

А. И. Алиева-Кенгерли,

директор ЦНБ НАН Азербайджана, кандидат филологических наук

С. В. Кондратьев,

научный сотрудник ЦНБ НАН Азербайджана

ПОДХОД К ОРГАНИЗАЦИИ НАУЧНЫХ ЭЛЕКТРОННЫХ БИБЛИОТЕК

Рассматривая методы разработки современных информационных систем нового поколения, теоретические и практические вопросы создания электронных библиотек, поддерживающих доступ к научным электронным коллекциям, авторы статьи обосновывают комбинированный вариант работ, предполагающий формирование как тематических, так и ведомственных электронных коллекций.

Ключевые слова: единое национальное информационное пространство, электронные коллекции, интеграция ресурсов, распределенные хранилища информационных ресурсов, международные стандарты и форматы, система метаданных.

In considering how to develop modern information systems of new generation, the theoretical and practical issues of creating digital libraries that support access to scientific digital collections, the authors justify a combined version of the work, involving the formation of both the thematic and institutional digital collections.

Keywords: union national information field, electronic collections, integration of the resources, the distributed storage of information resources. The international standards and formats, system of the metadata.

Национальная стратегия и Программа работ по развитию информационно-коммуникационных технологий (ИКТ) в Азербайджане одним из важных шагов построения информационного общества и обеспечения процесса перехода к обществу знаний определили формирование единого национального информационно-коммуникационного пространства, национальных информационных ресурсов (ИР). В числе задач формирования национальных ИР, включающих тематические (межведомственные) и ведомственные ИР, решаются задачи интеграции отдельных хранилищ информации и коллекций знаний.

Развитие сетевого взаимодействия и интеграция электронных ИР учреждений науки, образования и культуры – одно из направлений развития информационного общества. В Европе это направление явилось ключевым аспектом инициативы «i2010» по развитию «коллективной

памяти» на основе создания распределенной электронной коллекции печатных и аудиовизуальных материалов, фотографий, экспонатов музеев, исторических, архивных документов, музыкальных произведений, кинолент.

Среди основных задач развития национальных ИР Азербайджана сегодня: обеспечение широкого использования ИКТ в создании разнообразных электронных коллекций для сохранения, защиты и популяризации исторического, литературного и культурного наследия, включая широкое применение ИКТ в библиотеках, музеях и архивах; создание электронных библиотек-хранилищ научных коллекций по различным отраслям науки и новейшим технологиям; а также интеграция национальных электронных коллекций с международными ИР для обеспечения доступа к ним.

Среди ведомственных ИР Национальной Академии наук Азербайджана (НАНА): материалы краеведческого характера; ценные научные коллекции, архивы и базы данных, научные издания, публикации ученых, труды научных конференций; информацию о научных проектах, научных коллективах; отчеты о научно-исследовательской работе (НИР).

К ИР большого объема относятся: каталоги и фонды Центральной научной библиотеки и библиотечной сети НАНА; музейные ИР – научные коллекции, архивы и выставки Музея истории Азербайджана (свыше миллиона единиц хранения) и Музея национальной литературы НАНА. Для поддержки функционирования НАНА обрабатываются административные ИР – информация о структуре и кадрах НАНА, адресно-справочные сведения (см. сайт www.science.az); нормативно-правовые, финансовые и хозяйственно-административные документы.

Приведенный выше перечень ресурсов свидетельствует о том, что тематические и ведомственные ИР НАНА характеризуются неоднородностью, разнообразием представления, сверхбольшими объемами. Для их отображения требуются усилия по описанию отдельных электронных документов и коллекций в целом.

Задачи управления знаниями и интегрированной совокупностью ИР, включая научные коллекции свыше трех десятков научно-исследовательских институтов (НИИ) НАНА можно решать путем создания распределенных хранилищ информационных ресурсов:

✓ хранение **тематических ИР** было предложено решать путем создания Национальной научной электронной библиотеки (НЭБ) и централизованного хранилища электронных документов в ЦНБ НАНА [1–3] – с использованием технологии автоматизированной библиотечно-инфор-

мационной системы (АБИС «ИРБИС») для отражения в электронном каталоге публикаций в официальных источниках и создания полнотекстовых баз данных с доступом через Web-ИРБИС;

✓ создание **ведомственных ИР** для НИИ НАНА было предложено решать на основе единой концепции создания и ведения институционального репозитория [10], построенного на основе системы EPrints, инициативы открытых архивов и использования протокола OAI-PMH.

