

Algorithms for Integration of Unstructured Mathematical Documents into the Common Digital Space of Scientific Knowledge

Polina Gafurova¹[0000-0002-1544-155X], Alexander Elizarov²[0000-0003-2546-6897],
Evgeny Lipachev³[0000-0001-7789-2332]

¹⁻³ Institute of Information Technologies and Intelligent Systems, Kazan (Volga Region)
Federal University, Kremliovskaya ul., 35, Kazan, 420008, Russia
¹pogafurova@gmail.com, ²amelizarov@gmail.com,
³elipachev@gmail.com

Abstract. Algorithms for documents metadata formation of unstructured digital mathematical collections are proposed. They are based on the use of search queries to such open scientific resources of the World Wide Web as DBPedia, Wikidata, Wikipedia, Freebase. The developed algorithms made it possible to generate metadata in cases where it is very difficult or even impossible to extract it from document using text analytics methods. The results of applying these algorithms in the process of forming fundamental metadata sets of retro-collections included in the Lobachevskii Digital Mathematical Library ((Lobachevskii-DML, <https://lobachevskii-dml.ru/>) are presented. The composition of the main set of metadata complies with the requirements of the European Digital Mathematical Library (EuDML). The xml-language based on the Archive and Log Exchange Tag Package (NISO JATS) was used to represent the metadata.

Keywords¹: scientific information space, unified digital space of mathematical knowledge, digital mathematical library, digital mathematical library Lobachevskii-DML.

¹ CDSSK–2020: International Conference “Common Digital Space of Scientific Knowledge”, November 10–12, 2020, Moscow, Russia
EMAIL: pogafurova@gmail (Polina Gafurova); amelizarov@gmail.com (Alexander Elizarov); elipachev@gmail.com (Evgeny Lipachev);
ORCID: 0000-0002-1544-155X (Polina Gafurova); 0000-0003-2546-6897 (Alexander Elizarov); 0000-0001-7789-2332 (Evgeny Lipachev);



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0)



CEUR Workshop Proceedings (CEUR-WS.org)

1 Introduction

Currently, on the basis of the technologies of the World Wide Web, the Common Digital Space of Scientific Knowledge (CDSSK) is being actively formed. The goals of building such a space, the methods and technologies used for this, as well as precise definitions of terms used in scientific discussions are given in [1, 2].

The main directions of integration of mathematical knowledge are defined in the projects “Global Digital Mathematical Library” (GDML) and “World Digital Mathematical Library” (WDML) (see, for example, [3, 4]).

In [5–7], the specificity of obtaining mathematical results is noted: it is associated not only with painstaking calculations and operations with formulas, but also with the need for cooperation of large groups of mathematicians in long-term research, for example, in solving the problem of classifying simple finite groups. In the same works, to designate the area of creating methods and tools for supporting and integrating mathematical knowledge, the term BigMath (by analogy with the well-known term BigData) and 5 main directions of development of this area of mathematical research are proposed.

As noted in [4, 8], digital mathematical libraries are assigned the role of the main integrator of mathematical knowledge presented in scientific documents ever published. An overview of the most significant digital mathematical libraries is given in [8, 9].

Our interests are focused on building the Lobachevskii Digital Mathematical Library (Lobachevskii-DML, <https://lobachevskii-dml.ru/>). This library has been developing since 2017, its architecture is based on the principles of WDML [10, 11].

This article presents the results of research on the creation of meta-descriptions of documents of archival electronic mathematical collections that are part of Lobachevskii-DML. In this case, the schemes used in the aggregating mathematical libraries are used.

Section 2 highlights features of archival collections that limit the use of standard methods for extracting metadata.

Section 3 contains a description of the metadata schemas used to integrate relevant documents with aggregating digital libraries.

Section 4 presents algorithms for the formation of a fundamental set of metadata for documents of retro-collections of the Lobachevskii-DML. The information, which for a number of reasons turned out to be inaccessible for extraction by methods of textual and structural analysis, was supplemented from open scientific resources of the World Wide Web using a search query system.

