

Network Working Group
Internet-Draft
Obsoletes: [3023](#) (if approved)
Updates: [6839](#) (if approved)
Intended status: Standards Track
Expires: April 19, 2014

H. S. Thompson
University of Edinburgh
C. Lilley
W3C
October 16, 2013

XML Media Types
draft-ietf-appsawg-xml-mediatypes-03

Abstract

This specification standardizes three media types -- application/xml, application/xml-external-parsed-entity, and application/xml-dtd -- for use in exchanging network entities that are related to the Extensible Markup Language (XML) while defining text/xml and text/xml-external-parsed-entity as aliases for the respective application/ types. This specification also standardizes the '+xml' suffix for naming media types outside of these five types when those media types represent XML MIME entities.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on April 19, 2014.

Copyright Notice

Copyright (c) 2013 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents

carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1.	Introduction	3
2.	Notational Conventions	3
3.	XML Media Types	4
3.1.	Application/xml Registration	5
3.2.	Text/xml Registration	7
3.3.	Application/xml-external-parsed-entity Registration	7
3.4.	Text/xml-external-parsed-entity Registration	8
3.5.	Application/xml-dtd Registration	8
3.6.	Charset considerations	9
4.	The Byte Order Mark (BOM) and Charset Conversions	10
5.	Fragment Identifiers	11
6.	The Base URI	12
7.	XML Versions	13
8.	A Naming Convention for XML-Based Media Types	13
8.1.	Referencing	14
8.2.	+xml Structured Syntax Suffix Registration	15
9.	Examples	16
9.1.	UTF-8 Charset	17
9.2.	UTF-16 Charset	17
9.3.	Omitted Charset and 8-bit MIME entity	17
9.4.	Omitted Charset and 16-bit MIME entity	18
9.5.	Omitted Charset, no Internal Encoding Declaration and UTF-8 Entity	18
9.6.	UTF-16BE Charset	19
9.7.	Non-UTF Charset	19
9.8.	Omitted Charset with Internal Encoding Declaration	19
9.9.	INCONSISTENT EXAMPLE: Conflicting Charset and Internal Encoding Declaration	20
9.10.	INCONSISTENT EXAMPLE: Conflicting Charset and BOM	20
10.	IANA Considerations	21
11.	Security Considerations	21
12.	References	22
12.1.	Normative References	22
12.2.	Informative References	24
Appendix A.	Why Use the '+xml' Suffix for XML-Based MIME Types?	26
Appendix B.	Changes from RFC 3023	26
Appendix C.	Acknowledgements	26
	Authors' Addresses	26

1. Introduction

The World Wide Web Consortium has issued the Extensible Markup Language (XML) 1.0 [[XML](#)] and Extensible Markup Language (XML) 1.1 [[XML1.1](#)] specifications. To enable the exchange of XML network entities, this specification standardizes three media types -- application/xml, application/xml-external-parsed-entity, and application/xml-dtd and two aliases -- text/xml and text/xml-external-parsed-entity, as well as a naming convention for identifying XML-based MIME media types (using '+xml').

XML has been used as a foundation for other media types, including types in every branch of the IETF media types tree. To facilitate the processing of such types, and in line with the recognition in [[RFC6838](#)] of structured syntax name suffixes, a suffix of '+xml' is described in [Section 8](#). This will allow generic XML-based tools -- browsers, editors, search engines, and other processors -- to work with all XML-based media types.

This specification replaces [[RFC3023](#)]. Major differences are in the areas of alignment of charset handling for text/xml and text/xml-external-parsed-entity with application/xml, the addition of XPointer and XML Base as fragment identifiers and base URIs, respectively, integration of the XPointer Registry and updating of many references.

2. Notational Conventions

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 [[RFC2119](#)].

As defined in [[RFC2781](#)] (informative), the three character sets "utf-16", "utf-16le", and "utf-16be" are used to label UTF-16 text. In this specification, "the UTF-16 family" refers to those three character sets. By contrast, the phrases "utf-16" or UTF-16 in this specification refer specifically to the single charset "utf-16".

As sometimes happens between two communities, both MIME and XML have defined the term entity, with different meanings. [Section 2.4 of](#) [\[RFC2045\]](#) says:

"The term 'entity' refers specifically to the MIME-defined header fields and contents of either a message or one of the parts in the body of a multipart entity."

Section 4 of [[XML](#)] says:

"An XML document may consist of one or many storage units. These are called entities; they all have content and are all (except for the document entity and the external DTD subset) identified by entity name".

In this specification, "XML MIME entity" is defined as the latter (an XML entity) encapsulated in the former (a MIME entity).

Furthermore, XML provides for the naming and referencing of entities for purposes of inclusion and/or substitution. In this specification "XML-entity declaration/reference/..." is used to avoid confusion when referring to such cases.

3. XML Media Types

Registration information for media types for use with XML MIME entities is described in the sections below. Within the XML specification, such entities can be classified into four types. In the XML terminology, they are called "document entities", "external DTD subsets", "external parsed entities", and "external parameter entities". Appropriate usage for the types registered below is as follows:

document entities The media types `application/xml` or `text/xml` MAY be used.

external DTD subsets The media type `application/xml-dtd` SHOULD be used. The media types `application/xml` and `text/xml` MUST NOT be used.

external parsed entities The media types `application/xml-external-parsed-entity` or `text/xml-external-parsed-entity` SHOULD be used. The media types `application/xml` and `text/xml` MUST NOT be used unless the parsed entities are also well-formed "document entities" and are referenced as such.

external parameter entities The media type `application/xml-dtd` SHOULD be used. The media types `application/xml` and `text/xml` MUST NOT be used.

