

---

# TechNote 0009: XMP Extension Schemas in PDF/A-1

This TechNote explains how to construct an XMP extension schema for PDF/A-1. It complements TechNote 0008: Predefined XMP Properties in PDF/A-1 [5]. Both TechNotes should be used in combination as follows:

- If metadata requirements for PDF/A-1 documents arise, first consult the list of predefined XMP properties in TechNote 0008. Identify appropriate metadata schemas and properties which match the requirements. If no such properties are available proceed to the next step.
- If no predefined XMP schema can be used, an XMP extension schema for PDF/A-1 must be defined and included in the document's XMP metadata. This TechNote explains the details of this process, and presents an example of such an XMP extension schema and the corresponding description which is required in PDF/A-1.

In addition to talking about XMP extension schemas for PDF/A-1 this TechNote includes various general notes on XMP and Acrobat. These may be useful when working with XMP extension schemas.

Extension schemas can also add properties to the predefined XMP schemas. In order to add new properties to one of the predefined schemas you must also create an extension schema according to this TechNote. It is recommended to not include descriptions of predefined properties in such an extension schema container schema. However, if predefined properties are included in an extension schema it is recommended that PDF/A validators ignore such property descriptions.

**Note** The term "schema" in the context of XMP metadata for PDF/A-1 is unrelated to XML schema definitions (.xsd).

---

## 1 XMP in Acrobat 8

### 1.1 Displaying XMP with Document Properties in Acrobat 8

In order to display XMP metadata with Acrobat 8 proceed as follows:

- Click on File, Properties...
- In the resulting dialog, click Additional Metadata...
- In the list on the left side of the resulting panel click Advanced. Acrobat will display a tree structure with namespaces and corresponding elements/attributes.

Be warned that the dialog does not accurately reflect the XMP data in the PDF document. For example, the display will be different from the actual file contents in the following cases:

- If you open a document which contains invalid XMP, Acrobat 8.1 may silently discard the XMP and construct a minimum replacement XMP from the entries in the document information dictionary.
- The Advanced panel will present XMP entries which are not actually present in the PDF document, e.g. `tiff:DateTime` will automatically be created based on `xmp:CreateDate`.

### 1.2 Displaying XMP with Acrobat 8 Preflight

Unlike the Document Properties method mentioned above, the Preflight plugin displays an unmodified view of the XMP metadata in the PDF document. Proceed as follows:

- Open the document, choose Advanced, Preflight..., and execute one of the PDF/A-1a or PDF/A-1b Preflight profiles.
- Go to the Results panel, and navigate to Overview, Document information, Metadata, Advanced. Here you can see all namespaces used in the XMP along with the respective metadata entries.
- In the Preflight window you can alternatively use Options, Browse Internal PDF Structure..., click on Catalog, Metadata, XML stream.

However, this method will not reflect any unsaved XMP changes you may have applied in Acrobat, but the XMP contents of the unmodified PDF file.

### **1.3 Importing XMP into PDF/A with Acrobat 8**

In order to import an XMP stream into an existing PDF/A document with Acrobat 8.1 proceed as follows:

- Disable PDF/A mode since Acrobat does not allow any XMP modifications for documents which are viewed in PDF/A mode: Edit, Preferences, [General], Documents, View documents in PDF/A mode: Never.
- Choose File, Properties... (or Ctrl-D), Additional Metadata..., Advanced, Replace... or Append..., and select your XMP file.
- If Acrobat accepts the XMP, it will present all entries in the Advanced panel.

### **1.4 Validating XMP with Acrobat 8 Preflight**

The Acrobat 8.1 Preflight plugin checks the required PDF/A entries in the XMP metadata, and also checks whether the standard document information entries match the corresponding predefined XMP properties. However, the PDF/A profiles in Acrobat 8.1 Preflight do not check XMP extension schemas for PDF/A conformance.

## 2 Validating and Testing XMP

Here are some hints for validating and testing XMP metadata for use with PDF/A-1:

### **XMP must be wellformed XML**

Obviously, your XMP must constitute wellformed XML. You can check this with any XML validation tool.

### **RDF Validation**

Since XMP is a subset of RDF, XMP metadata must comprise valid RDF. You can therefore use the W3C online validator for RDF [8]. Valid XMP should not result in any validation errors.

As an aside, the W3C validator is also capable of creating nice graphs for RDF/XMP input.

