[Draft] Minutes of the Teleconference Meeting of

ISO/IEC JTC 1/SC 34/WG4, 2014-01-07

Rex Jaeschke (rex@RexJaeschke.com)

2014-01-08

1. Opening remarks

The meeting started at 21:05 GMT. The convener, Murata-san, welcomed everyone to the 57th teleconference of WG4.

2. Roll call of delegates

The following members were present during part or all of the meeting:

Name	Affiliation	Employer/Sponsor
Makoto Murata	WG4 Convener, JP	International University of Japan
Jirka Kosek	CZ, XML Guild liaison	Consultant
Rex Jaeschke	Ecma, Project Editor	Consultant
Jim Thatcher	Ecma, US	Microsoft
Caroline Arms	Ecma	Library of Congress
John Haug	Ecma, US	Microsoft
Chris Rae	Ecma	Microsoft
Francis Cave	GB	Francis Cave Digital Publishing

Present were 8 people representing 4 NBs, and 2 liaison.

3. Adoption of the agenda

The agenda (SC 34 N 1990) was adopted as published (with the closing time corrected).

4. Administration

Approval of Previous Meeting Minutes [WG4 N 0275]

The draft minutes were approved, as circulated.

Outstanding Action Items

1. Rex will apply the edits mentioned in these minutes creating WD2.0 of the MCE Part, which he'll circulate to those members of WG4 who were on this call, for a short review. **Done**

2. Rex will submit WD2.0 of the MCE Part to Kimura-san for a 60-day ballot. [Ballot runs from 2013-12-12 to 2014-02-12] **Done**

3. Murata-san will announce to the WG4 that we'll start work on the Part 2 revision on the next teleconference. **Done** in mail on 2013-12-05

Report from the WG4 Secretariat

The following NBs and liaisons have registered delegates to WG4: BR, CA, CH, CI, CN, CZ, DE, DK, Ecma, FI, FR, GB, IN, IT, JP, KR, NL, NO, OASIS, PL, US, W3C, XML Guild, and ZA. All requests for additions, deletions, and changes to the delegate list should be sent to the WG4 Secretariat (rex@RexJaeschke.com).

The WG4 email list is e-SC34-WG4@ecma-international.org. The document repository is at http://lucia.itscj.ipsj.or.jp/itscj/servlets/ScmDoc10?Com_Id=w4.

5. Defect Reports

The public, online DR log is at http://www.29500sc34comments.org/. Access individual DRs via the hyperlinks contained within the spreadsheet's left-most column.

DR 13-0011 "PML: Incorrect description for sld element"

After a short discussion, we agreed to move this to Last Call.

DR 13-0012 "SML: Generalization of Currency in Function Descriptions"

After a short discussion, we agreed to move this to Last Call.

DR 13-0013 "WML: omissions and inconsistencies in the specification of attributes"

DR 13-0014 "PML: omissions and inconsistencies in the specification of attributes"

There was a lengthy discussion. How shall we divide the issues into DRs? By ML? By functionality? Francis, John H., Chris, and Murata-san agreed to look at these and related issues, and to come up a plan of attack.

DR 13-0015 "Copy-paste errors"

The resolution looks good, but we agreed to defer the final adoption until the Berlin F2F.

6. Revising Part 2 (Open Packaging Conventions)

For the most part, we walked through the OPC-related DRs reminding ourselves of the main issues. In particular, there was a long discussion about non-ASCII characters in part names. While allowing non-ASCII characters seemed okay, perhaps a recommendation saying something like, "For maximum interoperability, only ASCII characters should be used."

On several occasions, we restated that, "OPC was the most widely implemented Part of 29500, so we need to be very careful about making changes, so that we don't break existing implementations".

Regarding DR 09-0293, "OPC: pack URI scheme", the internet draft has expired. We could document that scheme in OPC and ask IANA to reference our spec, or we could create a new RFC.

Reference was made to document WG4 N 0207, "Improving Part 2 in reply to DRs", from September 2011.

For convenience, what follows are copies of recent OPC-related mails posted to the WG4 list:

"OPC part names and references" thread

2013-12-24 Murata-san:

I am trying to implement Section 2 of the Japanese proposal (http://kikaku.itscj.ipsj.or.jp/sc34/wg4/archive/sc34-wg4-2011-0207.html) for improving part names and references.

While doing so, I studied the conversion from part references (which are relative) to part names again. Here are some random thoughts.

1) Leading "/"

In Seattle, we learned from OPC experts that references beginning with "/" and those not beginning with it are resolved differently.

Proposal: Explicitly state differences between these two types of references. A reference to RFC 3986 is not good enough.

