



XHTML Media Types

W3C Note 1 August 2002

This version:

<http://www.w3.org/TR/2002/NOTE-xhtml-media-types-20020801>
([HTML](#), [XHTML](#))

Latest version:

<http://www.w3.org/TR/xhtml-media-types>

Previous version:

<http://www.w3.org/TR/2002/NOTE-xhtml-media-types-20020430>

Editor:

[石川 雅康 \(ISHIKAWA Masayasu\)](#), W3C

[Copyright](#) ©2002 [W3C](#)® ([MIT](#), [INRIA](#), [Keio](#)), All Rights Reserved. W3C [liability](#), [trademark](#), [document use](#) and [software licensing](#) rules apply.

Abstract

This document summarizes the best current practice for using various Internet media types for serving various XHTML Family documents. [In summary](#), 'application/xhtml+xml' **SHOULD** be used for XHTML Family documents, and the use of 'text/html' **SHOULD** be limited to HTML-compatible XHTML 1.0 documents. 'application/xml' and 'text/xml' **MAY** also be used, but whenever appropriate, 'application/xhtml+xml' **SHOULD** be used rather than those generic [XML](#) media types.

Status of This Document

This section describes the status of this document at the time of its publication. Other documents may supersede this document. The latest status of this document series is maintained at the W3C.

This document is a Note made available by the World Wide Web Consortium (W3C) for your information. Publication of this Note by W3C indicates no endorsement by W3C or the W3C Team, or any W3C Members.

This document represents the consensus of the [W3C HTML Working Group](#) ([members only](#)) on the usage of Internet media types for various XHTML Family documents. However, this document is not intended to be a normative specification. Instead, it documents a set of recommendations to maximize the interoperability of XHTML documents with regard to Internet media types. This document *does not* address general issues on media types and namespaces, which is outside the scope of the [HTML Working Group charter](#) and will be dealt by the [Technical Architecture Group \(TAG\)](#).

Comments on this document may be sent to www-html-editor@w3.org ([archive](#)). Public discussion on this document may take place on the mailing list www-html@w3.org ([archive](#)).

This document has been produced as part of the [W3C HTML Activity](#). The goals of the HTML Working Group are discussed in the HTML Working Group charter. A [list of current W3C Recommendations](#) and other technical documents can be found at <http://www.w3.org/TR>.

Table of Contents

1. [Introduction](#)
2. [Terms and Definitions](#)
3. [Recommended Media Type Usage](#)
 - 3.1. ['text/html'](#)
 - 3.2. ['application/xhtml+xml'](#)
 - 3.3. ['application/xml'](#)
 - 3.4. ['text/xml'](#)
 - 3.5. [Summary](#)
- [References](#)
- [Changes from Previous Version](#)

1. Introduction

XHTML 1.0 [[XHTML1](#)] reformulated HTML 4 [[HTML4](#)] as an XML application, and Modularization of XHTML [[XHTMLM12N](#)] provided a means to define XHTML-based markup languages using XHTML modules, collectively called as "XHTML Family". However, due to historical reasons, a recommended way to serve such XHTML Family documents, in particular with regard to Internet media types, was quite unclear.

"5.1 Internet Media Type" of the first edition of [[XHTML1](#)] included the following vague statement:

As of the publication of this recommendation, the general recommended MIME labeling for XML-based applications has yet to be resolved.

However, XHTML Documents which follow the guidelines set forth in [Appendix C](#), "HTML Compatibility Guidelines" may be labeled with the Internet Media Type "text/html", as they are compatible with most HTML browsers. This document makes no recommendation about MIME labeling of other XHTML documents.

Meanwhile, after the publication of [[XHTML1](#)], an RFC for XML media types was revised and published as RFC 3023 [[RFC3023](#)], and it introduced the '+xml' suffix convention for XML-based media types. The 'application/xhtml+xml' media type [[RFC3236](#)] was registered following that convention. Now that there are at least four possibilities on media type labeling for XHTML Family documents - 'text/html', 'application/xhtml+xml', and generic XML media types 'application/xml' and 'text/xml'.

