



**HAL**  
open science

## The EuDML metadata schema

Claude Goutorbe, Thierry Bouche, Jean-Paul Jorda, Nicolas Houillon, Jiri Rákosnik

► **To cite this version:**

Claude Goutorbe, Thierry Bouche, Jean-Paul Jorda, Nicolas Houillon, Jiri Rákosnik. The EuDML metadata schema. [Technical Report] D3.6, Mathdoc. 2013, pp.32. hal-03765892

**HAL Id: hal-03765892**

**<https://hal.univ-grenoble-alpes.fr/hal-03765892>**

Submitted on 31 Aug 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## DELIVERABLE

**Project Acronym:** EuDML  
**Grant Agreement number:** 250503  
**Project Title:** The European Digital Mathematics Library

### D3.6: The EuDML metadata schema

Revision: 1.1 as of 28th January 2013

#### Authors:

Claude Goutorbe UJF/CMD  
Thierry Bouche UJF/CMD  
Jean-Paul Jorda EDPS  
Nicolas Houillon UJF/CMD

#### Contributors:

Jiří Rákosník MIAS

Project co-funded by the European Commission within the ICT Policy Support Programme		
Dissemination Level		
P	Public	✓
C	Confidential, only for members of the consortium and the Commission Services	

## Revision History

Revision	Date	Author	Organisation	Description
Draft	2013/01/15	Claude Goutorbe	UJF/CMD	First attempt
0.1	2013/01/16	Thierry Bouche	UJF/CMD	Version for internal discussion
0.9	2013/01/21	Nicolas Houillon	UJF/CMD	Checked everything and corrected examples
0.99	2013/01/22	Thierry Bouche	UJF/CMD	Final version after internal discussion and XML examples
1.0	2013/01/24	Thierry Bouche	UJF/CMD	Final version with updated refs to web site and web services
1.1	2013/01/28	Thierry Bouche	UJF/CMD	Final version taking into account Jiří's comments

### Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Standardization and upgrade of JATS article . . . . .	1
1.2	Main differences between initial and final versions of the EuDML metadata schema . . . . .	1
<b>2</b>	<b>The EuDML schema version 2.0 specification</b>	<b>2</b>
2.1	Supported document types . . . . .	2
2.2	The records types . . . . .	3
2.3	The EuDML journal article schema . . . . .	3
2.4	The EuDML book schema . . . . .	4
2.4.1	<b:front> : book level metadata . . . . .	5
2.4.2	<b:body>: book contents . . . . .	5
<b>3</b>	<b>Tagging best practices</b>	<b>6</b>
<b>4</b>	<b>Online examples</b>	<b>7</b>
	<b>Appendix: Tagging Best Practices</b>	<b>11</b>

## 1 Introduction

This document describes the final version of the EuDML metadata schema. It is a sequel to deliverable D3.2 [2] which describes the initial version of the EuDML schema. In order to be reasonably self-contained, it has an amount of overlap with D3.2 but does not reproduce the parts of this previous work that explained our initial design, as our work with EuDML metadata confirmed the conclusions in D3.2. For the rationale, expectations, and abstract modeling of the project's metadata schema, we refer thus explicitly to § 2-4 in D3.2.

### 1.1 Standardization and upgrade of JATS article

Since we released the initial version of the EuDML schema in D3.2, the NLM Journal Archiving and Interchange Tag Suite used as main reference for our work has been passed over to the NISO standard body which has published NISO Z39.96-2012 [4]. As a consequence, a new Journal Archiving and Interchange Tag Set schema, with NISO version 1.0 has been published [3].

The differences between the NLM 3.0 and NISO 1.0 versions of JATS are shown on the web page <http://jats.nlm.nih.gov/archiving/tag-library/1.0/n-zad2.html>.

Straightforward consequences of these changes are:

- The out-of-the-box article schema is now perfectly suited to our needs, as all the extensions we had adopted, or planned to adopt in the next revision of our schema are now standard.
- A number of nice features have been introduced that could help improve some parts of EuDML operation. For instance, the `<citation-alternatives>` element could be used to store different representations of the same citation (for instance as a result of some internal non-destructive matching/enhancing process, such as adding reference database links, or a structured (`<element-citation>`) version for an initially unstructured one. We do not expect to take full advantage of these possibilities during the current project, as it is now almost over. But we think this standard now offers room for all metadata enhancements and refinements we could think about in a foreseeable future.
- It is “fully backward compatible with NLM version 3.0” and thus with EuDML article 1.0 (see <http://jats.nlm.nih.gov/about.html>).
- The NCBI Book Tag Set is not part of the NISO standard and was thus not upgraded to that standard.

### 1.2 Main differences between initial and final versions of the EuDML metadata schema

Compared to initial schemas suite published in D3.2

- All new structures introduced in JATS NISO 1.0 are supported.

- The **article** document type is described with the standard article XSD provided by JATS NISO 1.0.
- Name spaces have changed.
- The **book** document type is described using a new schema defined from scratch based explicitly on JATS NISO 1.0 elements. This schema is very similar to EuDML v. 1.0 book structure, but the way it is defined has completely changed and backward compatibility has not been sought in this case.
- Multi-volume works are now not any more supported by a third XML record type (**mbook**), but through a special metadata element in the book schema.

## 2 The EuDML schema version 2.0 specification

### 2.1 Supported document types

The EuDML project is concerned with the identification and description of mathematical digital documents brought by associated content providers. It is understood as a reference library system providing discovery services to self-contained mathematical texts which have been scientifically validated and formally published (and are thus the kind of documents that are expected to be cited in the scientific literature): the *EuDML items*.

Specifically, the EuDML metadata schema must handle the following document types, that are explicitly supported:

- A journal article;
- a book, namely
  - a monograph (which might be a doctoral dissertation, a memoir...),
  - an edited book (a book that contains chapters or articles that have been written by different authors and collated by scientific editors, which might be a conference proceedings volume)—this can also be named a collective book,
  - a single volume from a multi-volume work;
- a part of a book such as a chapter, or an article in a proceedings volume;
- a multi-volume work.

The following do not constitute requirements on EuDML services and are thus not in the scope of a EuDML metadata schema.

- Material that is not considered as having been formally published (e.g. preprints, personal web pages, working paper...).
- Special provisions for papers not generally accessible online (e.g. on paper only, inhouse access only, library catalogue...).
- Version control for documents, as EuDML only considers works in published final form.
- Complicated author/contributor structures for documents such as non-alphabetical order, as this is of no significance in math publishing.

The description of access embargo periods (moving wall) and other licensing, access barriers, digital rights management issues, were included in the above list of exclusions in D3.2. After some discussion, it was agreed within the consortium that it could enhance the user experience if EuDML had some mean to tell users whether a given item is currently open access or not. As the adopted schemas allow to include this information in records (see appendix, § 5.3), we decided to accept and standardize it.

## 2.2 The records types

The schema as defined by this document is implemented in two XML schemas (XSD files) holding XML metadata for two major types of documents, namely journal articles and books. As a consequence:

- There is no separate schema for book parts (typically individual articles in a proceedings volume); these are described and exchanged within the whole book record they belong to.
- There is no separate schema for multi-volume works. Instead a book record may carry the description of the multi-volume work it belongs to, if any.

In the following we briefly describe these two schemas, and present EuDML recommended practices for tagging.

## 2.3 The EuDML journal article schema

During the course of this project, the NLM journal article archiving and interchange tag suite, which we had selected as the basis for the EuDML article schema, became a NISO standard. It appears that all the changes we had to incorporate in our version of the schema were taken into account by the new standard (see [2, § 7]). Therefore, the EuDML article schema is now exactly the Journal Archiving and Interchange Tag Library (NISO JATS version 1.0).

The official specification of this standard makes the remark that most of its users have no use for xml namespaces, therefore no target namespace is defined in the corresponding W3C xml schema file. Should an application require an XML namespace, it is recommended to use “<http://jats.nlm.nih.gov>” as the namespace URI. This is the case of the EuDML system, since the OAI-PMH protocol used for data exchange requires the use of XML namespaces.

Hence the formal specification of the journal article schema, located at <http://eudml.org/schema/2.0/eudml-article-2.0.xsd> adds the definition of this target namespace to the standard XML schema definition. There are no other modifications to the standard.

Structured documentation for this schema is available at <http://jats.nlm.nih.gov/archiving/tag-library/1.0/>.

## 2.4 The EuDML book schema

There is large amount of overlap between metadata needed to describe journal articles and that needed to describe books (including book parts).

In particular, all metadata needed to describe the intellectual content (see D3.2, § 4.2 [2]), as well as describing authorship are identical.

The EuDML book schema has been designed using the article schema as a base, retaining common elements and adding book-specific elements and attributes. Hence this schema uses elements residing in the journal article namespace (uri: <http://jats.nlm.nih.gov>, with namespace prefix “a” in the following description) and defines its own target namespace (uri: <http://eudml.org/schema/2.0/eudml-book>, with namespace prefix “b” in the following description) in which book-specific elements reside.