2 Digital Archival Mathematical Collections: Problems of Metadata Formation

We refer to archival digital collections containing documents (articles from periodicals, books, preprints, collections of conference reports), published in the period before the widespread introduction of information technologies, not only in the process of creating

a document by its authors, but also in subsequent stages of the publication life cycle (see, for example, [12–15]). The dissemination of scientific knowledge during this period (usually referred to as “pre-digital”) was carried out exclusively through print media. As a result, these documents, as a rule, do not contain the attributes required for publications distributed on the Web, such as subject classifiers, keywords, annotations, and author affiliations.

A separate category is formed by historical scientific collections (retro-collections), which include scientific documents published in periodicals before the beginning of the 20th century, and Russian-language scientific publications – before the spelling reform of 1918.

The works [16, 17] describe methods of forming metadata of digital collections using the tools of the metadata factory of the digital library Lobachevskii-DML. The implementation of these tools is based on methods for analyzing the structure of documents (see, for example, [18–23]). Also, standard text analysis algorithms were applied (see, for example, [24, 25]). In [26], the results of applying metadata factory services to documents of the Lobachevskii-DML library retro-collections are presented. Metadata normalization was carried out in accordance with the xml-schemas of the mandatory EuDML set [27, 28].

Let us name the most important problems of the formation of metadata of documents of retro-collections:

- variability of the structure of documents;
- lack of subject classifiers;
- lack of keywords characterizing the research area;
- lack of annotations to articles;
- authors of articles can be indicated both at the beginning of the article and on its last page;
- for the authors of articles, only the surname and the initial letter of the name can be indicated;
- the author's surname can be supplied with a title;
- there are abbreviations of the authors' surnames to initials;
- the place of work of the author is not indicated;
- named entities in the names of theorems and references to articles in the text are given in the original language;
- references to literature are often given in footnotes without a complete bibliographic description;
- Russian-language digital retro-collections contain documents using spelling prior to the 1918 Russian language reform.

3 Semantic representation of digital mathematical collections

This section describes the basic metadata schemas used in digital math libraries to integrate their documents with aggregated digital libraries.

As is known, the most common format for the presentation of metadata is the set of metadata elements of the Dublin Core [29, 30]. Note that an extension of this set is used in the Open Archives Initiative Protocol for Metadata Harvesting [31, 32].

The Journal Archiving and Interchange Tag Suite (NISO JATS) xml schemas have been developed to represent the metadata of articles in periodicals and are used in a number of digital mathematical libraries (see, for example, [16, 33–35]). In the EuDML project based on NISO JATS V1.0, three sets of metadata are distinguished: obligatory metadata, fundamental metadata, and supplemental metadata (see [28, 36, 37]). In Fig. 1 shows a fragment of the metadata of one of the articles of the digital collection of the journal “Proceedings of the Physics and Mathematics Society at Kazan University” (in Russian, in French).

```

<article-meta>
  <article-id>14_1_1_0</article-id>
  <title-group>
    <article-title xml:lang="fr">Rapport sur les travaux de M. Pieri</article-title>
  </title-group>
  <contrib-group>
    <contrib contrib-type="author">
      <name-alternatives>
        <name>
          <surname xml:lang="fr">Peano</surname>
          <string-name xml:lang="fr">G. Peano</string-name>
        </name>
      </name-alternatives>
    </contrib>
  </contrib-group>
  <volume>14</volume>
  <volume-series>2</volume-series>
  <pub-date>
    <year>1904</year>
  </pub-date>
  <issue>1</issue>
  <issue-part>1</issue-part>
</article-meta>

```

Fig. 1. Fragment of the metadata set of G. Peano's article “Rapport sur les travaux de M. Pieri” published in the journal “Proceedings of the Physics and Mathematics Society at Kazan University” (series 2, volume 14, number 1, 1904). The metadata is generated according to the NISO JATS schema using the Lobachevskii-DML metadata factory tools. Extracted from the text of the document: the title of the article in French, the surname and initial of the author, the numbers of the initial and final pages of the document.

