

Metadata and their sources

Oldřich Ulrych, Jiří Veselý
(Charles University, Prague)

Praha, June 10–12, 2008

Launching the DML-CZ
Czech Digital Mathematics Library

1. Preparation
2. Scanning
3. Optical Character Recognition
4. **Metadata harvesting**
5. Integration
6. Digital Library

Electronic publication :

Pictures of Czech Mathematicians

Czech folk songs

Czech fairy tales published in last two centuries

Interesting Czech mathematical sources

Electronic publication :

Pictures of Czech Mathematicians

Czech folk songs

Czech fairy tales published in last two centuries

Interesting Czech mathematical sources

What it could be?

Possible material:

Journals containing mathematics published Bohemia

Proceedings of conferences related to Bohemia

What it could be?

Possible material :

Journals containing mathematics published Bohemia

Proceedings of conferences related to Bohemia

Czech mathematical books

What it could be?

Possible material :

Journals containing mathematics published Bohemia

Proceedings of conferences related to Bohemia

Czech mathematical books

What it could be?

Possible material :

Journals containing mathematics published Bohemia

Proceedings of conferences related to Bohemia

Czech mathematical books

What it could be?



Books :

Collection of works of Bernard Bolzano
Related materials

What it could be?



Books :

Collection of works of Bernard Bolzano

Related materials

Some other books

What it could be?



Books :

Collection of works of Bernard Bolzano

Related materials

Some other books

What it could be?



Books :

Collection of works of Bernard Bolzano

Related materials

Some other books

For all journals we can distinguish :

Traditionally composed volumes/issues

Retro born digital volumes/issues

Digital born volumes/issues

UTF-8:

"all alphabets together"
advantages

UTF-8:

"all alphabets together"

advantages

complications

UTF-8:

"all alphabets together"

advantages

complications

UTF-8:

"all alphabets together"

advantages

complications

Jahrbuch über die Fortschritte der Mathematik (1868-1942)

Zentralblatt für Mathematik und ihre Grenzgebiete
(since 1931)

Mathematical Reviews
(since 1940)

Jahrbuch über die Fortschritte der Mathematik
(1868-1942)

Zentralblatt für Mathematik und ihre Grenzgebiete
(since 1931)

Mathematical Reviews
(since 1940)

Referativnyj zhurnal. Matematika
(1963-1994)

Jahrbuch über die Fortschritte der Mathematik
(1868-1942)

Zentralblatt für Mathematik und ihre Grenzgebiete
(since 1931)

Mathematical Reviews
(since 1940)

Referativnyj zhurnal. Matematika
(1963-1994)

ZMath

MathSciNet

Jahrbuch über die Fortschritte der Mathematik
(1868-1942)

Zentralblatt für Mathematik und ihre Grenzgebiete
(since 1931)

Mathematical Reviews
(since 1940)

Referativnyj zhurnal. Matematika
(1963-1994)

ZMath

MathSciNet

Jahrbuch über die Fortschritte der Mathematik
(1868-1942)

Zentralblatt für Mathematik und ihre Grenzgebiete
(since 1931)

Mathematical Reviews
(since 1940)

Referativnyj zhurnal. Matematika
(1963-1994)

ZMath

MathSciNet

Title

Author

Language

Classification (MSC)

id MathSciNet

id ZMath

id Jahrbuch

Type of article

Note(s)

Link(s)

Journal, Volume, Issue, Pages

Title

Language variants, English title obligatory, original title (not Russian one), T_EX syntax, correct graphics

Author

Title

Language variants, English title obligatory, original title (not Russian one), T_EX syntax, correct graphics

Author

All authors, full names, search facilities (more rich)

Title

Language variants, English title obligatory, original title (not Russian one), T_EX syntax, correct graphics

Author

All authors, full Names, search facilities (more rich)

Language

Title

Language variants, English title obligatory, original title (not Russian one), T_EX syntax, correct graphics

Author

All authors, full Names, search facilities (more rich)

Language

basically of the article (also titles, notes, ...)

Title

Language variants, English title obligatory, original title (not Russian one), T_EX syntax, correct graphics

Author

All authors, full Names, search facilities (more rich)

Language

basically of the article (also titles, notes, ...)

Classification (MSC)

Title

Language variants, English title obligatory, original title (not Russian one), T_EX syntax, correct graphics

Author

All authors, full Names, search facilities (more rich)

Language

basically of the article (also titles, notes, . . .)

