

Структура и фонды электронной библиотеки «Научное наследие России»

*Забровская И.Е., Кириллов С.А., Шабанов М.Б.
(Межведомственный суперкомпьютерный центр РАН)*

Архитектура проекта. Полный цикл создания электронной книги. Формы взаимодействия.

Электронная библиотека «Научное наследие России» [1] представляет собой распределенную информационную систему, которая объединяет множество дистанционно удаленных друг от друга участников, работающих на разном оборудовании, иногда по разным технологиям, но по общим правилам, которые определены централизованно для всех участников проекта.

Базовой единицей такой распределенной информационной системы являются типовые центры сканирования и обработки, размещенные в различных городах России. Основная задача этих центров — выполнять с помощью локальных ресурсов и двух централизованных серверов обработки данных полный цикл создания электронной книги.

Полный цикл создания электронной книги выполняется в режиме online на двух технологических серверах и завершается публикацией электронного издания на сайте электронной библиотеки «Научное наследие России» <http://e-heritage.ru>. Он состоит из следующих этапов:

1. Ввод набора метаданных, описывающих издание. В состав метаданных входят имена авторов, их биографии, год издания книги, издательство и т.д. [2]. Эти данные заносятся в специализированную базу данных, которая расположена на технологическом сервере электронной библиотеки «Научное наследие России» <http://meta.e-heritage.ru>.
2. Сканирование. Этап сканирования предназначен для перевода изображений страниц публикаций в цифровую форму.

3. Обработка отсканированных изображений. Этап необходим для повышения качества отображения страниц публикации, устранения дефектов изображения и преобразования цифровых данных в наиболее удобный для дальнейшей работы вид.
4. Верстка электронного издания. На этом этапе из отдельных изображений страниц осуществляется верстка электронной книги в режиме on-line на технологическом сервере верстки электронного издания <http://media.e-heritage.ru>.
5. Публикация электронного издания на сайте электронной библиотеки «Научное наследие России» <http://e-heritage.ru>.

Выполнить полный цикл создания электронной книги — идеальная задача, к которой должны стремиться все участники проекта. Однако часто возникает ситуация, когда участники располагают фондами, необходимыми для электронной библиотеки, но по разным причинам, например, из-за отсутствия оборудования, сотрудников и т.п. не могут выполнить полный цикл работ по созданию электронной публикации. В таком случае предусмотрен сокращенный цикл, по которому выполняется только часть работ, а завершение работ передается другим участникам проекта. Например, организация поставляет только биографии ученых и ссылки на внешние или внутренние источники (Архив РАН); организация осуществляет полный ввод метаданных, сканирование и графическую обработку публикации (ЦНБ УрО РАН); организация осуществляет полный ввод метаданных и сканирование публикации (Дарвиновский музей Москвы) и т.д. Первоначально, завершение всех работ централизованно принимал на себя отдел Информационных ресурсов и систем МСЦ РАН, который и сейчас контролирует весь процесс создания электронных книг от подачи заявок на сканирование до публикации электронной книги на Интернет-портале. Но в ходе развития проекта возникают различные формы межведомственного взаимодействия. Например, Государственный геологический музей им. В.И. Вернадского РАН осуществляет полный ввод метаданных и передает книги в центр сканирования и обработки БЕН РАН, который завершает цикл работ по со-

зданию электронной публикации. Другие формы горизонтальной кооперации реализованы в Санкт-Петербурге.

Участники проекта. Состав. География проекта. Ведомственная принадлежность.

На сегодняшний день контент для электронной библиотеки «Научное наследие России» поставляют 17 научных организаций, относящихся к 4-м ведомствам – 14 организаций представляют РАН, по одной — Российскую академию образования, Российскую академию сельскохозяйственных наук, один поставщик — муниципальное учреждение Правительства Москвы.

Большая часть организаций расположены в Москве и Московской области — 10 организаций, 5 организаций в Санкт-Петербурге, 1 организация в Ярославской области (пос. Борок), 1 организация в Свердловской области (г. Екатеринбург). Ниже приведен список участников проекта.

Москва и московская область.

1. Федеральное государственное бюджетное учреждение науки Межведомственный суперкомпьютерный центр Российской академии наук;
2. Федеральное государственное бюджетное учреждение науки Библиотека по естественным наукам Российской академии наук (БЕН РАН) и три её отдела: в Математическом институте им. В.А. Стеклова РАН, в Физическом институте им. П.Н. Лебедева РАН и в Пушинском научном центре РАН;
3. Федеральное государственное бюджетное учреждение науки Институт научной информации по общественным наукам Российской академии наук (ИНИОН);
4. Федеральное государственное бюджетное учреждение науки Институт славяноведения Российской академии наук;
5. Государственный геологический музей им. В.И. Вернадского Российской академии наук;

6. Федеральное государственное бюджетное учреждение науки Институт этнологии и антропологии им. Н.Н. Миклухо-Маклая Российской академии наук;
7. Федеральное государственное бюджетное учреждение науки Архив Российской академии наук.
8. Государственный Дарвиновский музей Правительства Москвы;
9. Федеральное государственное научное учреждение «Научная педагогическая библиотека имени К.Д. Ушинского» Российской академии образования;
10. Государственное научное учреждение Центральная научная сельскохозяйственная библиотека Российской академии сельскохозяйственных наук.