4 Algorithms for extracting metadata by the queries method to the Common Digital Scientific Space

This section provides algorithms for the formation of a fundamental set of metadata of the digital library Lobachevskii-DML retro-collections documents. Information, which for a number of reasons turned out to be inaccessible for extraction by methods of textual and structural analysis, is replenished from open scientific resources of the World Wide Web through the search query system.

Using the tools of the metadata factory of the digital library Lobachevskii-DML, digitized documents of retro collections are processed. As a result, it is possible to extract the following metadata from the texts:

- title of the publication (“journal-title-group” and “trans-title-group”),
- time and place of publication (“publisher”, “pub-date”),
- title of the article (“title-group”) in one of the languages (pre-reform Russian, English, German, French),
- the author's surname with initials or with only one initial (“contrib-group”),
- volume and issue number of a periodical (“volume”, “issue”),
- numbers of the first and last pages of the publication (“star-page”, “end-page”).

Processing footnotes in the text of an article allows you to obtain information about the title and authors (or only about one of the possible authors) of those articles to which the links point.

For additional information, in particular, for the formation of a fundamental set of metadata according to the EuDML schema, the following algorithm is proposed.

The available metadata is converted to csv format. The surnames of the authors and the titles of the articles are supplemented with variants of their writing in modern Russian (if pre-reform spelling is used in the document), and transliteration is also performed. This is necessary for the formation of templates for search queries.

At the next stage, search queries are formed, including the templates obtained at the previous stage.

Further, standard operations are performed: extracting, cleaning, integration, similarity, entity resolution, analysis of the resulting data (see, for example, [38]).

With the help of the generated search queries, you can clarify (or supplement) information about the names and patronymics of authors, place of work, years of their publication activity, add URLs of sites containing biographies and other data about authors.

The above approach turned out to be effective only in those cases when the authors of documents of retro-collections are reflected in the scientific network space.

To clarify, as well as replenish the already formed metadata, open scientific resources were used, in particular, Wikipedia, Wikidata, DBPedia and Freebase. To search and retrieve information from the Web, we used the wikipedia package tool (<https://wikipedia.readthedocs.io/en/latest/code.html>) and pywikibot (<https://doc.wikimedia.org/pywikibot/master/index.html>), as well as a system of SPARQL-queries to the specified network resources

Next, we present an algorithm for extracting metadata from the open scientific resources of the World Wide Web, with the processing of the data obtained (see Algorithm 1).

At the input of the algorithm, we are fed the set

$$\mathbf{M} = \{d_1.xml, d_2.xml, \dots, d_m.xml\}.$$

This set consists of files with metadata extracted from the documents of the retro collection and we present it in Fig. 1 fragment of one such file. The metadata contained in these files is generated using the Lobachevskii-DML metadata factory tools. As we pointed out earlier, these metadata sets are incomplete. As indicated in the schemes of integrating digital mathematical libraries, it is required to present a significantly larger amount of meta-information about scientific documents. In particular, this metadata is insufficient to form a fundamental set according to XMLSchema EuDML [28].

