

Литвинова Н. Н.

**Методические
рекомендации
для разработчиков
по организации работы
открытых репозиториев:
корректное создание
метаданных**



NEICON
ЭЛЕКТРОННАЯ ИНФОРМАЦИЯ



НАЦИОНАЛЬНЫЙ
АГРЕГАТОР
ОТКРЫТЫХ
РЕПОЗИТОРИЕВ

Москва
2020

УДК 025.135:004.65(083.132)

ББК 78.5+32.82

Л64

Л 64 Литвинова Н.Н. Методические рекомендации для разработчиков по организации работы открытых репозиторий: корректное создание метаданных. — М.: Ваше цифровое издательство, 2020. — 20 с.

Методические рекомендации подготовлены для использования в организациях, которые поддерживают собственные репозитории или только приступают к их разработке. Основная цель — рекомендовать единые правила формирования метаданных с целью передачи их в Национальный агрегатор открытых репозиторий (НОРА) и последующей интеграции с международными агрегаторами репозиторий. Рекомендации могут использоваться также в организациях, не передающих данные в НОРА, но заинтересованных в прямой передаче своего контента в международные агрегаторы BASE, OpenAIRE и CORE. Рассмотрена структура минимального набора полей метаданных и их развитие в сторону большей детализации в контексте возрастания роли открытого доступа к результатам научной деятельности.

Составление настоящих методических рекомендаций проводилось в рамках проекта «Открытая наука России», поддержанного грантом Президента Российской Федерации на развитие гражданского общества № 19-1-014574

УДК 025.135:004.65(083.132)

ББК 78.5+32.82

Это произведение доступно по лицензии Creative Commons
С указанием авторства 4.0 Всемирная



ISBN 978-5-6040408-9-8

© НП НЭИКОН, 2020

© Н. Н. Литвинова, текст

© Ваше цифровое
издательство, макет

Содержание

1. Введение	4
1.1 Термины и определения.....	4
1.2 Проблемная ситуация.....	4
2. Обоснование предлагаемых решений по представлению метаданных и перечень элементов метаданных	6
3. Форматы метаданных международных агрегаторов	17
3.1 BASE	17
3.2 OpenAIRE	17
3.3 CORE	18
Приложение 1	20

1. Введение

Цель разработки настоящей методики — рекомендовать единые правила описания ресурсов, загружаемых в репозитории организаций-участников Национального агрегатора открытых репозиторий (НОРА), для эффективного сбора метаданных в НОРА и международные агрегаторы репозиторий.

1.1 Термины и определения

Ресурс — в данном контексте — объект хранения в репозитории, включающий метаданные и (опционально) полный контент: полный текст документа, наборы данных, программный код, изображение, видео и подобные.

Метаданные — данные, описывающие содержание и структуру ресурса.

Элемент метаданных — дискретная единица метаданных.

Дублинское ядро — стандартизованный набор элементов метаданных для описания ресурсов, разработанный в рамках инициативы метаданных Дублинского ядра (Dublin Core Metadata Initiative; DCMI).

Схема описания метаданных (также формат метаданных) — логический план, показывающий отношения между элементами метаданных, обычно посредством установления правил использования и управления метаданными.

Протокол сбора метаданных — набор правил, определяющих механизм сбора записей, содержащих метаданные; в рамках инициативы открытых архивов (Open Archive Initiative, OAI) применяется протокол OAI-PMH.

1.2 Проблемная ситуация

В настоящее время каждый репозиторий-участник применяет свои правила и процедуры подготовки описаний ресурсов и соответствующих записей, ориентированные на конкретные цели, сформулированные внутри организации-держателя репозитория. Передача записей во внешние системы (интеграция с внешними системами), такие как агрегаторы репозиторий национально-го и международного уровней и поисковые системы Интернета, требует соблюдения определенных правил для того, чтобы запи-

си репозиториев-участников были представлены в них единообразно и корректно с учетом особенностей функционирования внешних систем.

