

ЭЛЕКТРОННЫЕ АРХИВЫ ДЛИТЕЛЬНОГО СРОКА ЖИЗНИ: МОДЕРНИЗАЦИЯ И ИНТЕГРАЦИЯ

А. Г. Марчук¹ [0000-0001-8455-725X], С. Н. Трошков² [0000-0003-2952-9509],

И. А. Крайнева³ [0000-0002-0601-9795]

¹⁻³Институт систем информатики им. А.П. Ершова СО РАН

¹mag@iis.nsk, ²kamronis@xtech.ru, ³cora@iis.nsk

Аннотация

За период около двадцати лет в Институте систем информатики им. А.П. Ершова Сибирского отделения РАН (ИСИ СО РАН) были созданы информационные системы исторической направленности: Электронный архив академика А.П. Ершова, Фотоархив Сибирского отделения РАН, Архив газеты «Наука в Сибири», Открытый архив СО РАН и др. У каждого из этих ресурсов есть своя специфика, но в целом их контент базируется на общей социальной и территориальной основе научной и общественной деятельности СО АН СССР/РАН и Новосибирского Академгородка. В статье рассмотрены некоторые проблемы интеграции/дезинтеграции разрозненных электронных ресурсов на общую платформу на базе имеющихся и создаваемых инструментов.

Ключевые слова: интеграция электронных ресурсов, качественная информация, открытые архивы, междисциплинарность, история науки, Сибирское отделение РАН, проприетарное ПО, Semantic Web, Drupal.

ВВЕДЕНИЕ

С середины 1990-х гг. интернет и электронные ресурсы стремительно захватывают пространство деловой, общественной и научной активности России. Центры использования информационных технологий в историко-культурных, научно-теоретических, историко-биографических и прочих исследованиях помимо Москвы (чл.-корр. Л.И. Бородкин, д. и. н. И.М. Гарскова, д. и. н. Ю.Ю. Юмашева), действуют в Новосибирске (к. и. н. Ю.П. Холюшкин, д. ф.-м. н. А.Г. Марчук), Барнауле (д. и. н. В.Н. Владимиров), Томске (д. и. н. С.А. Некрылов), Ижевске (д. филолог. н. В.А. Баранов), Перми (д. и. н. С.П. Корниенко), Краснояр-

ске (И.А. Кижнер) и др. Государственные архивы, которые, как правило, являются основным местом паломничества исследователей, проводят оцифровку научно-справочного аппарата [1].

В работе пойдет речь о цикле жизни, возможностях и перспективах подобных ресурсов.

1. ИНФОРМАЦИОННЫЕ СИСТЕМЫ ДЛЯ ИССЛЕДОВАНИЙ В ГУМАНИТАРНЫХ НАУКАХ

Под информационными системами (ИС) мы понимаем совокупность технического, программного, организационного и финансового обеспечения, а также персонала, способного обеспечивать работоспособность этого комплекса и выполнение проекта. Минимальное количество персонала в таком проекте по опыту ИСИ СО РАН – около 10 человек: программисты, историки, информационные специалисты (операторы), переводчики. Специалистами в области применения информационных технологий в гуманитарной сфере из Пермского университета предложена спецификация историко-ориентированных систем как «особого класса систем, предназначенных для хранения, организации исторической информации, обеспечения доступа к ней и ее аналитической обработки в соответствии с потребностями исторических исследований и/или образования» [2]. Особый интерес вызывают системы, содержащие помимо исторической информации исследовательский инструментарий (поисковый, аналитический, распознавание текста, внутритекстовые гиперссылки и др.). Авторы выделяют два подхода к созданию ИС: источник-ориентированный, когда основой системы является массив одного источника, а его структура становится моделью системы, и проблемно-ориентированный, когда модель строится на основе рассматриваемой предметной области.

