Citation Elements:
Bindings for GEDCOM X

19 May 2017

> *Editorial note —* This is an **exploratory draft** of a standard documenting the proposed usage of the FHISO Citation Elements standard in GEDCOM X. This document is not a FHISO standard and is not endorsed by the FHISO membership. It may be updated, replaced or obsoleted by other documents at any time.
>
> In particular, some examples in this draft use *citation elements* that are not even included in the draft Citation Element Vocabulary. These elements are very likely to be changed as the vocabulary progresses.

FHISO's suite of **Citation Elements** standard provides an extensible framework and vocabulary for encoding all the data about a genealogical *source* that might reasonably be included in a *formatted citation* to that *source*.

This information is represented as a sequence of *citation elements*, logically self-contained pieces of information about a *source*. This document defines an extension to the GEDCOM X which allows *citation elements* has be represented in the GEDCOM X data model and its JSON and XML serialisations; it also recommends the use of RDFa attributes in HTML as a way of tagging *citation elements* in any *formatted citations* that may be present in a GEDCOM X dataset in order that they may be maintained more easily.

Other documents in the suite of Citation Elements standards are as follows:

— **Citation Elements: General Concepts**. This standard defines the general concepts used in FHISO's suite of Citation Elements standards, and the basic framework and data model underpinning them.
— **Citation Elements: Vocabulary**. This standard defines a collection of *citation elements* allowing the representation of information normally found in *formatted citations* to diverse types of source.
— **Citation Elements: Bindings for RDFa**. This standard defines a means by which *citation elements* may be identified and tagged using RDFa attributes within HTML and XML *formatted citations*, allowing a computer to extract them in a systematic manner.
— **Citation Elements: Bindings for ELF**. This standard defines how *citation elements* should be represented in FHISO's Extensible Legacy Format (ELF), a format based on and compatible with GEDCOM 5.5.1, but with the addition of a new extensibility mechanism.

> *Editorial note —* Not all of these documents are yet at the stage of having a first public draft.

1 Introduction

1.1 Conventions used

Where this standard gives a specific technical meaning to a word or phrase, that word or phrase is formatted in bold text in its initial definition, and in italics when used elsewhere. The key words MUST, MUST NOT, REQUIRED, SHALL, SHALL NOT, SHOULD, SHOULD NOT, RECOMMENDED, NOT RECOMMENDED, MAY and OPTIONAL in this standard are to be interpreted as described in [RFC 2119].

An application is **conformant** with this standard if and only if it follows all the requirements and prohibitions contained in this document, as indicated by use of the words MUST, MUST NOT, REQUIRED, SHALL and SHALL NOT, and the relevant parts of its normative references. Standards referencing this standard MUST NOT loosen any of the requirements and prohibitions made by this standard, nor place additional requirements or prohibitions on the constructs defined herein.

> *Note* — Adding requirements or prohibitions is disallowed so as to preserve interoperability between applications: data generated by one *conformant* application must always be acceptable to another *conformant* application, regardless of what additional standards each may conform to.

This standard depends on the *Citation Elements: General Concepts* standard [CEV Concepts]. To be *conformant* with this standard, an application MUST also be conformant with [CEV Concepts]. Some words and phrases defined in that standard are used here without further definition.

> *Note* — Readers are advised to read at least the introduction to [CEV Concepts] before reading this standard.

Indented text in coloured boxes, such as preceding paragraph, does not form a normative part of this standard, and is labelled as either an example or a note.

> *Editorial note* — Editorial notes, such as this, are used to record outstanding issues, or points where there is not yet consensus; they will be resolved and removed for the final standard. Examples and notes will be retained in the standard.

For the purpose of exposition in this document, the following XML namespace prefix bindings are assumed.

```
gx:     http://gedcomx.org/v1/
cev:    http://terms.fhiso.org/sources/
html:   http://www.w3.org/1999/xhtml
xsd:    http://www.w3.org/2001/XMLSchema
```

> *Note* — No restriction is placed on the choice of namespace prefixes used in actual data conforming to this standard, nor do these represent the recommended choice of prefix.

