

## Фабрики метаданных в цифровых библиотеках

А.М. Елизаров<sup>1,2</sup>, Е.К. Липачёв<sup>1</sup>

<sup>1</sup> Казанский (Приволжский) федеральный университет,

<sup>2</sup> Казанский филиал Межведомственного суперкомпьютерного центра РАН

amelizarov@gmail.com, elipachev@gmail.com

### Аннотация

Обсуждены направления развития и использования цифровых технологий в научной деятельности на базе цифровых платформ, значение и роль цифровых библиотек в формировании таких платформ, а также проблемы обеспечения интегрированности (связности) извлекаемой информации.

Решен ряд задач, связанных с построением фабрики метаданных цифровой математической библиотеки Lobachevskii-DML. Под фабрикой метаданных понимается система взаимосвязанных программных инструментов, которые позволяют создавать, обрабатывать и хранить метаданные объектов цифровых библиотек, а также управлять ими и интегрировать создаваемые электронные коллекции в агрегирующие цифровые научные библиотеки.

Предложено использовать эту реализованную фабрику метаданных как элемент экосистемы любой научной цифровой библиотеки.

**Ключевые слова:** научная цифровая платформа, научная цифровая библиотека, экосистема цифровой библиотеки, метаданные объектов цифровых библиотек, фабрика метаданных, цифровая математическая библиотека Lobachevskii-DML

**Библиографическая ссылка:** Елизаров А.М., Липачёв Е.К. Фабрики метаданных в цифровых библиотеках // Информационное общество: образование, наука, культура и технологии будущего. Выпуск 4 (Труды XXIII Международной объединенной научной конференции «Интернет и современное общество», IMS-2020 (сборник научных статей). — СПб: Университет ИТМО, 2020. С. 25-33. DOI: 10.17586/2587-8557-2020-4-25-33

### 1. Цифровые платформы и цифровые библиотеки

Как известно, сегодня цифровая экономика понимается как экономика, основанная на процессах производства и использования цифровых технологий (см., например, [1]). Эти процессы в значительной степени реализуются в настоящее время на базе цифровых платформ, организованных в различных предметных областях и сферах деятельности. Такие платформы обладают своими наборами сервисов и позволяют решать разнообразные комплексы задач по развитию и использованию цифровых технологий. Само понятие цифровой платформы и некоторые особенности текущего этапа их развития проанализированы в [2], а в [3] даны оценки современного уровня развития цифровых платформ в России, включающие четыре основных компонента: определение многосторонних цифровых платформ, факторы развития цифровых платформ, бизнес-модели и динамика конкуренции. Отметим, что формирование цифровых платформ для исследований и разработок предусмотрено программой «Цифровая экономика Российской Федерации» [4, гл. 1], где отмечена необходимость создания цифровых платформ для фундаментальных и прикладных исследований, научно-исследовательских и опытно-конструкторских разработок, как раз и составляющих основные направления деятельности в области исследований и разработок.

С каждой создаваемой цифровой платформой, как правило, связана организация, которая выступает в роли оператора платформы и формирует вокруг нее собственную экосистему.

Еще одно направление развития и использования цифровых технологий в научной деятельности предусматривает организацию на современном уровне и с использованием цифровых технологий доступа к новейшим научным результатам, в частности, научным публикациям и наукометрической информации о них. Исторически это направление связано с формированием в мире цифровых (или электронных) библиотек, в том числе, научных, активное развитие которых началось в конце XX в. (см., например, [5–7]). Вообще под цифровыми (электронными) библиотеками любой направленности (не только научной) понимают модели сложных информационных систем, которые служат основой создания универсальных распределенных хранилищ знаний и снабжены средствами навигации и поиска в коллекциях разнородных электронных документов, включаемых в эти хранилища.

