**Caroline Arms notes on OPC WD3.5**

2018-09-09

**Page vii**

**Section on changes.** I assume this will be modified based after the separate document on the rationale for changes has been finalized.

The first bullet

“Removed the allowance for media type to be an empty string, as this conflicts with the definition of media type in RFC 2046 and the existing regular expression defined in Annex C.”

has at least two problems.

(1) The regular expression is no longer in Annex C, since schemas have been removed from this document.

(2) The use of “allowance” is awkward English. The word has several meanings, but is not typically used as a synonym for “permission.”

Is it the case that the previous version had an internal inconsistency.? It included “The value of the content type is permitted to be the empty string.” If the regular expression in the schema did not permit the empty string, then that inconsistency was a significant problem and RFC 2046 is of minor relevance. Perhaps there are other changes that are fixes for internal inconsistencies that could be grouped together.

**Page viii**

Need to mention Annex H and possibly the Bibliography.

Need to review Annex G, which (a) was not fixed for the 2012 edition (mentions 2011 edition) and (b) has links to Annex C that relate to a schema. Those links may need adjustment.

**Page 1 Scope**

Change “one or more interrelated byte stream (part)” to

 one or more interrelated byte streams (parts)

**Normative References**

**Page 2**

1.

I checked for further activity wrt Dublin Core Terms Namespace (ISO/NP 15836-2). I could find no public activity beyond formation of working group ISO/TC 46/SC4/WG16

2. Does the parenthesized note for the XML reference need to be updated?

**Page 3**

Table numbers in note for ZIP reference need to be changed to B-3 and B-4.

**Terms & Definitions**

Since I have not been party to the most recent discussions on the re-organization of this clause, I am relying on others to check the substance of this clause. Comments below are primarily editorial.

**Page 6**

Remove final period in 3.3.4

logical item is defined as “a non-interleaved part, the non-interleaved Media Types stream, a piece of an interleaved part, or a piece of the interleaved Media Types stream”

Personally, I find the use of “the” with the two Media Types stream phrases here awkward. We must be considering “logical item” in a general context, not within a particular package, since in a particular package there cannot be both a non-interleaved and an interleaved Media Types Stream. Hence, I would replace by the “a” in both spots.

**Page 7 – package implementer**

I would change “software that implements the physical input-output operations to a package according to the requirements and recommendations of this document”

to

“software that implements physical input-output operations on a package according to the requirements and recommendations of this document”

**Notational Conventions, page 8**

Is there an ISO convention for references to items in a bibliography? Are we following it? An example is on page 13, end of first para in 8.2.2.3.

**General Description**

**Page 9**

Add mention of Annex H and Bibliography.

**Overview**

**Page 11**

Should the final paragraph (about conformance) mention the schemas explicitly, since they are no longer part of “the text of this document”.

**Abstract Package Model**

**8.1 Page 12**

I would drop “two parts:” because a full package would actually have additional parts, e.g., for relationships.

**8.2.2.2 Pages 12/13**

The relationship of the definition for I18N segments (near top of page 13) to the preceding text is confusing, particularly because it isn’t a proper sentence. I propose moving that text before the 3 bullets that use it, either inserting the definition after “constraints” or making it a separate sentence or paragraph with a proper beginning.

**8.2.2.3 Page 14**

first para on page, beginning “This subclause further”

Change “does” to “do”

**8.5.1 Page 22**

I see a minor grammatical problem in the second sentence of the first para. A semicolon is used inappropriately. We haven’t introduced this problem – it is inherited from the 2012 version (and possibly earlier). I would probably replace the semicolon by a comma.

**8.5.1. Page 23**

There are still comments in the current draft. Here’s another attempt at explaining why I find the current text confusing. I think Murata-san is possibly trying to make two independent points in the first sentence without making it clear that they are independent.

Relationships have (and must have) Id attributes and such identifiers must be unique within a Relationships part.

One reason is that there can be more than one relationship from a source to the same target. An Id is a simple mechanism to distinguish them. See example in 8.5.4.5

I think the second point Murata-san wanted to make is that the Id value for a relationship might also be used within the source part associated with its parent Relationships part to indicate an exact location at which the relationship is relevant.

I finally realized that this use of a relationship Id might be something that Microsoft uses and checked Part 1. I find “A document part can have a relationship to a hyperlink only if that hyperlink's Relationship element’s Id attribute value is referenced explicitly by the document part’s XML” in clause 9.2. Since 8.5.1 is an informative clause, perhaps this or some other example can be included in OPC 8.5.1. I assume such usage would always be implementation-defined.