Настоящие методические рекомендации подготовлены с учетом требования интеграции репозиториев-участников с НОРА с последующей передачей записей из НОРА в международные агрегаторы. Соблюдая эти рекомендации, репозитории-участники смогут сформировать записи, содержащие метаданные, для сбора метаданных внешними системами таким образом, чтобы они были корректно интерпретированы этими системами. При этом держателям репозиториев не нужно кардинально менять сложившуюся практику описания ресурсов и представления их в репозитории организации; достаточно применять данные рекомендации при вводе новых записей и обеспечить преобразование данных для формирования записей, содержащих наборы метаданных, для сборщика метаданных по протоколу OAI-PMH.

Записи репозиториев-участников собираются для загрузки в НОРА; в свою очередь, НОРА предоставляет все записи для сборщиков метаданных международных агрегаторов. В настоящее время осуществляется сбор метаданных из НОРА в агрегаторы BASE и OpenAIRE.

BASE (Bielefeld Academic Search Engine) — глобальный агрегатор научной информации, включающий более 150 миллионов записей, поставляемых от более чем 7000 участников. Контент BASE индексируется глобальным сервисом дискавери EBSCO Discovery Service, который используется в сотнях зарубежных библиотек и десятках крупнейших библиотек России. BASE интегрирован с системой ORCID. Несколько десятков российских репозиториев, в том числе участники НОРА, зарегистрированы в BASE индивидуально и предоставляют свои записи для сборщика BASE по протоколу OAI-PMH.

OpenAIRE — агрегатор репозиториев, журналов и архивов, поддерживающих политику открытого доступа; начавшись как проект Европейского Союза, он вскоре приобрел международный характер и сейчас включает записи контент-провайдеров со всего мира. Содержит более 30 миллионов записей; в отличие от BASE, OpenAIRE объединяет записи на один объект, поступив-

шие от разных контент-провайдеров, поэтому объем записей этих двух агрегаторов сопоставим. Российские репозитории организаций представлены в OpenAIRE меньше, чем в BASE.

Команда HOPA обеспечивает также интеграцию репозиторий-участников с международным агрегатором CORE, регистрируя в нем репозитории. Сбор записей в CORE осуществляется непосредственно из этих репозиторий. HOPA не может быть источником сбора записей в CORE по правилам этого агрегатора, поскольку содержит только ссылки на полный контент (полные тексты и подобные), но не сам контент. Контент CORE индексируется глобальными сервисами дискавери Summon и Primo.

Репозитории-участники, самостоятельно регистрирующиеся в BASE и OpenAIRE, продолжают оставаться источниками поставки метаданных в эти агрегаторы и могут поставлять метаданные в других форматах, более детальных, чем предлагаемые для поставки в HOPA.

2. Обоснование предлагаемых решений по представлению метаданных и перечень элементов метаданных

Метаданные репозиторий, предоставляемые для сборщиков метаданных по протоколу OAI-PMH с учетом требований интеграции, должны генерироваться в тех форматах метаданных, которые воспринимают внешние системы. В настоящее время единым форматом метаданных для HOPA, BASE и OpenAIRE является классический набор из 15 элементов Dublin Core (DC) element set, регламентируемый стандартом ISO 15836-1:2017 “The Dublin Core Metadata Element Set — Part 1: Core elements” (префикс oai_dc). Этот набор позволяет описать объекты на довольно общем уровне, не позволяющем детализировать важные дополнительные аспекты, например структурированно описать название, том и номер журнала, в котором опубликована версия статьи, загружаемой в репозиторий, раскрыть роли лиц, подготовивших документ (редактор, переводчик и др.). Эти ограничения преодолеваются использованием других форматов метаданных, например DC с квалификаторами, специального формата для описания диссертаций и других. Международные агрегаторы разрабатывают специальные форматы метаданных для более детального описания объектов (кратко об этих форматах — в разделе 3). Наиболее продвинутый из них — формат OpenAIRE версии 4. Однако эти форматы специфичны для каждого агрегатора и не могут слу-

жить цели поставки метаданных в несколько агрегаторов одновременно.

