XML Base

W3C Recommendation 27 June 2001

This version:
http://www.w3.org/TR/2001/REC-xmlbase-20010627/ (available in HTML, XML)
Latest version:
http://www.w3.org/TR/xmlbase/
Previous versions:
http://www.w3.org/TR/2000/PR-xmlbase-20001220/
Editor:
Jonathan Marsh, Microsoft <jmarsh@microsoft.com>

Abstract

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

Status of this Document

This document has been reviewed by W3C Members and other interested parties and has been endorsed by the Director as a W3C Recommendation. It is a stable document and may be used as reference material or cited as a normative reference 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 has been produced by the W3C XML Linking Working Group as part of the XML Activity in the W3C Architecture Domain. For background on this work, please see the XML Activity Statement.

Please report possible errors in this document to the public email list www-xml-linking-comments@w3.org (archive at http://lists.w3.org/Archives/Public/www-xml-linking-comments/). Any confirmed errors will be documented in an list of errata available at http://www.w3.org/2001/06/xmlbase-errata.

The English version of this specification is the only normative version. Information

about translations of this document is available at http://www.w3.org/2001/06/xmlbase-translations.

A list of current W3C Recommendations and other technical documents can be found at http://www.w3.org/TR/.

Table of Contents

Appendices

---

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.

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 [IETF RFC 2119].]

The terms **base URI** and **relative URI** are used in this specification as they are defined in [IETF RFC 2396].

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 URI Reference as defined in RFC 2396 [IETF RFC 2396], after processing according to Section 3.1.

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'
      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</l
      </item>
      <item>
        <link xlink:type="simple" xlink:href="pick2.xml">Hot Pick #2</l
      </item>
      <item>
        <link xlink:type="simple" xlink:href="pick3.xml">Hot Pick #3</l
      </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"

3.1 URI Reference Encoding and Escaping

The set of characters allowed in `xml:base` attributes is the same as for XML,

namely [Unicode]. However, some Unicode characters are disallowed from URI references, and thus processors must encode and escape these characters to obtain a valid URI reference from the attribute value.

The disallowed characters include all non-ASCII characters, plus the excluded characters listed in Section 2.4 of [IETF RFC 2396], except for the number sign (#) and percent sign (%) characters and the square bracket characters re-allowed in [IETF RFC 2732]. Disallowed characters must be escaped as follows:

1. Each disallowed character is converted to UTF-8 [IETF RFC 2279] as one or more bytes.
2. Any bytes corresponding to a disallowed character are escaped with the URI escaping mechanism (that is, converted to %HH, where HH is the hexadecimal notation of the byte value).
3. The original character is replaced by the resulting character sequence.

4 Resolving Relative URIs

4.1 Relation to RFC 2396

RFC 2396 [IETF RFC 2396] 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 2396 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

2396 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 or external entity, if one exists, otherwise
3. the base URI of the document entity or external entity containing the element.

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 declared in an external entity. Such declarations might not be read by software which is based on a non-validating XML processor. Many XML applications fail to require validating processors. For correct operation with such applications, `xml:base` values should be provided either directly or via default attributes declared in the internal subset of the DTD.

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

IETF RFC 2119
RFC 2119: Key words for use in RFCs to Indicate Requirement Levels. Internet Engineering Task Force, 1997. (See http://www.ietf.org/rfc/rfc2119.txt.)
IETF RFC 2279
RFC 2279: UTF-8, a transformation format of ISO 10646. Internet Engineering Task Force, 1998. (See http://www.ietf.org/rfc/rfc2279.txt.)
IETF RFC 2396
RFC 2396: Uniform Resource Identifiers. Internet Engineering Task Force, 1995. (See http://www.ietf.org/rfc/rfc2396.txt.)
IETF RFC 2732
RFC 2732: Format for Literal IPv6 Addresses in URL's. Internet Engineering Task Force, 1999. (See http://www.ietf.org/rfc/rfc2732.txt.)
Unicode
The Unicode Standard. The Unicode Consortium. (See http://www.unicode.org/unicode/standard/standard.html.)
XML
Tim Bray, Jean Paoli, C.M. Sperberg-McQueen, and Eve Maler, editors. *Extensible Markup Language (XML) 1.0 (Second Edition).* World Wide Web Consortium, 2000. (See http://www.w3.org/TR/2000/REC-xml-20001006.)
XML Names
Tim Bray, Dave Hollander, and Andrew Layman, editors. *Namespaces in XML*. Textuality, Hewlett-Packard, and Microsoft. World Wide Web Consortium, 1999. (See http://www.w3.org/TR/1999/REC-xml-names-19990114/.)

