

Syntax and Vocabulary of the Academic Metadata Format

draft: 2004-06-10

The latest version of this document is available on the web at <http://amf.openlib.org/doc/asakusa.html>. This document contains contributions by Tim D. Brody, Zhuoan Jiao, Thomas Krichel and Simeon M. Warner. It is maintained by Thomas Krichel. We are grateful for helpful comments from José Manuel Barrueco Cruz, Christopher F. Baum and Ivan V. Kurmanov.

1 Introduction

This document is a draft for the Academic Metadata Format (AMF). AMF encodes descriptions of

- academics, i.e. people who produce scientific research;
- resources created and used by academics, at the moment these cover textual resources only but later non-textual resources like software and datasets may be added;
- organizations that are important in the academic world, including universities, research centers, academic publishers, scholarly societies, funding agencies etc.

AMF relies as much as possible on standard vocabulary by borrowing metadata terms from other vocabularies. The relevant vocabularies are

- Dublin Core Metadata Element Set, see DCMI (1999), labeled dc:
- Dublin Core Qualifiers, see DCMI (2000b), labeled dcq:
- DCMI Type Vocabulary, see DCMI (2000a), labeled dctype:
- San Antonio Community Profile, see AX (2004) (to be revised), labeled SAP:
- DC citation terms, see Apps (2002), labeled dccite:
- vCard, see Howes, Smith, and Dawson (1998) , and Ianella (2001), labeled vCard:

AMF is not definite and may change at any moment. The current specification is provided as a basis for experimental deployment only. During this test phase, the draft standard's files are maintained at <http://amf.openlib.org>. Work on the AMF is supported by the Open Archives Initiative.

The remainder of the document is organized as follows. In Section 2, we introduce the general markup of AMF. In Section 3, we describe the names and semantics of elements used by AMF. In Section 4, we discuss constraints on the contents of elements. Such value constraints are indicated by the use of italics in the description of element semantics in Section 3. In Section 5, we present optional attributes that may be useful to further qualify element contents. In Section 6 we give some examples.

2 AMF markup

AMF is encoded in XML. All XML elements defined in this document belong to the <http://amf.openlib.org> namespace and must be qualified accordingly. AMF data must be wrapped by an amf root

```
<amf xmlns="http://amf.openlib.org">
```

```
...
```

```
</amf>
```

This specification comes with an XML Schema To simplify using the schema to process the data, it is useful to include a `schemaLocation` hint attribute into the amf wrapping element. When used, `schemaLocation` attribute must be

in the `http://www.w3.org/2001/XMLSchema-instance` and it must quote the pair `http://amf.openlib.org`
`http://amf.openlib.org/2001/amf.xsd`

AMF is an open vocabulary in the sense that the AMF XML schema file allows to place elements from foreign vocabularies within the AMF vocabulary. This can be done on the in the contents of the wrapping `amf` element, or in any child contents of that element. Foreign element names must be namespace qualified.

By convention, the `xsi` namespace prefix is used for XML Schema instance namespace. Even when no `schemaLocation` attribute is provided, the XML Schema instance namespace declaration is necessary for `xsi:type` attribute if it is used, See section 5.6. A good AMF document might look like this:

```
<amf xmlns="http://amf.openlib.org"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xsi:schemaLocation="http://amf.openlib.org
                        http://amf.openlib.org/2001/amf.xsd">
  <text id="oai:arXiv:hep-lat/0008015"> <title> ... </title> ...
</text>
</amf>
```

In the example, `text` is a “noun” element. In general, the AMF root element must contain one or more nouns. Nouns are repeatable. There are four nouns:

<code>person</code>	a physical person
<code>organization</code>	an entity that has physical persons as its members
<code>text</code>	a <code>dctype:text</code>
<code>collection</code>	a <code>dctype:collection</code> of resources

Each instance of a noun element in AMF data that is not an empty element is called an AMF record. All child elements of AMF records are optional and repeatable. An AMF record admits two types of child elements.

The first type are “adjective” elements. Adjectives give further information about nouns. Some adjectives have a nested structure. In the example above, `title` is an adjective.