Осознавая ограничения базового набора элементов DC, агрегаторы дают некоторые рекомендации для представления содержательно важных элементов описания внутри принятых 15 элементов DC с тем, чтобы впоследствии выделить их в отдельные поля для хранения в базе и обеспечения поиска по ним. Например, “Golden rules for repository managers” BASE (https://www.base-search.net/about/en/faq_oai.php) рекомендует записывать ORCID автора в круглых скобках после его фамилии и инициалов. Есть и другие рекомендации. Мы учитываем эти рекомендации и предлагаем репозиториям-участникам НОРА в нескольких случаях использовать дополнительные поля невалифицированного DC с определенными правилами заполнения. Это позволит, с одной стороны, корректно загрузить записи репозитория-участников в НОРА с нужной степенью детализации, с другой стороны, преобразовать записи, поступившие в НОРА, для сбора метаданных международными агрегаторами с учетом их рекомендаций.

15 элементов DC для описания объектов максимально обобщены. Благодаря этому они могут использоваться для описания разных типов объектов на общем уровне. С другой стороны, семантика большинства полей размыта и оставляет простор разработчикам репозитория для интерпретации правил заполнения полей. Например, представляет трудности содержательное разграничение полей “source” и “relation”, “creator” и “contributor” и других. При поставке метаданных в репозиторий-агрегатор это приводит к противоречиям в представлении записей. Мы предлагаем настоящие рекомендации по заполнению полей для репозитория-участников, чтобы избежать противоречий и максимально унифицировать состав и наполнение полей.

Мы различаем **объекты репозитория** (ресурсы) и связанные с ними объекты. Это различие особенно важно для статей, книг, материалов конференций и других документов, для которых имеются опубликованные версии (version of record), которые мы называем **публикациями**.

В рекомендациях о заполнении полей значениями из фиксированных списков допустимых значений (словарей) мы ориентиру-

емся на словари, рекомендуемые COAR — Coalition of Open Access Repositories. Эти словари предлагают использовать OpenAIRE и CORE; BASE распознает термины этих словарей в большинстве случаев.

В приведенной ниже таблице приводится перечень элементов описания и соответствующих полей DC с пояснениями по наполнению полей и примерами. При этом имеется в виду наполнение полей для представления метаданных по протоколу OAI-PMH в формате oai_dc. Репозитории-участники могут генерировать этот формат из внутренних форматов своих репозиторий, которые в ряде случаев имеют более детализированную структуру описания объектов и соответствующих элементов метаданных.

Сначала перечислены шесть обязательных полей (отмечены знаком «плюс» в столбце «Обязательность»), затем последовательность полей соответствует принятому порядку элементов DC. Названия дополнительных полей помечены комментарием «дополнительное поле». Все поля могут повторяться.

Таблица 1. Перечень элементов описания ресурсов и соответствующих полей метаданных

Название элемента метаданных	Метка поля DC	Обязательность	Пояснения по наполнению поля	Примеры	Комментарий
Идентификатор объекта	dc:identifier	+	В этом поле должны быть идентификаторы, однозначно идентифицирующие объект репозитория , прежде всего, адрес URL; другие возможные идентификаторы — DOI объекта репозитория (то есть такой DOI, который разрешается в запись в репозитории, а не в опубликованную версию) или внутренний идентификатор в данном репозитории	<p><dc:identifier>https://openrepository.ru/article?id=632583</dc:identifier></p> <p><dc:identifier>https://www.repository.cam.ac.uk/handle/1810/284491</dc:identifier></p> <p><dc:identifier>https://dx.doi.org/10.17863/CAM.31867</dc:identifier></p>	<p>Это URI записи в репозитории Apollo (Cambridge University)</p> <p>Это DOI ресурса, размещенного в репозитории Apollo (Cambridge University), который разрешается в запись в этом репозитории</p>
Заглавие	dc:title	+	Заглавие документа, название изображения, программы, другого ресурса. Повторяемость поля позволяет ввести альтернативные заглавия, например на другом языке	<p><dc:title>The repository, the researcher, and the REF: "It's just compliance, compliance, compliance"</dc:title></p>	