While most of these new elements and attributes have been inspired by the NCBI Book Tag Set [1], the resulting schema differs from it in a number of ways, and will be described without referring to the original. As is the case with the journal article schema, elements that are used to encode the actual textual content of a book have been retained (since they may be used in future developments) but are not described here.

The formal definition of this schema can be found at the following location: <http://eudml.org/schema/2.0/eudml-book-2.0.xsd>.

Structured documentation is available at <http://eudml.mathdoc.fr/schemas/doc/book.html>.

In the following a high level view of the schema is given, readers needing a detailed description are referred to the above documentation.

The overall tree structure is similar to an article’s structure, with the following main parts:

```
<b:book>
  <b:front>...</b:front>
  <b:body> ...</b:body>
  <a:back> ...</a:back>
</b:book>
```

Note that the <a:back> element lives in the journal article namespace. In the context of EuDML, it is used to hold reference lists (i.e the list of works that are cited by the given item) and the content is the same for both document types.



### 2.4.1 <b:front> : book level metadata

Book level metadata is wrapped in the <b:front> container, which may contain the following book specific elements:

- <b:book-meta>. This is a container for book specific metadata, including the following elements:
  - <b:book-id> : book identifier, one or more,
  - <b:book-title-group> : container for all title related metadata,
  - all other metadata can be described by elements already defined in the “a” (article) namespace. They include in particular:
    - \* description of contributors (authors, editors, translators),
    - \* publishing information: publisher, date published, volume, edition, series,
    - \* abstracts, keywords, conference information in case of a proceeding volume,
    - \* links to the book’s text.
- <b:mbook-meta>. When a given book is actually part of a multi-volume work, this container element must be used to describe this work. Since a multi-volume work is actually a book that happens to have been published in several volumes, the content model for <b:mbook-meta> is the same as for <b:book-meta>. If present, this element must be the first child of the <b:front> element.

### 2.4.2 <b:body>: book contents

The body of a book contains the metadata for its constituent parts, when available, according to the following (recursive) representation.

```
<b:book-part>
  <b:book-part-meta> ... </b:book-part-meta>
  <b:body> ... </b:body>
  <a:back> ... </a:back>
</b:book-part>
```

The metadata for book parts is actually almost the same as for a journal article. A few elements have been renamed for consistency while serving exactly the same purpose: <b:book-part-id> should be used instead of <a:article-id>.

Various book part titles are grouped in a <b:title-group> element whose content model is slightly different from an article’s <a:title-group>, the <b:title> element should be used instead of <a:article-title>.

A book part may have its own reference list (as is the case of a proceedings article or a chapter in an edited book), the <a:back> element is thus allowed in a book part description.

The following is a simplified example of tagging a book part.

```
<b:book-part>
  <b:book-part-meta>
    <b:book-part-id pub-id-type="dmlcz-id">400227</b:book-part-id>
    <b:title-group>
      <b:title xml:lang="de">Bernard Bolzano's Schriften</b:title>
      <a:trans-title-group xml:lang="cs">
        <a:trans-title>Spisy Bernarda Bolzana</a:trans-title>
      </a:trans-title-group>
    </b:title-group>
    <a:contrib-group>
      <a:contrib contrib-type="author">
        <a:string-name>Bolzano, Bernard</a:string-name>
      </a:contrib>
    </a:contrib-group>
    <a:fpage>[1a]</a:fpage>
    <a:lpage>[1e]</a:lpage>
  </b:book-part-meta>
</b:book-part>
```

### 3 Tagging best practices

A distinct characteristic of the tag set we have selected is that it is not overly prescriptive and may be used in many ways. Although the documentation gives helpful hints and examples of use, interoperability in EuDML requires establishing a number of conventions that should be adhered to by content providers as well as the central system to enable the exchange of data in a most effective way.

This includes the initial uploading of data to the EuDML system, as well as the downloading by a content provider of its own data, possibly enhanced within the central system (e.g. MathML encoding of formulas, interlinking of documents).

A set of tagging recommendations (best practices) has been written so that content providers can deliver their data in a way that is consistent with the expectations of the EuDML system.

They include:

1. The definition of mandatory metadata elements, whose absence will cause the system to reject submitted data. They are the following:
  - at least one identifier (assigned by the content provider),
  - the original title,
  - the publication year,

- enough bibliographic data to uniquely identify the item's publication location among the published literature.  
For journal articles:
  - journal metadata,
  - issue number,
  - some way to unambiguously locate the article within the issue. This may be a page range, an article number or identifier (especially for electronic articles), or an ordinal position within the issue.
 For books:
  - author or scientific editor,
  - publisher;
- 2. conventions for tagging common bibliographic elements (titles, authors);
- 3. conventions for tagging supplementary bibliographic elements (abstracts, keywords);
- 4. conventions for dealing with various document types, e.g. using the book schema to encode monographs, conference proceedings, edited books, multi-volume works;
- 5. rules about the encoding of identifiers, assigned by the content provider as well as external services such as Zentralblatt MATH and Math Reviews indexing services;
- 6. how to deal with language and transliteration issues, since EuDML includes content written in many languages, using various scripts (e.g. latin, cyrillic as well as greek);
- 7. how to deal with the full text of documents: how to specify various access points (landing pages, pdf files location) and alternative full text formats (OCR, text extracted from a pdf file, tagged text);
- 8. how to encode reference lists, including linking from a reference to the actual work it describes;
- 9. how to handle mathematical expressions.

The rules are listed in the Appendix.

These rules and conventions have been embodied (as far as possible) by a Schematron specification that may be used by content providers to validate their data against them. A validation service is available at <http://eudml.mathdoc.fr/eudml-validation-demo/>.

## 4 Online examples

Conformant EuDML v2.0 XML examples are available from EuDML OAI-PMH server, see <https://project.eudml.org/oai-pmh-server>.

Examples of EuDML XML structures and coding practices can be found (in some non-documented variant) using the "NLM metadata via REST" services, at URLs according to the following pattern: [http://eudml.org/api/rest/\[eudml-id\]?format=eudml-nlm](http://eudml.org/api/rest/[eudml-id]?format=eudml-nlm), where [eudml-id] has the following format: `urn:eudml:doc:<eudml-short-id>`. and eudml-short-id is a number. These numbers can be gathered from the user interface of the EuDML system. An interactive demonstration of this service is available at <http://project.eudml.org/api-tester/restNlm>.

Here are random examples of the various item types available through EuDML that can be grabbed from the public OAI-PMH server:

- A journal article such as <https://eudml.org/doc/115934>: [XML link<sup>1</sup>](#);
- A book, of the following kinds:
  - a plain monograph such as <https://eudml.org/doc/219333>: [XML link<sup>2</sup>](#),
  - a doctoral dissertation such as <https://eudml.org/doc/192803>: [XML link<sup>3</sup>](#),
  - a memoir such as <https://eudml.org/doc/192658>: [XML link<sup>4</sup>](#);
  - an edited book (conference proceedings volume) such as <https://eudml.org/book/10052>: [XML link<sup>5</sup>](#)
  - a single volume from a multi-volume work such as <https://eudml.org/doc/202774>: [XML link<sup>6</sup>](#);

---

1. Full link: [https://eudml.org/oai/OAIHandler?verb=GetRecord&identifier=urn:eudml.eu:CEDRAM:urn:eudml.eu:CEDRAM\\_yadda:urn:eudml:doc:115934&metadataPrefix=eudml-article2](https://eudml.org/oai/OAIHandler?verb=GetRecord&identifier=urn:eudml.eu:CEDRAM:urn:eudml.eu:CEDRAM_yadda:urn:eudml:doc:115934&metadataPrefix=eudml-article2)

2. Full link: [https://eudml.org/oai/OAIHandler?verb=GetRecord&identifier=urn:eudml.eu:PLDML\\_book:urn:eudml.eu:PLDML\\_Book\\_yadda:urn:eudml:doc:219333&metadataPrefix=eudml-book2](https://eudml.org/oai/OAIHandler?verb=GetRecord&identifier=urn:eudml.eu:PLDML_book:urn:eudml.eu:PLDML_Book_yadda:urn:eudml:doc:219333&metadataPrefix=eudml-book2)

3. Full link: [https://eudml.org/oai/OAIHandler?verb=GetRecord&identifier=urn:eudml.eu:NUMDAM\\_book:urn:eudml.eu:NUMDAM\\_book\\_yadda:urn:eudml:doc:192803&metadataPrefix=eudml-book2](https://eudml.org/oai/OAIHandler?verb=GetRecord&identifier=urn:eudml.eu:NUMDAM_book:urn:eudml.eu:NUMDAM_book_yadda:urn:eudml:doc:192803&metadataPrefix=eudml-book2)