The second type are “verb” elements. Verbs relate one noun to other nouns. Each verb must have one or more nouns as children. Verbs must not have adjectives as direct children.

3 The elements of AMF

3.1 The person and organization nouns

The person noun element describes or refers to a physical person:

```
<person id="..."> ...adjectives and verbs... </person> or <person ref="..." />
```

The organization noun element describes or refers to an organization. An organization is a group of two or more persons:

```
<organization id="..."> ...adjectives and verbs... </organization> or <organization ref="..." />
```

Both nouns accept the same verbs and adjectives. Therefore they will be collectively referred to as the “p/o” noun in the remainder of this document.

3.1.1 The adjectives of the p/o noun

```
<name> unstructured full name, as vCard:FN </name>
<shortname> short name e.g. IMF, as vCard:NickName </shortname>
<familyname> full name, as vCard:N;FamilyName </familyname>
<givenname> given name, as vCard:N;GivenName </givenname>
<additionalname> additional name, as vCard:N;AdditionalName </additionalname>
<nameprefix> honorary prefix, as vCard:N;HonoraryPrefix </nameprefix>
<namesuffix> honorary name, as vCard:N;HonorarySuffix </namesuffix>
<date> date associated with the p/o, as dc:date </date>
<homepage> URL of homepage </homepage>
<postal> postal address as vCard:LABEL </postal>
<phone> telephone number as vCard:TEL;TYPE=pref,voice </phone>
<fax> fax number vCard:TEL;TYPE=pref,fax </fax>
<email> email as vCard:EMAIL;TYPE=internet,pref </email>
<identifier> for the p/o from a scheme that does not use AMF, as dc:identifier </identifier>
```

3.1.2 p/o to p/o verbs

<isreplacedby> an organization is replaced by another </isreplacedby>
<replaces> an organization replaces another </replaces>
<ispartof> an organization is a part of another </ispartof>
<haspart> an organization has another as a part </haspart>

3.1.3 p/o to text verbs

<isauthorof> as dc:creator </isauthorof>
<iseditorof> as dc:creator or dc:contributor </iseditorof>
<ispublisherof> in the sense of dc:publisher </ispublisherof>
<istranslatorof></istranslatorof>
<ismaintainerof> p/o who maintains metadata about the text </ismaintainerof>

3.1.4 p/o to collection verbs

<iseditorof> p/o responsible for the contents of the collection </iseditorof>
<ispublisherof> in the sense of dc:publisher </ispublisherof>
<ismaintainerof> p/o who maintains metadata about the collection </ismaintainerof>

3.2 The text noun

The text noun element describes or refers to a text, independent of its status. Thus a PhD thesis, an article in a learned journal, the transcript of a speech etc, are all texts:

<text id="..."> ...adjectives and verbs... </text> or <text ref="..." />

Note: A journal is not a text, it is a collection. A book may also be a collection if it contains papers by different authors.

3.2.1 The adjectives of the text noun

<title> as dc:title </title>
<abstract> as dcq:abstract </abstract>
<keywords> list of uncontrolled keywords, may be subject to a scheme vocabulary to be developed </keywords>
<classification> list of classification codes, see section </classification>
<copyright> a plain-text statement about the copyright, as dc:rights </copyright>
<status> a plain-text description of the status of the text, say published in a journal, presented at a conference etc.</status>
<comment> something about the text that is not the status, e.g. a dedication </comment>
<email> *email* for the text, not necessarily one of the authors or editors </email>
<date> *date* associated with the text </date>
<displaypage> *URL* of a page where access to the text is explained </displaypage>
<citation> unstructured full text of citation </citation>
<serial> container tag for structured serial access information that *citation* can provide
<journaltitle> title of serial, as SAP:title, dccite:journaltitle </journaltitle>
<journalabbreviatedtitle> abbreviated title of serial, as SAP:stitle, as dccite:journalabbreviatedtitle </journalabbreviatedtitle>
<journalidentifier> identifier (usually ISSN) of journal, as dccite:journalidentifier </journalidentifier>
<issuedate> *date* on the serial issue cover, as SAP:date, as dccite:cronology </issuedate>
<volume> as SAP:volume, as dccite:volume </volume>
<part> as SAP:part, as dccite:number </part>
<issue> as SAP:issue, as dccite:number </issue>
<season> season of publication (spring or summer or autumn or winter), as SAP:ssn, as dccite:cronology </season>
<quarter> quarter of publication (1 or 2 or 3 or 4), as SAP:quarter, , as dccite:cronology </quarter>
<startpage> number of the first page of the text in the serial issue, as SAP:spage </startpage>
<endpage> number of the last page of the text in the serial issue, as SAP:epage </endpage>
<pages> unstructured page data, as SAP:pages, as dccite:pagination </pages>
<articlenumber> article number, in the absence of pagination, as SAP:artnum </articlenumber>