В соответствии с данной классификацией ИС, созданные в ИСИ СО РАН, являются источник-ориентированными: Фотоархив СО РАН [3] вмещает на своей платформе два разных источника – фотодокументы и газету «Наука в Сибири», материалы, связанные тематически и органически, поскольку многие фото выполнены фотокорреспондентами – сотрудниками редакции газеты. Электронный архив А.П. Ершова и Открытый архив СО РАН [4] помимо сканированных документов содержат сканы фотографий и печатных научных трудов. Мы считаем

наши ИС источник-ориентированными и по другой причине: по наличию образов подлинных документов (полнотекстовые ИС), транскрипция которых является дополнением, позволяющим знакомиться с трудночитаемыми текстами. Кроме того, наши ИС являются интернет-ориентированными: то, что системы названы электронными, не дает представления о степени их открытости. И в завершение данной темы отметим, что наши ИС – это системы, позволяющие осуществлять описание документов с удаленных рабочих мест – распределенные. Итак, наши ИС – источник-ориентированные, поскольку позволяют знакомиться с образами подлинных документов (полнотекстовые), интернет-ориентированные распределенные системы. Источник-ориентированный подход демонстрируют создатели таких ресурсов, как «Берестяные грамоты» [5], «Подвиг народа» [6], Архив академика В.И. Вернадского [7] и др.

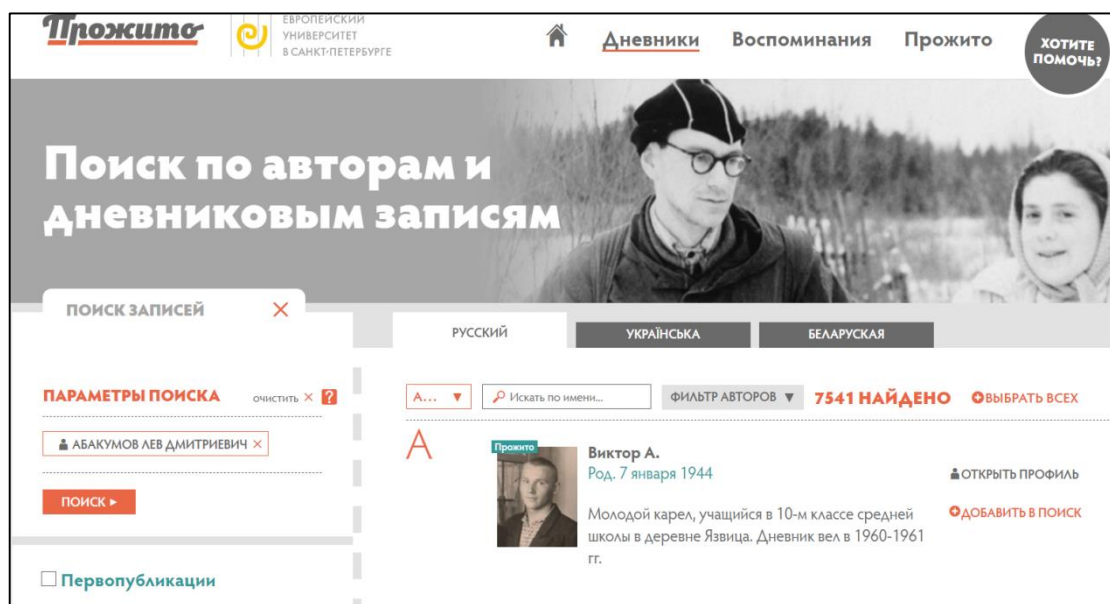


Рис. 1. Страница корпуса «Дневники» на сайте «Прожито»

Смешанный подход (проблемно-ориентированный и источник-ориентированный), на наш взгляд, демонстрируют те историко-ориентированные ИС, которые, помимо образов документов, таких как фотографии, карты-схемы, содержат преимущественно транскрипции рукописных документов. К таковым можно отнести, например, проекты «Бессмертный барак» [8], «Истмат» [9]. На «Истмате» для транскрибированных документов при-

ведены ссылки на архивные источники. Сайт «Прожито», на котором собраны личные дневники и воспоминания, содержит не только сканы артефактов, но и транскрипции уже опубликованных документов, а также ссылки на аналогичные источники в других ИС, в частности, в музейном Госкаталоге (Рис. 1) [10]. При консультационной поддержке ИСИ СО РАН выполнен проект «Электронный архив Ольги Михайловны Фрейденберг (1890–1955) [11]. О.М. Фрейденберг – филолог-классик, философ культуры, двоюродная сестра поэта Бориса Пастернака. В контент архива включено исследование, выполненное на его материалах.