С этой целью были проанализированы существующие стандарты и подходы для хранения ИР и доступа к информации различной тематики [4–11]:

- стандарты интероперабельности – Протокол удаленного доступа к ресурсам Z39.50 (ISO 23950) и подход инициативы открытых архивов OAI [7];

- стандарты и форматы кодирования ресурсов (форматов данных, например возможности postscript, PDF, TIFF, GIF), в т. ч. структурированной информации (языки SGML, XML);

- стандарты и форматы идентификации ресурсов (как URL и инициативы DOI и ARK, так и системы The Handle System, механизм постоянного адреса PURL);

- стандарты и форматы описания ресурсов (формат описания книг и других библиографических изданий ISBD, его машиночитаемый аналог MARC, формат для архивов EAD, описание на основе RDF и др.);

- стандарты архивирования данных (эталонная модель «Открытых архивных систем» OAIS и др. средства, позволяющие организовать долгосрочную сохранность) [5, 13];

- стандарты управления записями.

Использование международных стандартов и форматов – единственная стратегия, ведущая к цели при создании ИР больших объемов с унифицированным доступом к данным. При постоянном развитии ИКТ приверженность стандартам облегчает передачу информации между аппаратными средствами и платформами программного обеспечения, а также способствует интеграции и обмену ИР. Ресурсы, кодированные с применением открытых стандартов и форматов, имеют больший шанс на сохранение при длительном хранении данных, нежели ресурсы, кодированные при помощи локальных методов и форматов. Другой важнейшей задачей является интеграция ресурсов на основе единого описания на верхнем уровне – уровне коллекций [6].

Возникает вопрос относительно правильной организации цифровых коллекций с точки зрения интеграции информационного обслуживания

пользователей, а также: избирательного распространения информации, внедрения общих стандартов организации метаданных и обмена данными, вопросов секретности, безопасности данных, соблюдения авторских прав и др. В 2004 г. группа Framework Advisory Group под патронажем Организации национальных стандартов США по информации (NISO) подготовила издание основных правил для построения качественных цифровых коллекций.

С развитием процессов оцифровки ИР специалисты стали больше внимания уделять стандартизации действий разработчиков коллекций не во время фазы отладки системы, а действиям при формировании качественных коллекций, служащих потребностям различных групп пользователей. Планка совершенствования была установлена на таком уровне, чтобы достичь удобства и простоты использования коллекции, доступности ресурсов и релевантности ожиданиям пользователей.

Индикаторы качества при создании коллекций устанавливаются для следующих основных четырех сущностей [4]:

- коллекций (организованных групп объектов, экземпляров);
- объектов (цифровых материалов);
- метаданных (информации, описывающей объекты);
- проектов (инициатив по созданию и управлению коллекциями).

Принципы формирования научных коллекций включают:

- подробно разработанную, согласованную и документированную (до начала работ по оцифровке объектов) политику развития коллекции, максимально приближенную к целям и статусу организации;
- описание характеристик коллекции, в том числе охват, формат, ограничения по доступу, указание владельца коллекции и любой другой информации, важной для определения подлинности, целостности и интерпретируемости коллекции;
- поддержка во времени, финансирование проекта в течение длительного периода;
- обеспечение круглосуточной доступности и устранение препятствий при использовании;
- соблюдение прав интеллектуальной собственности;
- ведение статистики и применение методов библиометрии для анализа использования и разработки стратегий развития коллекции;
- и, наконец, хорошая коллекция вписывается в контекст основных национальных и международных инициатив по цифровым библиотекам.

Создание качественной коллекции ИР невозможно без качественного подхода к созданию системы метаданных.

Метаданные различных типов создаются во время каждого периода жизненного цикла информации. На стадии создания информации фиксируются метаданные об авторах, людях, внесших вклад в создание объекта, об источниках появления объекта. На уровне организации информации формируются метаданные по теме, истории опубликования сведений, целевой аудитории, для которой информация предназначена, а также общее описание объекта и другая информация, создаваемая каталогизаторами и систематизаторами. На стадии доступа и использования информации метаданные могут отражать привилегии по доступу к данным, сведения о праве на копирование, данные по сохранности. Важно помнить, что различные типы метаданных об объекте создаются различными людьми и на различных стадиях жизни информационного объекта.

Метаданные бывают трех основных типов. Описательные (дескриптивные) метаданные помогают пользователям найти объект, отличить один объект от другого, понять тематику и содержимое объекта. Административные метаданные помогают администраторам коллекции отслеживать управление файлами цифровых объектов, интеллектуальными правами, сохранностью. Структурные метаданные документируют взаимосвязи между объектами (статьями сборника, томами сериальных публикаций, страницами и главами книги).