1.2 GEDCOM X

GEDCOM X is defined as an abstract data model containing **data types** each of which have various **properties**. This terminology is defined in §1.3.1 of GEDCOM X. Additional *properties* which extend the functionality of a *data type* and which are not defined in the GEDCOM X specification are known as **extension properties**. As permitted in §5.3 of GEDCOM X, this standard defines one new *data types*, the `SourceCitation` *data type*, and adds an *extension properties* to the `CitationElement` *data types*

The GEDCOM X specification suite defines serialisations of the abstract GEDCOM X data model as XML and JSON. This standard defines how the *data types* and *extension properties* defined here are serialised in these formats, with examples of their use.

> *Editorial note —* At present none of the GEDCOM X standards have been formally released as standards. They are described as "stable drafts" and "may be subject to limited changes, but not backwards-incompatible changes".
>
> Liaison will be needed with the GEDCOM X team to determine whether it is sensible for FHISO to standardise a set of GEDCOM X bindings while GEDCOM X has not yet reached "standard" status.

2 The `CitationElement` Data Type

This standard defines a new `CitationElement` *data type* which is identified by the following IRI:

> `http://terms.fhiso.org/sources/CitationElement`

It represents a *citation element* and has four properties, the detailed semantics of which are defined in the Citation Elements standard.

Property	Description, data type and constraints
layer	an OPTIONAL *layer identifier* which is a `string`.
name	a REQUIRED *citation element name* in the form of an IRI per [RFC 3987].
lang	an OPTIONAL ISO *language tag* per [RFC 5646].
value	a REQUIRED *citation element value* which is a `string` containing plain text.

The form of the data in the *citation element*'s value *property* may be further constrained in a manner dependent on the particular value of the name *property*.

> *Example —* The `http://terms.fhiso.org/sources/publicationDate` *citation element* is defined in the Citation Elements standard. When a `CitationElement`'s name *property* is equal to this IRI, the value *property* MUST contain a date in the prescribed date format based on [ISO 8601].

Editorial note — It would be ideal if GEDCOM X were to change their draft spec to use IRIs instead of URIs. If they will not, and if the use of an IRI in the `name` property will impede its adoption in GEDCOM X, the `name` property can be defined to contain an URI by requiring the use of the the algorithm given in §3.1 of [RFC 3987] to convert IRIs to URIs, and §3.2 for the reverse mapping. In the vast majority of cases, including all those defined in the Citation Elements standard, IRIs will already be valid URIs and translation will be a no-op.

The conversion of an IRI to a URI and back again does not necessarily result in the original IRI, but the Citation Elements standard prohibits the use of IRIs which do not have this property.

The `xsd:anyURI` type which is used extensively in [GEDCOM X XML] refers to a URI in XML Schema 1.0 but is generalised to an IRI in XML Schema 1.1, despite the type still being called `xsd:anyURI`.

Editorial note — The `layer` and `value` properties are defined as a `string` which is inconsistently defined in GEDCOM X. §1.3.3 of GEDCOM X defines a **string** as "a finite-length sequence of *characters*", with a **character** being "an atomic unit of text as specified by ISO/IEC 10646". [ISO 10646] does not define a "unit of text", so presumably this refers to its definition of a *character* in §4.5 which the null character and all the C0 escape characteers.

However the [GEDCOM X XML] standard routinely serialises `string` as an `xsd:string`, for which GEDCOM X cites XML Schema 1.0. XML Schema defines an `xsd:string` as a sequence of characters matching the XML `Char` production, but XML Schema 1.0 cites XML 1.0's definition and XML Schema 1.1 cites XML 1.1's, and the two XML standards' defintiions differ. XML 1.1 allows all characters except the null character, where XML 1.0 also disallows all C0 escape characters except tab (U+0009), line feed (U+000A) and carriage return (U+000D). This means that GEDCOM X allows strings that cannot be serialised in XML.

GEDCOM X would benefit from clarifying precisely which of these characters are allowed in its string type. The W3C has gradually been updating standards to use the 1.1 definitions, and FHISO has followed this precedent. If GEDCOM X wishes to do the same, it should explicitly reference the XML Schema 1.1 standard (or say the most recent one) when referencing `xsd:string`, and should explicitly exclude the null character from the definition of a *character* in §1.3.3 of GEDCOM X.