В настоящее время цифровые библиотеки существуют во всех развитых странах мира. В России примерами таковых являются:

- цифровая библиотека Соционет (<https://socionet.ru>, год организации – 2000), которая обеспечила участие России в разработке международной онлайн-научно-образовательной инфраструктуры (первоначально – в области общественных наук, в настоящее время – во всех научных дисциплинах);
- крупнейшая российская цифровая научная библиотека eLibrary (<https://elibrary.ru/>, год организации – 2005), которая интегрирована с Российским индексом научного цитирования (РИНЦ);
- цифровая научная библиотека Cyberleninka (<https://cyberleninka.ru>, год организации – 2012), построенная на основе концепции открытой науки (Open Science) и входящая в пятерку крупнейших открытых научных архивов мира;
- общероссийский математический портал MathNet.RU (<http://www.mathnet.ru/>, год организации – 2006), содержащий архивы ведущих российских математических и физических журналов, журналов по информатике, коллекцию видеолекций, инструменты навигации и поиска, информационную систему управления редакционными процессами.

Кроме перечисленных в России создано большое число разнообразных цифровых библиотек, связанных с современными публикационными и наукометрическими сервисами. Примерами последних являются Mendeley (<https://www.mendeley.com>, год организации – 2008) – инструмент для управления персональной научной библиотекой и эффективной совместной научной работы; ИСТИНА (<https://istina.msu.ru/>, год организации – 2014) – инструмент сбора, систематизации, хранения и анализа наукометрической информации для подготовки и принятия управленческих решений.

Крупнейшими международными научными цифровыми библиотеками являются наукометрические базы данных:

- Web of Science (до 2014 года – Web of Knowledge) (<https://clarivate.com/webofsciencigroup/solutions/web-of-science/>) – возникла в 1961 году как продукт американской компании ISI (Institute for Scientific Information), позже принадлежала медиакорпорации Thomson Reuters и стала цифровой, с 2016 года принадлежит компании Clarivate Analytics; основной продукт – Web of Science Core Collection;
- Scopus (<https://www.scopus.com/>, год основания – 2004) – крупнейшая база данных, содержащая информацию о цитировании рецензируемой научной литературы.

Названные цифровые библиотеки (а также большой ряд других) играют огромную роль в ускорении оборота имеющегося знания и доступа к нему. Но без интернета, ставшего сегодня всеобъемлющей интегрированной информационной средой, извлечение информации из разного рода информационных источников (баз данных), каковыми являются разнообразные цифровые библиотеки, было бы невозможным. Одновременно возникает целый ряд серьезных проблем обеспечения интегрированности (связности)

извлекаемой информации. С этой точки зрения сужение всего пространства имеющейся информации дает возможность точнее специфицировать информацию и, следовательно, обеспечить более качественные доступ к ней и ее использование. Такое сужение обеспечено в рамках специализированных научных цифровых библиотек, организованных в конкретных предметных областях. Например, высокого уровня организации достигли математические цифровые библиотеки, история возникновения и развития которых представлена, например, в [8].

Таким образом, с одной стороны, в сфере науки и научных исследований в настоящее время сформировано значительное количество разнообразных научных цифровых библиотек, реализующих широкий спектр поисковых сервисов и обладающих собственной экосистемой. С другой стороны, примеры созданных научных цифровых платформ, реализующих собственные функционалы и сервисы взаимодействия пользователей в соответствии с базовыми определениями [2, 3], а также собственные бизнес-модели, в настоящее время отсутствуют. Представляется, что базой для построения таких цифровых платформ могут послужить научные цифровые библиотеки, имеющиеся (не у всех) экосистемы которых должны быть усовершенствованы (созданы). Одно из направлений такого усовершенствования обсуждено ниже.

## **2. Метаданные и навигация в научном информационном пространстве**

Хорошо известно, что навигация в информационном пространстве в значительной степени обеспечивается сегодня наличием и полнотой набора метаданных (данных о данных) документов, представленных в сети (например, [9–13]). Сегодня имеется достаточно много различных стандартов метаданных. Эти стандарты должны обеспечивать возможности interoperабельности с внешней средой, идентификации и интеграции информации, ее поиска в распределенной среде, быть открытыми и расширяемыми, ориентированными на современные семантические и цифровые технологии. Но даже при наличии таких стандартов реально обеспечить нужные свойства метаданных различных документов весьма затруднительно. Значительно проще задачи стандартизации метаданных решаются применительно к конкретной предметной области и на базе тех цифровых библиотек, которые созданы в этой области. Примером служит область математических и компьютерных наук, где создано значительное число цифровых библиотек, выполняющих разнообразные функции интеграции математических знаний. Особенности представления метаданных документов в различных цифровых математических библиотеках описаны в [8].

