of the data into meaningful information)

These elements must be encapsulated within an **Information Package** to be used within an OAIS, containing the information needed to aid in the preservation of the Content Information (e.g. manifest and package identifier) and to facilitate data search and retrieval.

There are several implementations available to provide an encapsulated method as the OAIS Reference Model does not specify any particular mechanism for packaging information. There are different categories of metadata for specific object types so there is no common standard fulfilling the needs of every digital object type (this can cause interoperability problems transferring metadata or objects between repositories).

XML Formatted Data Unit (**XFDU**), is the standard currently used by SAFE. This format is standardised by the CCSDS and consists of:

- an XML Manifest
- all files contained in the Manifest
- all files and XFDUs referenced from it

An XFDU manifest file is an XML document containing the following complex types/elements:

- Package Header: metadata that apply to the whole XFDU Package.
- Information Package Map: hierarchical view of the content of the XFDU.
- Data Object Section: contains any number of Data Object elements.
- Metadata Section: records all of the metadata for all items in the XFDU package.
- Behavior Section: may contain any number of behaviour objects (each behaviour object can be used to associate executable behaviours with one or more Content Units in the containing XFDU).

The following sections try to summarise the existing alternatives and the possible benefits on using a new data packaging approach in SAFE.

3 Data Packaging Formats Alternatives

Some standards have been identified as possible candidates to replace current XFDU implemented in SAFE, but only a few of them have been considered suitable according to the SAFE needs.

Thus for example, some formats like FOXML (simple format Oriented to Fedora repositories), OAI-ORE (a standard for Web resources), BagiIT (hierarchical file packaging format for storage and transfer of arbitrary digital content not yet standardised) ... have been discarded from the present analysis as they are not verifying the minimum SAFE needs.

Next sections provide an overview of the qualified candidates.

3.1 Metadata Encoding and Transmission Standard (METS)

3.1.1 Description

The Metadata Encoding and Transmission Standard (METS) is a data encoding and transmission specification, expressed in a standardised XML schema, that provides the means to convey the metadata necessary for both the management of digital objects within a repository and the exchange of such objects between repositories (or between repositories and their users).

METS is an initiative of a group of libraries and related agencies called Digital Library Federation¹ (DLF) establishing the use of information technologies to extend collections and services.

The Library of Congress² serves as maintenance agency for METS by hosting the website³ and providing additional support. Among other responsibilities, the Library of Congress is committed to maintaining editorial control over METS format, its XML Schema, the METS profiles and the official documentation.

3.1.2 Structure

METS is an open standard for encoding descriptive (information about the intellectual content of a digital object), administrative (information needed by the repository for the long-term management), and structural (information about the relationships between digital objects) metadata regarding objects within a digital library expressed using an XML schema.

The METS schema perspective is composed of:

- **METS Header:** containing metadata describing the METS document itself (creator, editor, etc...)

¹ <http://www.diglib.org/>

² <http://www.loc.gov/standards/>

³ <http://www.loc.gov/standards/mets/>

- Descriptive Metadata: this section may point to descriptive metadata external to the METS document, or contain internally embedded descriptive metadata, or both.
- Administrative Metadata: provides information regarding how the files were created and stored, intellectual property rights, metadata regarding the original source object, and provenance information.
- File Section: This section lists all files comprising the electronic versions of the digital object.
- Structural Map: It outlines a hierarchical structure for the digital object, and links the elements of that structure to content files and metadata that pertain to each element.
- Structural Links: It allows METS creators to record the existence of hyperlinks between nodes in the hierarchy outlined in the Structural Map.
- Behaviour: Used to associate executable behaviours with content in the METS object.

This standard is very flexible and can be adapted to different necessities. This flexibility is obviously an advantage, but it can be also considered as a disadvantage when transferring METS files between different institutions.

A common practice among METS users is required to be established in order to keep the flexibility level but facilitating their use by different actors. To this end, there are METS profiles developed along with a formal registration process that makes the profiles visible to others.

The description of the components of a METS profile is available at the METS website⁴.

3.2 MPEG-21 Digital Item Declaration Language (DIDL)

3.2.1 Description

MPEG-21 is an open standard-based multimedia framework designed to support content interoperability across communities. The content along with some additional relevant information is called Digital Item.