4. Full link: [https://eudml.org/oai/OAIHandler?verb=GetRecord&identifier=urn:eudml.eu:NUMDAM\\_book:urn:eudml.eu:NUMDAM\\_book\\_yadda:urn:eudml:doc:192658&metadataPrefix=eudml-book2](https://eudml.org/oai/OAIHandler?verb=GetRecord&identifier=urn:eudml.eu:NUMDAM_book:urn:eudml.eu:NUMDAM_book_yadda:urn:eudml:doc:192658&metadataPrefix=eudml-book2)

5. Full link: [https://eudml.org/oai/OAIHandler?verb=GetRecord&identifier=urn:eudml.eu:DML\\_CZ\\_Proceeding:urn:eudml.eu:DML\\_CZ\\_Proceeding\\_yadda:urn:eudml:book:10052&metadataPrefix=eudml-book2](https://eudml.org/oai/OAIHandler?verb=GetRecord&identifier=urn:eudml.eu:DML_CZ_Proceeding:urn:eudml.eu:DML_CZ_Proceeding_yadda:urn:eudml:book:10052&metadataPrefix=eudml-book2)

6. Full link: [https://eudml.org/oai/OAIHandler?verb=GetRecord&identifier=urn:eudml.eu:GDZ\\_Band:urn:eudml.eu:GDZ\\_Band\\_yadda:urn:eudml:doc:202774&metadataPrefix=eudml-book2](https://eudml.org/oai/OAIHandler?verb=GetRecord&identifier=urn:eudml.eu:GDZ_Band:urn:eudml.eu:GDZ_Band_yadda:urn:eudml:doc:202774&metadataPrefix=eudml-book2)

## References

- [1] Digital Archive of Journal Articles National Center for Biotechnology Information (NCBI) and National Library of Medicine (NLM). NCBI Book Tag Library version 3.0, November 2008. Full online documentation at <http://dtd.nlm.nih.gov/book/>.
- [2] Michael Jost, Thierry Bouche, Claude Goutorbe, and Jean-Paul Jorda. The EuDML metadata schema, November 2010. Deliverable D3.2 of EU CIP-ICT-PSP project 250503 EuDML: The European Digital Mathematics Library.
- [3] U.S. National Library of Medicine National Center for Biotechnology Information. Journal archiving and interchange tag library, niso jats version 1.0, August 2012. Full online documentation at <http://jats.nlm.nih.gov/1.0/>.
- [4] National Information Standards Organization. JATS: Journal Article Tag Suite, ANSI/NISO Z39.96-2012, August 2012. See <http://jats.niso.org/>.



## Appendix: EuDML metadata schema (final)/Tagging best practices

### Contents

- 1 Metadata for describing items
  - 1.1 Common bibliographic elements
    - 1.1.1 Title
    - 1.1.2 Contributors
      - 1.1.2.1 Editors
      - 1.1.2.2 Data associated with contributors
      - 1.1.2.3 Identifiers for contributors
    - 1.1.3 Page information
  - 2 Document types and how to deal with them
    - 2.1 Journal articles
    - 2.2 Conference proceedings articles
    - 2.3 Proceedings volume, edited books, monographs
    - 2.4 Multivolume works
  - 3 Describing the intellectual content
    - 3.1 Abstract
    - 3.2 MSC (Mathematics Subject Classification)
    - 3.3 Keywords
  - 4 How and where to add Identifiers
    - 4.1 EuDML identifiers
      - 4.1.1 EuDML identifiers in metadata
      - 4.1.2 EuDML identifiers in bibliographic references
    - 4.2 "Primary" identifier
      - 4.2.1 "Primary" identifiers in metadata
      - 4.2.2 "Primary" identifiers in bibliographic references
      - 4.2.3 Labels for the authorities assigning primary IDs
    - 4.3 "Document" identifiers
      - 4.3.1 How to use <self-uri>
      - 4.3.2 Value of <self-uri> element
    - 4.4 "External" identifiers
      - 4.4.1 How to use <ext-link> for external identifiers
      - 4.4.2 Where to use <ext-link> for external identifiers
  - 5 Full text
    - 5.1 Full NLM encoded item
    - 5.2 Other cases (OCR, TXT extracted from PDF, flat XML...)
      - 5.2.1 How to use <ext-link> for metadata full text
      - 5.2.2 Value of <ext-link> element
    - 5.3 Declaring when the full text is available as open access
  - 6 Bibliography (citations)
    - 6.1 Citation styles
      - 6.1.1 When to use mixed style
      - 6.1.2 When to use element style
    - 6.2 Value of @publication-type
    - 6.3 Identifiers and External links added by EuDML or local DML
    - 6.4 Authors structuring
  - 7 Language and transliteration issues
    - 7.1 Language
      - 7.1.1 Where xml:lang attribute should appear

- 7.1.2 Where xml:lang attribute may appear
    - 7.1.3 Where xml:lang attribute should not appear
  - 7.2 Transliteration
    - 7.2.1 Specifying transliteration scheme
    - 7.2.2 Transliterate author names
    - 7.2.3 Transliterate title
    - 7.2.4 Other elements
- 8 Mathematical expressions
- 9 Metadata producer

## Metadata for describing items

### Common bibliographic elements

#### Title

The article or book titles are to be placed in a **<title-group>** (<http://jats.nlm.nih.gov/archiving/tag-library/1.0/index.html?elem=title-group>) or **<book-title-group>** wrapper, which contains the original title (**<article-title>**, **<book-title>**), eventually the subtitle (**<subtitle>**), the translated and the transliterated versions (see Language and transliteration issues).

Item title is an obligatory element. However, in the very exceptional case where an item lacks a title, the **<title>** element should be omitted rather than left empty.

#### Contributors

Contributors is often expressed as an ordered list of authors. Author names may be provided in an unstructured way. In this case, it must be possible to add a structured version of the author name, provided by the metadata enhancer. See also #Transliteration for author names expressed in non latin characters.

Collective contributions should be tagged using **<collab>** instead of **<name>** (see: <http://jats.nlm.nih.gov/archiving/tag-library/1.0/index.html?elem=collab>). This is not expected to happen often in EuDML items.

#### **Example: Author name, with structured and unstructured version**

```
<contrib-group content-type="authors">
  <contrib contrib-type="author">
    <string-name>Dušan Bednařík</string-name> <!-- provided by provider -->
    <name> <!-- added by enhancer -->
    <surname>Bednařík</surname>
    <given-names>Dušan</given-names>
  </contrib>
  <contrib contrib-type="author">
    <string-name>Karel Pastor</string-name>
    <name>
      <surname>Pastor</surname>
      <given-names>Karel</given-names>
    </name>
    <email>pastor@inf.upol.cz</email>
  </contrib>
</contrib-group>
```



### Editors

To set information about *editors* (e.g for a proceedings, when *editors* replace *authors*), `<contrib>` element is used together with `@contrib-type` attribute set to `"editor"`.

In the case of an article published in a special issue, its *authors* (or *editors*) should be distinguished from the *issue editors*: `<contrib>` element is used together with `@contrib-type` attribute set to `"issue-editor"`.

### Data associated with contributors

Some of them may be useful to identify an author using a third-party author-identification service. These are handled as `<ext-link>` elements with `@content-type` attribute set to a service identifier.

### Identifiers for contributors

Contributors may be identified by:

- an email address
- EuDML author ID (similar to Zbl author identification?)
- Zentralblatt author ID e.g. **wiles.andrew-j** (<http://www.zentralblatt-math.org/zmath/en/search/?q=ai:wiles.andrew-j>)
- Maybe, in the future, by an ORCID (<http://www.orcid.org/>) identifier

#### **Example: Author description with identifiers link**

```
<contrib-group>
<contrib contrib-type="author">
  <ext-link content-type="zbl-author"
    xlink:href="http://www.zentralblatt-math.org/zmath/en/search/?q=ai:wiles.andrew-j">wiles.andrew-j</ext-link>
  <email>andrew.wiles@fermat.net</email>
  <name>
    <surname>Wiles</surname>
    <given-names>Andrew</given-names>
  </name>
</contrib>
</contrib-group>
```

### Page information

Obligatory metadata requires that each item be endowed with enough metadata to identify it uniquely among the whole literature. This implies that the full bibliographic references of its container be present, as well as the distinctive location of that item within its container: this is typically a page range for items originated in paper or in an electronic format modelled on paper, but it can as well be something else when the model deviates from the continuous page model.

Anyway, an item should always have `<fpage>` and `<lpage>` elements (the value being the string which is printed on the first and last pages: it can be a Roman numeral such as V, an arabic numeral as well as anything else). When content of `<fpage>` does not allow to sort properly articles within an single journal issue, it is advised to use the `@seq` attribute to provide the default ordering. The default value is considered to be equal to 0 and is not required if not considered useful for an item.