### **Use XMP in Acrobat 8**

Work with XMP metadata in Acrobat 8 as discussed in chapter 1, "XMP in Acrobat 8" (page 2). Import XMP into a PDF/A document with Acrobat 8; Acrobat will reject malformed or invalid XMP. However, the message `unknown error` which is displayed in many cases may not be very helpful for isolating the problem.

## 3 RDF Syntax Issues

Since XMP is based on a subset of RDF, the RDF syntax is relevant for XMP. In this section we will point out several relevant aspects of RDF syntax rules.

### 3.1 RDF Shorthand Notation using Property Attributes

The RDF Syntax Specification [7] mandates in section 2.5 "Property Attributes":

*When a property element's content is string literal, it may be possible to use it as an XML attribute on the containing node element. This abbreviation is known as a Property Attribute and can be applied to any node element.*

The resulting shorthand notation with property attributes can also be used in XMP, and must be supported by PDF/A-1 conforming software.

For example, the following denotes the PDF/A-1 conformance entries using standard element syntax:

```
<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 following variant uses the equivalent attribute syntax for the `pdfaid` entries. Both conformance entries are written as an attribute for the containing `rdf:Description` element:

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

The property attribute variant results in somewhat smaller output.

### 3.2 Omitting Blank Nodes with `rdf:parseType="Resource"`

The RDF Syntax Specification [7] mandates in section 2.11 "Omitting Blank Nodes: `rdf:parseType="Resource"`":

*Blank nodes (not RDF URI reference nodes) in RDF graphs can be written in a form that allows the `<rdf:Description>` `</rdf:Description>` pair to be omitted. The omission is done by putting an `rdf:parseType="Resource"` attribute on the containing property element.*

Note that `rdf:parseType` cannot be used with property attributes, but only property elements.

For example, in the following excerpt the `rdf:Description` node contains two elements, but no attributes:

```
<acmeemail:From>
  <rdf:Description>
    <mailto:name>John Doe</mailto:name>
    <mailto:mailto>john@acme.com</mailto:mailto>
  </rdf:Description>
</acmeemail:From>
```

The `rdf:Description` node can be abbreviated by replacing the element with a `rdf:parseType` attribute in the `acmeemail:From` as follows:

```
<acmeemail:From rdf:parseType="Resource">
  <mailto:name>John Doe</mailto:name>
  <mailto:mailto>john@acme.com</mailto:mailto>
</acmeemail:From>
```

The `rdf:parseType` variant is recommended since it results in shorter XMP output.

### 3.3 Unqualified RDF Attributes are deprecated

The RDF Syntax Specification [7] mandates in section 6.1.4 "Attribute Event":

*Several RDF attributes, including `about`, `parseType`, and a few others can be written without any namespace prefix. ... New documents SHOULD NOT use these unqualified attributes and applications MAY choose to warn when the unqualified form is seen in a document.*

For example, the following construct is considered deprecated, and should not be used:

```
<rdf:Description about="">
```

The following variant is preferred over the above since it avoids the deprecated use of an unqualified `about` attribute:

```
<rdf:Description rdf:about="">
```

While Acrobat 8 accepts unqualified `about` attributes, the W3C Validator warns as follows:

Error: {W102} unqualified use of `rdf:about` is deprecated.

Unqualified `about` attributes are occasionally used in examples in the XMP Specification [4], but should generally be avoided in XMP for PDF/A-1.

---

## 3.4 Ordering

The PDF/A-1 standard does not require any specific ordering of the elements in the XMP metadata. In particular, the description of an extension schema could be provided after its use. XMP readers with PDF/A-1 support should be prepared to deal with properties in extension schemas which appear before the corresponding schema description in the XMP:

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
...
<!-- Use the XMP extension schema -->
<rdf:Description rdf:about=""
  xmlns:acmeemail="http://www.acme.com/ns/email/1/" >
...
</rdf:Description>

<!-- XMP extension schema container schema -->
<rdf:Description rdf:about=""
  xmlns:pdfaExtension="http://www.aiim.org/pdfa/ns/extension/" >

  <!-- Container for all embedded extension schema descriptions -->
  <pdfaExtension:schemas>
    ...
  </pdfaExtension:schemas>
</rdf:Description>

</rdf:RDF>
```

## 4 PDF/A-1 Extension Schema Container Schema

This section describes the mechanism for defining XMP extension schemas which is provided by PDF/A-1. It should be read in parallel with the example in chapter 5, "Extension Schema Example" (page 12).

