

УДК 004.021 + 004.42

## БАЗОВЫЕ СЕРВИСЫ ФАБРИКИ МЕТАДАННЫХ ЦИФРОВОЙ МАТЕМАТИЧЕСКОЙ БИБЛИОТЕКИ LOBACHEVSKII-DML

П. О. Гафурова<sup>1[0000-0002-1544-155X]</sup>, А. М. Елизаров<sup>2[0000-0003-2546-6897]</sup>,  
Е. К. Липачёв<sup>3[0000-0001-7789-2332]</sup>

<sup>1-3</sup>*Высшая школа информационных технологий и интеллектуальных систем Казанского федерального университета, ул. Кремлевская, 35, г. Казань, 420008*

<sup>2,3</sup>*Институт математики и механики им. Н.И. Лобачевского Казанского федерального университета, ул. Кремлевская, 35, г. Казань, 420008*

<sup>1</sup>pogafurova@gmail.com, <sup>2</sup>amelizarov@gmail.com, <sup>3</sup>elipachev@gmail.com

### **Аннотация**

Решен ряд задач, связанных с построением фабрики метаданных цифровой математической библиотеки Lobachevskii-DML. Под фабрикой метаданных понимается система взаимосвязанных программных инструментов, направленных на создание, обработку, хранение и управление метаданными объектов цифровых библиотек и позволяющих интегрировать создаваемые электронные коллекции в агрегирующие цифровые научные библиотеки. С целью выбора оптимальных таких программных инструментов из существующих и их модернизации:

- обсуждены особенности представления метаданных документов различных электронных коллекций, связанные как с применяемыми форматами, так и с изменениями состава и полноты набора метаданных в течение всего времени издания соответствующего научного журнала;
- представлены и охарактеризованы программные инструменты управления научным контентом и методы организации автоматизированной интеграции репозиториев математических документов с другими информационными системами;
- обсуждена такая важная функция фабрики метаданных цифровой библиотеки, как нормализация метаданных в соответствии с форматами других агрегирующих библиотек.

В результате разработки фабрики метаданных цифровой математической библиотеки Lobachevskii-DML предложена система сервисов автоматизированного формирования метаданных электронных математических коллекций; разработан xml-язык представления метаданных, основанный на Journal Archiving and Interchange Tag Suite (NISO JATS); созданы программные инструменты нормализации метаданных электронных коллекций научных документов в форматах, разработанных международными организациями – агрегаторами ресурсов по математике и Computer Science; разработан алгоритм приведения метаданных к формату oai\_dc и генерации структуры архивов для импорта в цифровое хранилище DSpace; предложены и реализованы методы интеграции электронных математических коллекций Казанского университета в отечественные и зарубежные цифровые математические библиотеки.

*Ключевые слова:* цифровые библиотеки, цифровая математическая библиотека, формирование метаданных, извлечение метаданных, нормализация метаданных, фабрика метаданных, NISO JATS, семантические связи, Lobachevskii-DML.

## **ВВЕДЕНИЕ**

Как известно, навигация в научном информационном пространстве в значительной степени обеспечивается сегодня наличием и полнотой набора метаданных документов, представленных в сети [1–5]. Важной составляющей информационного научного пространства являются цифровые библиотеки (см., например, [6]). В области математических наук создано значительное число цифровых библиотек, выполняющих разнообразные функции интеграции математических знаний [7–12]. Обзор специфики и функциональных возможностей ряда существующих цифровых математических библиотек содержится в [13].

Комитетом по электронной информационной коммуникации (Committee on Electronic Information Communication, CEIC) Международного математического союза (International Mathematical Union, IMU) в 2002 году был подготовлен документ [14], определяющий лучшие практики распространения результатов математических исследований, которые учитывают новые достижения в области цифровых коммуникаций. Российских математиков в CEIC представлял академик А.Б. Жижченко. Среди первых цифровых математических библиотек, созданных в

соответствии с практиками, обозначенными в [14], наиболее крупными являются MathNet.Ru (<http://www.mathnet.ru/>), Numdam (<http://www.numdam.org/>) и DML-CZ (Czech Digital Mathematics Library, <https://dml.cz/>). Результаты их формирования и развития к 2013 году представлены в работах [15–19]. Одновременно для информационной поддержки цифровых библиотек, в том числе математических, стали разрабатываться методы обработки документов, основанные на семантических связях объектов, выделенных из их контента [20–27].

Заметным шагом в цифровизации математических знаний стали инициативы “World Digital Mathematics Library” (WDML, «Всемирная цифровая математическая библиотека», <https://www.mathunion.org/ceic/library/world-digital-mathematics-library-wdml>, 2012 год; предысторию этой инициативы подробно описал Peter J. Olver, см. [http://www-users.math.umn.edu/~olver/t\\_/wdmlb.pdf](http://www-users.math.umn.edu/~olver/t_/wdmlb.pdf)) и “The Global Digital Mathematics Library” (GDML, «Глобальная цифровая математическая библиотека», 2014 год). Эти инициативы направлены на создание методологии интеграции математических знаний [28–31]. Основная их цель состоит в разработке принципов построения информационной среды, предоставляющей доступ ко всем когда-либо опубликованным работам по математике с помощью системы интеллектуальных агентов, обеспечивающих навигацию в информационном пространстве.

В направлении, обозначенном названными выше инициативами, инициирован и реализован ряд значимых международных проектов. Например, проект “The European Digital Mathematics Library” – «Европейская Цифровая Математическая Библиотека» (EuDML, <https://initiative.eudml.org/>) – направлен на интеграцию математических ресурсов европейских цифровых библиотек [32–34]. Далее, благодаря реализации проекта MathNet.Ru (<http://www.math-net.ru/>) оцифрованы, снабжены метаданными и представлены в открытый доступ архивы многих российских математических научных журналов и других изданий. На портале этого проекта предложены к использованию разработанные в рамках проведенных исследований методы навигации и расширенного поиска по математическому контенту, а также указаны возможности организации системы связей с международными библиографическими базами данных [15–17]. В Казанском университете,

начиная с 2017 года, в соответствии с базовыми принципами WDML и GDML создается цифровая математическая библиотека Lobachevskii Digital Mathematical Library (Lobachevskii-DML, <https://lobachevskii-dml.ru/>) [35–37].

В настоящее время в связи со значительным ростом объемов научных публикаций появилась необходимость разработки специализированных методов автоматизированной обработки больших массивов документов [20, 25, 38–41]. Одним из результатов, полученных в этом направлении, является ряд разработанных семантических методов обработки математического контента [42–48]. Семантическая составляющая большинства используемых здесь методов основана на применении специализированных онтологий. Отметим, в частности, спроектированную и развиваемую в настоящее время цифровую экосистему OntoMath [45, 46]: она включает семантические поисковые инструменты [47, 48], онтологии профессиональной [42] и образовательной математики [49–51], рекомендательные системы, ориентированные на специфику математического контента [52–57], и набор инструментов для работы с математическими документами. Все перечисленное составляет основу фабрики метаданных цифровой математической библиотеки, разрабатываемой нами в настоящее время.

Мы используем термин *фабрика метаданных цифровой библиотеки* (*metadata factory of digital library*) в том же смысле, в каком он использован в [25], а именно: фабрика метаданных – это система взаимосвязанных программных инструментов, направленных на создание, обработку, хранение и управление метаданными объектов цифровых библиотек и позволяющих интегрировать создаваемые электронные коллекции в агрегирующие цифровые научные библиотеки. С помощью этих инструментов преимущественно в автоматизированном режиме выполняются такие операции, как выделение объектов и связей между ними, экстракция метаданных из различных источников и конкретных документов, верификация, уточнение, улучшение, нормализация в различных форматах и гармонизация метаданных с помощью ручного редактирования или автоматизированных агентов, хранение и связывание метаданных с внешними базами данных. В случае цифровой математической библиотеки к перечисленным инструментам добавляется ряд специализированных, таких, например, как преобразование в формат MathML, разметка математических формул и организация поиска по ним.

---

Для обозначения методов формирования и преобразования метаданных документов в соответствии с правилами и XML-схемами цифровых библиотек и наукометрических баз данных мы используем термин *нормализация* (см. [25, 58–60]).

В настоящей статье представлены базовые сервисы фабрики метаданных цифровой математической библиотеки Lobachevskii-DML. Для решения задачи интеграции информационных ресурсов предложены методы преобразования метаданных электронных коллекций и содержащихся в них документов по DTD-правилам и XML-схемам Journal Archiving and Interchange Tag Suite (NISO JATS) различных версий [61–66]. Так, например, для интеграции с ресурсами, сформированными в рамках проекта EuDML, необходимо, чтобы метаданные интегрируемых электронных коллекций были представлены в соответствии с EuDML-схемой на основе NLM JATS версии 1.0. Одновременно должен быть создан отдельный OAIPMH-сервер для реализации EuDML-харвестинга. Отметим также, что разработанный метод нормализации метаданных электронных коллекций цифровой библиотеки Lobachevskii-DML по правилам NISO JATS послужил основой для формирования обязательного и фундаментального наборов метаданных по схемам EuDML. Кроме того, в статье приведен алгоритм автоматизированной подготовки метаданных электронной коллекции статьей журнала «Электронные библиотеки» (“Russian Digital Libraries Journal”, <https://elbib.ru/>) по правилам библиографической базы по компьютерным наукам “dblp computer science bibliography” (DBLP, <https://dblp.uni-trier.de/>).

Статья организована следующим образом.

В Разделе 1 обсуждены особенности представления метаданных документов различных электронных коллекций, связанные не только с применяемыми форматами, но и с изменениями состава и полноты набора метаданных в течение всего времени издания соответствующего научного журнала.

В Разделе 2 представлены программные инструменты управления научным контентом. Эти инструменты используются и фабрикой метаданных для создания, обработки, хранения и управления метаданными электронными документами и позволяют интегрировать создаваемые электронные коллекции в агрегирующие цифровые научные библиотеки.

В разделе 3 затронуты вопросы организации автоматизированной интеграции репозитория математических документов с другими информационными системами.

В разделе 4 обсуждена такая важная функция фабрики метаданных цифровой библиотеки, как нормализация метаданных в соответствии с форматами других агрегирующих библиотек. Реализация этой функции позволяет организовать взаимодействие сервисов как в рамках самой цифровой библиотеки, так и с внешними библиотеками и базами данных, для чего необходимо учитывать используемые в них форматы метаданных.

В разделе 5 решен ряд задач, связанных с построением фабрики метаданных цифровой математической библиотеки Lobachevskii-DML.

В Заключении сформулированы выводы из проведенных исследований и намечены планы их развития.

### **1. ОСОБЕННОСТИ ПРЕДСТАВЛЕНИЯ МЕТАДАНЫХ ДОКУМЕНТОВ В РАЗЛИЧНЫХ ЦИФРОВЫХ МАТЕМАТИЧЕСКИХ БИБЛИОТЕКАХ**

В настоящее время многие наукометрические базы данных осуществляют индексацию статей, опубликованных в ведущих математических журналах за достаточно большие временные периоды их издания. Эти базы данных предъявляют различные требования к набору метаданных включаемых документов, а также схемам их представления (см., например, [67, 68]). Вместе с тем, необходимо отметить, что, как правило, пока еще не индексируются такие новые формы публикаций, как презентации, научные блоги и видеолекции – важные компоненты современных цифровых библиотек [69–71]. В последнее время электронные научные журналы стали поддерживать мультимедийный контент – многие их статьи содержат, помимо текста, формул и графики, также анимационные вставки [72, 73].

Цифровые математические библиотеки при формировании электронных коллекций, входящих в них, используют различные форматы метаданных. Это объясняется тем, что многие такие коллекции образуются из статей, опубликованных в академических журналах и оформленных в соответствии с правилами, уста-

новленными в них и отличающимися требованиями к используемым метаданным. Эти отличия касаются, прежде всего, состава метаданных и их формата и более всего заметны в архивных коллекциях научных журналов.

Наборы метаданных статей даже одного научного журнала в зависимости от года их публикации существенно отличаются. Примером служит архив статей журнала «Известия высших учебных заведений. Математика» (“Russian Mathematics (Izvestiya VUZ. Matematika)”, <https://kpfu.ru/science/nauchnye-izdaniya/ivrm>), статьи которого составляют одну из электронных коллекций цифровой библиотеки Lobachevskii-DML. Этот журнал издается с 1957 года, а относительно полным набором метаданных сопровождаются только статьи, опубликованные в нем, начиная с 2010 года. В статьях, опубликованных до 2008 года, отсутствуют ключевые слова и аннотации. История расширения набора метаданных указанного журнала приведена в Таблице 1.

**Таблица 1.** Состав метаданных статей журнала «Известия вузов. Математика» в период 1957–2020 годов

Год	Состав метаданных
1957–1965	Только основные метаданные: ФИО авторов, название статьи, организация (университет, институт) или город (например, Москва), дата поступления статьи
1965–1969	Начиная со второго номера 1965 года, статьи дополнительно снабжены классификаторами УДК
1970–2007	Из многих статей можно выделить блок с кратким описанием (аннотацию)
2008–2020	Начиная с третьего номера 2008 года, появились блоки «Аннотация» и «Ключевые слова», указаны e-mail адреса авторов и адреса организаций (не всегда)

Отметим, что архивы журнала «Известия вузов. Математика», как и ряда других отечественных научных журналов, оцифрованы и размещены на портале

Mathnet.Ru, причем набор метаданных включенных статей расширен по сравнению с тем, который был изначальным: в дополнение к метаданным, приведенным в Таблице 1, указаны ссылки на английские переводы статей (с 1975 года), выделены блоки литературы, списки статей, ссылающихся на данную статью, а также ссылки на описания статьи в реферативных и наукометрических базах данных MathSciNet и Scopus.

