УДК 023.5:004.658.2

DOI: 10.32340/2414-9101-2019-3-83-86

Н. С. Редькина, доктор педагогических наук

Государственная публичная научно-техническая библиотека
Сибирского отделения Российской академии наук (Новосибирск, Россия),
Новосибирский государственный педагогический университет
(Новосибирск, Россия)
to@spsl.nsc.ru

ПОДГОТОВКА БИБЛИОТЕКАРЕЙ В ОБЛАСТИ УПРАВЛЕНИЯ ИССЛЕДОВАТЕЛЬСКИМИ ДАННЫМИ

Аннотация. Охарактеризована роль специализированных программных средств управления крупными массивами исследовательских данных в контексте организации научного труда современного учёного. Представлено описание основных этапов жизненного цикла процесса управления исследовательскими данными; рассмотрен ряд потенциальных рисков, сопряжённых с долгосрочным хранением данных в электронном формате, а также некоторые преимущества и недостатки свободного обмена данными между учёными, коллективами исследователей. Автор излагает собственный взгляд на состав базовых компетенций современного библиотекаря-профессионала в области управления потоками цифровых данных и информационной поддержки научного труда (в частности, в сфере визуализации данных, интеллектуального анализа контента крупных коллекций электронных документов и цифровых репозиториях и др.).

Ключевые слова: *библиотекари, профессиональные компетенции библиотекаря, управление исследовательскими данными, жизненный цикл данных, метаданные, идентификаторы данных, открытый доступ, репозитории данных.*

Учёные прикладывают значительные физические и интеллектуальные усилия для сбора, сопоставления и анализа данных, управления ими перед публикацией результатов; данные остаются ценным ресурсом даже после окончания проекта, так как будущим исследователям не требуется собирать их заново. Мировым научным сообществом одобрены принципы FAIR (Findable – легкий поиск определённых наборов данных, Accessible – удобный доступ (по условиям доступа и возможностям хранения в течение длительного времени, Interoperable – совместимость с другими наборами данных или программным обеспечением, Re-usable – многократное использование в последующих исследованиях и дальнейшей обработке). Для того чтобы данные соответствовали принципам FAIR, они должны быть доступны для поиска с помощью метаданных, идентифицируемыми и локализуемыми с помощью стандартного механизма идентификации, такого как постоянные и уникальные «цифровые идентификаторы объектов» (DOI – Digital Object Identifiers) и др.

Все больше научных журналов предлагают модели для улучшения видимости данных, совместного и открытого доступа к ним, что повышает прозрачность и достоверность исследований, а также позволяет критически оценивать полученные результаты. Журналы, требующие наличия данных в качестве предварительного условия для публикации, – это «Nature», «PLOS», «Science» и др. Главная задача политики этих журналов – облегчить читателям доступ к данным для лучшего понимания научных результатов, оценки преимуществ и определения достоверности исследования. Учитывая разнообразие типов данных и способов их представления, журналы рекомендуют хранить данные в открытых хранилищах или приводить основные данные в тексте статьи,

а дополнительную информацию предоставлять по запросу. Так, издательство «Elsevier» предлагает связывать данные со статьей, загружая соответствующий компьютерный код и данные в репозиторий (например, «Chemical Data Collections»).

Управление исследовательскими данными (англ. Research Data Management, RDM), является неотъемлемой частью жизненного цикла научного проекта, включает в себя сбор, документирование, хранение, создание резервных копий, совместное использование, обеспечение целостно-

сти, безопасности, управление версиями, надёжное планирование и стратегическое управление данными. Выбор корректных форматов данных (структурированных и неструктурированных), онтологий и программных средств, необходимых для проведения экспериментов или создания набора данных, является важным этапом исследовательского цикла. Соответствующие стандартам форматы и имена файлов гарантируют, что данные могут быть идентифицированы и доступны в будущем. Данные нередко требуют пояснений, поэтому они должны сопровождаться метаданными (информацией, которая описывает данные). Использование соответствующих методов хранения и резервного копирования помогает защитить данные исследований от возможных потерь, а также обеспечивает доступ к ним в долгосрочной перспективе.

Для того чтобы управлять данными исследований необходимо знать основные понятия и этапы жизненного цикла данных, преимущества подготовки плана управления данными, идентификаторы и требования к цитированию данных, требования к совместному использованию данных, возможные риски по долговременному сохранению цифровых данных, преимущества и проблемы, связанными с обменом исследовательскими данными, различные типы открытых лицензий и возможности лицензирования данных, надежные репозитории данных и др.