Note that [[RFC3023](#)] (which this specification obsoletes) recommended the use of `text/xml` and `text/xml-external-parsed-entity` for document entities and external parsed entities, respectively, but described charset handling which differed from common implementation practice. These media types are still commonly used, and this specification aligns the charset handling with industry practice.

Note that [[RFC2376](#)] (which is obsolete) allowed application/xml and text/xml to be used for any of the four types, although in practice it is likely to have been rare.

Neither external DTD subsets nor external parameter entities parse as XML documents, and while some XML document entities may be used as external parsed entities and vice versa, there are many cases where the two are not interchangeable. XML also has unparsed entities, internal parsed entities, and internal parameter entities, but they are not XML MIME entities.

Compared to [[RFC2376](#)] or [[RFC3023](#)], this specification alters the charset handling of text/xml and text/xml-external-parsed-entity, treating them no differently from the respective application/ types, however application/xml and application/xml-external-parsed-entity are still RECOMMENDED, to avoid possible confusion based on the earlier distinction. The former confusion around the question of default character sets for the text/xml... types has been resolved by [[HTTPbis](#)] changing [[RFC2616](#)] by removing the ISO-8859-1 default and not defining any default at all, as well as [[RFC6657](#)] updating [[RFC2046](#)] to remove the US-ASCII default. See [Section 3.6](#) for the now-unified approach to the charset parameter which results.

XML provides a general framework for defining sequences of structured data. It is often appropriate to define new media types that use XML but define a specific application of XML, due to domain-specific display, editing, security considerations or runtime information. Furthermore, such media types may allow UTF-8 or UTF-16 only and prohibit other character sets. This specification does not prohibit such media types and in fact expects them to proliferate. However, developers of such media types are RECOMMENDED to use this specification as a basis for their registration. See [Section 8](#) for more detailed recommendations on using the '+xml' suffix for registration of such media types.

An XML document labeled as application/xml or text/xml, or with a '+xml' media type, might contain namespace declarations, stylesheet-linking processing instructions (PIs), schema information, or other declarations that might be used to suggest how the document is to be processed. For example, a document might have the XHTML namespace and a reference to a CSS stylesheet. Such a document might be handled by applications that would use this information to dispatch the document for appropriate processing.

[3.1. Application/xml Registration](#)

Type name: application

Subtype name: xml

Required parameters: none

Optional parameters: charset

See [Section 3.6](#).

Encoding considerations: Depending on the charset encoding used, XML MIME entities may consist of 7bit, 8bit or binary data [[RFC6838](#)]. For 7-bit transports, 7bit data, for example data with charset encoding US-ASCII, does not require content-transfer-encoding, but 8bit or binary data, for example data with charset encoding UTF-8 or UTF-16, MUST be content-transfer-encoded in quoted-printable or base64. For 8-bit clean transport (e.g. 8BITMIME [[RFC6152](#)], ESMTP or NNTP [[RFC3977](#)]), 7bit or 8bit data, for example data with charset encoding UTF-8 or US-ASCII, does not require content-transfer-encoding, but binary data, for example data with a charset encoding from the UTF-16 family, MUST be content-transfer-encoded in base64. For binary clean transports (e.g. HTTP [[RFC2616](#)]), no content-transfer-encoding is necessary (or even possible, in the case of HTTP) for 7bit, 8bit or binary data.

Security considerations: See [Section 11](#).

Interoperability considerations: XML has proven to be interoperable across both generic and task-specific applications and for import and export from multiple XML authoring and editing tools. For maximum interoperability, validating processors are recommended. Although non-validating processors may be more efficient, they are not required to handle all features of XML. For further information, see sub-[section 2.9](#) "Standalone Document Declaration" and [section 5](#) "Conformance" of [[XML](#)] .

Published specification: Extensible Markup Language (XML) 1.0 (Fifth Edition) [[XML](#)] or subsequent editions or versions thereof.

Applications that use this media type: XML is device-, platform-, and vendor-neutral and is supported by a wide range of generic XML tools (editors, parsers, Web agents, ...), generic and task-specific applications.

Additional information:

Magic number(s): None.

Although no byte sequences can be counted on to always be present, XML MIME entities in ASCII-compatible character sets

(including UTF-8) often begin with hexadecimal 3C 3F 78 6D 6C ("<?xml"), and those in UTF-16 often begin with hexadecimal FE FF 00 3C 00 3F 00 78 00 6D 00 6C or FF FE 3C 00 3F 00 78 00 6D 00 6C 00 (the Byte Order Mark (BOM) followed by "<?xml"). For more information, see [Appendix F](#) of [\[XML\]](#).

File extension(s): .xml

Macintosh File Type Code(s): "TEXT"

Base URI: See [Section 6](#)

Person and email address for further information: See Authors' Addresses section

Intended usage: COMMON

Author: See Authors' Addresses section

Change controller: The XML specification is a work product of the World Wide Web Consortium's XML Working Group

[3.2.](#) Text/xml Registration

text/xml is an alias for application/xml, as defined in [Section 3.1](#) above.

[3.3.](#) Application/xml-external-parsed-entity Registration

Type name: application

Subtype name: xml-external-parsed-entity

Required parameters: none

Optional parameters: charset

See [Section 3.6](#).

Encoding considerations: Same as application/xml as described in [Section 3.1](#).

Security considerations: See [Section 11](#).

Interoperability considerations: XML external parsed entities are as interoperable as XML documents, though they have a less tightly constrained structure and therefore need to be referenced by XML documents for proper handling by XML processors. Similarly, XML

documents cannot be reliably used as external parsed entities because external parsed entities are prohibited from having standalone document declarations or DTDs. Identifying XML external parsed entities with their own content type should enhance interoperability of both XML documents and XML external parsed entities.

Published specification: Same as application/xml as described in [Section 3.1](#).