Расширение набора используемых метаданных, описанное выше, является типичным практически для всех научных журналов. В последнее время в ряде научных журналов обязательным является указание ORCID (Open Researcher and Contributor ID) каждого автора [74]. Этот идентификатор предназначен для решения проблемы возможной неоднозначности представления имен и фамилий авторов публикаций в сетях научной коммуникации.

Для пополнения набора метаданных разрабатываются методы извлечения метаданных из документов [20, 25, 39, 68]. Также возникла необходимость в методах нормализации метаданных, позволяющих преобразовать уже созданные метаданные в форматы наукометрических баз данных (методы нормализации метаданных подробно описаны ниже). Отметим также, что обязательным условием участия в интеграционных проектах, таких, например, как EuDML, является предоставление наборов метаданных, сформированных по схемам, установленным в этих проектах.

Статьи из ведущих российских математических журналов переводятся на английский язык и, как правило, имеют библиографическое описание, не совпадающее с имеющимся в русскоязычном издании. В Таблице 2 приведен пример, характеризующий возникающие при этом проблемы. Русскоязычный источник в списке литературы англоязычной статьи либо транслитерируется, либо указывается ссылка на переводную версию документа. Однако, как представлено в Таблице 2, библиографическое описание переведенной статьи может отличаться от транслитерированного описания не только в названии статьи – могут быть совсем иными название журнала, а также номер выпуска и диапазоны страниц статьи. Таким образом, транслитерированная статья и ее переводная версия могут быть восприняты как разные документы. В настоящее время схемы метаданных, основанные на NISO JATS и используемые в цифровой библиотеке EuDML, не позво-



ляют соединить в рамках одного метаописания статью, опубликованную на русском языке, и ее переводную версию на английском языке [75, 76]. В коллекциях Lobachevskii-DML, а также цифровых библиотеках eLibrary.ru и MathNet.Ru такие статьи представлены как дубликаты одного документа.

При формировании электронной коллекции статей научного журнала приходится также учитывать, что в определенные периоды времени этот журнал мог выпускаться под другим названием. Так, например, переводная версия журнала «Известия высших учебных заведений. Математика» до 1991 года имела название “Soviet Mathematics (Izvestiya VUZ. Matematika)”, а, начиная с 1992 года, – “Russian Mathematics (Izvestiya VUZ. Matematika)”.

**Таблица 2.** Как отличаются метаданные в статье и её переводных версиях

Цитирование на русском языке	А.М. Елизаров, А.Б. Жижченко, Н.Г. Жильцов, А.В. Кириллович, Е.К. Липачёв, «Онтологии математического знания и рекомендательная система для коллекций физико-математических документов», Докл. РАН, 467:4 (2016), 392–395
Транслитерированное цитирование (русскоязычный источник в англоязычной статье)	A.M. Elizarov, A.B. Zhizhchenko, N.G. Zhiltsov, A.V. Kirillovich, E.K. Lipachev, “Ontologii matematicheskogo znaniya i rekomendatelnaya sistema dlya kollektсий fiziko-matematicheskikh dokumentov”, Dokl. RAN, 467:4 (2016), 392–395
Цитирование переведенной статьи	A.M. Elizarov, A.B. Zhizhchenko, N.G. Zhiltsov, A.V. Kirillovich, E.K. Lipachev, “Mathematical knowledge ontologies and recommender systems for collections of documents in physics and mathematics”, Dokl. Math., 93:2 (2016), 231–233

Представление правильного варианта цитирования статьи важно при включении таких документов в международные агрегирующие базы данных (например, Scopus), так как между транслитерированным и переведенным названиями

практически нет связи. Отметим, что сегодня во многих журналах для каждой публикуемой статьи приводится готовый вариант ее библиографического описания для последующего цитирования другими авторами. Это описание также является частью метаданных этого документа.

Одним из недостатков формата представления метаданных, предложенного EuDML, является отсутствие средств поддержки нескольких версий одной и той же публикации. Поэтому с использованием этого формата нельзя описать не только новые формы публикации, например, динамические и живые публикации, но и приходится описывать переведенные статьи как различные статьи в разных журналах. В Таблице 2, приведенной выше, показано, как статья и ее перевод на английский язык могут быть представлены как совершенно разные документы. В этом контексте также встает вопрос о препринтах: считать ли ссылку на препринт статьи, размещенный в свободном доступе, ссылкой на статью в цифровой библиотеке; возможно ли ассоциированное размещение в цифровой библиотеке ссылок на оригинальный текст статьи и ее препринт.

Таким образом, при формировании инструментов фабрики метаданных, отвечающих за выбор формата метаданных для описания электронной коллекции, включаемой в цифровую библиотеку, обязательно нужно учитывать историю развития индексируемого научного издания и связи исходных и переводных версий соответствующих документов.

Другой важной составляющей любой цифровой библиотеки являются программные инструменты управления научным контентом. Эти инструменты используются и фабрикой метаданных для создания, обработки, хранения и управления метаданными электронных документов и позволяют интегрировать создаваемые электронные коллекции в агрегирующие цифровые научные библиотеки. Опишем подробнее существующие решения.

## **2. ПРОГРАММНЫЕ ИНСТРУМЕНТЫ УПРАВЛЕНИЯ НАУЧНЫМ КОНТЕНТОМ**

Создание цифровой библиотеки и последующее расширение её функциональных возможностей предполагают решение целого ряда трудоемких задач, связанных, в том числе, с управлением научным контентом (см., например, [6]).

Структура наиболее известных цифровых математических библиотек и разработанные в них сервисы управления математическим контентом обсуждены в [12, 14].

Существующие цифровые библиотеки, а также агрегаторы научных знаний предлагают ряд программных инструментов работы с контентом, прежде всего, сервисы поиска в электронных коллекциях. Например, средства семантического поиска документов представлены на сайте проекта EuDML (<https://initiative.eudml.org/>). Здесь же размещены демонстрационные версии инструментов, разработанных для обслуживания EuDML (см. Таблицу 3).

**Таблица 3:** Демонстрационные версии инструментов EuDML

Название инструмента	Цель создания	Режим доступа и возможности демо-версии
PDF Text Extractor	Экстракция и распознавание текста PDF-документа	ДемOVERсия ( <a href="https://initiative.eudml.org/EuDmlAnalysisDemo/">https://initiative.eudml.org/EuDmlAnalysisDemo/</a> ) извлекает текст только из распознанных документов
Maxtract	Создан для анализа PDF-документа и выделения в нем математических конструкций в виде списка MathML	В демоверсию ( <a href="https://www.cs.bham.ac.uk/research/groupings/reasoning/sdag/eudml-demo.php">https://www.cs.bham.ac.uk/research/groupings/reasoning/sdag/eudml-demo.php</a> ) включены 10 страниц (с. 11–20) книги “Semi-Riemann Geometry and General Relativity – Schlomo Sternberg, 2003”. Есть возможность получить в MathML-формате всю страницу или только определенные строки
TeX2NLM	Принимает на входе T <sub>E</sub> X-строку в UTF-8-кодировке и возвращает то же содержи-	Доступ не предоставлен

	мое в T <sub>E</sub> X- и MathML-представлениях в соответствии со структурой EuDML NLM	
Enhance NLM <sub>TeX</sub> -MathML	Пакетный инструмент, работающий с действительными XML-документами и предназначенный для обновления метаданных с помощью (представления) MathML для любой формулы, написанной на T <sub>E</sub> X, при условии, что функции T <sub>E</sub> X известны компилятору	Доступ не предоставлен
Plain Text Reference Segmenter	Выделяет библиографические ссылки из текста	Совмещен с сервисом PDF Text Extractor. Выделение не всегда корректно ( <a href="https://initiative.eudml.org/EuDmlAnalysisDemo/">https://initiative.eudml.org/EuDmlAnalysisDemo/</a> )
Bibliographic Reference Parser	Разделяет библиографические ссылки, выделяет такие метаданные, как имена и фамилии авторов, названия публикации, год публикации и т. д.	Совмещен с сервисом PDF Text Extractor. Работает не всегда корректно ( <a href="https://initiative.eudml.org/EuDmlAnalysisDemo/">https://initiative.eudml.org/EuDmlAnalysisDemo/</a> )
Find similar articles via Gensim	Поиск сходства между статьями в коллекциях arXiv.org с использованием библиотеки Gensim	<a href="https://mir.fi.muni.cz/eudmldemo/gensim-arxiv/">https://mir.fi.muni.cz/eudmldemo/gensim-arxiv/</a> )
MlaS4gensim demo	Формирует списки терминов и математических формул по их встречаемости в тек-	Словари не полностью очищены от стоп-слов. Пример из полного списка терминов данной библиотеки.

сте, анализируя статьи физико-математической направленности в arxiv.org	23659	anybody	35	
	49385	anyhow	91	
	67484	anymore	518	
	82158	anyone	157	
	88230	anything		999
	4840	anytime	18	
	18032	anyway	716	
	86702	anywhere		408
	https://mir.fi.muni.cz/eudmldemo/mias4gensim/			

Более подробно назначение и функциональные возможности приведенных в таблице программных инструментов описаны в [77–79]. Как видно из Таблицы 3, часть перечисленных инструментов имеет непосредственное отношение к соответствующей фабрике метаданных.

Рассмотрим теперь программные инструменты, представленные в национальных цифровых математических библиотеках.

В рамках проекта «Общероссийский математический портал MathNet.Ru» разработаны сервисы поиска статей, персоналий и организаций, обеспечен полный доступ к статьям. Статьи связаны ссылками с их англоязычными версиями и дополнены соответствующими метаданными. Разработан пакет `amsbib`, предназначенный для оформления библиографии в  $\text{T}_\text{E}\text{X}$ -нотации, а также сервис `MiRef` для работы со списками литературы. Выделены связи со статьями, цитирующими данную статью. Построена гибкая система учета дублирований при цитировании. Разработана и поддерживается система видеопубликаций. Предложена система электронного документооборота редакций журналов [15–17].

Новая платформа французской цифровой математической библиотеки Numdam (<http://www.numdam.org/>) содержит фабрику метаданных, включающую сервисы нормализации метаданных, их корректировки и улучшения, а также инструменты работы с формулами в  $\text{T}_\text{E}\text{X}$ - и MathML-форматах [25].

С 2005 года развивается проект чешской цифровой математической библиотеки DML-CZ (<https://dml.cz/>). Как отмечено в [19], при создании этой библиотеки были учтены успешные решения, реализованные ранее в проекте Numdam. Первоочередная задача проекта DML-CZ состояла в создании цифрового архива

математической литературы, изданной в Чехии. Эта цифровая библиотека содержит практически все математические журналы, изданные чешскими издателями с девятнадцатого века. С самого начала проекта разработчики этой цифровой библиотеки рассматривали её как один из строительных блоков для предполагаемой глобальной DML. Программные инструменты были созданы с учётом этого обстоятельства, и это можно заметить по списку разработчиков сервисов EuDML (см., например, [80]). Названная цифровая библиотека интегрирована в EuDML.

Одной из важнейших задач проекта DML-CZ было также создание комплексной программной системы управления контентом цифровой библиотеки. Эта система содержит программные инструменты, адаптированные к потребностям DML-CZ. Основным инструментом является редактор метаданных [18], объединяющий все действия, связанные с обработкой цифрового контента, созданием метаданных и взаимосвязанными источниками информации.

Отметим, что в DML-CZ реализованы сервисы, учитывающие математическую специфику контента, в частности, имеется возможность классификации документов по кодам AMS Math Classification [38, 80].

### **3. ОРГАНИЗАЦИЯ ХРАНИЛИЩА ЦИФРОВЫХ КОЛЛЕКЦИЙ**

Одной из важных задач при работе с цифровыми математическими библиотеками является автоматизированная интеграция репозиторий математических документов с другими информационными системами. Такой процесс основан на модели агрегирования и распространений метаданных. Модель OAI Protocol for Metadata Harvesting (далее OAI-PMH) [81] поддерживается большинством систем, предназначенных для хранения информационных ресурсов.

Для организации работы с OAI-PMH необходимо использовать систему поддержки цифрового хранилища. Обзор различных цифровых хранилищ приведен в [82]. Наиболее известными из них являются DSpace, Eprints, Fedora и Greenstone. Некоторые библиотеки имеют специализированные методы харвестинга метаданных из других хранилищ, в этом случае необходимо, чтобы у поставщиков данных были инструменты и сервисы, которые позволяют распространять метаданные.

Цифровые математические библиотеки DML-CZ, EuDML и Numdam базируются на информационной платформе DSpace. Это инструмент с открытым исходным кодом, предназначенный для реализации цифровых библиотек и репозиторий [83]. DSpace предлагает большинство базовых функций цифровых библиотек и сервисов, включая пользовательский интерфейс, индексирование, поиск документов, просмотр источников информации, постоянная идентификация документов, предоставление метаданных для сбора данных по протоколу OAI-PMH, поддержка долгосрочного сохранения цифровых данных.