B References (Non-Normative)

HTML 4.01
Dave Raggett, Arnaud Le Hors, Ian Jacobs, editors. *HTML 4.01 Specification*. World Wide Web Consortium, 1999. (See http://www.w3.org/TR/1999/REC-html401-19991224/.)
XLink
Steve DeRose, Eve Maler, David Orchard, and Ben Trafford, editors. *XML Linking Language (XLink).* World Wide Web Consortium, 2000. (See http://www.w3.org/TR/2001/REC-xlink-20010627/.)
XML Datatypes
Paul V. Biron, Ashok Malhotra, editors. *XML Schema Part 2: Datatypes.* World Wide Web Consortium Working Draft. (See http://www.w3.org/TR/2001/REC-xmlschema-2-20010502/.)
XHTML
Steven Pemberton, et al. *XHTML(TM) 1.0: The Extensible HyperText Markup Language.* World Wide Web Consortium, 2000. (See http://www.w3.org/TR/2000/REC-xhtml1-20000126/.)
XML Infoset
John Cowan and Richard Tobin, editors. *XML Information Set*. World Wide Web Consortium, 1999. (See http://www.w3.org/TR/2001/CR-xml-infoset-20010514/.)
XPath
James Clark and Steven DeRose, editors. *XML Path Language* World Wide Web Consortium, 1999. (See http://www.w3.org/TR/1999/REC-xpath-

19991116.)

XSLT

James Clark, editor. *XSL Transformations*. World Wide Web Consortium, 1999. (See http://www.w3.org/TR/1999/REC-xslt-19991116.)

C Impacts on Other Standards (Non-Normative)

XML Base defines a mechanism for embedding base URI information within an XML document. It does not define a mechanism to recognize which content or attribute values might contain URIs. This is only known by the specifications or applications assigning semantics to the vocabulary.

It is the intention of XML Base that future specifications and revisions of XML vocabularies identify which parts of the XML document are considered to be URIs, and provide normative reference to this specification in order to ensure that relative URIs are treated consistently across XML documents.

The impacts of XML Base on other standards (as of the publication date of this document) are described below.

- XML 1.0 [XML] uses URI references in the system identifiers for external entities. Since these declarations appear outside of the document element (in an internal subset or external DTD), the scoping rules for `xml:base` prevent these URIs from being affected by the value of `xml:base`.
- The XML Infoset [XML Infoset] defines the base URI property of element information items. The latest Infoset specification supports XML Base for purposes of determining the value of this property. Interfaces, applications, and specifications referencing this infoset property will support XML Base natively.
- Namespaces in XML [XML Names] uses URI references, which as currently defined should not be resolved relative to the base URI defined by `xml:base` for the purposes of namespace identification. Higher level processes which dereference namespace URIs are not covered by the namespaces specification and might at their option specify that `xml:base` is honored for the purposes of fetching resources at those URIs.
- The XPath [XPath] data model preserves neither base URI information nor the boundaries of external entities and thus is insufficient to support resolution of relative URI references within these entities to be resolved correctly. This includes relative URI references in `xml:base` attributes.
- The XSLT [XSLT] extensions to the XPath data model do provide for base URI information to be retained, but defines this information in a way that precludes support for XML Base. Future XSLT versions might want to require support for XML Base.
- XML Schema Part 2: Datatypes [XML Datatypes] defines a `uriReference` primitive datatype. The XML Datatypes specification might want to require that applications recognizing this datatype and resolving such URIs be aware of XML Base.
- The XLink [XLink] specification requires support for XML Base.
- XHTML [XHTML] uses URI references beyond those expressible in XLink. These URI references might be resolved by an application relative to the

base URI defined by XML Base. The XHTML specification might want to describe their level of support for XML Base.