There is no similar incompatibility with JSON strings. A string in JSON may contain arbitrary characters, including the null character if suitably escaped, so there are no GEDCOM X strings that cannot be serialised in JSON.

2.1 JSON serialisation

The `CitationElement` *data type* is serialised in JSON as an object, with each *property* being represented by a JSON member with the same name.

> *Example* — A *citation element* containing the title of Christian Settipani's book *Les ancêtres de Charlemagne* would be serialised in JSON as follows:
>
> ```
> { "name":  "http://terms.fhiso.org/sources/title",
>   "value": "Les ancêtres de Charlemagne",
>   "lang":  "fr" }
> ```
>
> The IRI in the `name` property identifies the element as containing the title of a source. The language tag "`fr`" is code assigned to the French language in [ISO 639]. It indicates the book's title is written in French. It does not indicate that the book itself is written in French, nor that the researcher who created the *citation element* was working in French. The *citation element* in this example does not contain a *layer identifier*.

The `name` and `lang` *properties* MUST, if present, be serialised as JSON strings. The `value` *property* MUST be serialised as either a JSON string or a JSON integer. For the purpose of this standard, a JSON integer is defined as a JSON number, as defined in §6 of [RFC 7159], but without a fractional or exponential component. A JSON integer may be negative. The `value` *property* SHOULD NOT be serialised as a JSON integer unless the *citation elememt* is defined as having an integer value.

> *Example* — The Citation Elements standard says that the *citation element* stating that a 2nd edition consulted SHOULD have the value "2" rather than "2nd" or "`second edition`". The element's value however is defined as a string rather than an integer, which is to allow descriptive edition labels like "`large print`". It therefore *should not* be serialised with a JSON integer as this incorrect example does:
>
> ```
> { "name": "http://terms.fhiso.org/sources/edition",
>   "value": 2 }
> ```

> *Editorial note* — The format of the *layer identifier* is undecided, but it will be serialised either as a JSON string or a JSON integer.

> *Editorial note* — The language about the *citation elememt* having an integer value is imprecise due to there not being any such elements yet, and the Citation Elements spec not yet having a definition of an integer.

2.2 XML serialisation

The `CitationElement` *data type* is serialised in XML as an XML element. It is formally the responsibility of the parent structure to define the name of this element, but in all instances in this standard the element is named `<cev:element>`. Each of its *properties* is serialised as either an attribute on this element or a child element of it, as indicated below.

Property	XML representation
`layer`	A `layer` attribute of type to be determined.
`name`	A `name` attribute of type `xsd:anyURI`.
`lang`	An `xml:lang` attribute of type `xsd:language`. This is a standard XML attribute defined in §2.12 of [XML].
`value`	A `<cev:value>` child element whose content is of a type that depends on the value of the `name` attribute, but may safely be parsed as `xsd:string`.

Editorial note — The format of the `layer` attribute has yet to be determined. It may be `xsd:token` though `xsd:integer` is also possible.

Editorial note — This serialisation uses two XML element in FHISO's `cev:` namespace: `<cev:element>` and `<cev:value>`. It would be better if these could both be in the `gx:` namespace, the latter reusing the existing `<gx:value>` element. This will result in a more natural XML serialisation, and leave FHISO's namespace unpopulated with XML element and free for future. Obviously any such use of the GEDCOM X namespace needs permission from the GEDCOM X project team.

Example — The *citation element* containing the title of Settipani's book *Les ancêtres de Charlemagne* would be serialised in XML as follows:

```
<cev:element xml:lang="fr"
             name="http://terms.fhiso.org/sources/title">
  <cev:value>Les ancêtres de Charlemagne</cev:value>
</cev:element>
```

This is the exact XML analogue of the JSON example given above.

Editorial note — There is no compelling technical reason why `layer` and `name` must be attributes rather than child elements. However the 'Principle of Readability' in IBM's *Principles of XML design* series suggests they should be attributes:

> If the information is intended to be read and understood by a person, use elements. In general this guideline places prose in element content. If the information is most readily understood and digested by a machine, use attributes.