Поиск

Поиск по описанию: Город: Дата начала: Ключевые слова:

Например: 11.06.2021

Дата окончания: Например: 11.06.2021

▼ Поиск по человеку

Автор: X

Адресат: X

Персона: X

Сортировать по: Порядок сортировки:

Документов найдено: 19.

[Письмо Д. Маккарти](#) 10.12.1968
содержащее шуточную песенку, посвященную Э. Хоару по случаю избрания его профессором Королевского университета

[Письмо Д. Маккарти -> А.П. Ершов](#) 14.06.1969
Рассказывает о своих делах после возвращения из Новосибирска. /Текст рукописный./

[Письмо Д. Маккарти -> А.П. Ершов](#) 13.11.1969
Рассказывает об использовании компьютера для написания писем и деловых бумаг, о своих успехах в прыжках с парашютом в Новосибирск.

[Письмо Д. Маккарти -> А.П. Ершов](#) 06.10.1971
Автор описывает свою поездку на Восток (Иран, Индия, Япония); рассказывает о проектах, с которыми имел возможность встретиться в Лондоне; если придет приглашение, приедет в Новосибирск в конце марта. /Текст рукописный./

[Письмо Д. Маккарти -> А.П. Ершов](#) 20.08.1972
Сообщает о своих планах быть в Ленинграде и Москве в начале сентября. Надеется увидеться с адресатом в след

Рис. 2. Пример поискового запроса с указанием параметров автор, тип документа, дата в Электронном архиве академика А.П. Ершова

2. ВКЛАД ИСИ СО РАН

Методологической основой использования информационных технологий в исследовательской работе гуманитариев является междисциплинарность. С течением времени сформирован инструментальный аспект коммуникативных процессов, которые осуществляются при посредстве ИТ, сетевой организации вычислительной техники и поддерживают доступность контента наследия, размещенного в Сети. Рабочий инструментарий привлечения ИТ в архивную работу – электронная историческая фактография, технология и метод – совокупность приемов, на основе которых создаются информационные системы (ИС) для размещения массива разнородных документов в Сети, систематизации их путем установления связей между сущностями, отраженными в документах. Документы могут цитироваться как электронной ссылкой, так и указанием на дело и лист в архиве (в частности, в архиве А.П. Ершова, в других архивах – если документы поступили из государственных хранилищ). Совершенствуется и подход к технике сканирования и визуализации артефактов: в последнее время мы стремимся к визуализации этапов графической обработки документов, в частности, фотографий. На страницу документа в электронном архиве помещаются не только отреставрированный вариант фотографии, но и исходное сканированное изображение, а также изображение обратной стороны фотографии без графической обработки (Рис. 3).

Очевидно, уже нет необходимости доказывать нужность и полезность открытых архивных источников, особенно в условиях значительного ограничения социальной мобильности. И чем полнее будут архивные ИС, тем более продуктивной станет исследовательская работа. В том случае, когда ИС заполняется достаточно репрезентативными и разнообразными источниками, она должна быть снабжена прецизионным поисковым инструментарием, а сами источники, помимо этого, должны быть систематизированы. Открытые архивы СО РАН в данный момент имеют несколько уровней систематизации: фонд, коллекции фонда, подколлекции коллекций (или группы и подгруппы, как в архиве А.П. Ершова). Поисковая же система довольно проста: поиск осуществляется по одному из параметров: типу документа, имени персоны, названию организации, ключевому слову. Поиск по типу документа не дает прецизионного результата, поскольку к

одному типу документов могут относиться артефакты разных фондообразователей. Инструментарий архива академика А.П. Ершова позволяет ограничивать поиск хронологическими рамками, именем персоны, типом документа, то, что и требуется в идеале (Рис. 2). Задача интеграции нескольких архивных ресурсов на одной платформе должна учитывать это обстоятельство, иначе массив станет слабо изучаемым. Чем больше поисковых возможностей будет у исследователя, тем эффективней он сможет использовать возможности открытых архивов.

[Семейные фотографии](#) » [Де-Метцы \(1890-1941\)](#) » Портрет Маргариты Де-Метц

Портрет Маргариты Де-Метц

Дата: 1904-09

Описание документа: Размер фотографии - 100 x 140 мм, общий размер с паспорту - 110 x 165 мм

1. Исходник.
2. После графической обработки.
3. Обратная сторона.



