

ДОКУМЕНТАЛЬНЫЕ ИСТОЧНИКИ ИНФОРМАЦИИ

УДК 025.355:004.77:001(470+571); 027.2:002.6(470+571)

Н.А. Мазов, В.Н. Гуреев, С.В. Синявина

Актуальность создания полнофункционального электронного архива печатных версий русских научных журналов*

Представлен обзор международных и отечественных инициатив по оцифровке архивных выпусков научных журналов. Проведен анализ доступности цифровых архивов русских журналов на примере области наук о Земле: демонстрируется малая глубина доступных метаданных и полных текстов, хотя работа по оцифровке ведется. Показаны возможные пути к подготовке и отражению метаданных и полных текстов в открытом доступе на материале двух ведущих отечественных журналов по наукам о Земле – «Геология и геофизика» (переводная версия – “Russian Geology and Geophysics”) и «Записки Горного института» (“Journal of Mining Institute”). Особое внимание акцентируется на роли научных библиотек в подготовке архивов.

Ключевые слова: научный журнал, журнальный архив, сканирование, оцифровка, цифровой архив, научная библиотека, открытый доступ, полные тексты

DOI: 10.36535/0548-0019-2024-10-3

ВВЕДЕНИЕ

Проблема развития русских научных журналов с определенной периодичностью актуализируется в отечественном информационном пространстве. Так, за прошедшее десятилетие дважды была реализована целевая поддержка лучших изданий (ФЦП 14.597.11.0003 в 2014 г. и ФЦП 14.597.11.0035 в 2017 г.), существенно возросло число отечественных журналов, представленных в авторитетной библиографической базе данных *Scopus*. В другой системе – *Web of Science* – кроме общего увеличения числа русских журналов была сформирована и до 2022 г. функционировала отдельная база данных лучших русских журналов *Russian Science Citation Index*. За последние два года значительная часть реализованных отечественными и зарубежными специалистами информационных инициатив в России оказалась под угрозой разрушения или уже разрушена. *Scopus* и *Web of Science* ныне недоступны в России, что существенно затрудняет оценку редакциями позиций своих журналов в этих системах и, таким образом, повышает риски падения их рейтингов [1].

С трудностями происходит переход на собственные, независимые от зарубежных систем, критерии оценки научных журналов.

Таким образом, развитие отечественной научной периодики в 2022 г. снова стало актуальной и обсуждаемой темой. На официальном уровне повышенное внимание, к сожалению, уделялось лишь адаптации к новым условиям системы оценки научных результатов, которая бы не зависела от международных наукометрических баз данных. На уровне профессионального сообщества обсуждались более значимые для развития русской научной периодики вопросы повышения качества русских журналов, в том числе за счет усиления института рецензирования [2], расширения круга авторов и читателей, соответствия международным стандартам публикационной этики. При этом отмечалось, что необходимо продолжать перенимать лучшие международные практики издательского дела [3].

Одной из таких практик, незаслуженно редко упоминаемой при обсуждении проблем русской периодики, является подготовка электронных журнальных архивов. Как правило, этой проблемы нет у изданий, открывшихся уже в эпоху электронной подготовки выпусков. Однако для большинства авторитетных русских журналов это до сих пор нере-

* Исследование выполнено по проектам ГПНТБ СО РАН (122040600059-7) и ИНГТ СО РАН (FWZZ-2022-0028).

шенная задача, имеющая много негативных последствий. В большей степени это касается областей знания, где устаревание информации идет медленно (например, в гуманитарных науках); в теоретических направлениях науки, где результаты могут оказаться востребованными практикой спустя много лет (социальные науки); в областях с детальным описанием фактологических данных (например, результаты геологических экспедиций).

Проблему усугубляет продолжающееся закрытие научных библиотек и прекращение подписок на тематические коллекции при исследовательских организациях, что часто делает недоступным единственный источник получения научных сведений из печатных выпусков журналов. Менее очевидным негативным следствием является возможный плагиат из неоцифрованных коллекций, возможность выявления которого целиком зависит от начитанности рецензента, поскольку все системы его обнаружения работают только с электронными текстами. Главное же отрицательное последствие – это выпадение из научного оборота большого пласта знаний, которым можно было бы пользоваться и который мог бы в значительной мере способствовать развитию науки. Такая ситуация может вести к повторному финансированию ранее проведенных, но забытых и «потерянных» исследований. В наукометрии известен феномен «спящих красавиц» – статей, которые цитируются, т.е. начинают быть востребованными, спустя годы [4]. Но условием для раскрытия потенциала таких значимых работ должна быть их доступность в цифровом виде, желательно в открытом доступе, на развитие которого сейчас направлено множество инициатив. Из последних среди них можно упомянуть Рекомендации ЮНЕСКО по открытой науке (<https://www.unesco.org/en/open-science/about?hub=686>), Глобальную исследовательскую инициативу по открытой науке (<https://hal-lara.archives-ouvertes.fr/hal-04554338>), инициативу «Открытые принципы» компании *Digital Science* (<https://www.digital-science.com/resource/open-principles/>), Барселонскую декларацию об открытой научной информации (<https://barcelona-declaration.org/>) и др.

Отдельно следует отметить, что кроме отсутствия в электронной среде полных текстов российских журналов зачастую отсутствуют и их метаданные, т.е. информация не представлена даже на уровне описаний.

В свою очередь, наличие оцифрованных архивов несет множество преимуществ и издательскому, и научному, и библиотечному сообществам:

- в работе журналов – ранее недоступная информация может привлечь новую читательскую аудиторию и авторов, включая исследователей, занимающихся вопросами истории развития области наук;
- в выявлении некорректных заимствований – пополнение массивов научных данных существенно повышает функциональность систем по обнаружению плагиата, которые бессильны в случае заимствований из отсутствующих в сети текстов;
- в науке – доступные для поиска большие пласты ранее полученных результатов и данных могут дать импульс новым исследованиям с учетом обновившейся приборной базы, новых методов обработки данных. Международные проекты оцифровки

журнальных архивов показали крайне высокий интерес научного сообщества к архивным выпускам журналов [5];

- в адаптации контента к новым моделям чтения – обращение к бумажным копиям ежегодно идет на спад, поскольку поисковые стратегии читателей перешли в цифровую среду еще в 2000-х гг. [6]. Поскольку теперь читателям менее важно, в каком именно журнале находится текст, а поиск проводится по большим массивам информации, стало важным не столько отражать контент на сайте журнала, сколько включать его в большие коллекции с обеспечением поискового функционала [7];

- в проверке данных (фактчекинге) – архивные выпуски журналов формируют фактологическую основу для проведения исследований по истории науки и науковедению;

- в работе научных библиотек – обладание электронными архивами снижает необходимость в постоянном увеличении площадей для хранения бумажных версий.

В настоящей статье подчеркивается необходимость подготовки цифровых архивных версий отечественных журналов – на уровне метаданных и полных текстов (в цифровом html-формате и/или pdf-файле для скачивания и распространения), – что необходимо для повышения престижа российской науки в международном электронном информационном пространстве [8].

При разработке базы данных с нуля или выборе платформы для размещения оцифрованных документов становится важным обеспечение технической возможности формирования xml-выгрузки. Она необходима для интеграции с другими системами для упрощения дальнейшего размещения метаданных статей в различные тематические и библиографические указатели, национальные электронно-библиотечные системы (ЭБС) или даже международные платформы, такие как *Scopus*, *Dimensions*, *Lens*, *DOAJ*.

После краткого обзора международных инициатив по подготовке электронных архивов журналов мы приведем результаты анализа по наличию архивов российских периодических изданий, для примера используя выборку журналов по наукам о Земле. В практической части статьи будут описаны подходы к созданию полных архивных версий двух наиболее авторитетных отечественных журналов по наукам о Земле – «Геология и геофизика» и «Записки Горного института».

КРАТКИЙ ОБЗОР ИНИЦИАТИВ ПО ПОДГОТОВКЕ ЭЛЕКТРОННЫХ ЖУРНАЛЬНЫХ АРХИВОВ

Перевод в электронный вид именно периодических, а не книжных изданий, исторически оказался реализуем и востребован в наибольшей мере, что обусловлено рядом причин:

- для издателей, как уже отмечалось, представление электронного архива расширяет круг читателей и авторов, что ведет к повышению престижности и рейтинговых показателей (например, при размещении архивных выпусков в eLibrary.ru повы-

шается возможность получения статистики из-за возможности количественного мониторинга цитирований, включая их распределение по типам документов, авторам);

- для читателей во многих научных областях знания, особенно в естественно-научных, медицинских и технических, журнальные статьи являются основным источником получения информации, поэтому цифровые копии журнальных статей становятся наиболее востребованными типами документов;

- для ответственных за подготовку электронных архивов в случае журналов существенно упрощается проблема лицензирования: поскольку каждое крупное издательство издает от десятков до сотен научных журналов, сокращается объем юридических вопросов по оформлению авторских прав на оцифрованные архивы, права на каждый из которых принадлежат лишь одному издателю, тогда как с книжными изданиями такие вопросы необходимо решать фактически по каждой книге отдельно;

- для библиотек значительной оказывается экономия места, поскольку журнальная литература занимает от четверти до трети всего пространства научных библиотек.

