XML Base (Second Edition)

W3C Recommendation 28 January 2009

This version:
http://www.w3.org/TR/2009/REC-xmlbase-20090128/
Latest version:
http://www.w3.org/TR/xmlbase/
Previous versions:
http://www.w3.org/TR/2001/REC-xmlbase-20010627/
http://www.w3.org/TR/2008/PER-xmlbase-20080320/
Editors:
Jonathan Marsh, Microsoft <jmarsh@microsoft.com>
Richard Tobin, University of Edinburgh <richard@inf.ed.ac.uk>

Please refer to the **errata** for this document, which may include normative corrections.

See also **translations**.

This document is also available in these non-normative formats: HTML with diff markup and XML.

Abstract

This document describes a facility, similar to that of HTML BASE, for defining base URIs for parts of XML documents.

Status of this Document

This section describes the status of this document at the time of its publication. Other documents may supersede this document. A list of current W3C publications and the latest revision of this technical report can be found in the W3C technical reports index at http://www.w3.org/TR/.

This document has been produced by the W3C XML Core Working Group as part of the W3C XML Activity. The English version of this specification is the only normative version. However, for translations of this document, see http://www.w3.org/2003/03/Translations/byTechnology?technology=xmlbase

This document is an Edited Recommendation of the W3C. It supersedes the previous W3C Recommendation of 27 June 2001. This second edition is not a new version of XML Base; its purpose is to clarify a number of issues that have become apparent since the first edition was published. Some of these were first published as separate errata (http://www.w3.org/2001/06/xmlbase-errata), others were published in a public editor's draft in November 2006 (http://www.w3.org/XML/2006/11/xmlbase-2e/Overview.html), and a PER in December 2006 (http://www.w3.org/TR/2006/PER-xmlbase-20061220/). The changes are summarized in an appendix.

Please report errors in this document to the public mailing list www-xml-linking-comments@w3.org; public archives are available.

There is no implementation report or test suite for this specification, but there is a document describing methods of testing XML Base conformance.

This document has been reviewed by W3C Members, by software developers, and by other W3C groups and interested parties, and is endorsed by the Director as a W3C Recommendation. It is a stable document and may be used as reference material or cited from another document. W3C's role in making the Recommendation is to draw attention to the specification and to promote its widespread deployment. This enhances the functionality and interoperability of the Web.

This document is governed by the 24 January 2002 CPP as amended by the W3C Patent Policy Transition Procedure. W3C maintains a public list of any patent disclosures made in connection with the deliverables of the group; that page also includes instructions for disclosing a patent. An individual who has actual knowledge of a patent which the individual believes contains Essential Claim(s) must disclose the information in accordance with section 6 of the W3C Patent Policy.

Table of Contents

1 Introduction
2 Terminology
3 xml:base Attribute
  3.1 URI Reference Encoding and Escaping
4 Resolving Relative URIs
  4.1 Relation to RFC 3986
  4.2 Granularity of base URI information
  4.3 Matching URIs with base URIs
  4.4 Interpretation of same-document references
5 Conformance

Appendices

A References
B References (Non-Normative)
C Impacts on Other Standards (Non-Normative)
D Changes since the first edition (Non-Normative)

---

1 Introduction

The XML Linking Language [XLink] defines Extensible Markup Language (XML) 1.0 [XML] constructs to describe links between resources. One of the stated requirements on XLink is to support HTML [HTML 4.01] linking constructs in a generic way. The HTML BASE element is one such construct which the XLink Working Group has considered. BASE allows authors to explicitly specify a document's base URI for the purpose of resolving relative URIs in links to external images, applets, form-processing programs, style sheets, and so on.

This document describes a mechanism for providing base URI services to XLink, but as a modular specification so that other XML applications benefiting from additional control over relative URIs but not built upon XLink can also make use of it. The syntax consists of a single XML attribute named `xml:base`.

The deployment of XML Base is through normative reference by new specifications, for example XLink and the XML Infoset. Applications and specifications built upon these new technologies will natively support XML Base. The behavior of `xml:base` attributes in applications based on specifications that do not have direct or indirect normative reference to XML Base is undefined.

This specification does not attempt to specify which text strings in a document are to be interpreted as URIs. That is the responsibility of each XML vocabulary. The question addressed by this specification is: given a relative URI in an XML document, what base URI is it resolved against?

It is expected that a future RFC for XML Media Types will specify XML Base as the mechanism for establishing base URIs in the media types it defines.

2 Terminology

[Definition: The key words **must**, **must not**, **required**, **shall**, **shall not**, **should**, **should not**, **recommended**, **may**, and **optional** in this specification are to be interpreted as described in [RFC 2119].]