Once I am sure I understand Murata-san’s intent completely, I can probably suggest some clearer wording.

[Aside: I see that OPC clause 12 mentions relationship Ids wrt digital signatures , but haven’t dug in to that.]

**8.5.3.1 Page 25**

In para at top of page, there is a reference to Annex C that provides little value given that the schemas are no longer there. opc-relationships.xsd is really only “listed” in Annex C, not “described.”

**8.5.3.3 Pages 25/26**

Many references to §C.x for individual attributes and elements seem awkward given that the schemas are no longer actually in Annex C. The links that seem intended to go to a location for the schemas aren’t working for me.

**8.5.4 Pages 26/27**

I would drop the commas after “Consider a package”in 8.5.4.1 and 8.5.4.2. [Compare with examples in 8.5.2.2.]

**8.5.4.3 Page 28**

This subclause starts “The figure below” but there is no figure.

**Physical Package Model**

**9.1 Page 31**

Reference to Annex E should be to Annex F.

I would drop “the” before “Physical package model design considerations” in second sentence of first para.

I would change “An example physical package is shown” to “An example physical package is described”

**9.2.3.3 Page 34**

I would drop “to be” in sentence beginning “The Types element”

**9.2.3.4 Page 34**

I would change the semicolon in first sentence to a period and capitalize the following letter T.

**9.2.4 Interleaving, Page 36**

I found the first two paragraphs awkward. In trying to reword I noticed the change in order of clauses 9.2.4 and 9.2.5. [Note: The switch in order for 9.2.4 and 9.2.5 happened between drafts I have dated 2018-04-03 and 2018-04-17]

Looking at differences between OPC WD3.5 and the 2012 published version, there are some issues that seem substantive to me.

OPC 2012 indicates that pieces ”should” be stored in natural order for optimal efficiency.

OPC WD3.5 states that pieces “shall” be stored in natural order.

OPC 2012 states that pieces are named using a “unique mapping from the part name” and implies that it is this mapping that allows pieces to be joined together in the original order.

OPC WD3.5 does not mention the naming for pieces in 9.2.4 and might be read to imply there that it is the ordering of the pieces in the package that permits the joining to get the original stream. [See “Pieces in physical packages shall occur in their natural order for optimal efficiency. Pieces can later be joined together, forming the original stream.”]

Are these changes intentional?

I certainly think that the naming convention for interleaved pieces needs to be mentioned in 9.2.4 with a pointer to 9.2.5.

Suggested replacement for the first two paragraphs of 9.2.4:

==>

When mapping an abstract package to a physical package, the data stream of a part or the Media Types stream may be broken into pieces, which can be interleaved with pieces of other parts or with whole parts. Pieces can later be joined together, forming the original stream, based on the use of suffixes to derive piece names, as specified in 9.2.5.2. Pieces in physical packages shall occur in their natural order for optimal efficiency.

This document does not require a physical package model to support interleaving.

==>

As I have pointed out before, the diagrams in this subclause are not always legible because the text and striped background interfere with each other. I see this problem on the screen as well as when printed.

**9.2.5 Page 37**

First para, first sentence. Doesn’t adequately motivate the need for logical items.

First para, final sentence: Not immediately clear what “it” refers to.

Rewording suggestion

==>

A mapping from an abstract package to a physical package shall use logical items as intermediate objects in order to permit interleaving (see 9.2.4). If a part or the Media Types stream is not interleaved, it shall be a logical item. Otherwise, each piece constructed from it shall be a logical item.

==>

9.2.5.2 seems to be missing the equivalent of

LogicalItemName = PrefixName [SuffixName]

It wasn’t immediately clear to me that there is no separator between prefix and suffix. I realize that “The prefix of a logical item name is the result of removing a suffix, if any, from the logical item name” implies that, but it needed re-reading several times for me to understand why such a convoluted explanation was needed. Examples would help.

Since the diagram of part names and logical item names has been dropped, I would at least like to see a textual example for a complete sequence of pieces.

**9.3.3 and 9.3.4 Page 39**

9.3.3 Says “ZIP item names are Unicode strings. … The support of non-ASCII ZIP item names is not required.”

 AND

9.3.4 says Percent-encode every non-ASCII character

This struck me as slightly odd. If we are saying that non-ASCII ZIP item names should not be used in OPC ZIP-based packages, then clearly support for non-ASCII ZIP item names is not required. I think some rewording may be warranted for clarifications, but I‘m not certain of the substance.