Потенциальные выгоды, однако, сопровождаются повышенными трудностями, связанными с необходимостью постатейного индексирования журнального контента для обеспечения эффективного поиска, а в случае подготовки архивов для индексирования в библиометрических системах дополнительно требуется наиболее затратная по времени индексация еще и всех пристатейных ссылок.

Из групп организаций, которые могли взять бы на себя ответственность по подготовке архивов, изначально лежал выбор между издательствами и библиотеками. Если говорить об издателях, то их функции связаны прежде всего с управлением процессами производства и распространения знаний, взаимодействием с учеными для обеспечения процессов рецензирования, с анализом новых научных направлений для открытия новых журналов, привлечением новых авторов. Создание и хранение архивов выпадает из области преимущественно коммерческих интересов издателей, отчего они зачастую даже не хранят бумажные копии за неимением места [7, 9]. Примечательно, что опросы среди издателей указывали на отсутствие продуманных политик относительно сканирования журналов [10], а осознав сложность и дороговизну процессов оцифровки текстов и подготовки метаданных, многие издательства после изначального сопротивления передаче лицензий библиотекам на обеспечение постоянного доступа к оцифрованному архивному контенту в итоге пошли на уступки.

Научные библиотеки оказались наиболее инициативными и подготовленными к созданию цифровых журнальных архивов [7], хотя и отмечалась некоторая размытость степени их ответственности [11], а также трудности в стандартизации политики архивирования [10]. Ведущая роль библиотек обусловлена несколькими факторами:

- 1) именно в библиотеках хранятся и доступны для перевода в электронный вид полные комплекты подшивок научных журналов, которых зачастую нет

даже в самих издательствах. Важную роль играет электронная доставка документов между различными библиотеками, которая позволяет формировать архивы необходимых документов, где неизбежные лакуны можно было бы заполнить в оперативном режиме;

- 2) сотрудники библиотек в наибольшей мере обладают необходимыми компетенциями в подготовке метаданных, описывающих журнальный контент для обеспечения поисковых возможностей;

- 3) технологическое преимущество библиотек перед издателями: основой электронных метаописаний архивных журнальных статей во многих случаях становятся автоматизированные библиотечно-информационные системы (АБИС), разработанные для ведения библиотечных каталогов и нацеленные прежде всего на реализацию эффективного поиска;

- 4) библиотеки демонстрируют большую надежность и стабильность, тогда как издатели из-за коммерческой выгоды могут сливаться, менять платформы, закрывать отдельные направления и пр.;

- 5) экономический аспект, поскольку библиотеки сталкиваются с проблемой нехватки помещений для хранения печатных изданий.

На международном уровне наиболее заметной инициативой в создании архивов научных журналов стал крупномасштабный проект *JSTOR*, подразумевавший сканирование сразу нескольких десятков научных журналов на полную глубину архива. Он был реализован в 1990-х гг. при поддержке Фонда Э. Меллона преимущественно силами библиотеки Университета Мичиган (США), а далее перерос в самостоятельный проект, охватывающий все большее число издателей.

Катализатором проекта стал кризис физического пространства в библиотеках для хранения бумажной продукции, что было вызвано бурным ростом числа научных журналов по всему миру [12, 13]. Кроме того, проблемой стало постоянное повышение издательствами цен на журнальную продукцию, что подталкивало библиотеки к действиям по созданию консорциумов и к идее общего доступа к литературе сразу для нескольких организаций [5]. В то же время компьютеры университетов стали объединяться в сети, и возможность доступа к журнальному контенту с любого компьютера воспринималась крайне передовой. А потенциальный совместный доступ к архивам по типу единого многократного использования каталожных записей *OCLC* позволял не только экономить огромные средства на постройку новых помещений, но и освобождать уже имеющиеся площади, сокращать затраты на каталогизацию, хранение и выдачу литературы. Следует отметить, что идея сетевого доступа пришла на смену предыдущим попыткам использования микрофильмов и компакт-дисков, которые требовали дополнительного оборудования и были неудобны читателям.

Этапы подготовки к созданию электронного архива

Поскольку проект *JSTOR* и ряд других, менее масштабных инициатив того же времени, были первым подобным опытом в истории, в них с нуля решались самые разные задачи, подробно перечисленные

в работе [10]. В сгруппированном виде их можно свести к следующему.

1. Подготовка к работе по созданию электронных архивов:

- *лицензирование оцифрованного контента* и переговоры с издательствами для выработки подходов к лицензированию. В рамках этой задачи в числе прочих решались вопросы по временной границе формирования архивов, которая не затрагивала бы последние выпуски журналов, чтобы не подрывать устоявшиеся модели продаж издателями своей продукции. Эта модель, однако, позволяла постоянно пополнять архив относительно свежими выпусками журналов, которые с течением времени выходили из «окна» подписки, что создало основу компромисса интересов между библиотеками и издателями;

- *финансирование работ по оцифровке архивов* и подготовке метаданных, разработка модели распределения доходов от подписки на архивы. Для каждого из проектов важно было выйти на устойчивую модель дальнейшего развития уже без участия грантодателей. Библиотеками, занятыми оцифровкой архивов, велись переговоры с библиотечными консорциумами по вопросам подписки на архивы, притом что стоимость подписки должна была окупить затраты на поддержание бумажных фондов, от которых в случае подписки на электронные архивы предоставлялась возможность отказаться. В проекте *JSTOR*, в частности, затраты на архивирование были распределены приблизительно между 900 библиотеками США;

- *определение тематических приоритетов* для выработки порядка сканирования архивов. Кроме определения наиболее востребованных научных областей, здесь же решались более частные задачи выбора типов сканируемых документов по каждому названию журнала. Например, в научных журналах помимо собственно статей публикуются сведения о редколлегиях, объявления о вакансиях и конференциях, поздравления и некрологи, рекламная продукция, которая может быть неважна с научной точки зрения, но безусловно может быть полезна историкам науки. Известно, что даже сейчас подготавливаемые параллельно бумажные и электронные макеты нередко различаются.

2. Разработка концепции подготовки метаданных. Предварительно исследуются основные этапы развития журнала и выделяется публикационная концепция каждого из этапов для обеспечения унификации в представлении данных:

- *полнота метаданных* – тогда как в одних случаях создаются библиографические описания журнальных статей с минимальным числом полей, фактически не обеспечивающих эффективный поиск, в других организуются полноценные базы данных, где индексируется максимально возможная информация, устанавливаются гиперссылки, в записи добавляется внешняя информация, например, по идентификаторам статей в разных библиографических системах. Индексация мест работы авторов, аннотаций и ключевых слов, сведения о финансировании, разобранные списки литературы дают крайне значимую информацию для эффективного поиска;

- *точность метаданных* – где необходимо принимать решение, следует ли исправлять или каким-то образом пометать ошибки и опечатки, например, в метаданных или списках литературы.