6

> In general this guideline means that information tokens that are not natural language go in attributes.
>
> URLs are used as a particular example with the recommendation that URLs are placed in attributes.

3   The `SourceCitation` Data Type

GEDCOM X defines a `SourceCitation` *data type* which is identified by the following IRI:

`http://gedcomx.org/v1/SourceCitation`

It is used to represent citations to sources. It has two *properties*, to which this standard adds a *extension property* named `elements`:

Property	Description, data type and constraints
`lang`	an OPTIONAL ISO *language tag* per [RFC 5646].
`value`	an OPTIONAL *formatted citation* which is a `string` containing either plain text or a fragment of XHTML.
`elements`	an OPTIONAL *citation element set* represented by a list of *citation elements*, each represented by the `CitationElement` *data type*. Order *must* be preserved, except as explicitly allowed by the Citation Elements standard.

> *Editorial note* — GEDCOM X refers to the value of the `lang` property as a "locale tag", but this term is not used in [RFC 5646]. A language tag matching the `langtag` production may contain region or script subtags, such as de-CH for Swiss German or ro-Cyrl for Romanian written using Cyrillic letters. Presumably this is what is meant by a "locale tag".

> *Editorial note* — Do *citation elements* really belong in the `SourceCitation`? At one level it seems obviously right: the elements are part of the citation and logically belong there. But the `SourceCitation` is really the representation of a *formatted citation*, and a single `SourceDescription` can have several `SourceCitations` differing in language or citation style. The *citation elements* are not style-dependent and only rarely language-dependent, so it seems more logical that they belong in the `SourceDescription` to avoid duplication. A future draft may well move them there.
>
> Two other slightly related suggestions.
>
> — The `SourceCitation` should have a `style` property to identify style variants, e.g. Chicago vs MLA.
> — A `SourceReference` also should be able to contain `SourceCitations` and `CitationElements`, as this is where a page number logically belongs, unless it is GEDCOM X's intention that every record, page, etc., should have its own `SourceDescription`.

> Neither of these are in scope for FHISO's current work, but are worth suggesting to the GEDCOM X team.

The `value` and `elements` *properties* contain alternative representations of the same underlying information: the former as a *formatted citation* designed to be read and understood by a person; the latter as a *citation element set* designed also to be digested by a machine.

When a `SourceCitation` has both `lang` and `elements` *properties* present, in addition to specifying the language of the `value` *property*, the `lang` *property* also provides a default language tag for each *citation element* in the `elements` *property*.

> *Note* — The additional use of `lang` as the default language tag for each *citation element* is a direct consequence of how GEDCOM X uses the standard `xml:lang` attribute. The XML serialisation of GEDCOM X requires the `xml:lang` attribute containing the `lang` *property* to be placed on the `<gx:citation>` element. As §2.12 of [XML] says this attribute applies to all content, direct or otherwise, it must apply to each *citation element*.

3.1   JSON serialisation

The `SourceCitation` *data type* is serialised in JSON as an object, with each *property* being represented by a JSON member with the same name. The value of the `elements` member should be a JSON array of objects, each of which is a JSON serialisation of a `CitationElement` *data type*.

> *Example* — A simplified citation to *Les ancêtres de Charlemagne* could be represented in JSON as follows:
>
> ```
> { "lang":  "en",
>   "value": "Christian Settipani, Les ancêtres de Charlemagne, 2nd
>   → ed",
>   "elements": [
>     { "name":  "http://terms.fhiso.org/sources/authorName",
>       "value": "Settipani, Christian" },
>     { "name":  "http://terms.fhiso.org/sources/title",
>       "value": "Les ancêtres de Charlemagne",
>       "lang":  "fr" },
>     { "name":  "http://terms.fhiso.org/sources/edition",
>       "value": "2" } ] }
> ```
>
> This example contains both a *formatted citation* and three *citation elements* representing the same information.
>
> The *formatted citation* is correctly tagged with the language code en denoting English. This is because, even though the book's title is French, the citation as a whole is in English. Had the citation been writen in French, the edition would have been written "2e éd" rather than "2nd ed".