MPEG-21 is organized into several independent parts, where "Part 2", is the Digital Item Declaration (DID) abstract data model (ISO/IEC 21000-2:2003). This data model specifies the Digital Item Declaration Language (DIDL) for the standard representation in XML of a Digital Item. The entities of the DID model are represented in XML by the elements and attributes of DIDL.

Moving Picture Experts Group (MPEG) is an ISO/IEC working group in charge of standards developments for coded representation of digital audio, video and related data since 1988.

⁴ http://www.loc.gov/standards/mets/profile_docs/components.html

3.2.2 Structure

In the MPEG-21 framework, Digital Items are modelled according to the MPEG-21 DID standard. This standard specifies the declaration of Digital Items in three distinct sections:

- Abstract Model: A set of abstract terms and concepts declaring Digital Items. MPEG-21 DID defines a normative for XML representation of Digital Items based on the Model.
- Representation of the Model in XML: The description of an XML syntax for each of the entities defined in the Abstract Model. This XML-based syntax is referred to as the DIDL. A DID represented according to the DIDL syntax is referred to as a DIDL document.
- XML Schema⁵: A W3C XML Schema specifying the DIDL syntax and constrains for the structure of DIDL documents.

This DID model does not define a language in and of itself. Instead, the model provides a common set of abstract concepts and terms that can be used to define such a scheme, or to perform mappings. Some of these concepts are:

- **Container**: is a structure that allows *items* and/or *containers* to be grouped
- **Item**: is a grouping of sub-*items* and/or *components* that are bound to relevant *descriptors*. These *descriptors* contain information about the *item*.
- **Component**: is the binding of a *resource* to a set of *descriptors*. These *descriptors* are information concerning all or part of the specific *resource* instance.
- **Descriptor**: associates information with the enclosing entity. This information may be a *component*, or a textual *statement*.
- **Resource**: is an individually identifiable Asset such as a video or audio clip, an image, or a textual Asset. It may also potentially be a physical object.
- **Statement**: is a literal textual value that contains information, but not an Asset.

Unlike METS and XFDU, a DIDL document is a single hierarchical XML 1.0 tree of digital resources, with metadata nested within the objects to which they refer. [1]

DIDL documents hold at least one container. These containers can be recursively defined and each one can hold one or more Items (each item can hold one or more Components or Items).

⁵ "didl.xsd" at: http://standards.iso.org/ittf/PubliclyAvailableStandards/MPEG-21_schema_files/did/

Components can hold one or more Resources that are the leaf nodes in the data model; they may contain URIs for data objects (PDFs, MPEGs, HTML, etc...) or the data object itself.

3.3 IMS Content Packaging (IMS-CP)

3.3.1 Description

Content Packages are the standard way of describing learning content that can be read by many educational programs. IMS Content Packaging (IMS-CP) is the most widely used format.

The IMS Content Packaging public specification is used to define the data format for the exchange of learning resources (or learning objects) from one program to another, facilitating easier delivery, reuse and sharing of materials. The specification focuses on the packaging and transport of resources, but doesn't determine the nature of those resources. The specification allows adopters to gather, structure and aggregate content in an unlimited variety of formats.

IMS Content Packaging Specification has been developed by the IMS Global Learning Consortium⁶ and it is initially designed for use with physical file packages (typically in ZIP format), but suitable for logical file packages as well in future versions.

The IMS Global Learning Consortium (IMS GLC) is a global, non-profit, member organization attempting to enable the growth and impact of learning technology in the education and corporate learning sectors worldwide.

Current specification (v1.2) is undergoing standardisation by ISO/IEC.

3.3.2 Structure

The IMS Content Package is composed by two major elements:

- Manifest: an XML file describing the package containing the following sections:
 - Metadata: XML element describing the whole manifest
 - Organisation: XML element describing the organizations of the content within a manifest
 - Resources: XML element containing references to all resources and media elements needed for a manifest (including metadata and references to external files)
 - Sub-Manifest: one or more optional nested manifest.
- File Resources: The actual media elements (text files, images, ...) organised in sub-directories.