Нам представляется, что существенным элементом экосистемы любой цифровой библиотеки должна стать фабрика метаданных этой библиотеки. Как и в [8], мы используем термин «фабрика метаданных цифровой библиотеки» (*metadata factory of digital library*) в следующем смысле (ср. с [14]): фабрика метаданных – это система взаимосвязанных программных инструментов, направленных на создание, обработку, хранение и управление метаданными объектами 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. Отметим, что, как и в [8], термин «нормализация» использован здесь для обозначения методов формирования и преобразования метаданных документов в соответствии с правилами и XML-схемами цифровых библиотек и наукометрических баз 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.

### 3. Фабрика метаданных

Укажем основные задачи, которые должны быть решены в рамках фабрики метаданных цифровой библиотеки. Прежде всего, фабрика метаданных должна обеспечить автоматизированную интеграцию имеющихся репозиториях (хранилищ соответствующих документов) с другими информационными системами. Этот процесс может быть основан на известных моделях агрегирования и распространения метаданных. Одной из таких моделей является OAI Protocol for Metadata Harvesting (<http://www.openarchives.org/OAI/openarchivesprotocol.html>, далее OAI-PMH). Модель OAI-PMH поддерживается большинством систем, предназначенных для хранения информационных ресурсов. Для организации работы с моделью OAI-PMH необходимо использовать цифровые хранилища и соответствующие системы их поддержки. В настоящее время существует целый ряд таких хранилищ, а наиболее известными из них являются DSpace, Eprints, Fedora и Greenstone. Отметим, что некоторые цифровые библиотеки используют специализированные методы харвестинга (сбора) метаданных из других хранилищ. В этом случае необходимо, чтобы у поставщиков данных были инструменты и сервисы, которые позволяют распространять метаданные.

Для организации взаимодействия сервисов как в рамках цифровой библиотеки, так и с внешними библиотеками и базами данных, необходимо учитывать те форматы метаданных, которые в них используются. Даже в одной цифровой библиотеке программные инструменты работают с несколькими форматами метаданных, что связано как с особенностями формирования цифрового контента, так и требованиями, которые предъявляют агрегирующие цифровые библиотеки и известные наукометрические базы данных, названные выше. Отметим лишь наиболее распространенные форматы метаданных, с которыми приходится иметь дело при организации взаимодействия сервисов в цифровых библиотеках (полные их описания доступны в интернете). Прежде всего, это формат Dublin Core и его расширения, формат каталогизации MARC, форматы библиографических ссылок RIS (Research Information Systems), AMSBib, XML-форматы РИНЦ и NISO JATS.

Далее, следующей важной функцией фабрики метаданных любой цифровой библиотеки является нормализация метаданных (в указанном выше смысле) в соответствии с форматами других агрегирующих библиотек. Например, модель OAI-PMH и соответствующий протокол требуют обязательного включения в описание ресурса набора метаданных в нотации oai\_dc, который основан на формате Dublin Core, включающем в себя 15 элементов для описания любого цифрового ресурса. Отметим, что нотация oai\_dc использует только названия ограниченного количества тегов Dublin Core.

Функции, описанные выше, были реализованы нами при построении фабрики метаданных цифровой математической библиотеки Lobachevskii-DML (<https://lobachevskii-dml.ru/>). Как и в случае любой цифровой научной библиотеки, формирование библиотеки Lobachevskii-DML и соответствующей фабрики метаданных потребовало привлечения как ранее созданных, так и новых разработанных нами технологических решений управления научным контентом. В результате метаданные документов цифровых математических коллекций с помощью фабрики метаданных цифровой библиотеки Lobachevskii-DML формируются в несколько этапов.