Санкт-Петербург.

1. Федеральное государственное бюджетное учреждение науки Межведомственный суперкомпьютерный центр, Санкт-Петербургский филиал МСЦ РАН;
2. Федеральное государственное бюджетное учреждение науки Библиотека Российской академии наук (БАН);
3. Федеральное государственное бюджетное учреждение науки Архив РАН, Санкт-Петербургский филиал;
4. Федеральное государственное бюджетное учреждение науки Институт цитологии РАН;
5. Федеральное государственное бюджетное учреждение науки Институт русской литературы РАН (Пушкинский Дом).

Ярославская область.

1. Отдел страхового документального фонда Архива РАН (ОСДФ АРАН), Ярославская область, Некоузский район, пос. Борок.

Свердловская область.

1. Федеральное государственное бюджетное учреждение науки Центральная научная библиотека Уральского отделения РАН (Екатеринбург).

В момент основания электронной библиотеки «Научное наследие России» в 2006 году проект насчитывал всего 6 участников, количество участников постоянно увеличивается. Это определяется, с нашей точки зрения, следующими основными факторами: во-первых, растет интерес и желание участвовать в проекте со стороны научных организаций России; у них есть желание рассказать об ученых, работавших в этих организациях, есть интерес к новым технологиям, которые предлагает электронная библиотека «Научное наследие России»; второй фактор обусловлен потребностями и ростом самой электронной библиотеки «Научное наследие России». Проект ощущает нехватку поставок контента по некоторым направлениям естественнонаучных знаний, например, астрономии, некоторым областям химии, физики и биологии. Это заставляет организаторов ЭБ привлекать новые научные организации, обладающие соответствующими фондами научных трудов и архивных материалов.

Характеристика библиотечных фондов. Состав библиотеки.

Информация по авторам (персоналиям).

Главным приоритетом для информации, включаемой в ЭБ «Научное наследие России», является персона. Другими словами, сначала мы выбираем ученого, создаем биографическую справку и затем помещаем в ЭБ «Научное наследие России» его научные труды, архивные данные, сведения о посвящённых ему музейных коллекциях, а также другую информацию. Поэтому основной особенностью нашего проекта является *требование создавать развернутую биографическую справку по всем, без исключения, авторам, произведения которых включены в ЭБ «Научное наследие России».*

Одновременно с вводом в технологическую базу данных сведений об учёном ведется работа по поиску относящихся к нему внешних источников информации, ссылки на которые вводятся в ЭБ «Научное наследие России».

В мае 2013 года поля «Внешние ссылки» и «Музейная информация» содержали данные, относящиеся к 186 учёным.

Похожим образом построена работа с Архивом РАН. Кроме качественных биографий и фотографий учёных, которые введены непосредственно в технологическую базу данных ЭБ «Научное наследие России», Архив РАН предоставил ссылки на свой сайт (поле «Архивная информация» заполнено для 322 учёных).

Таблица 1. «Область знаний» дает представление о том, в каких областях знаний специализировались учёные, представленные в нашей электронной библиотеке.

Область знаний	%
Сельскохозяйственные науки	0,81%
Астрономия	1,27%
Политические науки	1,73%
Механика	2,15%
Медицина	2,32%
Техника	3,55%
Государство и право, юридические науки	4,53%
Математика	4,76%
Химия	5,11%
Философия	6,32%
Физика	7,72%
Общественные науки	8,22%
Науки о Земле	8,78%
Биология	8,97%
Религия, атеизм	9,28%
Другие науки	24,47%

Всего авторов в технологической базе данных 4745, из них загружено на сайт ЭБ 2804.

Информация по количеству книг.

С точки зрения текущего производственного процесса (готовности) фонды электронной библиотеки «Научное наследие России» состоят из двух частей: опубликованных и находящихся в работе. На 20 мая 2013 года были введены в систему метаданные на 16871 изданий, из них прошли полную проверку и загружены на сайт 9722.

Распределение публикаций по языкам изданий.