Applications which use this media type: Same as application/xml as described in [Section 3.1](#).

Additional information:

Magic number(s): Same as application/xml as described in [Section 3.1](#).

File extension(s): .xml or .ent

Macintosh File Type Code(s): "TEXT"

Base URI: See [Section 6](#)

Person and email address for further information: See Authors' Addresses section.

Intended usage: COMMON

Author: See Authors' Addresses section.

Change controller: The XML specification is a work product of the World Wide Web Consortium's XML Working Group

[3.4.](#) Text/xml-external-parsed-entity Registration

text/xml-external-parsed-entity is an alias for application/xml-external-parsed-entity, as defined in [Section 3.3](#) above.

[3.5.](#) Application/xml-dtd Registration

Type name: application

Subtype name: xml-dtd

Required parameters: none

Optional parameters: charset

See [Section 3.6](#).

Encoding considerations: Same as [Section 3.1](#).

Security considerations: See [Section 11](#).

Interoperability considerations: XML DTDs have proven to be interoperable by DTD authoring tools and XML validators, among others.

Published specification: Same as application/xml as described in [Section 3.1](#).

Applications which use this media type: DTD authoring tools handle external DTD subsets as well as external parameter entities. XML validators may also access external DTD subsets and external parameter entities.

Additional information:

Magic number(s): Same as application/xml as described in [Section 3.1](#).

File extension(s): .dtd or .mod

Macintosh File Type Code(s): "TEXT"

Person and email address for further information: See Authors' Addresses section.

Intended usage: COMMON

Author: See Authors' Addresses section.

Change controller: The XML specification is a work product of the World Wide Web Consortium's XML Working Group

[3.6](#). Charset considerations

When a charset parameter is specified for an XML MIME entity which contains in-band encoding information, that is, either a BOM ([Section 4](#)) or an XML encoding declaration or both, the normative component of the [\[XML\]](#) specification leaves the question open as to which should be taken to be authoritative in the case of conflict. In its (non-normative) [Appendix F](#) it defers to this specification:

[T]he preferred method of handling conflict should be specified as part of the higher-level protocol used to deliver XML. In particular, please refer to [IETF [RFC 3023](#)] or its successor

All processors SHOULD treat a BOM ([Section 4](#)) as authoritative if it is present in an XML MIME entity. In the absence of a BOM ([Section 4](#)), all processors SHOULD treat the charset parameter as authoritative. [Section 4.3.3](#) of the [XML] specification does *not* make it an error for the charset parameter and the XML encoding declaration to be inconsistent.

XML-aware processors SHOULD supply a charset parameter and/or an appropriate BOM with non-UTF-8-encoded XML MIME entities which lack an encoding declaration, or whose encoding declaration is known to be incorrect (for example, as a result of transcoding).

The charset parameter MUST NOT be used unless the charset is reliably known. This information will be used by all processors to determine authoritatively the charset of the XML MIME entity in the absence of a BOM.

"utf-8" [[RFC3629](#)] and "utf-16" [[RFC2781](#)] are the recommended values, representing the UTF-8 and UTF-16 character sets, respectively. These character sets are preferred since they are supported by all conforming processors of [XML].

If an entity of one of the types defined above is received where the charset parameter is omitted, no information is being provided about the charset by the MIME Content-Type header. Conforming XML processors MUST follow the requirements in section 4.3.3 of [XML] that directly address this contingency. MIME processors that are not XML processors SHOULD NOT assume a default charset if the charset parameter is omitted from such an entity.

4. The Byte Order Mark (BOM) and Charset Conversions

Section 4.3.3 of [XML] specifies that XML MIME entities in the charset "utf-16" MUST begin with a byte order mark (BOM), which is a hexadecimal octet sequence 0xFE 0xFF (or 0xFF 0xFE, depending on endian). The XML Recommendation further states that the BOM is an encoding signature, and is not part of either the markup or the character data of the XML document.

Due to the presence of the BOM, applications that convert XML from "utf-16" to an encoding other than "utf-8" MUST strip the BOM before conversion. Similarly, when converting from another encoding into "utf-16", the BOM MUST be added after conversion is complete unless the original encoding was "utf-8" and a BOM was already present, in which case it will have been transcoded into a "utf-16" BOM already.

Section 4.3.3 of [\[XML\]](#) also allows for XML MIME entities in the charset "utf-8" to begin with a byte order mark (BOM), which is a hexadecimal octet sequence 0xEF 0xBB 0xBF, also defined to be an encoding signature, and not part of either the markup or the character data of the XML document.

Applications that convert XML from "utf-8" to an encoding other than "utf-16" MUST strip the BOM, if present, before conversion. Applications which convert XML into "utf-8" SHOULD add a BOM after conversion is complete.

In addition to the charset "utf-16", [\[RFC2781\]](#) introduces "utf-16le" (little endian) and "utf-16be" (big endian) as well. The BOM is prohibited for these character sets. When an XML MIME entity is encoded in "utf-16le" or "utf-16be", it MUST NOT begin with the BOM but SHOULD contain an in-band XML encoding declaration. Conversion from "utf-16" or "utf-8" to "utf-16be" or "utf-16le" and conversion in the other direction MUST strip or add the appropriate BOM, respectively.

5. Fragment Identifiers

Uniform Resource Identifiers (URIs) can contain fragment identifiers (see [Section 3.5 of \[RFC3986\]](#)). Specifying the syntax and semantics of fragment identifiers is devolved by [\[RFC3986\]](#) to the appropriate media type registration.