На этапе *препроцессорной обработки* соответствующая коллекция документов обрабатывается программными инструментами с целью приведения к виду, пригодному для дальнейшей автоматической обработки [15]. Для этого производится кластеризация документов по стиливому сходству, а затем выполняется приведение стиливых

конструкций, используемых в документе, к шаблонным. Например, .tex-коллекции могут содержать документы, в которых для оформления названия статьи используется не только команда `\title{}`, но и `\tit{}`, `\ArticleNAME{}` и другие. Более того, название статьи, список авторов, ключевые слова и другие блоки, необходимые для включения в состав метаданных, могут и не оформляться командами – в этих случаях производится попытка автоматически найти такие блоки по местоположению в тексте и шрифтовому оформлению. Такой подход применяется и к большинству коллекций документов, созданных в офисных форматах [16].

На этапе препроцессинга часть файлов не удается обработать автоматически, например, когда для обрабатываемого .tex-документа отсутствуют стилевой файл или файл с авторскими макроопределениями, на которые имеются ссылки в этом документе. Из таких файлов формируется набор, который приходится корректировать полуавтоматическими инструментами или вручную. После этого препроцессорная обработка проводится повторно, а файлы, отбракованные на этом этапе несколько раз, исправляются вручную.

Следующим этапом является *формирование набора основных метаданных*. Из документов извлекаются строки с названием статьи и списком авторов. Производятся поиск и выделение из текста кодов предметной классификации, блока ключевых слов, аффилиации авторов и аннотации к статье, если они содержатся в документе. Также в состав метаданных каждого документа включается его URL-ссылка в коллекции цифровой библиотеки. Сформированные метаданные сохраняются в xml-файле в соответствии с DTD-правилами и XML-схемами, принятыми в цифровой библиотеке.

На этапе *улучшения и уточнения метаданных* используются программные инструменты, с помощью которых исправляются простые орфографические ошибки и опечатки в названии статьи, списке авторов и ключевых словах. Производятся улучшение метаданных, в частности, транслитерирование названий статей, дополнение сокращений полными наименованиями («СПб» – «Санкт-Петербург», «LJM» – «Lobachevskii J. Math.», «Lobachevskii Journal of Mathematics»). С помощью сервисов уточнения проверяются URL-ссылки, и в состав метаданных включается дата обращения к веб-ресурсу. Формульные фрагменты в названиях статей и аннотациях к ним преобразуются в MathML-код.

Не все метаданные можно получить поиском соответствующих блоков в документе и последующим их экстрагированием из текста. Ключевые слова, коды классификаторов и другие данные определяются только как результат текстового и семантического анализа документа [17]. В фабрике метаданных эти операции выполняются на этапе формирования дополнительных метаданных. В рамках проекта создания цифровой математической библиотеки Lobachevskii-DML нами разработаны инструменты, позволяющие автоматизировать ряд операций этого этапа [18, 19].

Одной из функций фабрики метаданных является преобразование наборов метаданных в соответствии с XML-схемами и DTD-правилами агрегирующих библиотек и библиографических баз – этот процесс обозначен выше, как этап *нормализации метаданных*. В фабрике метаданных цифровой математической библиотеки Lobachevskii-DML реализованы методы нормализации соответствующих метаданных в форматы РИНЦ [20] и Библиографической базы по компьютерным наукам “dblp computer science bibliography” (DBLP, <https://dblp.uni-trier.de/>) [21, 22]. Кроме того, разработаны методы формирования обязательного и фундаментального наборов метаданных по XML-схемам, используемым Европейской математической библиотекой EuDML (<https://initiative.eudml.org/>) [22].

Подробно инструменты фабрики метаданных цифровой математической библиотеки Lobachevskii-DML, уже реализованные нами, описаны в [8].

## Заключение

Решен ряд задач, связанных с построением фабрики метаданных цифровой математической библиотеки Lobachevskii-DML. Основные полученные результаты состоят в следующем: разработаны система сервисов автоматизированного формирования метаданных электронных математических коллекций и 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; предложен алгоритм приведения метаданных к формату oai\_dc и генерации структуры архивов для импорта в цифровое хранилище DSpace; разработаны и реализованы методы интеграции электронных математических коллекций Казанского университета в отечественные и зарубежные цифровые математические библиотеки.