Для решения наших задач важным обстоятельством является совместимость DSpace с издательской платформой Open Journal System (OJS) [84], что позволяет реализовать модель «хранилище + система» работы научного журнала, включая операции с архивом журнала. Обмен данными происходит через OAI-PMH сервер, что позволяет автоматически производить харвестинг метаданных.

#### **4. НОРМАЛИЗАЦИЯ МЕТАДААННЫХ**

Для организации взаимодействия сервисов как в рамках цифровой библиотеки, так и с внешними библиотеками и базами данных, прежде всего, требуется учитывать используемые в них форматы метаданных. Даже в одной цифровой библиотеке программные инструменты работают с несколькими форматами метаданных, что связано с особенностями формирования цифрового контента (например, методов оцифровки), а также требованиями агрегирующих цифровых библиотек и наукометрических баз данных. Отметим лишь наиболее распространенные форматы метаданных, с которыми приходится иметь дело при организации взаимодействия сервисов в цифровых математических библиотеках.

Прежде всего, это формат Dublin Core и его расширения [85], формат каталогизации MARC [86, 87], формат библиографических ссылок RIS (Research Information Systems), AMSBib [15], XML-форматы РИНЦ [88–90], NISO JATS [61–66], форматы на основе схем DBLP [91].

Одной из функций фабрики метаданных является нормализация метаданных в соответствии с форматами других агрегирующих библиотек. Как было отмечено во Введении, термином «нормализация метаданных» обозначают систему сервисов преобразования или формирования метаданных в соответствии с регла-

ментирующими XML-схемами или DTD-правилами целевой цифровой библиотеки [25, 58–60, 68, 92–95]. Так, например, протокол OAI-PMH требует обязательного включения в описание ресурса набора метаданных в нотации oai\_dc, которая основана на Dublin Core. Данная нотация использует только названия ограниченного количества тегов Dublin Core [85].

Для описания статей из математических журналов в цифровой математической библиотеке EuDML применены XML-схемы NISO JATS V1.0 [62], а общая схема метаданных этой цифровой библиотеки описана в [75, 76]. Выделены три набора метаданных: обязательные (obligatory metadata), фундаментальные (fundamental metadata) и дополнительные (supplemental metadata). Из них минимальным по составу является обязательный набор метаданных, в который включены: название статьи на языке оригинала, список авторов, библиография, уникальный идентификатор статьи, например, doi и URL полного текста статьи. Фундаментальный набор метаданных дополнительно к обязательным метаданным включает аннотацию статьи и ключевые слова.

Специализированной базой данных по компьютерным наукам является DBLP (<https://dblp.uni-trier.de/>) [91]. Необходимым условием включения электронных коллекций в эту базу данных являются реорганизация и нормализация метаданных документов включаемой цифровой библиотеки. Передача метаданных электронных коллекций (отдельные статьи не индексируются) в базу DBLP осуществляется автоматическими средствами либо вручную. Схема описания статьи из электронной коллекции включает в себя такие метаданные, как название, ФИО авторов, номера страниц, год выпуска и название конференции/журнала, в котором она была размещена. Индексируемый сборник конференции имеет такие метаданные: ФИО редакторов, название конференции (полное и сокращенное), место проведения конференции.

Определенные затруднения при подготовке метаданных в DBLP возникают с коллекциями, содержащими статьи, не имеющие переводов или, хотя бы, транслитерацию на английский язык таких важных метаданных, как название статьи, список авторов, набор ключевых слов, аффилиация и аннотация. Соответствующие сервисы, автоматизирующие эти процессы, либо являются частью фабрики метаданных цифровой библиотеки, либо онлайн-сервисами других цифровых библиотек.



Отметим, что процесс подготовки метаданных в формате eLibrary.ru автоматизирован (см. [87–90, 96]), а соответствующие сервисы включены в фабрику метаданных цифровой математической библиотеки Lobachevskii-DML, в частности, с их помощью формируются метаданные журнала «Электронные библиотеки».

## **5. ФАБРИКА МЕТАДААННЫХ ЦИФРОВОЙ МАТЕМАТИЧЕСКОЙ БИБЛИОТЕКИ**

В настоящем разделе мы предлагаем решения ряда задач, связанных с построением фабрики метаданных в рамках проекта создания цифровой математической библиотеки Lobachevskii-DML [35–37]. Как и в случае любой цифровой научной библиотеки, формирование библиотеки Lobachevskii-DML и соответствующей фабрики метаданных потребовало привлечения ранее созданных, а также разработки новых технологических решений управления научным контентом.

На Рис. 1 представлена IDEF0-схема фабрики метаданных цифровой библиотеки Lobachevskii-DML.

На этапе препроцессорной обработки выполняется отсев тех документов, которые не получилось обработать в автоматическом режиме, с указанием возникших проблем в файле отчета, сформированном автоматически. Также на этом этапе производятся исправление некоторых ошибок орфографии, неправильного выбора регистра, а также удаление лишних пробелов и знаков.

На этапе экстракции метаданных обрабатываются полные тексты документов, используются шаблоны поиска обязательных метаданных.

На этапе верификации метаданных выполняется проверка полноты и соответствия состава выделенных метаданных установленным правилам, записанным в виде DTD-файлов или XML-схем. После прохождения этапа верификации возможны три варианта дальнейших действий: дополнительная экстракция необходимых и дополнительных метаданных; повторная верификация метаданных и выдача отчета о том, что документ недостаточен для получения требуемых метаданных; переход к финальному действию – нормализации метаданных.

Экстракция дополнительных метаданных направлена на извлечение метаданных не только из самого документа, но и с помощью внешних ресурсов (например, персональных страниц авторов).

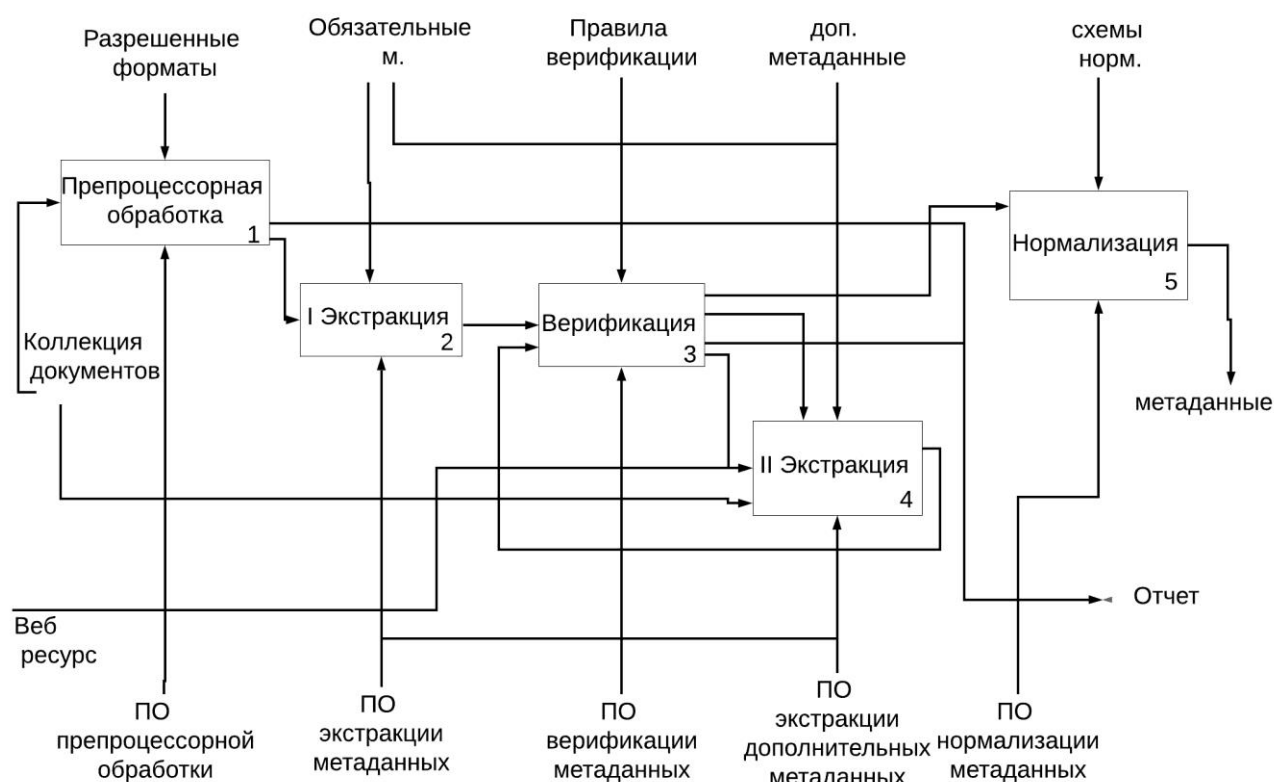


Рис. 1. Этапы формирования метаданных в фабрике метаданных цифровой математической библиотеки

Фабрика метаданных подразумевает, что последние собираются для конкретной цифровой библиотеки. Однако модель такой фабрики может быть распространена на процесс подготовки метаданных и для сайтов-агрегаторов математических знаний. В этом случае изменяются правила верификации метаданных, а также форматы, в которых будет происходить нормализация (некоторые результаты в этом направлении представлены в работах [92, 93]).

Ниже представлены некоторые из уже реализованных нами инструментов фабрики метаданных.

### 5.1. ЭКСТРАКЦИЯ И НОРМАЛИЗАЦИЯ МЕТАДААННЫХ

Размещение метаданных в интернете привело к тому, что одним из их источников могут стать веб-страницы сайта-агрегатора метаданных или самой цифровой библиотеки. Таким образом, при формировании фундаментального набора метаданных электронных коллекций, а также при получении дополнительных метаданных необходимо использовать метаданные, хранящиеся на

внешних ресурсах. Эта задача сопряжена с задачами поиска информации в агрегирующих базах данных и цифровых библиотеках, некоторые из которых частично закрыты для доступа или прерывают соединение, позволяя скачивать только ограниченное количество метаданных. При поиске метаданных на страницах сайтов-агрегаторов нужно также понимать и учитывать, что выбор и порядок поиска в таких источниках должны быть определены заранее, так как некоторые источники хранят информацию только по конкретной тематике (например, библиографическая база данных DBLP) или же неполный список метаданных.

Один из частных случаев экстракции метаданных с сайтов-агрегаторов разработан нами на основе сайта проекта MathNet. Цель созданного программного приложения – выделение и запись метаданных статьи на русском и английском языках с дальнейшей нормализацией по формату, принятому в EuDML. Основные шаги алгоритма экстракции и нормализации метаданных на примере коллекции журнала «Известия вузов. Математика» (“Russian Mathematics”) приведены в [94].

Задача перевода метаданных из одного формата в другой часто связана с задачами дополнения или улучшения метаданных. Так, например, в цифровой библиотеке Lobachevskii DML возникла необходимость перевода метаданных электронной коллекции статей журнала «Электронные библиотеки» (“Russian Digital Libraries Journal”, <https://elbib.ru/>) в формат базы данных DBLP. Процесс перевода включал семантическую транслитерацию имен и фамилий авторов статей. Исходные наборы метаданных, использованных при переводе в названной формат, были сформированы автоматически с помощью программных инструментов, разработанных в редакции журнала «Электронные библиотеки» (<http://ojs.kpfu.ru/index.php/elbib>), и средств программной платформы OJS [84], на которой функционирует данный журнал. Алгоритм перевода этих метаданных в формат DBLP был успешно реализован, подробно он подробно в [94, 95].

Одним из самых распространенных форматов метаданных, принятых в цифровых библиотеках, является формат Dublin Core. В соответствии с концепцией гармонизации метаданных, цифровые библиотеки должны иметь возможность создавать для своих электронных коллекций метаданные в различных форматах. Поэтому была поставлена задача создания инструмента для автоматизированного перевода имеющихся метаданных в формат Dublin Core. Данная задача была

решена на примере коллекции «Трудов Математического центра им. Н.И. Лобачевского» (далее – «Труды»). Эта коллекция насчитывает более 60 томов на русском и английском языках, имеющиеся метаданные статей разнородны.

Прежде всего, была произведена кластеризация, в результате которой соответствующие тома «Трудов» были разделены на классы по сходству их структуры и оформления. Для каждого класса был разработан набор паттернов регулярных выражений, задающих правила поиска информационных блоков. Далее производилась обработка массива файлов «Трудов» с целью выделения метаданных, описывающих как том в целом, так и статьи, входящие в него. В частности, определялись номера страниц всех статей каждого тома. С помощью методов текстового анализа из документов электронной коллекции были выделены термины, из которых были образованы наборы ключевых слов для включения в состав метаданных [97, 98]. Следующий этап включал процедуры разделения каждого тома «Трудов» на отдельные статьи. Далее следовал этап нормализации метаданных. Алгоритмы 1 и 2 нормализации метаданных в различные форматы представлены ниже.