> The language of the authorName *citation element* defaults to en, as this is the value of the SourceCitation's lang *property*. This may or may not be what was intended: the author is French but his name would not normally be altered in translation to English. The explicit *language tag* is necessary on the title *citation element*, as the title is clearly French.

3.2 XML serialisation

The SourceCitation *data type* is serialised in XML as an XML element. It is formally the responsibility of the parent structure to define the name of this element, but in every instance of its use in the GEDCOM X standard the element is named <gx:citation>. Each of its *properties* is serialised as either an attribute on this element or a child element of it, as indicated below.

Property	XML representation
lang	An xml:lang attribute of type xsd:language. This is a standard XML attribute defined in §2.12 of [XML].
value	A <gx:value> child element whose content is of type xsd:string.
elements	A sequence of <cev:element> child elements each of which is of type CitationElement.

> *Example* — The simplified citation to *Les ancêtres de Charlemagne* could be represented in XML as follows:
>
> ```
> <gx:citation xml:lang="en">
>   <gx:value>Christian Settipani, Les ancêtres de Charlemagne,
>     2nd ed.</gx:value>
>   <cev:element name="http://terms.fhiso.org/sources/authorName"
>     >Settipani, Christian</cev:element>
>   <cev:element name="http://terms.fhiso.org/sources/title"
>     xml:lang="fr">Les ancêtres de Charlemagne</cev:element>
>   <cev:element name="http://terms.fhiso.org/sources/edition"
>     >2</cev:element>
> </gx:citation>
> ```
>
> This is the exact XML analogue of the last JSON example. In this case, the application of xml:lang="en" to the enclosed *citation elements* follows from the definition of the xml:lang attribute in the XML standard.

3.3 XHTML value properties

GEDCOM X says the *formatted citation* in the value *property* of the SourceCitation *data type* MAY be a fragment of XHTML.

Editorial note — The extact form of the `value` *property* is confused in the current GEDCOM X drafts. It may be a piece of plain text or a fragment of XHTML, but the use of XHTML that is underspecified.

The GEDCOM X draft says it MAY be an [XHTML] `<cite>` element, and if so that "the element MUST represent the title of a work". But [XHTML] defines (by reference to §9.2.1 of [HTML4]) the `<cite>` as just "a citation or a reference to other sources", and [HTML5] says "it must include the title of the work or the name of the author (person, people or organization) or an URL reference, which may be in an abbreviated form as per the conventions used for the addition of citation metadata." Restricting the use of `<cite>` as GEDCOM X does prevents its use for sources that have no title and introduces an unnecessary incompatibility with HTML. It seems clear that HTML intends the `<cite>` element to be used for any *formatted citation*, and it is suggested that GEDCOM X adopts this too. Yet the default behaviour of browsers is to render a `<cite>` element in italics, as just a title would be, so it is perhaps best to remove all reference to `<cite>` from GEDCOM X. This is consistent with current practice in the FamilySearch API which is to include HTML markup but not to enclosed in a `<cite>` element, nor to include such an element.

When HTML in a GEDCOM X `value` property is serialised as XML, the [GEDCOM X XML] specification gives no hint as to whether the HTML markup must be escaped as a single string. Experimentation with the FamilySearch API suggests it is supposed to be, for example:

```
<gx:citation xml:lang="en">
  <gx:value><![CDATA[[
    Christian Settipani, <i>Les ancêtres de Charlemagne</i>
  ]]></gx:value>
</gx:citation>
```

This format is consistent with how HTML must be treated in JSON, but is unnatural from an XML point of view. Perhaps GEDCOM X could follow the example of the [RDF XML] spec by allowing unescaped XHTML as a property value when a `parseType="Literal"` attribute is given?