Модель фабрики метаданных, представленная выше, реализована в конкретной цифровой математической библиотеке и поэтому учитывает специфику обрабатываемого контента. Вместе с тем, она может быть использована как элемент экосистемы любой научной цифровой библиотеки.

Работа выполнена в рамках программы развития Регионального научно-образовательного математического центра Приволжского федерального округа, номер соглашения № 075-02-2020-1478/1.

Проведенные исследования выполнены при частичной финансовой поддержке Российского фонда фундаментальных исследований (РФФИ) в рамках проекта № 18-29-03086, РФФИ и Правительства Республики Татарстан в рамках проекта № 18-47-160012.

## Литература

- [1] Ершова Т.В. Концептуализация предметной области «цифровая экономика» как основа развития ее понятийного аппарата // Информационное общество. 2019. Вып. 6. С. 34–41.
- [2] Ершова Т.В., Хохлов Ю.Е. Цифровые платформы для исследований и разработок // Информационное общество. 2017. Вып. 6. С. 17–24.
- [3] Ефферин Я.Ю., Россото К.М., Хохлов Ю.Е. Цифровые платформы в России: конкуренция между национальными и зарубежными многосторонними платформами стимулирует экономический рост и инновации // Информационное общество. 2019. Вып. 1-2. С. 16–34.
- [4] Программа «Цифровая экономика Российской Федерации». Утверждена распоряжением Правительства РФ от 28 июля 2017 г. № 1632. URL: <http://government.ru/docs/28653/> (дата обращения: 21.03.2020).
- [5] Армс В. Электронные библиотеки. М.: ПИК ВИНТИ, 2001. 274 с. Пер. с англ. изд. W.Y. Arms. Digital libraries. Cambridge; London, 2000.
- [6] Антопольский А.Б., Майстрович Т.В. Электронные библиотеки: принципы создания. М.: Либерия-Бибинформ, 2007. 288 с.
- [7] Xie L., Matusiak K.K. Discover Digital Libraries: Theory and Practice. Elsevier Inc., 2016. 388 p.
- [8] Гафурова П.О., Елизаров А.М., Липачев Е.К. Базовые сервисы фабрики метаданных цифровой математической библиотеки Lobachevskii-dml // Электронные библиотеки. 2020. Т. 23, №3. С. 336–381.
- [9] Sicilia M.-A. (Ed.) Handbook of Metadata, Semantics and Ontologies. World Scientific Publishing Co. Pte. Ltd., 2014. 579 p.
- [10] 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/ (дата обращения: 21.03.2020).
- [11] 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>.
- [12] Коголовский 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> (дата обращения: 21.03.2020).
- [13] Коголовский М.Р. Метаданные в компьютерных системах // Программирование. 2013. Т. 39 (4). С. 28–46. URL: <http://www.ipr-ras.ru/articles/kogalov13-03.pdf> (дата обращения: 21.03.2020).
- [14] 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/> (дата обращения: 21.03.2020).
- [15] Елизаров А.М., Жильцов Н.Г., Кириллович А.В., Липачев Е.К., Невзорова О.А. Программный комплекс автоматического препроцессинга цифровых коллекций математических документов // Свидетельство о регистрации программы для ЭВМ RU 2019610241, 09.01.2019. Заявка № 2018664769 от 18.12.2018.
- [16] Елизаров А.М., Липачёв Е.К., Хайдаров Ш.М. Программный комплекс выделения метаданных из коллекций физико-математических документов, представленных в формате OpenXML // Свидетельство о регистрации программы для ЭВМ RU 2016618802, 08.08.2016. Заявка № 2016616511 от 21.06.2016.
- [17] 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: <http://keldysh.ru/abrau/2017/73.pdf>. doi:10.20948/abrau-2017-73 (дата обращения: 21.03.2020).
- [18] Герасимов А.Н., Елизаров А.М., Липачев Е.К., Хайдаров Ш.М. Методы автоматизированного извлечения метаданных научных публикаций для библиографических и реферативных баз цитирования // Информационное общество: образование, наука, культура и технологии будущего сборник научных статей. Труды XIX Международной объединенной научной конференции «Интернет и современное общество». Санкт-Петербург, 2016. С. 41–48.
- [19] 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.
- [20] 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.
- [21] Гафурова П.О., Елизаров А.М., Липачев 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: <http://keldysh.ru/abrau/2019/theses/28.pdf> doi:10.20948/abrau-2019-28 (дата обращения: 21.03.2020).