The syntax and semantics of fragment identifiers for the XML media types defined in this specification are based on the [\[XPointerFramework\]](#) W3C Recommendation. It allows simple names, and more complex constructions based on named schemes. When the syntax of a fragment identifier part of any URI or IRI with a retrieved media type governed by this specification conforms to the syntax specified in [\[XPointerFramework\]](#), conforming applications MUST interpret such fragment identifiers as designating that part of the retrieved representation specified by [\[XPointerFramework\]](#) and whatever other specifications define any XPointer schemes used. Conforming applications MUST support the 'element' scheme as defined in [\[XPointerElement\]](#), but need not support other schemes.

If an XPointer error is reported in the attempt to process the part, this specification does not define an interpretation for the part.

A registry of XPointer schemes [[XPtrReg](#)] is maintained at the W3C. Document authors SHOULD NOT use unregistered schemes. Scheme authors SHOULD register their schemes ([[XPtrRegPolicy](#)] describes requirements and procedures for doing so).

See [Section 8.1](#) for additional requirements which apply when an XML-based media type follows the naming convention '+xml'.

If [[XPointerFramework](#)] and [[XPointerElement](#)] are inappropriate for some XML-based media type, it SHOULD NOT follow the naming convention '+xml'.

When a URI has a fragment identifier, it is encoded by a limited subset of the repertoire of US-ASCII [[ASCII](#)] characters, as defined in [[RFC3986](#)].

6. The Base URI

[Section 5.1 of \[RFC3986\]](#) specifies that the semantics of a relative URI reference embedded in a MIME entity is dependent on the base URI. The base URI is established by (1) the base URI embedded in content, (2) the base URI from the encapsulating entity, (3) the base URI from the Retrieval URI, or (4) the default base URI, in order of precedence. [[RFC3986](#)] further specifies that the mechanism for embedding the base URI is dependent on the media type.

This specification accordingly provides the following media type dependent mechanism for embedding the base URI in a MIME entity of type application/xml, text/xml, application/xml-external-parsed-entity or text/xml-external-parsed-entity: An XML MIME entity MAY use the xml:base attribute, as described in detail in [[XMLBase](#)], to establish a base URI for that entity.

Note that the base URI itself might be embedded in a different MIME entity, since the default value for the xml:base attribute can be specified in an external DTD subset or external parameter entity. Since conforming XML processors need not always read and process external entities, the effect of such an external default is uncertain and therefore its use is NOT RECOMMENDED.

7. XML Versions

application/xml, application/xml-external-parsed-entity, and application/xml-dtd, text/xml and text/xml-external-parsed-entity are to be used with [\[XML\]](#). In all examples herein where version="1.0" is shown, it is understood that version="1.1" may also be used, providing the content does indeed conform to [\[XML1.1\]](#).

The normative requirement of this specification upon XML documents and processors is to follow the requirements of [\[XML\]](#), section 4.3.3. Except for minor clarifications, that section is substantially identical from the first edition to the current (5th) edition of XML 1.0, and for XML 1.1 1st or 2nd edition [\[XML1.1\]](#). Therefore, references herein to [\[XML\]](#) may be interpreted as referencing any existing version or edition of XML, or any subsequent edition or version which makes no incompatible changes to that section.

Specifications and recommendations based on or referring to this RFC SHOULD indicate any limitations on the particular versions or editions of XML to be used.

8. A Naming Convention for XML-Based Media Types

This section supersedes the earlier registration of the '+xml' suffix [\[RFC6839\]](#).

This specification recommends the use of a naming convention (a suffix of '+xml') for identifying XML-based media types, in line with the recognition in [\[RFC6838\]](#) of structured syntax name suffixes. This allows the use of generic XML processors and technologies on a wide variety of different XML document types at a minimum cost, using existing frameworks for media type registration.

When a new media type is introduced for an XML-based format, the name of the media type SHOULD end with '+xml' unless generic XML processing is in some way inappropriate for documents of the new type. This convention will allow applications that can process XML generically to detect that the MIME entity is supposed to be an XML document, verify this assumption by invoking some XML processor, and then process the XML document accordingly. Applications may match for types that represent XML MIME entities by comparing the subtype to the pattern '*/*+xml'. (However note that 4 of the 5 media types defined in this specification -- text/xml, application/xml, text/xml-external-parsed-entity, and application/xml-external-parsed-entity -- also represent XML MIME entities while not conforming to the '*/*+xml' pattern.)

NOTE: [Section 5.3.2](#) HTTPbis [[HTTPbis](#)] does not support Accept headers of the form "Accept: */*+xml" and so this header MUST NOT be used in this way.

Media types following the naming convention '+xml' SHOULD introduce the charset parameter for consistency, since XML-generic processing applies the same program for any such media type. However, there are some cases that the charset parameter need not be introduced. For example:

When an XML-based media type is restricted to UTF-8, it is not necessary to introduce the charset parameter. "UTF-8 only" is a generic principle and UTF-8 is the default of XML.

When an XML-based media type is restricted to UTF-8 and UTF-16, it might not be unreasonable to omit the charset parameter. Neither UTF-8 nor UTF-16 require in-band XML encoding declarations.

XML generic processing is not always appropriate for XML-based media types. For example, authors of some such media types may wish that the types remain entirely opaque except to applications that are specifically designed to deal with that media type. By NOT following the naming convention '+xml', such media types can avoid XML-generic processing. Since generic processing will be useful in many cases, however -- including in some situations that are difficult to predict ahead of time -- the '+xml' convention is to be preferred unless there is some particularly compelling reason not to.

The registration process for specific '+xml' media types is described in [[RFC6838](#)]. The registrar for the IETF tree will encourage new XML-based media type registrations in the IETF tree to follow this guideline. Registrars for other trees SHOULD follow this convention in order to ensure maximum interoperability of their XML-based documents. Similarly, media subtypes that do not represent XML MIME entities MUST NOT be allowed to register with a '+xml' suffix.