GEDCOM X contains no mechanism for determining whether the `<value>` element is plain text or HTML. An application must rely on heuristics, such as looking to see whether all instances of < (U+003C) are part of a well-formed XML tag, but this is undesirable as < and > have uses, particularly in linguistics and critical editions, that may look superficially like XML tags. This problem would be best solved by adding an OPTIONAL `textType` *property* to the `SourceCitation` data type (orthogonal to `parseType` discussed above, which is not a property of the data type, just a piece of syntactic sugar). This *property* indicates whether the `value` property is plain text or XHTML. Such a property is already used elsewhere in GEDCOM X and its values are defined in §1.3.8 of GEDCOM X.

Including a `textType` *property* will also define the dialect of XHTML that is used in a *formatted citation*. It is currently unclear whether the requirements of §1.3.8 apply to this `value` element. It would make sense for them to apply uniformly to all uses of XHTML within GEDCOM X.

GEDCOM X underspecifies exactly what syntactic form XHTML text may take: specifically, must it have a top-level element? It would make sense if, again following the example of [RDF XML], it were defined to match the `content` production of [XML]. This allows fragments of XML without a single top-level element, which is consistent with its current use in the FamilySearch API.

A related question is whether the HTML must conform to [XML Names]. Conformance with it is currently implicit in GEDCOM X by virtue of it referencing [XHTML], yet the FamilySearch API fails to declare the XHTML namespace, which is contrary to [XHTML]. A pragmatic solution would be to allow non-XML parsing per [HTML5] when the data is in a string, and XML-compatible XHTML parsing when it is included with parseType="Literal". This would make the `xmlns` declaration optional when HTML is escaped and legitimise the present behaviour of the FamilySearch API.

With the suggestions outlined in this note, the example above could be written:

```
<gx:citation xml:lang="en" textType="xhtml">
  <gx:value xmlns="http://www.w3.org/1999/xhtml"
            parseType="Literal">
    Christian Settipani, <i>Les ancêtres de Charlemagne</i>
  </gx:value>
</gx:citation>
```

Regardless of the above argument for using HTML5's parsing, to be forwards compatible, GEDCOM X should give consideration to using a definition of HTML other than [XHTML]. Requiring The registration of the `text/html` the in [IANA MIME] database generally refers to the most recent stable version of HTML, currently 5.1. Saying that XHTML text in GEDCOM X MUST be valid XML content in a `text/html` document is perhaps an option.

FHISO would benefit from working with GEDCOM X towards a solution to these problems, perhaps by the TSC submitting written feedback to the GEDCOM X project on FHISO's behalf.

Applications that conform to this FHISO standard MUST allow the attributes listed in §5 of [RDFa Core] to be present on XHTML elements in the `value` *property*. Applications MAY reject, ignore or remove uses of these attributes that does comply with [RDFa Core].

> *Note —* This does not require applications to parse and understand those attributes, but an application MUST NOT treat correctly-used RDFa attributes as syntax errors, and MUST NOT strip them from the XHTML other than at the explicit requested of the user.

When new *formatted citations* are created, this FHISO standard recommends that they SHOULD be formatted in XHTML with their constituent *citation elements* marked up as described in FHISO's [CEV RDFa] standard.

> *Example —* In the previous example, the *formatted citation* "Christian Settipani, Les ancêtres de Charlemagne, 2nd ed." was encoded as plain text, but this standard recommends the use of HTML with RDFa attributes. This very same *formatted citation* is used as an example in the [CEV RDFa] standard, where it is correct mark up with RDFa attributes is explained. Including it in a `value` property in XML yields the following:
>
> ```
>     <gx:citation xml:lang="en">
>       <gx:value><![CDATA[[
>         <span xmlns="http://www.w3.org/1999/xhtml" lang="en"
>               vocab="http://terms.fhiso.org/sources/" typeof="Source">
>           <span property="authorName"
>                 content="Settipani, Christian">Christian
>     → Settipani</span>,
>           <i property="title" lang="fr">Les ancêtres de Charlemagne</i>,
>           <span property="edition" content="2">2nd ed.</span>
>         </span>
>       ]]></gx:value>
>     </gx:citation>
> ```

If a `SourceCitation` has a `value` property containing HTML, an application MAY parse it according to the rules in [CEV RDFa] to extract *citation elements*. If any *citation elements* are found, each extracted *citation element* MAY be added to the *citation element set* in the `elements` property if and only if it does not have the same *layer identifier*, *citation element name* and *language tag* as a *citation element* that was in the `elements` *citation element set* before the extraction began.