⁶ <http://www.imsglobal.org>

XML manifest file wrapped up inside a zip file. The learning content itself is either included in the zip file if it is HTML (or other media that can run on its own) or else is referenced as a URL from within the manifest.

4 Data Packaging Formats Assessment

The formats described above have been analysed according the following parameters that are considered relevant for the SAFE project:

- Standardisation
- Open source specification
- Documentation
- Flexibility
- Compliance with OAIS
- Widespread use

The parameters described above have been assessed according the following criteria:

- (0) : Not compliant with SAFE needs
- (1) : Roughly compliant with SAFE
- (2) : Partially compliant with SAFE needs
- (3) : Fully compliant with SAFE needs

4.1 Standardisation

The use of standard specifications has been identified as one of the major requirements for SAFE. In that sense, all the suitable candidates identified in this trade-off are standards.

MPEG-21 DIDL and XFDU are ISO standards (XFDU is standardised by CCSDS which is an ISO TC20/SC13 standards body) therefore the international acceptance is granted.

IMS CP is undergoing standardisation by ISO. This standardisation process was initially expected by 2009 but there is no official news about the status of this process nowadays.

METS is standardised by the U.S. Library of Congress by the Network Development and MARC Standards.

	XFDU	METS	MPEG-21	IMS-CP
Standardisation	3	3	3	1

Table 4--: Standardisation assessment

4.2 Open Source Specification

Other relevant aspect to be considered in the selection of a new data packaging format for SAFE is the need to use open source solutions to

assure the format availability in the long-term that is no licenses restrictions will impact in the preservation of the data.

The specifications for all data packaging formats described in this trade-off, are open source, and they are available free of charge for public access.

	XFDU	METS	MPEG-21	IMS-CP
Open Source Specification	3	3	3	3

Table 4-: Open Source Specification assessment

4.3 Documentation

The data packaging format to be used by SAFE should be well specified and the appropriate documentation should be freely accessible to be used in the long-term.

Examples, schemas and tutorial documentation for the data packaging formats described in this trade-off, are all available free of charge for public access.

The analysed documentation seems to be enough for implementing any of the mechanism within SAFE.

	XFDU	METS	MPEG-21	IMS-CP
Documentation	3	3	3	3

Table 4-: Documentation assessment

4.4 Flexibility

SAFE products will hold different data objects coming from several missions therefore a flexible data packaging format is a must.

METS allows considerable flexibility in the structuring of the administrative metadata as it does not define a vocabulary or syntax for encoding metadata. Administrative metadata can be expressed according to many current community defined standards, or locally produced XML schemas.

In this aspect XFDU is very similar to METS, but is better designed for complex data sets as it allows recursive inclusion of XFDU files in the same unit. XFDU is more data-centric rather than document-centric (as METS). It scales better than METS to very large file structures because it can reference external XFDU documents.

In the other hand, the MPEG-21 DID provides a set of abstract terms and concepts to form a useful model for defining different Digital Items. The goal of this model is to be as flexible and general as possible. This model helps to provide a common set of abstract concepts and terms that can be used for defining or mapping existing schemes.

The IMS CP specification is independent from the nature of the physical files and allows to gather, structure, and aggregate content in an unlimited variety of formats.

	XFDU	METS	MPEG-21	IMS-CP
Flexibility	3	3	3	3

Table 4-: Flexibility assessment

4.5 OAIS compliance

XFDU was developed by CCSDS as an expression of the metadata ontology (classes, attributes, relations ...) described in their OAIS model. Thus XFDU provides predefined metadata categories and classes via enumerated attributes following the OAIS information model.

METS is very close to XFDU, but the structure of METS documents does not strictly reproduce the structure of the OAIS Information Model.

MPEG-21 DIDL can be aligned with the OAIS Reference Model [2] but it is only a mapping approach as the OAIS ontology is not followed.

IMS-CP provides the following types of packages: stand-alone package, bare manifest, meta-package, archive package and the specialized package. This classification seems not be compliant with the OAIS model. No mapping with OAIS has been found.