[8.1.](#) Referencing

Registrations for new XML-based media types under top-level types SHOULD, in specifying the charset parameter and encoding considerations, define them as: "Same as [charset parameter / encoding considerations] of application/xml as specified in RFC XXXX."

Enabling the charset parameter is RECOMMENDED, since this information can be used by XML processors to determine authoritatively the charset of the XML MIME entity in the absence of a BOM. If there are some reasons not to follow this advice, they SHOULD be included as

part of the registration. As shown above, two such reasons are "UTF-8 only" or "UTF-8 or UTF-16 only".

These registrations SHOULD specify that the XML-based media type being registered has all of the security considerations described in RFC XXXX plus any additional considerations specific to that media type.

These registrations SHOULD also make reference to RFC XXXX in specifying magic numbers, base URIs, and use of the BOM.

These registrations MAY reference the application/xml registration in RFC XXXX in specifying interoperability considerations, if these considerations are not overridden by issues specific to that media type.

8.2. +xml Structured Syntax Suffix Registration

Name: Extensible Markup Language (XML)

+suffix: +xml

Reference: This specification

Encoding considerations: Same as [Section 3.1](#).

Fragment identifier considerations: Registrations which use this '+xml' convention MUST also make reference to RFC XXXX, specifically [Section 5](#), in specifying fragment identifier syntax and semantics, and they MAY restrict the syntax to a specified subset of schemes, except that they MUST NOT disallow barenames or 'element' scheme pointers. They MAY further require support for other registered schemes. They also MAY add additional syntax (which MUST NOT overlap with [[XPointerFramework](#)] syntax) together with associated semantics, and MAY add additional semantics for barenamed XPointers which, as provided for in [Section 5](#), will only apply when this specification does not define an interpretation.

In practice these constraints imply that for a fragment identifier addressed to an instance of a specific "xxx/yyy+xml" type, there are three cases:

For fragment identifiers matching the syntax defined in [[XPointerFramework](#)], where the fragment identifier resolves per the rules specified there, then process as specified there;

For fragment identifiers matching the syntax defined in [[XPointerFramework](#)], where the fragment identifier does not resolve per the rules specified there, then process as specified in "xxx/yyy+xml";

For fragment identifiers not matching the syntax defined in [[XPointerFramework](#)], then process as specified in "xxx/yyy+xml". A fragment identifier of the form "xywh=160,120,320,240", as defined in [[MediaFrag](#)], which might be used in a URI for an XML-encoded image, would fall in this category.

Interoperability considerations: Same as [Section 3.1](#). See above, and also [Section 3.6](#), for guidelines on the use of the 'charset' parameter.

Security considerations: See [Section 11](#).

Contact: See Authors' Addresses section.

Author: See Authors' Addresses section.

Change controller: The XML specification is a work product of the World Wide Web Consortium's XML Working Group.

9. Examples

The examples below give the charset portion, if any, of the value of the MIME Content-type header and the XML declaration or Text declaration (which includes the encoding declaration) inside the XML MIME entity. For UTF-16 examples, the Byte Order Mark character appropriately UTF-16-encoded is denoted as "{BOM}", and the XML or Text declaration is assumed to come at the beginning of the XML MIME entity, immediately following the encoded BOM. Note that other MIME headers may be present, and the XML MIME entity may contain other data in addition to the XML declaration; the examples focus on the Content-type header and the encoding declaration for clarity.

All the examples below apply to all five media types declared above in [Section 3](#), as well as to any media types declared using the '+xml' convention (with the exception of the examples involving the charset parameter for any such media types which to not enable its use). See the XML MIME entities table ([Section 3](#), Paragraph 2) for discussion of which types are appropriate for which varieties of XML MIME entities.

This section is non-normative. In particular, note that all [[RFC2119](#)] language herein reproduces or summarizes the consequences

of normative statements already made above, and has no independent normative force, and accordingly does not appear in uppercase.

9.1. UTF-8 Charset

Content-type charset: charset="utf-8"

```
<?xml version="1.0" encoding="utf-8"?>
```

This is the recommended encoding for use with all the media types defined in this specification. Since the charset parameter is provided and there is no BOM, both MIME and XML processors must treat the enclosed entity as UTF-8 encoded.

If sent using a 7-bit transport (e.g. SMTP [[RFC5321](#)]), the XML MIME entity must use a content-transfer-encoding of either quoted-printable or base64. For an 8-bit clean transport (e.g. 8BITMIME, ESMTP or NNTP), or a binary clean transport (e.g. HTTP), no content-transfer-encoding is necessary (or even possible, in the case of HTTP).

9.2. UTF-16 Charset

Content-type charset: charset="utf-16"

```
{BOM}<?xml version="1.0" encoding="utf-16"?>
```

or

```
{BOM}<?xml version="1.0"?>
```

For application/... cases, if sent using a 7-bit transport (e.g. SMTP) or an 8-bit clean transport (e.g. 8BITMIME, ESMTP or NNTP), the XML MIME entity must be encoded in quoted-printable or base64; for a binary clean transport (e.g. HTTP), no content-transfer-encoding is necessary (or even possible, in the case of HTTP).

As described in [[RFC2781](#)], the UTF-16 family must not be used with media types under the top-level type "text" except over HTTP or HTTPS (see [section 19.4.2 of \[RFC2616\]](#) for details). Hence this example is only possible in text/... cases when the XML MIME entity is transmitted via HTTP or HTTPS, which use a MIME-like mechanism and are binary-clean protocols, hence do not perform CR and LF transformations and allow NUL octets. Since HTTP is binary clean, no content-transfer-encoding is necessary (or even possible).