Algorithm 1: Extracting metadata from open scientific Web sources

```

1. load M=[d1.xml, d2.xml,..., dm.xml]
2. for d in M:
    # Parsing XML-tree:
3.   md=parse(d).getroot()
4.   results_list=[]
    # Find a group of tags with data about the authors
    # (NISO JATS schema):
5.   for authors in md.findall("./contrib_group/
                               [@content_type='authors']"):
        # Select a group of tags for each author:
6.       for author in md.findall("./contrib/
                                   [@contrib_type='author']"):
            # Find a tag indicating the author's surname and initials:
7.           name_author_in_paper=author.find('string-name')
            # Translation and transliteration of the author's surname and initials:
8.           if language(name_author_in_paper)=='ru':
9.               name_author_en=transliterate()
10.          else
11.              name_author_ru=translate_ru()
12.              if language(name_author_in_paper)=='ru-old':
13.                  name_author_ru=translate_ru_old()
14.                  name_author_en=transliterate()
            # Form a list of patterns for search queries:
15.           patterns=pattern_list()
            # Selecting and connecting to a network access point
            # (Wikipedia, Wikidata, DBPedia):
16.           point=point_connect()
17.           results=[]
            # Run searches with each of the templates:
18.           for p in patterns:
19.               result=search(p)
                # Processing search results:
20.               extracting(result)
21.               cleaning(result)
22.               similarity(result)
                #adding metainformation from the web
23.               results.append(result)
24.           end for
            # Writing new metadata according to XMLSchema rules:
25.           normalization(results)
            # Adding metainformation of author d in results list
26.           results_list.append(results)
27.       end for

```

```

28. end for
    # Save information about paper
29. save(result_list)
30. end for

```

The following figures (Figs. 2 and 3) shows a fragment of the received metadata and the result of their normalization.

```

<article-meta>
  <article-id>14_1_1_0</article-id>
  <title-group>
    <article-title xml:lang="fr">Rapport sur les
      travaux de M. Pieri</article-title>
  </title-group>
  <contrib-group>
    <contrib contrib-type="author">
      <name-alternatives>
        <name>
          <surname xml:lang="fr">Peano</surname>
          <surname xml:lang="ru">Пеано</surname>
          <given-names xml:lang="fr">G.</given-names>
          <given-names xml:lang="it">Giuseppe
            </given-names>
          <given-names xml:lang="ru">Джузеппе
            </given-names>
        </name>
        <string-name xml:lang="fr">G. Peano
          </string-name>
        <string-name xml:lang="it">Giuseppe Peano
          </string-name>
        <string-name xml:lang="ru">Джузеппе Пеано
          </string-name>
      </name-alternatives>
      <aff xml:lang="fr">University of Turin</aff>
    </contrib>
  </contrib-group>
  <volume>14</volume>
  <volume-series>2</volume-series>
  <pub-date>
    <year>1904</year>
  </pub-date>
  <issue>1</issue>
  <issue-part>1</issue-part>
</article-meta>

```

Fig. 2. A fragment of the fundamental set of metadata formed by Algorithm 1. The previously formed set (see Fig. 1) is supplemented with information about the author of the article.

E	F	G	H	I	J	K	L
автор	исходное название	исходное название статьи	WikidataURI	MathN ZbMATHAuthorID	OpenLibraryID		
H. Poincare	Rapport sur les travaux de M. Hilbert	Q81082		poincare.henri	OL7476098A		
P. Mansion	Rapport sur les travaux de M. Barbarin	null		mansion.paul	OL3775794A		
C. A. Laisant	Rapport sur les travaux de M. Lemoine	Q25318		laisant.ch-a	OL2426857A		
G. Peano	Rapport sur les travaux de M. Pieri	Q191029		peano.giuseppe	OL32329A		

Fig. 3. Additional metadata extracted from online sources. In particular, the url of pages containing references to the document in question was obtained.

5 Conclusion

As you know, digital mathematical libraries are assigned today the role of the main integrator of mathematical knowledge presented in scientific documents ever published. Our interests are focused on building a digital library, the Lobachevskii Digital Mathematical Library, based on the principles of WDML. This paper presents the results of research on the creation of meta-descriptions of documents of archival digital mathematical collections included in Lobachevskii-DML. Using search queries to such open scientific resources of the World Wide Web as DBPedia, Wikidata, Wikipedia, Freebase, algorithms for the formation of metadata of documents of unstructured digital mathematical collections are proposed. This metadata meets the requirements of “The European Digital Mathematics Library”. An xml language based on the Journal Archiving and Interchange Tag Suite was used to represent metadata. The developed algorithms made it possible to generate metadata in cases where it is impossible to extract them from document texts using text analytics methods.

