

## Earth Observation Payload Data Ground Systems Infrastructure Evolution 2011-2014


LTDP SAFE

### Representation information in XML Trade-Off

Ref: PDGS-SAFE-GMV-TN-12/0066

Version: 1.0

Date: 7th May 2012

<b>Author</b>	Adrián Sanz, GMV	07/05/2012
<b>Reviewer</b>		X 
<b>Approver</b>		A. Sanz 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 20710/12 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

<b>1</b>	<b>Introduction .....</b>	<b>4</b>
1.1	Purpose.....	4
1.2	Scope .....	4
1.3	Document Status .....	4
1.4	Applicable Documents .....	4
1.5	Reference Documents.....	5
1.6	Acronyms and Abbreviations .....	5
<b>2</b>	<b>XML schema.....</b>	<b>6</b>
<b>3</b>	<b>Representation information schema.....</b>	<b>7</b>
<b>4</b>	<b>Validation .....</b>	<b>8</b>
<b>5</b>	<b>Conclusions .....</b>	<b>10</b>

## List of Tables

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

# 1 Introduction

## 1.1 Purpose

This trade-off will analyse the feasibility of using an alternative approach to the representation language, based on instances of a common XML schema, instead of representation of the binary data with a single XML schema currently used by SAFE.

This analysis is needed as results of the Software Requirements Review (SRR) process outcomes where new requirements for the SAFE core specification were identified. Some of these new requirements imply changes in the SAFE format design and requires some previous analysis to assess the correct implementation.

In this case, it was identified significant constraints for the representation language used by SAFE as it shall be based on XML and it shall use a clean implementation in accordance with XML state-of-the-art usage:

- Representation language shall be defined as XML schema
- Product representation is an XML schema instance

## 1.2 Scope

Before considering any decision on the representation language to be used by SAFE, it is expected to analyse if a schema validation for the representation language is really needed.

The analysis result presented in this document will be used as starting point for design decisions before updating the SAFE Core Specification documents. These core documents will be reviewed during the PDR-C afterwards.

The main objective of the PDR-C is to validate the correct implementation of the SAFE core Specification according to the new SAFE requirements included in the Software System Specification document (SSS).

## 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.eportal.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

**Table 1-: Applicable Documents**

## 1.5 Reference Documents

Acronym	Title	Reference	Issue
[SSS]	Software System Specification	SAFE-GMV-SSS-001	1.0
[OGF-DFDL]	OGF-DFDL Standard	<a href="http://www.ogf.org/dfdl/">http://www.ogf.org/dfdl/</a>	
[XML_W3C]	XML Schema	<a href="http://www.w3.org/XML/Schema">http://www.w3.org/XML/Schema</a>	
[RELAX NG]	Information technology – Document Schema Definition Language (DSDL) – Part 2: Regular-grammar-based validation – RELAX NG	ISO/IEC 19757-2:2003(E)	
[Schematron]	Schematron	<a href="http://www.schematron.com">http://www.schematron.com</a>	

**Table 1-: Reference Documents**

## 1.6 Acronyms and Abbreviations

Acronym	Meaning
DFDL	Data Format Definition Language
DOM	Document Object Model
DTD	Document Type Definition
EO	Earth Observation
LTDP	Long Term Data Preservation
PDGS	Payload Data Ground System
PDR-C	Preliminary Design Review - Core
SAFE	Satellite Archive Format for Europe
SAX	Simple API for XML
SDF	Structured Data File
SGML	Standard Generalized Markup Language
SRR	Software Requirements Review
SSS	Software System Specification
W3C	World Wide Web Consortium
XML	eXtensible Markup Language
XSD	XML Schema Definition

**Table 1-: Acronyms**

## 2 XML schema

An XML schema describes a XML document, usually expressed in terms of constraints on the structure and content of documents of that type (beyond the basic syntactical constraints imposed by XML itself). These constraints are generally expressed using some combination of grammatical rules and more specialized rules such as uniqueness and referential integrity constraints.

XML schemas can be specified using different languages called "XML Schema Definition Languages" (XSDL) like for example:

- DTD: the first formalized standard (inherited from SGML), but its use is not widespread extended.
- W3C XML Schema: (Usually known as "XML Schema" with capital "S") is supported by W3C and is currently the de facto standard for describing XML documents.
- RELAX NG: provides a simpler definition and validation framework than XML Schema, making it easier to use and implement
- Schematron: allows many kinds of structures to be represented which are inconvenient and difficult in grammar-based schema languages.

### 3 Representation information schema

One of the needs of SAFE is to be able to represent binary data, hence a representation information schema (unlike XML schema) to describe non-XML data is required.

To this end, these representation information schemas are XML schemas augmented with special "annotations" used to describe non-XML data.

According the W3C XML Schema specification, the annotation fields can be used to add comments (documentation tag) or to add stylesheets and instructions for other tools (appinfo tag) in a XML Schema (this is an established approach that is already being used today in commercial systems).