### 4.1 Extension Schema Summary

The table below summarizes the schemas which must be used to construct an extension schema description in PDF/A-1. They are collectively called "PDF/A-1 extension schema container schema". Unlike predefined XMP schemas, the namespace prefixes are not only preferred, but required.

schema name and description	namespace URI	required namespace prefix
PDF/A extension schema container schema <sup>1</sup>	<a href="http://www.aiim.org/pdfa/ns/extension/">http://www.aiim.org/pdfa/ns/extension/</a>	pdfaExtension
PDF/A field type schema	<a href="http://www.aiim.org/pdfa/ns/field#">http://www.aiim.org/pdfa/ns/field#</a> <sup>2</sup>	pdfaField
PDF/A property value type	<a href="http://www.aiim.org/pdfa/ns/property#">http://www.aiim.org/pdfa/ns/property#</a> <sup>2</sup>	pdfaProperty
PDF/A schema value type	<a href="http://www.aiim.org/pdfa/ns/schema#">http://www.aiim.org/pdfa/ns/schema#</a>	pdfaSchema
PDF/A ValueType value type	<a href="http://www.aiim.org/pdfa/ns/type#">http://www.aiim.org/pdfa/ns/type#</a> <sup>2</sup>	pdfaType

1. The description of this schema is missing in ISO 19005-1, and has been added in [2].
2. This namespace URI is incorrectly described in ISO 19005-1, and has been corrected in [2].

### 4.2 PDF/A Extension Schema Container Schema

This schema is required for defining XMP extension schemas. It is a container with one or more extension schemas. The description of this schema is missing in ISO 19005-1, and was added in [2].

Schema namespace URI: <http://www.aiim.org/pdfa/ns/extension/>

Required schema namespace prefix: pdfaExtension

property name	value type	PDF/A-1 description	notes
pdfaExtension:schemas	bag Schema	Container for all embedded extension schema descriptions	All extension schemas used in the document must be defined here according to section 4.3.



### 4.3 PDF/A Schema Value Type

This schema describes a single extension schema which may comprise an arbitrary number of properties.

Schema namespace URI: <http://www.aiim.org/pdfa/ns/schema#>

Required schema namespace prefix: pdfaSchema

property name	value type	PDF/A-1 description	notes
pdfaSchema:namespaceURI	URI	Schema namespace URI	Unique URI which describes the schema
pdfaSchema:prefix	Text	Preferred schema namespace prefix	This prefix can be used in addition to the predefined XMP namespace prefixes.
pdfaSchema:property	seq Property	Description of schema properties	All properties in the extension schema must be defined here according to section 4.4.
pdfaSchema:schema	Text	Optional description of schema	Human-readable text
pdfaSchema:valueType	seq ValueType	Description of schema-specific value types	All types which are used in the extension schema, but are not defined in the XMP Specification must be defined here according to section 4.5. This property is required if custom types are used in the extension schema. If no custom types are used it may be absent or empty.

### 4.4 PDF/A Property Value Type

This schema describes a single property.

Schema namespace URI: <http://www.aiim.org/pdfa/ns/property#>

Required schema namespace prefix: pdfaProperty

property name	value type	PDF/A-1 description	notes
pdfaProperty:category	Closed Choice of Text	Property category: internal or external	Internal properties are created automatically from document content. External properties are based on user input.
pdfaProperty:description	Text	Description of the property	Human-readable text

property name	value type	PDF/A-1 description	notes
pdfaProperty:name	Text	Property name	The property names comprise the vocabulary defined by the schema. Property names must be valid XML element names.
pdfaProperty:valueType	Open Choice of Text	Value type of the property, drawn from XMP Specification, or an embedded PDF/A extension schema value type	Predefined XMP type names or names of custom types according to section 4.5 can be used.

## 4.5 PDF/A ValueType Value Type

The PDF/A ValueType schema is required for all property value types which are not defined in the XMP 2004 specification [4], i.e. for value types outside of the following list:

- Array types (these are container types which may contain one or more fields): Alt, Bag, Seq
- Basic value types: Boolean, (open and closed) Choice, Date, Dimensions, Integer, Lang Alt, Locale, MIMETYPE, ProperName, Real, Text, Thumbnail, URI, URL, XPath
- Media Management value types: AgentName, RenditionClass, Resource-Event, ResourceRef, Version
- Basic Job/Workflow value type: Job
- EXIF schema value types: Flash, CFAPattern, DeviceSettings, GPSCoordinate, OECF/SFR, Rational

Schema namespace URI: <http://www.aiim.org/pdfa/ns/type#>

Required schema namespace prefix: pdfaType

property name	value type	PDF/A-1 description	notes
pdfaType:description	Text	Description of the property value type	Human-readable text
pdfaType:field	seq Field	Optional description of the structured fields	Separate entries are required for all fields in a structured type.
pdfaType:namespaceURI	URI	Property value type field namespace URI	
pdfaType:prefix	Text	Preferred value type field namespace prefix	
pdfaType:type	Text	Property value type name	

## 4.6 PDF/A Field Value Type

This schema describes a field in a structured type. It is very similar to the PDF/A Property Value Type schema (see chapter 4.4), but defines a field in a structure instead of a property.