**Алгоритм 1: Нормализация метаданных по схемам DBLP**

```
1: Загрузить VolCollection    \\ коллекция xml файлов метаописаний
2: for each volume in VolCollection do
3   for each paper in volume do
4     Считать из paper значения тегов: author's_names, title,
page_numbers, year_of_issue, url, volume
5     Считать cite_page из https://elbib.ru/en/year/+year
6     Split cite_page
7     Выделить metadata: author's_names in English, url in
elbib.kpfu.ru.
8     Split author's_names
9     Answer:=Form(author's_name, Transliteration(name), ti-
tle, page_numbers, url, volume);
10.    Записать Answer in dblp.xml
11.  end for
12. end for
```

### Алгоритм 2: Нормализация метаданных в формат Dublin Core

```
1: Загрузить VolCollection                \\ коллекция xml файлов
2: for each volume from VolCollection do
3:   Прочитать numbervolume из названия файла volume
4:   Считать из файла info.csv переменные publisher, issue, year
5:   Papers:=new string List
6:   for each paper from volume do
7:     Считать из paper переменные: author's names, title,
page numbers
8:     Paper:=Formoai_dc(authors's names, title, issue
year,           page           numbers,           publisher)
\\ функция формирует строковое описание статьи в формате Dublin Core
9:     Papers.Add(Paper)
10:  end for
11:  Создать файл с названием numbervolume
12:  for each paper from Papers do
13:    Записать paper в файл numbervolume.txt
14:  end for
15: end for
```

## 5.2. ХРАНЕНИЕ ЭЛЕКТРОННЫХ КОЛЛЕКЦИЙ

Одной из основных функций фабрики метаданных является подготовка электронных коллекций для загрузки в цифровую библиотеку. При использовании системы DSspace нужно было решить задачу автоматического формирования файлов и их загрузки в DSspace. Загрузка метаданных в DSspace производилась следующим образом.

Формируется файл-таблица в формате csv (Comma-Separated Values), в который записываются метаданные, подготовленные по схеме Dublin Core. Используется также способ передачи архива в формате Simple Archive Format. Кроме того, используются возможности загрузки метаданных через консоль и ручного ввода метаданных на сайте цифровой библиотеки.

Как показал наш опыт, наиболее рационально использовать загрузку архивом: основными достоинствами являются простота загрузки всех файлов одним архивом, а также возможность загрузить не только метаданные, но и сами файлы.

Далее встает задача нормализации метаданных в тот формат, который принят в DSpace. Подробный обзор методов подготовки метаданных приведен [94, 99].

## **ЗАКЛЮЧЕНИЕ**

С целью интеграции электронных математических коллекций Казанского университета в международное научное информационное пространство разработаны алгоритмы и инструменты создания, обработки, хранения и управления метаданными объектов этих электронных коллекций, позволяющие включать создаваемые электронные коллекции в цифровую математическую библиотеку Lobachevskii-DML. Названные алгоритмы и инструменты составляют фабрику метаданных библиотеки Lobachevskii-DML и обеспечивают формирование метаданных этих коллекций и документов, входящих в них, в соответствии с форматами международных цифровых математических библиотек и наукометрических баз данных, а также дают возможность организовать взаимодействие названных сервисов как в рамках самой цифровой библиотеки, так и с внешними библиотеками и базами данных. Дальнейшее направление развития заключается в совершенствовании созданной фабрики метаданных и разработке возможности ее использования в любых научных цифровых библиотеках.

Работа выполнена при частичной финансовой поддержке Российского фонда фундаментальных исследований и Правительства Республики Татарстан в рамках проекта № 18-47-160012 и в рамках программы развития Регионального научно-образовательного математического центра Приволжского федерального округа, номер соглашения № 075-02-2020-1478/1. Настоящая статья содержит также результаты, полученные в рамках проекта «Мониторинг и стандартизация развития и использования технологий хранения и анализа больших данных в цифровой экономике Российской Федерации», выполняемого в рамках реализации Программы Центра компетенций Национальной технологической инициативы «Центр хранения и анализа больших данных», поддерживаемого Министерством науки и высшего образования Российской Федерации по Договору Московского государственного университета имени М.В. Ломоносова с Фондом поддержки проектов Национальной технологической инициативы от 15.08.2019 № 7/1251/2019.

## СПИСОК ЛИТЕРАТУРЫ

1. *Sicilia M.-A. (Ed) Handbook of Metadata, Semantics and Ontologies.* World Scientific Publishing Co. Pte. Ltd., 2014. 579 p.
2. *Alemu G., Stevens B. An Emergent Theory of Digital Library Metadata. Enrich then Filter.* Chandos Publishing is an imprint of Elsevier. 2015, 121 p. URL: <http://store.elsevier.com/An-Emergent-Theory-of-Digital-Library-Metadata/Getaneh-Alemu/isbn-9780081003855/>.
3. *Gartner R. Metadata. Shaping Knowledge from Antiquity to the Semantic Web.* Springer International Publishing Switzerland, 2016. 118 p. URL: <https://doi.org/10.1007/978-3-319-40893-4>.
4. *Когаловский М.Р. Метаданные, их свойства, функции, классификация и средства представления // Труды 14-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» — RCDL-2012, Переславль-Залесский, Россия, 15–18 октября 2012 г. Переславль-Залесский, 2012. С. 3–14.* URL: <http://rcdl.ru/doc/2012/paper3.pdf>.
5. *Когаловский М.Р. Метаданные в компьютерных системах // Программирование. 2013. Т. 39 (4). С. 28–46.* URL: <http://www.ipr-ras.ru/articles/kogalov13-03.pdf>.
6. *Xie I., Matusiak K.K. Discover Digital Libraries: Theory and Practice.* Elsevier Inc., 2016. 388 p.
7. *Jackson A. The Digital Mathematics Library // Notices Amer. Math. Soc. 2003. V. 50. P. 918–923.*
8. *Bouche T. Introducing the mini-DML project // ECM4 Satellite Conference EMANI/DML. 2004. P. 19–29.*
9. *Borwein J.M., Rocha E.M., Rodrigues J.F. (eds.) Communicating Mathematics in the Digital Era.* Taylor & Francis, 2008. 325 p.
10. *Bouche T. Some Thoughts on the Near-Future Digital Mathematics Library // Towards a Digital Mathematics Library. Masaryk University, 2008. P. 3–15.* URL: <https://eudml.org/doc/221606>.
11. *Bouche T. Digital Mathematics Libraries: The Good, the Bad, the Ugly // Math. Comput. Sci. 2010. V. 3. P. 227–241.* URL: <https://doi.org/10.1007/s11786-010-0029-2>.

12. *Bouche T.* The Digital Mathematics Library as of 2014 // *Notices Amer. Math. Soc.* 2014. V. 61. No 9. P. 1085–1088.

13. *Elizarov A.M., Lipachev E.K., Zuev D.S.* Digital Mathematical Libraries: Overview of Implementations and Content Management Services // *CEUR Workshop Proceedings.* 2017. V. 2022. P. 317–325.

14. Committee on Electronic Information Communication of the International Mathematical Union, Best current practices: Recommendations on electronic information communication // *Notices of the AMS.* 2002. 49(8), pp. 922–925. URL: <http://ams.org/notices/200208/commpractices.pdf>.

15. *Жижченко А.Б., Изаак А.Д.* Информационная система Math-Net.Ru. Применение современных технологий в научной работе математика // *Успехи математических наук.* 2007. Т. 62, №5 (377). С. 107–132. URL: <https://doi.org/10.4213/rm8147>. URL: <http://www.mathnet.ru/links/c59aff2f134382372f88aa415a76755f/rm8147.pdf>.

16. *Жижченко А.Б., Изаак А.Д.* Информационная система Math-Net.Ru. Современное состояние и перспективы развития. Импакт-факторы российских математических журналов // *Успехи математических наук.* 2009. Т. 64, №4 (388). С. 195–204. URL: <https://doi.org/10.4213/rm9312>; <http://www.mathnet.ru/links/e27ab619eaefe03fe79d663468ddd3a0/rm9312.pdf>.

17. *Chebukov D.E., Izaak A.D., Misyurina O.G., Pupyrev Yu.A., Zhizhchenko A.B.* Math-Net.Ru as a Digital Archive of the Russian Mathematical Knowledge from the XIX Century to Today. *Intelligent Computer Mathematics // Lecture Notes in Computer Science.* 2013. V. 7961. P. 344–348. URL: [https://doi.org/10.1007/978-3-642-39320-4\\_26](https://doi.org/10.1007/978-3-642-39320-4_26).

18. *Bartošek M., Kovář P., Šárfa M.* DML-CZ Metadata Editor // In: Sojka P. (ed.) *Towards Digital Mathematics Library.* Masaryk University, 2010. P. 139–151. URL: [https://dml.cz/bitstream/handle/10338.dmlcz/702537/DML\\_001-2008-1\\_17.pdf](https://dml.cz/bitstream/handle/10338.dmlcz/702537/DML_001-2008-1_17.pdf).

19. *Bartošek M., Rákosník J.* DML-CZ: The Experience of a Medium-Sized Digital Mathematics Library // *Notices of the AMS.* 2013. V. 60. No 8. P. 1028–1033. URL: <http://dx.doi.org/10.1090/noti1031>.

20. *Bouche T.* Toward a digital mathematics library? // In: Borwein J.M., Rocha E.M., Rodrigues J.F. (Eds) *Communicating Mathematics in the Digital Era.* Taylor & Francis, 2008. P. 47–73. URL: <https://hal.archives-ouvertes.fr/hal-00347682>.

---



21. *Lange C.* Ontologies and Languages for Representing Mathematical Knowledge on the Semantic Web // *Semantic Web*. 2013. V. 4. No. 2. P. 119–158. URL: <https://content.iospress.com/articles/semantic-web/sw059>.

22. *Серебряков В.А.* Что такое семантическая цифровая библиотека // Труды 16-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» — RCDL-2014, Дубна, Россия, 13–16 октября 2014 г. Дубна: ОИЯИ, 2014. С. 1–5. URL: [http://rcdl.ru/doc/2014/paper/RCDL2014\\_021-25.pdf](http://rcdl.ru/doc/2014/paper/RCDL2014_021-25.pdf).

23. *Елизаров А.М., Кириллович А.В., Липачёв Е.К., Невзорова О.А.* Управление математическими знаниями: онтологические модели и цифровые технологии // *Аналитика и управление данными в областях с интенсивным использованием данных: сборник статей XVIII Междун. конф. DAMDID/RCDL'2016*. М.: ФИЦ ИУ РАН, 2016. С. 95–101.

24. *Elizarov A.M., Kirillovich A.V., Lipachev E.K., Nevzorova O.A.* Mathematical Knowledge Management: Ontological Models and Digital Technology // *CEUR Workshop Proceedings*. 2016. V. 1752. P. 44–50. URL: <http://ceur-ws.org/Vol-1752/paper08.pdf>

25. *Bouche T., Labbe O.* The New Numdam Platform // In: Geuvers H., England M., Hasan O., Rabe F., Teschke O. (Eds) *Intelligent Computer Mathematics. CICM 2017. Lecture Notes in Computer Science*. Vol. 10383. Springer, Cham, 2017. P. 70–82. URL: [https://doi.org/10.1007/978-3-319-62075-6\\_6](https://doi.org/10.1007/978-3-319-62075-6_6). <https://zenodo.org/record/581405/>.

26. *Sadegh A., Lange C., Vidal M.E., Auer S.* Integration of Scholarly Communication Metadata using Knowledge Graphs // *International Conference on Theory and Practice of Digital Libraries*. 2017. P. 328–341.

27. *Ataeva O.M., Serebryakov V.A.* Information Model of LibMeta Digital Library // *Lobachevskii J. of Mathematics*. 2019. V. 40. No 7. P. 861–875. URL: <https://doi.org/10.1134/S1995080219070035>.

28. *Developing a 21st Century Global Library for Mathematics Research* // Washington: The National Academies Press, 2014. 142 p. doi:10.17226/18619.

29. *Olver P.J.* The World Digital Mathematics Library: Report of a Panel Discussion // *Proceedings of the International Congress of Mathematicians, August 13–*

21, 2014, Seoul, Korea. Kyung Moon SA, 1. 2014. P. 773–785.

30. *Watt S.* How to Build a Global Digital Mathematics Library // 18th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC). 2016. P. 37–40.

31. *Ion P.D.F., Watt S.M.* The Global Digital Mathematics Library and the International Mathematical Knowledge Trust // ICM 2017: Intelligent Computer Mathematics, 2017. Lecture Notes in Artificial Intelligence. 2017. V. 10383. P. 56–69. URL: [https://doi.org/10.1007/978-3-319-62075-6\\_5](https://doi.org/10.1007/978-3-319-62075-6_5).

32. *Bouche T.* Reviving the free public scientific library in the digital age? The EuDML project // In: Kaiser K., Krantz S.G., Wegner B. (Eds) Topics and Issues in Electronic Publishing JMM/AMS Special Session, FIZ Karlsruhe. 2013. P. 57–80. URL: <https://www.emis.de/proceedings/TIEP2013/05bouche.pdf>.

33. *Bouche T., Rákosník J.* Report on the EuDML External Cooperation Model // In: Kaiser K., Krantz S.G., Wegner B. (Eds) Topics and Issues in Electronic Publishing, JMM, Special Session. San Diego. 2013. P. 99–108. URL: [https://www.emis.de/proceedings/TIEP2013/07bouche\\_rakosnik.pdf](https://www.emis.de/proceedings/TIEP2013/07bouche_rakosnik.pdf).

34. *Sylwestrzak W., Borbinha J., Bouche T., Nowiński A., Sojka P.* EuDML – Towards the European Digital Mathematics Library // In: Sojka P.(ed.) Towards a Digital Mathematics Library. Masaryk University, 2010. P. 11–26. URL: <https://eudml.org/doc/220786>.