Название элемента метаданных	Метка поля DC	Обязательность	Пояснения по наполнению поля	Примеры	Комментарий
Автор, создатель	dc:creator	+	<p>Индивидуальный или коллективный автор, создатель. При наличии ORCID автора его можно:</p> <p>а) записать после имени автора в это поле в виде ссылки, разрешающейся в ORCID автора, заключенной в квадратные скобки;</p> <p>б) если такой вариант неудобен для держателя репозитория, то это поле может содержать только имена авторов; тогда ORCID следует включить в дополнительное поле dc:creatorORCID (см. описание ниже)</p>	<pre><dc:creator>Иванов П.И.</dc:creator></pre> <pre><dc:creator>Robinson, Neil[https://orcid.org/0000-0002-0893-2190]</dc:creator></pre>	<p>Запись, содержащая ORCID, позволит репозиториям-агрегаторам выделить его по формальному признаку и занести в отдельное поисковое поле</p>
ORCID	dc:creatorORCID (дополнительное поле)	-	<p>ORCID автора, если он не был указан в поле «Автор». ORCID записывается после имени автора в квадратных скобках в виде ссылки, разрешающейся в ORCID автора</p>	<pre><dc:creator>Robinson, Neil[https://orcid.org/0000-0002-0893-2190]</dc:creator></pre>	<p>Запись, содержащая ORCID, позволит репозиториям-агрегаторам выделить его по формальному признаку и занести в отдельное поисковое поле</p>

Название элемента метаданных	Метка поля DC	Обязательность	Пояснения по наполнению поля	Примеры	Комментарий
Тип ресурса	dc:type	+	<p>Вид, жанр ресурса. Обычно выбирается из списка типов, которые предлагаются при вводе описаний в конкретной системе. OpenAIRE, версия 4, а также CORE рекомендуют использовать словарь COAR "Resource Type Vocabulary", версия 2, 2019 год. Англоязычные термины этого словаря в их иерархии приведены в Приложении 1. Полный текст многоязычного словаря — по адресу http://purl.org/coar_resource_type/c_3e5a</p>	<pre><dc:type>journal article</dc:type> <dc:type>report</dc:type> <dc:type>dataset</dc:type></pre>	<p>В международных агрегаторах тип ресурса определяет его отнесение к крупной категории, в пределах которой ведется поиск или которая служит фильтром результатов поиска. Особенно важно это для OpenAIRE, где поиск по умолчанию ведется в категории "Publications", куда попадают ресурсы категории "Text". Использование англоязычных терминов этого словаря первого-второго уровня иерархии обеспечит адекватную интерпретацию типа ресурса и в агрегаторе BASE. Поле dc:type может также использоваться для представления информации о версии публикации, включенной в репозиторий. Версии публикаций обозначаются одним из 8 стандартизованных терминов словаря COAR "Version Type Vocabulary" (http://vocabularies.coar-repositories.org/documentation/version_types).</p>

Название элемента метаданных	Метка поля DC	Обязательность	Пояснения по наполнению поля	Примеры	Комментарий
Дата	dc:date	+	Дата создания ресурса в формате, предусмотренном стандартом ISO 8601	<p><dc:date>2015-05-27</dc:date></p> <p><dc:date>2012</dc:date></p>	<p>В нескольких случаях в полях можно записать другие даты, имеющие отношение к жизненному циклу ресурса или публикации, версией которой он является. Однако отсутствие квалификаторов не позволяет задать семантику этих дат, то есть отразить события, с которыми они связаны, например дату принятия авторской рукописи в журнал и другие</p>

Название элемента метаданных	Метка поля DC	Обязательность	Пояснения по наполнению поля	Примеры	Комментарий
<p>Правовой статус</p>	<p>dc:rights</p>	<p>+</p>	<p>В этом поле должно быть записано одно из четырех допустимых значений, описывающих права доступа к объекту. Допустимые значения из словаря COAR: open access — означает свободный доступ любых пользователей к метаданным и полным текстам); restricted access — означает наличие ограничений в доступе, например доступ только для сотрудников организации; embargoed access — означает временный доступ к полным текстам после периода эмбарго; metadata only access — означает доступ к метаданным, но не к полному тексту. Если держатели репозитория не хотят использовать англоязычные термины при вводе и выводе информации, можно ввести соответствующие русскоязычные термины в стандартной форме; для представления метаданных в OAI-PMH из НОРА в международные агрегаторы они будут заменены на англоязычные:</p> <p>открытый доступ доступ с эмбарго ограниченный доступ доступ только к метаданным</p> <p>В другие повторы поля можно вводить любую информацию, относящуюся к правам правообладателей, как в виде текста, так и в виде ссылки на текст документа, регламентирующего права</p>	<p>Примеры стандартных формулировок: <dc:rights>restricted access</dc:rights> <dc:rights>open access</dc:rights></p> <p>Примеры других формулировок: <dc:rights>Copyright © Российская государственная библиотека 2020 Все права защищены</dc:rights></p> <p><dc:rights>Creative Commons Attribution 4.0 International</dc:rights></p> <p><dc:rights>http://www.rii-oxh.net/licenses/all-rights-reserved</dc:rights></p>	<p>Обязательность поля обусловлена требованиями международными агрегаторов, которые группируют записи по значениям этого поля и используют эти значения для фильтрации результатов поиска. Отсутствие в записи этого поля с принятыми значениями ведет к тому, что все записям присваивается неопределенный статус. Для пользователей агрегаторов отсутствие статуса открытого доступа подразумевает, что они не смогут обратиться к полному контенту ресурса</p>