3. Техническое обеспечение подготовки, обработки и последующей поддержки электронных архивов:

- *выбор аппаратного и программного обеспечения (ПО) для сканирования* – как правило, часто используется потоковое сканирование с предварительным «разбором» журналов. В отдельных случаях сканируются лишь страницы с оглавлениями, что позволяет создавать минимальный объем метаданных для поиска, а печатный журнальный фонд в библиотеке перенести в отдаленное и менее дорогое для хранения место [5];

- *выбор формата сохранения данных* – необходимо определять качество сохранения данных, формат и/или форматы, если их несколько для разных частей документов. Здесь важна периодическая проверка форматов хранения данных и физических носителей на соответствие текущим стандартам, поскольку со временем ПО для корректного прочтения тех или иных данных перестает поддерживаться [9]. При том, что ПО неразрывно связано с аппаратным обеспечением, следует учитывать совместимость ПО и операционных систем, разрабатывать эмуляторы для имитации устаревшего ПО, учитывать срок службы физических носителей. Известна история Бюро переписи населения США, которое в 1976 г. безуспешно пыталось получить доступ к исторически значимым файлам переписи населения 1960 г. Файлы хранились на ленточном накопителе *UNIVAC*, который устарел к середине семидесятых годов [7]. Специалисты по информационному делу справедливо отмечают, что человечество по-прежнему может читать шумерские таблички, но сталкивается с существенными трудностями при чтении 3,5-дюймовой дискеты, созданной 30 лет назад;

- *контроль целостности и качества файлов* – требует периодических проверок и наличия специализированного ПО;

- *выбор типа доступа к данным* – как правило, делается выбор между открытым и закрытым хранением данных. К открытому архиву обычно предъявляется больше требований: более строгое лицензирование; поддержка современных интерфейсов, удобных пользователям; отслеживание статистики использования. Хранение без широкого доступа, ограниченное IP-адресами библиотеки, намного дешевле, и при таком подходе реализуется страховая функция по сохранению материалов, а не функция постоянного обслуживания пользователей;

- *выбор локации хранения электронных архивов* – необходимо решить вопрос о централизованном либо распределенном характере хранения архивов. В целом рекомендуется некоторая избыточность, при которой создается несколько копий, хранящихся в разных локациях и за которые несут ответственность разные организации, на случай повреждения какого-либо из архивов. Для организации открытого доступа следует создавать зеркальные сайты в различных ре-

гионах мира, повышающие скорость получения данных. Обратим внимание на имеющий негативные последствия отказ издателей в предоставлении архивов на компакт-дисках в дополнение к сетевой версии. Именно этот путь мог бы гарантировать доступ к архивам при любых обстоятельствах, в том числе таких, с которыми сталкивается Россия в последние два года. Несмотря на юридически оформленные обязательства зарубежных издательств предоставлять вечный доступ к оплаченным коллекциям, закрытой во многих случаях оказывается вся подписка, в том числе оплаченная прежде. Отметим, что подобные опасения по поводу электронного доступа звучали и за рубежом, причем еще в 2000-е гг. Характер гарантий доступа определялся как существующий лишь в теории [10, 11, 14], а библиотеки уже тогда периодически сообщали о потере доступа к оплаченным ранее цифровым коллекциям [15].

Другие проекты по сохранению фонда периодических изданий

Многие из перечисленных выше вопросов остаются актуальными для стран, где создание электронных архивов еще не завершено или не проводилось, поэтому опыт *JSTOR* и других проектов оказывается полезным для изучения. К менее масштабным инициативам, реализуемым как отдельными библиотеками, так и их консорциумами, можно отнести:

- *LOCKSS*, реализуемый Стэнфордским университетом (<https://www.lockss.org/>);
- *MUSE*, направленный на архивирование журналов гуманитарных и общественных дисциплин (<https://muse.jhu.edu/>);
- *Electronic Collections Online*, проводимый в *OCLC* и прекративший существование в 2011 г.;
- *E-Depot* – проект Национальной библиотеки Нидерландов по сохранению архива издательства *Elsevier*; совместный проект библиотеки Йельского университета и *Elsevier* по сохранению цифровых копий журналов издательства [16], а также еще один проект *Elsevier* – *TULIP*, направленный на подготовку и доставку идентичных бумажным цифровых копий журналов [17];
- *DigiZeitschriften*, представляющий собой базу данных немецких научных журналов по гуманитарным наукам (<https://www.digizeitschriften.de/>);
- *Gallica*, результатом которого стала цифровая библиотека Национальной библиотеки Франции (<https://gallica.bnf.fr/accueil/de/content/accueil-de?mode=desktop>).

Своеобразным ответвлением от *JSTOR* стал проект *Portico* (<https://www.portico.org/>), целью которого является не предоставление немедленного доступа онлайн, но гарантированное сохранение оцифрованного контента в случае форс-мажорных обстоятельств. Например, если закрывается издательство, журнал переходит к другому издателю с иной политикой распространения или происходят необратимые сбои на веб-платформе, то необходимыми становятся ранее созданные резервные копии, к которым могут обращаться научные библиотеки. При этом библиотеки продолжают играть важную роль хранителей

информации даже в эпоху изначально электронной подготовки журналов, например, в рамках совместного проекта Фонда Э. Меллона и Совета по библиотекам и информационным ресурсам (*CLIR*) [9].

Отметим, что кроме библиотек в отдельных случаях оцифровку выполняли сами издательства из разных отраслей науки. Из российских примеров можно упомянуть:

- «Успехи физических наук», <https://ufn.ru/ru/articles/> (учредитель ФГБУН Физический институт им. П.Н. Лебедева РАН), где представлен архив с 1918 г. с блоком для цитирования, pdf статей, присвоенными номерами DOI и информацией о статистике использования;
- «Казанский медицинский журнал», <https://kazanmedjournal.ru/kazanmedj/issue/archive> (учредители ФГБОУ ВО «Казанский государственный медицинский университет», ООО «Эко-Вектор»), где представлен архив с 1901 г. с полными метаданными, pdf- и html-версиями текста, присвоенными DOI, информацией о статистике использования и указанием перекрестных ссылок из новых статей на старые.

В то же время многие небольшие издательства, преимущественно научных сообществ, до сих пор не смогли себе позволить этот затратный процесс.

Создание электронных архивов журналов в России развивается медленнее, чем хотелось бы, и реализуется по большей части бессистемно, силами отдельных библиотек, организаций-учредителей или издательств. Из централизованных начинаний можно упомянуть Федеральную целевую программу «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2014–2020 годы», по которой в 2020 г. был объявлен конкурс на создание электронных архивов по семи научным направлениям (<https://tass.ru/obschestvo/8916299>). Среди поддержанных направлений были «Математика, физика, информационные технологии» (отражены на математическом портале <https://www.mathnet.ru/>), «Медицина» (<http://медархив.эко-вектор.пф/>), «Международные отношения и историко-филологические науки» (<https://arxiv.gaugn.ru/>), «Общественные науки» (<https://arxiv.gaugn.ru/>), «Науки о Земле и энергетика» (<https://doc365.ru/>), «Химия, биология и физиология» (<http://physchembio.ru/>), «Сельскохозяйственные науки» (<http://www.agriculture.ru/index>). О результатах некоторых из созданных архивов сообщалось в открытой печати [8, 18], тогда как об иных общедоступной информации либо не было, либо обозначенные по итогам гранта сайты оказывались вне доступа (например, в случае сельскохозяйственных наук, а также наук о Земле), что позволяет усомниться в эффективности последнего проекта. Интересен опыт БЕН РАН [19], в которой был полностью оцифрован, хотя и без метаданных, архив одного из наиболее востребованных научных журналов – «Доклады Академии наук».

Следует обратить внимание на присутствие в России высокопрофессиональных компаний по оцифровке различных архивов, таких как «ЭЛАР» (<https://elar.ru/>), которые активно сотрудничают с библиотеками, когда сканируют каталоги, книги и журнальные коллекции.

ЭЛЕКТРОННЫЕ АРХИВЫ РОССИЙСКИХ ЖУРНАЛОВ ПО НАУКАМ О ЗЕМЛЕ

Чтобы определить состояние подготовки архивных выпусков российских журналов, мы проанализировали область наук о Земле как одну из наиболее развитых дисциплин, в том числе хорошо представленных на международном уровне [20, 21]. В составленный перечень вошли наиболее авторитетные и широко известные в международном информационном пространстве российские журналы, которые удовлетворяли следующим условиям:

- оригинальная или переводная версия входила в одну из 14 тематических категорий общего направления *Geosciences* в *Journal Citation Reports* версии 2022 г.;

- имели переводную англоязычную версию, либо изначально издавались на английском языке.

Обоим условиям из общего числа 908 журналов по наукам о Земле, представленных на платформе *Journal Citation Reports*, удовлетворили 27 из 34 российских журналов (7 журналов не имели англоязычной версии). По базам данных *Web of Science Core Collection*, *Scopus*, РИНЦ и сайтам журналов была определена глубина доступных на текущий момент метаданных, полных текстов и идентификаторов DOI, а также время полужизни цитируемых статей для выявления минимально необходимой глубины востребованного архива. С учетом возраста журнала была рассчитана доля доступного в электронном виде контента в виде метаданных и полных текстов. Собранные авторами сведения доступны в открытом виде по ссылке: <https://doi.org/10.5281/zenodo.11551274>.

Большинство журналов с богатой историей систематически и без лакун фиксируются на уровне метаданных, прежде всего в РИНЦ, приблизительно с 2000-х г., т.е. с момента перехода на электронную подготовку. Глубже онлайн-информация о тех или иных изданиях, преимущественно об их переводных версиях, находится лишь в зарубежных базах данных, как в политематических (*Web of Science*), так и в узконаправленных (*GeoRef*, *SciFinder*). За рубежом достаточно полно представлены и некоторые оригинальные русскоязычные версии: например, «Доклады Академии наук» фактически в полном объеме и за весь период – в *Web of Science*. Поскольку зарубежные базы данных приводят перевод метаданных, это делает обратную трансформацию невозможной, но за неимением иных сведений в электронном виде даже такая информация может оказаться полезной.