Основная масса публикаций издана на русском языке. Кроме русского языка в Библиотеке широко представлены книги на латинском языке, как правило, это российские академические издания XVIII века, и на французском языке, который являлся языком международного научного общения в XIX веке. Издания на английском и немецком языках равномерно представляют XVIII, XIX и XX века. Распределение публикаций по языкам издания представлено в табл. 2.

Таблица 2.

<i>Язык</i>	<i>Кол-во изданий в базе технологического сервера метаданных</i>	<i>Кол-во изданий на сайте ЭБ</i>
русский	13928	7642
французский	762	565
немецкий	436	311
английский	347	154
латинский	671	544
прочие	249	75

Распределение публикаций по времени издания.

В электронной библиотеке «Научное наследие России» широко представлены все периоды развития научной мысли в России до середины 20 века (табл. 3). Резкое сокращение работ, со второй пол. XX века по настоящее время обусловлено, прежде всего, законом об авторском праве, которым руководствуются поставщики

контента. Все отраженные в ЭБ издания этого периода представлены институтами, поставляющими работы своих выдающихся ученых с согласия автора или наследников.

Таблица 3.

<i>Период</i>	<i>Кол-во изданий в технологической базе данных</i>	<i>Кол-во изданий на сайте ЭБ</i>
1618-1699	10	0
1700-1799	1405	971
1800-1899	4196	2033
1900-1949	9158	5709
1949-1999	1662	890
2000-2012	227	113

Распределение книг по типу изданий.

Таблица распределения публикаций по типу издания.

<i>Тип издания</i>	<i>Кол-во изданий в технологической базе данных</i>
монография	6875
том многотомника	2446
выпуск сериального издания	705
сборник статей	601
доклад	87
сборник трудов конференций	48
письмо	20
рукопись	12
прочие виды публикаций	2609
статья из журнала, сборника трудов	2371
отдельный оттиск	632

Распределение публикаций по научной тематике (рубрикам ГРНТИ).

Для систематизации и поиска изданий в электронном каталоге ЭБ «Научное наследие России» применяется Государственный рубрикатор научно-технической информации (ГРНТИ). При вводе публикации на технологи-

ческий сервер электронного каталога ЭБ «Научное наследие России» <http://meta.e-heritage.ru> используются все три уровня иерархии рубрикатора, где: первый уровень — это тематический раздел, который в свою очередь делится на классы (второй уровень), а каждый класс на подклассы (третий уровень). Затем, при публикации электронного издания на сайте ЭБ «Научное наследие России» (<http://e-heritage.ru>), иерархия рубрикатора ГРНТИ обобщается до двух уровней.

При анализе таблицы следует учитывать, что одно издание может иметь несколько значений рубрик ГРНТИ из разных тематических разделов (Пример А.), так и в рамках одного тематического раздела (Пример Б).

Пример: А.

Жиров, Н.Ф. Люминофоры : (Светящиеся твердые составы)/ Н.Ф. Жиров ; Под ред. С.И. Вавилова и Б.Я. Свешникова. — М. : Гос. изд-во оборон. пром-сти, 1940. Издание определяется следующими рубриками ГРНТИ: 29.31.23; 61.69.39; 61.31.53; 47.09.43; 38.41.19

Пример: Б.

Лепехин Иван Иванович. Дневные записки путешествия доктора Имп. академии наук адъюнкта Ивана Лепехина по разным провинциям Российского государства в 1768 и 1769 году : Ч.2. — 1770. — СПб. : Имп. акад. наук, 1802.

Издание определяется следующими рубриками ГРНТИ: 34.29.00; 34.29.35; 34.33.00

Особенно тщательно в нашем проекте (две и более рубрики ГРНТИ на одно издание) рубрицированы издания по истории, философии и религии, что оказывает влияние на цифры приведенной ниже таблицы.

Таблица распределения публикаций по рубрикам ГРНТИ.

<i>Рубрика ГРНТИ</i>	<i>Кол-во рубрик в базе технологического сервера мета-данных в%</i>	<i>Кол-во рубрик на сайте ЭБ в%</i>
ДРУГИЕ	5,15	9,08
РЕЛИГИЯ. АТЕИЗМ	12,56	14,11

ФИЛОСОФИЯ	8,04	12,13
ФИЗИКА	7,24	6,93
ЛИТЕРАТУРА. ЛИТЕРАТУРОВЕДЕНИЕ	9,23	6,91
МАТЕМАТИКА	5,76	6,46
ГЕОЛОГИЯ	5,88	5,81
БИОЛОГИЯ	4,99	5,76
ХИМИЯ	4,64	4,84
ИСТОРИЯ. ИСТОРИЧЕСКИЕ НАУКИ	9,43	4,78
ЯЗЫКОЗНАНИЕ	4,43	3,44
СЕЛЬСКОЕ И ЛЕСНОЕ ХОЗЯЙСТВО	2,52	3,15
ЭКОНОМИКА. ЭКОНОМИЧЕСКИЕ НАУКИ	6,47	2,84
МЕХАНИКА	2,19	2,69
ГОСУДАРСТВО И ПРАВО. ЮРИДИЧЕСКИЕ НАУКИ	4,02	2,45
ПОЛИТИКА. ПОЛИТИЧЕСКИЕ НАУКИ	1,31	1,99
ГЕОГРАФИЯ	1,03	1,8
АСТРОНОМИЯ	1,39	1,64
КУЛЬТУРА. КУЛЬТУРОЛОГИЯ	1,04	1,46
ГЕОФИЗИКА	1,09	1,37
НАУКОВЕДЕНИЕ	1,48	1,25