Управление исследовательскими данными связано с организацией данных с самого начала исследовательского цикла и до архивирования полученных результатов и их публикации. Жизненный цикл данных состоит из нескольких этапов:

- 1) Планирование и разработка исследовательского проекта План управления данными.
- 2) Сбор данных (либо новых первичных данных, либо уже существующих). Определение форматов файлов данных.
- 3) Анализ данных, генерирование или поиск дополнительных данных, необходимых для проекта.
- 4) Отбор данных и определение периода их хранения.
- 5) Документирование. Обработка и структурирование данных, подготовка метаданных, организация файлов. Проверка требований к документированию и ограничений для учреждений и спонсоров исследования.
- 6) Архивирование. Обзор существующих источников данных, изучение различных аспектов по архивированию (возможные затраты, согласие авторов исследования и конфиденциальность, определение потенциальных пользователей данных и др.). Создание резервных копий.
- 7) Организация доступа. Публикация данных. Передача в архив данных, выбранных для хранения. Определение условий доступа. Цитирование данных.
- 8) Хранение и обеспечение сохранности с учётом меняющейся технологической среды. Организация долгосрочного хранения во время и после окончания работы над исследовательским проектом (зависит от наличия соответствующей технической инфраструктуры).
- 9) Повторное использование. Определение условий для повторного использования данных, предоставление доступа для других исследователей.

Для того чтобы данные исследования были организованы, структурированы и доступны, необходим План управления данными (англ. Data Management Plan, DMP) – формальный документ, в котором описаны:

- типы данных, которые будут получены во время исследований,
- политика использования данных (финансирование, институциональная и юридическая стороны применения данных),
- методы управления данными (резервное копирование, хранение, контроль доступа, архивирование),
- требуемые средства и оборудование (пространство на жестком диске, резервный сервер, репозиторий),
- обмен данными,
- этические и юридические вопросы или ограничения на совместное использование данных,
- возможности повторного использования и обеспечение долгосрочного сохранения.

Открытый доступ к данным исследований является основополагающим принципом программы Horizon 2020 [1]. Исследователи, подающие заявки на финансирование по этой программе,

должны разработать план управления данными, в котором излагается стратегия сбора, хранения и доступности данных, созданных в рамках проектов, финансируемых Европейским Советом.

Учёные считают, что библиотекари-исследователи становятся незаменимым партнёром исследовательского процесса [2]. Для того чтобы идти в ногу со временем, библиотекари должны предоставлять исследовательские данные, оказывать услуги, получать новые навыки в области анализа данных и визуализации, интеллектуального анализа контента и др. Аналогичной позиции придерживаются и австралийские учёные К. Howard и Н. Partridge [3], которые считают, что помимо определённых общих навыков и знаний, включая такие как решение проблем, критическое мышление, критический анализ, письменное и устное общение, адаптивность и лидерство, специалисты должны изучать в образовательной программе курс управления информацией (данными). N. Upadhyay [4] полагает, что библиотекари будут играть жизненно важную роль в оказании помощи академическим и исследовательским сообществам, разработке и внедрению методик описания данных, технологий эффективного хранения, управления и поиска.

Управление исследовательскими данными – одно из перспективных направлений в деятельности библиотек. Необходимые знания, которыми должен обладать библиотекарь, отражены в Руководстве по управлению исследовательскими данными, представленном на сайте ГПНТБ СО РАН [5], состоящем из следующих разделов:

– «Введение»;	– «Метаданные»;
– «Управление исследовательскими данными»;	– «Идентификаторы данных»;
– «Жизненный цикл данных»;	– «Безопасность и защита данных»;
– «План управления данными»;	– «Хранение данных»;
– «Шаблоны плана управления данными»;	– «Обмен и архивирование»;
– «Классификация данных»;	– «Цитирование данных»;
– «Форматы файлов данных»;	– «Примеры цитирования данных»;
– «Организация файлов»;	– «Авторское право и конфиденциальность»;
	– «Специалисты по RDM».

Каждый раздел детально раскрывает особенности работы с данными на всех этапах жизненного цикла данных. К примеру, в разделе «Метаданные» отражены три категории метаданных: 1) описательные – общие сведения о данных, такие как название, автор, реферат, ключевые слова, которые помогают пользователям находить источники в интернете через поиск; 2) административные – данные о способах сохранения, дате создания, управления правами, программном обеспечении, форматах файлов, проверках целостности файлов; 3) структурные – сведения о связи данных друг с другом для облегчения навигации, например схема, описывающая отношения между таблицами в базе данных. Отмечено, что поиск информации через метаданные будет более эффективным, если использовать стандартизированные термины и фразы при помощи контролируемых словарей, файлов полномочий, классификаций, тезаурусов, онтологий и таксономий (например, BARTOC, Базельский реестр тезаурусов, онтологий и классификаций). Кроме того, указано, что стандарты метаданных варьируются в зависимости от дисциплины, поэтому рекомендуется выбрать стандарт метаданных, соответствующий дисциплинарным требованиям или типу данных, которые будут генерироваться / собираться:

- Dublin Core – словарь (семантическая сеть) основных понятий английского языка, предназначенный для унификации метаданных для описания широчайшего диапазона ресурсов.
- DDI (Data Documentation Initiative – международный стандарт для описания исследований, опросов, статистических файлов данных и информации в области социальных наук.
- EML (Ecological Metadata Language) – стандарт специально для экологических дисциплин.
- ISO 19115 и FGDC-CSDGM – стандарты, используемые для описания географической информации (карт, файлов географических информационных систем (ГИС), изображений и других ресурсов данных на основе местоположения). Содержат информацию об идентификации, объеме, качестве, пространственных и временных аспектах, содержании, пространственной привязке, изображении, распределении и других свойствах цифровых географических данных и услуг.
- MINSEQE (MINimal information about high throughput SEQuencing Experiments) – стандарт по геномам.

– FITS (Flexible Image Transport System) – цифровой стандарт для описания файлов по астрономии.

– MIBBI – это набор руководящих принципов для представления данных, полученных соответствующими методами в биологических и биомедицинских исследованиях (например, <https://fairsharing.org/collection/MIBBI>).

Данные описываются в соответствии с различными стандартами метаданных, которые предназначены для документирования содержимого файлов, раскрытия технических характеристик файлов, выражения отношений между файлами в наборе данных и др. Важным стандартом по подготовке метаданных является DataCite Metadata Schema, используемый для публикации и цитирования исследовательских данных [6].

Представленный в Руководстве материал определяет направления в развитии услуг по управлению данными, а также новые компетенции библиотечных специалистов в этой области (например, поиск информации о планах управления данными, стандартах метаданных или практиках цитирования, подготовке данных для архивирования, политиках цифрового сохранения, безопасности и стратегиях обмена данными).

Список литературы

1. H2020 Programme: Guidelines on FAIR Data Management in Horizon 2020. – Text : electronic // European Commission. Directorate-General for Research and Innovation. – Brussels. – URL: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf (05.05.2019).
2. Borrego, Á. Librarians as Research Partners: Their Contribution to the Scholarly Endeavour Beyond Library and Information Science. – Text : electronic / Á Borrego, J. Ardanuy, C. Urbano // The Journal of Academic Librarianship. – 2018. – Vol. 44, is. 5. – P. 663–670. – (02.05.2019). – DOI: <https://doi.org/10.1016/j.acalib.2018.07.012>.
3. Howard, K. Passion trumps pay: A study of the future skills requirements of information professionals in galleries, libraries, archives and museums in Australia. – Text : electronic / K. Howard, H. Partridge, H. Hughes, G. Oliver // Information Research. – 2016. – Vol. 21, is. 2. – 22 p. – URL: <https://files.eric.ed.gov/fulltext/EJ1104373.pdf> (05.05.2019).
4. Upadhyay, N. Trends that Will Affect Technology and Resource Decision in Academic Libraries in near Future. – Text : electronic // 4th International Symposium on Emerging Trends and Technologies in Libraries and Information Services, ETTLIS 2015. – India ; Noida ; 2015. – P. 75–79.
5. Руководство по управлению исследовательскими данными. – Текст : электронный // Гос. публ. науч.-техн. б-ка Сиб. отд-я Рос. академии наук. – Новосибирск, 2019. – URL: <http://www.spsl.nsc.ru/naukresursy-i-uslugi-gpntb-so-ran-dlya-nauki-i-biznesae-i-biznesu/rdm> (дата обращения: 05.05.2019).
6. Smaele de M. DataCite Metadata Schema Documentation for the Publication and Citation of Research Data / M. Smaele [and others] // DataCite Metadata Schema [site]. – URL: https://schema.datacite.org/meta/kernel-4.1/doc/DataCite-MetadataKernel_v4.1.pdf (05.05.2019).

Natalia S. Red'kina, Dr. of Pedagogical Science
State Public Scientific and Technical Library
of the Siberian Branch of the Russian Academy of Sciences (Novosibirsk, Russia)
Novosibirsk State Pedagogical University (Novosibirsk, Russia)
to@spsl.nsc.ru

TEACHING OF LIBRARIANS' SKILLS IN RESEARCH DATA MANAGEMENT

Abstract. The article clears up a significance of specific software application for managing big data resources within the system of information support of a today's scientist's research labor. The author gives a description for the key phases of a life cycle of research data management process, considers risks related with long-term conservation of e-data, some advantages and disadvantages of free data change between researchers, teams of scientists. Also, the author outlines her own view on a set of basic competences of a modern professional librarian in managing streams of e-data, information support of research work (in particular, in data visualization, content analysis of large e-document collections and repositories).

Key words: librarians, professional competences of a librarian, research data management, life cycle of data, metadata, data identifiers, open access, information repository.