**Example: Pages ordered by seq**

```
<fpage seq="0">i</fpage>
<lpag>iii</lpag>
.....
<fpage seq="1">v</fpage>
<lpag>v</lpag>
.....
<fpage seq="2">1</fpage>
<lpag>13</lpag>
```

Sometimes, all articles from a given journal start in page 1. This may happen for two different reasons: either the journal style assigns article numbers to distinguish between them, or the journal is electronic only and affects unique identifiers to each published article. In the first case, JATS DTD provides a *@seq* attribute to the **<issue>** element which allows to locate the article in the sequence of all the issue's articles

**Example: Articles distinguished by their order within a single issue**

```
<issue seq="24"/>
<fpage>1</fpage>
<lpag>153</lpag>
```

In the second case, the **<elocation-id>** element is meant to store the electronic unique identifier of the given item.

**Example: Articles identified by unique ID**

```
<elocation-id>015204</elocation-id>
```

## Document types and how to deal with them

- **General tree structure** is already described in the JATS DTD documentation (<http://jats.nlm.nih.gov/archiving/tag-library/1.0/>) and therefore *is not detailed here*
- **General description of the tags and attributes** are already described in the JATS DTD documentation (<http://jats.nlm.nih.gov/archiving/tag-library/1.0/>) and therefore *is not discussed here*

### Journal articles

Journal articles do not differ in EuDML schema specification from plain NISO JATS DTD v. 1.0 as long as obligatory metadata is involved. It is mandatory to insert the journal description in **<journal-meta>**, a provider identifier for the journal, its title, and an issn. It is mandatory to insert volume and issue numbers if relevant, as well as date of publication.

As a rule of thumb, do not try to overload the **<journal-meta>** element by information or identifiers that might not be constant (or just constant) over the whole journal run, such as external identifiers to services that might have a different view on the same journal run. It is considered safer to insert such references as **<ext-link>** at the **<article-meta>** level.

### Conference proceedings articles

Conference proceedings articles are dealt with as journal articles if the proceedings volume is published in a journal special (or regular) issue. The bibliographic metadata is thus standard journal article metadata. The conference details are provided by the **<conference>** element in **<article-meta>**.

If they are published in a separate volume, the volume is encoded with the book schema, and each conference article is encoded as a **<book-part>** inside the **<body>** of that book.

## Proceedings volume, edited books, monographs

These are all encoded with the EuDML book schema. Each **<book-part-meta>** element can contain a **<book-part-id>** element serving the same purpose as the **<article-id>** element, with same rules for its *@pub-id-type* attribute.

The **<conference>** element has also been allowed in **<book-meta>**.

An isbn, publisher, and year of publication is obligatory for all flavours of books.

The *@book-type* attribute of the top level element **<book>** should be set to the following values:

"monograph" for monographs

"volume" for a volume of a multi-volume monograph. Such a book must contain a **<mbook-meta>** element describing the multi-volume work itself.

"proceedings" for a book that is a conference proceedings.

"proceedings-volume" for a book that is a volume in a multi-volume conference proceedings.

"edited-book" for a book that is a collection of independent chapters (typically with different authors, the book itself having scientific editors)

Note: A "proceedings" is in principle nothing more than an "edited-book" with a **<conference>** element, but it was thought that it could be handy to have this distinction as some handling could be optimized for proceedings books in a conference series.

## Multivolume works

Multivolume works are not encoded with a separate schema. Instead, each volume of a multivolume work must include an **<mbook-meta>** element (as the first child of its **<front>** element).

They can be proceedings (such as those of ICM), edited books (such as collected works on a dedicated subject), single authors' œuvre spanning multiple volumes.

The **<mbook-meta>** element has the same content model as the **<book-meta>** element.

## Describing the intellectual content

### Abstract

Abstract and translated abstract are respectively tagged with **<abstract>** and **<trans-abstract>** element. See also #Language and transliteration issues

#### *Example:* Article in French, with an abstract translated in english

```
<abstract>
  <p>Dans cet article nous proposons différents algorithmes pour résoudre une nouvelle
  classe de problèmes variationnels non convexes. Cette classe généralise plusieurs types d'inégalités
  variationnelles (Cho et al. (2000), Noor (1992), Zeng (1998), Stampacchia (1964)) du cas convexe
  au cas non convexe. La sensibilité de cette classe de problèmes variationnels non convexes a été aussi
  étudiée.</p>
</abstract>
<trans-abstract xml:lang="en">
  <p>In this paper we propose several algorithms of the projection
  type to solve a new class of nonconvex variational problems. This class generalizes many types
  of variational inequalities (Cho et al. (2000), Noor (1992), Zeng (1998), Stampacchia (1964))
  from the convex case to the nonconvex case. The sensitivity of this class of nonconvex variational
  problems is also studied.</p>
</trans-abstract>
```

## MSC (Mathematics Subject Classification)

Mathematics Subject Classification (<http://www.zentralblatt-math.org/msc/>) codes must be inserted using element `<kwd-group>` with attribute `@kwd-group-type` set to the actual scheme: "msc" following by the year (e.g. "**msc2000**").

*When the actual scheme is unknown, the attribute value must be set to "msc".*

More information about MSC :

- <http://202.38.126.65/mirror/www.ams.org/msc/msc-changes.html>
- <http://www.ams.org/mathweb/mi-classifications.html>

### Example: MSC codes

```
<kwd-group kwd-group-type="msc2000">
  <kwd>53C05</kwd>
  <kwd>53C30</kwd>
  <kwd>22E60</kwd>
</kwd-group>
```

### Identified revisions of the MSC :

2010: **msc2010**  
 2000: **msc2000**  
 1991: **msc1991**  
 1985: **msc1985**  
 1980: **msc1980**  
 1970: **msc1970**

## Keywords

Keywords must be inserted in a `<kwd-group>` container. If the keywords' language is not the article's language, it must be set in `@xml:lang` attribute. If keywords in different languages are available, they must be grouped by language in different keyword groups.

### Example: Keywords (article in english, keywords in french and english)

```
<kwd-group xml:lang="fr">
  <kwd>dépendences entre critères</kwd>
  <kwd>modélisation des préférences</kwd>
  <kwd>procédures interactives</kwd>
  <kwd>intégrale de Choquet</kwd>
  <kwd>méthodes ELECTRE</kwd>
</kwd-group>
<kwd-group>
  <kwd>dependencies between criteria</kwd>
  <kwd>preference modelling</kwd>
  <kwd>interactive procedures</kwd>
  <kwd>Choquet integral</kwd>
  <kwd>ELECTRE methods</kwd>
</kwd-group>
```

## How and where to add Identifiers

### EuDML identifiers

EuDML will assign identifiers to items, containers and authors. XML documents dispatched by EuDML will hold these identifiers.

### EuDML identifiers in metadata

Except for author and affiliations, <\*-id> elements must be used:

EuDML ID	In metadata	
	Element	Parent Element
Multi-book ID	< <b>book-id pub-id-type="eudml-id"</b> >	< <b>mbook-meta</b> >
Book ID	< <b>book-id pub-id-type="eudml-id"</b> >	< <b>book-meta</b> >
Journal ID	< <b>journal-id journal-id-type="eudml-id"</b> >	< <b>journal-meta</b> >
Journal issue ID	< <b>issue-id pub-id-type="eudml-id"</b> >	< <b>article-meta</b> >
Article ID	< <b>article-id pub-id-type="eudml-id"</b> >	< <b>article-meta</b> >
Book part ID	< <b>book-part-id pub-id-type="eudml-id"</b> >	< <b>book-part-meta</b> >
Author ID	< <b>ext-link ext-link-type="eudml-id"</b> >	< <b>contrib</b> >
Affiliation ID	< <b>ext-link ext-link-type="eudml-id"</b> >	< <b>aff</b> >

### EuDML identifiers in bibliographic references

In bibliographic references, EuDML identifiers must be added using <ext-link> elements with an @*ext-link-type* set to "eudml-id", the same way than external identifiers. See below.

#### "Primary" identifier

They identify an item (a journal article, a chapter, a contribution) or a container (a volume, a journal, a book, a multi-book). They are assigned by the the editor (DOI, PII and editor specific identifier) or by the local DML.

There is no need to associate this type of ID with URL for one of the following reason:

- they have a widely recognized resolution mechanism (e.g the DOI),
- they are used only inside the EuDML community (e.g the local DML identifiers),
- there is no resolution mechanism available (ISSN, ISBN).

If an item record is generated by consolidating metadata from many sources, it should hold each corresponding identifier.