Название элемента метаданных	Метка поля DC	Обязательность	Пояснения по наполнению поля	Примеры	Комментарий
Тема	dc:subject	-	Ключевые слова, предметные рубрики, классификационные индексы, раскрывающие тематику ресурса	<pre><dc:subject>синезеленые водоросли</dc:subject> <dc:subject>ecology</dc:subject></pre>	Для того, чтобы тематика ресурса располагалась между другими агрегаторами хотя бы на общепонимаемом уровне, рекомендуется вводить ключевые слова — названия дисциплин на английском языке, например biology, mathematics
Описание	dc:description	-	Описание содержания ресурса. Для текстовых ресурсов типичное наполнение этого поля — аннотация. В описании можно отразить любые аспекты содержания и использования ресурса	<pre><dc:description>Анализ зарубежного и отечественного опыта по развитию университетского научного бизнеса</dc:description></pre>	В повторах поля можно ввести другие аспекты описания ресурса, например ролевые
Издатель	dc:publisher	-	Организация или лицо, обеспечившее публикацию, издатель. Для ресурса, имеющего версию, опубликованную в издании, предпочтительно записывать издателя опубликованной версии. Например, для размещенной в репозитории авторской версии статьи, финальная версия которой опубликована в журнале издательства Sage Publications, записывается название этого издательства	<pre><dc:publisher>Финансы и кредит</dc:publisher> <dc:publisher>Sage Publications</dc:publisher></pre>	В некоторых репозиториях рекомендовано записывать в это поле организацию-держателя репозитория; тем самым выделяется единый подход к включению в метаданные только характеристик объекта репозитория

Название элемента метаданных	Метка поля DC	Обязательность	Пояснения по наполнению поля	Примеры	Комментарий
Совтор, соисполнитель	dc:contributor	-	Имена, названия лиц и организаций, которые внесли определенный вклад в содержательную подготовку ресурса: редакторов, переводчиков, научных руководителей и других	<dc:contributor>Сергеева, Н. Н.</dc:contributor>	
Формат ресурса	dc:format	-	Характеристики ресурса, относящиеся к его размеру (например, in folio для книг), длительности (для звуковых и видеофайлов, цифровой репрезентации). Для последних рекомендуется использовать обозначения форматов файлов согласно Internet Media Types (MIME)	<dc:format>application/pdf</dc:format> <dc:format>application/postscript</dc:format>	Используется для описания полноконтентного ресурса, для которого создана запись в репозитории; для документальной информации — полный текст документа
Источник	dc:source	-	Источник происхождения ресурса, например печатный оригинал, оцифрованная версия которого находится в репозитории. Название целого для его частей, размещенных в репозитории: журнала для журнальных статей, книги для глав из книг и подобных	<dc:source>Quaestio Rossica, 7(2), 383-392</dc:source>	Содержательные рекомендации по заполнению этого поля в различных репозиториях очень противоречивы. Часто здесь записываются только названия или идентификаторы изданий, с которых произво­дилась оцифровка, а название целого (журнала, книги) для его частей рекомендуют записывать в dc:relation. Однако мы рекомендуем другой вариант, поскольку ориентируемся на интеграцию с международными агрегаторами: в частности, BASE требует включать данные о журнале, в котором опубликована статья, именно в Source