2) Resolution of relative URI references

Neither OPC part references (or Unicode strings as specified in Annex A) nor OPC part names contain schemes (e.g., http:). Should we nevertheless rely on resolution of relative URI references for the conversion from OPC part references to OPC part names? In other words, should we first create absolute URIs thus introducing schemes and then construct OPC part names by removing schemes?

Proposal: Stop relying on resolution of relative URI references.

Rather, introduce "base OPC part name", which is the OPC part name of the containing OPC part, and introduce a procedure for merging base OPC part names and OPC part references. This processing model does not have to touch schemes.

3) xml:base

Do we really have to allow xml:base (and other similar mechanisms) to change the interpretation of OPC part references? If such a mechanism specifies irrelevant URIs such as http://www.example.com, how should we interpret OPC part references?

Proposal: Stop using xml:base (and other similar mechanisms).

2013-12-28 Murata-san:

The more I think about OPC, the more confused I am.

I have thought that references to OPC parts ("Unicode string" in Annex A of OPC) can contain non-ASCII characters and that such non-ASCII characters are percent-encoded before referenced OPC parts are located. I have also thought that references to OPC parts are resolved relative to containing OPC parts when they do not begin with "/".

However, my experiment with .Net in F# appears to show I am mistaken. It reports errors if references to OPC parts contain non-ASCII characters. It also reports errors if references to OPC parts do not begin with "/".

I plan to manually edit OOXML documents and XPS documents and handle them by MS-Office and XPS viewers.

Here is my F# program.

```
open System.IO.Packaging
open System
let readOPC() =
    let package = Package.Open("f:test.opc", IO.FileMode.Open)
    let uri = new Uri(Uri.EscapeUriString "/f あ/f1", UriKind.Relative)
    let part =  package.GetPart(uri)
    let enum = part.GetRelationships().GetEnumerator()
    while (enum.MoveNext()) do
      let relship = enum.Current
      let targetURI = relship.TargetUri
      try
        let targetPart = package.GetPart(targetURI)
        let s = targetPart.GetStream()
        System.Console.WriteLine("Success: {0} {1}", targetURI, s.ReadByte()) with | :?
System.ArgumentException ->
System.Console.WriteLine("Error: {0}", targetURI)
    package.Close()
readOPC()
```

2014-01-02 John Haug:

For my part, it's going to take some time to ramp up mentally on the URI/IRI/LEIRI and relative/absolute/etc. nuances we've discussed previously before I can comment halfway intelligently on this topic.

Since it came to mind, here are other references people should recall.

1. E-mail from June 2011 and attachment (Murata-san's comments on Annex A conversion steps) -

http://mailman.vse.cz/pipermail/sc34wg4/2011-June/002305.html

2. WG 4 N 0207 from Sept 2011 - http://kikaku.itscj.ipsj.or.jp/sc34/wg4/archive/sc34-wg4-2011-0207.html

3. E-mail thread from May 2013 "DRs related to Part names" - last mail in the thread at

http://mailman.vse.cz/pipermail/sc34wg4/2013-May/002908.html

2014-01-06 Murata-san:

I did some more experiments using MS Office 2007 and .Net.

Here is my understanding.

- MS Office 2007 converts %HH to characters at least when %HH represents ASCII characters.
- MS Office 2007 resolves absolute-path references (which begins with "/") correctly.
- MS Office 2007 resolves relative-path references (which does not begin with "/") correctly.
- .Net (Package.GetPart) recognizes neither relative-path references nor %HH

I think that we should limit our concern to MS Office. The .Net implementation of OPC does not implement Annex A of Part 2 at all.

2014-01-06 Murata-san:

2014/1/3 John Haug <johnhaug@exchange.microsoft.com>:

> For my part, it's going to take some time to ramp up mentally on the URI/IRI/LEIRI and relative/absolute/etc. nuances we've discussed previously before I can comment halfway intelligently on this topic.

Let me give a very short overview.

- URIs do not allow non-ASCII. In other words, URIs are ASCII strings.
- IRIs allow non-ASCII characters.
- LEIRIs allow some more characters, which are arguably problematic but also arguably useful.
- Conversion from LEIRIs to IRIs and that from IRIs to URIs are done by %HH.
- Absolute-path references begin with "/", while relative-path references don't.
- Both absolute-path references and relative-path references are relative references (i.e., no "http://" or some other scheme).

2014-01-06 Murata-san:

Here are some further experiments.

Summary:

MS Word 2007 does not allow non-ASCII characters within part names even if they are percent-encoded. %HH in OPC part references are decoded as long as they represent ASCII characters.

Experiments:

First, I replaced "document.xml" in a WML document by "%E3%81%82.xml".