Сайты многих журналов, к сожалению, малоинформативны, малофункциональны и не содержат архивов. Особенно удручающая ситуация с наиболее авторитетными изданиями Российской академии наук, которые в большинстве случаев вообще не имеют собственных сайтов, а платформы с контентом меняются довольно часто. Так, описания и (при наличии доступа) полные тексты журналов с 2013 по 2018 гг. находятся на платформе ЭБС «Наука» (<https://www.libnauka.ru/>), выпуски за 2019 г. – на платформе «Эко-Вектор» (<https://journals.eco-vector.com/>), выпус-

ки за несколько последующих лет – на сайте издательства «Российские научные академические журналы» (<https://sciencejournals.ru/>), а за последние два года – на Национальной платформе периодических научных изданий РЦНИ (<https://journals.rcsi.science/>), что делает архив распределенным и крайне неудобным в использовании. Среди русскоязычных версий нашей выборки выделяются три журнала: «Геология и геофизика» (Сибирское отделение РАН), «Записки Горного института» (Санкт-Петербургский горный университет императрицы Екатерины II) и «Литология и полезные ископаемые» (РАН), причем работа по созданию архивов во всех случаях проведена силами учредителей журналов.

Ситуация с переводными (или изначально англоязычными) версиями выглядит лучше. Например, для переводных версий журналов «Записки Горного института», «Метеорология и гидрология», «Основания, фундаменты и механика грунтов», «Солнечно-земная физика», «Физико-технические проблемы разработки полезных ископаемых», а также изначально англоязычного журнала “*Russian Journal of Earth Sciences*” электронные архивы с метаописаниями и полными текстами доступны с первых выпусков.

Важно отметить, что существенно лучшая представленность и доступность переводных версий большинства журналов РАН обеспечивается издательством *Pleiades Publishing*, располагающем журналами на платформе *SpringerLink*. С точки зрения подписчиков в лице библиотек и читателей в лице ученых это безусловно удобно, поскольку доступ осуществляется на единой платформе и с соблюдением всех современных стандартов, включая html- и pdf-версии статей, функционирующие DOI, оперативность размещения на платформе и далее – в индексирующих библиографических системах, представляющих широкие поисковые возможности. Парадоксальным, но для многих привычным является полугодовое опережение издания переводных версий статей по сравнению с оригинальными. Именно техническое превосходство, постоянство и доступность зарубежных платформ, на которых расположены переводные версии журналов, привели к ситуации, при которой цитируемая полужизнь журнальных статей, т.е. усредненная глубина наиболее востребованного архива, в 1,5–2 раза выше у переводных версий по сравнению с оригинальными (<https://doi.org/10.5281/zenodo.11551274>).

Таким образом, оригинальные версии в значительной мере уступают переводным по наличию html-версий, числу обращений, просмотров, загрузок, а во многих случаях и цитирований, поскольку международные библиографические системы предпочитают индексировать именно англоязычные варианты. На это накладывается нерешенный конфликт между РАН и *Pleiades Publishing* по поводу авторских прав на названия и в целом на сопровождение и поддержку переводных версий, который можно оценивать по-разному, но где в целом прослеживается недальновидность предшествующих действий РАН при передаче прав.

Большинству журналов еще предстоит работа как по созданию метаданных, так и по оцифровке полных текстов. Завершена работа лишь у ничтожно малого числа изданий. Крайне слабое внимание журналы уделяют присвоению идентификаторов DOI, причем не только ныне недоступным в электронном виде выпускам, но и размещенным в информационных системах. Многие журналы стали присваивать DOI лишь с начала 2010-х гг., тогда как электронные архивы появились у них на десятилетие раньше. Необходимо сделать оговорку по поводу текущей невозможности получения DOI отдельными журналами, издаваемыми в российских организациях, попавших под санкции. Это приводит к полному выпадению сведений о новых выпусках из многих систем, основанных на данных Crossref, и, как следствие, к потере части читателей. Стоит отметить, что число таких изданий невелико, и выразить надежду, что в обозримом будущем найдутся решения, способные разделить публикационные стандарты и целостность распространения научной периодики от геополитики.

Стратегии подготовки журнальных архивов в России, как показал проведенный нами анализ, также различны. Если одни начинают работу с подготовки метаданных, то другие первостепенное внимание уделяют оцифровке архивов для обеспечения доступа к полным текстам. Выставленные же в открытый доступ полные тексты оказываются либо распознанными, либо в виде картинок, могут быть разбиты построчно или представлены в виде отдельных номеров. Однако, как уже отмечалось, наличие оцифрованных версий журналов без подготовки метаданных плохо решает задачу обеспечения поисковых возможностей по архивным статьям. Единственное, что получает пользователь, – это возможность просмотра страниц с оглавлениями журналов со своего рабочего места; сама же необходимость рутинного визуального просмотра никуда не исчезает.

В качестве примера можно привести журнал «Тихоокеанская геология», где цифровой архив текстов (выпусков целиком, без разделения на статьи) подготовлен, но не сопровождается метаданными, что исключает поисковые возможности. Другой пример – журнал «Основания, фундаменты и механика грунтов», у которого наличествует почти полный архив метаданных, но проиндексированы лишь два поля – авторы и названия статей. Подобные полумеры не делают журналы более доступными и включенными в научный оборот, чем прежде, хотя являются важным этапом на пути к созданию полнофункциональных полнотекстовых архивов с достаточными для поиска библиографическими описаниями.

Таким образом, по разным журналам, хотя бы в какой-то мере представившим информацию о своем архиве, у пользователей для работы имеются совершенно различные поисковые и функциональные возможности. Впрочем, такая ситуация характерна и для многих зарубежных журналов. Далее представлены подходы, использованные нами при подготовке полных метаданных и полных электронных текстов для журналов «Геология и геофизика» и «Записки Горного института».

ТЕХНОЛОГИЧЕСКИЙ КОМПЛЕКС ПОДГОТОВКИ ЭЛЕКТРОННОГО АРХИВА

Журналы «Геология и геофизика» и «Записки Горного института» представляют собой издания с достаточно длительной историей, их объединяют как тематические научные направления, так, соответственно, и одинаковые проблемы в публикационной сфере: обособленность российских журналов от других стран, развитие практик авторского права, появление движения по публикационной этике и открытому доступу, необходимость перевода метаданных на английский язык и вхождения в международные базы данных.

Для расчета приблизительных объемов работы, которая может потребоваться другим журналам, в таблице представлены общие сведения по формированию электронных архивов журналов «Геология и геофизика» и «Записки Горного института».

Пример журнала «Геология и геофизика» / «Russian Geology and Geophysics»

Технологический комплекс подготовки метаданных и полных текстов журнала «Геология и геофизика» и их представления в веб-среде был создан и апробирован в информационно-библиотечном (ныне – информационно-аналитическом) центре Института нефтегазовой геологии и геофизики им. А.А. Трофимука (ИНГГ) СО РАН. Этот журнал является одним из ведущих изданий в области наук о Земле не только в России, но и за рубежом. Ровесник Сибирского отделения Российской академии наук, он начал издаваться в 1960-м г. и недавно отметил свое 65-летие.

Электронный архив русскоязычной версии журнала на сайте издательства Сибирского отделения РАН доступен с 2001 г. Метаописания более ранних русскоязычных выпусков – с 1986 по 2000 гг. – можно найти в двух базах данных «Геология» и «Геофизика» реферативного журнала ВИНТИ РАН, где к стандартному описанию добавлены ключевые слова и тематические рубрики ГРНТИ и ВИНТИ. Тем не менее формат представления библиографической информации в РЖ ВИНТИ РАН характеризуется ограниченным набором полей и полным отсутствием пристатейной библиографии публикаций, что существенно сужает поисковые возможности.

В ИНГГ СО РАН – одном из четырех учредителей журнала «Геология и геофизика» – работа по созданию базы данных метаописаний публикаций основной русскоязычной версии журнала, пристатейной литературы и подготовке pdf-версий публикаций была инициативно начата в 2010 г. и в основной части завершена в 2012 г., а после перерыва вышла на финальную стадию в 2022 г. Прежде всего эта работа стала возможной из-за накопленного коллективом информационно-библиотечного центра опыта подготовки и представления метаданных в виде базы данных, содержащей труды сотрудников ИНГГ СО РАН, а также длительного опыта эксплуатации баз данных ВИНТИ РАН.

