# Scope, Cross-Linking, Classification:

# how EuDML and Zentralblatt MATH

# can support and improve each other

## Olaf Teschke, FIZ Karlsruhe/Zentralblatt MATH

**Zentralblatt MATH**

Panel discussion, Kraków, July 2nd, 2012

**FIZ Karlsruhe**

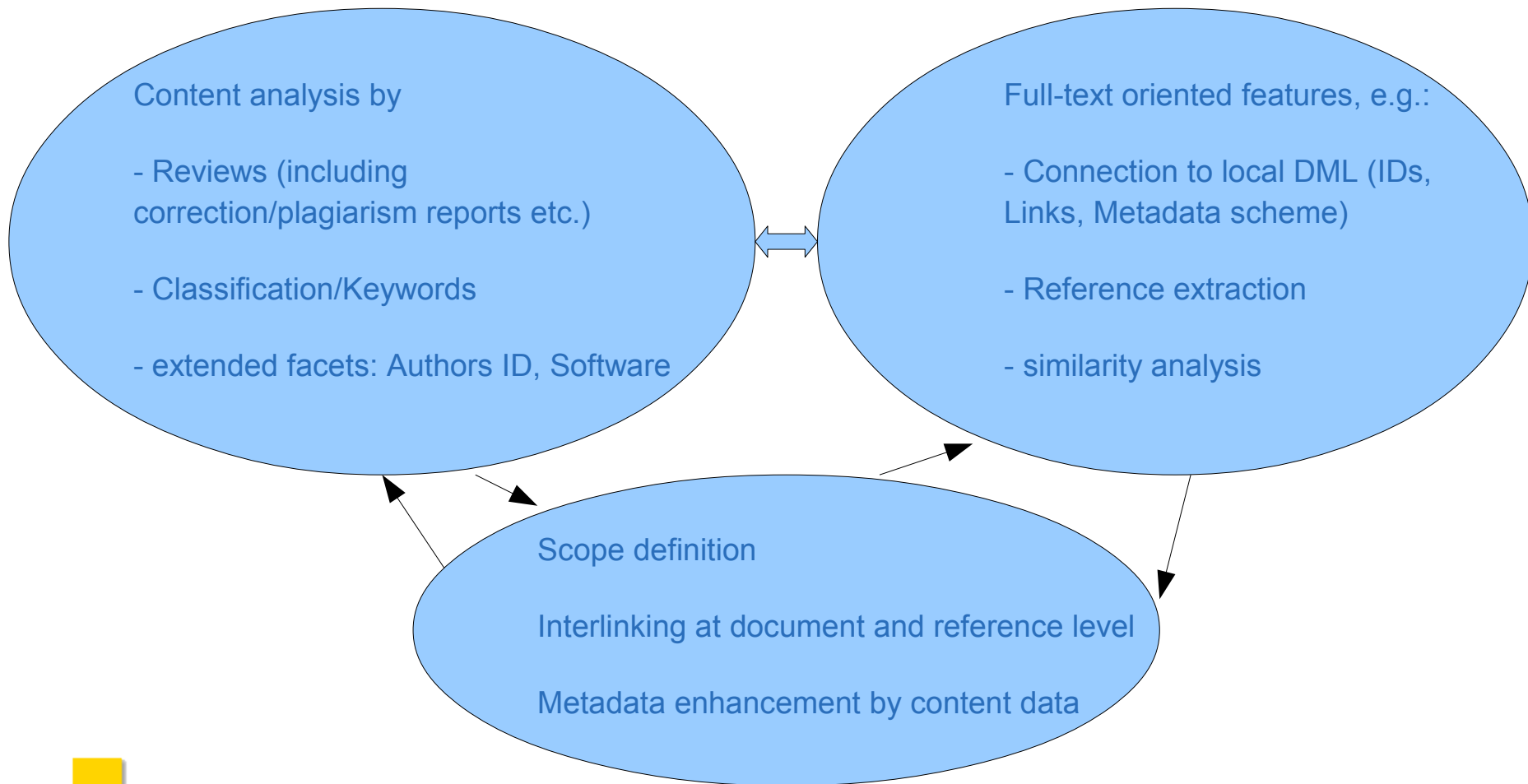Leibniz Institute for Information Infrastructure

# DML Challenges

**1. What is needed for an integration of DMLs?**

a) Digitization of content (now available for most of recent and a large part of archival content)
b) Technical standards for acessibility of mathematical content (MathML, Formula readers...) } Ongoing work in EuDML
c) Protocols for availability of future content

**2. What are the main challenges?**

a) Content: What should be included?
b) Presentation, interfaces, links } supported by zbMATH
c) Copyright (still open and probably more complicated than ever)
d) Archiving

**Zentralblatt MATH**

Panel discussion, Kraków, July 2nd, 2012

**FIZ Karlsruhe**

Leibniz Institute for Information Infrastructure

# Review database ↔ DML

**Content analysis by**

- Reviews (including correction/plagiarism reports etc.)

