

## **Memorandum**

---

**To** :

**From** : MED/C/SZ

**Copy** :

**Ref.** : EUM/OPS/DOC/12/3875, v1

---

**Date** : 12 November 2012

---

**Subject** : **Comments on PDGS-SAFE-GMV-TN-12-0171\_SimplDesign\_v1.0.pdf**

#	Category	Location	Comment	Proposed resolution
1.	major	5, p14 and general	<p>The conclusion reached in this document is not to externalize the representation information. (“it is concluded that the externalisation of the representation information is not recommended for SAFE”)</p> <p>This is clearly in contrast to what has been discussed and agreed previously in the other meetings (SRR, PDR-C).</p> <p>If that is the case, then the remainder of the document could have been cut much shorter, and specifically the sections about the registry, etc., are not understood in this context. Specifically, all the section talking about the simplification, complete section 4, etc.</p>	Discuss with all stakeholders and reviewers from the previous events (SRR, PDR-C), if externalisation shall be pursued or not.

**Comments on PDGS-SAFE-GMV-TN-12-0171\_SimplDesign\_v1.0.pdf**

2.	major	general	<p>This document introduces the need for a registry to map logical and physical paths when externalizing the representation information, and it proposes designs for such a registry on a high level.</p> <p>It is not understood why this document deals with this topic. The trade-off shall be about “simplification”. The registry as presented seems to be in a different context altogether.</p> <p>Moreover, several concepts of the registry are not understood and the presented approaches are partly wrong or not analyzed properly.</p>	Remove content and design for registry from this trade-off.
3.	normal	1.1, p5	<p>“The purpose of this trade-off is to simplify the SAFE structure reducing to a <b>minimum the number of files</b> comprising the SAFE Package”.</p> <p>IMO this is not the purpose of the trade-off.</p> <p>See as well follow on “The aim of this analysis is to evaluate if it is possible to reduce the number of SAFE Packages and files comprising the SAFE Packages[...]”.</p> <p>IMO the latter was what was requested, minimize the number of SAFE packages for the representation information, only.</p> <p>This document then, however, tries to address the question if representation information shall be externalised or not.</p>	Re-consider scope of trade-off
4.	major	general	<p>As another comment on the topic of SAFE packages for Representation Information: It is to be analyzed if Representation Information needs to be persisted in the archive <u>at all as SAFE packages</u>.</p> <p>SAFE packages are meant to keep all information together which are needed for interpretation of the content, that is, not only the data files, but as well metadata (files) and auxiliary data (e.g. indexes).</p> <p>If now Representation Information as one vital component is handles as SAFE packages as well, it is questioned if this makes sense. Why</p>	Discuss and possibly re-consider approach of storing externalized Representation Information as SAFE packages.

			would one need a manifest for this? Why would one need the Representation Information for Representation Information?	
5.	normal	2.2.1, p15	<p>“Furthermore, each SAFE Package included in the Representation Information must have a Manifest so that the number of Manifest files will be very high, as many as XML Schema files.”</p> <p>Surely, each SAFE package must have a manifest file. It is not understood though why this is mentioned here. As well, in comparison to the number of SAFE packages for the EO products, the number of SAFE packages for representation information files seems to be small (negligible).</p> <p>Reading the mentioned sentence it appears as if this would be a disadvantage, but this cannot be seen.</p>	Consider removing this sentence.
6.	normal	2.3, p16	<p>“The Simplification trade-off takes as base the SAFE 2.0 Draft. This new trade-off will try to add some improvements proposed during the PDR-C.”</p> <p>Looking at the conclusion reached (don’t externalize) this seems to be not really met.</p>	Comment, consider re-phrasing.
7.	normal	2.3, p16	<p>“On the other hand, after the development of the Externalisation trade-off many questions were raised about how the information from different SAFE Packages relate to each other. Thus, a detailed study on different approaches will be developed here”</p> <p>I must admit I don’t find those different approaches in this document.</p>	Align document content.
8.	normal	3, p17	<p>“The goal of the Simplification trade-off is to decrease the number of XML Schemas”</p> <p>I’m not sure this is the goal.</p>	Re-consider scope of trade-off.
9.	normal	3, p17	<p>“This document provides: new strategies for versioning, references between components of different packages and simplify the process for the product specialisations”</p>	Re-consider scope of trade-off.