Принципы формирования метаданных научных коллекций следующие:

а) метаданные должны соответствовать материалам (объектам) в коллекции, пользователям коллекции и быть предназначены как для текущих нужд, так и для вероятного использования цифрового объекта в будущем. Существует множество (более 15) разработанных схем метаданных для цифровых объектов, веб-сайтов, и электронных ресурсов, используемых образовательными, научными и культурными учреждениями. Часто к материалам в данной коллекции применима не одна схема, а несколько, например форматы MARC, Dublin Core, METS используются как для нужд локального хранилища ИР, так и для целей обмена данными с другими провайдерами ИР;

б) качественные метаданные поддерживают интероперабельность. Это достигается двумя путями – либо использованием одной и той же стандартной схемы метаданных всеми организациями – участниками создания коллекций по данной тематике (например, MARC+DC в библиотечной среде), либо путем создания средств трансформирования записей одной системы в выбранную другую схему на основании таблицы соответствия (например, система EPrints [15] позволяет не только

поддерживать институциональный репозиторий со множеством форматов для оцифрованных данных, но и осуществлять выгрузку экземпляров-единиц хранения в форматы Dublin Core и METS). Другой путь поддержки интероперабельности – использование форматов метаданных и протокола «сбора урожая» – Инициативы открытых архивов (the Open Archives Initiative – Protocol for Metadata Harvesting, OAI-PMH). Контрастирующий подход, принятый OAI, заключается в том, что все провайдеры данных переводят все их «родные» метаданные к виду общего основного набора элементов и выставляют это для «сбора урожая» с помощью протокола Metadata Harvesting Protocol (MHP). Протокол используется для обмена структурными (мета)данными в ЭБ, музеях, репозиториях электронных препринтов, исследовательских проектах, корпоративных интранет-сетях;

в) качественные метаданные включают контроль авторитетности и стандарты для контента, включая использование управляемых словарей (классификационных схем), помогающих пользователю при описании объекта и при его поиске, а также служащих для установления связей с другими объектами. Например, использование УДК и ГРНТИ служит целям классификации информации в системе «ИРБИС», а использование классификации диссертаций с указанием кодов тематики в институциональном репозитории;

г) качественные метаданные включают четко определенные правила условий использования цифрового объекта. Под условиями здесь понимаются интеллектуальные, авторские права, разрешения, ограничения использования. Пользователь должен быть хорошо информирован в отношении прав на использование материалов цифровой коллекции или конкретного объекта. Должны быть отражены также технические требования, требования к программному обеспечению просмотра, чтения данных;

д) качественные метаданные поддерживают долгосрочное управление объектами коллекций. Административные данные помогают управлять ресурсами. Они могут включать данные по созданию ресурсов и доступу к ним, архивированию контента (рисунок). При создании институционального репозитория пользователи делятся на создателей контента, редакторов и администраторов репозитория. Технические данные отражают форму фиксации информации, физический формат, размер файла, контрольную сумму, частотные характеристики и т. п.;

е) записи качественных метаданных сами являются объектами и должны соответствовать качеству объектов коллекции, включая авторитетность,

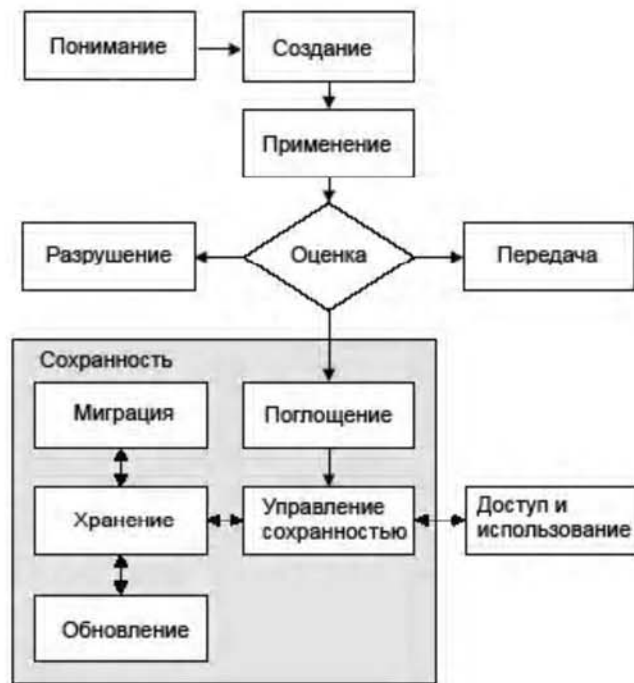


Рис. Жизненный цикл информации

подлинность, архивируемость, постоянство и иную уникальную идентификацию.