The terms **base URI** and **relative URI** are used in this specification as they are defined in [RFC 3986].

3 xml:base Attribute

The attribute xml:base may be inserted in XML documents to specify a base URI other than the base URI of the document or external entity. The value of this attribute is interpreted as a Legacy Extended IRI (LEIRI) as defined in the W3C Note "Legacy extended IRIs for XML resource identification" [LEIRI] .

In namespace-aware XML processors, the "xml" prefix is bound to the namespace name http://www.w3.org/XML/1998/namespace as described in Namespaces in XML [XML Names]. Note that xml:base can be still used by non-namespace-aware processors.

An example of xml:base in a simple document containing XLinks follows. XLink normatively references XML Base for interpretation of relative URI references in xlink:href attributes.

```
<?xml version="1.0"?>
<doc xml:base="http://example.org/today/"
     xmlns:xlink="http://www.w3.org/1999/xlink">
  <head>
    <title>Virtual Library</title>
  </head>
  <body>
    <paragraph>See <link xlink:type="simple" xlink:href="new.xml">what's
      new</link>!</paragraph>
    <paragraph>Check out the hot picks of the day!</paragraph>
    <olist xml:base="/hotpicks/">
      <item>
        <link xlink:type="simple" xlink:href="pick1.xml">Hot Pick #1</link>
      </item>
      <item>
        <link xlink:type="simple" xlink:href="pick2.xml">Hot Pick #2</link>
      </item>
      <item>
        <link xlink:type="simple" xlink:href="pick3.xml">Hot Pick #3</link>
      </item>
    </olist>
  </body>
</doc>
```

The URIs in this example resolve to full URIs as follows:

- "what's new" resolves to the URI "http://example.org/today/new.xml"
- "Hot Pick #1" resolves to the URI "http://example.org/hotpicks/pick1.xml"
- "Hot Pick #2" resolves to the URI "http://example.org/hotpicks/pick2.xml"
- "Hot Pick #3" resolves to the URI "http://example.org/hotpicks/pick3.xml"

Note:

This specification does not give the xml:base attribute any special status as far as XML validity is concerned. In a valid document the attribute must be declared in the DTD, and similar considerations apply to other schema languages.

3.1 URI Reference Encoding and Escaping

The value of an `xml:base` attribute is a Legacy Extended IRI (LEIRI) and may contain characters not allowed in URIs. (However, some characters allowed in LEIRIs are not legal XML characters, and cannot therefore appear in `xml:base` values.)

In accordance with the principle that percent-encoding must occur as late as possible in the processing chain, applications which provide access to the base URI of an element should calculate and return the value without escaping.

In the example below, the base URI of element `e2` should be returned as "http://example.org/wine/rosé".

```
<?xml version="1.0"?>
<e1 xml:base="http://example.org/wine/">
  <e2 xml:base="rosé"/>
</e1>
```

4 Resolving Relative URIs

4.1 Relation to RFC 3986

RFC 3986 [RFC 3986] provides for base URI information to be embedded within a document. The rules for determining the base URI can be summarized as follows (highest priority to lowest):

1. The base URI is embedded in the document's content.
2. The base URI is that of the encapsulating entity (message, document, or none).
3. The base URI is the URI used to retrieve the entity.
4. The base URI is defined by the context of the application.

Note:

The term "entity" in points #2 and #3 above uses the RFC 3986 meaning of the term. Elsewhere in this document the term "entity" is used in the XML sense.

This document specifies the details of rule #1 for embedding base URI information in the specific case of XML documents.

4.2 Granularity of base URI information

Relative URIs appearing in an XML document are always resolved relative to either an element, a document entity, or an external entity. There is no provision for finer granularity, such as per-attribute, per-character, or per-entity base information. Neither internal entities, whether declared in the internal subset or in an external DTD, nor freestanding text (text not enclosed in an element) in an external entity, are considered to set a base URI separate from the base URI in scope for the entity reference.

The base URI of a document entity or an external entity is determined by RFC 3986 rules, namely, that the base URI is the URI used to retrieve the document entity or external entity.

The base URI of an element is:

1. the base URI specified by an `xml:base` attribute on the element, if one exists, otherwise
2. the base URI of the element's parent element within the document entity or external entity, if one exists, otherwise
3. the base URI of the document entity or external entity containing the element.

   Note:

   It follows that the base URI specified by an `xml:base` attribute is inherited by descendant elements within the same entity until another element with an `xml:base` attribute is encountered.

The base URI of an element bearing an `xml:base` attribute with a value that is not a valid Legacy Extended IRI is application dependent.