- [22] 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.

## Digital Library Metadata Factories

A.M. Elizarov <sup>1,2</sup>, E.K. Lipachev <sup>1</sup>

<sup>1</sup> Kazan (Volga Region) Federal University,

<sup>2</sup> Joint Supercomputer Center of the Russian Academy of Sciences, Kazan

We discuss the directions of development and use of digital technologies in scientific activities based on digital platforms, the importance and role of digital libraries in the formation of such platforms, as well as the problems of ensuring the integration (connectivity) of the extracted information. We have solved a number of problems associated with building a metadata factory for the Lobachevskii-DML digital mathematical library. We understand by a metadata factory a system of interconnected software tools that allow you to create, process and store metadata of digital library objects, as well as manage them and integrate the created digital collections into aggregating digital scientific libraries. We propose to use this implemented metadata factory as an element of the ecosystem of any scientific digital library.

**Keywords:** digital science platform; digital science library; digital library ecosystem; metadata of digital library objects; metadata factory; digital math library Lobachevskii-DML

**Reference for citation:** Elizarov A.M., Lipachev E.K. Digital Library Metadata Factories // Book Title.Vol. 4 (Proceedings of the XXIII International Joint Scientific Conference «Internet and Modern Society», IMS-2020, St. Petersburg, June 17-20, 2020). - St. Petersburg: ITMO University, 2020. P. 25 – 33. DOI: 10.17586/0000-0000-2020-4-25-33