Общие сведения о журналах

Критерии	Название журнала	
	Геология и геофизика / Russian Geology and Geophysics	Записки Горного института / Zapiski Gornogo Instituta
Год создания, возраст	1960, 65 лет (оригинальная версия) 1974, 50 лет (переводная версия)	1907, 118 лет
Количество изданных томов	65 (12 выпусков в томе)	269 (в некоторые года от 2 до 5 выпусков в томе)
Сканируемый архив	1960–2004	1907–2011
Оцифровано страниц	более 100 тыс.	более 50 тыс.
Оцифровано статей	более 9500	более 8200
Типы статей, размещенных в архиве	Научные статьи	Все типы публикаций
Архивация и глубина архива	Глубина архива: до 35 лет: – Научная электронная библиотека eLibrary.ru – ВИНТИ РАН (базы данных «Геофизика» и «Геология») – GeoScienceWorld – Crossref	Глубина архива: 5–10 лет: – Научная электронная библиотека eLibrary.ru – КиберЛенинка – Издательство «ЛАНЬ» – IPR SMART – «Айбукс» – Znanium – УБД «ИВИС» – DOAJ Seal – Crossref, DataCite
Данные, размещенные на странице статьи	Метаданные	– Метаданные – Ссылка для цитирования – Список похожих статей – При наличии перевода размещены иностранские версии на английской странице сайта
Представление полного текста	pdf-версии статей с 1960 г.	pdf-версии статей с 1907 г. html-версии статей с 2021 г.

Основная цель создания такой базы данных – представление в электронном, доступном для широкого использования виде научных знаний из журнала, а задача – формирование полнотекстового архива с полным метаописанием публикаций за период 1960–2004 гг., поскольку затем редакция перешла на подготовку электронных версий журнала в дополнение к печатной. Иллюстративный материал по особенностям сканирования архивных выпусков обоих журналов доступен по ссылке <https://doi.org/10.5281/zenodo.11551274>. Эта работа была разбита на несколько этапов:

1) первым этапом стала подготовка к оцифровке печатного архива, что включало проверку ежегодных журнальных подшивок на полноту и сохранность, поскольку для создания электронных образов требовались хорошо сохранившиеся страницы. В случае обнаружения лакун или непригодных к сканированию экземпляров проводились поиски выпусков журналов в других научных библиотеках;

2) в плане отбора контента было принято решение сканировать только научные статьи; вступительное слово редактора, поздравительные материалы и некрологи были исключены;

3) ошибки и опечатки в основных метаданных (авторы, заглавия, аннотации и списки литературы)

было решено исправлять по мере их обнаружения, однако специально такая задача не ставилась;

4) сканирование высокого разрешения с использованием потокового сканера было проведено усилиями сотрудников ГПНТБ СО РАН с предварительным постраничным «разбором» журналов, предполагавшим «обрубание» переплета. На этом же этапе с использованием прилагаемого к сканеру программного обеспечения было проведено автоматическое распознавание текста;

5) следующим этапом стало глобальное графическое редактирование полученных образов страниц для увеличения контрастности и читаемости текста, что также предполагало «отбеливание» сканированного фона страниц, поскольку журнал длительное время печатался на серой бумаге;

6) далее формировались файлы статей из отдельных страниц. Был использован современный подход, предполагающий начало каждой статьи с новой страницы. С этой целью со страниц, где пересекались окончание и начало двух статей, удалялась соответственно либо завершающая, либо начальная часть. Работы этого этапа были проведены специалистами отдела информационных технологий ИНГГ СО РАН. Физическое хранение данных обеспечено собственным библиотечным сервером, веб-зеркалом на

основном сайте Института (<http://www.ipgg.sbras.ru/ru/cooperation/journals/gg>);

7) наиболее трудоемким и затратным по времени этапом стала работа по подготовке метаданных на основе информации с титульных страниц соответствующих статей. Отметим, что важность и сложность этого этапа, требующая привлечения квалифицированных библиографов, отмечалась и специалистами издательства *Elsevier* в их проекте *TULIP* по переводу бумажных версий в электронные [17]. Этап предполагал существенный объем визуальной проверки точности распознанного текста, верификацию специальных символов из названий статей, правки смешения символов латинского и кириллического алфавитов для обеспечения функционального поиска по метаданным. Работа проводилась в среде автоматизированной библиотечно-информационной системы *CDS-ISIS*. Кроме основных стандартных метаданных, включающих авторов, название публикации и выходные данные, в метаописания добавлялись ключевые слова, аннотации, affiliations авторов, а также тематические коды ГРНТИ. Для всего набора метаданных также приводились соответствующие переводные и транслитерированные (в случае авторов и affiliations) сведения, которые в разных выпусках могли располагаться как в самих статьях, так и на отдельных страницах – в конце выпуска, что требовало дополнительной работы по соотнесению переводных метаданных с основными;

8) не менее трудоемким этапом была подготовка пристатейных списков используемой литературы для каждой публикации. Они были сформированы в виде отдельной базы данных, где каждая ссылка представлялась как стандартизованное библиографическое описание. Кроме этого, для каждой цитируемой ссылки указывался тип документа, язык и прочие доступные сведения, обеспечивающие возможность последующего поиска по каждому описанному элементу;

9) завершающим этапом стал ряд финальных глобальных правок базы данных, в том числе чистовая проверка орфографии, целостности сформированных pdf-файлов публикаций и их соответствия метаописаниям, проверка полноты метаописания на основе визуального сличения метаданных с титульными страницами публикаций. Для геологических журналов эта работа особенно важна из-за нередкого наличия в бумажных версиях различных картографических вкладок разного масштаба, которые при потоковом сканировании выпадали.

Таким образом, была создана зарегистрированная в Роспатенте база данных со связанными в ней различными элементами¹, которая включает библиографические описания публикаций, их рефераты, пристатейные списки литературы и полные тексты, т.е. максимально отражает сведения о журнале. Основные поля базы данных являются поисковыми, что позволяет выводить, например, все публикации

определенного автора, все публикации, цитирующие определенный источник, все публикации конкретной организации и многое другое. Учитывая важность представления метаданных и полных текстов журнала в открытом информационном пространстве, электронный архив журнала в текущее время представлен на веб-странице ИНГГ СО РАН (<http://www.ipgg.sbras.ru/ru/science/journals/gg>).

С 1974 г. издается англоязычная версия этого журнала – *“Russian Geology and Geophysics”*, в полной мере представляющая переводы всех русскоязычных публикаций. До 2018 г. переводная версия была размещена на платформе *ScienceDirect* издательства *Elsevier*, а в текущий момент перешла на основную международную платформу по наукам о Земле *GeoScienceWorld* (<https://pubs.geoscienceworld.org/rgg>). К сожалению, печатные выпуски переводной версии до 1990 г. не поставлялись издательством, отвечавшим за перевод, в Россию и сейчас физически недоступны ни в одной из библиотек нашей страны, что затрудняет их перевод в электронный вид. На платформе *GeoScienceWorld*, представляющей журнал *“Russian Geology and Geophysics”*, на текущий момент отражены его выпуски, начиная с 2002 г. Ведется работа по оцифровке и подготовке метаданных за период 1991 – 2001 гг.

Пример журнала «Записки Горного института» / “Journal of Mining Institute”

Технологический комплекс подготовки метаданных и полных текстов журнала «Записки Горного института» был выполнен в двух подразделениях – Главной библиотеке и Издательском доме Санкт-Петербургского горного университета императрицы Екатерины II (учредитель и издатель). Период проведения работ разбивался по времени в зависимости от окончания работ одного этапа, технических возможностей и загруженности сотрудников подразделений.

Журнал «Записки Горного института» – один из старейших ведущих российских и международных рецензируемых изданий по наукам о Земле и техническим наукам [22, 23]. В периоды военных столкновений журнал не издавался (революция, ВОВ и пр.).

Исходя из возраста журнала можно сделать вывод о разнообразии подходов к требованиям к научной периодике за последние 100 лет в плане тематик научных исследований, изменений (изначально отсутствия) требований к периодичности издания, объему тома или статей в томе, структуре статьи, подходе к указанию использованных источников и, конечно, объему и отображению метаданных статей. В связи с этим возникло много трудностей в понимании унификации данных.

По архиву можно увидеть изменение решений руководства Горного университета по перспективам развития журнала, что продиктовано общими подходами к публикации материалов в России и мире:

- 1907–1921 гг. – издание фундаментальных трудов, монографий в виде одной или нескольких глав в составе одного тома. В свою очередь, тома включали в себя несколько номеров;

¹ Мазов Н.А., Гуреев В.Н. IP4G электронный архив метаданных журнала «Геология и геофизика»: Свидетельство о государственной регистрации программы для ЭВМ // Свид-во о прогр. 2020621522; RU; № 2020621328, заявл. 14.08.2020, опубл. 25.08.2020, ИНГГ СО РАН.