#### "Primary" identifiers in metadata

Outside of bibliographic references, "primary" IDs must be set in the following elements (depending of the context, see below):

< <b>issn</b> > ( <a href="http://jats.nlm.nih.gov/archiving/tag-library/1.0/index.html?elem=issn">http://jats.nlm.nih.gov/archiving/tag-library/1.0/index.html?elem=issn</a> )	ISSN must be set or added. It's a common practice to assign a different ISSN for the print version and the online version of the journal. In this case, the @ <i>pub-type</i> ( <a href="http://jats.nlm.nih.gov/archiving/tag-library/1.0/index.html?attr=pub-type">http://jats.nlm.nih.gov/archiving/tag-library/1.0/index.html?attr=pub-type</a> ) attribute must be used, with a value of either " <b>ppub</b> " or " <b>epub</b> ".
< <b>isbn</b> > ( <a href="http://jats.nlm.nih.gov/archiving/tag-library/1.0/index.html?elem=isbn">http://jats.nlm.nih.gov/archiving/tag-library/1.0/index.html?elem=isbn</a> )	For some journals, in particular those publishing conference proceedings, an ISBN can be assigned to an issue. Books and multi-books should have ISBNs.
< <b>book-id</b> > ( <a href="http://eudml.mathdoc.fr/schemas/doc/book.html#element_book-id">http://eudml.mathdoc.fr/schemas/doc/book.html#element_book-id</a> )	Used to store the book ID as defined by the local DML and by the publisher. The attribute @ <i>pub-id-type</i> must be set according to the defined values
< <b>book-part-id</b> > ( <a href="http://eudml.mathdoc.fr/schemas/doc/book.html#element_book-part-id">http://eudml.mathdoc.fr/schemas/doc/book.html#element_book-part-id</a> ) ( <i>EuDML book DTD extension</i> )	Used to store the book part (chapter, contribution) ID as defined by the local DML and by the publisher. The attribute @ <i>pub-id-type</i> ( <i>part of EuDML book DTD</i> )

	<i>extension</i> ) must be set according to the defined values
< <b>journal-id</b> > ( http://jats.nlm.nih.gov/archiving/tag-library/1.0/index.html?elem=journal-id)	Used to store the journal ID as defined by the local DML and by the publisher. The attribute <i>@journal-id-type</i> must be set according to the defined values
< <b>issue-id</b> > ( http://jats.nlm.nih.gov/archiving/tag-library/1.0/index.html?elem=issue-id)	Used to store the issue ID as defined by the local DML and by the publisher. The attribute <i>@pub-id-type</i> must be set according to the defined values
< <b>article-id</b> > ( http://jats.nlm.nih.gov/archiving/tag-library/1.0/index.html?elem=article-id)	Used to store the article ID as defined by the local DML and by the publisher. The attribute <i>@pub-id-type</i> must be set according to the defined values

The <**\*-id**> elements must always be characterized by an attribute *@\*-id-type* whose value must be chosen according to the table below

#### "Primary" identifiers in bibliographic references

In the list of bibliographic references, the following element must be used, to the extent possible:

- <**issn**>
- <**isbn**>
- <**ext-link**>

<**ext-link**> element must always be characterized by an attribute *@ext-link-type* whose value must be chosen according to the table below

#### Labels for the authorities assigning primary IDs

These labels must be used in attributes such as *@journal-id-type* or *@pub-id-type* (see below).

Authority	<i>@*-id-type</i> value
EuDML ID	" <b>eudml-id</b> "
Local DML ID	" <b>partner-id</b> " (Example : " <b>numdam-id</b> "). For <i>partner</i> labels, see the two tables below
ID of the publisher	" <b>publisher-id</b> "
URL of the primary resource	" <b>url</b> "
Digital Object Identifier	" <b>doi</b> "
Publisher Item Identifier	" <b>pii</b> "
Directory of Open Access Journals	" <b>doaj</b> "

**List of assigned content partner's IDs**

<b>EuDML Partner</b>	<b>Content</b>	<b>ID</b>	<b>URL</b>
<b>IST/BNP</b>	Digitised <i>Portugaliae Mathematica</i>	pm	<a href="http://purl.pt/index/pmath/PT/index.html">http://purl.pt/index/pmath/PT/index.html</a>
<b>CMD</b>	NUMDAM	numdam	<a href="http://www.numdam.org/?lang=en">http://www.numdam.org/?lang=en</a>
<b>CMD</b>	CEDRAM	cedram	<a href="http://www.cedram.org/?lang=en">http://www.cedram.org/?lang=en</a>
<b>FIZ</b>	ELibM	elibm	<a href="http://www.emis.de/journals/">http://www.emis.de/journals/</a>
<b>MU+IMAS</b>	DML-CZ	dmlcz	<a href="http://dml.cz/">http://dml.cz/</a>
<b>ICM</b>	DML-PL	dmlpl	<a href="http://pldml.icm.edu.pl/mathbwn/">http://pldml.icm.edu.pl/mathbwn/</a>
<b>CSIC+USC</b>	DML-E	dmlle	<a href="http://dmlle.cindoc.csic.es/en/portada_en.php">http://dmlle.cindoc.csic.es/en/portada_en.php</a>
<b>EDPS</b>	EDP Sciences math. Journals	edps	<a href="http://publications.edpsciences.org/">http://publications.edpsciences.org/</a>
<b>IMI-BAS</b>	BulDML	buldml	<a href="http://sci-gems.math.bas.bg/jspui/handle/10525/2">http://sci-gems.math.bas.bg/jspui/handle/10525/2</a>
<b>IU</b>	HDML	hdml	<a href="http://www.hdml.gr">http://www.hdml.gr</a>

<b>Associated Partner</b>	<b>Content</b>	<b>ID</b>	<b>URL</b>
<b>SUB Goe</b>	GDZ Mathematica & RusDML	gdz	<a href="http://gdz.sub.uni-goettingen.de/">http://gdz.sub.uni-goettingen.de/</a>
<b>SIMAI/UMI</b>	BDIM	bdim	<a href="http://www.bdim.eu/">http://www.bdim.eu/</a>
<b>Uni Belgr</b>	Virtual Library	virtlib	<a href="http://elibrary.matf.bg.ac.rs/">http://elibrary.matf.bg.ac.rs/</a>
<b>EMS-PH</b>	JEMS	emsph	<a href="http://www.ems-ph.org/journals/journal.php?jrn=jems">http://www.ems-ph.org/journals/journal.php?jrn=jems</a>

**Example: Identifiers for a journal**

```
<journal-meta>
  <journal-id journal-id-type="eudml-id">M2AN</journal-id>
  <journal-id journal-id-type="numdam-id">M2AN</journal-id>
  <journal-id journal-id-type="url">http://www.esaim-m2an.org/</journal-id>
  <journal-title-group>
    <journal-title>ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique</journal-title>
    <abbrev-journal-title abbrev-type="short-title">ESAIM, Math. Model. Numer. Anal.</abbrev-journal-title></journal-title-group>
  <issn pub-type="ppub">0764-583X</issn>
  <issn pub-type="ppub">1290-3841</issn>
  <publisher>
    <publisher-name>EDP Sciences</publisher-name>
    <publisher-loc>Paris</publisher-loc>
  </publisher>
</journal-meta>
```

**"Document" identifiers**

"Document" identifiers are links to the different versions pertaining to an item or a container on the provider's web site (the PDF version, the full HTML version, etc.).

To set the value of this link, the **<self-uri>** (<http://jats.nlm.nih.gov/archiving/tag-library/1.0/index.html?elem=self-uri>) elements must be used in **<mbook-meta>**, **<book-meta>**, **<book-part-meta>** or **<article-meta>**.

In EuDML, **<self-uri>** elements are typically used to build a list of accessible documents related to the item.

### How to use <self-uri>

- At least one such <self-uri> is mandatory for each EuDML item
- The "main access" (aka "landing page", or persistent URI) to the item must be the first one
- When available, the MIME type should be provided in the *@content-type* attribute (except for the "main access")
- In order to ease the translation process, element value should be one of the labels listed below.

### Value of <self-uri> element

Value of <self-uri>		mime-type
<b>Access to full text</b>	Main access to the full text (typically a persistent URL associated with a resolver service)	
Full (PDF)	Direct link to the full text in PDF format	application/pdf
Full (PS)	Direct link to the full text in PostScript format	application/ps
Full (DVI)	Direct link to the full text in DVI format	application/x-dvi
Full (DjVu)	Direct link to the full text in DjVu ( <a href="http://www.djvu.org/">http://www.djvu.org/</a> )	image/x.djvu
Full (HTML)	Direct link to the full text in HTML	text/html
Full (XML)	Direct link to the full text in XML (whatever the XML format is...)	text/xml
Abstract	Direct access to the abstract, probably in HTML...	text/html
Abstract/References	Direct access to the abstract AND the references, probably in HTML...	text/html
References (BibTeX)	Direct access to the references in BibTeX format	application/x-bibtex
References	Direct access to the references in HTML format	text/html

**Note:** This list will need to be updated as needed...  
There are also cases of DVI, PS.GZ, DVI.GZ... (see elibm, e.g.)