Название элемента метаданных	Метка поля DC	Обязательность	Пояснения по наполнению поля	Примеры	Комментарий
Язык	dc:language	-	Коды языков согласно стандарту ISO 639-3. Возможно также использование кодов по стандарту ISO 639-1 и ISO 639-2	<pre><dc:language>rus</dc:language> <dc:language>eng</dc:language></pre>	Хотя поле не является обязательным, его заполнение крайне желательно, чтобы обеспечить фильтрацию результатов поиска по языку в международных агрегаторах. Русскоязычным пользователям удобно ограничивать поиск материалами на русском языке
Отношение	dc:relation	-	Названия или идентификаторы ресурсов, с которыми связан объект описания: переводы, различные версии. Используется также для ссылок из списка использованной литературы (citations)	<pre><dc:relation>Cranfield Library report</dc:relation> <dc:relation>Zemtsov V. Moscow under Napoleon: social experience of cross- cultural dialogue. Part 2 / V. Zemtsov // Quaestio Rossica. — 2016. — Iss. 4. — P. 223-242. </dc:relation></pre>	В примере имеется в виду, что ресурс явился источником подготовки библиотечного отчета. Элемент из списка использованной литературы
DOI связанного объекта	dc:doi (дополнительное поле)	-	В это поле записываются DOI связанных объектов: опубликованной версии статьи, записи в другом репозитории, перевода	<pre><dc:doi>https:// dx.doi.org/10.1002/ srhc.201800513</dc:doi></pre>	Это DOI опубликованной статьи, авторская версия которой, прилегающая к публикации, размещена в репозитории ArXiv (см. пример заполнения поля dc:identifier)
Охват	dc:coverage	-	Временной период, географическая сфера действия	<pre><dc:coverage>2010-2020</dc:coverage> <dc:coverage>Дин- стия Мин</dc:coverage></pre>	Семантически близко элементу «Тема»

3. Форматы метаданных международных агрегаторов

3.1 BASE (<https://www.base-search.net/>)

BASE собирает данные по протоколу OAI-PMH только в формате простого (неквалифицированного) Dublin Core. Правила подготовки метаданных для передачи в BASE кратко описаны в документе “Golden Rules for Repository Managers” (https://www.base-search.net/about/en/faq_oai.php). В нем изложены некоторые приемы формализованного представления информации, для которой нет отдельных полей в структуре DC. В частности, предлагается включение ORCID авторов в поле dc:creator. Отмечается, что при обработке метаданных в них программно отыскиваются формализованно оформленные DOI и ORCID независимо от того, в каких элементах метаданных они присутствуют.

3.2 OpenAIRE (<https://explore.openaire.eu/>)

Рекомендации по подготовке метаданных для предоставления их по протоколу OAI-PMH изложены в нескольких версиях руководств (Guidelines). Действующая версия, принятая в 2013 г., — третья. Ей на смену идет опубликованная в ноябре 2018 г. четвертая версия.

Текст третьей версии руководства для менеджеров документальных репозиториях (“Guidelines for Literature Repository Managers”) доступен по адресу: https://guidelines.openaire.eu/en/latest/literature/index_guidelines-lit_v3.html, адрес для загрузки документа в pdf — [doi:10.5281/zenodo.1487968](https://doi.org/10.5281/zenodo.1487968)

Основные положения руководства следующие.

1. Предоставлять записи для сбора их по протоколу OAI-PMH рекомендуется в специальной коллекции (set) openaire, префикс метаданных — oai_dc; как показывает практика, при отсутствии этой коллекции данные собираются из коллекций простого DC с пользовательскими названиями, префикс метаданных oai_dc.
2. Применяется стандартный формат Dublin Core; для структурированного представления некоторых важных данных (например, данных о финансирующих организациях) предлагается использовать значения стандартных полей DC, оформленных по специальному синтаксису, который позволяет разобрать эти значения на элементы для хранения

и поиска. Например, для ввода данных о проектах рекомендуется использовать в поле `dc:relation` значения вида `info:eu-repo/grantAgreement/Funder/FundingProgram/ProjectID/`, а для ввода данных о версии статьи — значения из 5 фиксированных терминов в поле `dc:type`, таких как `info:eu-repo/semantics/acceptedVersion` и других.