Научные коллекции ИР весьма разнообразны, динамичны, специфичны, отражают многообразие сфер исследований и особенностей исследовательской деятельности, а также различия в методологии и «технологиях» исследований в разных областях науки, дивергенцию круга пользователей.

При предложенном подходе к организации ИР создаются как персональные коллекции отдельных ученых и коллективов исследовательских лабораторий (в рамках институциональных репозитариев НИИ), так и коллекции, имеющие национальную или

мировую значимость и предоставляемые для глобального доступа на основе тематических ИР [16–20].

Персональная коллекция может со временем приобрести высокий уровень значимости для многих ученых, и в этой связи потребуются обеспечить к ней более широкий доступ. В отличие от ИР других видов коллекций, в научных коллекциях могут быть представлены сведения не о реальных процессах и явлениях, а гипотетические данные или данные компьютерных экспериментов с гипотетическими исследовательскими моделями, порой даже альтернативные (противоречивые) сведения об исследуемых объектах, явлениях или процессах. В таких ситуациях, конечно, неправомерно вести речь об интегральной целостности данных в коллекции. Для научных коллекций ИР не является необычной неполнота и/или нечеткость представляемых ими сведений [12].

Коллекции в систематизирующих научных дисциплинах (ботаника, зоология, минералогия и др.) основываются на различных классификаторах, сама разработка которых является существенным элементом проводимого научного исследования. Динамика таких коллекций, как правило, является односторонней – они лишь пополняются. В областях науки, где отводится важное место компьютерному моделированию (некоторые разделы математики, механики, физики, биологии, экономико-математические исследования и др.), организация коллекций должна быть приспособлена для использования их ИР в различных модельных компьютерных экспериментах и предоставлять пользователю данные из коллекций в форме, удобной для интерпретации исходных данных и результатов исследований [12].

Коллекции ИР в некоторых областях наук о Земле (в геофизике, океанологии, физике атмосферы [14]), в экономических исследованиях должны поддерживать пространственные и/или темпоральные свойства данных. В космических исследованиях коллекции хранятся огромные объемы данных (в т. ч. радиотелеметрии). В таких областях науки, как география, экология, демография, региональная экономика, значительную часть ИР коллекций составляют картографические данные на основе ГИС-технологий. Основными видами ИР в химических коллекциях являются графически представленные структурные формулы соединений, спектрограммы, текстовые описания качественных свойств веществ и их количественные характеристики [12].

Модель OAIS (Open Archival Information System) оказалась полезной в качестве модели высокого уровня для структурирования

долговременного хранилища организации (института) [17–18]. Цель хранилища в контексте НИИ – сбор и распространение результатов исследований учреждения, научных работ, диссертаций, презентаций и других оцифрованных материалов. Модель OAIS может быть использована для конфигурирования хранилища одновременно для нескольких учреждений путем дезагрегирования служб хранилища [18].

Литература

1. *Алиева-Кенгерли А. И.* О создании Национальной Электронной библиотеки в НАНА / А. И. Алиева-Кенгерли, С. В. Кондратьев // Науч. тр. ; ЦНБ НАНА. – Баку, 2004. – Вып. 5. – С. 7–17.

2. *Алиева-Кенгерли А. И.* К вопросу о создании Национальной электронной библиотеки в НАН Азербайджана / А. И. Алиева-Кенгерли, С. В. Кондратьев // Библиотеки национальных академий наук: проблемы функционирования, тенденции развития : сб. – К. : Нац. б-ка Украины им. В. И. Вернадского, 2006. – Вып. 4. – С. 157–164.

3. *Alguliyev R.* About Functions and Content of the National Electronic Library / R. Alguliyev, S. Kondratyev // Proceedings of the IV International Conference «Internet – Education – Science – 2004». – Vinnytsia : UNIVERSUM-Vinnytsia. – 2004. – Vol. 1. – P. 138–141.

4. *NISO.* Understanding metadata [Electronic resource]. – Bethesda, MD : NISO Press, [2004]. – Mode of access: <http://www.niso.org/publications/press/UnderstandingMetadata.pdf>. – Title from the screen.

5. Consultative Committee for space data systems. Reference model for an open archival information system (OAIS) [Electronic resource]. CCSDS 650.0-B-1. Blue Book. Iss. 1. – 2002, January. – Mode of access: <http://www.classic.ccsds.org/documents/pdf/CCSDS-650.0-B-1.pdf>. – Title from the screen; *ISO 14721:2003.* Space data and information transfer systems – Open archival information system – Reference model [Electronic resource]. – Mode of access: http://www.iso.org/iso/catalogue_detail.htm?Cnumber=24683. – Title from the screen.