Производственный ритм проекта.

Производственный ритм нашего проекта задают наши поставщики. Как видим по таблице, наиболее продуктивны периоды февраль-май и октябрь-декабрь. Ежегодно количество поставляемых книг увеличивалось на 6-8% по сравнению с предыдущим годом. Первая половина 2013 года была очень продуктивной. В январе-мае 2013 года количество поставленных сверстанных изданий увеличилось на 37% по сравнению с предыдущим периодом.

Таблица поставок сверстанных изданий.

Месяц/Год	2010	2011	2012	2013
Январь	78	143	119	229
Февраль	152	303	215	230
Март	235	160	237	323
Апрель	241	189	169	322

Май	246	156	199	209
Июнь	153	169	227	
Июль	296	145	154	
Август	54	105	181	
Сентябрь	111	173	222	
Октябрь	119	385	237	
Ноябрь	412	250	433	
Декабрь	240	296	284	
Итого	2337	2474	2677	1313

Вклад участников проекта в 2010-2013 году.

Как видно из таблицы, наибольшее количество изданий (33%), в период 2010-2013 гг., поставил ИНИОН РАН, сканирующий исключительно собственные фонды.

Поставки от БЕН РАН (22%) включают собственные фонды и фонды Государственного геологического музея им. В.И. Вернадского РАН.

Библиотеки СПб (17%) — объединение учреждений Санкт-Петербурга: Института цитологии РАН, СПб филиала Архива РАН и Санкт-Петербургского филиала МСЦ РАН, который активно сотрудничает с библиотекой Физико-технический института имени А.Ф. Иоффе.

МСЦ РАН (16%) оцифровывает и верстает издания, полученные из разных источников: это книги и документы, которые приносят авторы или наследники, издания переданные нам по обмену, но основная масса (более 80%) изданий, сверстанных МСЦ РАН, поступила из фондов Библиотеки Российской академии наук (БАН) (Санкт-Петербург).

Таблица поставленных сверстанных изданий в 2010-2013 году участниками проекта.

<i>Наименование участника</i>	<i>Количество поставленных книг</i>	<i>Доля в%</i>
ИНИОН	2896	33
БЕН РАН	1907	22
Библиотеки СПб	1524	17
МСЦ РАН	1398	16
ИРЛИ РАН	462	5

МИАН им. В.А. Стеклова	239	3
Другие участники	375	4

Заключение.

Проект показывает хорошую динамику роста. Поставки контента постоянно увеличиваются. В этом году прирост поставленных сверстанных изданий составит не менее 10-15%. Этот результат достигается как за счет совершенствования технологий сканирования и разработки новых алгоритмов работы сотрудников, так и вследствие развития взаимодействия и горизонтальной кооперации между участниками проекта.

Дальнейшее развитие проекта требует изменений в техническом, технологическом, организационном плане.

В связи с постоянно возрастающими объемами материалов необходимо наращивать мощность систем, участвующих в процессе производства электронных книг. В частности, провести плановое обновление парка рабочих станций, решать вопросы сервисного обслуживания оборудования. Переход к полноцветному сканированию предъявляет новые требования к программному обеспечению проекта. В первую очередь необходимо модернизировать систему верстки электронных изданий (программу NNMetaDis).

В организационном плане необходимо, не сокращая объём поставок материалов, относящихся к общественным наукам, увеличить долю естественнонаучных трудов. Для этого, например, привлечь к проекту новые научные организации, обладающие фондами научных трудов соответствующей тематики.

Литература

1. Калёнов Н.Е., Савин Г.И., Серебряков В.А., Сотников А.Н. Принципы построения и формирования электронной библиотеки «Научное наследие России» // Программные продукты и системы, 2012. — №4(100). — С. 30-40.
2. Калёнов Н.Е., Савин Г.И., Сотников А.Н. Электронная библиотека «Научное наследие России»: технология наполнения // Новые технологии в информационном обеспечении науки: Сборник научных трудов / Калёнов Н.Е. (ред). — М.: Научный мир, 2007. — С. 40-48.