Zc 455_081 Zc 455_082 Zc 455_083

Отраженные персонажи: [Воскобойникова \(Де-Метц\) Маргарита Георгиевна](#)

Геоинформация: [Киев](#)

Источник поступления: [Шиплюк Екатерина Владимировна](#)

Рис. 3. Визуализация фотодокументов в Открытом архиве СО РАН

3. ЖИЗНЕННЫЙ ЦИКЛ И МОДЕРНИЗАЦИЯ ИНФОРМАЦИОННЫХ СИСТЕМ

Самой «древней» информационной системой, созданной и развивающейся в нашем коллективе, является информационная система «Библиотека А.П. Ершова» [13]. Она была создана для ЭВМ БЭСМ-6 еще в эпоху перфокарт. И ее создание, и дальнейшие переработки выполнялись силами энтузиастов. Эксплуатация, включая пополнение фондов и исправление ошибок и неточностей, выполнялась сотрудниками Отдела научно-технической информации (ОНТИ) ВЦ АН СССР/ ИСИ СО РАН. В 1990-х годах «Библиотека А.П. Ершова» была перенесе-

на на MS DOS с использованием СУБД FoxPro 2. С помощью этой системы велся каталог книг и журналов, учет читателей, формировались списки новых поступлений. В таком виде она прослужила ИСИ почти 30 лет.

Несмотря на актуальность и функциональность приложения, окружение, в котором оно было создано, устарело, и дальнейшая поддержка и развитие оказались невыполнимы. Кроме того, приложение не поддерживало структуризацию данных и словари. Все данные были строкового типа, заполнялись вручную библиотекарем, что привело к массе ошибочных и дублирующихся данных.

В 2018 был выполнен реинжиниринг приложения «Библиотека А.П. Ершова» <http://lib.iis.nsk.su/> на основе свободно распространяемой веб-платформы Drupal. В ходе работ была выполнена полуавтоматическая коррекция ошибок в именах авторов и названиях фондов, осуществлены миграция данных с сохранением модели данных и имплементация удобных современных интерфейсов [12].

Первым интернет-ориентированным электронным архивом, созданным и поддерживаемым в ИСИ СО РАН, был Архив академика А.П. Ершова <http://ershov.iis.nsk.su> [13]. Он был создан по классической схеме веб-приложения, построенного на реляционной базе данных, имеющего публичный интерфейс (frontend) и интерфейс редактирования (backend). Был проделан значительный объем работ по сканированию и описанию документов, содержащихся в более чем 500 папках «бумажного» архива, хранимого в ИСИ. Технология структуризации внесения, редактирования информации, представления документов и дополнительной информации пользователям оказалась удачной и в дальнейшем. В архив были погружены дополнительные сегменты, уже не связанные напрямую с папками, сформированными А.П. Ершовым. Это архивы ВНТК «Старт», ИСИ СО РАН, включая конференции PSI, а также архив члена-корреспондента АН СССР Святослава Сергеевича Лаврова.

Проект был разработан в 2000 году при поддержке Microsoft Research с использованием технологий Майкрософт и столкнулся с одной из частых проблем для приложений с использованием проприетарного ПО. Спустя 15 лет после разработки проекта актуальность его сохранилась, но дальнейшая поддержка и разработка были затруднены в связи с истечением лицензий на проприе-

тарное ПО. Поэтому в 2016 году было принято решение о миграции приложения на свободно распространяемое ПО. Важным условием миграции было сохранение исходной модели данных архива, представляющей историческую ценность как одной из первых моделей данных для электронных архивов. Миграция на свободно распространяемую веб-платформу Drupal прошла успешно, помимо переноса данных был усовершенствован интерфейс пользователя и библиотекаря, была реализована поддержка массовой загрузки изображений.

Для проекта Фотоархива СО РАН (Фотолетопись) <http://soran1957.ru> была подготовлена технология, основанная на Semantic Web [14]. Также была сформирована онтология – система структуризации данных, в дальнейшем получившая оформление в виде онтологии неспецифических сущностей [14]. Новыми проблемами, преодолеваемыми в проекте, были не только проблемы нового подхода к структуризации, но и обработка и представление фото- и видеоматериалов, решение задач модульности информации и защиты данных от случайных и злонамеренных искажений. Как и многие другие архивы, фотоархив постоянно пополняется новыми материалами. Это делается с помощью группы backend-технологий и интерфейсов.