- Classification/Keywords

- extended facets: Authors ID, Software

**Full-text oriented features, e.g.:**

- Connection to local DML (IDs, Links, Metadata scheme)

- Reference extraction

- similarity analysis

**Scope definition**

Interlinking at document and reference level

Metadata enhancement by content data

**Zentralblatt MATH**

Panel discussion, Kraków, July 2nd, 2012

**FIZ Karlsruhe**

Leibniz Institute for Information Infrastructure

# Content: Growth of mathematics literature (zbMATH, 06/12)

Zentralblatt MATH

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure
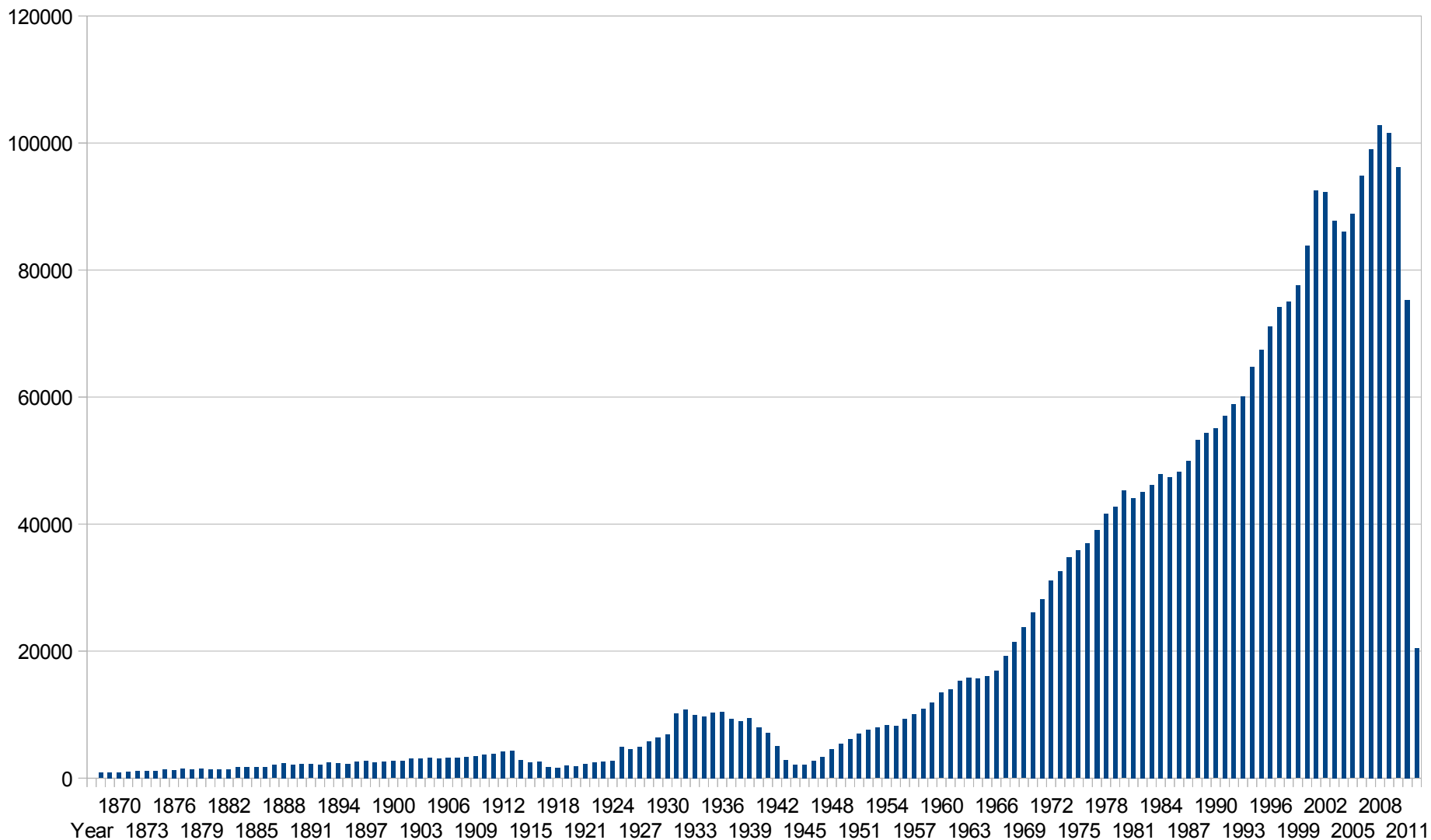
# Zentralblatt MATH provides for EuDML

