



Content Standards and Processes

Elsevier
Radarweg 29
1043 NX Amsterdam
The Netherlands

Phone: +31 20 485 2316
Fax: +31 20 485 3266
Email: j.migchielsen@elsevier.com

Release note of the AnI 5.12 Output DTD and Schemas

AnI_5.12_Output_Schemas_release_notes.doc — version 1.10, 25 November 2015, by Jos Migchielsen

version 0.1	26 February 2008	First draft
version 0.2	11 March 2008	Second draft, remarks from HDa
version 1.0	18 March 2008	Final version, correction from HDa
version 1.1	28 November 2008	Patch 5.12p1 described in Section 7.1
version 1.2	19 December 2008	Patch 5.12p2 described in Section 7.2
version 1.3	7 October 2009	Patch 5.12p3 described in Section 7.3
version 1.4	13 October 2010	Patches 5.12p4 and p5 described in Sections 7.4 and 7.5
version 1.5	22 June 2011	Patch 5.12p6 described in Section 7.6
version 1.6	15 June 2012	Patches 5.12p7 and p8 described in Sections 7.7 and 7.8
version 1.7	30 September 2013	Version 5.12.1 described in Section 8
version 1.8	18 June 2015	Version 5.12.2 described in Section 9
version 1.9	30 July 2015	Version 5.12.3 described in Section 10
version 1.11	25 November 2015	Version 5.12.4 described in Section 11

Contents

1. Introduction	2
2. The AnI DTD version 5.12	2
3. Backwards compatibility	2
4. The AnI Schema version 5.12	3
5. The AnI Product Schemas version 5.12	3
6. Changes	5
6.1. Removed entities	5
6.2. New entities	5
6.3. Removed elements	5
6.4. New elements and attributes	5
7. Patches	7
7.1. Version 5.12p1	7
7.2. Version 5.12p2	7
7.3. Version 5.12p3	7
7.4. Version 5.12p4	8
7.5. Version 5.12p5	8
7.6. Version 5.12p6	8
7.7. Version 5.12p7	8
7.8. Version 5.12p8	8
8. Version 5.12.1	8
9. Version 5.12.2	9
10. Version 5.12.3	9
11. Version 5.12.4	9

1. Introduction

Currently Opsbank is using the Abstract and Indexing DTD 5.11 for its XML output. For Opsbank II it was decided to update this DTD and create an equivalent schema as well as a limited number of schemas for different products. This document describes some features of the DTD and schemas and lists all the changes.

Note: The DTD will not be made available to external customers.

2. The AnI DTD version 5.12

Like DTD 5.11 the new DTD contains one top-element, the doctype, `bibdataset`. An output file begins as follows:

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE bibdataset PUBLIC
    "-//ES//DTD abstracting and indexing DTD version 5.12//EN//XML"
    "ani512.dtd">
```

The top-element has three mandatory attributes, `xmlns`, `xmlns:ait` and `xmlns:ce`. As the attributes have fixed values they need not be mentioned. They define the namespaces that are used in the DTD and the corresponding prefixes. The namespaces are:

```
http://www.elsevier.com/xml/ani/ani
http://www.elsevier.com/xml/ani/ait
http://www.elsevier.com/xml/ani/common
```

They are named using URIs – these are abstract names not pointing to any page on the Elsevier corporate website. The first namespace is the so-called target namespace. Any element without a namespace prefix is in that namespace. These namespaces differ from the ones in the previous DTD where, moreover, the target namespace was actually not used.

Two optional attributes were added to the top-element: `xmlns:xsi` and `xsi:schemaLocation`. By adding these two attributes a file conforming to the schema will also conform to the DTD.

It is expected that customers use the public-ID string in the doctype declaration and an XML catalogue to find the DTD on their system.

3. Backwards compatibility

One of the requirements was that the new DTD was backwards compatible. *Backwards compatibility* means that applications that can handle documents conforming to a certain version, can also handle documents conforming to a previous version. Unfortunately, this is not the case as in DTD 5.12 several elements were removed. Also, in DTD 5.11 most elements did not belong to a namespace whereas in DTD 5.12 every element belongs to a namespace.

In practice, files that conform to DTD 5.11 will conform to DTD 5.12 as the elements that were removed were never actually used. All new elements and attributes were added as optional ones.

There are other changes that endanger the backwards compatibility. For instance, an attribute `country` will no longer contain a two-letter code taken from a list of values (based on the ISO3166 standard). Instead the type of the attribute was changed to CDATA and it will contain the three-letter code that Opsbank II uses. Customers will have to be prepared for this.

Another change is that more citation information will be present in *all* output. Citation information is present in subelements of `head`. The following elements for instance will be present when possible: `author-keywords` and `figure-information` in `head/citation-info`, `conferenceinfo` and `publisher` in `head/source`.

4. The AnI Schema version 5.12

The DTD was converted to an equivalent XML schema. When using schemas elements from different namespaces cannot be declared in one schema. This resulted in five schemas: `ani512.xsd` (the main schema), `ait.xsd`, `ce.xsd`, `ani-ce.xsd` and `xml.xsd`.

An output file will typically begin as follows:

```
<?xml version="1.0" encoding="UTF-8"?>
<bibdataset xmlns="http://www.elsevier.com/xml/ani/ani"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:ce="http://www.elsevier.com/xml/ani/common"
  xmlns:ait="http://www.elsevier.com/xml/ani/ait"
  xsi:schemaLocation="http://www.elsevier.com/xml/ani/ani
    http://www.elsevier.com/xml/ani/ani512.xsd">
```

The three mandatory attributes of the top-element, `xmlns`, `xmlns:ait` and `xmlns:ce`, define the namespaces that are used in the schema and the corresponding prefixes. The namespaces are:

```
http://www.elsevier.com/xml/ani/ani
http://www.elsevier.com/xml/ani/ait
http://www.elsevier.com/xml/ani/common
```

They are named using URIs – these are abstract names not pointing to any page on the Elsevier corporate website. The first namespace is the so-called target namespace. Any element without a namespace prefix is in that namespace.

It is expected that customers use the second part of the content of attribute `xsi:schemaLocation` and an XML catalogue to find the schema and other schemas that are imported.

5. The AnI Product Schemas version 5.12

Product schemas were created for fifteen different products. Below is a list of the products and the corresponding schemas.

Compendex	<code>compendex.xsd</code>
Elsevier_BIOBASE	<code>elsevier_biobase.xsd</code>

EMBASE	embase.xsd
EMBASE_COM	embase_com.xsd (*)
EMBiology	embiology.xsd (*)
EMCare	emcare.xsd
EMCare_COM	emcare_com.xsd (*)
EMSCOPES	emscopes.xsd (*)
EnCompass_LIT	encompass_lit.xsd
EnCompass_PAT	encompass_pat.xsd
Fluidex	fluidex.xsd
GEOBASE	geobase.xsd
Medline	medline.xsd (*)
Paperchem	paperchem.xsd
WorldTextiles	worldtextiles.xsd (*)

The schemas marked with (*) are for internal use only.

An output file that conforms to a product schema will typically begin as follows:

```
<?xml version="1.0" encoding="UTF-8"?>
<bibdataset xmlns="http://www.elsevier.com/xml/ani/ani"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:ce="http://www.elsevier.com/xml/ani/common"
  xmlns:ait="http://www.elsevier.com/xml/ani/ait"
  xsi:schemaLocation="http://www.elsevier.com/xml/ani/ani
    http://www.elsevier.com/xml/ani/compindex.xsd">
```

Again, the second part of the content of attribute `xsi:schemaLocation` and an XML catalogue are expected to be used to find the schema and the other schemas that are needed.

A product schema redefines simple and complex types in the main schema (`ani512.xsd`). For instance, the GEOBASE product schema contains the following:

```
<xs:complexType name="enhancementType">
  <xs:complexContent>
    <xs:restriction base="ani:enhancementType">
      <xs:sequence>
        <xs:element minOccurs="0" ref="ani:descriptorgroup"/>
        <xs:element minOccurs="0" ref="ani:classificationgroup"/>
      </xs:sequence>
    </xs:restriction>
  </xs:complexContent>
</xs:complexType>
```

One can see that the complex type `enhancementType` is restricted and can only contain elements `descriptorgroup` and `classificationgroup`. That is, in a file that conforms to the GEOBASE product schema, element `enhancement` can only have those two elements as subelement and not e.g. element `chemicalgroup`.

6. Changes

The changes below are with respect to version 5.11 of the DTD. They are of course carried over in the schemas. Where appropriate a remark on the schemas is added.

6.1. Removed entities

The ISO3166 (countries) and ISO369 (languages) entities were removed from the DTD. The following attributes were made of type CDATA: `affiliation/@country`, `country/@iso-code`, `source/@country`, `conflocation/@country`, `citation-language/@xml:lang`, `abstract-language/@xml:lang`, `titletext/@xml:lang`, `abstract/@xml:lang` and `translated-sourcetitle/@xml:lang`. The corresponding attributes in the schema are untyped.

6.2. New entities

Entity `%yesno;` was added. The entity `%Date.att;` in the schemas replaced by an attribute group `date.att` in the `ce` namespace.

The values for the different enhancement types are now placed in entities. These new entities are: `%chemicalsourcetype.att;`, `%classificationtype.att;`, `%descriptortype.att;`, `%manufacturertype.att;`, `%sequencebanktype.att;` (renamed from `%Sequencebanks.ref;`) and `%tradenametype.att;`. In the schemas there are corresponding simple types, e.g. `descriptortype.att`. If a product contains any of these enhancements these simple types are usually restricted in the corresponding product schema.

6.3. Removed elements

The following elements were removed: `degree`, `document-delivery`, `documentid`, `edition`, `formatinfo`, `linkinfo`, `medium`, `previous`, `price`, `publication-notes`, `reprint`, `season` (from `%PubDate.ref;`) `sectionauthor`, `sectionauthortext`, `sectiontitle`, `service`, `toc` and `tocentry`.

6.4. New elements and attributes

Top element `bibdataset`: Optional attributes `xmlns:xsi` and `xsi:schemaLocation` were added to top-level element `bibdataset`; this makes it possible to validate a schema file against the DTD.

Classifications: Element `classification` contained a classification code or a description of a classification. To be able to distinguish between these two, elements `classification-code` and `classification-description` were added. For example:

```
<classificationgroup>
  <classifications type="EMCLASS">
    <classification>
      <classification-code>18</classification-code>
      <classification-description>Cardiovascular Diseases and
        Cardiovascular Surgery</classification-description>
    </classification>
    <classification>
```

```

    <classification-code>3</classification-code>
    <classification-description>
      Endocrinology</classification-description>
  </classification>
  <classification>
    <classification-code>37</classification-code>
    <classification-description>Drug Literature
      Index</classification-description>
  </classification>
</classifications>
</classificationgroup>

```

Autoposting: To support so-called “autoposting” element `API-descriptorgroup` and 25 other elements were added. `API-descriptorgroup` is an optional subelement of element `enhancement`. Autoposting is described in a separate document.

Grants: To support the capturing of grant information elements `grantlist`, `grant`, `grant-id`, `grant-acronym` and `grant-agency` were added. `grantlist` is an optional subelement of element `head` and can be used in all product schemas.

Conferences: To be able to capture more conference information elements `conftheme`, `conforganization`, `confseriestitle` and `confURL` were added. They are all optional subelements of `confevent`.

Designated countries: Element `designatedcountries` (a subelement of `patent`) can now contain more than one subelement `country`.

CAS registry numbers: Element `chemical/cas-registry-number` was made optional.

Patents: Patent information is now considered to be citation information and hence element `patent` was added as optional subelement of `citation-info`. For reasons of backwards compatibility element `patent` was not removed as subelement from `enhancement`. In the product schemas `patent` cannot appear as subelement of `enhancement`.

Enzymes: Element `enzyme-commission-number` was added as subelement of `chemical`.

Itemfilelinks: Element `itemfilelinks` was added as subelement of `bibrecord`. The element consists of one or more elements `itemfilelink`.

Links: The content model of elements `link`, `sublink` and `subsublink` was changed to allow the use of element `sup` and `inf`. Also, an optional attribute `linknr` was added to the three elements. Note that these attributes are added for future expansion, they will not appear in the output until further notice.

External source: Element `external-source` was added as optional subelement of `item-info`.

Descriptors: An optional attribute `name` was added to element `descriptors`.

7. Patches

When a patch to the output DTD and schemas is published, all version numbers of the DTD and the output schemas (excluding the support schemas `ce.xsd`, `ani-xsd`, `ait.xsd` and `xml.xsd`) are changed to 5.12p1, 5.12p2, etc. The names of the DTD and the schemas are not changed and therefore neither will their invocations (in the `xsi:schemaLocation` attribute of the top element). The DTD and schemas in use should be replaced by the new versions.

7.1. Version 5.12p1

- The following ten values were added to entity `%descriptortype.att` in the DTD and to simple type `descriptortype.att` in the output schema: `DID`, `GLY`, `OZR`, `SPO`, `SUA`, `SID`, `BLS`, `CSS`, `CSW`, and `CSL`. The restrictions to this simple type in the output product schemas were not changed.
- In the following eight output product schemas the restriction to simple type `sequencebanktype.att` was removed: `elsevier_biobase.xsd`, `medline.xsd`, `emscopes.xsd`, `emcare.xsd`, `emcare_com.xsd`, `embiology.xsd`, `embase.xsd` and `embase_com.xsd`. This effectively means that `EMBL` now becomes a possible value for attribute `sequencebank/@name` in those eight products.
- The default values for attributes `titletext/@xml:lang` and `abstract/@xml:lang` were removed from the DTD and the output schema. `OpsbankII` will output three-letter country codes.
- The declaration of the unused element `API-CAS-nr` was removed from the DTD and the output schema.
- A new element was declared, `API-LTM-group`, and the model of element `API-LTM` was changed to contain this element. These changes were done in the DTD and in the output schema and affect only those customers who receive material enhanced with “autoposting”.