Поскольку количество заявок на хранение архивов в цифровом формате стало расти, был сформирован подход и создана технология мультиархивной системы Открытого архива Сибирского отделения СО АН СССР/РАН (<http://odasib.ru>). Основа технологических решений была прежней, изменились интерфейсы, изменилась в сторону детализации структура описания документов. Например, документы теперь рассматриваются как состоящие из частей (разделы, страницы, сканы страниц), авторство расширено до разделения автор-получатель. Были созданы удобные средства для информационных операторов, которые ведут обработку больших объемов сканированных страниц. Отдельные архивы были квалифицированы как «фонды», это соответствует архивной терминологии. В настоящее время в Открытый архив погружено 25 фондов, ведется работа над еще несколькими. На Рис. 4 изображена схема работы с документами разработанных электронных архивов.

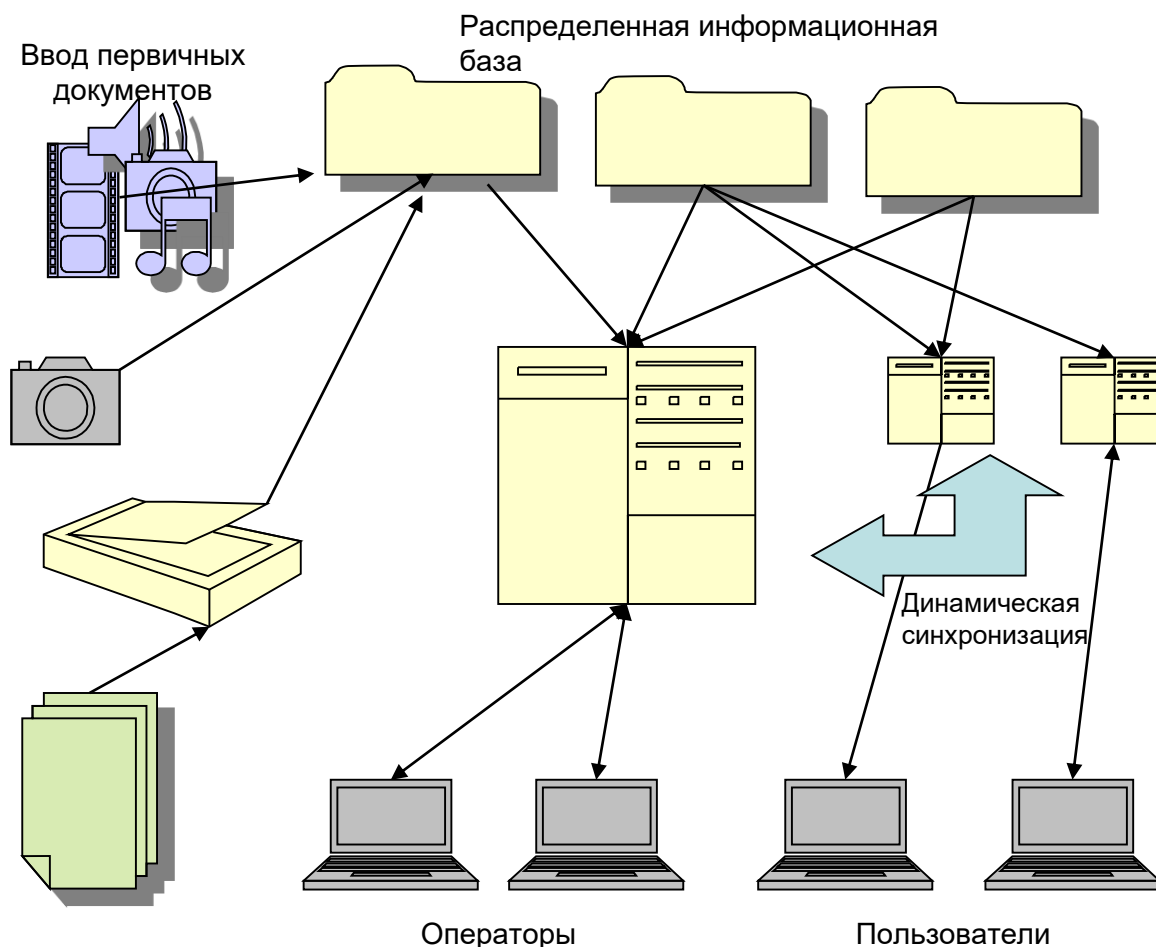


Рис. 4. Схема комплекса работы с электронными документными архивами

Еще одна проблема – физическая сохранность данных. Магнитные носители деградируют, могут быть повреждены в результате физического износа и внешних воздействий. Пока мы не сталкивались с серьезными вирусными заражениями, но это также должно быть учтено. Нельзя сказать, что сохранность данных сегодня обеспечивается нами на современном надежном уровне – принятые решения громоздки и требуют затрат. Основой обеспечения надежности хранения и функционирования электронных архивов является серверный пул машин и устройств, предоставляющий виртуальные машины для разработчиков и эксплуататоров, созданный, развивающийся и функционирующий в ИСИ СО РАН. Собственно, здесь уместно было бы использование специализированных решений или облачной инфраструктуры. Однако пока мы хотим полностью контролировать наши данные, соблюдая необходимую гибкость в решениях. В слу-

