ISO/IEC 29500 Strict/Transitional Forms

Discussion Document for ISO/IEC JTC 1/SC 34/WG 4

Author: Alex BROWN

Contributors: Jesper Lund STOCHOLM Shawn VILLARON Dave WELSH Mohamed ZERGAOUI

1.0 Background

This document was produced by members of WG 4. Its purpose is to inform onward discussion on the maintenance of ISO/IEC 29500.

1.1 Standardization

1.1.1 Ecma-376

There was only one set of schemas in the Ecma-376 standard and the concept of "strict" and "transitional" did not, at the time, exist.

1.1.2 The DIS 29500 Ballot and its Resolution

During the Fast Track standardization of Ecma-376 many commentators and participants took particular exception to the legacy features contained in the specification, most notoriously the *autoSpaceLikeWord95* and similar elements ("compatibility settings") that were particular to certain application-specific behaviour. Consequently many of the comments submitted in the DIS ballot requested removal or amendment of these features.

In response, Ecma proposed that many of these features were to be described as "deprecated" — however the standard as a whole still kept these deprecated parts together with non-deprecated parts in both the narrative text and schemas.

During the Ballot Resolution Meeting the decision was taken (Resolution 10) to split the deprecated features out of the body text and place them in a new annex:

Requests have been made to implement a more formal separation of "deprecated" features and to avoid the term "deprecated". Canada pro-

poses meeting these requirements by introducing strict and transitional conformance classes. Strict and transitional conformance classes determine verifiably different types of documents. The primary difference is that features from the proposed Annex A, "Selected Transitional Migration Features," are prohibited from strict documents, but are allowed in transitional documents. [...]

The intent [...] is to enable a transitional period during which existing binary documents being migrated to DIS 29500 can make use of legacy features to preserve their fidelity, while noting that new documents should not use them.

[...] This annex is normative for the current edition of the Standard, but not guaranteed to be part of the Standard in future revisions. The intent is to enable the future DIS 29500 maintenance group to choose, at a later date, to remove this set of features from a revised version of DIS 29500.

Originally this was intended to form the basis of an annex to the standard, but during the subsequent partitioning of the text during the BRM it became one of the Parts of IS 29500, Part 4. Thus the language of the Canadian resolution was repeated in the subsequent US Resolution 15 that partitioned the standard into four parts.

However, due to an oversight on Ecma's part these words were not incorporated into the published text, and so *as published* IS 29500 does not contain the BRM-recommended prohibition on using the Transitional form for new documents. This fault was noted in a comment from Denmark, "DR 09-0192 — BRM Resolution not implemented (Transitional Migration Features)", and a correction is included in the DCORs currently under ballot.

1.1.3 ISO/IEC 29500

In its first year as a published standard SC 34/WG 4 has maintained IS 29500, and while for the most part its maintenance activity has concerned small-scale technical fixes, for some questions (like those surrounding the Namespace change and date representation) it has been clear that an understanding of the relationship between the Strict and Transitional forms would help efficient and coherent decision-making. In an effort to achieve this WG 4 intends to arrive at such an understanding and to document it so as to guide its onward maintenance.

As published, Part 4 is an "editing list" that is to be read as patching Part 1. Thus the Transitional form is a "virtual text" to be made by mentally performing this patching operation. A consequence of this inheritance relationship is that Part 4 (and hence the Transitional form itself) automatically gains any new modifications made to Part 1, unless text is specifically introduced into Part 4 to derogate from such modifications.

2.0 Considerations

2.1 New documents should not use them?

One of the provisions of ISO/IEC 29500 is that new documents should not use legacy features (see the text of BRM Resolutions 10 and 15).

However, the precise definition of a "new document" is problematic:

- Are documents created by converting legacy binary documents "new"?
- Are documents created by legacy software (and some converters) "new"?
- Are documents created by modifying old OOXML documents "new"?
- Are documents created by new software but used by legacy software "new"?

There are some scenarios where WG 4 could reasonably agree a document was "new"; for example a document created by a fresh installation of new software and intended to be used by new software only. Likewise we can reasonably assert that all the documents that pre-date even Ecma-376 are "legacy" ones. However, there is a big gray area between these.

2.2 MCE

MCE is currently specified in ISO/IEC 29500-3:2008, but Parts 1 or 4 do not normatively reference MCE. While it seems likely there will be consensus that the Standard should be amended or corrected so that MCE becomes usable in conformant documents and implementations, the option here is *how* exactly that should be done. Specifically, should MCE be allowed just for OOXML Strict, or for both Strict and Transitional forms?

2.3 Date/Time Representation

While being a vexed topic in its own right, the issue of date representation in OOXML crystalizes several aspects of the Strict/Transitional debate.

The original published 29500 Standard could be read as prohibiting the use of serial date values in both Strict and Transitional versions, yet such a prohibition was clearly at odds with the stated scope of the standard since the ISO 8601 data system is not capable of loss-lessly representing the existing corpus of legacy documents, most notoriously the "date value" of 29 February 1900¹.