	XFDU	METS	MPEG-21	IMS-CP
OAIS compliance	3	2	1	0

Table 4-: OAIS Compliance assessment

4.6 Worldwide usage

The implementation of a data packaging format extensively used within the community is considered convenient for SAFE in order to facilitate the SAFE products manipulation by new users not familiarised with SAFE or reluctant to adopt the SAFE format in their archives.

According to this, MPEG-21 format is the most extended format among the formats considered in this trade-off. MPEG-21 is closely supported by the industry and it is used by numerous companies to develop multimedia applications. However MPEG-21 has not applied to large science data collections yet.

IMS-CP and METS are extensively used but only among the targeted users for whom the packaging format was developed i.e. IMS-CP for e-learning, METS for digital libraries.

On the contrary, XFDU is not very extended due to the standard complexity only used by the space / scientific community.

	XFDU	METS	MPEG-21	IMS-CP
Worldwide usage	1	2	3	3

Table 4-: Worldwide usage assessment

5 Conclusions

Three data packaging standards have been analysed as possible candidates to replace XFDU in SAFE. The outcomes of this analysis can be summarised as follows:

METS

- Standardised by the U.S. Library of Congress
- Open source specification
- Public and free of charge documentation available
- Flexible format
- OAIS compliant but not strictly reproducing OAIS Information Model
- Extensively used by the digital libraries community

MPEG-21 DIDL

- ISO standard
- Open source specification
- Public and free of charge documentation available
- Flexible format
- The format can be aligned with OAIS Information Model (mapping)
- Extensively used in multimedia

IMS-CP

- Not ISO standard yet
- Open source specification
- Public and free of charge documentation available
- Flexible format
- Not aligned with OAIS
- Extensively used in educational environments

XFDU

- ISO standard
- Open source specification
- Public and free of charge documentation available
- Flexible format, allowing complex data sets

- Compliant with the OAIS information model (predefined metadata categories and classes).
- Not extensively used due to its complexity

The following table provides a quantitative approach and tries to show at which level the SAFE needs are covered by the analysed standards:

	XFDU	METS	MPEG-21	IMS-CP
Standardisation	3	3	3	1
Open source specification	3	3	3	3
Documentation	3	3	3	3
Flexibility	3	3	3	3
OAIS compliance	3	2	1	0
Worldwide usage	1	2	3	3
Mean	2.67	2.67	2.67	2.17

Where

- (0) : Language is not compliant with SAFE needs
- (1) : Language is roughly compliant with SAFE
- (2) : Language is partially compliant with SAFE needs
- (3) : Language is fully compliant with SAFE needs

Table 5-: Data Packaging format comparison summary table

According the presented results and the information provided in previous sections, there are no clear advantages on replacing the XFDU format currently implemented in SAFE by any of the identified candidates. **XFDU is still recommended as the most adequate packaging format for SAFE.**

6 References

- [1] Alex Ball; UKOLN, University of Bath; February 2006; "*Briefing Paper: the OAIS Reference Model*". (<http://www.ukoln.ac.uk/projects/grand-challenge/papers/oaisBriefing.pdf>)
- [2] Jeroen Bekaert; Xiaoming Liu; and Herbert Van de Sompel; Sept. 2005; "*Representing Digital Assets for Long-Term Preservation using MPEG-21 DID*". (<http://arxiv.org/abs/cs.DL/0509084>)
- [3] JEROEN BEKAERT; HERBERT VAN DE SOMPEL; Aug 2005; "*Representing Digital Assets using MPEG-21 Digital Item Declaration*". (<http://arxiv.org/abs/cs/0508065>)

Appendix A XFDU

A complete set of sample files can be found at:

<http://public.ccsds.org/publications/archive/661x0b1.pdf>

Appendix B METS

A complete set of sample files can be found at:

<http://www.loc.gov/standards/mets/mets-examples.html>

Appendix C MPEG-21 DIDL

A complete set of sample files can be found at:

<http://mpeg.chiariglione.org/technologies/mpeg-21/mp21-did/index.htm>

Appendix D IMS-CP

A complete set of sample files can be found at:

<http://www.imsglobal.org/content/packaging/>

Change Record

Issue	Revision	Date	Change Status	Origin
1	Pr1	18/05/2012	First Version	Adrian Sanz GMV

< End of Document >