- ориентировочно 1922–1991 гг. – тематики статей менялись в зависимости от государственной политики, наблюдались сбои в нумерациях томов (вероятно, утвержденные тома публиковались по мере прохождения цензуры публикуемых материалов в СССР);

- 1999–2014 гг. – многие тома содержали публикации материалов по итогам научных конференций (привлечение новых авторов и распространение журнала);

- 2015–2016 гг. – публикация статей с обязательным соавторством профессорского состава Горного университета (поднятие уровня научных статей);

- 2016–2018 гг. – 1) выход из «домашнего» журнала с привлечением авторов из российских и иностранных организаций и вузов; 2) создание сайта журнала и начало размещения архива; 3) с 219 по 260 том – присвоение номеров DOI (прекращено из-за наложения санкций США на вузы и научно-исследовательские институты РФ, которые специализируются на добыче природных ископаемых, 19 мая 2023 г.) (<https://ofac.treasury.gov/faqs/added/2023-05-19>); 4) с 220 тома полнотекстовый перевод статей на английский язык; 5) сужение тематических направлений с мультидисциплинарного на направления «технические науки» и «науки о Земле» (решение о вступлении в *Scopus*, *Web of Science*, а также распространение в другие тематические базы данных и ЭБС);

- 2021 г. – наст. вр. – изменение пула стран и аффилиаций авторов с европейских на азиатские государства. Как следствие, изменение тематики исследований (переход после введения санкций против РФ);

- 2022 г. – наст. вр. – создание тематических томов с приглашенными редакторами (повышение научного уровня статей из-за конкуренции, привлечение новых авторов и рецензентов);

- 2023 г. – наст. вр. – введена публикация статей в формате *Online First* – полноценная публикация до размещения в печатный том. Преимущество в публикации статей по готовности, а также URL-ссылка останется неизменной после распределения в очередной том.

Для удобства поиска в печатном архиве в журнале были созданы указатели статей [24–27], а также тематические подборки [28, 29].

Организация электронного архива была начата в 2014 г. по инициативе ректора Горного университета, и при разработке сайта журнала единственная цель создания архива была обозначена сразу: все статьи любого типа должны быть размещены в открытом доступе. Но не было поставлено дополнительных целей, основных задач и сроков.

Работа выполнялась несколькими отделами и разбилась на стадии.

I. Главная библиотека:

- оцифровка – сканирование томов на обычном постраничном сканере – 2009 – 2014 гг. (<https://pmi.spmi.ru/pmi/announcement/view/44>);

- перевод метаданных на английский язык – нанятый переводчик работал с печатным архивом и выполнял перевод метаданных томов 144 (1999 г.) – 182 (2009 г.) ориентировочно в период 2012–2014 гг.

II. Издательский дом:

- распознавание файлов – при распознавании были различные последствия сканирования, которые не предполагали возможности дальнейшей цифровой расшифровки. Журналы сканировались в исходном виде без обрезания корешка (при наличии единственного экземпляра), из-за чего в объемных томах отсканированный текст оказывался усеченным или шел «волной», часть страниц были высветленные, некоторые, наоборот, затемненные или мутные – многие неполадки возникали из-за качества печати или возраста бумажного издания. Это мешало работе при распознавании, поэтому приходилось вручную набирать текст метаданных. Еще один усугубляющий фактор для дальнейшего распознавания включал наличие текстов с дореволюционной орфографией, которые приходилось «переводить» на современный манер, к тому же без каких-либо метаданных;

- размещение метаданных и файлов постатейно на официальном сайте журнала в 2016–2020 гг.; появление понимания о работе поисковых систем и о стандартизации массива произошло со временем, но самые заметные опечатки исправлялись сразу. Из библиотеки передавались целые pdf-файлы каждого тома, далее они распознавались и разделялись на статьи – эта монотонная работа приводила к пропуску некоторых публикаций. В кабинет статьи на сайте прикреплялся файл конкретной статьи с сопоставлением метаданных по общему тому, но без проверки каждого файла перед размещением. Таким образом, иногда происходило дублирование или смещение файлов, и оставшиеся файлы статей тома размещались некорректно;

- присвоение DOI на статьи с 2016 г. После наложения санкций на учредителя (май 2023 г.) произошел переход на указание *eLibrary document number (EDN)*;

- верстка текстов статей в html-версиях с содержанием оглавлений, формулами с использованием *MathJax*, иллюстрациями и ссылками на библиографические источники – с середины 2020 г.;

- корректировка данных. Сформированы факторы для проверки (подробно описаны далее), с 2023 г. по настоящее время;

- формирование модулей для xml-выгрузки для других баз – это подготовка к завершающему этапу – размещению полного архива метаданных в Научной электронной библиотеке (РИНЦ) с гиперссылками на официальный сайт журнала.

Изначально информация размещалась в том виде, как была получена из библиотеки, без учета сложности индексации и SEO-продвижения. Затруднения при создании архива возникали по следующим причинам:

- выполнение работ разными подразделениями;
- смена сотрудников в течение проведения работ;

- деятельность, совмещенная с основной работой сотрудника;

- непонимание важности унификации данных в целом (например, аффилиации авторов);

- недопонимание объема и типа метаданных;

- перемещение сайта с платформы OJS2 (время работы 2016 – 2019 гг.) на версию OJS3 (различия в представлении статей, дата публикации заполнялась автоматически при переносе данных).

После полного размещения архива на сайте журнала для исправления и унификации были выделены следующие основные позиции:

- названия статей – проверка соответствия оригиналу, унификация шрифта (начиная с заглавных букв, остальное строчными, за исключением имен и названий);

- авторы – проверка указания всех авторов, написания их фамилий и инициалов, а также создание профилей активных авторов (с кем редакция работает в текущее время) и размещение всех статей в их профили – для дальнейшей настройки поиска в архиве «по автору» при наличии однофамильцев;

- сведения об авторах – указание имеющихся данных о должности, ученой степени и звании, аффилиации, контактных данных. Во многих статьях первых 40–60 лет издания журнала у автора указана только должность – идентификацию во многих случаях невозможно произвести;

- аффилиация – на протяжении издания журнала многие организации претерпевали реорганизацию и переименование, менялись требования к указанию названий аффилирующих организаций. Поэтому принято решение указывать, например, университеты без обозначения ФГБОУ, исключать сокращения и аббревиатуры, для часто повторяющихся организаций основные названия указывать через аббревиатуры. Например, для Горного университета было выделено несколько основных, резко различающихся названий (менялось 12 раз): «Ленинградский ...», «... университет/институт», «Национальный минерально-сырьевой университет «Горный»;

- аннотации – большое количество статей не имело аннотаций, в этом случае были использованы первые вводные 1–2 абзаца текста для введения читателя в содержание статьи. В ином случае для написания аннотации необходимо было бы привлекать специалистов. Процесс изучения статьи, написания аннотации и выделения ключевых слов мог бы затянуться на годы, а также возник бы немаловажный вопрос сохранения авторского права, при этом в спешке можно было неправильно интерпретировать сведения о статье и ввести в заблуждение читателей;

- ключевые слова – аналогично с аннотациями к выделению этого типа данных не было требований, и в последующих томах поле оставляли пустым;

- список литературы – в более современных статьях список литературы проверялся на правильность переноса данных. При углублении в архив стало понятно, что требований к списку литературы также не существовало. Было принято решение о выделении в список литературы источников из сносок или непосредственно из текста, поскольку они и являются базой для обзора вопроса о предмете изучения в статье. Далее эти списки литературы также индексируются в системе сайта, чтобы была возможность по поисковому запросу выдавать их в результатах поиска;

- первая и последняя страницы текста – перепроверка из pdf-файла;

- сами pdf-файлы – проверка наличия файла статьи и тома, правильности загруженного файла в кабинет статьи, а также проверка пропущенных страниц при сканировании для пополнения этих материалов;

- тип статьи – при формировании нормативов для более современных статей стало обязательным указывать тип статьи, который изначально не обозначался. Принято решение выделять следующие типы: обзорная статья, научная статья, слово редакции, биография (сюда входят юбилеи, памятные даты, воспоминания и пр.);

- перевод метаданных – был использован перевод метаданных, полученных библиотекой: если в статье уже есть перевод метаданных, то его вычитывают, грубые ошибки исправляют без изменения стиля автора; для статей, имеющих метаданные на одном языке, их переводят (русский – английский / английский – русский). Для обозначения и унификации некоторых данных, например, должностей, собраний глоссарий, который пополняется по мере работы с текстом;