Schema namespace URI: <http://www.aiim.org/pdfa/ns/field#>

Required schema namespace prefix: pdfaField

property name	value type	PDF/A-1 description	notes
pdfaField:description	Text	Field description	Human-readable text
pdfaField:name	Text	Field name	Field names must be valid XML element names.
pdfaField:valueType	Open Choice of Text	Field value type, drawn from XMP Specification 2004, or an embedded PDF/A value type extension schema	Predefined XMP type names or names of custom types according to section 4.5 can be used.

## 5 Extension Schema Example

### 5.1 Description of the sample Extension Schema

The XMP 2004 specification [4] explains in section "Creating Custom Schemas":

*To define a new schema, you should write a human-readable schema specification document. Your specification document should include:*

- *A unique name for your schema in the form of a URI and a preferred prefix.*
- *A table containing the name of each property, the value type, and the description of the property.*

Let's prepare our extension schema example according to these recommendations. We will construct a sample schema for use in a fictitious e-mail archiving application. For each stored PDF containing one e-mail information about the delivery date, the sender and the receiver shall be recorded in the XMP extension schema. The company name ACME is assumed.

Human-readable schema description: ACME's schema for describing attributes of archived e-mails

Schema name: ACME E-Mail Schema

Schema namespace URI: <http://www.acme.com/ns/email/1/>

Preferred schema namespace prefix: `acmeemail`

The properties of this schema are provided in the table below:

property name	value type	description
acmeemail:Delivery-Date	Date	date of email delivery
acmeemail:From	mailaddress	sender email address
acmeemail:To	mailaddress	receiver email address

The mailaddress data type is a structured type containing the following fields:

field name	value type	description
add:name	Text	plaintext name
add:mailto	Text	email address

## 5.2 Resulting XMP

The XMP extension schema for our sample application is presented below. While it does include the required PDF/A-1 identification entries, it does not include any other metadata entries.

**Note** Some entries may span multiple lines in the readable version below, but this shall not represent linebreak characters in the XMP.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">

<!-- Predefined properties (add pdf:Producer etc. as appropriate) -->
<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>

<!-- XMP extension schema container schema -->
<rdf:Description rdf:about=""
  xmlns:pdfaExtension="http://www.aiim.org/pdfa/ns/extension/"
  xmlns:pdfaSchema="http://www.aiim.org/pdfa/ns/schema#"
  xmlns:pdfaProperty="http://www.aiim.org/pdfa/ns/property#"
  xmlns:pdfaType="http://www.aiim.org/pdfa/ns/type#"
  xmlns:pdfaField="http://www.aiim.org/pdfa/ns/field#" >

<!-- Container for all embedded extension schema descriptions -->
<pdfaExtension:schemas>
  <rdf:Bag>
    <rdf:li rdf:parseType="Resource">
      <!-- Optional description of schema -->
      <pdfaSchema:schema>ACME E-Mail Schema</pdfaSchema:schema>

      <!-- Schema namespace URI -->

<pdfaSchema:namespaceURI>http://www.acme.com/ns/email/1/</pdfaSchema:namespaceURI>

      <!-- Preferred schema namespace prefix -->
      <pdfaSchema:prefix>acmeemail</pdfaSchema:prefix>

<!-- Description of the schema properties -->
<pdfaSchema:property>
  <rdf:Seq>
    <rdf:li rdf:parseType="Resource">
      <pdfaProperty:name>Delivery-Date</pdfaProperty:name>
      <pdfaProperty:valueType>Date</pdfaProperty:valueType>
      <pdfaProperty:category>internal</pdfaProperty:category>
      <pdfaProperty:description>date of email
delivery</pdfaProperty:description>
    </rdf:li>
    <rdf:li rdf:parseType="Resource">
```

```

        <pdfaProperty:name>From</pdfaProperty:name>

<pdfaProperty:valueType>mailaddress</pdfaProperty:valueType>
    <pdfaProperty:category>internal</pdfaProperty:category>
    <pdfaProperty:description>sender email
address</pdfaProperty:description>
    </rdf:li>
    <rdf:li rdf:parseType="Resource">
        <pdfaProperty:name>To</pdfaProperty:name>