This document summarizes the best current practice for using those various Internet media types for various XHTML Family documents.

2. Terms and Definitions

The key words "**MUST**", "**MUST NOT**", "**REQUIRED**", "**SHALL**", "**SHALL NOT**", "**SHOULD**", "**SHOULD NOT**", "**RECOMMENDED**", "**MAY**", and "**OPTIONAL**" in this document are to be interpreted as described in RFC 2119 [[RFC2119](#)].

XHTML

The Extensible HyperText Markup Language. XHTML is *not* the name of a single, monolithic markup language, but the name of a family of document types which collectively form this markup language. The namespace URI for XHTML is

<http://www.w3.org/1999/xhtml>.

Note: Future version of XHTML might use different namespace.

XHTML Family document type

A document type which belongs to a family of XHTML document types. Such document types include [[XHTML1](#)], and XHTML Host Language document types such as XHTML 1.1 [[XHTML11](#)] and XHTML Basic [[XHTMLBasic](#)]. Elements and attributes in those document types belong to the XHTML namespace (except those from the XML namespace, such as `xml:lang`), but an XHTML Family document type **MAY** also include elements and attributes from other namespaces, such as MathML [[MathML2](#)].

XHTML Host Language document type

A document type which conforms to the "XHTML Host Language Document Type Conformance" as defined in section 3.1 of [[XHTMLM12N](#)].

XHTML Integration Set document type

A document type which conforms to the "XHTML Integration Set Document Type Conformance" as defined in section 3.2 of [[XHTMLM12N](#)].

3. Recommended Media Type Usage

This section summarizes which Internet media type **SHOULD** be used for which XHTML Family document for which purpose.

3.1. 'text/html'

The 'text/html' media type [[RFC2854](#)] is primarily for HTML, not for XHTML. In general, this media type is **NOT** suitable for XHTML. However, as [[RFC2854](#)] says, "[\[XHTML1\]](#) defines a profile of use of XHTML which is compatible with HTML 4.01 and which may also be labeled as text/html".

[[XHTML1](#)], Appendix C "HTML Compatibility Guidelines" summarizes "*design guidelines for authors who wish their XHTML documents to render on existing HTML user agents*".

The use of 'text/html' for XHTML **SHOULD** be limited for the purpose of rendering on existing HTML user agents, and **SHOULD** be limited to [[XHTML1](#)] documents which follow the HTML Compatibility Guidelines. In particular, 'text/html' is **NOT** suitable for XHTML Family document types that adds elements and attributes from foreign namespaces, such as XHTML+MathML [[XHTML+MathML](#)].

XHTML documents served as 'text/html' will not be processed as XML [[XML10](#)], e.g. well-formedness errors may not be detected by user agents. Also be aware that HTML rules will be applied for DOM and style sheets (see C.11 and C.13 of [[XHTML1](#)] respectively).

Authors should also be careful about character encoding issues. A typical misunderstanding is that since an XHTML document is an XML document, the character encoding of an XHTML document should be treated as UTF-8 or UTF-16 in the absence of an explicit character encoding information. This is **NOT** the case when an XHTML

document is served as 'text/html'. "6. Charset default rules" of [RFC2854](#) notes as follows:

The use of an explicit charset parameter is strongly recommended. While [\[MIME\]](#) specifies "The default character set, which must be assumed in the absence of a charset parameter, is US-ASCII." [\[HTTP\]](#) Section 3.7.1, defines that "media subtypes of the 'text' type are defined to have a default charset value of 'ISO-8859-1'". Section 19.3 of [\[HTTP\]](#) gives additional guidelines. Using an explicit charset parameter will help avoid confusion.

Using an explicit charset parameter also takes into account that the overwhelming majority of deployed browsers are set to use something else than 'ISO-8859-1' as the default; the actual default is either a corporate character encoding or character encodings widely deployed in a certain national or regional community. For further considerations, please also see Section 5.2 of [\[HTML40\]](#).