This work was financially supported by the Russian Science Foundation (project No. 21-11-00105).

References

1. Kalenov, N.E., Savin, G.I, Sotnikov, A.N.: The Architecture of the Common Digital Space of Scientific Knowledge. In: N.E. Kalenov, A.N. Sotnikov (Eds.). *Edinoe cifrovoe prostranstvo nauchnyh znaniy: problemy i resheniya: sbornik nauchnyh trudov.* Direktmedia Publishing. P. 7–16. Moscow, Berlin (2021). <https://doi.org/10.23681/610687>
2. Ataeva, O.M., Kalenov, N.E., Serebryakov, V.A, Sotnikov A.N.: Functionality of a Common Digital Space of Scientific Knowledge. In: N.E. Kalenov, A.N. Sotnikov (Eds.). *Edinoe cifrovoe prostranstvo nauchnyh znaniy: problemy i resheniya: sbornik nauchnyh trudov,* Direktmedia Publishing. P. 89–107. Moscow, Berlin (2021). <https://doi.org/10.23681/610687>
3. Ion, P.D.F, Watt, S.M.: The Global Digital Mathematics Library and the International Mathematical Knowledge Trust. In: H. Geuvers, M. England, O. Hasan, F. Rabe, O. Teschke (Eds.). *Intelligent Computer Mathematics – CICM 2017, volume 10383 of Lecture Notes in Computer Science,* Springer, Cham. P. 56–69 (2017). https://doi.org/10.1007/978-3-319-62075-6_5

4. Developing a 21st Century Global Library for Mathematics Research, DC: The National Academies Press, Washington. <https://doi.org/10.17226/18619>
5. Carette, J., Farmer, W.M., Kohlhase, M., Rabe, F.: Big Math and the One-Brain Barrier A Position Paper and Architecture Proposal. arXiv:1904.10405v2. 2019. URL: <https://arxiv.org/pdf/1904.10405.pdf>, last accessed 2021/09/09
6. Bercic, K., Carette, J., Farmer, W.M., Kohlhase, M., Müller, D., Rabe, F., Sharoda, Y.: The Space of Mathematical Software Systems – A Survey of Paradigmatic Systems. arXiv:2002.04955v1. 2020. URL: <https://arxiv.org/pdf/2002.04955.pdf>, last accessed 2021/09/09
7. Carette, J., Farmer, W.M., Kohlhase, M., Rabe, F.: Big Math and the One-Brain Barrier: The Tetrapod Model of Mathematical Knowledge. *Math. Intelligencer* 43. P. 78–87 (2021). <https://doi.org/10.1007/s00283-020-10006-0>
8. Borwein, J.M., Rocha, E.M., Rodrigues, J.F. (Ed.): *Communicating Mathematics in the Digital Era*. A K Peters, Ltd., CRC Press (2008).
9. Elizarov, A.M., Lipachev, E.K., Zuev, D.S.: Digital Mathematical Libraries: Overview of Implementations and Content Management Services. In: *CEUR Workshop Proceedings 2022*. P. 317–325 (2017). URL: <http://ceur-ws.org/Vol-2022/paper49.pdf>, last accessed 2021/07/24
10. Elizarov, A.M., Lipachev, E.K.: Lobachevskii DML: Towards a Semantic Digital Mathematical Library of Kazan University. In: *CEUR Workshop Proceedings 2022*. P. 326–333 (2017). URL: <http://ceur-ws.org/Vol-2022/paper50.pdf>, last accessed 2021/07/24
11. Elizarov, A.M., Lipachev, E.K.: Big Math Methods in Lobachevskii-DML Digital Library. In: *CEUR Workshop Proceedings 2523*. P. 59–72 (2019). URL: <http://ceur-ws.org/Vol-2523/invited08.pdf>, last accessed 2021/07/24
12. Elizarov A.M., Zuev D.S., Lipachev E.K. Lifecycle Management of Electronic Publications in Information Systems Scientific Journal. *Proceedings of Voronezh State University. Series: Systems Analysis and Information Technologies*. (4). P. 81-88 (2014).
13. Bartling S., Friesike S.: Towards Another Scientific Revolution. In: Bartling S., Friesike S. (Eds.) *Opening Science. The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing*. Springer International Publishing. 2014. P. 3–15 (2014). https://doi.org/10.1007/978-3-319-00026-8_1
14. Binfield, P.: Novel Scholarly Journal Concepts. In: Bartling S., Friesike S. (Eds.) *Opening Science. The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing*. Springer International Publishing. P. 155–163 (2014). https://doi.org/10.1007/978-3-319-00026-8_10
15. Henitiuk, V., O’Sullivan, C.: Aims and Scope: Journal Identity and Twenty-First-Century Scholarly Publishing. In: *Translation and Academic Journals. The Evolving Landscape of Scholarly Publishing*. Sun Y. (Ed.) Palgrave Macmillan US. P. 15–35 (2015). https://doi.org/10.1057/9781137522092_2
16. Gafurova, P.O., Elizarov, A.M., Lipachev, E.K.: Basic Services of Factory Metadata Digital Mathematical Library Lobachevskii-DML. *Russian Digital Libraries Journal* 23 (3). P. 336–381 (2020). <https://doi.org/10.26907/1562-5419-2020-23-3-336-381>
17. Elizarov, A.M., Lipachev, E.K.: Digital Library Metadata Factories. In: *CEUR Workshop Proceedings 2813*. P. 13–21 (2021). URL: <http://ceur-ws.org/Vol-2813/rpaper01.pdf>, last accessed 2021/07/24

