
TechNote 0001: PDF/A-1 and Namespaces

Note The information provided in this TechNote has been published by ISO as PDF/A-1 Technical Corrigendum 1 (see [4]) in April 2007. However, this TechNote predates Technical Corrigendum 1 and we still provide it for consistency.

A PDF/A-1 file must be identified as such by the presence of a certain entry in its XMP document metadata. In order to be able to retrieve this entry reliably, it is essential that all syntactical requirements for storing metadata are followed precisely. In early implementations of PDF/A tools this was not always the case due to several reasons.

1 Implicit Provisions

The PDF/A-1 standard refers to several other standards and specifications and makes them as mandatory as if they were directly part of the standard. One such specification is the XMP specification [1], which defines how metadata has to be structured according to the XMP syntax. XMP in turn refers to the RDF specification [2] which describes the underpinnings of the XMP syntax (in a way XMP can be seen as a special case of RDF). Among the provisions in RDF there is one that requires that a namespace URI must end with a slash character (or with a number sign '#' followed by a fragment identifier).

While this applies to any namespace URI in a PDF's XMP metadata, it is especially critical for the namespace in which the PDF/A-1 identifying entry is stored. Adobe Acrobat Professional Distiller 7.07 (note that Acrobat 7.0 supported only a draft version of PDF/A) erroneously used a namespace URI for the PDF/A-1 identifier with a `.html` suffix. The Preflight Plugin of Acrobat 7.07 matches Distiller's behavior, and only (and incorrectly so) accepts the PDF/A URI namespace if it has a `.html` suffix:

```
http://www.aiim.org/pdfa/ns/id.html Wrong, but used by Acrobat 7.07!
```

This caused some confusion among those developers who assumed that if they mirrored the behavior of Acrobat 7.07 they would be doing the right thing.

Note that the namespace used for creating and validating PDF/A in Acrobat 8 changed as described below. Acrobat 8 therefore fully complies with the recommendations in this Technical Note.

2 Inaccuracies in the Wording of the PDF/A-1 Standard

The original version of ISO 19005-1, as published in October 2005, mentions the URI namespace URI for the PDF/A-1 identifying entry inaccurately and omits the trailing slash:

```
http://www.aiim.org/pdfa/ns/id Wrong, but published in ISO 19005-1!
```

This led most developers not familiar with the requirements of the underlying XMP and RDF specifications to believe that this is the way the namespace has to be written. Thus a few early PDF/A-1 creation tools produced PDF/A-1 files with PDF/A-1 namespace URIs which are incorrect.

The PDF/A committee in ISO (more precisely: TC 171, SC2, WG5) soon recognized this issue and started working on a corrigendum which addresses this and a few other inaccuracies. This corrigendum has been published in April 2007. As it is obvious that the correct way of using a namespace URI requires the trailing slash to be included in the URI, it is recommended that all PDF/A tools honor this requirement even before the corrigendum has been published. Thus the correct PDF/A-1 namespace URI must look exactly like this:

```
http://www.aiim.org/pdfa/ns/id/ Correct; only this should be used!
```

The complete XMP portion in a PDF/A-1 document's metadata which identifies the document as a PDF/A-1 file should look as in one of the two following variants (the examples indicate that the document is a PDF/A-1b file). The first variant uses the element syntax for the `pdfaid` entries:

```
<rdf:Description rdf:about=""
  xmlns:pdfaid="http://www.aiim.org/pdfa/ns/id/"
>
  <pdfaid:part>1</pdfaid:part>
  <pdfaid:conformance>B</pdfaid:conformance>
</rdf:Description>
```

The second variant uses the attribute syntax for the `pdfaid` entries:

```
<rdf:Description rdf:about=""
  xmlns:pdfaid="http://www.aiim.org/pdfa/ns/id/"
  pdfaid:part="1"
  pdfaid:conformance="B"
/>
```

The PDF/A Competence Center strongly recommends to use either the element or attribute syntax for the part and conformance entries, and to avoid mixtures of both. Note that variations in the XML syntax are of course allowed, e.g. whitespace or choice of quotation characters.

Bibliography

- [1] XMP specification (January 2004)
www.aiim.org/documents/standards/xmpspecification.pdf
- [2] RDF specification
www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/
- [3] ISO 19005-1: Document management — Electronic document file format for long-term preservation (PDF/A-1) — Part 1: Use of PDF 1.4 (PDF/A-1)
www.iso.ch
- [4] ISO 19005-1: Document management — Electronic document file format for long-term preservation — Part 1: Use of PDF 1.4 (PDF/A-1). Technical Corrigendum 1
www.iso.ch
- [5] PDF 1.4 specification
www.aiim.org/documents/standards/PDFreference.pdf

Copyright and Usage

Copyright © 2006-2008 PDF/A Competence Center, www.pdfa.org
You can link to the original location of this document. However, redistributing this document is only allowed with written approval.

Please contact info@pdfa.org if you have any questions regarding the contents of this TechNote or the redistribution policy.

Status of this Document

2006-12-06 First released version

2007-04-16 Re-released without any changes in content to fix problems in the PDF document's XMP metadata caused by two bugs in Acrobat Distiller 8

2008-03-14 Update:

- Updated formatting and added references to the entries in the bibliography
- Added note explaining the relationship of this TechNote and Corrigendum 1
- Bibliography: replaced an obsolete reference with a reference to Technical Corrigendum 1
- Copyright and Usage: added contact info