4 References

4.1 Normative references

[CEV Concepts]

    FHISO (Family History Information Standards Organisation). *Citation Elements: General Concepts.* Exploratory draft of standard. See http://tech.fhiso.org/drafts/cev-concepts.

[CEV RDFa]

    FHISO (Family History Information Standards Organisation). *Citation Elements: Bindings for RDFa.* Exploratory draft of standard. See http://tech.fhiso.org/drafts/cev-rdfa-bindings.

GEDCOM X

    Intellectual Reserve Inc. *The GEDCOM X Conceptual Model.* Stable draft, accessed May 2017. See http://gedcomx.org/.

[GEDCOM X JSON]

    Intellectual Reserve Inc. *The GEDCOM X JSON Serialization Format.* Stable draft, accessed May 2017. See http://gedcomx.org/.

[GEDCOM X XML]

    Intellectual Reserve Inc. *The GEDCOM X XML Serialization Format.* Stable draft, accessed May 2017. See http://gedcomx.org/.

[RFC 2119]

    IETF (Internet Engineering Task Force). *RFC 2119: Key words for use in RFCs to Indicate Requirement Levels.* BCP 14. Scott Bradner, 1997. See http://tools.ietf.org/html/rfc2119.

[RFC 3987]

    IETF (Internet Engineering Task Force). *RFC 3987: Internationalized Resource Identifiers (IRIs).* Martin Duerst and Michel Suignard, 2005. See http://tools.ietf.org/html/rfc3987.

[RFC 5646]

    IETF (Internet Engineering Task Force). *RFC 5646: Tags for Identifying Languages.* BCP 47. Addison Phillips and Mark Davis, eds., 2009. See http://tools.ietf.org/html/rfc5646.

[RFC 7159]

    IETF (Internet Engineering Task Force). *The JavaScript Object Notation (JSON) Data Interchange Format* Tim Bray, ed., 2014. (See http://tools.ietf.org/html/rfc7159.)

4.2   Other references

[HTML4]

W3C (World Wide Web Consortium). *HTML 4.01 Specification.* W3C Recommendation, 24 Dec 1999. See http://www.w3.org/TR/html4.

[HTML5]

W3C (World Wide Web Consortium). *HTML5: A vocabulary and associated APIs for HTML and XHTML.* W3C Recommendation, 28 Oct 2014. See http://www.w3.org/TR/html5/.

[IANA MIME]

IANA (Internet Assigned Numbers Authority). *Media Types.* See http://www.iana.org/assignments/media-types/.

[ISO 639]

ISO (International Organization for Standardization). *ISO 639-1:2002. Codes for the representation of names of languages — Part 1: Alpha-2 code.* 2002.

[ISO 8601]

ISO (Internation Organization for Standardization). *ISO 8601:2004. Data elements and interchange formats — Information interchange — Representation of dates and times.* 2004.

[ISO 10646]

ISO (International Organization for Standardization). *ISO/IEC 10646:2014. Information technology — Universal Coded Character Set (UCS).* 2014.

[RDF XML]

W3C (World Wide Web Consortium). *RDF 1.1 XML Syntax.* W3C Recommendation, 25 Feb 2014. See http://www.w3.org/TR/rdf-syntax-grammar/.

[XHTML]

W3C (World Wide Web Consortium). *XHTML™ 1.0 The Extensible HyperText Markup Language (Second Edition): A Reformulation of HTML 4 in XML 1.0.* W3C Recommendation, 26 Jan 2000, revised 1 Aug 2002. See https://www.w3.org/TR/xhtml1/.

[XML]

W3C (World Wide Web Consortium). *Extensible Markup Language (XML) 1.0 (Fifth Edition).* W3C Recommendation, 26 Nov 2008. See https://www.w3.org/TR/REC-xml/.

[XML Names]

W3 (World Wide Web Consortium). *Namespaces in XML 1.0 (Third Edition).* W3C Recommendation, 8 Dec 2009. See https://www.w3.org/TR/REC-xml-names/.