**Comments on PDGS-SAFE-GMV-TN-12-0171\_SimplDesign\_v1.0.pdf**

			That's all nice to know but wasn't the goal of the trade-off, IMO. Moreover, I don't really see the mentioned topics properly (in-depth) addressed.	
10.	major	3.1.1.1, p17	<p>"The main structure of SAFE Packages seen in version SAFE 2.0 Draft is still valid but this one adds a new package to treat the EO metadata Collections and a new Registry Package to treat the references."</p> <p>If the main structure is still valid, then I don't see any simplification.</p> <p>Additionally, it is jumped to a conclusion here that a Registry Package is needed. I disagree with this premature conclusion.</p> <p>Again, IMO, the scope of the trade-off was to simplify the number of Representation Information (Packages) needed.</p> <p>The whole aspect of Registry could have been left out of this document and should ideally go to a separate document, if at all.</p>	Consider comment.
11.	major	3.1.1.1, p19	<p>"The registry could be implemented as a XML file and stored in a SAFE Registry Package, where each one should be based on a XML Schema file to comply with OASIS. Therefore, a new SAFE Registry Schema Base Package should be defined containing the representation information (XSD file). In addition, it is advised to have a SAFE Registry by Mission Level (0, 1 or 2) because the size of the file could grow exponentially."</p> <p>Several aspects are not understood here:</p> <ol style="list-style-type: none"> <li>1. Why we need the registry in a SAFE package? The registry is IMO not part of SAFE.</li> <li>2. The mentioned exponential growth cannot be seen. However, if done via XML file, it is supported that the information in the registry is separated by some level. To manage this via XML file, though, is not favoured. It doesn't bring any advantage only disadvantages, IMO.</li> </ol>	Consider comment.
12.	major	3.1.1.1, p19	"For instance, in the database case, this would mean that the registry in	Consider comment.

**Comments on PDGS-SAFE-GMV-TN-12-0171\_SimplDesign\_v1.0.pdf**

			<p>this case will be a centralized database (that has to be preserved) with all the mappings between logical and physical paths stored.”</p> <p>The mentioned mapping should be part (black-box) of the archive system used and is IMO not a part of the SAFE standard; the “registry” should be disregarded, thus.</p>	
13.	normal	3.1.1.2, p19	<p>“Grouping all DFDL XML Schemas of the Binary files belonging to the same EO Product into one SAFE Representation Information Package.”</p> <p>Likely “EO Product Type” is meant instead of “EO Product”. Otherwise it wouldn’t make sense.</p>	Consider comment.
14.	normal	3.1.1.2, p19/20	<p>“Grouping all OGC EOP O&amp;M XML Schema Base Types (eop.xsd, atm.xsd, etc.) into one SAFE Representation Information Package”, etc.</p> <p>This seems to be the solution for the issue discussed at PDR-C, i.e. too many representation information packages, that is: to group the representation information files into meaningful packages.</p> <p>If that is all basically what leads already to a significant simplification, the rest of the document is not understood.</p>	Consider comment.
15.	major	3.1.2, p25	<p>The document takes it as a given that there will be a registry, discussing pros and cons under that assumption.</p> <p>Again this is IMO a detail of the employed archive system and shall be treated as a black-box, at least from the perspective of the SAFE standard.</p>	Consider comment.
16.	major	3.2	<p>This section discusses IMO aspects completely out of scope of this trade-off.</p>	Consider comment.
17.	normal	3.2.2	<p>The representation storage node (REPSN) is just one possible solution for the storage of a registry.</p> <p>Again, why a specific design is discussed here if the trade-off shall</p>	Consider comment.