</serial>
 <file> a container for full-text file information; it may be repeated for each component file.
 <url> *URL* for the file itself </url>
 <function> the function of the file in the text, e.g. main text, appendix </function>
 <format> same as dc:format, encoded in Internet Media Types, see IANA (2001)</format>
 <restriction> text that explains access to the file, as dc:rights </restriction>
 </file>
 <reference> a container for a reference made by the text
 <literal> text of reference </literal>
 <context> to context of the citation </context>
 </reference>
 <type> a *text type* </type>
 <identifier> an identifier for the text from a scheme that does not use AMF, as dc:identifier </identifier>

3.2.2 Text to person/organization verbs

<hasauthor> </hasauthor>
 <haseditor> </haseditor>
 <haspublisher> in the sense of dc:publisher </haspublisher>
 <hassupervisor> </hassupervisor>
 <hastranslator> </hastranslator>
 <hasmaintainer> </hasmaintainer>

3.2.3 Text to text verbs

<iserratumof> </iserratumof>
 <haserratum> </haserratum>
 <isaddendumto> </isaddendumto>
 <hasaddendum> </hasaddendum>
 <isreviewof> </isreviewof>
 <hasreview> </hasreview>
 <iscommenton> </iscommenton>
 <hascomment> </hascomment>
 <istranslationof> </istranslationof>
 <hastranslation> </hastranslation>
 <isreplacedby> as dcq:isReplacedBy </isreplacedby>
 <replaces>as dcq:replaces </replaces>
 <ispartof> as dcq:isPartOf </ispartof>
 <haspart> as dcq:hasPart </haspart>
 <isreferencedby> for example cited by another, as dcq:isReferencedBy text </isreferencedby>
 <references> e.g. cites another text, as dcq:references </references>
 <isversionof> points to an earlier text that current text is developed from, as dcq:isVersionOf </isversionof>
 <hasversion> points to a later text developed from the current text, as dcq:hasVersion </hasversion>
 <isformatof> points to an original text with the same intellectual contents in a different format,
 as dcq:isFormatOf </isformatof>
 <hasformat> points to a derived text with the same intellectual contents in a different format,
 as dcq:hasFormat </hasformat>

3.2.4 Text to collection verbs

<ispartof> a text belongs to a collection, as dcq:isPartOf </ispartof>

3.3 The collection noun

The collection noun element is used whenever statements about a set of several texts are being made. This can be a classification collection, (i.e. all the texts that have the subject classification code), a serial, the papers presented at a conference etc.:

<collection id="..."> ...adjectives and verbs... </collection> or <collection ref="..." />

3.3.1 The adjectives of the collection noun

<title> same as a journal title, conference title etc. </title>
<abbreviatedtitle> abbreviation, e.g. PRL </abbreviatedtitle>
<description> a plain text description of the collection, as dc:description </description>
<homepage> *URL* for humans to read more about the collection </homepage>
<accesspoint> *URL* for machines to access the collection </accesspoint>
<type> a *collection type* </type>
<identifier> an identifier for the collection from a scheme that does not use AMF, as dc:identifier </identifier>

3.3.2 Collection to collection verbs

<isreplacedby> as dcq:isReplacedBy </isreplacedby>
<replaces> as dcq:replaces </replaces>
<ispartof> as dcq:isPartOf </ispartof>
<haspart> as dcq:hasPart </haspart>

3.3.3 Collection to text verbs

<haspart> as dcq:haspart </haspart>

3.3.4 Collection to p/o verbs

<haseditor> </haseditor>
<haspublisher> in the sense of dc:publisher </haspublisher>
<hasmaintainer> p/o who maintain metadata about the collection </hasmaintainer>

4 Element value constraints

The values of some of the elements are restricted. These content types are listed here.