#### Example: <self-uri> elements for an article

```
<self-uri xlink:href="http://aif.cedram.org/item?id=AIF_2007__57_4_1359_0">Access to full text</self-uri>
<self-uri content-type="text/html" xlink:href="http://aif.cedram.org/item?id=AIF_2007__57_4_1359_0">Abstract/References</self-uri>
<self-uri content-type="application/pdf" xlink:href="http://aif.cedram.org/cedram-bin/article/AIF_2007__57_4_1359_0.pdf">Full (PDF)</self-uri>
```

### "External" identifiers

"External" identifiers are identifiers of other authorities, such as Zentralblatt MATH, Math. Reviews, Crossref, which assign IDs to articles, authors, journals or books. External identifiers must be set using <ext-link> (<http://jats.nlm.nih.gov/archiving/tag-library/1.0/index.html?elem=ext-link>) elements. We describe below how and where to use these elements.

#### How to use <ext-link> for external identifiers

<ext-link> must be characterized with a *@ext-link-type* attribute identifying the authority, according to the following table:

Database	<i>@ext-link-type</i> value		
	For authors	For items (article/book /book-part)	For journals
Zentralblatt MATH	"zbl-author-id"	"zbl-item-id"	"zbl-journal-id"



( <a href="http://www.zentralblatt-math.org/zmath/en/">http://www.zentralblatt-math.org/zmath/en/</a> ) Value example:	kolmogorov.andrey-n ( <a href="http://www.zentralblatt-math.org/zmath/en/search/?q=ai:kolmogorov.andrey-n">http://www.zentralblatt-math.org/zmath/en/search/?q=ai:kolmogorov.andrey-n</a> )	05770630 ( <a href="http://www.zentralblatt-math.org/zmath/en/search/?q=an:pre05770630&amp;format=complete">http://www.zentralblatt-math.org/zmath/en/search/?q=an:pre05770630&amp;format=complete</a> )	00000134 ( <a href="http://www.zentralblatt-math.org/zmath/en/journal/?an=00000134">http://www.zentralblatt-math.org/zmath/en/journal/?an=00000134</a> )
Jahrbuch JFM ( <a href="http://www.emis.de/MATH/JFM/JFM.html">http://www.emis.de/MATH/JFM/JFM.html</a> ) Value example:	" <b>jfm-author-id</b> "  N/A	" <b>jfm-item-id</b> "  29.0522.01 ( <a href="http://www.emis.de/cgi-bin/JFM-item?29.0522.01">http://www.emis.de/cgi-bin/JFM-item?29.0522.01</a> )	" <b>jfm-journal-id</b> "  N/A
MathSciNet ( <a href="http://www.ams.org/mathscinet/">http://www.ams.org/mathscinet/</a> ) Value example:	" <b>mr-author-id</b> "  104340 ( <a href="http://www.ams.org/mathscinet/search/author.html?mrauthid=104340">http://www.ams.org/mathscinet/search/author.html?mrauthid=104340</a> )	" <b>mr-item-id</b> "  2605837 ( <a href="http://www.ams.org/mathscinet-getitem?mr=2605837">http://www.ams.org/mathscinet-getitem?mr=2605837</a> )	" <b>mr-journal-id</b> "  J_Math_Phys ( <a href="http://www.ams.org/mathscinet/search/journaldoc.html?cn=J_Math_Phys">http://www.ams.org/mathscinet/search/journaldoc.html?cn=J_Math_Phys</a> )

### Where to use <ext-link> for external identifiers

To identify..	... in the context of ...	... add a <ext-link> in
a journal	metadata	< <b>article-meta</b> > ( <a href="http://jats.nlm.nih.gov/archiving/tag-library/1.0/index.html?elem=article-meta">http://jats.nlm.nih.gov/archiving/tag-library/1.0/index.html?elem=article-meta</a> ) (forbidden by the DTD in < <b>journal-meta</b> >)
a journal	reference	< <b>source</b> > ( <a href="http://jats.nlm.nih.gov/archiving/tag-library/1.0/index.html?elem=source">http://jats.nlm.nih.gov/archiving/tag-library/1.0/index.html?elem=source</a> )
a book	metadata	< <b>book-meta</b> > ( <a href="http://eudml.mathdoc.fr/schemas/doc/book.html#element_book-meta">http://eudml.mathdoc.fr/schemas/doc/book.html#element_book-meta</a> )
a book	reference	< <b>source</b> > ( <a href="http://jats.nlm.nih.gov/archiving/tag-library/1.0/index.html?elem=source">http://jats.nlm.nih.gov/archiving/tag-library/1.0/index.html?elem=source</a> )
an article	metadata	< <b>article-meta</b> > ( <a href="http://jats.nlm.nih.gov/archiving/tag-library/1.0/index.html?elem=article-meta">http://jats.nlm.nih.gov/archiving/tag-library/1.0/index.html?elem=article-meta</a> )
an article	reference	< <b>mixed-citation</b> > or < <b>element-citation</b> > <i>at the beginning of the citation</i> (before all other elements or text)
a book part	metadata	< <b>book-part-meta</b> > ( <a href="http://eudml.mathdoc.fr/schemas/doc/book.html#element_book-part-meta">http://eudml.mathdoc.fr/schemas/doc/book.html#element_book-part-meta</a> )
a book part	reference	< <b>chapter-title</b> > ( <a href="http://jats.nlm.nih.gov/archiving/tag-library/1.0/index.html?elem=chapter-title">http://jats.nlm.nih.gov/archiving/tag-library/1.0/index.html?elem=chapter-title</a> )
an author	metadata	< <b>contrib</b> > ( <a href="http://jats.nlm.nih.gov/archiving/tag-library/1.0/index.html?elem=contrib">http://jats.nlm.nih.gov/archiving/tag-library/1.0/index.html?elem=contrib</a> )
an author	reference	< <b>string-name</b> >

## Full text

In the scope of EuDML, item's full texts can exist in many different flavours.

The most basic full text is the end-user file (usually a PDF, considered here data rather than metadata) and it is mandatorily registered with a <**self-uri**> element, see section on Document identifiers.

However, for indexing or accessibility purposes, we need other flavours of full texts, that can be more textual, or more machine-oriented representations of the item's content.

### Full NLM encoded item

The royal road to contributing such "metadata" full texts to the EuDML central system should be to just embed it into the <**body**> element with NLM encoding. A EuDML-ready OAI-PMH server could serve three formats for

each document type (which can be mechanically derived from the richest one): eudml-articlefull (resp. eudml-bookfull; EuDML with NLM-encoded full text in **<body>**), eudml-article (resp. eudml-book; same with **<body>** omitted: just front and back matter), and the mandatory oai\_dc.

### Other cases (OCR, TXT extracted from PDF, flat XML...)

When the full text is not available in NLM encoding, it should be made available to the system in the general frame of "document" identifiers with links to the relevant versions pertaining to an item. In order to remove these informations from a user interface, we must agree on special values so that the consuming system easily distinguishes user formats of the full text from machine-oriented ones.

To set the value of this link, the **<ext-link>** element must be used in **<mbook-meta>**, **<book-meta>**, **<book-part-meta>** or **<article-meta>**.

#### How to use **<ext-link>** for metadata full text

- The MIME type should be provided through **@ext-link-type** attribute in the following way `eudml-fulltext: <standard value of mime-type>`.
- The rights licensed to EuDML are listed in the **@specific-use** attribute, as a comma separated list of fixed keywords: `index, enhance, serve` is currently the full list. The right is granted to EuDML if and only if the corresponding keyword appears in the list.

**Note:** it is now under discussion whether maintaining `index` and `enhance` as separate rights is meaningful. Maybe a better scheme would be:

- `process` means that any internal processing of the harvested full-text is granted to EuDML (index text, OCR, math-extract, structure extract, index math, etc.). Given the nature of EuDML as a pilot project, this could be considered an implied default.
  - `serve` means that the result of processing might be served again by eudml itself. This will probably need further refinements, as the number of processes and the number or variations of output formats cannot be determined from the contributed formats.
- In addition, element value should be one of the labels listed below.