- дата публикации статьи – ставится в зависимости от указания даты в конкретной статье, в остальных случаях – сверяется с обложкой тома (выходные данные);

- тематические теги статьи – формируются для дальнейшей настройки поиска по архиву – выдачи похожих статей и фильтру – по тематической области научных исследований. Изначально в журнале публиковались статьи по областям наук, по которым журналы еще не издавались [23], затем в тома набирались статьи по мере подачи и формировались не по темам, далее тома посвящались конференциям, и названия были длинные и соответствовали названиям секций. Таких тегов выделено 12 по основным тематикам журнала или ранее опубликованным дисциплинам: геология, горное дело, нефтегазовое дело, металлургия и обогащение, энергетика и электротехника, машиностроение, геоэкология, охрана труда, отраслевая экономика, гуманитарные науки, социальные науки, слово редактора. Принято решение к каждой статье указывать не более трех тегов. Например, раздел «Проблемы геодинамической и экологической безопасности при освоении месторождений нефти и газа, их хранении и транспортировке» будет внесен в «Нефтегазовое дело», «Геоэкология», и автору, который проводит поиск на сайте, будет проще найти материал, воспользовавшись поиском по ключевым словам и по дате публикации или автору.

Проверка и внесение исправлений в распознанные и размещенные сведения необходимы для корректности формирования блока цитирования (указан под ключевыми словами и на сайте, и в pdf-версии статьи) для удобства использования авторами. Блок цитирования имеет несколько распространенных форматов ссылок (APA, Chicago, Harvard и др.), в том числе созданный личный формат журнала «PMI», который используется в русскоязычной версии.

ОБСУЖДЕНИЕ

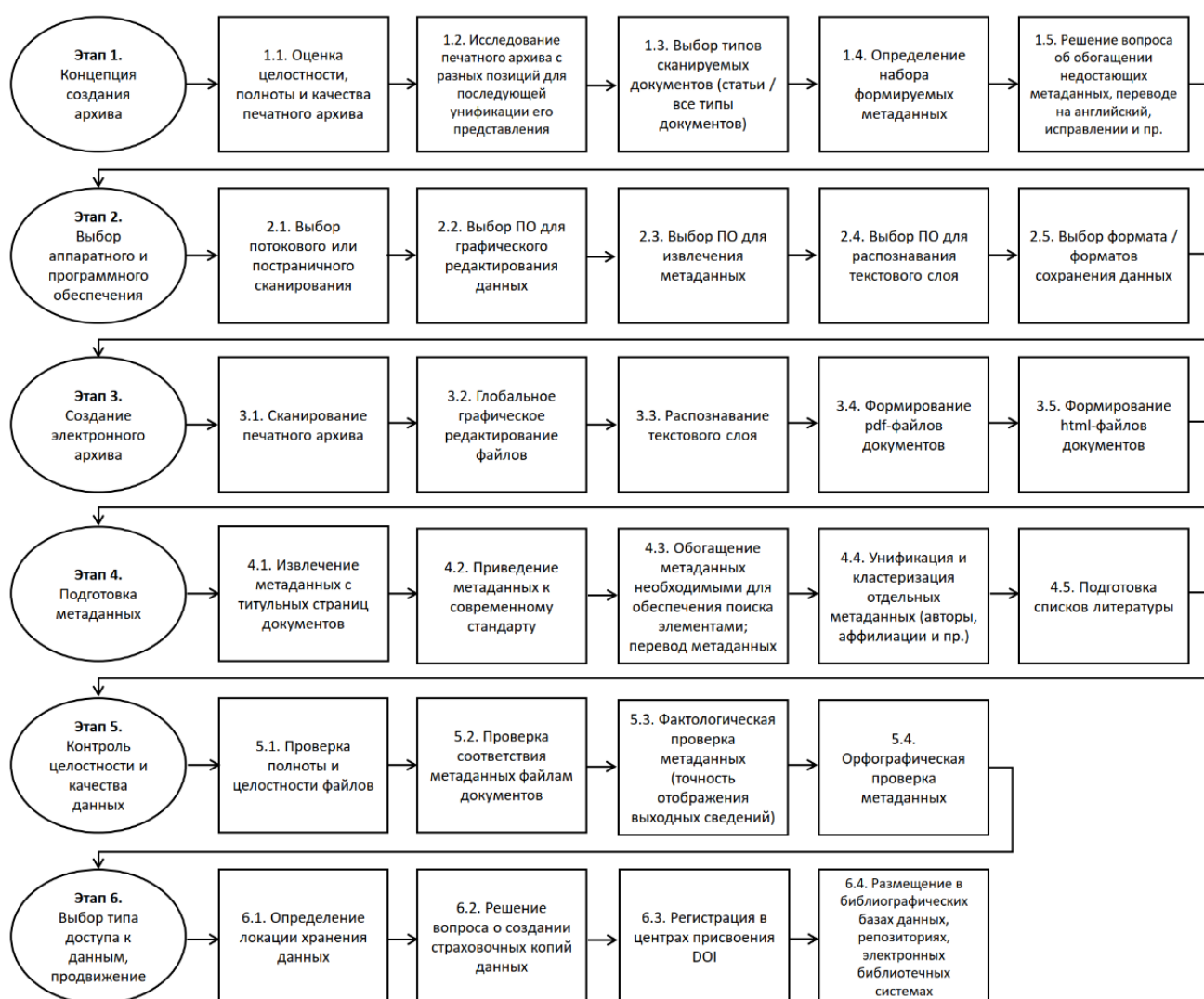
Работа над созданием архива двух представленных отечественных журналов заняла около 10 лет разнообразных по своим функциям действий, потребовала привлечения множества специалистов, решения нетипичных вопросов, на которые не могли ответить ни библиографы, ни менеджеры баз данных, так как не было понимания всех возможностей современных поисковых систем и запросов читателей и авторов.

На рисунке показана предлагаемая нами обобщенная схема этапов создания полнофункционального электронного архива научного журнала. Любые редакции, издательства и библиотеки могут использовать эту схему в качестве ориентира для работы.

Журнал «Геология и геофизика» из ближайших целей определяет для себя присвоение DOI всему архиву оригинальной версии и размещение метаданных с полными текстами на платформе Научной электронной библиотеки eLibrary.ru. Ведется подготовка аналогичного архива для англоязычной версии журнала за 1991–2002 гг. (период, за который имеется печатный архив) и поиск в зарубежных библиотеках более ранних выпусков, вплоть до 1974 г.

Журнал «Записки Горного института» по перечисленным позициям обрабатывает тома по десятилетиям, далее планирует выборочную проверку статей и подготовку xml-файла для размещения в библиотеке eLibrary.ru. Окончательный этап (параллельный с выгрузкой в eLibrary.ru) предполагает создание функционального поисковика на официальном сайте для еще более удобного и качественного поиска статей.

После завершающего этапа создания архива статей важно продолжать его продвигать различными способами (дайджесты, поздравления, рассылки, посты в соцсетях), чтобы массив новой информации не перебил желание искать старые публикации в огромном потоке новых. К тому же многие ученые могут и не предполагать о наличии в каком-то конкретном журнале архива большой глубины. Другой важный момент – это сбор статистики с ЭБС, где размещается журнал (посещение, скачивания, подписки на коллекцию), продолжение или разрыв договора и мониторинг новых баз, а также отслеживание репозиторий других стран для размещения своего журнала.



Этапы создания полнофункционального электронного архива научного журнала

Уже сейчас нашу работу можно считать успешной, поскольку по данным *Scopus* на статьи журнала «Записки Горного института» за 1909–2016 гг. (имеющие метаданные, pdf полного текста, аннотацию только на русском языке) – более 500 ссылок, а на момент начала создания сайта и подготовки журнала к индексации в *Scopus* их было не более 60. В редакцию журнала поступают сообщения с указанием на опечатки в размещенном архиве (сейчас проводится верификация данных архива), что свидетельствует о заинтересованности ученых в исследованиях тех лет.

В ряде современных статей про великих ученых-основоположников, историю развития научной школы или организации встречаются ссылки на статьи, опубликованные в начале прошлого столетия, по транслитерированному названию журнала “*Zapiski Gornogo Instituta*”.

Подобные работы также проводились на сформированном архиве журнала «Геология и геофизика» [30].

ЗАКЛЮЧЕНИЕ

С учетом слабой представленности в электронном пространстве архивов российских научных журналов, этой проблеме следует уделять большее внимание.