4.1 The *date* type

The *date* adjective is of the form yyyy[-mm[-dd]], where [] encloses optional components. For details, see the *date* type definition of XML Schema.

4.2 The *URL* type

The value must be a valid Uniform Resource Locator.

4.3 The *email* type

The value must be a valid email address.

4.4 The *collection type* type

The *collection noun* covers a wide variety of things in AMF. It is useful to indicate the type of a collection through a controlled vocabulary.

book	as SAP:book	classification	a classification scheme
proceedings	conference proceedings	serial	a serial of texts
journal	as SAP:journal	archive	an archive of documents

4.5 The *text type* type

The *text noun* covers a wide variety of things in AMF. It is useful to indicate the type of a text through a controlled vocabulary.

book	as SAP:book	article	as SAP:article
conferencepaper	as SAP:proceeding	preprint	also covers working papers and technical reports
bookitem	as SAP:bookitem		as SAP:preprint
code	computer code component	as DC:software	

5 Attributes of elements

5.1 The `id` attribute

All AMF records (i.e. non-empty nouns) may have an `id` attribute. The value must be an XML Name. If a value is set for a particular record, it is assumed that within the scope of a collection of AMF records, the record is uniquely identified by the value of this attribute.

5.2 The `ref` attribute

Any noun—be it empty or not—may carry a `ref` attribute. If it is present, its value is identical to the `id` attribute of another record. AMF `ref` attributes may be resolved to records that have the appropriate identifiers. The details of the resolution algorithm are outside the scope of AMF.

A noun may carry a `ref` and an `id` attribute. If that is the case, the value of the `ref` attribute is ignored.

5.3 The `from` and `until` attributes

All verbs admit two additional attributes: `from` and `until`. The values of these attributes must be of the type *date*. These attributes indicate a time span for which the relationship holds. The dates are inclusive. Example:

```
<iseditorof from="1999-01-01" until="2000-01-01"> ... </iseditorof>
```

5.4 The `xml:lang` attribute

All adjective elements have an optional attribute called `xml:lang`. It takes the same syntax as in the XML 1.0 specification. It uses values for `xml:lang` from <http://www.w3.org/TR/2000/WD-xml-2e-20000814#sec-lang-tag>.

As a general rule, the `xml:lang` attribute refers to the value of the element content. For example `<title xml:lang="fr">Robin des Bois</title>` does not mean that the text is a French translation of the adventures of Robin Hood.

The only exception occurs when an element content has the *URL* type. Then, the resource referenced by the URL is supposed to have the human language indicated by the value of the `xml:lang` attribute.

5.5 The `event` attribute

All *date* elements may have an optional attribute `event` that indicates what happened on the date. The admissible values are

<code>created</code>	text was first written, as <code>dcq:created</code> or person was born, as <code>vCard:BDAY</code>
<code>available</code>	date where a person was alive or a resource is available, as <code>dcq:available</code>
<code>issued</code>	the formal publication date of a text, as <code>dcq:issued</code>
<code>modified</code>	the date a resource was changed, as <code>dcq:modified</code>

Example: `<date event="created">2000-03</date>`

5.6 The `type` attribute

The `identifier`, `classification` and `keyword` adjectives may have an `type` attribute in the XML Schema Instance namespace, i.e. <http://www.w3.org/2001/XMLSchema-instance>

What needs to be said here for consistency, is that it must be "type" attribute in the XML Schema Instance namespace, namely >>>

. In that case, controlled values for the value have been registered with AMF. The AMF Controlled vocabulary document lists all the controlled vocabularies.