35. *Elizarov A.M., Lipachev E.K.* Lobachevskii DML: Towards a Semantic Digital Mathematical Library of Kazan University // CEUR Workshop Proceedings. 2017. V. 2022. P. 326–333.

36. *Елизаров А.М., Липачёв Е.К.* Семантические методы и инструменты электронной математической библиотеки Lobachevskii-DML // Научный сервис в сети Интернет: труды XIX Всероссийской научной конференции (18–23 сентября 2017 г., г. Новороссийск). М.: ИПМ им. М.В. Келдыша, 2017. С. 130–136. URL: <https://doi.org/10.20948/abrau-2017-73>. <http://keldysh.ru/abrau/2017/73.pdf>.

37. *Елизаров А.М., Липачёв Е.К., Хайдаров Ш.М.* Структура и сервисы цифровой математической библиотеки Lobachevskii-DML // Ученые записки Института социально-гуманитарных знаний. 2017. № 1 (15). С. 215–220.

38. *Růžička M., Sojka P., Krejčíř V.* Towards Machine-Actionable Modules of a Digital Mathematics Library // In: Carette J., Aspinall D., Lange C., Sojka P., Windsteiger

W. (eds) Intelligent Computer Mathematics. CICM 2013. Lecture Notes in Computer Science. 2013. V. 7961. P. 263–277. URL: [https://doi.org/10.1007/978-3-642-39320-4\\_17](https://doi.org/10.1007/978-3-642-39320-4_17).

39. *Елизаров А.М., Липачёв Е.К., Хайдаров Ш.М.* Семантический анализ больших коллекций научных документов // Труды международной конференции по компьютерной и когнитивной лингвистике TEL-2016. Казань: Изд-во Казан. унта, 2016. С. 21–25.

40. *Елизаров А.М., Липачёв Е.К., Хайдаров Ш.М.* Автоматизированная система сервисов обработки больших коллекций научных документов // Аналитика и управление данными в областях с интенсивным использованием данных: сборник статей XVIII Междун. конф. DAMDID/RCDL'2016. М.: ФИЦ ИУ РАН, 2016. С. 109–115.

41. *Elizarov A.M., Khaydarov Sh.M., Lipachev E.K.* Scientific Documents Ontologies for Semantic Representation of Digital Libraries // Proc. of the 2nd Russia and Pacific Conf. on Computer Technology and Applications. 2017. P. 1–5. URL: <https://doi.org/10.1109/RPC.2017.8168064>.

42. *Nevzorova O., Zhiltsov N., Kirillovich A., Lipachev E.* OntoMathPRO Ontology: A Linked Data Hub for Mathematics // Klinov P., Mouromstev D. (eds.) Proceedings of the 5th International Conference on Knowledge Engineering and Semantic Web (KESW 2014). Communications in Computer and Information Science, Springer, Cham, 2014. V. 468. P. 105–119. URL: [https://doi.org/10.1007/978-3-319-11716-4\\_9](https://doi.org/10.1007/978-3-319-11716-4_9).

43. *Elizarov A., Kirillovich A., Lipachev E., Nevzorova O., Solovyev V., Zhiltsov N.* Mathematical Knowledge Representation: Semantic Models and Formalisms // Lobachevskii J. of Mathematics. 2014. V. 35. No 4. P. 347–353. URL: <https://doi.org/10.1134/S1995080214040143>.

44. *Елизаров А.М., Липачёв Е.К., Невзорова О.А., Соловьев В.Д.* Методы и средства семантического структурирования электронных математических документов // Доклады РАН. 2014. Т. 457, № 6. С. 642–645. URL: <https://doi.org/10.7868/S0869565214240049>.

45. *Елизаров А.М., Жильцов Н.Г., Кириллович А.В., Липачёв Е.К., Невзорова О.А.* Экосистема ONTOMATH и проект Всемирной цифровой математической

библиотеки // Труды международной конференции по компьютерной и когнитивной лингвистике TEL-2016. Казань: Изд-во Казан. ун-та, 2016. С. 25–28.

46. *Elizarov A.M., Kirillovich A.V., Lipachev E.K., Nevzorova O.A.* Digital Ecosystem OntoMath: Mathematical Knowledge Analytics and Management // Communications in Computer and Information Science. Springer. 2017. V. 70. P. 33–46. URL: [https://doi.org/10.1007/978-3-319-57135-5\\_3](https://doi.org/10.1007/978-3-319-57135-5_3).

47. *Elizarov A., Kirillovich A., Lipachev E., Nevzorova O.* Semantic Formula Search in Digital Mathematical Libraries // Proceedings of the 2nd Russia and Pacific Conference on Computer Technology and Applications (RPC 2017). IEEE, 2017. P. 39–43. URL: <https://doi.org/10.1109/RPC.2017.8168063>.

48. *Birialtsev E., Gusenkov A., Zhibrik O., Gusenkova P., Palacheva Y.* Search in Collections of Mathematical Articles // In: Rocha Á., Adeli H., Reis L.P., Costanzo S. (eds) Trends and Advances in Information Systems and Technologies. WorldCIST'18 2018. Advances in Intelligent Systems and Computing, Springer, Cham, 2018. V. 745. P. 561–567. URL: [https://doi.org/10.1007/978-3-319-77703-0\\_55](https://doi.org/10.1007/978-3-319-77703-0_55).

49. *Шакирова Л.Р., Фалилеева М.В., Кириллович А.В., Липачев Е.К., Невзорова О.А., Невзоров В.Н.* Образовательная математическая онтология OntoMathEdu: структура и отношения // Научный сервис в сети Интернет: труды XXI Всероссийской научной конференции (23–28 сентября 2019 г., г. Новороссийск). М.: ИПМ им. М.В. Келдыша, 2019. С. 653–661. URL: <https://doi.org/10.20948/abrau-2019-84>. <http://keldysh.ru/abrau/2019/theses/84.pdf>.

50. *Shakirova L., Falileeva M., Kirillovich A., Lipachev E.* Modeling and evaluation of the mathematical educational ontology //CEUR Workshop Proceedings. 2020. V. 2543. P. 305–319.

51. *Kirillovich A., Shakirova L., Falileeva M., Lipachev E.* Towards an Educational Mathematical Ontology // L. Gómez Chova, et al. (eds). Proceedings of the 13th International Technology, Education and Development Conference (INTED2019), Valencia, Spain, March 11th–13th, 2019. IATED, 2019. P. 6823–6829.

52. *Елизаров А.М., Жильцов Н.Г., Кириллович А.В., Липачёв Е.К.* Семантическое аннотирование в системе управления физико-математическим контентом // Научный сервис в сети Интернет: труды XVII Всероссийской научной конференции (21–26 сентября 2015 г., г. Новороссийск). М.: ИПМ им. М.В. Келдыша, 2015. С. 98–103. URL: [https://kpfu.ru//staff\\_files/F1890276653/Elizarov\\_at\\_all\\_.pdf](https://kpfu.ru//staff_files/F1890276653/Elizarov_at_all_.pdf).

---

53. *Елизаров А.М., Жильцов Н.Г., Кириллович А.В., Липачёв Е.К.* Терминологическое аннотирование и рекомендательный сервис в системе управления физико-математическим контентом // Труды XVII Международной конференции DAMDID/RCDL'2015 «Аналитика и управление данными в областях с интенсивным использованием данных». Обнинск: ИАТЭ НИЯУ МИФИ, 2015. С. 357–350. URL: [https://kpfu.ru//staff\\_files/F684099727/damdid2015\\_paper\\_Elizarov\\_new.pdf](https://kpfu.ru//staff_files/F684099727/damdid2015_paper_Elizarov_new.pdf).

54. *Елизаров А.М., Жижченко А.Б., Жильцов Н.Г., Кириллович А.В., Липачёв Е.К.* Онтологии математического знания и рекомендательная система для коллекций физико-математических документов // Доклады РАН. 2016. Т. 467, № 4. С. 392–395. URL: <https://doi.org/10.7868/S0869565216100042>.

55. *Хайдаров Ш.М., Ямалутдинова Г.Ш.* Алгоритм формирования словарей рекомендующей системы подбора классификаторов научной информации // Ученые записки Института социально-гуманитарных знаний. 2017. № 1 (15). С. 552–557.

56. *Хайдаров Ш.М., Ямалутдинова Г.Ш.* Рекомендательная система классификации физико-математических документов // Научный сервис в сети Интернет: труды XX Всероссийской научной конференции (17–22 сентября 2018 г., г. Новороссийск). М.: ИПМ им. М.В. Келдыша, 2018. С. 480–486. URL: <https://doi.org/10.20948/abrau-2018-57>. <http://keldysh.ru/abrau/2018/theses/57.pdf>.

57. *Khaydarov S.M., Yamalutdinova G.S.* Recommender system of physical and mathematical documents classification // CEUR Workshop Proceedings. 2018. V. 2260. P. 480–486.

58. *Harper C.* Metadata normalization: a case study in Primo and linked open data in libraries // Metadata Working Group Forum, Cornell, 2008. URL: <https://ecommons.cornell.edu/handle/1813/10920>.

59. *Koh J., Hong D., Gupta R., Whitehouse K., Wang H., Agarwal Y.* Plaster: An Integration, Benchmark, and Development Framework for Metadata Normalization Methods // BuildSys'18: Proceedings of the 5th Conference on Systems for Built Environments. 2018. P. 1–10. URL: <https://doi.org/10.1145/3276774.3276794>.

60. *Koh J., Balaji B., Sengupta D., McAuley J., Gupta R., Agarwal Y.* Scrabble: Transferrable Semi-Automated Semantic Metadata Normalization using Intermediate

Representation // BuildSys'18: Proceedings of the 5th Conference on Systems for Built Environments. 2018. P. 11–20. URL: <https://doi.org/10.1145/3276774.3276795>.

61. Journal Article Tag Suite. URL: <https://jats.nlm.nih.gov/about.html>.

62. ANSI/NISO Z39.96-2012, JATS: Journal Article Tag Suite (ver. 1.0). URL: [https://groups.niso.org/apps/group\\_public/download.php/10904/z39.96-2012.pdf](https://groups.niso.org/apps/group_public/download.php/10904/z39.96-2012.pdf).

63. Journal Archiving and Interchange Tag Set, ver. 1.1. URL: <https://jats.nlm.nih.gov/publishing/1.1/>.

64. Journal Archiving and Interchange Tag Set, ver. 1.2. URL: <https://jats.nlm.nih.gov/archiving/1.2/>.

65. JATS: Journal Article Tag Suite. National Information Standards Organization, ver. 1.2 (ANSI/NISO Z39.96-2019). 8 Febr. 2019. 652 p. URL: [https://groups.niso.org/apps/group\\_public/download.php/21030/ANSI-NISO-Z39.96-2019.pdf](https://groups.niso.org/apps/group_public/download.php/21030/ANSI-NISO-Z39.96-2019.pdf).

66. Journal Publishing Tag Library NISO JATS, version 1.3d1 (ANSI/NISO Z39.96-2019). October 2019. URL: <https://jats.nlm.nih.gov/archiving/1.3d1/>.

67. *Heller L., The R., Bartling S.* Dynamic Publication Formats and Collaborative Authoring // In: Bartling S. and Friesike S. (eds.). *Opening Science*. Springer Cham, 2014. P. 191–211. URL: [https://doi.org/10.1007/978-3-319-00026-8\\_13](https://doi.org/10.1007/978-3-319-00026-8_13).

68. *Елизаров А.М., Зайцева Н.В., Зуев Д.С., Липачёв Е.К., Хайдаров Ш.М.* Сервисы формирования метаданных цифровых документов в форматах международных наукометрических баз данных // *Научный сервис в сети Интернет: труды XX Всероссийской научной конференции (17–22 сентября 2018 г., г. Новороссийск)*. М.: ИПМ им. М.В. Келдыша, 2018. P. 75–185. URL: <https://doi.org/10.20948/abrau-2018-53>. <http://keldysh.ru/abrau/2018/theses/53.pdf>.

69. *Кириллович А.В.* Информационная архитектура блогов // *Электронные библиотеки*. 2017. Т. 20. № 2. С. 147–162. URL: <https://elbib.ru/article/view/418/504>.

70. *Елизаров А.М., Кириллович А.В., Липачёв Е.К.* Блоги в системе научных коммуникаций // *Ученые записки Института социально-гуманитарных знаний*. 2017. №1 (15). С. 209–214.

71. *Гафурова П.О.* Форумы в системе научных коммуникаций // *Сборник конференции: Информационные технологии в образовании и науке (ИТОН-2018)*. 2018. С. 102–103.

72. Борисов Н.В., Захаркина В.В., Мбого И.А., Прокудин Д.Е., Щербаков П.П. Проблемы создания онлайн научного журнала с мультимедиа контентом // Научный сервис в сети Интернет: труды XXI Всероссийской научной конференции (23–28 сентября 2019 г., г. Новороссийск). М.: ИПМ им. М.В. Келдыша, 2019. С. 153–165. URL: <https://doi.org/10.20948/abrau-2019-87>. <http://keldysh.ru/abrau/2019/theses/87.pdf>.

73. Borisov N.V., Zakharkina N.V., Mbogo I.A., Prokudin D.E., Scherbakov P.P. Challenges of Publishing Online Scholarly Journals with Multimedia Content // CEUR Workshop Proceedings. 2020. V. 2543. P. 93–102. URL: <http://ceur-ws.org/Vol-2543/rpaper09.pdf>.