"5.2.2 Specifying the character encoding" of the HTML 4 specification [\[HTML4\]](#) also notes that "*user agents must not assume any default value for the "charset" parameter*". Therefore, authors **SHOULD NOT** assume any default value for an XHTML document served as 'text/html', and as mentioned in [\[RFC2854\]](#), the use of an explicit charset parameter is **STRONGLY RECOMMENDED**. When it is difficult to specify an explicit charset parameter through a higher-level protocol, authors **SHOULD** include the XML declaration (e.g. `<?xml version="1.0" encoding="EUC-JP"?>`) and a meta http-equiv statement (e.g. `<meta http-equiv="Content-Type" content="text/html; charset=EUC-JP" />`). See "C.9. Character Encoding" of [\[XHTML1\]](#) for details.

3.2. 'application/xhtml+xml'

The 'application/xhtml+xml' media type [\[RFC3236\]](#) is **the** media type for XHTML Family document types, and in particular it is suitable for XHTML Host Language document types. XHTML Family document types suitable for this media type include [\[XHTML1\]](#), [\[XHTMLBasic\]](#), [\[XHTML11\]](#) and [\[XHTML+MathML\]](#). An XHTML Host Language document type that adds elements and attributes from foreign namespaces **MAY** identify its profile with the 'profile' optional parameter or other means such as the "Content-features" MIME header described in RFC 2912 [\[RFC2912\]](#). Each namespace **SHOULD** be explicitly identified through namespace declaration [\[XMLNS\]](#). This document does not preclude the registration of its own media type for specific XHTML Host Language document type.

In general, this media type is **NOT** suitable for XHTML Integration Set document types. This document does not define which media type should be used for XHTML Integration Set document types.

'application/xhtml+xml' **SHOULD** be used for serving XHTML documents to XHTML user agents. Authors who wish to support both XHTML and HTML user agents **MAY** utilize content negotiation by serving HTML documents as 'text/html' and XHTML documents as 'application/xhtml+xml'. Also note that it is not necessary for XHTML documents served as 'application/xhtml+xml' to follow the HTML Compatibility Guidelines.

When serving an XHTML document with this media type, authors **SHOULD** include the XML stylesheet processing instruction [\[XMLstyle\]](#) to associate style sheets.

As for character encoding issues, as mentioned in "6. Charset default rules" of [\[RFC3236\]](#), 'application/xhtml+xml' has the same considerations as 'application/xml'.

See [section 3.3](#) for details.

3.3. 'application/xml'

The 'application/xml' media type [\[RFC3023\]](#) is a generic media type for XML documents, and the definition of 'application/xml' does not preclude serving XHTML documents as that media type. Any XHTML Family document **MAY** be served as 'application/xml'.

However, authors should be aware that such a document may not always be processed as XHTML (e.g. hyperlinks may not be recognized), depending on user agents. Generic XML processors might recognize it as just an XML document which includes elements and attributes from the XHTML namespace (and others), and may not have a priori knowledge what to do with such a document beyond they can do for generic XML documents.

Authors **SHOULD** explicitly identify the XHTML namespace through the namespace declaration when they serve an XHTML Family document as 'application/xml' to facilitate the chance for reliable processing. The XML stylesheet PI **SHOULD** be used to associate style sheets.

Whenever appropriate, 'application/xhtml+xml' **SHOULD** be used rather than 'application/xml'.

As for character encoding issues, "3.2 Application/xml Registration" of [\[RFC3023\]](#) says that *"the use of the charset parameter is **STRONGLY RECOMMENDED**"*, and also specifies a rule that *"[i]f an application/xml entity is received where the charset parameter is omitted, no information is being provided about the charset by the MIME Content-Type header"*. This means that conforming XML processors **MUST** follow the requirements described in section 4.3.3 of [\[XML10\]](#).

Therefore, while it is **STRONGLY RECOMMENDED** to specify an explicit charset parameter through a higher-level protocol, authors **SHOULD** include the XML declaration (e.g. `<?xml version="1.0" encoding="EUC-JP"?>`). Note that a meta http-equiv statement will not be recognized by XML processors, and authors **SHOULD NOT** include such a statement in an XHTML document served as 'application/xml' (and 'application/xhtml+xml' as well for that matter).