9.3. Omitted Charset and 8-bit MIME entity

Content-type charset: [none]

```
<?xml version="1.0" encoding="iso-8859-1"?>
```

Since the charset parameter is not provided in the Content-Type header and there is no BOM, XML processors must treat the "iso-8859-1" encoding as authoritative. XML-unaware MIME processors should make no assumptions about the charset of the XML MIME entity.

9.4. Omitted Charset and 16-bit MIME entity

Content-type charset: [none]

```
{BOM}<?xml version="1.0" encoding="utf-16"?>
```

or

```
{BOM}<?xml version="1.0"?>
```

This example shows a 16-bit MIME entity with no charset parameter. However since there is a BOM all processors must treat the entity as UTF-16-encoded.

Omitting the charset parameter is not recommended for application/... when used with transports other than HTTP or HTTPS. text/... should not be used for 16-bit MIME with transports other than HTTP or HTTPS (see discussion above ([Section 9.2](#), Paragraph 6)).

9.5. Omitted Charset, no Internal Encoding Declaration and UTF-8 Entity

Content-type charset: [none]

```
<?xml version='1.0'?>
```

In this example, the charset parameter has been omitted, there is no internal encoding declaration, and there is no BOM. Since there is no BOM or charset parameter, the XML processor follows the requirements in [section 4.3.3](#), and optionally applies the mechanism described in [Appendix F](#) (which is non-normative) of [\[XML\]](#) to determine the charset encoding of UTF-8. Although the XML MIME entity does not contain an encoding declaration, the encoding actually is UTF-8, so this is still a conforming XML MIME entity.

An XML-unaware MIME processor should make no assumptions about the charset of the XML MIME entity.

See [Section 9.1](#) for transport-related issues for UTF-8 XML MIME entities.

9.6. UTF-16BE Charset

Content-type charset: charset="utf-16be"

```
<?xml version='1.0' encoding='utf-16be'?>
```

Observe that the BOM does not exist. Since the charset parameter is provided and there is no BOM, MIME and XML processors must treat the enclosed entity as UTF-16BE encoded.

See also the additional considerations in the UTF-16 example ([Section 9.2](#)) above.

9.7. Non-UTF Charset

Content-type charset: charset="iso-2022-kr"

```
<?xml version="1.0" encoding="iso-2022-kr"?>
```

This example shows the use of a non-UTF charset (in this case Hangul, but this example is intended to cover all non-UTF-family character sets). Since the charset parameter is provided and there is no BOM, all processors must treat the enclosed entity as encoded per [RFC 1557](#).

Since ISO-2022-KR [[RFC1557](#)] has been defined to use only 7 bits of data, no content-transfer-encoding is necessary with any transport: for character sets needing 8 or more bits, considerations such as those discussed above ([Section 9.1](#), [Section 9.2](#)) would apply.

9.8. Omitted Charset with Internal Encoding Declaration

Content-type charset: [none]

```
<?xml version='1.0' encoding="iso-10646-ucs-4"?>
```

In this example, the charset parameter has been omitted, and there is no BOM. However, the XML MIME entity does have an encoding declaration inside the XML MIME entity that specifies the entity's charset. Following the requirements in [section 4.3.3](#), and optionally applying the mechanism described in [Appendix F](#) (non-normative) of [[XML](#)], the XML processor determines the charset encoding of the XML MIME entity (in this example, UCS-4).

An XML-unaware MIME processor should make no assumptions about the charset of the XML MIME entity.

For character sets needing 8 or more bits, considerations such as those discussed above ([Section 9.1](#), [Section 9.2](#)) would apply

9.9. INCONSISTENT EXAMPLE: Conflicting Charset and Internal Encoding Declaration

Content-type charset: charset="iso-8859-1"

```
<?xml version="1.0" encoding="utf-8"?>
```

Although the charset parameter is provided in the Content-Type header and there is no BOM and the charset parameter differs from the XML encoding declaration, MIME and XML processors will interoperate. Since the charset parameter is authoritative in the absence of a BOM, all processors will treat the enclosed entity as iso-8859-1 encoded. That is, the "UTF-8" encoding declaration will be ignored.

Processors generating XML MIME entities must not label conflicting charset information between the MIME Content-Type and the XML declaration unless they have definitive information about the actual encoding, for example as a result of systematic transcoding. In particular, the addition by servers of an explicit, site-wide charset default has frequently lead to interoperability problems for XML documents.

9.10. INCONSISTENT EXAMPLE: Conflicting Charset and BOM

Content-type charset: charset="iso-8859-1"

```
{BOM}<?xml version="1.0"?>
```

Although the charset parameter is provided in the Content-Type header, there is a BOM, so MIME and XML processors may not interoperate. Since the BOM parameter is authoritative for XML processors, they will treat the enclosed entity as UTF-16-encoded. That is, the "iso-8859-1" charset parameter will be ignored. XML-unaware MIME processors on the other hand may be unaware of the BOM and so treat the entity as encoded in iso-8859-1.

Processors generating XML MIME entities must not label conflicting charset information between the MIME Content-Type and an entity-initial BOM.

10. IANA Considerations

As described in [Section 8](#), this specification updates the [\[RFC6839\]](#) registration for XML-based MIME types (the "+xml" types).

11. Security Considerations

XML MIME entities contain information which may be parsed and further processed by the recipient. These entities may contain, and recipients may permit, explicit system level commands to be executed while processing the data. To the extent that a recipient application executes arbitrary command strings from within XML MIME entities, they may be at risk.

In general, any information stored outside of the direct control of the user -- including CSS style sheets, XSL transformations, XML-entity declarations, and DTDs -- can be a source of insecurity, by either obvious or subtle means. For example, a tiny "whiteout attack" modification made to a "master" style sheet could make words in critical locations disappear in user documents, without directly modifying the user document or the stylesheet it references. Thus, the security of any XML document is vitally dependent on all of the documents recursively referenced by that document.