#### Value of **<ext-link>** element

Value of <b>&lt;ext-link&gt;</b>		mime-type	note
MetaFull (PDF)	Direct link to the full text in PDF format	application/pdf	must be unencrypted, without any extra material (cover page, copyright notice...)
MetaFull (TXT)	Direct link to the full text in raw TXT format (UTF-8 encoding)	text/plain	
MetaFull (HTML)	Direct link to the full text in HTML format	text/html	
MetaFull (XMLEuDMLFullText)	Direct link to the full text in EuDML XML format	application/xml	using eudml-fulltext schema (aka untagged UTF-8 text and math in MathML)
MetaFull (XML)	Direct link to the full text in some XML format (not recommended as DTD info is missing)	application/xml	Can be reduced to eudml-full-text by stripping off all tags outside of <b>&lt;math&gt;</b> element
MetaFull (XHTML)	Direct link to the full text in XHTML format (with MathML)	application/xhtml+xml	
MetaFull (XHTMLFromInfy)	Direct link to the full text in XHTML format (with MathML) generated by Infy	application/xhtml+xml	

MetaFull (XMLviaEudricsFromInfty)	Direct link to the full text in XML format generated by Eudrics from Infty latex output	application/xml	
MetaFull (TeXFromInfty)	Direct link to the full text in (La)TeX format generated by Infty	application/x-tex	
MetaFull (HRTeXFromInfty)	Direct link to the full text in Human Readable TeX format generated by Infty	application/x-tex	
MetaFull (IMLFromInfty)	Direct link to the full text in IML format generated by Infty	application/x-impl	
MetaFull (KMLFromInfty)	Direct link to the full text in KML format generated by Infty	application/x-kml	

**Note:** This list will need to be updated as needed...

**Example: Possible <ext-link> elements for an article's metadata full text**

```
<ext-link ext-link-type="eudml-fulltext:application/pdf" specific-use="index,enhance,serve" xlink:href="http://your.secret.url/to/full/text.pdf">MetaFull (PDF)</ext-link>
<ext-link ext-link-type="eudml-fulltext:application/txt" specific-use="index,enhance" xlink:href="http://your.secret.url/to/full/text.txt">MetaFull (TXT)</ext-link>
<ext-link ext-link-type="eudml-fulltext:application/xml" specific-use="index,enhance" xlink:href="http://your.secret.url/to/full/text.xml">MetaFull (XML)</ext-link>
<ext-link ext-link-type="eudml-fulltext:application/pdf" specific-use="index,enhance,serve" xlink:href="http://dml.cz/bitstream/handle/10338.dmlcz/121822/CasPestMa
<ext-link ext-link-type="eudml-fulltext:application/xhtml+xml" specific-use="index,enhance" href="http://your.secret.url/to/full/121822_infty.xhtml">MetaFull (XHTML)
<ext-link ext-link-type="eudml-fulltext:application/xml" specific-use="index,enhance" href="http://your.secret.url/to/full/121822_infty.xml">MetaFull (XMLviaEudricsF
<ext-link ext-link-type="eudml-fulltext:application/x-impl" specific-use="index,enhance" href="http://your.secret.url/to/full/121822_infty.impl">MetaFull (IMLFromInfty)
<ext-link ext-link-type="eudml-fulltext:application/x-kml" specific-use="index,enhance" href="http://your.secret.url/to/full/121822_infty.kml">MetaFull (KMLFromInfty)
<ext-link ext-link-type="eudml-fulltext:application/x-tex" specific-use="index,enhance" href="http://your.secret.url/to/full/121822_infty.tex">MetaFull (TeXFromInfty)
<ext-link ext-link-type="eudml-fulltext:application/x-tex" specific-use="index,enhance" href="http://your.secret.url/to/full/121822_infty.txt">MetaFull (HRTeXFromInfty)
```

**Declaring when the full text is available as open access**

In EuDML, all contributed items have a digital full text published under an eventual open access licence. Hence each full text has a date associated to it after which that full text will be freely downloadable.

This info can be delivered using the following standard NLM construct. If policy changes, the OAI-PMH record must be updated.

**Example: Declaring open access date for the full text**

```
<article-meta>
<history>
<date date-type="open-access" iso-8601-date="2015-01-01" />
</history>
</article-meta>
```

## Bibliography (citations)

See the JATS DTD documentation ( <http://jats.nlm.nih.gov/archiving/tag-library/1.0/n-v9g2.html> ).

Citations are listed as **<ref>** elements in a **<ref-list>** container (see example). Citations as gathered from providers data and are more or less structured. This structure may be enhanced by EuDML system.

Because the DTD documentation is quite comprehensive, this documentation only focuses on some points:

- The choice of the citation style
- The possible values of the **@publication-type** attribute.
- The linking inside bibliographic items
- The authors structuring

### Citation styles

Two citation styles are offered by the DTD (<http://jats.nlm.nih.gov/archiving/tag-library/1.0/n-8zg2.html>) : *mixed style* (element **<mixed-citation>**) and *element style* (element **<element-citation>**). The *mixed style* can be used for unstructured citations, while the *element style* can't. For structured citations, we could use both style. In order to preserve as much information as possible, we suggest the following rules:

#### When to use *mixed style*

*Mixed style* must be used when the original citation string must or can be preserved:

- the citation is not structured
- The citation is only partially structured (examples: authors are not structured; in a book, only the book title and the authors are structured)
- The citation is structured in a mixed way.

#### When to use *element style*

*Element style* must be used when the original citation is stored as in a database without presentation information, more generally when the original citation string is not known or cannot be reconstructed from the information available.

It is not possible in NLM DTD to store side-by-side an element and a mixed citation for the same citation. We could decide that a single **<ref>** element only contains **one** reference, but can contain one or both of **<element-citation>** and **<mixed-citation>**. This would have the same benefits as having **<string-name>** together with **<name>** elements. However, this "best practice" would clearly break compatibility with recommended JATS practice where each **<\*-citation>** element in a **<ref>** encodes a distinct reference. That would however fit better our understanding and reserve room for expected enhancer module's functionalities, as one could imagine that a matching module could find the DOI, or Zbl-ID from the string citation, thus be able to add a structured version of the same citation together with more **<ext-link>** elements.

### Value of @publication-type

When the type of the cited work is known, the attribute **@publication-type** of either **<mixed-citation>** or **<element-citation>** must be set to one of the following values:

<b>@publication-type value</b>	<b>Meaning</b>
<b>article</b>	Journal article
<b>book</b>	Book or book series
<b>conf-proc</b>	Conference proceedings
<b>thesis</b>	Work written as part of the completion of an advanced degree
<b>web</b>	Website
<b>other</b>	None of the listed types.

## Identifiers and External links added by EuDML or local DML

Identifiers for the cited resource must be added using **<ext-link>** elements at the beginning of the reference. See How and where to add Identifiers

## Authors structuring

It is recommended to always use the **<string-name>** element to store authors in citations, as this allows to add an **<ext-link>** to possible author identifier systems (including foreseen EuDML's).

## Language and transliteration issues

### Language

Language information are provided by the standard **@xml:lang** (<http://jats.nlm.nih.gov/archiving/tag-library/1.0/index.html?attr=xml:lang>) attribute.

See also <http://www.w3.org/International/questions/qa-choosing-language-tags> for background information about the value of this attribute.

We suggest to follow the rule defined in the NLM JATS documentation:

- Default value is 'EN' for some elements
- Language value is inherited for other elements (this is the behaviour specified for **@xml:lang** in the XML specification (<http://www.w3.org/TR/xml/>))

### Where xml:lang attribute should appear

- In the root element **<article>** or **<book>** (this determines the item's main language)
- In **<trans-title-group>** elements
- In **<trans-abstract>** elements

If the attribute is not set, then it is implicitly set to 'en'.

### Where xml:lang attribute may appear

**@xml:lang** may appear in the following element, when translations are available:

- In **<kwd-group>**
- In **<issue-title>**

If not present, the language for these elements is assumed to be the article language, as set in the root element.

### Where xml:lang attribute should *not* appear

- In **<trans-title>** and **<trans-sub-title>**: the attribute must be present in the parent element **<trans-title-group>** instead.
- In **<abstract>**: for translated abstracts, use **<trans-abstract>** instead. It means only one **<abstract>** should be present in the document.

## Transliteration

For items written in non-latin language, transliterated versions of titles, author names, etc., may be available, or added by EuDML. This section describes how to add transliteration information depending on the context.

### Specifying transliteration scheme

When known, the transliteration scheme must be specified using **@content-type** or another available attribute (see below). There are potentially many transliteration systems that can be used or may have been used in the past by local DMLs (see e.g. ([http://en.wikipedia.org/wiki/Romanization\\_of\\_Russian](http://en.wikipedia.org/wiki/Romanization_of_Russian))). Some transliteration schema

and the corresponding value for the attribute are gathered in this table:

Transliteration scheme	Attribute value
<i>unknown scheme</i>	translit:unknown
ALAC-LC 97 ( <a href="http://www.loc.gov/catdir/cpsa/roman.html">http://www.loc.gov/catdir/cpsa/roman.html</a> )	translit:alcl97
ISO/R 9:1968 (1968) (Cyrillic)	translit:iso9:1968
ISO 9:1995/GOST 7.79 (2002) (Cyrillic)	translit:iso9:1995
ISO 843 (Greek)	translit:iso843

### Transliterate author names

**Example: Notation of a transliterate author name**

```
<contrib-group content-type="authors">
<contrib contrib-type="author">
<name>
<surname>Егизарян</surname>
<given-names>К.М. </given-names>
</name>
<name content-type="translit:iso9:1968">
<surname>Egizarjan</surname>
<given-names>K.M. </given-names>
</name>
<name content-type="translit:iso9:1995">
<surname>Egizarân</surname>
<given-names>K.M. </given-names>
</name>
</contrib>
</contrib-group>
```

### Transliterate title

Transliterate titles must be provided using **<alt-title>**.