1. Решающую роль в подготовке журнальных архивов могут сыграть научные библиотеки, поскольку они обладают высоким профессиональным опытом работы с научной информацией, нередко имеют необходимое оборудование для сканирования и программное обеспечение для распознавания оцифрованных образов и верстки pdf- и html-версий, а также полные архивы российских периодических изданий. Кроме того, в этой работе библиотеки реализуют одну из их основных функций – сохранение информации.

2. Концептуально полезным является перевод метаданных архивных статей на английский язык (при отсутствии переводной версии издания). Этим повышается внимание к журналу и российской науке, так как для иностранных специалистов создается возможность находить тематические статьи и получать полные тексты для дальнейшего изучения, даже с помощью машинного перевода.

3. Вопросы авторского права на электронный архив, финансовых затрат на присвоение идентификаторов DOI, содержания технологической инфраструктуры для хранения цифровых копий, безусловно, должны решаться совместными усилиями всех заинтересованных участников.

4. На примере двух ведущих изданий по наукам о Земле – журналов «Геология и геофизика» и «Записки Горного института» показано, что все задачи по разработке полнофункциональных электронных архивов печатных версий с созданием поисковых возможностей вполне реализуемы при четкой постановке цели и сроков, определении обязательных полей для обработки данных и назначении ответственных за эту работу.

СПИСОК ЛИТЕРАТУРЫ

1. Гуреев В.Н., Кириллова О.В., Мазов Н.А. Оценка представленности и индексации российских научных журналов в Web of Science Core Collection // Научный редактор и издатель. – 2023. – Т. 8, № 2. – С. 84-98.
2. Гуреев В.Н., Мазов Н.А., Метелкин Д.В. О некоторых причинах перехода российских авторов в зарубежные журналы // Управление наукой: теория и практика. – 2022. – Т. 4, № 3. – С. 20-34.
3. Кириллова О.В. Подводя своеобразные итоги десятилетия и рисуя планы... // Научный редактор и издатель. – 2022. – Т. 7, № 1. – С. 8-11.
4. van Raan A.F.J. Sleeping Beauties in science // *Scientometrics*. – 2004. – Vol. 59, № 3. – P. 467-472.
5. Schonfeld R.C. The Idea at Denison, the Project at Mellon // *JSTOR. A History*. – Scottsdale: Princeton University Press, 2003. – 405 p.
6. Tenopir C., King D.W., Edwards S., Wu L. Electronic journals and changes in scholarly article seeking and reading patterns // *Aslib Proceedings*. – 2009. – Vol. 61, № 1. – P. 5-32.
7. Keyhani A. Creating an electronic archive: Who should do it and why? // *Serials Librarian*. – 1998. – Vol. 34, № 1-2. – P. 213-224.
8. Соколова И.С. Электронный архив научных журналов РАН по тематическому направлению «Химия, Биология и Физиология» *PhysChemBio.ru* // Научно-техническая информация. Сер. 1. – 2021. – № 5. – С. 31-35.
9. Flecker D. Preserving Scholarly E-Journals // *D-Lib Magazine*. – 2001. – Т. 7, № 9. DOI: 10.1045/september2001-flecker.
10. Hughes J.A. Issues and concerns with the archiving of electronic journals // *Science and Technology Libraries*. – 2004. – Vol. 22, № 3-4. – P. 113-136.
11. Waters D.J. Urgent action needed to preserve scholarly electronic journals // *Digital Library Federation*. – 2005. – URL: <https://www.arl.org/wp-content/uploads/2005/10/ejournal-preservation-15oct05.pdf> (дата обращения: 13.06.2024).
12. Mabe M. The growth and number of journals // *Serials*. – 2003. – Vol. 16, № 2. – P. 191-197.
13. Mabe M., Amin M. Growth dynamics of scholarly and scientific journals // *Scientometrics*. – 2001. – Vol. 51, № 1. – P. 147-162.
14. Machovec G. E-journal Archives and Preservation An Executive Overview // *The Charleston Advisor*. – 2006. – Vol. 7, № 4. – P. 51.
15. Balnaves E., Chehade M. Smart Client approaches to digital archiving of e-journals // *Electronic Library*. – 2009. – Vol. 27, № 5. – P. 837-845.
16. YEA: The Yale electronic archive one year of progress. Report on the digital preservation planning project. – New Haven: Federation D. L., 2002. – 77 p.
17. Mostert P. TULIP AT Elsevier Science // *Library Hi Tech*. – 1995. – Vol. 13, № 4. – P. 25-30.
18. Объявление о создании нового Электронного архива научных публикаций по тематике «Химия, биология и физиология» // *Журнал эволюционной биохимии и физиологии*. – 2021. – Т. 57, № 2. – С. 93.
19. Лобанова А.С. Создание электронной коллекции «Доклады Академии наук» в БЕН РАН // Сборник докладов 26-й Международной конференции «Информационные технологии, компьютерные системы и издательская продукция для библиотек».

- лиотек» (LIBCOM-2022) (20–25 ноября 2022 г., Суздаль). – Москва: ГПНТБ России, 2022. – С. 100–106.
20. Mazov N.A., Gureyev V.N. Nontraditional approaches to assessing journal importance: Case study of Russian journals on Earth sciences // *Serials Review*. – 2020. – Vol. 46, № 1. – P. 10–20.
21. Мазов Н.А., Гуреев В.Н., Эпов М.И. Российские публикации и журналы по наукам о Земле в международных базах данных // *Вестник Российской академии наук*. – 2015. – Т. 85. – № 1. – С. 26–31; Mazov N.A., Gureev V.N., Epov M.I. Russian publications and journals on Earth sciences in international databases // *Herald of the Russian Academy of Sciences*. – 2015. – Т. 85, № 1. – С. 20–25.
22. Бокий Б.В. 50 лет «Запискам Горного института» (Историческая справка) // *Записки Горного института*. – 1959. – Т. 40. – С. 3.
23. Ребещенкова И.Г. «Записки Горного института»: их история и модернизация (к 110-летию выхода в свет) // *Записки Горного института*. – 2016. – Т. 221. – С. 773.
24. Магнус О.А. Указатель статей, напечатанных в I–X тт. «Записок Ленинградского горного института» (1907–1936) с алфавитным указателем авторов их // *Записки Горного института*. – 1937. – Т. 10, № 3. – С. 93–104. – URL: <https://pmi.spmi.ru/pmi/article/view/15584>
25. Кумурджи М.И. Систематический указатель статей, опубликованных в «Записках Ленинградского Горного института» с I по XXXVII т. (1907–1958 гг.) // *Записки Горного института*. – 1959. – Т. 40. – С. 127.
26. Карха Ж.В. Библиографический указатель статей, напечатанных в томах 35–63 «Записок» // *Записки Горного института*. – 1991. – Т. 127. – С. 3–104.
27. Карха Ж.В., Иванова Я.А. Библиографический указатель статей, напечатанных в томах 64–114 «Записок» // *Записки Горного института*. – 1992. – Т. 131. – С. 3–122.
28. Журавский А.М. Математика и механика в «Записках Ленинградского горного института» // *Записки Горного института*. – 1959. – Т. 40. – С. 47.
29. Шафрановский И.И. Кристаллография, минералогия и петрография в «Записках Ленинградского горного института» // Там же. – С. 47.
30. Мазов Н.А., Гуреев В.Н. Ретроспективный библиометрический анализ журнала «Геология и геофизика» за 1960–2012 гг. как основа отражения исследований в области наук о Земле // *Проблемы наукометрии: состояние и перспективы развития: Тезисы докладов Международной конференции (10–12 октября 2013 г., Москва)*. – Москва: Институт проблем развития науки РАН, 2013. – С. 84–86.

Материал поступил в редакцию 13.06.24.

Сведения об авторах

МАЗОВ Николай Алексеевич – кандидат технических наук, ведущий научный сотрудник информационно-аналитического центра Института нефтегазовой геологии и геофизики им. А.А. Трофимука СО РАН; Государственная публичная научно-техническая библиотека СО РАН, г. Новосибирск
e-mail: MazovNA@ipgg.sbras.ru

ГУРЕЕВ Вадим Николаевич – кандидат педагогических наук, заведующий информационно-аналитическим центром Института нефтегазовой геологии и геофизики им. А.А. Трофимука СО РАН; Государственная публичная научно-техническая библиотека СО РАН, г. Новосибирск
e-mail: GureyevVN@ipgg.sbras.ru

СИНЯВИНА Светлана Викторовна – кандидат технических наук, директор издательского дома, Санкт-Петербургский горный университет императрицы Екатерины II, г. Санкт-Петербург
e-mail: Sinyavina_SV@pers.spmi.ru