The XML-entity lists and DTDs for XHTML 1.0 [\[XHTML\]](#), for instance, are likely to be a commonly used set of information. Many developers will use and trust them, few of whom will know much about the level of security on the W3C's servers, or on any similarly trusted repository.

The simplest attack involves adding declarations that break validation. Adding extraneous declarations to a list of character XML-entities can effectively "break the contract" used by documents. A tiny change that produces a fatal error in a DTD could halt XML processing on a large scale. Extraneous declarations are fairly obvious, but more sophisticated tricks, like changing attributes from being optional to required, can be difficult to track down. Perhaps the most dangerous option available to attackers, when external DTD subsets or external parameter entities or other externally-specified defaulting is involved, is redefining default values for attributes: e.g. if developers have relied on defaulted attributes for security, a relatively small change might expose enormous quantities of information.

Apart from the structural possibilities, another option, "XML-entity spoofing," can be used to insert text into documents, vandalizing and perhaps conveying an unintended message. Because XML permits multiple XML-entity declarations, and the first declaration takes

precedence, it is possible to insert malicious content where an XML-entity reference is used, such as by inserting the full text of Winnie the Pooh in place of every occurrence of —.

Security considerations will vary by domain of use. For example, XML medical records will have much more stringent privacy and security considerations than XML library metadata. Similarly, use of XML as a parameter marshalling syntax necessitates a case by case security review.

XML may also have some of the same security concerns as plain text. Like plain text, XML can contain escape sequences that, when displayed, have the potential to change the display processor environment in ways that adversely affect subsequent operations. Possible effects include, but are not limited to, locking the keyboard, changing display parameters so subsequent displayed text is unreadable, or even changing display parameters to deliberately obscure or distort subsequent displayed material so that its meaning is lost or altered. Display processors SHOULD either filter such material from displayed text or else make sure to reset all important settings after a given display operation is complete.

Some terminal devices have keys whose output, when pressed, can be changed by sending the display processor a character sequence. If this is possible the display of a text object containing such character sequences could reprogram keys to perform some illicit or dangerous action when the key is subsequently pressed by the user. In some cases not only can keys be programmed, they can be triggered remotely, making it possible for a text display operation to directly perform some unwanted action. As such, the ability to program keys SHOULD be blocked either by filtering or by disabling the ability to program keys entirely.

Note that it is also possible to construct XML documents that make use of what XML terms "[XML-]entity references" to construct repeated expansions of text. Recursive expansions are prohibited by [\[XML\]](#) and XML processors are required to detect them. However, even non-recursive expansions may cause problems with the finite computing resources of computers, if they are performed many times. (XML-entity A consists of 100 copies of XML-entity B, which in turn consists of 100 copies of XML-entity C, and so on)

[12.](#) References

[12.1.](#) Normative References

- [RFC2045] Freed, N. and N. Borenstein, "Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies", [RFC 2045](#), November 1996.
- [RFC2046] Freed, N. and N. Borenstein, "Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types", [RFC 2046](#), November 1996.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [RFC2616] Fielding, R., Gettys, J., Mogul, J., Nielsen, H., Masinter, L., Leach, P., and T. Berners-Lee, "Hypertext Transfer Protocol -- HTTP/1.1", [RFC 2616](#), June 1999.
- [RFC2781] Hoffman, P. and F. Yergeau, "UTF-16, an encoding of ISO 10646", [RFC 2781](#), February 2000.
- [RFC3986] Berners-Lee, T., Fielding, R., and L. Masinter, "Uniform Resource Identifiers (URI): Generic Syntax.", [RFC 3986](#), January 2005.
- [RFC3987] Dueerst, M. and M. Suignard, "Internationalized Resource Identifiers (IRIs)", [RFC 3987](#), July 2005.
- [RFC6657] Melnikov, A. and J. Reschke, "Update to MIME regarding "charset" Parameter Handling in Textual Media Types", [RFC 6657](#), July 2012,
<<http://www.rfc-editor.org/rfc/rfc6657.txt>>.
- [RFC6838] Freed, N., Klensin, J., and T. Hansen, "Media Type Specifications and Registration Procedures", [BCP 13](#), [RFC 6838](#), January 2013.
- [RFC6839] Hansen, T. and A. Melnikov, "Additional Media Type Structured Syntax Suffixes", [RFC 6839](#), January 2013.
- [XML1.1] Bray, T., Paoli, J., Sperberg-McQueen, C.M., Maler, E., Yergeau, F., and J. Cowan, "Extensible Markup Language (XML) 1.1 (Second Edition)", W3C Recommendation REC-xml, September 2006,
<<http://www.w3.org/TR/2006/REC-xml11-20060816/>>.
- Latest version available at
- [XMLBase] Marsh, J. and R. Tobin, "XML Base (Second Edition)", W3C Recommendation REC-xmlbase-20090128, January 2009,
<<http://www.w3.org/TR/2009/REC-xmlbase-20090128/>>.

Latest version available at

- [XML] Bray, T., Paoli, J., Sperberg-McQueen, C.M., Maler, E., and F. Yergeau, "Extensible Markup Language (XML) 1.0 (Fifth Edition)", W3C Recommendation REC-xml, November 2008, <<http://www.w3.org/TR/2008/REC-xml-20081126/>>.

Latest version available at

- [XPointerElement] Grosso, P., Maler, E., Marsh, J., and N. Walsh, "XPointer element() Scheme", W3C Recommendation REC-XPointer-Element, March 2003, <<http://www.w3.org/TR/2003/REC-xptr-element-20030325/>>.