4.3 Matching URIs with base URIs

The base URI corresponding to a given relative URI appearing in an XML document is determined as follows:

- The base URI for a URI reference appearing in text content is the base URI of the element containing the text.
- The base URI for a URI reference appearing in an `xml:base` attribute is the base URI of the parent element of the element bearing the `xml:base` attribute, if one exists within the document entity or external entity, otherwise the base URI of the document entity or external entity containing the element.
- The base URI for a URI reference appearing in any other attribute value, including default attribute values, is the base URI of the element bearing the attribute.

- The base URI for a URI reference appearing in the content of a processing instruction is the base URI of the parent element of the processing instruction, if one exists within the document entity or external entity, otherwise the base URI of the document entity or external entity containing the processing instruction.

Note:

The presence of `xml:base` attributes might lead to unexpected results in the case where the attribute value is provided, not directly in the XML document entity, but via a default attribute. For instance, such a declaration in an external entity might not be read by software which is based on a non-validating XML processor. Defaulting attributes through an external mechanism such as XML Schema may also lead to unexpected results; even if a validating processor is used by the application, the addition of defaulted attributes subsequent to creation of the infoset can cause xml:base attributes to get out of sync with the [base URI] infoset property. For these reasons, `xml:base` values should be provided either directly in the XML document instance or via default attributes declared in the internal subset of the DTD.

4.4 Interpretation of same-document references

RFC 3986 defines certain relative URI references, in particular the empty string and those of the form `#fragment`, as *same-document references*. Dereferencing of same-document references is handled specially. However, their use as the value of an `xml:base` attribute does not involve dereferencing, and XML Base processors should resolve them in the usual way. In particular, `xml:base=""` does not reset the base URI to that of the containing document.

Note:

Some existing processors *do* treat these `xml:base` values as resetting the base URI to that of the containing document, so the use of such values is strongly discouraged.

5 Conformance

An application conforms to XML Base if it calculates base URIs in accordance with the conditions set forth in this specification.

A References

RFC 2119
RFC 2119: Key words for use in RFCs to Indicate Requirement Levels. *Internet Engineering Task Force, 1997.*

RFC 3986
RFC 3986: Uniform Resource Identifier (URI): Generic Syntax. *Internet Engineering Task Force, 2005.*

LEIRI
Legacy extended IRIs for XML resource identification. Henry S. Thompson, Richard Tobin, and Norman Walsh, editors. *World Wide Web Consortium.*
XML
Extensible Markup Language (XML) 1.0. Tim Bray et al. *World Wide Web Consortium.*
XML Names
Namespaces in XML 1.0.. Tim Bray et al. *World Wide Web Consortium.*

B References (Non-Normative)

HTML 4.01
HTML 4.01 Specification. Dave Raggett, Arnaud Le Hors, Ian Jacobs, editors. *World Wide Web Consortium, 1999.*
XLink
XML Linking Language (XLink). Steve DeRose, Eve Maler, David Orchard, and Ben Trafford, editors. *World Wide Web Consortium, 2000.*
XML Datatypes
XML Schema Part 2: Datatypes. Paul V. Biron, Ashok Malhotra, editors. *World Wide Web Consortium Working Draft.*
XHTML
XHTML(TM) 1.0: The Extensible HyperText Markup Language. Steven Pemberton, et al. *World Wide Web Consortium, 2000.*
XML Infoset
XML Information Set. John Cowan and Richard Tobin, editors. *World Wide Web Consortium, 1999.*
XPath
XML Path Language James Clark and Steven DeRose, editors. *World Wide Web Consortium, 1999.*
XSLT
XSL Transformations. James Clark, editor. *World Wide Web Consortium, 1999.*

C Impacts on Other Standards (Non-Normative)

This section has been deleted.

D Changes since the first edition (Non-Normative)

1. The published errata (see http://www.w3.org/2001/06/xmlbase-errata) have been incorporated;

2. The definition of URI reference has been switched from RFC2396 to 3986;

3. The xml:base attribute has been redescribed as a Legacy Extended IRI, but this does not change its syntax (the December 2006 PER used the term "XML Resource Identifier" which was to be defined in an XLink revision, but that plan has been superseded by the definition of LEIRI in the W3C Note);

4. Implementations are now encouraged to return base “URIs” without escaping non-URI characters;

5. The meanings of xml:base="" and xml:base="#frag" have been clarified;

6. The expected reference to XML Base in the forthcoming XML Media Types RFC (“son of 3023”) has been noted;

7. It has been clarified that normal validity rules apply to the xml:base attribute;

8. The out-of-date appendix describing effects on other standards has been removed;

9. Various minor editorial changes have been made.