## References

- [1] Ershova T.V. The Conceptualization of the Subject Area ‘Digital Economy’ as the Basis for the Development of Its Terminological Framework // Information society. 2019. № 6. P. 34–41.
- [2] Ershova T.V., Hohlov Yu.E. Digital Research & Development Platforms // Information society. 2017. № 6. P. 17–24.
- [3] Eferin Ya.Yu., Rossoto K.M., Hohlov Yu.E. Digital Platforms in Russia: Competition between National and Foreign Multisided Platforms Stimulates Growth and Innovation // Information society. 2019. № 1-2. P. 16–34.
- [4] Programma «Cifrovaya ekonomika Rossijskoj Federacii». Utverzhdena rasporyazheniem Pravitel'stva RF ot 28 iyulya 2017 g. № 1632. URL: <http://government.ru/docs/28653/>. (data obrashcheniya: 21.03.2020).
- [5] Arms W.Y. Digital libraries. Cambridge; London, 2000.
- [6] Antopol'skij A.B., Majstrovich T.V. Elektronnye biblioteki: principy sozdaniya. M.: Liberiya-Bibinform, 2007. 288 s.
- [7] Xie I., Matusiak K.K. Discover Digital Libraries: Theory and Practice. Elsevier Inc., 2016. 388 p.
- [8] Gafurova P.O., Elizarov A.M., Lipachev E.K. Basic Services of Factory Metadata Digital Mathematical Library Lobachevskii-DML // Russian Digital Libraries Journal. 2020. V. 23, No. 3. P. 336–381.
- [9] Sicilia M.-A. (Ed.) Handbook of Metadata, Semantics and Ontologies. World Scientific Publishing Co. Pte. Ltd., 2014. 579 p.
- [10] 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/ (data obrashcheniya: 21.03.2020).
- [11] 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>.
- [12] Kogalovsky M.R. Metadannye, ih svojstva, funkcii, klassifikaciya i sredstva predstavleniya // Trudy 14-j Vserossijskoj nauchnoj konferencii «Elektronnye biblioteki: perspektivnye metody i tekhnologii, elektronnye kollekcii» – RCDL-2012, Pereslavl'-Zalesskij, Rossiya, 15–18 oktyabrya 2012 g. Pereslavl'-Zalesskij, 2012. S. 3–14. URL: <http://rcdl.ru/doc/2012/paper3.pdf>. (data obrashcheniya: 21.03.2020).
- [13] Kogalovsky M.R. Metadata in Computer Systems // Programming and Computer Software. 2013. V. 39. No. 4. P. 182–193. URL: <http://www.ipr-ras.ru/articles/kogalov13-03.pdf>. (data obrashcheniya: 21.03.2020).
- [14] 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/>. (data obrashcheniya: 21.03.2020).
- [15] Elizarov A.M., Zhil'cov N.G., Kirillovich A.V., Lipachev E.K., Nevzorova O.A. Programmnyj kompleks avtomaticheskogo preprocessinga cifrovyyh kollekciy matematicheskikh dokumentov // Svidetel'stvo o registracii programmy dlya EVM RU 2019610241, 09.01.2019. Zayavka № 2018664769 ot 18.12.2018.
- [16] Elizarov A.M., Lipachev E.K., Khajdarov Sh.M. Programmnyj kompleks vydeleniya metadannyh iz kollekciy fiziko-matematicheskikh dokumentov, predstavlenyyh v formate OpenXML // Svidetel'stvo o registracii programmy dlya EVM RU 2016618802, 08.08.2016. Zayavka № 2016616511 ot 21.06.2016.
- [17] Elizarov A.M., Lipachev E.K. Semanticheskie metody i instrumenty elektronnoj matematicheskoy biblioteki Lobachevskii-DML // Nauchnyj servis v seti Internet: trudy XIX Vserossijskoj nauchnoj konferencii (18–23 sentyabrya 2017 g., g. Novorossiysk). M.: IPM im. M.V. Keldysha, 2017. S. 130–136. URL: <http://keldysh.ru/abrau/2017/73.pdf>. doi:10.20948/abrau-2017-73. (data obrashcheniya: 21.03.2020).
- [18] Gerasimov A.N., Elizarov A.M., Lipachev E.K., Khajdarov S.M. Automated methods of metadata extraction from scientific publications for bibliographic databases // Informacionnoe obshchestvo: obrazovanie, nauka, kul'tura i tekhnologii budushchego sbornik nauchnyh statej. Trudy XIX Mezhdunarodnoj ob"edinennoj nauchnoj konferencii "Internet i sovremennoe obshchestvo". Sankt-Peterburg, 2016. S. 41–48.
- [19] Elizarov A.M., Zajceva N.V., Lipachev E.K., Khajdarov Sh.M. Programma avtomatizirovannogo formirovaniya metadannyh dokumentov cifrovoy matematicheskoy biblioteki Lobachevskii DML // Svidetel'stvo o registracii programmy dlya EVM RU 2019611328, 24.01.2019. Zayavka № 2019610406 ot 15.01.2019.
- [20] Elizarov A.M., Lipachev E.K., Khajdarov Sh.M. Programma avtomatizirovannogo formirovaniya metadannyh v formate Rossijskogo Indeksa Nauchnogo Citirovaniya dlya statej zhurnala "Elektronnye biblioteki" // Svidetel'stvo o registracii programmy dlya EVM RU 2018612458, 16.02.2018. Zayavka № 2017663206 ot 19.12.2017.
- [21] 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 (23–28 sentyabrya 2019 g., g. Novorossiysk). M.: IPM im. M.V. Keldysha, 2019. S. 234–244. URL: <http://keldysh.ru/abrau/2019/theses/28.pdf>. doi:10.20948/abrau-2019-28 (data obrashcheniya: 21.03.2020).
- [22] Gafurova P.O., Elizarov A.M., Lipachev E.K., Khammatova D.M. Metadata Normalization Methods in the Digital Mathematical Library // CEUR Workshop Proceedings. 2020. Vol. 2543. P. 136–148.