74. ORCID. Connecting Research and Researchers. URL: <https://orcid.org/>.

75. EuDML metadata schema specification (v2.0–final). URL: <https://initiative.eudml.org/eudml-metadata-schema-specification-v20-final>.

76. 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>.

77. D7.2: Toolset for Image and Text Processing and Metadata Editing – Initial release. URL: <http://www.mathdoc.fr/publis/d7.2-v1.0.pdf>.

78. D7.3: Toolset for Image and Text Processing and Metadata Enhancements – Value release. URL: <http://www.mathdoc.fr/publis/d7.3.pdf>.

79. D7.4: Toolset for Image and Text Processing and Metadata Enhancements – Final Release. URL: <https://wiki.eudml.eu/mediawiki/eudml/images/D7.4-v1.0.pdf>.

80. Řehůřek R., Sojka P. Automated Classification and Categorization of Mathematical Knowledge // In: Autexier S., Campbell J., Rubio J., Sorge V., Suzuki M., Wiedijk F. (Eds) Intelligent Computer Mathematics. CICM 2008. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, 2008. V. 5144. P. 543–557. URL: [https://doi.org/10.1007/978-3-540-85110-3\\_44](https://doi.org/10.1007/978-3-540-85110-3_44).

81. Open Archives Initiative Protocol for Metadata Harvesting. URL: <http://www.openarchives.org/OAI/openarchivesprotocol.html>.

82. Федотов А.М., Байдавлетов А.Т., Жижимов О.Л., Самбетбаева М.А., Федотова О.А. Цифровой репозиторий в научно-образовательной информационной системе // Вестник НГУ. Серия: Информационные технологии. 2015. Т. 13.

№ 3. С. 68–86. URL: <https://cyberleninka.ru/article/n/tsifrovoy-repozitoriy-v-nauchno-obrazovatelnoy-informatsionnoy-sisteme>.

83. *Krejčíř V.* Building the Czech Digital Mathematics Library upon DSpace System // In: Sojka P. (Ed) DML 2008. Towards Digital Mathematics Library. Brno: Masaryk University, 2008. P. 117–126. URL: [https://dml.cz/bitstream/handle/10338.dmlcz/702539/DML\\_001-2008-1\\_14.pdf](https://dml.cz/bitstream/handle/10338.dmlcz/702539/DML_001-2008-1_14.pdf).

84. *MacGregor J., Stranack K., Willinsky J.* The Public Knowledge Project: Open Source Tools for Open Access to Scholarly Communication // 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. 165–175. URL: [https://doi.org/10.1007/978-3-319-00026-8\\_3](https://doi.org/10.1007/978-3-319-00026-8_3).

85. Expressing Dublin Core metadata using XML. URL: <https://www.dublincore.org/specifications/dublin-core/dc-xml/>.

86. What is a MARC record and why is it important? URL: <https://www.loc.gov/marc/umb/um01to06.html>.

87. *MARC Standarts.* URL: <http://www.loc.gov/marc/>.

88. *Елизаров А.М., Липачёв Е.К., Хайдаров Ш.М.* Программа автоматизированного формирования метаданных в формате Российского индекса научного цитирования для статей журнала «Электронные библиотеки». Свидетельство о регистрации программы для ЭВМ RU 2018612458, 16.02.2018. Заявка № 2017663206 от 19.12.2017. URL: <https://www.elibrary.ru/item.asp?id=39291526>.

89. *Герасимов А.Н., Елизаров А.М., Липачёв Е.К.* Формирование метаданных для международных баз цитирования в системе управления электронными научными журналами // Электронные библиотеки. 2015. Т. 18. № 1–2. С. 6–31. URL: <https://elbib.ru/article/view/356/447>.

90. *Ахметов Д.Ю., Елизаров А.М., Липачёв Е.К.* Сервис-ориентированная информационная система научного журнала «Электронные библиотеки» // Электронные библиотеки. 2016. Т. 19. № 1. С. 2–39. URL: <https://elbib.ru/article/view/377/468>.

91. *Ley M.* DBLP – Some Lessons Learned // Proceedings of the VLDB Endowment. 2009. V. 2 (2). P. 1493–1500.

92. *Гафурова П.О.* Методы нормализации метаданных цифровых матема-



тических библиотек // В книге: Ломоносов-2019. Сборник тезисов XXVI Международной научной конференции студентов, аспирантов и молодых ученых; секция «Вычислительная математика и кибернетика». 2019. С. 162–164.

93. *Гафурова П.О., Елизаров А.М., Липачёв Е.К.* Методы нормализации метаданных электронных математических коллекций // Ученые записки Института социально-гуманитарных знаний. 2019. Т. 17. № 1. С. 141–148.

94. *Gafurova P.O., Elizarov A.M., Lipachev E.K., Khammatova D.M.* Metadata Normalization Methods in the Digital Mathematical Library // CEUR Workshop Proceedings. 2020. V. 2543. P. 136–148.

95. *Гафурова П.О., Елизаров А.М., Липачёв Е.К., Хамматова Д.М.* Методы формирования и нормализации метаданных в цифровой математической библиотеке // Научный сервис в сети Интернет: труды XXI Всероссийской научной конференции (23–28 сентября 2019 г., г. Новороссийск). М.: ИПМ им. М.В. Келдыша, 2019. С. 234–244. URL: <https://doi.org/10.20948/abrau-2019-28>. <http://keldysh.ru/abrau/2019/theses/28.pdf>.

96. *Елизаров А.М., Зайцева Н.В., Липачёв Е.К., Хайдаров Ш.М.* Программа автоматизированного формирования метаданных документов цифровой математической библиотеки Lobachevskii DML. Свидетельство о регистрации программы для ЭВМ RU 2019611328, 24.01.2019. Заявка № 2019610406 от 15.01.2019. URL: <https://www.elibrary.ru/item.asp?id=39309871>.

97. *Батыршина Р.Р., Елизаров А.М., Липачёв Е.К.* Организация коллекций цифровой математической библиотеки методами семантического анализа // Научный сервис в сети Интернет: труды XXI Всероссийской научной конференции (23–28 сентября 2019 г., г. Новороссийск). М.: ИПМ им. М.В. Келдыша, 2019. С. 85–90. URL: <https://doi.org/10.20948/abrau-2019-97>. <http://keldysh.ru/abrau/2019/theses/97.pdf>.

98. *Elizarov A.M., Lipachev E.K.* Methods of Processing Large Collections of Scientific Documents and the Formation of Digital Mathematical Library // CEUR Workshop Proceedings. 2020. V. 2543. P. 354–360.

99. *Гафурова П.О., Липачёв Е.К.* Методы семантического представления математических коллекций цифровой библиотеки Lobachevskii-DML // Труды Математического центра им. Н.И. Лобачевского. 2018. Т. 56. С. 90–93.

## BASIC SERVICES OF FACTORY METADATA DIGITAL MATHEMATICAL LIBRARY LOBACHEVSKII-DML

P. O. Gafurova <sup>1</sup>, A. M. Elizarov <sup>2</sup>, E. K. Lipachev <sup>3</sup>

<sup>1-3</sup> Higher School of Information Technologies and Intelligent Systems, Kazan (Volga Region) Federal University, ul. Kremlyovskaya, 35, Kazan, 420008

<sup>2,3</sup> N. I. Lobachevskii Institute of Mathematics and Mechanics, Kazan (Volga Region) Federal University, ul. Kremlyovskaya, 35, Kazan, 420008

<sup>1</sup>pogafurova@gmail.com, <sup>2</sup>amelizarov@gmail.com, <sup>3</sup>elipachev@gmail.com

### **Abstract**

A number of problems related to the construction of the metadata factory of the digital mathematical library Lobachevskii-DML have been solved. By metadata factory we mean a system of interconnected software tools aimed at creating, processing, storing and managing metadata of digital library objects and allowing integrating created electronic collections into aggregating digital scientific libraries. In order to select the optimal such software tools from existing ones and their modernization:

- we discussed the features of the presentation of the metadata of documents of various electronic collections related both to the formats used and to changes in the composition and completeness of the set of metadata throughout the entire publication of the corresponding scientific journal;
- we presented and characterized software tools for managing scientific content and methods for organizing automated integration of repositories of mathematical documents with other information systems;
- we discussed such an important function of the digital library metadata factory as the normalization of metadata in accordance with the formats of other aggregating libraries.

As a result of the development of the metadata factory of the digital mathematical library Lobachevskii-DML, we proposed a system of services for the automated generation of metadata for electronic mathematical collections; we have developed an xml metadata presentation language based on the Journal Archiving and

Interchange Tag Suite (NISO JATS); we have created software tools for normalizing metadata of electronic collections of scientific documents in formats developed by international organizations – aggregators of resources in mathematics and Computer Science; we have developed an algorithm for converting metadata to oai\_dc format and generating the archive structure for import into DSpace digital storage; we have proposed and implemented methods for integrating electronic mathematical collections of Kazan University into domestic and foreign digital mathematical libraries.

**Keywords:** *digital libraries, digital mathematical library, metadata generation, metadata extraction, metadata normalization, metadata factory, NISO JATS, semantic relationships, Lobachevskii-DML.*

## REFERENCES

1. *Sicilia M.-A. (Ed) Handbook of Metadata, Semantics and Ontologies.* World Scientific Publishing Co. Pte. Ltd., 2014. 579 p.
2. *Alemu G., Stevens B. An Emergent Theory of Digital Library Metadata.* Enrich then Filter. Chandos Publishing is an imprint of Elsevier. 2015, 121 p. URL: <http://store.elsevier.com/An-Emergent-Theory-of-Digital-Library-Metadata/Getaneh-Alemu/isbn-9780081003855/>.
3. *Gartner R. Metadata. Shaping Knowledge from Antiquity to the Semantic Web.* Springer International Publishing Switzerland, 2016. 118 p. URL: <https://doi.org/10.1007/978-3-319-40893-4>.
4. *Kogalovsky Mikhail. Metadata, their Properties, Functions and Classifications // CEUR Workshop Proceedings.* 2012. V. 934. P. 3–14. URL: <http://ceur-ws.org/Vol-934/paper3.pdf>.
5. *Kogalovsky M.R. Metadata in Computer Systems // Programming and Computer Software.* 2013. V. 39 (4). P. 182–193.
6. *Xie I., Matusiak K.K. Discover Digital Libraries: Theory and Practice.* Elsevier Inc., 2016. 388 p.
7. *Jackson A. The Digital Mathematics Library // Notices Amer. Math. Soc.* 2003. V. 50. P. 918–923.
8. *Bouche T. Introducing the mini-DML project // ECM4 Satellite Conference EMANI/DML.* 2004. P. 19–29.
9. *Borwein J.M., Rocha E.M., Rodrigues J.F. (eds.) Communicating*

Mathematics in the Digital Era. Taylor & Francis, 2008. 325 p.

10. *Bouche T.* Some Thoughts on the Near-Future Digital Mathematics Library // Towards a Digital Mathematics Library. Masaryk University, 2008. P. 3–15. URL: <https://eudml.org/doc/221606>.

11. *Bouche T.* Digital Mathematics Libraries: The Good, the Bad, the Ugly // Math. Comput. Sci. 2010. V. 3. P. 227–241. URL: <https://doi.org/10.1007/s11786-010-0029-2>.

12. *Bouche T.* The Digital Mathematics Library as of 2014 // Notices Amer. Math. Soc. 2014. V. 61. No 9. P. 1085–1088.

13. *Elizarov A.M., Lipachev E.K., Zuev D.S.* Digital Mathematical Libraries: Overview of Implementations and Content Management Services // CEUR Workshop Proceedings. 2017. V. 2022. P. 317–325.

14. Committee on Electronic Information Communication of the International Mathematical Union, Best current practices: Recommendations on electronic information communication // Notices of the AMS. 2002. 49(8), pp. 922–925. URL: <http://ams.org/notices/200208/commpractices.pdf>.

15. *Zhizhchenko A.B., Izaak A.D.* The information system Math-Net.Ru. Application of contemporary technologies in the scientific work of mathematicians // Russian Math. Surveys. 2007. V. 62 (5). P. 943–966. <http://dx.doi.org/10.1070/RM2007v062n05ABEH004455>.

16. *Zhizhchenko A.B., Izaak A.D.* The information system Math-Net.Ru. Current state and prospects. The impact factors of Russian mathematics journals // Russian Math. Surveys. 2009. V. 64 (4). P. 775–784. <http://dx.doi.org/10.1070/RM2009v064n04ABEH004638>.

17. *Chebukov D.E., Izaak A.D., Misyurina O.G., Pupyrev Yu.A., Zhizhchenko A.B.* Math-Net.Ru as a Digital Archive of the Russian Mathematical Knowledge from the XIX Century to Today. Intelligent Computer Mathematics // Lecture Notes in Computer Science. 2013. V. 7961. P. 344–348. [https://doi.org/10.1007/978-3-642-39320-4\\_26](https://doi.org/10.1007/978-3-642-39320-4_26).

18. *Bartošek M., Kovář P., Šárfa M.* DML-CZ Metadata Editor // In: Sojka P. (ed.) Towards Digital Mathematics Library. Masaryk University, 2010. P. 139–151. URL: [https://dml.cz/bitstream/handle/10338.dmlcz/702537/DML\\_001-2008-1\\_17.pdf](https://dml.cz/bitstream/handle/10338.dmlcz/702537/DML_001-2008-1_17.pdf).