<pdfaProperty:valueType>mailaddress</pdfaProperty:valueType>
    <pdfaProperty:category>internal</pdfaProperty:category>
    <pdfaProperty:description>receiver email
address</pdfaProperty:description>
    </rdf:li>
    </rdf:Seq>
</pdfaSchema:property>

<!-- Description of schema-specific value types -->
<pdfaSchema:valueType>
    <rdf:Seq>
        <rdf:li rdf:parseType="Resource">
            <pdfaType:type>mailaddress</pdfaType:type>

<pdfaType:namespaceURI>http://www.acme.com/ns/email/1/mailaddress/</pd
faType:namespaceURI>
    <pdfaType:prefix>add</pdfaType:prefix>
    <pdfaType:description>email
address</pdfaType:description>

    <!-- Optional description of the structured fields -->
    <pdfaType:field>
        <rdf:Seq>
            <rdf:li rdf:parseType="Resource">
                <pdfaField:name>name</pdfaField:name>
                <pdfaField:valueType>Text</pdfaField:valueType>
                <pdfaField:description>plaintext
name</pdfaField:description>
            </rdf:li>
            <rdf:li rdf:parseType="Resource">
                <pdfaField:name>mailto</pdfaField:name>
                <pdfaField:valueType>Text</pdfaField:valueType>
                <pdfaField:description>email
address</pdfaField:description>
            </rdf:li>
        </rdf:Seq>
    </pdfaType:field>
    </rdf:li>
    </rdf:Seq>
    </pdfaSchema:valueType>
</rdf:li>
</rdf:Bag>
</pdfaExtension:schemas>

```

```
</rdf:Description>
```

```
<!-- Use the XMP extension schema -->
```

```
<rdf:Description rdf:about=""
```

```
  xmlns:acmeemail="http://www.acme.com/ns/email/1/"
```

```
  xmlns:add="http://www.acme.com/ns/email/1/mailaddress/" >
```

```
  <acmeemail:Delivery-Date>2008-01-  
21T21:37:06+01:00</acmeemail:Delivery-Date>
```

```
  <acmeemail:From rdf:parseType="Resource">  
    <add:name>John Doe</add:name>  
    <add:mailto>john@acme.com</add:mailto>  
  </acmeemail:From>
```

```
  <acmeemail:To rdf:parseType="Resource">  
    <add:name>Jane Miller</add:name>  
    <add:mailto>jane@acme.com</add:mailto>  
  </acmeemail:To>  
</rdf:Description>
```

```
</rdf:RDF>
```

---

## Bibliography

- [1] ISO 19005-1: Document management — Electronic document file format for long-term preservation — Part 1: Use of PDF 1.4 (PDF/A-1)  
[www.iso.ch](http://www.iso.ch)
- [2] 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, published 2007-04-01  
[www.iso.ch](http://www.iso.ch)
- [3] PDF Reference: Adobe Portable Document Format, Version 1.4, Adobe Systems Incorporated – 3rd ed. (ISBN 0-201-75839-3).  
[www.aiim.org/documents/standards/PDFreference.pdf](http://www.aiim.org/documents/standards/PDFreference.pdf)
- [4] XMP Specification, January 2004, Adobe Systems Incorporated.  
[www.aiim.org/documents/standards/xmpspecification.pdf](http://www.aiim.org/documents/standards/xmpspecification.pdf)
- [5] TechNote 0008: Predefined XMP Properties in PDF/A-1, PDF/A Competence Center  
[www.pdfa.org/doku.php?id=pdfa:en:techdoc](http://www.pdfa.org/doku.php?id=pdfa:en:techdoc)
- [6] Resource Description Framework (RDF) Specification and other resources, World Wide Web Consortium (W3C)  
[www.w3.org/RDF/](http://www.w3.org/RDF/)
- [7] RDF/XML Syntax Specification (Revised), World Wide Web Consortium (W3C)  
[www.w3.org/TR/rdf-syntax-grammar](http://www.w3.org/TR/rdf-syntax-grammar)
- [8] W3C Validation Service for RDF, World Wide Web Consortium (W3C)  
[www.w3.org/RDF/Validator/](http://www.w3.org/RDF/Validator/)

## Copyright and Usage

Copyright © 2008 PDF/A Competence Center, [www.pdfa.org](http://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](mailto:info@pdfa.org) if you have any questions regarding the contents of this TechNote or the redistribution policy.

## Status of this Document

2008-03-20 First released version