**Comments on PDGS-SAFE-GMV-TN-12-0171\_SimplDesign\_v1.0.pdf**

			discuss a different aspect in the first view?	
18.	normal	3.2.3	<p>This whole section is not understood.</p> <p>Does it make sense to assume that auxiliary data is stored in non-SAFE formats? If yes, then IMO one of the aspects why to use SAFE is lost completely.</p> <p>Why now as well a provenance registry is needed? Are the authors clear about the fact that when millions of files are to be connected via such a mechanism the XML approach will be not workable anymore? And again, it should be the task of the archive to provide the linkage, not a task of SAFE. Leave the design to the archive designers.</p> <p>It seems as well that the particular aspect of the governance will be discussed in just another trade-off – why then to mention it here?</p>	Consider comment.
19.	major	3.3, p30	<p>“The reference between an XML metadata component and an XML Schema component is defined in the “schemaLocation” attribute of the XML file”</p> <p>This is simply wrong! See W3C definition:</p> <p>“In an instance document, the attribute xsi:schemaLocation provides hints from the author to a processor regarding the location of schema documents. The author warrants that these schema documents are relevant to checking the validity of the document content, on a namespace by namespace basis[...]The first member of each pair is a namespace name, and the second member of the pair is a hint describing where to find an appropriate schema document for that namespace. The presence of these hints does not require the processor to obtain or use the cited schema documents, and the processor is free to use other schemas obtained by any suitable means, or to use no schema at all.”</p> <p>It is more true to say that the reference between XML and its associated structure definition is via (versionized) namespaces.</p>	Consider comment.

**Comments on PDGS-SAFE-GMV-TN-12-0171\_SimplDesign\_v1.0.pdf**

20.	major	3.3	SAFE is about the AIP not the DIP, so the whole discussion presented here about DIP is not necessary and confusing.	Consider comment.
21.	major	3.3	The purpose of this section is not understood. Shall it propose how to model different versions of each representation information? How to store them? It appears as if the conclusion reached is that we need just 2 representation information packages for the complete SAFE suite, one containing all versions of all DFDL files and one containing all versions of all EOP schema files.	Consider comment.
22.	major	3.3.1	<p>“Scenario 1: The SAFE EO Product Package is related to different versions of SAFE Representation Information Package (e.g. 1.0, 2.0, 3.0, etc.)”</p> <p>This is simply a non-existent scenario. One EO Product Package can belong to one and only one version of Representation information.</p> <p>Additionally, how this would be modelled in the manifest of the mentioned EO Product Package?</p>	Consider comment.
23.	normal	3.4	What is this section for, what is its purpose?	Please comment.
24.	normal	4.1, p36	<p>“1. To avoid breaking namespace-aware applications with each new version of an XML Schema, the same namespace should be used for all versions”.</p> <p>Disagreed, it is common practice to include the version in the namespace to be able to clearly distinguish (break the flow) different versions.</p>	Consider comment.
25.	normal	4.1, p36	<p>“2. Each new version of a XML schema should be located in a different directory in the SAFE Representation Information Package”</p> <p>Disagreed, it’s up to the archive to design how and where data is stored, this is of no concern for SAFE.</p>	Consider comment.
26.	normal	4.1, p36	Recommendation #3, see previous note on schemaLocation.	Consider comment.
27.	normal	4.1, p36	“4. It is recommended to store all the SAFE Packages for a specific	Consider comment.

**Comments on PDGS-SAFE-GMV-TN-12-0171\_SimplDesign\_v1.0.pdf**

			mission and level (0, 1 or 2) at the same level directory in the archive” Disagreed, it’s up to the archive to design how and where data is stored, this is of no concern for SAFE.	
28.	normal	4.1, p36	Recommendation #5, this is just one way of ensuring consistency – up to the implementation.	Consider comment.
29.	normal	4.1, p36	Recommendation #6, this is not a recommendation but just a statement of fact / repetition of well known information.	Consider comment.
30.	normal	4.1, p37	“9. It is recommended as well to store in the registry the location and name of the last applicable representation schema version for each product type.” It is not understood why this is needed nor recommended.	Consider comment.
31.	normal	4.1, p37	Recommendation #10, this seems to be banal.	Consider comment.
32.	normal	6.1, p42	The relationship mentioned between logical path and physical path as “0..*” is not understood. “it can be zero if the component addressed has not been preserved in SAFE format” → then it is not SAFE compliant, or? “It can be more than one when it refers to a whole set of files of the same type” → how then one can know which one to use?	Consider comment.
33.	normal	6.1.2	The presented DB design models not a 1 to 0..* relationship but vice versa.	Consider comment.