Rationale for the OPC Revision

1. Part names, references, pack URIs, base URIs, resolution of relative references, and non-ASCII characters

The biggest motivation for revising OPC is to address issues around part names, references, pack URIs, base URIs, resolution of relative references, and non-ASCII characters. Numbers/Preview/QuickLook of Apple did have a bug (see this bug report for openpyxl) due to these issues.

1.1 Defect Reports

[09–0280](#)	OPC: Non–ASCII characters in Part Names	Clause 4 and Annex A disallows non–ASCII characters in part names, while 9.1.1.1 allows them.
[09–0283](#)	OPC: Inconsistencies between Clause 9.1 and Annex A	There are duplications in §9.1 and §A. Furthermore, the terminology in §9.1 and that in §A are slightly different. For example, "Part IRI" and "Part URI" in §9.1 are never used in §A.

09-0284	OPC: part-URI and part-IRI grammar productions	It is not clear where in the BNFs in RFC 3986, RFC 3987, or Appendix A the non-terminals part-URI and part-IRI occur.
09-0285	OPC: Use of Terms "Part URI" and "Part IRI"	The term "Part IRI" is very misleading. It sounds like a particular type of IRI, but it actually means those parts of IRIs which specify OPC part names. Likewise, the term "Part URI" is also misleading.
09-0286	OPC: The syntax of "references"	The syntax of "references" is never clearly stated. What is a reference? Is it a part of relative LEIRI, IRI, or URI references?
09-0291	OPC: Use of term "Unicode string"	It is not clear why the term "Unicode string" has to be introduced here. Moreover, its syntax is quite unclear.
09-0292	OPC: Space characters in part names	It is not clear whether the space character is allowed as part of OPC part name.
09-0293	OPC: pack URI scheme	Although the pack URI scheme has been registered as a provisional scheme at IANA, its definition appears in an Internet Draft rather than an RFC. The latest Internet Draft has expired in August 2009. Furthermore, the registration of the pack scheme at IANA has been changed from "provisional" to "historical" since the Internet Draft mistakenly allowed ":" as part of an authority.
10-0015	OPC: Relationship Markup	It is not clear how relative URIs (which are values of the Target attribute) are resolved.

1.2 History

[Harmonizing OPC with Web Addresses and ZIP](), SC34/WG4 N0148 (2010-09)

This personal contribution by Murata studies the use of non-ASCII characters in OOXML, IETF URI/IRI RFCs, and W3C LEIRI Note.

[Improving Part 2 in reply to DRs](), SC34/WG4 N0207 (2011-09)

This Japanese national body contribution studied relevant DRs including those listed above and asserted that a revision is needed for addressing them.

[Minutes of the Bellevue WG4 meeting]() (2013-06-17/20)

In the Bellevue meeting (2013-06-17/20), WG4 extensively discussed these DRs and even studied MS Office and .Net implementations. Chris Rae, John Haug, Jim Thatcher, and original OPC designers were involved in this discussion. The current OPC draft is based on consensus in this meeting.

1.3 Difficulties

Why is this topic so hard? There are several reasons:

- RFC 3986, RFC 3987, and WHATWG URL Spec are not as good as we hope.
- Relative references in non-Relationships parts and those in Relationships parts need different base URIs (when the target mode is not external).
- OOXML documents and XPS documents use different conventions for referencing parts.
- MS Office and .Net exhibit different behaviors.
- The first edition of ISO/IEC 29500-2 specifies behaviors different from any of those mentioned above.
- Non-ASCII characters were introduced after the DIS ballot in a hurry.

1.4 Solutions and Remaining Issues

WG4 has addressed most of the problems identified by the above DRs by thoroughly rewriting the clause for the Abstract Package Model. In particular, 1) a new subclause "Resolving Relative References" has been added; 2) part Relationship parts and package Relationship parts are distinguished; 3) base IRIs are clearly defined for part Relationship parts and package Relationship parts depending on the target mode; and, 4) Pack URIs are defined in Clause 8 rather than an annex.

However, DRs 09-0286 and 09-0291 have not been completely addressed, since WG4 does not know the behaviors of all implementations of OPC. DR 09-0293 requires further work at IANA after the publication of the revised OPC.

2. Addressing conformance issues (leftover from the BRM)

In the BRM, it was agreed that OPC conformance is purely syntactical. However, for the lack of time, many requirements in OPC were *not* rewritten as requirements on data, as implied by this sentence in the published 29500-2: "Conformance requirements written as requirements for package implementers (e.g., M1.1) are document conformance requirements".

2.1 Defect Reports

13–0002	OPC: Issues with Conformance Guidelines
14–0001	OPC: Annex H Cleanup

2.2 Solutions

Wherever possible, WG4 has rewritten requirements on programs as those on data.

The informative annex for summarizing guidelines for meeting conformance has been dropped.

The clause for conformance was moved from Clause 2 to Clause 6, since ISO/IEC editing directives require that Clause 2 be "Normative References".

3. Clarifications

3.1 Terminology

WG4 removed terms that are not used by any normative clauses and then reorganized the remaining terms by subclauses.

3.2 Physical Package Model

The clause for physical packages has been renamed as "Clause 9 Physical Package Model". Interleaving is introduced before logical item names. Percent-encoding and un-percent encoding of non-ASCII characters are explicitly introduced in Subclause 9.3.

3.3 Core Properties

Diagrams are replaced with normative prose, which clearly specifies requirements.

3.4 Digital Signatures

As part of this revision, the addition of XAdES digital signatures was planned. However, WG4 finally decided that this revision does not introduce any new features. Lots of clarifications were made though, as requested by DRs:

10-0043	OPC: Non-ambiguity of DC identifiers
10-0048	OPC: Processing model for handling ZIP encryption
11-0029	OPC: Do not copy text or schemas from W3C XML Signature
11-0030	OPC: Obsolete version of W3C XML Digital Signature 1.0
11-0031	OPC: Use official RELAX NG schemas from W3C
12-0001	OPC: Correct Spelling of "relationship part"

Note: The relationship type for digital signatures in MS Office documents and that in the currently-published 29500-2 are different. The ongoing revision fixed this problem by updating the relationship type.

4. Misc

The clause for acronyms and abbreviations was dropped since it does not make sense to for an ISO/IEC standard to define "ISO" and "IEC".

Schemas are not incorporated but rather referenced by URI.