19. *Bartošek M., Rákosník J.* DML-CZ: The Experience of a Medium-Sized Digital Mathematics Library // Notices of the AMS. 2013. V. 60. No 8. P. 1028–1033.

---

URL: <http://dx.doi.org/10.1090/noti1031>.

20. *Bouche T.* Toward a digital mathematics library? // In: Borwein J.M., Rocha E.M., Rodrigues J.F. (Eds) *Communicating Mathematics in the Digital Era*. – Taylor & Francis, 2008. P. 47–73. URL: <https://hal.archives-ouvertes.fr/hal-00347682>.

21. *Lange C.* Ontologies and Languages for Representing Mathematical Knowledge on the Semantic Web // *Semantic Web*. 2013. V. 4. No. 2. P. 119–158. URL: <https://content.iospress.com/articles/semantic-web/sw059>.

22. *Serebryakov Vladimir.* Semantic digital libraries. What is it? // *CEUR Workshop Proceedings*. 2014. V. 1297. P. 1–5. URL: [http://ceur-ws.org/Vol-1297/1-5\\_paper-1.pdf](http://ceur-ws.org/Vol-1297/1-5_paper-1.pdf).

23. *Elizarov A.M., Kirillovich A.V., Lipachev E.K., Nevzorova O.A.* Mathematical knowledge management: ontological models and digital technology // *Analitika i upravlenie danny`mi v oblastiakh s intensivny`m ispol`zovaniem danny`x: sbornik statej XVIII Mezhdun. konf. DAMDID/RCDL'2016*. M.: FICz IU RAN, 2016. S. 95–101..

24. *Elizarov A.M., Kirillovich A.V., Lipachev E.K., Nevzorova O.A.* Mathematical Knowledge Management: Ontological Models and Digital Technology // *CEUR Workshop Proceedings*. 2016. V. 1752. P. 44–50. URL: <http://ceur-ws.org/Vol-1752/paper08.pdf>

25. *Bouche T., Labbe O.* The New Numdam Platform // In: Geuvers H., England M., Hasan O., Rabe F., Teschke O. (Eds) *Intelligent Computer Mathematics. CICM 2017. Lecture Notes in Computer Science*. Vol. 10383. Springer, Cham, 2017. P. 70–82. URL: [https://doi.org/10.1007/978-3-319-62075-6\\_6](https://doi.org/10.1007/978-3-319-62075-6_6). <https://zenodo.org/record/581405/>.

26. *Sadegh A., Lange C., Vidal M.E., Auer S.* Integration of Scholarly Communication Metadata using Knowledge Graphs // *International Conference on Theory and Practice of Digital Libraries*. 2017. P. 328–341.

27. *Ataeva O.M., Serebryakov V.A.* Information Model of LibMeta Digital Library // *Lobachevskii J. of Mathematics*. 2019. V. 40. No 7. P. 861–875. URL: <https://doi.org/10.1134/S1995080219070035>.

28. *Developing a 21st Century Global Library for Mathematics Research* // Washington: The National Academies Press, 2014. 142 p. doi:10.17226/18619.

29. *Olver P.J.* The World Digital Mathematics Library: Report of a Panel

Discussion // Proceedings of the International Congress of Mathematicians, August 13–21, 2014, Seoul, Korea. Kyung Moon SA, 1. 2014. P. 773–785.

30. *Watt S.* How to Build a Global Digital Mathematics Library // 18th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC). 2016. P. 37–40.

31. *Ion P.D.F., Watt S.M.* The Global Digital Mathematics Library and the International Mathematical Knowledge Trust // ICM 2017: Intelligent Computer Mathematics, 2017. Lecture Notes in Artificial Intelligence. 2017. V. 10383. P. 56–69. URL: [https://doi.org/10.1007/978-3-319-62075-6\\_5](https://doi.org/10.1007/978-3-319-62075-6_5).

32. *Bouche T.* Reviving the free public scientific library in the digital age? The EuDML project // In: Kaiser K., Krantz S.G., Wegner B. (Eds) Topics and Issues in Electronic Publishing JMM/AMS Special Session, FIZ Karlsruhe. 2013. P. 57–80. URL: <https://www.emis.de/proceedings/TIEP2013/05bouche.pdf>.

33. *Bouche T., Rákosník J.* Report on the EuDML External Cooperation Model // In: Kaiser K., Krantz S.G., Wegner B. (Eds) Topics and Issues in Electronic Publishing, JMM, Special Session. San Diego. 2013. P. 99–108. URL: [https://www.emis.de/proceedings/TIEP2013/07bouche\\_rakosnik.pdf](https://www.emis.de/proceedings/TIEP2013/07bouche_rakosnik.pdf).

34. *Sylwestrzak W., Borbinha J., Bouche T., Nowiński A., Sojka P.* EuDML – Towards the European Digital Mathematics Library // In: Sojka P.(ed.) Towards a Digital Mathematics Library. Masaryk University, 2010. P. 11–26. URL: <https://eudml.org/doc/220786>.

35. *Elizarov A.M., Lipachev E.K.* Lobachevskii DML: Towards a Semantic Digital Mathematical Library of Kazan University // CEUR Workshop Proceedings. 2017. V. 2022. P. 326–333.

36. *Elizarov A.M., Lipachev E.K.* Semanticheskie metody` i instrumenty` e`lektronnoj matematicheskoy biblioteki Lobachevskii-DML // Nauchny`j servis v seti Internet: trudy` XIX Vserossijskoj nauchnoj konferencii (18–23 sentyabrya 2017 g., g. Novorossijsk). M.: IPM im. M.V. Keldy`sha, 2017. S. 130–136. <https://doi.org/10.20948/abrau-2017-73>. URL: <http://keldysh.ru/abrau/2017/73.pdf>.

37. *Elizarov A.M., Lipachev E.K., Khaydarov S.M.* Structure and Services Digital Mathematical Library Lobachevskii-DML // Ucheny`e zapiski ISGZ. 2017. № 1 (15). S. 215–220..

38. *Růžička M., Sojka P., Krejčíř V.* Towards Machine-Actionable Modules of a

Digital Mathematics Library // In: Carette J., Aspinall D., Lange C., Sojka P., Windsteiger W. (eds) Intelligent Computer Mathematics. CICM 2013. Lecture Notes in Computer Science. 2013. V. 7961. P. 263–277. URL: [https://doi.org/10.1007/978-3-642-39320-4\\_17](https://doi.org/10.1007/978-3-642-39320-4_17).

39. *Elizarov A.M., Lipachev E.K., Khaidarov Sh.M.* Semanticheskij analiz bol'shix kollekcij nauchnyx dokumentov // Trudy mezhdunarodnoj konferencii po komp'yuternoj i kognitivnoj lingvistike TEL-2016. Kazan': Izd-vo Kazan. un-ta, 2016. S. 21–25.

40. *Elizarov A.M., Lipachev E.K., Khaidarov Sh.M.* Automated Processing Service System of Large Collections of Scientific Documents // CEUR Workshop Proceedings. 2016. V. 1752. P. 58–64.

41. *Elizarov A.M., Khaidarov Sh.M., Lipachev E.K.* Scientific Documents Ontologies for Semantic Representation of Digital Libraries // Proc. of the 2nd Russia and Pacific Conf. on Computer Technology and Applications. 2017. P. 1–5. URL: <https://doi.org/10.1109/RPC.2017.8168064>.

42. *Nevzorova O., Zhiltsov N., Kirillovich A., Lipachev E.* OntoMathPRO Ontology: A Linked Data Hub for Mathematics // Klinov P., Mouromstev D. (eds.) Proceedings of the 5th International Conference on Knowledge Engineering and Semantic Web (KESW 2014). Communications in Computer and Information Science, Springer, Cham, 2014. V. 468. P. 105–119. URL: [https://doi.org/10.1007/978-3-319-11716-4\\_9](https://doi.org/10.1007/978-3-319-11716-4_9).

43. *Elizarov A., Kirillovich A., Lipachev E., Nevzorova O., Solovyev V., Zhiltsov N.* Mathematical Knowledge Representation: Semantic Models and Formalisms // Lobachevskii J. of Mathematics. 2014. V. 35. No 4. P. 347–353. URL: <https://doi.org/10.1134/S1995080214040143>.

44. *Elizarov A.M., Lipachev E.K., Nevzorova O.A., Solov'ev V.D.* Methods and Means for Semantic Structuring of Electronic Mathematical Documents // Doklady Mathematics. 2014. 90 (1). P. 521–524. URL: <https://doi.org/10.1134/S1064562414050275>.

45. *Elizarov A.M., Zhiltsov N.G., Kirillovich A.V., Lipachev E.K., Nevzorova O.A.* E-kosistema ONTOMATH i proekt Vsemirnoj cifrovoj matematicheskoy biblioteki // Trudy mezhdunarodnoj konferencii po komp'yuternoj i kognitivnoj lingvistike TEL-

2016. Kazan` : Izd-vo Kazan. un-ta, 2016. S. 25–28.

46. *Elizarov A.M., Kirillovich A.V., Lipachev E.K., Nevzorova O.A.* Digital Ecosystem OntoMath: Mathematical Knowledge Analytics and Management // Communications in Computer and Information Science. Springer. 2017. V. 70. P. 33–46. URL: [https://doi.org/10.1007/978-3-319-57135-5\\_3](https://doi.org/10.1007/978-3-319-57135-5_3).

47. *Elizarov A., Kirillovich A., Lipachev E., Nevzorova O.* Semantic Formula Search in Digital Mathematical Libraries // Proceedings of the 2nd Russia and Pacific Conference on Computer Technology and Applications (RPC 2017). IEEE, 2017. P. 39–43. URL: <https://doi.org/10.1109/RPC.2017.8168063>.

48. *Birialtsev E., Gusenkov A., Zhibrik O., Gusenkova P., Palacheva Y.* Search in Collections of Mathematical Articles // In: Rocha Á., Adeli H., Reis L.P., Costanzo S. (eds) Trends and Advances in Information Systems and Technologies. WorldCIST'18 2018. Advances in Intelligent Systems and Computing, Springer, Cham, 2018. V. 745. P. 561–567. URL: [https://doi.org/10.1007/978-3-319-77703-0\\_55](https://doi.org/10.1007/978-3-319-77703-0_55).

49. *Shakirova L.R., Falileeva M.V., Kirillovich A.V., Lipachev E.K., Nevzorova O.A., Nevzorov V.N.* OntoMathEdu – Educational Mathematical Ontology: Structure and Relationships // Nauchny`j servis v seti Internet: trudy` XXI Vserossijskoj nauchnoj konferencii (23–28 sentyabrya 2019 g., g. Novorossijsk). M.: IPM im. M.V. Keldy`sha, 2019. S. 653–661. <https://doi.org/10.20948/abrau-2019-84>.

50. *Shakirova L., Falileeva M., Kirillovich A., Lipachev E.* Modeling and evaluation of the mathematical educational ontology //CEUR Workshop Proceedings. 2020. V. 2543. P. 305–319.

51. *Kirillovich A., Shakirova L., Falileeva M., Lipachev E.* Towards an Educational Mathematical Ontology // L. Gómez Chova, et al. (eds). Proceedings of the 13th International Technology, Education and Development Conference (INTED2019), Valencia, Spain, March 11th–13th, 2019. IATED, 2019. P. 6823–6829.

52. *Elizarov A.M., Zhiltsov N.G., Kirillovich A.V., Lipachev E.K.* Semanticheskoe annotirovanie v sisteme upravleniya fiziko-matematicheskim kontentom // Nauchny`j servis v seti Internet: trudy` XVII Vserossijskoj nauchnoj konferencii (21–26 sentyabrya 2015 g., g. Novorossijsk). M.: IPM im. M.V. Keldy`sha, 2015. S. 98–103. URL: [https://kpfu.ru//staff\\_files/F1890276653/Elizarov\\_at\\_all\\_.pdf](https://kpfu.ru//staff_files/F1890276653/Elizarov_at_all_.pdf).

53. *Elizarov A.M., Zhiltsov N.G., Kirillovich A.V., Lipachev E.K.* Mathematical Scholarly Papers Recommendation Service // Trudy` XVII Mezhdunarodnoj konferencii



DAMDID/RCDL'2015 «Analitika i upravlenie danny`mi v oblastiakh s intensivny`m ispol`zovaniem danny`x». Obninsk: IATE` NIYaU MIFI, 2015. S. 357–350. URL: [https://kpfu.ru//staff\\_files/F684099727/damdid2015\\_paper\\_Elizarov\\_new.pdf](https://kpfu.ru//staff_files/F684099727/damdid2015_paper_Elizarov_new.pdf).

54. *Elizarov A.M., Kirillovich A.V., Lipachev E.K., Zhizhchenko A.B., Zhil'tsov N.G.* Mathematical Knowledge Ontologies and Recommender Systems for Collections of Documents in Physics and Mathematics // *Doklady Mathematics*. 2016. 93 (2). P. 1–3. <https://doi.org/10.1134/S1064562416020174>.

55. *Khaydarov S.M., Yamalytdinova G.S.* Algorithm for Forming the Dictionary of the Recommender System of Selecting Classifiers of Scientific Information // *Ucheny`e zapiski ISGZ*. 2017. № 1 (15). S. 552–557.

