

# Format Background Document: XHTML 1.0

**Date:** July 12, 2005

**Author:** Grace Carpenter

## GDFR data

**Canonical identifier** info:gdfr/fred/f/xhtmll

**Description** Extensible HyperText Markup Language (XHTML 1.0)

### **Alias**

**Type** MIME media type

**Value** application/xhtml+xml

**Note** An xhtml document may be associated with one of four MIME types; 'application/xhtml+xml' is the type recommended by the W3C, but it may cause browser problems

### **Alias**

**Type** MIME media type

**Value** text/html

**Note** W3C states that "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 XHTML 1 documents which follow the HTML Compatibility Guidelines."

### **Alias**

**Type** MIME media type

**Value** application/xml

**Note** Although the W3C says "Any XHTML Family document may be served as 'application/xml'", it recommends that "Whenever appropriate, 'application/xhtml+xml' should be used rather than 'application/xml'".

### **Alias**

**Type** MIME media type

**Value** text/xml

**Note** W3C recommends 'application/xhtml+xml' over this more generic type.

**Version** 1.0

### **Legal or recognized owner**

**Name** World Wide Web Consortium

**Organization type** Non-profit entity

**Web site** <http://www.w3.org>

### **Relationship**

**Type of relation** Other (see notes)

#### **Target format identifier**

**Type** GDFR format identifier

**Value** info:gdfr/fred/f/html

**Note** W3C calls this a "reformulation" of HTML4 that "subsets" and "extends" HTML4

#### **Target registry identifier**

**Type** GDFR registry identifier

**Value** FRED

### **Relationship**

**Type of relation** Subtype of target

#### **Target format identifier**

**Type** GDFR format identifier

**Value** info:gdf/fred/f/xml  
**Target registry identifier**  
**Type** GDFR registry identifier  
**Value** FRED

#### Specification

**Document title** XHTML(TM) 1.0 The Extensible HyperText Markup Language  
**Document type** Article  
**Access regime** Unrestricted access  
**Identifier** <http://www.w3.org/TR/xhtml1/>

#### Signature

**Signature obligation** Optional  
**External signature type** File extension  
**Signature value** .xht

#### Signature

**Signature obligation** Optional  
**External signature type** File extension  
**Signature value** .xhtml

#### Signature

**Signature obligation** Optional  
**External signature type** File extension  
**Signature value** .html

#### Signature

**Signature obligation** Optional  
**Signature position** Fixed position (requires offset)  
**Byte offset** 0  
**Signature value** 0x3C 0x3F 0x78 0x6D 0x6C  
**Note** Most XML docs in ASCII-compatible char sets (including UTF-8) begin with the XML prolog “X?xml”, although it isn't strictly necessary.

#### Signature

**Signature obligation** Optional  
**Signature position** Fixed position (requires offset)  
**Byte offset** 1  
**Signature value** 0x003C 0x003F 0x0078 0x006D 0x006C  
**Note** This signature represents the XML prolog (“<?xml”) in a UTF-16 little-endian document. It is not strictly necessary.

#### Signature

**Signature obligation** Optional  
**Signature position** Fixed position (requires offset)  
**Byte offset** 1  
**Signature value** 0x3C00 0x3F00 0x7800 0x6D00 0x6C00  
**Note** This signature represents the XML prolog (“<?xml”) in a UTF-16 big-endian document. It is not strictly necessary.

### Non-GDFR Data

#### 1. General

**1.1 Description (long):** The W3C describes XHTML as “a reformulation of HTML 4 as an XML 1.0 application” [W3C 2000]. XHTML 1.0 is distinguished from HTML 4 primarily by its requirement of conformance to specification (put another way, documents must be well-formed in order to be considered XHTML), and by its emphasis on separation of style and content.

XHTML 1.0 has since been succeeded by XHTML 1.1, which “represents a departure from both HTML 4 and XHTML 1.0. Most significant is the removal of features that were deprecated. In general, the strategy is to define a markup language that is rich in structural functionality, but that relies upon style sheets for presentation.” [W3C 2001]) Because there is still a wealth of HTML 4.0 (and earlier) documents, I have chosen to focus on XHTML 1.0 in this document.

**1.2 Content type:** text

## **2. Category-specific**

### **General Technical**

**3.1 Encoding:** “XML is a text format usually encoded in Unicode. All ‘XML processors’ must accept the UTF-8 and UTF-16 encodings of ISO/IEC 10646. An XML document can be encoded in any character encoding as long as the character set is identified in the XML declaration (<?xml encoding=’EUC-JP’>), otherwise it is assumed to be UTF-8 or UTF-16.” [Goethals]

**3.2 Byte Order:** Platform-dependant

**3.3 Encryption:** .

**3.4 Human readable:** yes

## **4. Sustainability**

**4.1 Proprietary:** No. XHTML is an open and fully documented standard.

**4.2 Owner documentation:** In addition to the spec, the W3C provides one or two tutorials, but not much additional documentation for users/web designers.

**4.3 Other documentation:** There are a huge number of XHTML resources on the Web.

**4.4 Adoption:** XHTML has existed for several years, but the vast majority of material on the web is HTML, rather than XHTML. This is due partly to the fact that most browsers do not require XHTML, and the browser with the biggest market share--Internet Explorer—is not XHTML-aware. This means that although it is capable of rendering most XHTML, there are some potential problems. The issue most frequently mentioned is that IE doesn’t support the application/xhtml+xml MIME type, so instead of rendering pages with this MIME type, it displays the file download prompt. There seems to be a gradual awakening to the advantages of XHTML over HTML, such as consistency and adherence to standards, but most estimates of the amount of material on the web that is not XHTML compliant put it at over 90%.

**4.5 Competition:** There is no real alternative to XHTML, but the prevalence of HTML and the support of browsers for HTML allow web-page designers to continue to use HTML, even though it is basically an obsolete version of XHTML.

Although there are other markup languages (MathML, for instance), there is no alternative to HTML/XHTML for web page mark-up.

**4.6 Licensing and patent claims:**

**4.7 Other preservation issues:** A big issue in the preservation of XHTML—or any XML document—is the existence of external references. See [Goethals] for a discussion of this issue.

## **5. Lifecycle**

**5.1 Version Duration:** 1.5 years

**5.2 Version History:**

XHTML 1.0 Jan 2000

XHTML 1.1 May 2001

**5.3 Expected Newer Versions:**

XHTML 2.0 working draft Jul 2004

## **6. Local Use (in Dspace at MIT)**

## **7. References**

[Goethals] Goethals, Andrea, “Action Plan Background: XML 1.0”, FCLA; 9 June 2003  
[http://www.fcla.edu/digitalArchive/pdfs/action\\_plan\\_bgrounds/xml\\_1\\_0.pdf](http://www.fcla.edu/digitalArchive/pdfs/action_plan_bgrounds/xml_1_0.pdf)

[W3C 2000] W3C, “XHTML™ 1.0 The Extensible Hypertext Markup Language (Second Edition), A Reformulation of HTML 4 in XML 1.0” W3C Recommendation; 26 January 2000, revised 1 August 2002  
<http://www.w3.org/TR/xhtml1/>

[W3C 2001] W3C, “XHTML™ 1.1 – Module-based XHTML” W3C Recommendation; 31 May 2001”  
<http://www.w3.org/TR/xhtml11/>