Latest version available at

- [XPointerFramework] Grosso, P., Maler, E., Marsh, J., and N. Walsh, "XPointer Framework", W3C Recommendation REC-XPointer-Framework, March 2003, <<http://www.w3.org/TR/2003/REC-xptr-framework-20030325/>>.

Latest version available at

- [XPtrRegPolicy] Hazael-Massieux, D., "XPointer Scheme Name Registry Policy", 2005, <<http://www.w3.org/2005/04/xpointer-policy.html>>.

- [XPtrReg] Hazael-Massieux, D., "XPointer Registry", 2005, <<http://www.w3.org/2005/04/xpointer-schemes/>>.

12.2. Informative References

- [ASCII] American National Standards Institute, "Coded Character Set -- 7-bit American Standard Code for Information Interchange", ANSI X3.4, 1986.
- [HTTPbis] Fielding, R., "Hypertext Transfer Protocol (HTTP/1.1) [revised]", ietf-httpbis-p2-semantic (work in progress), September 2013.
- [ISO8859] ISO, "ISO-8859. International Standard -- Information Processing -- 8-bit Single-Byte Coded Graphic Character Sets -- Part 1: Latin alphabet No. 1, ISO-8859-1:1987", 1987.

[MediaFragments]

Troncy, R., Mannens, E., Pfeiffer, S., and D. Van Deursen, "Media Fragments URI 1.0 (basic)", W3C Recommendation media-frag, September 2012, <<http://www.w3.org/TR/2012/REC-media-frag-20120925/>>.

Latest version available at

- [RFC1557] Choi, U., Chon, K., and H. Park, "Korean Character Encoding for Internet Messages", [RFC 1557](#), December 1993.
- [RFC2130] Weider, C., Cecilia Preston, C., Simonsen, K., Alvestrand, H., Atkinson, R., Crispin, M., and P. Svanberg, "The Report of the IAB Character Set Workshop held 29 February - 1 March, 1996", [RFC 2130](#), April 1997.
- [RFC2376] Whitehead, E. and M. Murata, "XML Media Types", [RFC 2376](#), July 1998.
- [RFC2703] Klyne, G., "Protocol-independent Content Negotiation Framework", [RFC 2703](#), September 1999.
- [RFC3023] Murata, M., St.Laurent, S., and D. Kohn, "XML Media Types", [RFC 3023](#), January 2001.
- [RFC3629] Yergeau, F., "UTF-8, a transformation format of ISO 10646", [RFC 3629](#), November 2003.
- [RFC3977] Feather, B., "Network News Transfer Protocol", [RFC 3977](#), October 2006.
- [RFC5321] Klensin, J., "Simple Mail Transfer Protocol", [RFC 5321](#), October 2008.
- [RFC6152] Klensin, J., Freed, N., Rose, M., and D. Crocker, "SMTP Service Extension for 8-bit MIME Transport", [RFC 6152](#), March 2011.
- [TAGMIME] Bray, T., Ed., "Internet Media Type registration, consistency of use", April 2004, <<http://www.w3.org/2001/tag/2004/0430-mime>>.
- [XHTML] Pemberton, S. and et al, "XHTML 1.0: The Extensible HyperText Markup Language", W3C Recommendation xhtml1, December 1999, <<http://www.w3.org/TR/2000/REC-xhtml1-20000126/>>.

Latest version available at

[Appendix A.](#) Why Use the '+xml' Suffix for XML-Based MIME Types?

[RFC3023] contains a detailed discussion of the (at the time) novel use of a suffix, a practice which has since become widespread. Interested parties are referred to [\[RFC3023\]](#), [Appendix A](#).

[Appendix B.](#) Changes from [RFC 3023](#)

There are numerous and significant differences between this specification and [\[RFC3023\]](#), which it obsoletes. This appendix summarizes the major differences only.

First, XPointer ([\[XPointerFramework\]](#) and [\[XPointerElement\]](#)) has been added as fragment identifier syntax for "application/xml", and the XPointer Registry ([\[XPtrReg\]](#)) mentioned. Second, [\[XMLBase\]](#) has been added as a mechanism for specifying base URIs. Third, the language regarding character sets was updated to correspond to the W3C TAG finding Internet Media Type registration, consistency of use [\[TAGMIME\]](#). Fourth, many references are updated, and the existence of and relevance of the spec. to XML 1.1 acknowledged. Finally, a number of justifications and contextualizations which were appropriate when XML was new have been removed, including the whole of the original [Appendix A](#).

[Appendix C.](#) Acknowledgements

MURATA Makoto (FAMILY Given) and Alexey Melnikov made early and important contributions to the effort to revise [\[RFC3023\]](#).

This specification reflects the input of numerous participants to the ietf-xml-mime@imc.org, xml-mime@ietf.org and apps-discuss@ietf.org mailing lists, though any errors are the responsibility of the authors. Special thanks to:

Mark Baker, James Clark, Dan Connolly, Martin Duerst, Ned Freed, Yaron Goland, Bjoern Hoehrmann, Rick Jelliffe, Murray S. Kucherawy, Larry Masinter, David Megginson, S. Moonesamy, Keith Moore, Chris Newman, Gavin Nicol, Julian Reschke, Marshall Rose, Jim Whitehead, Erik Wilde and participants of the XML activity and the TAG at the W3C.

Jim Whitehead and Simon St.Laurent were editors of [\[RFC2376\]](#) and [\[RFC3023\]](#), respectively.

Authors' Addresses

Henry S. Thompson
University of Edinburgh

Email: ht@inf.ed.ac.uk

URI: <http://www.ltg.ed.ac.uk/~ht/>

Chris Lilley
World Wide Web Consortium
2004, Route des Lucioles - B.P. 93 06902
Sophia Antipolis Cedex
France

Email: chris@w3.org

URI: <http://www.w3.org/People/chris/>