18. Biryaltsev E., Elizarov A., Zhil'tsov N., Lipachev E., Nevzorova O., Solov'ev V.: Methods for Analyzing Semantic Data of Electronic Collections in Mathematics, Automatic Documentation and Mathematical Linguistics 48 (2). P. 81–85 (2014). <https://doi.org/10.3103/S000510551402006X>
19. Tkaczyk D., Tarnawski B., Bolikowski Ł. Structured Affiliations Extraction from Scientific Literature. D-Lib Magazine. 21(11/12) (2015). <https://doi.org/10.1045/november2015-tkaczyk>
20. Ronzano F., Saggion, H.Dr.: Inventor Framework: Extracting Structured Information from Scientific Publications. In: Japkowicz N., Matwin S. (Eds.) Discovery Science. Lecture Notes in Computer Science. Springer, Cham. 9356 (2015). https://doi.org/10.1007/978-3-319-24282-8_18
21. Elizarov, A.M., Lipachev, E.K., Khaidarov, S.M.: Automated Processing Service System of Large Collections of Scientific Documents. CEUR Workshop Proceedings. 1752. P. 58–64 (2016).
22. Tkaczyk, D.: New Methods for Metadata Extraction from Scientific Literature. arXiv:1710.10201v1. 2017. URL: <https://arxiv.org/pdf/1710.10201v1.pdf>, last accessed 2021/09/09
23. Elizarov, A.M., Khaidarov, Sh.M., Lipachev, E.K.: Scientific Documents Ontologies for Semantic Representation of Digital Libraries. In: Proceedings of the 2nd Russia and Pacific Conference on Computer Technology and Applications (RPC 2017). IEEE. P. 1–5 (2017). <https://doi.org/10.1109/RPC.2017.8168064>
24. Ingersoll, G.S., Morton, T.S., Farris, A.L.: Taming Text. How to Find, Organize, and Manipulate It, Manning Publications Co., 2013.
25. Lane, H., Hapke, H., Howard, C.: Natural Language Processing in Action: Understanding, analyzing, and generating text with Python. Manning Publications, 2019.
26. Gafurova, P.O., Elizarov, A.M., Lipachev, E.K.: Algorithms for Formation of Metadata Mathematical Retro Collections Based on Analysis of Structural Features of Documents. Russian Digital Libraries Journal. 24 (2), P. 238–271 (2021). <https://doi.org/10.26907/1562-5419-2021-24-2-238-270>. URL: <https://elbib.ru/article/view/671/764>, last accessed 2021/09/09
27. Jost, M., Bouche, T., Goutorbe, C., Jorda, J.P.: D3.2: The EuDML metadata schema. Revision: 1.6 as of 15th December 2010. URL: <http://www.mathdoc.fr/publis/d3.2-v1.6.pdf>, last accessed 2021/09/09
28. EuDML metadata schema specification (v2.0–final). URL: <https://initiative.eudml.org/eudml-metadata-schema-specification-v20-final>, last accessed 2021/09/09
29. Dublin Core™ Metadata Initiative. URL: <http://dublincore.org/>, last accessed 2021/09/09
30. Xie, I., Matusiak, K.K. Discover Digital Libraries: Theory and Practice. Elsevier Inc. 388 p. (2016).
31. The Open Archives Initiative Protocol for Metadata Harvesting. URL: <https://www.openarchives.org/OAI/openarchivesprotocol.html>, last accessed 2021/09/09
32. Müller, U.: OAI-PMH Implementation. URL: <http://eprints.rclis.org/4586/1/tutorial3muller.pdf>, last accessed 2021/09/09
33. Bouche, T., Labbe, O.: The New Numdam Platform. In: H. Geuvers, M. England, O. Hasan, F. Rabe, O. Teschke (Eds.), Intelligent Computer Mathematics – CICM 2017, volume 10383 of Lecture Notes in Computer Science, Springer, Cham. P. 70–82 (2017). https://doi.org/10.1007/978-3-319-62075-6_6