Classification (MSC)

all codes available (databases, article, other sources)

Title

Language variants, English title obligatory, original title (not Russian one), T_EX syntax, correct graphics

Author

All authors, full Names, search facilities (more rich)

Language

basically of the article (also titles, notes, . . .)

Classification (MSC)

all codes available (databases, article, other sources)

Title

Language variants, English title obligatory, original title (not Russian one), T_EX syntax, correct graphics

Author

All authors, full Names, search facilities (more rich)

Language

basically of the article (also titles, notes, . . .)

Classification (MSC)

all codes available (databases, article, other sources)

id MathSciNet

id ZMath

id Jahrbuch

always when available (a key to review, other related articles ...)

Type of article

id MathSciNet

id ZMath

id Jahrbuch

always when available (a key to review, other related articles ...)

Type of article

mathematics, physics, history (obituaries), reviews,
political material

id MathSciNet

id ZMath

id Jahrbuch

always when available (a key to review, other related articles ...)

Type of article

mathematics, physics, history (obituaries), reviews,
political material

Note(s)

id MathSciNet

id ZMath

id Jahrbuch

always when available (a key to review, other related articles ...)

Type of article

mathematics, physics, history (obituaries), reviews,
political material

Note(s)

some important message, (private notes)

id MathSciNet

id ZMath

id Jahrbuch

always when available (a key to review, other related articles ...)

Type of article

mathematics, physics, history (obituaries), reviews,
political material

Note(s)

some important message, (private notes)

Link(s)

id MathSciNet

id ZMath

id Jahrbuch

always when available (a key to review, other related articles ...)

Type of article

mathematics, physics, history (obituaries), reviews,
political material

Note(s)

some important message, (private notes)

Link(s)

continuations, corrections, remarks on the article

id MathSciNet

id ZMath

id Jahrbuch

always when available (a key to review, other related articles ...)

Type of article

mathematics, physics, history (obituaries), reviews,
political material

Note(s)

some important message, (private notes)

Link(s)

continuations, corrections, remarks on the article

Data as precise as possible!

id MathSciNet

id ZMath

id Jahrbuch

always when available (a key to review, other related articles ...)

Type of article

mathematics, physics, history (obituaries), reviews,
political material

Note(s)

some important message, (private notes)

Link(s)

continuations, corrections, remarks on the article

Data as precise as possible !

id MathSciNet

id ZMath

id Jahrbuch

always when available (a key to review, other related articles ...)

Type of article

mathematics, physics, history (obituaries), reviews,
political material

Note(s)

some important message, (private notes)

Link(s)

continuations, corrections, remarks on the article

Data as precise as possible !

None of available sources contains all data needed
(index for CMUC, contents in electronic form,...)

Any journal/proceedings requires specific approach

Certain percentage of data must be done from scratch

None of available sources contains all data needed
(index for CMUC, contents in electronic form, . . .)

Any journal/proceedings requires specific approach

Certain percentage of data must be done from scratch

Combination of approaches gives the best result

None of available sources contains all data needed
(index for CMUC, contents in electronic form, . . .)

Any journal/proceedings requires specific approach

Certain percentage of data must be done from scratch

Combination of approaches gives the best result

Numerous checking is a necessity

None of available sources contains all data needed

(index for CMUC, contents in electronic form, . . .)

Any journal/proceedings requires specific approach

Certain percentage of data must be done from scratch

Combination of approaches gives the best result

Numerous checking is a necessity

None of available sources contains all data needed

(index for CMUC, contents in electronic form, . . .)

Any journal/proceedings requires specific approach

Certain percentage of data must be done from scratch

Combination of approaches gives the best result

Numerous checking is a necessity

Czechoslovak Mathematical Journal

(1951 – 1991; 41 Vol.)

MR: 2080 **Zbl:** 2088 **Jb:** 0 **2298**

approx. 590 titles translated ($\approx 25\%$)

Commentationes Mathematicae Universitatis Carolinae

(1960 – 1990; 31 Vol.)

MR: 1803 **Zbl:** 1756 **Jb:** 0 **1901**

approx. 220 titles translated ($\approx 12\%$)

Časopis pro pěstování matematiky a fysiky

(1872 – 1950; 75 Vol.)

MR: 20 **Zbl:** 183 **Jb:** 1213 **3884**

approx. 3700 titles to be translated ($\approx 95\%$)

Czechoslovak Mathematical Journal

(1951 – 1991; 41 Vol.)

MR: 2080 **Zbl:** 2088 **Jb:** 0 **2298**

approx. 590 titles translated ($\approx 25\%$)

Commentationes Mathematicae Universitatis Carolinae

(1960 – 1990; 31 Vol.)

