

Metadata for Digital Mathematical Objects

Thomas Fischer, SUB Göttingen



Overview

- Stakeholders: Who is interested?
- Metadata: To what purpose?
- Metadata beyond mathematics
- General rules
- Interoperability



Stakeholders

- Scientists
- Publishers
- Digitization Centers
- Reference Journals
- Libraries
- Archives



Stakeholders

- **Scientists:** Find research-related information
- **Publishers:** Sustain revenue
- **Digitization Centers:** Keep track of digitized material
- **Reference Journals:** Link to digitized material
- **Libraries:** ensure availability
- **Archives:** ensure longevity (e.g. by redundancy)



Purposes of Metadata

- Resource Discovery
- Interconnection
- Identification
- Collection Management
- Long term preservation



Purposes of Metadata

- **Resource Discovery:** Name, title, classification, ...
- **Interconnection:** Bibliographic reference
- **Identification:** Relate different identifiers
- **Collection Management:** Complex related entities
- **Long term preservation:** Technical information



Beyond Mathematics I

The products of digitization in mathematics should adhere to general standards.

While the choice of the internal metadata scheme is entirely up to the individual project, for interoperability standards should be supported:

- Support the OAI Protocol for Metadata Harvesting
- Support metadata standards based on the Dublin Core Metadata Initiative's elements set



Beyond Mathematics II

The products of digitization in mathematics should fit into the general digital landscape.

Two important examples are:

- Metadata for long-term preservation are developed by PREMIS
(see <http://www.oclc.org/research/projects/pmwg/>)
- OCLC builds a general “Registry for Digital Masters”. Digitized Mathematics should register there too (heavily MARC based, see <http://www.oclc.org/digitalpreservation/why/digitalregistry/>)



General Metadata Problems

- Names:
 - Order of names: John Miller or Miller, John
 - Sort order: where to put Andrew Lloyd Webber
 - Transliteration: Cebycev or Tschebyscheff?
- Journals:
 - Long and abbr. Form: uniqueness, 1-1
 - History and Development (merger, branching...)
- Digital Objects:
 - Unique and persistent identifiers needed (reliable and inexpensive)

General solution: Authority files, unique identifiers for journals, PURL, DOI, URN etc.



But:

The digitization projects cannot solve (all) the problems of library science!

- Be aware of the problems
- Try to avoid mistakes
- Be open for better solutions

And:

- Don't wait for solutions!

In general: higher granularity is better.



Use of Metadata

- Reference Journals: connect citation with identifier and review (w/ classification)
- Publishers: create citation, connect with metadata (author, title, abstract ...)
- Digitization Centers: create digital object, connect with citation
- Library: connect citation with availability



Consequence

- The citation is the central item that binds different sources of metadata together
- The stakeholder should provide mechanisms to relate citation to their identifiers (Publishers, Reference Journals, Digitization Centers, Libraries)
(Preferably with some error tolerance)



General rules: Bibliographic Data

As long as there is no general standard, the **internal** metadata should be rich enough to support emerging standards.

- Support Unicode for interoperability with foreign material
- Use separate fields for first and last name(s) (and the full name if irregular) of authors
- Record the original author and title of the article, provide translations and/or transliterations if useful
- For journals, provide long and short titles (probably using internal or external identifiers like ISSN)
- Use separate fields for volume, year, page numbers of the article

General rules: Identifiers

Bibliographic identifiers

- Provide ISBN, ISSN if available (or register new ones for old material)

Object identifiers

- Create unique (locally) and persistent identifiers for the digital resource
- Base the URLs of available versions on this identifier

Mathematical identifiers

- Provide links to Math Reviews, Zentralblatt Math and possibly other review journals
(excellent for verification of bibliographic data)



Interoperability

Digitization of mathematical journals is only useful to the different stakeholders if it ties in with the general network of mathematics

- For OAI service, provide minidml as a data format as well as different sets for different journals (and probably for MSC first two digits)
- In the web interface, make your items searchable by all the mentioned identifiers (this would make linking from reference journals essentially automatic)



Thank you for your attention!

Thomas Fischer

State and University Library (SUB) Göttingen

fischer@mail.sub.uni-goettingen.de