**9.3.4 and 9.3.5 Pages 39/40**

I find “Mapping …shall perform…” awkward.

9.3.5 uses “part names” in one place when it should probably use “logical item names”.

9.3.5 has “for each ZIP item” that has no equivalent in 9.3.4.

I suggest:

 The following steps shall be performed, in order, to map each logical item name to a ZIP item name:

 and

 The following steps shall be performed, in order, to map each ZIP item name to a logical item name:

**9.3.6 Page 40**

I don’t think we should refer to “ZIP Appnote.txt” but probably to “the Zip File Format Specification (as used in Normative references) or to version 6.2.0 explicitly.

**9.3.7 Page 40/41**

Because a sentence has been removed, the note at the top of page 41 referring to “this requirement” is unclear. I would introduce an equivalent to the dropped requirement, yielding:

 “Logical item names in a ZIP file for the Media Types stream, or its pieces if interleaved, shall not be mapped to a part name. [Note: Bracket characters "[" and "]" were chosen for the Media Types stream name specifically because these characters violate the part naming grammar, thus reinforcing this requirement. end note]”

The note seems important given all the complexities involved in the Media Types stream not being considered a part. I don’t think the note can go elsewhere, because “[Content\_Types].xml” as the name of the Media Types stream only appears here (and in the new informative Annex H).

**10.2 Page 43**

First sentence as cumbersome because of all the occurrences of “part.” Suggested rewording for first two paragraphs:

==>

A package may contain at most one Core Properties part.

The content of the Core Properties part shall be an XML document that satisfies the requirements specified in 8.2.5. Its media type shall be the Core Properties part media type, as defined in Annex E.

==>

Last sentence

 “The namespaces for the properties in this table in the Open Packaging Conventions domain are defined in Annex E.”

refers to “this table” but there is no table in this sub-clause.

The sentence was moved from 10.1 to 10.2 when 10.1 was made informative. The table referred to is therefore Table 10-1, which is now only informative.

The sentence in OPC 2012 is “The namespace for the properties in this table in the Open Packaging Conventions domain are defined in Annex F.” Notice that namespace is singular not plural. In OPC 2012, the word “domain” was used as heading for the column now headed “Specification.” OPC 2012 Annex F just lists the namespace that applies to the OPC-specific properties (<http://schemas.openxmlformats.org/package/2006/metadata/core-properties>). I don’t see any documentation for the namespaces for Dublin Core elements in OPC 2012 Annex F.

Did we change the sentence to have namespaces in the plural because we expected to document the two DC namespaces in Annex E? Currently we have their namespaces in notes in 10.3.3 and 10.3.4

I assume this needs further discussion as to intent and don’t propose a change.

**10.3.5 Page 45**

Has lots of references to Annex D which should be to Annex C and need to be part of the general consideration needed of links to schemas.

**12 Digital Signatures**

My suggestions here are purely editorial.

**12.2 Page 50**

First sentence needs “s” at end of “represent”

This clause uses “OPC” as the subject of several sentences. We haven’t done that elsewhere. I’m not complaining, just noting.

**12.4.4 Page 52**

I would drop the first “or” in the first sentence.

**12.5.1 General Page 52**

I found the first paragraph slightly awkward to parse.

Suggested rewording (if my interpretation is correct):

The following subclauses cover OPC-specific restrictions and extensions to “XML-Signature Syntax and Processing”. Subclauses follow for elements defined for OPC-specific use or for which OPC introduces restrictions. Elements defined in “XML-Signature Syntax and Processing” (such as X509Certificate) for which no subclause is provided below are allowed in OPC packages without restriction.

**12.7 Page 57**

First para, last sentence. “refer” should be “refers”

Should “pds” be in different style?

Would a pointer to annex E for the namespace make sense?

**Annex C**

Links to schema documentation are not working for me.

I think all links to Annex C (including references of the form §C.x) need review as to whether replacement links or wording changes are warranted. See, for example, the attribute tables in 9.2.3.2.

**Annex G**

Needs review for consistency within entire document. Refers to 2011 published version (probably in error).

**Annex H**

**Page 83**

I would like to see “example” either in **H.1** or in title of **H.2**. Alternatively, ifthis is the only content for Annex H, the entire Annex could be titled “Example Package.”

Diagram needs “document.xml” after “part Relationships part associated with”

**Page 84**

**H.3**

Replace “Except” in second para. with “With the exception of”