MR: 1803 **Zbl:** 1756 **Jb:** 0 **1901**

approx. 220 titles translated ($\approx 12\%$)

Časopis pro pěstování matematiky a fysiky

(1872 – 1950; 75 Vol.)

MR: 20 **Zbl:** 183 **Jb:** 1213 **3884**

approx. 3700 titles to be translated ($\approx 95\%$)

Czechoslovak Mathematical Journal

(1951 – 1991; 41 Vol.)

MR: 2080 **Zbl:** 2088 **Jb:** 0 **2298**

approx. 590 titles translated ($\approx 25\%$)

Commentationes Mathematicae Universitatis Carolinae

(1960 – 1990; 31 Vol.)

MR: 1803 **Zbl:** 1756 **Jb:** 0 **1901**

approx. 220 titles translated ($\approx 12\%$)

Časopis pro pěstování matematiky a fysiky

(1872 – 1950; 75 Vol.)

MR: 20 **Zbl:** 183 **Jb:** 1213 **3884**

approx. 3700 titles to be translated ($\approx 95\%$)

Czechoslovak Mathematical Journal

(1951 – 1991; 41 Vol.)

MR: 2080 **Zbl:** 2088 **Jb:** 0 **2298**

approx. 590 titles translated ($\approx 25\%$)

Commentationes Mathematicae Universitatis Carolinae

(1960 – 1990; 31 Vol.)

MR: 1803 **Zbl:** 1756 **Jb:** 0 **1901**

approx. 220 titles translated ($\approx 12\%$)

Časopis pro pěstování matematiky a fysiky

(1872 – 1950; 75 Vol.)

MR: 20 **Zbl:** 183 **Jb:** 1213 **3884**

approx. 3700 titles to be translated ($\approx 95\%$)

Metadata created as described above:

Input of *metadata editor*

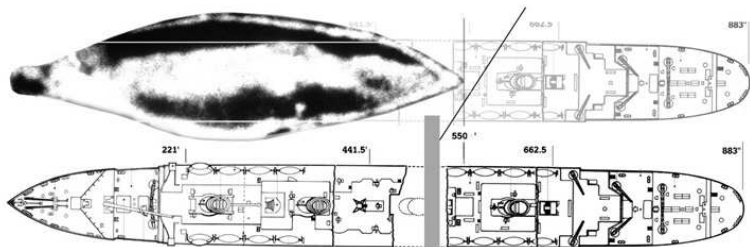
Comparison with activities in other countries

Metadata created as described above:

Input of *metadata editor*

Comparison with activities in other countries

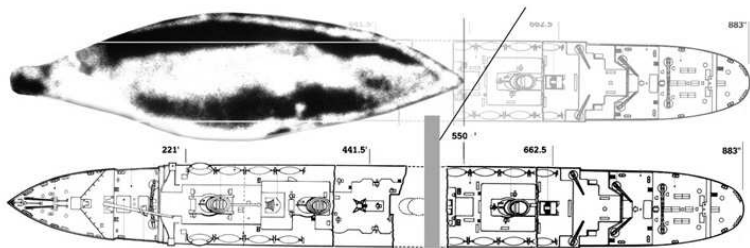
Many much better approaches exist



Ark mold formation at űzengili (Drupinar site) compared to RMS Titanic copyright © David A Deal 1998

Noah's ark was produced by amateurs while Titanic was created by highly experienced experts!

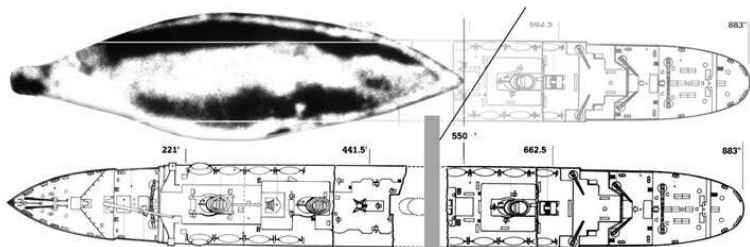
Many much better approaches exist



Ark mold formation at ūzengili (Drupinar site) compared to RMS Titanic copyright © David A Deal 1998

Noah's ark was produced by amateurs while Titanic was created by highly experienced experts!

Many much better approaches exist



Ark mold formation at űzenĝili (Drupinar site) compared to RMS Titanic copyright © David A Deal 1998

Noah's ark was produced by amateurs while Titanic was created by highly experienced experts!