чае с модернизацией электронного архива А.П. Ершова описанное решение также не выглядит решающим все проблемы обеспечения долгого срока эксплуатации. Те эксперименты, которые будут описаны ниже, являются, в том числе, попыткой определения направления технологической эволюции архива.

Настоящим полигоном новых решений явился электронный архив Летних школ юных программистов (ЛШЮП), ежегодно проводимых ИСИ для талантливых школьников. При создании данного архива был пройден путь технологических решений от базы данных в XML и интерфейсов, сформированных средствами XSLT (в то время Semantic Web еще был неизвестен) до вполне современных веб-приложений с RDF и OWL, ASP.NET и т. д. В проект внесли свою лепту школьники, обучавшиеся в ЛШЮП. Архив постоянно пополняется и используется (<http://mag.iis.nsk.su/syp>).

За прошедшие годы проводились различные эксперименты с архивами. Например, к 50-летию Механико-математического факультета НГУ были произведены сбор мультимедийных материалов и данных, их погружение в Фотоархив СО РАН, создан отдельный юбилейный сайт.

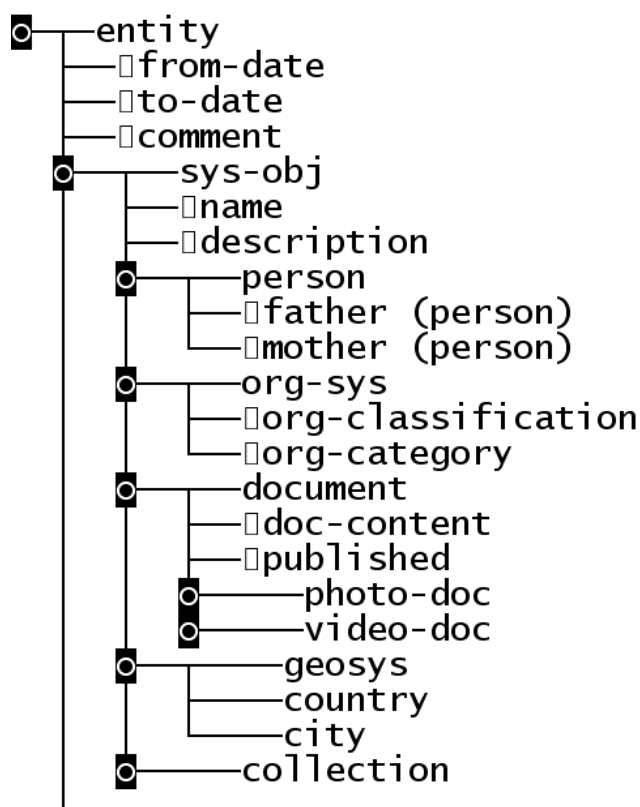
4. ИНТЕГРАЦИЯ: ЦЕЛИ И ЗАДАЧИ

Электронные архивы преследуют цели сбора и хранения данных и информации, а также предоставления удобного доступа к информации для специалистов и общественности. При этом сохранение данных в каком-то смысле является существенно более значимым, чем остальное. Основу архивов составляет тот или иной документный фонд. По информации, содержащейся в документах, архивы уникальны, но заметно пересекаются по таким параметрам, как авторы, адресаты, организации и события. Возникает вопрос о возможности полной, значительной или частичной интеграции разных архивов. Задача интеграции носит не только частный характер относительно упомянутых информационных собраний, но и представляет более общий методический интерес.