Текст четвертой версии руководства для менеджеров документальных репозиториях (“Guidelines for Literature Repository Managers”) доступен по адресу: <https://openaire-guidelines-for-literature-repository-managers.readthedocs.io/en/v4.0.0>, адрес для загрузки документа в pdf — [doi:10.5281/zenodo.1299203](https://doi.org/10.5281/zenodo.1299203)

В четвертой версии сделан важный шаг в направлении усложнения структуры метаданных с целью адекватного представления различных данных, прежде вводимых недифференцированно в поля DC широкой семантики. Для этого применяется три пространства имен (namespaces): DC, DataCite и OpenAIRE. Так, отдельные поля предусмотрены для ввода данных о проектах (`oaire:fundingReference`), альтернативных идентификаторах (`datacite:alternateIdentifier`) и других. Для представления контролируемых значений полей теперь применяются словари COAR (COAR Controlled Vocabularies).

Очевидно, что новая структура метаданных позволяет гораздо более детально раскрыть контент ресурсов репозиториях, которые приобретают все большее значение в связи с развитием открытого доступа и повышением роли репозиториях как хранилищ авторитетной научной информации в свободном доступе. Поэтому некоторые репозитории уже внедрили рекомендации четвертой версии, например португальский агрегатор научных репозиториях RCAAP (<http://projeto.rcaap.pt/index.php/lang-en>).

3.3 CORE (<https://core.ac.uk/>)

CORE собирает записи репозиториях-участников по протоколу OAI-PMH с адресов, зарегистрированных в CORE. Сбор метаданных из NOPA не осуществляется.

Структура метаданных репозитория описана в документе по адресу: <https://core.ac.uk/ref-audit/#recommendations>. Эти рекомендации в некоторой степени ориентированы на репозитории университетов Великобритании, поэтому для представле-

ния метаданных используется формат RIOXX (<http://www.riox.net/>), в котором выделены специальные поля для записи ORCID авторов, DOI опубликованной версии статьи, размещенной в репозитории, и детализированного представления дат (дата принятия рукописи в печать, дата публикации). В этом формате используются как метки DC, так и метки rioxterms для специальных полей. Формат совместим с форматом OpenAIRE версии 3: разработана таблица соответствия полей двух форматов — Crosswalk from RIOXX 2/0 to OpenAIRE 3.0. (https://www.riox.net/mappings/crosswalk_riox_2_0_openaire_3_0/).

Приложение 1.

Термины словаря COAR “Resource Type Vocabulary” в иерархической структуре

- -- interactive resource
 - * -- website
- -- dataset
 - * -- interview
- -- image
 - * -- moving image
 - ◇ -- video
 - * -- still image
- -- other
- -- software
 - * -- research software
- -- workflow
- -- cartographic material
 - * -- map
- -- sound
 - * -- musical composition
- -- text
 - * -- annotation
 - * -- bibliography
 - * -- book
 - ◇ -- book part
 - * -- conference object
 - ◇ -- conference proceedings
 - * -- conference paper
 - * -- conference poster
 - ◇ -- conference paper not in proceedings
 - ◇ -- conference poster not in proceedings

- * -- lecture
- * -- letter
- * -- periodical
 - ◇ -- journal
 - * -- contribution to journal
 - » -- journal article
 - -- data paper
 - -- review article
 - -- research article
 - -- corrigendum
 - -- software paper
 - » -- editorial
 - » -- letter to the editor
 - ◇ -- newspaper
 - * -- newspaper article
 - ◇ -- magazine
 - * -- patent
 - * -- preprint
 - * -- report
 - ◇ -- report part
 - ◇ -- internal report
 - ◇ -- memorandum
 - ◇ -- other type of report
 - ◇ -- policy report
 - ◇ -- project deliverable
 - * -- data management plan
 - ◇ -- report to funding agency
 - ◇ -- research report
 - ◇ -- technical report
 - * -- research proposal
 - * -- review
 - ◇ -- book review
 - * -- technical documentation
 - * -- working paper

- * -- thesis
 - ◇ -- bachelor thesis
 - ◇ -- doctoral thesis
 - ◇ -- master thesis
- * -- musical notation
- * -- blog post
- * -- manuscript

- -- learning object
- -- clinical trial
- -- clinical study