With this approach, the representation information language can be thought of as if it were XML data even though the representation is a much smaller and more efficient one. Additionally, this approach will make it possible to implement APIs similar to DOM and SAX and to simplify data exportation to XML documents.

Thus, a representation information schema is a well-formed XML file i.e. it is an XML instance document which conforms to all the well-formedness constrains described in the W3C XML Recommendation.

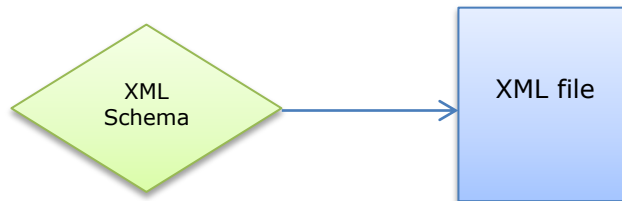
Representation information schemas can be specified using different languages like for example: SDF or DFDL

- SDF: allows describing text or binary data using dedicated markups directly inside a XML Schema so that it can be parsed as a structured hierarchy of nodes.
- DFDL: allows any text or binary data to be read from its native format by leveraging the W3C XML Schema Definition Language (XSDL).

## 4 Validation

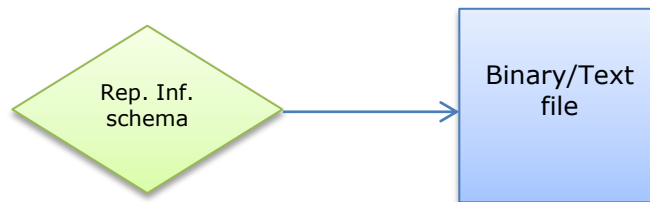
XML validation is the process of checking a XML document to identify if it conforms to a XML schema i.e. the document is well-formed (it follows the basic rules of XML) and valid (it follows the rules defined by a specific XML schema).

An XML document will be only considered valid, if it satisfies the requirements of the schema with which they have been associated.



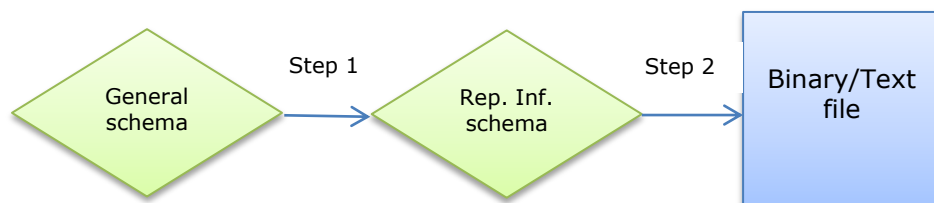
**Figure 4-: XML Schema representation**

Within SAFE products, the representation information schema does not apply to an XML document only, but also to a binary or a text file. Hence standard validation definition in these cases cannot be assumed:



**Figure 4-: Representation information schema**

The alternative proposed for the binary case during the SRR is depicted below and it is based on instances of a general schema, instead of representing the binary/Text data with a single schema:



**Figure 4-: Alternative for representation information schema**

An XML standard validation (Step 1) for the representation information schemas would only allow identifying syntactical correctness of those schemas, but the validation for the whole SAFE product (which implies Step 1 and Step 2) should include also a validation of the binary file (not only the representation information schema).

The binary files stored in a SAFE product will differ significantly one from the other therefore a "pure" XML Schema (General Schema in Figure 4-) should require a very general high level schema (at the same level of a grammar



for an XSDL) providing very little added-value to the final validation of the whole product.

To validate the binary/text file (Step 2) a semantic validation would be still needed to assure that it is possible to extract the correct data (blocks, order, fields, ...) from the representation information schema.

## 5 Conclusions

As the binary/text file within a SAFE product to be represented is not an XML file, there is no need to assume a XML standard validation mechanism to be compliant with W3C.

However, representation information schemas used in SAFE can be specified using existing languages like DFDL or SDF as they are compliant with W3C. Both languages make use of the annotation field. According to the W3C XML Schema specification, this field can be used to report comments (documentation tag) but also instructions for application (appinfo tag).

These schemas are not only used for file structure representation, but also to make it easy to convert information contained in a file (binary or text) into a corresponding XML document (containing the representation of the input data file) by making possible to implement APIs similar to DOM and SAX.

A syntactic validation can be implemented with specific existing tools for the representation schemas described for example in DFDL or SDF but it can be also developed to be included as part of the SAFE toolset.

According to the information presented in this trade-off, there is no need to develop from scratch a new language to specify representation information schemas because it won't provide relevant added value and, on the contrary, would have less expressive power respect to the languages already available.

## Change Record

Issue	Revision	Date	Change Status	Origin
1	Pr1	07/05/2012	First version	GMV

**< End of Document >**