56. *Khaydarov Sh.M., Yamalutdinova G.Sh.* Rekomendatel`naya sistema klassifikatsii fiziko-matematicheskix dokumentov // *Nauchny`j servis v seti Internet: trudy` XX Vserossijskoj nauchnoj konferencii (17–22 sentyabrya 2018 g., g. Novorossijsk)*. M.: IPM im. M.V. Keldy`sha, 2018. S. 480–486 <https://doi.org/10.20948/abrau-2018-57>. URL: <http://keldysh.ru/abrau/2018/theses/57.pdf>.

57. *Khaydarov S.M., Yamalutdinova G.S.* Recommender system of physical and mathematical documents classification // *CEUR Workshop Proceedings*. 2018. V. 2260. P. 480–486.

58. *Harper C.* Metadata normalization: a case study in Primo and linked open data in libraries // *Metadata Working Group Forum, Cornell, 2008*. URL: <https://ecommons.cornell.edu/handle/1813/10920>.

59. *Koh J., Hong D., Gupta R., Whitehouse K., Wang H., Agarwal Y.* Plaster: An Integration, Benchmark, and Development Framework for Metadata Normalization Methods // *BuildSys'18: Proceedings of the 5th Conference on Systems for Built Environments*. 2018. P. 1–10. URL: <https://doi.org/10.1145/3276774.3276794>.

60. *Koh J., Balaji B., Sengupta D., McAuley J., Gupta R., Agarwal Y.* Scrabble: Transferrable Semi-Automated Semantic Metadata Normalization using Intermediate Representation // *BuildSys'18: Proceedings of the 5th Conference on Systems for Built Environments*. 2018. P. 11–20. URL: <https://doi.org/10.1145/3276774.3276795>.

61. Journal Article Tag Suite. URL: <https://jats.nlm.nih.gov/about.html>.

62. ANSI/NISO Z39.96-2012, JATS: Journal Article Tag Suite (ver. 1.0). URL: [https://groups.niso.org/apps/group\\_public/download.php/10904/z39.96-2012.pdf](https://groups.niso.org/apps/group_public/download.php/10904/z39.96-2012.pdf).

63. Journal Archiving and Interchange Tag Set, ver. 1.1. URL: <https://jats.nlm.nih.gov/publishing/1.1/>.

64. Journal Archiving and Interchange Tag Set, ver. 1.2. URL: <https://jats.nlm.nih.gov/archiving/1.2/>.

65. JATS: Journal Article Tag Suite. National Information Standards Organization, ver. 1.2 (ANSI/NISO Z39.96-2019). 8 Febr. 2019. 652 p. URL: [https://groups.niso.org/apps/group\\_public/download.php/21030/ANSI-NISO-Z39.96-2019.pdf](https://groups.niso.org/apps/group_public/download.php/21030/ANSI-NISO-Z39.96-2019.pdf).

66. Journal Publishing Tag Library NISO JATS, version 1.3d1 (ANSI/NISO Z39.96-2019). October 2019. URL: <https://jats.nlm.nih.gov/archiving/1.3d1/>.

67. *Heller L., The R., Bartling S.* Dynamic Publication Formats and Collaborative Authoring // In: Bartling S. and Friesike S. (eds.). *Opening Science*. Springer Cham, 2014. P. 191–211. URL: [https://doi.org/10.1007/978-3-319-00026-8\\_13](https://doi.org/10.1007/978-3-319-00026-8_13).

68. *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 // *CEUR Workshop Proceedings*. 2018. V. 2260. P. 175–185.

69. *Kirillovich A.V.* Information Architecture of Blogs // *Russian Digital Libraries Journal*. 2017. V. 20. No 2. P. 147–162. URL: <https://elbib.ru/article/view/418/504>

70. *Elizarov A.M., Kirillovich A.V., Lipachev E.K.* Blogs in Scientific Communications Systems // *Ucheny`e zapiski ISGZ*. 2017. №1 (15). S. 209–214.

71. *Gafurova P.O.* Forumy` v sisteme nauchny`x kommunikacij // *Sbornik konferencii: Informacionny`e texnologii v obrazovanii i nauke (ITON-2018)*. 2018. S. 102–103.

72. *Borisov N.V., Zakharkina N.V., Mbogo I.A., Prokudin D.E., Scherbakov P.P.* Problems of creating an online scientific journal with multimedia content // *Nauchny`j servis v seti Internet: trudy` XXI Vserossijskoj nauchnoj konferencii (23–28 sentyabrya 2019 g., g. Novorossijsk)*. M.: IPM im. M.V. Keldy`sha, 2019. S. 153–165. <https://doi.org/10.20948/abrau-2019-87>. URL: <http://keldysh.ru/abrau/2019/theses/87.pdf>

73. *Borisov N.V., Zakharkina N.V., Mbogo I.A., Prokudin D.E., Scherbakov P.P.* Challenges of Publishing Online Scholarly Journals with Multimedia Content // *CEUR Workshop Proceedings*. 2020. V. 2543. P. 93–102. URL: <http://ceur-ws.org/Vol->

2543/rpaper09.pdf.

74. ORCID. Connecting Research and Researchers. URL: <https://orcid.org/>.
75. EuDML metadata schema specification (v2.0–final). URL: <https://initiative.eudml.org/eudml-metadata-schema-specification-v20-final>.
76. *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>.
77. D7.2: Toolset for Image and Text Processing and Metadata Editing – Initial release. URL: <http://www.mathdoc.fr/publis/d7.2-v1.0.pdf>.
78. D7.3: Toolset for Image and Text Processing and Metadata Enhancements – Value release. URL: <http://www.mathdoc.fr/publis/d7.3.pdf>.
79. D7.4: Toolset for Image and Text Processing and Metadata Enhancements – Final Release. URL: <https://wiki.eudml.eu/mediawiki/eudml/images/D7.4-v1.0.pdf>.
80. *Řehůřek R., Sojka P.* Automated Classification and Categorization of Mathematical Knowledge // In: Autexier S., Campbell J., Rubio J., Sorge V., Suzuki M., Wiedijk F. (Eds) Intelligent Computer Mathematics. CICM 2008. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, 2008. V. 5144. P. 543–557. URL: [https://doi.org/10.1007/978-3-540-85110-3\\_44](https://doi.org/10.1007/978-3-540-85110-3_44).
81. Open Archives Initiative Protocol for Metadata Harvesting. URL: <http://www.openarchives.org/OAI/openarchivesprotocol.html>.
82. *Fedotov A.M., Baidavletov A.T., Zhizhimov O.L., Sambetbayeva M.A., Fedotova O.A.* Digital repository of scientific and educational information system // Vestnik NSU. Series: Information Technologies. 2015. V. 13. No 3. P. 68–86. URL: <https://cyberleninka.ru/article/n/tsifrovoy-repozitoriy-v-nauchno-obrazovatelnoy-informatsionnoy-sisteme>.
83. *Krejčíř V.* Building the Czech Digital Mathematics Library upon DSpace System // In: Sojka P. (Ed) DML 2008. Towards Digital Mathematics Library. Brno: Masaryk University, 2008. P. 117–126. URL: [https://dml.cz/bitstream/handle/10338.dmlcz/702539/DML\\_001-2008-1\\_14.pdf](https://dml.cz/bitstream/handle/10338.dmlcz/702539/DML_001-2008-1_14.pdf).
84. *MacGregor J., Stranack K., Willinsky J.* The Public Knowledge Project: Open Source Tools for Open Access to Scholarly Communication // 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. 165–175. URL: [https://doi.org/10.1007/978-3-319-00026-8\\_3](https://doi.org/10.1007/978-3-319-00026-8_3).

85. Expressing Dublin Core metadata using XML. URL: <https://www.dublincore.org/specifications/dublin-core/dc-xml/>.

86. What is a MARC record and why is it important? URL: <https://www.loc.gov/marc/umb/um01to06.html>.

87. *MARC Standarts*. URL: <http://www.loc.gov/marc/>.

88. *Elizarov A.M., Lipachev E.K., Khaidarov S.M.* Programma avtomatizirovannogo formirovaniya metadanny`x v formate Rossijskogo indeksa nauchnogo citirovaniya dlya statej zhurnala "E`lektronny`e biblioteki". Svidetel`stvo o registracii programmy` dlya E`VM RU 2018612458, 16.02.2018. Zayavka № 2017663206 ot 19.12.2017. <https://www.elibrary.ru/item.asp?id=39291526>.

89. *Gerasimov A.N., Elizarov A.M., Lipachev E.K.* Subsystem of Formation Metadata for Science Index Databases on Management Platform Electronic Scientific Journals // *Russian Digital Libraries Journal*. 2015. V. 18. No 1–2. P. 6–31. URL: <https://elbib.ru/article/view/356/447>.

90. *Akhmetov D.Yu., Elizarov A.M., Lipachev E.K.* Service-oriented Information System of "Russian Digital Libraries Journal" // *Russian Digital Libraries Journal*. 2016. V. 19. No 1. P. 2–39. URL: <https://elbib.ru/article/view/377/468>.

91. *Ley M.* DBLP – Some Lessons Learned // *Proceedings of the VLDB Endowment*. 2009. V. 2 (2). P. 1493–1500.

92. *Gafurova P.O.* Metody` normalizacii metadanny`x cifrovyy`x matematicheskix bibliotek // V knige: Lomonosov-2019 Sbornik tezisov XXVI Mezhdunarodnoj nauchnoj konferencii studentov, aspirantov i molody`x uchenyy`x; sekciya «Vy`chislitel`naya matematika i kibernetika». 2019. S. 162–164.

93. *Gafurova P.O., Elizarov A.M., Lipachev E.K.* Methods of Normalization of Metadata Digital Mathematical Collections // *Uchenyy`e zapiski ISGZ*. 2019. T. 17. № 1. S. 141–148.

94. *Gafurova P.O., Elizarov A.M., Lipachev E.K., Khammatova D.M.* Metadata Normalization Methods in the Digital Mathematical Library // *CEUR Workshop Proceedings*. 2020. V. 2543. P. 136–148.

95. *Gafurova P.O., Elizarov A.M., Lipachev E.K., Khammatova D.M.* Methods of Formation and Normalization of Metadata in the Digital Mathematical Library //

Nauchnyj servis v seti Internet: trudy XXI Vserossijskoj nauchnoj konferencii. 2019. S. 234–244. <https://doi.org/10.20948/abrau-2019-28>. URL: <http://keldysh.ru/abrau/2019/theses/28.pdf>.

96. *Elizarov A.M., Zaitseva N.V., Lipachev E.K., Khaidarov S.M.* Programma avtomatizirovannogo formirovaniya metadannyx dokumentov cifrovoj matematicheskoj biblioteki Lobachevskii DML. Svidetel'stvo o registracii programmy dlya E`VM RU 2019611328, 24.01.2019. Zayavka № 2019610406 ot 15.01.2019. <https://www.elibrary.ru/item.asp?id=39309871>.

97. *Batyrshina R.R Elizarov A.M., Lipachev E.K* Organization of Digital Mathematics Library Collections by Semantic Analysis Methods // Nauchnyj servis v seti Internet: trudy` XXI Vserossijskoj nauchnoj konferencii (23–28 sentyabrya 2019 g., g. Novorossijsk). M.: IPM im. M.V. Keldy`sha, 2019. S. 85–90. <https://doi.org/10.20948/abrau-2019-97>. URL: <http://keldysh.ru/abrau/2019/theses/97.pdf>.

98. *Elizarov A.M., Lipachev E.K.* Methods of Processing Large Collections of Scientific Documents and the Formation of Digital Mathematical Library // CEUR Workshop Proseedings. 2020. V. 2543. P. 354–360.

99. *Gafurova P.O., Lipachev E.K.* Methods for the Semantic Representation of Mathematical Collections of the Lobachevskii Digital Mathematics Library // Trudy matematicheskogo centra im. N.I. Lobachevskogo. 2018. T. 56. S. 90–93.

## СВЕДЕНИЯ ОБ АВТОРАХ



**ГАФУРОВА Полина Олеговна** – магистр математики, аспирант Высшей школы информационных технологий и информационных систем Казанского (Приволжского) федерального университета.

**Polina GAFUROVA** – Magister of Mathematics, Kazan (Volga Region) Federal University. Current scientific interests: data mining, recommender systems, cloud computing, knowledge extraction technologies.

email: pogafurova@gmail.com; ORCID: 0000-0002-1544-155X



**ЕЛИЗАРОВ Александр Михайлович** – доктор физико-математических наук, профессор, заслуженный деятель науки Республики Татарстан, профессор кафедры программной инженерии Высшей школы информационных технологий и интеллектуальных систем Казанского (Приволжского) федерального университета.

**Alexander ELIZAROV** – Doctor of Physics and Mathematics, Professor, Honored Worker of Science of the Republic of Tatarstan, Kazan Federal University. Current scientific interests: data mining, recommender systems, cloud computing, knowledge extraction technologies.

email: amelizarov@gmail.com; ORCID: 0000-0003-2546-6897



**ЛИПАЧЁВ Евгений Константинович** – кандидат физико-математических наук, доцент кафедры Интеллектуальных технологий поиска Высшей школы информационных технологий и интеллектуальных систем Казанского (Приволжского) федерального университета.

**Evgeny Konstantinovich LIPACHEV** – Candidate of Physics and Mathematics, Associate Professor, Kazan Federal University. Current scientific interests: data mining, recommender systems, cloud computing, knowledge extraction technologies.

email: elipachev@gmail.com; ORCID: 0000-0001-7789-2332

*Материал поступил в редакцию 21 ноября 2019 года*

---