3.4. 'text/xml'

The 'text/xml' media type [\[RFC3023\]](#) is an another generic media type for XML documents, and the definition of 'text/xml' does not preclude serving XHTML documents as that media type, either. Any XHTML Family document **MAY** be served as 'text/xml'. The considerations for 'application/xml' also apply to 'text/xml'. Whenever appropriate, 'application/xhtml+xml' **SHOULD** be used rather than 'text/xml'.

Authors should also be aware of the difference between 'application/xml' (and for that matter 'application/xhtml+xml' as well) and 'text/xml' with regard to the treatment of character encoding. According to "3.1 Text/xml Registration" of [\[RFC3023\]](#), *"if a text/xml entity is received with the charset parameter omitted, MIME processors and XML processors **MUST** use the default charset value of "us-ascii"[\[ASCII\]](#)"*. This default value is authoritative over the encoding information specified in the XML declaration, or the XML default encodings of UTF-8 and UTF-16 when no encoding declaration is supplied, so omitting the charset parameter of a 'text/xml' entity might cause an unexpected result. As mentioned in [\[RFC3023\]](#), the use of the charset parameter is **STRONGLY RECOMMENDED**.

3.5. Summary

The following table summarizes recommendation to content authors for labeling XHTML documents. HTML 4 is also listed for comparison purpose.

Media types summary for serving XHTML documents

Media type	HTML 4	XHTML 1.0 (HTML compatible)	XHTML 1.0 (other)	XHTML Basic / 1.1	XHTML+MathML
text/html	SHOULD	MAY	SHOULD NOT	SHOULD NOT	SHOULD NOT
application/xhtml+xml	MUST NOT	SHOULD	SHOULD	SHOULD	SHOULD
application/xml	MUST NOT	MAY	MAY	MAY	MAY
text/xml	MUST NOT	MAY	MAY	MAY	MAY

References

[ASCII]

"Information Systems -- Coded Character Sets -- 7-Bit American National Standard Code for Information Interchange (7-Bit ASCII)", ANSI X3.4-1986, 1986.

[HTML4]

"[HTML 4.01 Specification](#)", W3C Recommendation, D. Raggett, A. Le Hors, I. Jacobs, eds., 24 December 1999. Available at:

<http://www.w3.org/TR/1999/REC-html401-19991224>

The [latest version of HTML 4.01](#) is available at: <http://www.w3.org/TR/html401>

The [latest version of HTML 4](#) is available at: <http://www.w3.org/TR/html4>

[HTML40]

"[HTML 4.0 Specification](#)", W3C Recommendation, D. Raggett, A. Le Hors, I. Jacobs, eds., 18 December 1997, revised on 24 April 1998. Available at <http://www.w3.org/TR/1998/REC-html40-19980424>

[HTTP]

"[Hypertext Transfer Protocol -- HTTP/1.1](#)", RFC 2616, J. Gettys, R. Fielding, J. Mogul, H. Frystyk, L. Masinter, P. Leach and T. Berners-Lee, June 1999. Available at: <http://www.rfc-editor.org/rfc/rfc2616.txt>

[MathML2]

"[Mathematical Markup Language \(MathML\) Version 2.0](#)", W3C Recommendation, D. Carlisle, P. Ion, R. Miner, N. Poppelier, eds., 21 February 2001. Available at: <http://www.w3.org/TR/2001/REC-MathML2-20010221>

The [latest version](#) is available at: <http://www.w3.org/TR/MathML2>

[MIME]

"[Multipurpose Internet Mail Extensions \(MIME\) Part Two: Media Types](#)", RFC 2046, N. Freed, N. Borenstein, November 1996. Available at: <http://www.rfc-editor.org/rfc/rfc2046.txt>

[RFC2119]

"[Key words for use in RFCs to Indicate Requirement Levels](#)", RFC 2119, S. Bradner, March 1997. Available at: <http://www.rfc-editor.org/rfc/rfc2119.txt>

[RFC2854]

"[The 'text/html' Media Type](#)", RFC 2854, D. Connolly, L. Masinter, June 2000.