Specifically:

- Renamed the file "document.xml" under the directory "word" as "%E3%81%82.xml"

- Renamed the file "document.xml.rels" under the directory "word/_rels" as "%E3%81%82.xml.rels"

- Replaced "word/document.xml" in the file "_rels/.rels" by "word/%E3%81%82.xml"

- Replaced "/word/document.xml" in "[Content_Types].xml" by "/word/%E3%81%82.xml"

Then, MS Word 2007 cannot open the revised WML document.

Second, I used "a" instead of "%E3%81%82" in the above four changes.

Then, the document opened successfully.

Third, I replaced "a.xml" in the file "_rels/.rels" by "%61.xml". I also percent-encoded some other references (values of Relationship/@Target).

Then, the document opened successfully.

Fourth, just in case, I tried verbatim U+3042 (HIRAGANA LETTER A) rather than %E3%81%82. As expected, the document does not open.

My conclusions:

- Non-ASCII characters in part names are not allowed even if they are percent-encoded.
- %HH in values of Relationship/@Target are decoded as long as they represent ASCII characters.

2014-01-06 Murata-san:

I can think of four design choices. Which do you prefer?

Option 1: No non-ASCII characters. No %HH. No I18N.

Option 2 (Office 2007): No non-ASCII characters. %HH in OPC part references are allowed but they are required to represent ASCII characters. (Note: What's the point of allowing %HH only for representing ASCII?) No I18N.

Option 3: Verbatim non-ASCII characters everywhere. Note: it remains to decide when and where %HH should be allowed. I18N based on the latest version of ZIP.

Option 4 (close to current wording in 29500-2): Verbatim non-ASCII characters are allowed as part of OPC part names and OPC part references. But ZIP item names are ASCII. Conversion to OPC part names to ZIP item names invokes percent encoding. I18N without requiring the latest version of ZIP. Note: it remains to decide when and where %HH should be allowed.

2014-01-06 Toshiya-san:

Thank you very much for difficult experiments.

Considering the results, I think

The option 1 (pure ASCII, no %HH) is the best if the standard prioritizes for the document interchange between the existing (i.e. including some bugs) implementations.

The option 3 is the second best, because it would be the most expected situation.

2014-01-07 Chris Rae:

Am I right in understanding that option 1 effectively subsumes option 2? I.e. "hello%20there" is valid in both cases, but in the second case %20 represents a space and in the first case, it's just the literal characters?

2014-01-07 Chris Rae:

I have a feeling that some of the sticking points we discovered with regard to relative references were related to XPS, but I can't remember exactly what the details were. I'll do some investigation.

We'll have to tread somewhat carefully here, as OPC is the most widely implemented part of ISO/IEC 29500.

2014-01-07 Jim Thatcher:

It would help me consider which of the options proposed by Murata-san I would advocate if I had a good idea of what use cases we are trying to cover. Does anyone want to take a stab at describing use cases that our work on Part 2 is intended to enable?

2014-01-07 John Haug:

Yes, coming in the mail I'm sending shortly. I also needed to get a better understanding of the big picture before getting lost in the details, which may change based on a better understanding of the goals and past discussions.

OPC changes between ECMA 1st edition & 29500:2008

2014-01-07 John Haug:

I did a diff of the OPC parts in ECMA-376 1st ed and 29500:2008 and found these categories of changes:

- Addition of Part IRI syntax subsections in (what is now) clause 9
- Removal of diagrams and prose duplicating content from the xmldsig-core W3C standard
- Scattered minor substantive (clarifications, mostly) and editorial changes, also including the following larger changes:
    - Added XML and RELAX NG schema to the document that were previously electronic-only (unknown if any were changed)
    - Related changes to requirements tables in (what is now) Annex H

This confirms that the set of changes to add Part IRI (mostly within 9.1.1 Part Names but with related changes needed elsewhere) is the primary content needing improvement. It also shows that some xmldsig duplication was already removed, so the remaining content needs to be carefully reviewed to see if there is a good reason it remains.

Proposed overview/scope of work for OPC revision

2014-01-07 John Haug:

To reach consensus on scope and to better understand what we're intending to accomplish before we get lost in details, may I suggest the following as the list of topics we want to address in our changes to Part 2? All open DRs marked for Part 2 should be included below. Anything to add/remove/change before we agree on what we're going to do?

I'm going to try to organize a set of reference info, more thorough than the short list I sent out quickly on Thursday, to include numerous of old e-mail threads, documents, etc. If I am able to do so, I may put it on a new Wiki page on the WG 4 Assembla wiki (https://www.assembla.com/spaces/IS29500/wiki). I would also add the below, plus changes.