6. *Powel A.* RSLP collection description [Electronic resource] / Andy Powell, Michael Heaney, Lorcan Dempsey // D-Lib Magazine. – 2000. – Vol. 6, № 9. – Mode of access: <http://www.dlib.org/dlib/september00/powell/09powell.html>. – Title from the screen.

7. *Lynch C.* Metadata harvesting and the open archives initiative [Electronic resource] / C. Lynch // ARL Bimonthly Report. – 2001. – № 217. – P. 1–9. – Mode of access: <http://www.arl.org/newsltr/217/pmh.html>. – Title from the screen.

8. OAI-rights white paper [Electronic resource] / Lagoze C., Nelson M., Van de Sompel H., Warner S. – 2003. – Mode of access: <http://www.openarchives.org/documents/OAIRightsWhitePaper>. – Title from the screen. – Retrieved November 20, 2004.

9. *Hussein Suleman.* A framework for building open digital libraries [Electronic resource] / Hussein Suleman, Edward A. Fox // D-Lib Magazine. – 2001. – Vol. 7, № 12. – Mode of access: <http://www.dlib.org/dlib/december01/12contents.html>. – Title from the screen.

10. RLG/OCLC Working Group on digital archive attributes. Trusted digital repositories: attributes and responsibilities [Electronic resource] : report. – Mountain View, CA : Research Libraries Group, 2002. – 62 p. – Mode of access: <http://www.rlg.org/longterm/repositories.pdf>. – Title from the screen.

11. The Open Archives Initiative (OAI). Protocol for metadata harvesting and guidelines for implementers [Electronic resource]. – Mode of access: <http://www.openarchives.org>. – Title from the screen.

12. *Когаловский М. П.* Научные коллекции информационных ресурсов в электронных библиотеках [Электронный ресурс] / М. П. Когаловский // Труды I Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, коллекции». – СПб. : Изд. Санкт-Петербург. ун-та, 1999. – Режим доступа: <http://www.cemi.rssi.ru/mei/articles/dlib.htm>. – Загл. с экрана.

13. *Van de Sompel H.* (2006). Access interfaces for open archival information systems based on the OAI-PMH and the OpenURL Framework for context-sensitive services [Electronic resource] / H. Van de Sompel. – Mode of access: <http://arxiv.org/abs/cs.DL/0509090>. – Title from the screen. – Retrieved on June 9, 2007 from Cornell University Library arXiv e-prints service.

14. Applying the OAIS standard to CCLRC's British Atmospheric Data Centre and the Atlas Petabyte Storage Service : paper presented at UK e-Science Programme All Hands Meeting (AHM2004), Nottingham, UK / D. Corney et al. – Mode of access: <http://epubs.cclrc.ac.uk/bitstream/487/156.pdf>. – Title from the screen. – Retrieved on June 9, 2007.

15. EPrints: Supporting Open Access. (n.d.) [Electronic resource]. – Mode of access: <http://www.eprints.org>. – Title from the screen. – Retrieved on December 29, 2009

16. Digital preservation service provider models for institutional repositories: towards distributed services [Electronic resource] / S. Hitchcock, T. Brody, J. M. N. Hey, L. Carr. – Mode of access: <http://preserv.eprints.org/papers/models/models-paper.html>. – Title from the screen. – Retrieved on June 9, 2007.

17. *Knight G.* A problem shared...Modelling OAIS compliance for distributed services [Electronic resource] / G. Knight. – Mode of access: http://www.sherpadp.org.uk/presentations/dcc_presentation2006.pdf. – Title from the screen.

18. *Knight G.* Modelling OAIS Compliance for Disaggregated Preservation Services [Electronic resource] / G. Knight, M. Hedges // The International journal of digital curation. – 2007. – Iss. 1, vol. 2. – Mode of access: <http://www.ijdc.net/index.php/ijdc/article/viewFile/25/14>. – Title from the screen.

19. *Thibodeau K.* What constitutes success in a digital repository? [Electronic resource] / K. Thibodeau. – Mode of access: <http://sil.unc.edu/events/2006jcdl/digitalcuration/Thibodeau-JCDLWorkshop2006.pdf>. – Title from the screen. – Retrieved on June 9, 2007.

20. *UKOLN.* DepositAPI Report, 2006, February 27 [Electronic resource]. – Mode of access: http://www.ukoln.ac.uk/repositories/digirep/index/DepositAPI_report. – Title from the screen. – Retrieved on June 9, 2007 27f7127c24218e1