**Example: Notation of a transliterate article title**

```
<title-group>
<article-title>Об инвариантных аффинных связностях на алгебре Ли группы Ли</article-title>
<trans-title-group>
<trans-title xml:lang="en">Invariant affine connections in the Lie algebra of a Lie group</trans-title>
</trans-title-group>
<alt-title alt-title-type="translit:iso9:1968">Ob invariantnykh affinykh svyaznostyakh na algebre Li gruppy Li</alt-title>
<alt-title alt-title-type="translit:iso9:1995">Ob invariantnykh affinykh svâznostâh na algebre Li gruppy Li</alt-title>
</title-group>
```

### Other elements

Other elements may contain transliterate information: **<book-title>**, **<journal-title>**, **<issue-title>**. In these cases, an additional element must be added with a **@content-type** attribute value set to relevant transliteration scheme.

## Mathematical expressions

Inline and display mathematical formulae are expressed respectively with **<inline-formula>** (<http://jats.nlm.nih.gov/archiving/tag-library/1.0/index.html?elem=inline-formula>) and **<disp-formula>** (<http://jats.nlm.nih.gov/archiving/tag-library/1.0/index.html?elem=disp-formula>) elements. Both MathML and (La)TeX version of the same formula can be wrapped up using **<alternative>** element. This mechanism will be extended so that other versions (accessible, aural) can be stored in a similar fashion.

It is recommended to attach a unique ID to each formula to ease further processing.

**Example: Parallel versions of a mathematical formula**

```

<abstract>
<p>We examine a class of modular functions for
<inline-formula id="eqido">
<alternatives>
<mml:math xmlns:mml="http://www.w3.org/1998/Math/MathML">
<mml:mrow>
<mml:msup>
<mml:mi> $\Gamma$ </mml:mi>
<mml:mn>0</mml:mn>
</mml:msup>
<mml:mrow>
<mml:mo>(</mml:mo>
<mml:mi>N</mml:mi>
<mml:mo>)</mml:mo>
</mml:mrow>
</mml:mrow>
</mml:math>
<tex-math>\Gamma^0(N)</tex-math>
</alternatives>
</inline-formula>
whose values generate ring class fields of imaginary quadratic orders. This fact leads to a new algorithm for constructing elliptic curves with complex multiplication. The
<inline-formula id="eqid1">
<alternatives>
<mml:math xmlns:mml="http://www.w3.org/1998/Math/MathML">
<mml:mrow>
<mml:msub>
<mml:mi>X</mml:mi>
<mml:mn>0</mml:mn>
</mml:msub>
<mml:mrow>
<mml:mo>(</mml:mo>
<mml:mi>N</mml:mi>
<mml:mo>)</mml:mo>
</mml:mrow>
</mml:mrow>
</mml:math>
<tex-math>X_0(N)</tex-math>
</alternatives>
</inline-formula>
is not zero are overcome by computing certain modular polynomials.</p>
<p>Being a product of four
<inline-formula id="eqid2">
<alternatives>
<mml:math xmlns:mml="http://www.w3.org/1998/Math/MathML">
<mml:mi> $\eta$ </mml:mi>
</mml:math>
<tex-math>\eta</tex-math>
</alternatives>
</inline-formula>-functions, the proposed modular functions can be viewed as a natural generalisation of the functions examined by Weber and usually e
</abstract>

```

Note: For TeX notation, it is recommended to put compilable code into **<tex-math>** element. This means that the switch to math mode should be part of the element's value, and control sequences should be standard (standard meaning: macros defined in plain and latex formats, possibly with ams extensions). It is also important because some environments such as `multline` change the internal grammar of their content while switching to math, and this would be lost. The question whether tex code should be put literally in a CDATA or as an XML string with some characters such as `&` or `<` escaped is still open.

**Example: Parallel versions of a mathematical formula**

```

<display-formula id="eqido">
<alternatives>
<mml:math xmlns:mml="http://www.w3.org/1998/Math/MathML" mode="display">
<mml:mtable displaystyle="true">
<mml:mtr>
<mml:mtd columnalign="left">
<mml:mrow>
<mml:mrow>
<mml:mo>(</mml:mo>
<mml:mi>x</mml:mi>
<mml:mo>,</mml:mo>
<mml:mi>y</mml:mi>
<mml:mo>)</mml:mo>
</mml:mrow>
<mml:mo>&Element;</mml:mo>
</mml:mrow>

```

```

<mml:mo>&lbrace;</mml:mo>
<mml:mrow>
<mml:mo></mml:mo>
<mml:mn>0</mml:mn>
<mml:mo>&pm;</mml:mo>
<mml:mn>1</mml:mn>
<mml:mo></mml:mo>
</mml:mrow>
<mml:mo></mml:mo>
<mml:mo></mml:mo>
<mml:mn>4</mml:mn>
</mml:mrow>
<mml:msub>
<mml:mi>F</mml:mi>
<mml:mrow>
<mml:mn>2</mml:mn>
<mml:mi>k</mml:mi>
<mml:mo>+</mml:mo>
<mml:mn>1</mml:mn>
</mml:mrow>
</mml:msub>
<mml:msub>
<mml:mi>F</mml:mi>
<mml:mrow>
<mml:mn>2</mml:mn>
<mml:mi>k</mml:mi>
<mml:mo>+</mml:mo>
<mml:mn>2</mml:mn>
</mml:mrow>
</mml:msub>
<mml:msub>
<mml:mi>F</mml:mi>
<mml:mrow>
<mml:mn>2</mml:mn>
<mml:mi>k</mml:mi>
<mml:mo>+</mml:mo>
<mml:mn>3</mml:mn>
</mml:mrow>
</mml:msub>
<mml:mo>&pm;</mml:mo>
<mml:mfenced separators="" open="(" close="")">
<mml:mn>2</mml:mn>
<mml:msub>
<mml:mi>F</mml:mi>
<mml:mrow>
<mml:mn>2</mml:mn>
<mml:mi>k</mml:mi>
<mml:mo>+</mml:mo>
<mml:mn>1</mml:mn>
</mml:mrow>
</mml:msub>
<mml:msub>
<mml:mi>F</mml:mi>
<mml:mrow>
<mml:mn>2</mml:mn>
<mml:mi>k</mml:mi>
<mml:mo>+</mml:mo>
<mml:mn>2</mml:mn>
</mml:mrow>
</mml:msub>
<mml:mo>-</mml:mo>
<mml:mn>1</mml:mn>
</mml:mfenced>
</mml:mrow>
</mml:mtd>
</mml:mtr>
<mml:mtr>
<mml:mtd columnalign="right">
<mml:mrow>
<mml:mo>x</mml:mo>
<mml:mfenced separators="" open="(" close="")">
<mml:mn>2</mml:mn>
<mml:msubsup>
<mml:mi>F</mml:mi>
<mml:mrow>
<mml:mn>2</mml:mn>
<mml:mi>k</mml:mi>
<mml:mo>+</mml:mo>
<mml:mn>2</mml:mn>
</mml:mrow>
<mml:mn>2</mml:mn>
</mml:msubsup>
<mml:mo>+</mml:mo>
<mml:mn>1</mml:mn>
</mml:mfenced>
<mml:mfenced separators="" open="(" close="")">
<mml:mn>2</mml:mn>
<mml:msub>
<mml:mi>F</mml:mi>
<mml:mrow>
<mml:mn>2</mml:mn>

```



```

<mml:mi>k</mml:mi>
<mml:mo>+</mml:mo>
<mml:mn>2</mml:mn>
</mml:mrow>
</mml:msub>
<mml:msub>
<mml:mi>F</mml:mi>
<mml:mrow>
<mml:mn>2</mml:mn>
<mml:mi>k</mml:mi>
<mml:mo>+</mml:mo>
<mml:mn>3</mml:mn>
</mml:mrow>
</mml:msub>
<mml:mo>+</mml:mo>
<mml:mn>1</mml:mn>
</mml:mfenced>
<mml:mrow>
<mml:mo>&rbrace;</mml:mo>
<mml:mo>.</mml:mo>
</mml:mrow>
</mml:mrow>
</mml:td>
</mml:tr>
</mml:table>
</mml:math>
<tex-math>\begin{multline*}(x, y) \in \{(0 \pm 1), (4F_{2k+1} F_{2k+2} F_{2k+3} \pm \left( 2F_{2k+1} F_{2k+2} - 1 \right)) \times \left( 2F_{2k+2} \pm 1 \right) \pm \left( 2F_{2k+2} F_{2k+3} \pm 1 \right)\}.\end{multline*}</tex-math>
</alternatives>
</display-formula>

```

### Metadata producer

It is recommended to identify the origin of the metadata in the following way: put in **<custom-meta-group>** a **<custom-meta>** element where the key/value pair is provider/partner.

**Example: Identifying which partner produced metadata**

```

<custom-meta-group>
<custom-meta>
<meta-name>provider</meta-name>
<meta-value>pm</meta-value>
</custom-meta>
</custom-meta-group>

```