Available at: <http://www.rfc-editor.org/rfc/rfc2854.txt>

[RFC2912]

"[Indicating Media Features for MIME Content](#)", RFC 2912, G. Klyne, September 2000. Available at: <http://www.rfc-editor.org/rfc/rfc2912.txt>

[RFC3023]

"[XML Media Types](#)", RFC3023, M. Murata, S. St.Laurent, D. Kohn, January 2001. Available at: <http://www.rfc-editor.org/rfc/rfc3023.txt>

[RFC3236]

"[The 'application/xhtml+xml' Media Type](#)", RFC 3236, M. Baker, P. Stark, January 2002. Available at: <http://www.rfc-editor.org/rfc/rfc3236.txt>

[XHTML1]

"[XHTML™ 1.0 The Extensible HyperText Markup Language \(Second Edition\): A Reformulation of HTML 4 in XML 1.0](#)", W3C Recommendation, S. Pemberton et al., August 2002. Available at: <http://www.w3.org/TR/2002/REC-xhtml1-20020801>
The [first edition](#) is available at: <http://www.w3.org/TR/2000/REC-xhtml1-20000126>
The [latest version](#) is available at: <http://www.w3.org/TR/xhtml1>

[XHTML11]

"[XHTML™ 1.1 - Module-based XHTML](#)", W3C Recommendation, M. Altheim, S. McCarron, eds., 31 May 2001. Available at: <http://www.w3.org/TR/2001/REC-xhtml11-20010531>
The [latest version](#) is available at: <http://www.w3.org/TR/xhtml11>

[XHTMLBasic]

"[XHTML™ Basic](#)", W3C Recommendation, M. Baker, M. Ishikawa, S. Matsui, P. Stark, T. Wugofski, T. Yamakami, eds., 19 December 2000. Available at: <http://www.w3.org/TR/2000/REC-xhtml-basic-20001219>
The [latest version](#) is available at: <http://www.w3.org/TR/xhtml1-basic>

[XHTMLM12N]

"[Modularization of XHTML™](#)", W3C Recommendation, M. Altheim, F. Boumphrey, S. Dooley, S. McCarron, S. Schnitzenbaumer, T. Wugofski, eds., 10 April 2001. Available at: <http://www.w3.org/TR/2001/REC-xhtml-modularization-20010410>
The [latest version](#) is at: <http://www.w3.org/TR/xhtml1-modularization>

[XHTML+MathML]

"[XHTML plus Math 1.1 DTD](#)", "A.2 MathML as a DTD Module", Mathematical Markup Language (MathML) Version 2.0. Available at: <http://www.w3.org/TR/MathML2/dtd/xhtml1-math11-f.dtd>

[XML10]

"[Extensible Markup Language \(XML\) 1.0 Specification \(Second Edition\)](#)", T. Bray, J. Paoli, C. M. Sperberg-McQueen, E. Maler, eds., 6 October 2000. Available at: <http://www.w3.org/TR/2000/REC-xml-20001006>
The [latest version](#) is available at: <http://www.w3.org/TR/REC-xml>

[XMLNS]

"[Namespaces in XML](#)", T. Bray, D. Hollander, A. Layman, eds., 14 January 1999. Available at: <http://www.w3.org/TR/1999/REC-xml-names-19990114>
The [latest version](#) is available at: <http://www.w3.org/TR/REC-xml-names>

[XMLstyle]

"[Associating Style Sheets with XML documents Version 1.0](#)", W3C Recommendation, J. Clark, ed., 29 June 1999. Available at: <http://www.w3.org/1999/06/REC-xml-stylesheet-19990629>
The [latest version](#) is available at: <http://www.w3.org/TR/xml-stylesheet>

Changes from Previous Version

In [3.5. Summary](#), changed 'text/html' for HTML 4 as **SHOULD** rather than **MAY**.

Updated [reference to XHTML 1.0](#) to refer to the Second Edition.