1. Combining IRI/URI

- 09-0280 – general DR about sections of OPC assuming non-ASCII is disallowed
- 09-0281 – terminology; all uses of "part" should specify what kind of part is meant
- 09-0283 – similar to 09-0280
- 09-0284 – terminology; need to re-associate the ABNF term names with the prose for part names
- 09-0285 – terminology; use of "part IRI" and "part URI"
- 09-0286 – need to update 9.2 (Part Addressing) to specify the format of a reference, based on the changes to be made to 9.1.1 (Part Names)
- 09-0288 – same as 09-0286, but for 9.3.2 (Relationship Markup)
- 09-0291 – terminology; specify what "Unicode string" means, based on the changes to be made to 9.1.1
- 09-0292 – which characters are allowed in a part name (e.g., whitespace, delimiters, special characters)
- 13-0002 – fix previous changes made to Annex H for introduction of non-ASCII support
- Add cautionary language noting that implementations might only support ASCII/URIs? Anything to note in Part 1?

2. Relative/absolute referencing

- 10-0015 – clarify "source part" for relationships
- Add cautionary language about detailed interpretations of relative vs. absolute references as defined in RFC 3986?

3. Pack URI

- 09-0293 – decide how, and whether, to handle IANA registration
  - (see also historic info, including issues with Pack URI, here:
    http://www.iana.org/assignments/uri-schemes/uri-schemes.xhtml,
    http://www.iana.org/assignments/uri-schemes/historic/pack)

4. Digital signatures

- 11-0029 – is content from xmldsig-core duplicated in 13.2.4 (Digital Signature Markup)?
- 11-0030 – update normative reference to xmldsig-core?
- 11-0031 – update reference to xmlsec schema?
- What, if anything, needs to be said about ensuring XAdES use is allowed?

5. Versioning

- 09-0168 – cannot distinguish between ECMA 1st ed OPC package and later versions (e.g., 29500 supports non-ASCII)

6. Editorial misc

- 12-0001 – unify capitalization of "relationship part"
- Remove reference to Part 3 from clause 8 (Overview)

OPC: Relevant documents

2013-01-08 Murata-san:

Here are references to relevant documents. Hope this helps.

1) URIs, IRIs, and LEIRIs.

http://www.w3.org/International/articles/idn-and-iri/

An Introduction to Multilingual Web Addresses http://www.w3.org/International/articles/idn-and-iri/

Uniform Resource Identifier (URI): Generic Syntax

http://tools.ietf.org/search/rfc3986

Internationalized Resource Identifiers (IRIs)

http://tools.ietf.org/search/rfc3987

LEIRI

http://www.w3.org/TR/leiri/

2) File names

Naming Files, Paths, and Namespaces

http://technet.microsoft.com/en-us/library/cc757759(v=ws.10).aspx

Wikipedia: Filename

http://en.wikipedia.org/wiki/Filename

3) Digital signature and encryption

[MS-OFFCRYPTO]

http://msdn.microsoft.com/en-us/library/cc313071(v=office.12).aspx

XML Signature Syntax and Processing Version 1.1 http://www.w3.org/TR/xmldsig-core1/

XML Signature Syntax and Processing (Second Edition) http://www.w3.org/TR/xmldsig-core/

XML Encryption Syntax and Processing Version 1.1 http://www.w3.org/TR/xmlenc-core1/

XML Encryption Syntax and Processing

http://www.w3.org/TR/xmlenc-core/

XML Security RELAX NG Schemas

http://www.w3.org/TR/xmlsec-rngschema/

7. Other Business

Thanking Host

We thanked Microsoft and Chris Rae for hosting the teleconference.

8. Future meetings

Face-to-Face Meetings:

- 2014-03-03/07, Berlin, DE (with other WGs)

    Monday morning: 09:00 WG1
    Monday afternoon: WG5
    Tuesday morning: WG4
    Tuesday afternoon: SC 34 re-organization ad hoc
    Wednesday: WG4 (continued)
    Thursday: WG4 (continued)

Friday morning: WG4 (continued)

- 2014-06-16/20, Prague, CZ (with other WGs)

    Monday morning: 10:00 WG1

    Monday afternoon: WG5

    Tuesday morning: WG4

    Tuesday afternoon: SC 34 re-organization ad hoc

    Wednesday: WG4 (continued)

    Thursday: WG4 (continued)

    Friday morning: WG4 (continued)

- 2014-09-22/26, Kyoto, JP (with other WGs, and Opening/Closing Plenaries)

Teleconferences:

- 2014-02-04 (Tue/Wed), 21:00 GMT (US/PT 13:00, GB 21:00, DE/DK/FR/CZ 22:00, JP 06:00 next day)

9. Adjournment

Adjourned by unanimous consent at 23:00.