- URL: <https://zenodo.org/record/581405>, last accessed 2021/07/24
34. Elizarov, A.M., Zaitseva, N.V., Zuev, D.S, Lipachev, E.K, Khaidarov, S.M.: Services for Formation of Digital Documents Metadata in the Formats of International Science-based Databases. In: CEUR Workshop Proceedings. 2260. P. 175–185 (2018). URL: http://ceur-ws.org/Vol-2260/53_175-185.pdf, last accessed 2021/07/24.
 35. Gafurova, P.O., Elizarov, A.M., Lipachev, E.K., Khammatova, D.M.: Metadata Normalization Methods in the Digital Mathematical Library. In: CEUR Workshop Proceedings. 2543. P. 136–148 (2020). URL: <http://ceur-ws.org/Vol-2543/rpaper13.pdf>, last accessed 2021/07/24
 36. Sylwestrzak, W., Borbinha, J., Bouche, T., Nowinski, A., Sojka, P.: EuDML – Towards the European Digital Mathematics Library. In: P. Sojka (Ed.): Towards a Digital Mathematics Library, Paris, July 7 – 8th 2010, Masaryk University Press, Brno. P. 11–26 (2010). URL: http://dml.cz/bitstream/handle/10338.dmlcz/702569/DML_003-2010-1_5.pdf, last accessed 2021/07/24
 37. Bouche, T.: Reviving the Free Public Scientific Library in the Digital Age? The EuDML Project. In: K. Kaiser, S. Krantz, B. Wegner (Eds.): Topics and Issues in Electronic Publishing, JMM, Special Session, San Diego. P. 57–80 (2013). URL: <http://www.emis.de/proceedings/TIEP2013/05bouche.pdf>, last accessed 2021/07/24
 38. Knoblock, C.A., Szekely, P.: A scalable architecture for extracting, aligning, linking, and visualizing multi-Int data. Proc. SPIE 9499, Next-Generation Analyst III, 949907 (15 May 2015). <https://doi.org/10.1117/12.2177119>.