- In the DTD the attribute `weight` was added to element `classification`. The attribute is of type `CDATA` and will contain one of the following four texts: “?”, “-“, “+”, “++”.
- In the main output schema the attribute `weight`, with the four possible values “?”, “-“, “+”, “++”, was added to element `classification`. This was done by means of introducing a new complex type, `classificationType`, and making `classification` of that type. The attribute `weight` is removed in every output product schema, except for `encompass_lit.xsd` and `encompass_pat.xsd`, by redefining the new complex type. This effectively means that attribute `classification/@weight` is only possible in the two `EnCompass` products.

7.2. Version 5.12p2

- Default values for attributes `titletext/@xml:lang` and `abstract/@xml:lang` were reintroduced. For both attributes the default value is `eng`.
- A new element was declared, `API-LT-group`, and the model of element `API-LT` was changed to contain this element. These changes were done in the DTD and in the output schema and affect only those customers who receive material enhanced with “autoposting”.

7.3. Version 5.12p3

- The value `cb`, conference abstract, was added to entity `%Citationtype.ref` in the DTD and to simple type `citationtype.att` in the output schema.

7.4. Version 5.12p4

- Values DRA and MEA were added to parameter entity %descriptortype.att; in the DTD and to simple type descriptortype.att in the schema as well as to the restrictions to this simple type in the output product schemas embase.xsd, embase_com.xsd and medline.xsd.
- Values CTN and TRP were added to the restriction of simple type descriptortype.att in the output schema embase_com.xsd.
- Value EMCLASS was added to the restriction of simple type classificationtype.att in the output schema embase_com.xsd.

7.5. Version 5.12p5

- Added values CBA, CBB, CBC, CBE, CNC to parameter entity %descriptortype.att; in the DTD and to simple type descriptortype.att in the schema.
- Added values CBNBGEO, CBNBSCOPE, CBNBSECTOR, CBNBSIC to parameter entity %classificationtype.att; in the DTD and to simple type classificationtype.att in the schema.
- Added values n and w to parameter entity %Sourcetype.ref; in the DTD to simple type sourcetype.att in the schema.
- Changed entity %PubDate.ref; in the DTD and group pubdateType in the schema.

7.6. Version 5.12p6

- Added attribute weight to element sublink (three possible values “a”, “b” and “not-defined” which is the default) in the DTD and the schema.

7.7. Version 5.12p7

- Added attribute seq to element ipc-code in the DTD (CDATA) and the schema (type positiveInteger).

7.8. Version 5.12p8

- Added values ba and rf to simple type citationtype.att in the schema, and to parameter entity Citationtype.ref in the DTD.
- Added values AUTARC-CONFIGURATION, CONFIGURATION-VALUE, PATAUTO, PATENT-FAMILY, PATMANU, REAXYS-PRIORITY, SATURATION and SUPPORT-INFORMATION to simple type classificationtype.att in the schema, and to parameter entity classificationtype.att in the DTD.
- Added values AUT, MGS, MLA, MMS, MSC, MUD and REA to simple type descriptortype.att in the schema, and to parameter entity descriptortype.att in the DTD.

8. Version 5.12.1

In September 2013 ORCID's and grant agency IDs were added to the schemas and the DTD. To signify the importance of these changes a new version number was introduced, v5.12.1. Additionally an ISO code and a new descriptor type were added. The new version remains backwards compatible. The changes are:

- Added optional attribute `author/@orcid` to the DTD (CDATA) and the schema.
- Added optional (new) element `grant/grant-agency-id` to the DTD (PCDATA) and the schema (type `string`).
- Added optional attribute `grant-agency/@iso-code` to the DTD (CDATA) and the schema.
- Added value `MDV` to entity `descriptortype.att` in the DTD and to simple type `descriptortype.att` in the schema.

9. Version 5.12.2

In June 2015 element `enhancement` received two optional attributes, `status` and `type`, to support new index deliveries.

10. Version 5.12.3

In July 2015 element `API-ATM/LT-count` was made optional.

11. Version 5.12.4

In November 2015 optional element `grantlist/grant-text` was added. It has an attribute `xml:lang` with default value `eng`.