6 Examples

None of the following examples is fictitious. However, the description of the items that is made through the examples may not be complete, to conserve space.

```
<amf>
  <text id="bible"> <title>The Holy Bible</title> </text>

  <text><title>The book of Genesis</title><ispartof>
    <text ref="bible"/></ispartof></text>
</amf>
```

```

<amf>
  <organization id="RePEc:edi:ocddfr">
    <name xml:lang="en">Organization for Economic Development and Cooperation</name>

    <shortname xml:lang="en">OECD</shortname>
    <name xml:lang="fr">Organisation de Cooperation de
    Develloppement Economiques</name>
    <shortname xml:lang="fr">OCDE</shortname>
    <homepage xml:lang="en">http://www.oecd.org/ </homepage>
    <homepage xml:lang="fr">http://www.oecd.org/index-fr.htm</homepage>
    <haspart><organization ref="RePEc:edi:edoecfr">
      <ispublisherof><collection ref="RePEc:oed:ocdec">
        <haspart>
          <text>
            <hasauthor>
              <person ref="RePEc_per_1956-06-20_GIUSEPPE_NICOLETTI"/>
            </hasauthor>
            <title>
              REGULATION IN SERVICES: OECD PATTERNS AND ECONOMIC IMPLICATIONS
            </title>
            <abstract>
              The paper looks at patterns of regulation in service industries
              and explores their implications for service performance.
            </abstract>
            <abstract xml:lang="fr">
              Cette étude analyse les approches règlementaires dans les
              secteurs des services et explore leurs implications pour
              les performances sectorielles dans les pays de l'OCDE.
            </abstract>
            <file>
              <url>http://www.olis.oecd.org/olis/2001doc.nsf/linkto/eco-wkp(2001)13</url
            >
            <format>application/pdf</format>
          </file>
        </text></haspart>
      </collection></ispublisherof>
    </organization>
  </haspart>
</organization>
</amf>

```

```

<amf>
  <collection id="csfhrd">
    <title>Classification Scheme for Human Rights Documentation</title>
    <homepage>http://www.huridocs.org/clasengl.htm</homepage>
    <haseditor>
      <person>
        <name>Ivana Caccia</name>
      </person>
    </haseditor>
    <haspart>
      <collection id="csfhrd:GEN_II.10"><title>
        natural justice </title></collection>
      <collection id="csfhrd:GEN_II.20"><title>
        universality / relativism </title></collection>
      <collection id="csfhrd:GEN_II.30"><title>
        philosophy & human rights </title></collection>
      <collection id="csfhrd:GEN_II.40"><title>

```

```
    political theories & human rights </title>
  <haspart>
    <collection id="csfhrd:GEN_II.41"><title>democracy</title></collection>
    <collection id="csfhrd:GEN_II.42"><title>liberalism</title></collection>
    <collection id="csfhrd:GEN_II.45"><title>marxism</title></collection>
  </haspart></collection>
</haspart></collection>
</amf>
```

7 Instructions for processing software

Processing software for AMF may preprocess the *email* and *URL* types. For the *email* type, it may check against a regular expression of email addresses and remove the part of the value that does not match that regular expression. For the *url* type, it may remove whitespace as described in RFC-1738.

References

- Apps, Ann (2002). A citation Element Refinement for dc:identifier. available <http://epub.mimas.ac.uk/DC/citproposal.html>.
- AX, NISO Committee (2004). The OpenURL Framework for Context-Sensitive Services (Initial Registry Content). draft available at http://library.caltech.edu/openurl/StandardDocuments/Z39_88_Pt2_ballot%20final.pdf.
- DCMI (1999). Dublin Core Metadata Element Set, Version 1.1: Reference Description. available at <http://www.dublincore.org/documents/dces/>.
- DCMI (2000a). DCMI Type Vocabulary. available at <http://www.dublincore.org/documents/2000/07/11/dcmi-type-vocabulary/>.
- DCMI (2000b). Dublin Core Qualifiers. available at <http://www.dublincore.org/documents/dcmes-qualifiers/>.
- Howes, Tim, Mark Smith, and Frank Dawson (1998). vCard MIME Directory Profile. RFC 2426, available at <http://www.imc.org/rfc2426>.
- IANA (2001). Media Types. available at <http://www.isi.edu/in-notes/iana/assignments/media-types/media-types>.
- Ianella, Renato (2001). Representing vCard Objects in RDF/XML. available at <http://www.w3.org/TR/vcard-rdf>.