**1) Scope → completeness of mathematics content:**

a) Recent/future content:

By now > 500 Journals/year apply for indexing (many without a critical mass of math content)

→ define the scope to guarantee both quality and comprehensiveness  (Use MSC!)

b) Archival content:

Europe's math treasures are scattered in different sources often hidden in libraries

→ use historical data (e.g., from Jahrbuch) to identify missing historical content

**2) Interfaces / Links**

Reference databases provide identifiers extending the scope beyond EuDML
→ by now, there are about 80.000 internal EuDML links but > 2 Million external links from references to CrossRef, MathSciNet, Zentralblatt MATH generated by interfaces
→ adds additional information from reviews, author profiles...

**3) Semantic Information; metadata enhancement**

**Zentralblatt MATH**

Panel discussion, Kraków, July 2nd, 2012

**FIZ Karlsruhe**

Leibniz Institute for Information Infrastructure

# ZBMATH author profiles



**Author's** details:

Different spellings+frequency

Author id,

Number/kind of publications

Areas of activity (by MSC)

Journals

Coauthors

Timeline

# EuDML provides for Zentralblatt MATH

**(Almost) everything useful related to full texts**

a) Stable IDs and links:

By now, many DMLs are linked from Zentralblatt MATH (e.g., > 30.000 Links zu ElibM, NUMDAM + a numer of smaller DML)

EuDML creates unique IDs and links to no less than 230.000 scattered open access articles

b) Unified metadata scheme, metadara enhancement in both directions (e.g., sources, Original titles/translations)

c) Interfaces to enhance semantic information in both directions:

Starting point: MSC in SKOS (cf. EMS Newsletter, June 2012)

Intellectual classifications + keywords from zbMATH ↔ semantics extracted from full texts in EuDML

MathML conversion enables further options (formula semantics/formula search)

**Zentralblatt MATH**

Panel discussion, Kraków, July 2nd, 2012

**FIZ Karlsruhe**

Leibniz Institute for Information Infrastructure

# zbMATH/DML - semantics

**Example:** The most frequent key phrases (of the length 4) for the MSC classes 13 and 14 derived from Zentralblatt MATH

Typically, the number of keyphrases for each MSC class is huge (>> 10.000)
MSC 13 (4 word groups)
====================
332 principal polarized abelian variety
187 smooth complex projective variety
99 complete discrete valuation ring
58 connected reductive algebraic group
49 smooth complex projective surface
47 smooth complex projective curve
41 finite dimensional vector space
35 connected linear algebraic group
34 principal polarized abelian surface
33 algebraic closed residue field

....
→ These lists could be both enhanced by full-text information and serve for datamining / (semi)automatic classification

**Zentralblatt MATH**

Panel discussion, Kraków, July 2nd, 2012

**FIZ Karlsruhe**

Leibniz Institute for Information Infrastructure

# Don't forget the rich history!

| | 2010-2011 | 2000-2009 | 1990-1999 | 1980-1989 | 1970-1979 | ⋯ | 1820-1829 | 1810-1819 |
|---|---|---|---|---|---|---|---|---|
| **2010-2011** | 2834 | 173024 | 87575 | 39909 | 23047 | ⋯ | 79 | 65 |
| **2000-2009** | 22 | 514740 | 518498 | 213640 | 118891 | ⋯ | 430 | 384 |
| **1990-1999** | 0 | 19 | 223020 | 349027 | 156424 | ⋯ | 218 | 186 |

A brief look at references in Zentralblatt MATH shows:

- Math results are relevant after centuries

- Present measures (which go back at most 10 years) forget even the peak period of influence in mathematics

**FIZ Karlsruhe**

Leibniz Institute for Information Infrastructure