The current Part 4 DCOR ballot decisively addresses the question of whether serial date value representation should be included in the Transitional form, but regardless of the outcome of the ballot serious issues remain to be addressed in IS 29500's use of ISO 8601 dates, as has been stated in Defect Reports GB DR 09-0081 and GB DR 09-0082.

2.4 Consolidating the text

As noted above, the text of the Transitional form is currently a virtual text. To construct it,

^{1.} This follows a convention established by a bug in the early spreadsheet software Lotus 1-2-3.

Part 1 is patched with Part 4 (which is, essentially an editing list). Opinions are divided on how usable this makes the Transitional form as a specification.

2.5 Stabilization

On the occasion of a Standard's five-yearly review¹ it may be formally designated as "stabilized" (i.e. that it "will not require further maintenance of any sort").

The question arises: will the Transitional form be a suitable candidate for stabilization in the future (2013, 2018, ...)?

There is an important interaction between this question and the question of consolidating the text mentioned above. If WG 4 chooses to maintain the "editing list" nature of Part 4, and wishes to have the option of keeping the feature-set of the Transitional form distinct, they can stablize Part 4 but cannot stablize the Transitional form, since changes in Part 1 will automatically change the Transitional form.

2.6 Withdrawal

Another option for either or both of the Strict and Transitional forms is that they are withdrawn after review (2013, 2018, ...). Some think the Transitional form is not suitable for withdrawal, as a vast corpus of legacy documents it is intended to represent will remain somewhat in existence; however others think the purpose of the Transitional form was merely to act as a transient documentation format, and that it will have served its purpose by the time of review. Nevertheless, there is room for debate over how long a suitable transition window is.

The Strict form presents different questions. Lack of implementation is a *prima facie* justification for withdrawal, and to date there are no implementations of it. If this situation persists it may, or may not, be that it is best withdrawn to prevent marketplace confusion.

2.7 The subset relationship

Since the effect of BRM resolution 10 was to take a set of items of the standard and remove them from the Strict form, a natural consequence is that – at that time – the Strict form became a subset of the Transitional form. After standardization it also emerged that this subset relationship was also true of the XML schemas.

The subset relationship was not explored during the DIS standardization process, and is not formally described in the standard. Whether it persists will depend on the answer to the other questions posed by this document.

The subset relationship is proposed to be broken in the amendments currently under ballot (in the FPDAM for Part 1). These will move the declarations of the Strict schemas to a set of Namespaces that differs from that used by the Transitional form.

^{1.} See JTC 1 Directives (5th Edition) 15.6 for a description of stabilized Standards. Note it is not entirely clear what "review cycle" means in 15.6.1 - does this mean an entire 5 years must pass after a modification before stabilization can occur? Or just a review itself?

2.8 Versioning

One of the challenges facing WG 4 is how to version the format. The amendments currently under ballot provide the opportunity to provide versioning of the Strict form, since a new Namespace makes Strict files distinguishable from Transitional files and the lack of an installed base means that versioning attributes may be specified and respected. However the Standard does not provide any mechanism for legacy applications that consume Transitional files to detect modifications to that format: of course this is impossible since the format is already released in its Ecma-376 form, and consumed by software which cannot detect versioning information. As was agreed in Prague, *we cannot travel in time*.

Faced with this problem before, it might be noted that WG4's preference was to cause deliberate format breakage (by changing the Namespace) rather than risk silent data loss. Note however the change was made for a conformance class not in widespread usage.

2.9 Extensions

If NBs propose extensions for OOXML the question will arise whether they should be applied to the Strict or Transitional forms, or to both of them.

How should any such extensions be added: by amending the schema/text; using MCE, new OPC parts, or other mechanisms?

2.10 Deprecation in International Standards

Other ISO and IEC standards use the concept of deprecation to make plain the intended future maintenance plan for provisions in Standards. For example ISO/IEC 18026 contains a normative annex for a "change and deprecation plan" for its registered concepts.

If WG 4 wished to deprecate the Transitional form then it would need to define precisely what was meant by deprecation (ideally based on established practice for other ISO/IEC standards). Providing a "change and deprecation plan" would, in such circumstances, be helpful for users.

The term "transitional" is not normatively defined by IS 29500.

2.11 "Strict" and "Transitional" as W3C terms

The W3C produced transitional variants of HTML (one with an SGML, one an XML specification). The comment at the head of the transitional DTD states:

This is the HTML 4.01 Transitional DTD, which includes presentation attributes and elements that W3C expects to phase out as support for style sheets matures. Authors should use the Strict DTD when possible, but may use the Transitional DTD when support for presentation attribute and elements is required.

In its intent this appears consistent with the terminology of IS 29500: a transitional format is one which is expected to be phased out; creators should use the strict variant where possible.

W3C did not amend the transitional formats except once to fix minor errors.¹ These were one-time formats to be used as a fixed transition point for users. There was no ongoing subset relationship in this case.

^{1.} Version 4.01 fixed "minor errors" found in 4.0; see http://www.w3.org/MarkUp/#html4