Чтобы понять возможный характер интеграции архивов, необходимо изучить интересы их пользователей. Поскольку доступ к архивной информации в электронных архивах осуществляется гораздо быстрее и удобнее, чем в классических «бумажных», круг пользователей существенно расширяется. Если ранее основными пользователями были ученые-историки, получающие в документных

архивах информацию для исследовательских целей, то теперь архивами легко могут пользоваться и те, кто задают вопросы «а кто это такой?», «а что это такое?», «а как это выглядит?». В этом числе могут быть пользователи, интересующиеся историей своей семьи, города, страны и т. д.

В наших архивах для структуризации документов и данных в основном используется вариант онтологии неспецифических сущностей [14]. На Рис. 5 изображена в виде дерева система классов и отношений (properties), составляющих основу этой онтологии.



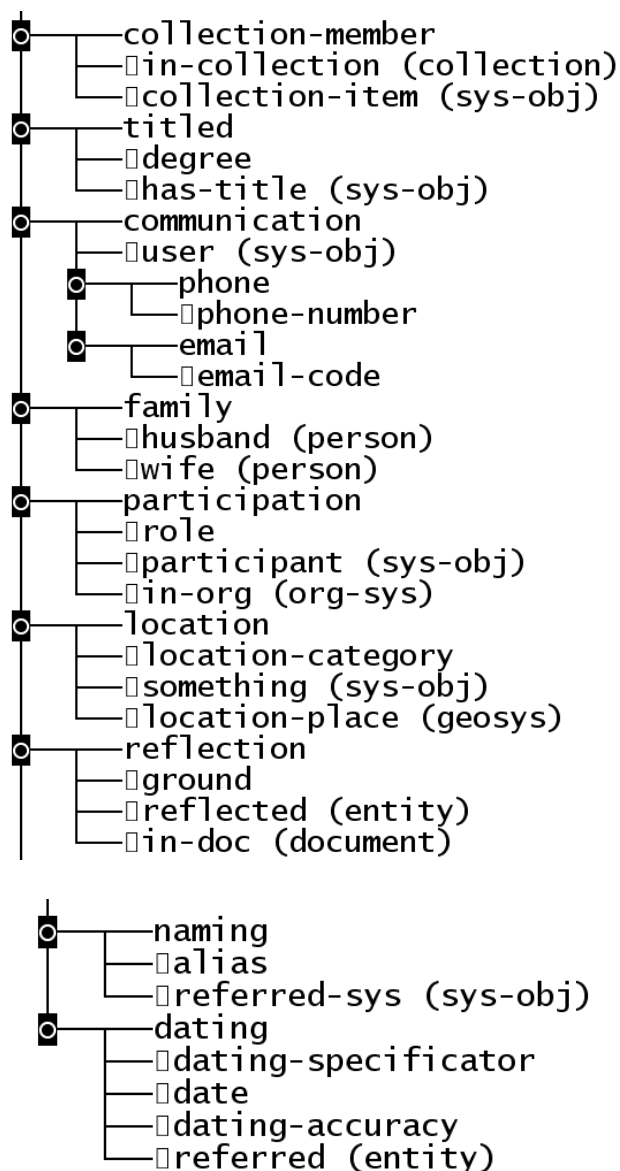


Рис. 5. Основные классы и отношения онтологии неспецифических сущностей

Отталкиваясь от документа той или иной коллекции, создатели архива фиксируют авторов документа, фигурирующие в документе персонажи, организации, географические пункты. Соответственно, через людей, организации и геоинформацию происходит дополнительная структуризация, которая может оказаться полезной для пользователя. Таким образом, интеграция может носить общий (максимальный объем запрашиваемой информации), частный (узкая проблема) и проблемно-ориентированный характер (специальный тематический запрос). Одновременно, как уже было показано, существует потребность в

разбиении единого архива на части, так сказать, «дезинтеграции» архива. Как ни странно, «ломать» в чем-то даже сложнее, чем «строить», т. е. объединять. Дело в том, что объединенный архив интегрирует мультимедиа, документы и базу данных. Несмотря на то, что предприняты значительные усилия для обеспечения модульности хранимых документов через технологию кассет и сегментирования базы данных через fog-сегменты, данные могут пересекаться и реально пересекаются. Технология разбиения так же, как и технология слияния, формируется и оформляется. Задача разбиения архива на части, как правило, включает еще и требование обеспечить дальнейшее развитие архива в части пополнения и изменения документного набора и базы данных. Копия части архива (подколлекции или коллекции, созданной по тематическому запросу) на конкретный момент представляется малополезной в случае «живого» архива, т. е. архива, который продолжает пополняться и редактироваться.

5. ИНТЕГРАЦИЯ КАК ЭКСПЕРИМЕНТ

Целью описанных экспериментов являлось выяснение возможности сопряжения и взаимного обогащения документных и информационных фондов, которые уже по отдельности находятся в эксплуатации. Также интерес представляет методика выполнения той или иной интеграции.

В первом развернутом эксперименте была предпринята попытка расширения информации, предоставляемой пользователю Электронного архива академика А.П. Ершова (ЭАЕ). В ходе эксперимента мы хотели дополнить информацию из ЭАЕ вставками из другой информационной системы. Для этого нужно было в информационную систему-донор добавить веб-сервис, позволяющий динамически отвечать на запросы от веб-приложения или веб-страницы и формировать ответ в виде HTML-разметки с запрашиваемой информацией справки. Аналогичный подход используется в некоторых других системах. В ответное сообщение помещаются фотография, информация об именах, датах жизни, основных степенях и наградах, профессиональном пути. В эксперименте информация берется из фотоархива СО РАН.

Что это дает? С точки зрения документного архива ЭАЕ является (почти) законченной композицией. Там есть информация о персонах, являющихся автора-

ми или адресатами документов. Есть организации и другие самостоятельные (не принадлежащие конкретному архиву) объекты. Информация в этих сущностях, зафиксированная на момент создания архива, может со временем изменяться. Кроме того, в ЭАЕ почти полностью отсутствует фотографический материал о персонажах. Мы предположили, что имеющаяся в ЭАЕ информация должна дополняться в результате интеграции с другими источниками. Пример такого эксперимента можно увидеть на Рис. 6.


документы	об архиве	о проекте	А.П. Ершов	in English
Юрий Леонидович Ершов Дата рождения: 01.05.1940 Город: Новосибирск		Ершов Юрий Леонидович 01-05-1940		
на: 01.01.1970 доктор физико-математических наук		1966 Доктор физико-математических наук 24-11-1970 Член-корреспондент по Отделению математики		
на: 19.09.1973 декан механико-математического факультета НГУ		07-12-1991 Академик по Секции математики, механики, информатики (математика)		
Связанные документы		1992 Лауреат премии им. А.И.Мальцева РАН 2003 Лауреат Государственной премии РФ		
Назад		1958 студент механико-математического		

Рис. 6. Информация о персоне из фотоархива СО РАН в Электронном архиве академика А.П. Ершова.

Участником другого эксперимента также являлся ЭАЕ. В этом случае рассматривалась не сама информационная система, а ее база данных. Задачей эксперимента было погружение ЭАЕ в Открытый архив СО РАН (ОА) в качестве отдельного фонда. Если переработать электронный архив на использование принципов и технологий Semantic Web, использовать онтологию неспецифических сущностей, с использованием кассет, лог-файлов базы данных, то подобное погружение осуществляется довольно просто: необходимо было сделать отождествление сущностей и осуществить привязку корня композиции к коллекции «Фонды». На данном этапе эксперимента экстракция базы данных из ЭАЕ в Открытый архив осуществлялась вручную, но этот процесс возможно автоматизировать в дальнейшем.

Эксперимент показал, что подобное погружение стороннего фонда успеш-

но осуществляется без значительных трудозатрат. С одной стороны, ЭАЕ в качестве фонда дополняется возможностями использования навигации и поиска, заданных для Открытого архива, с другой стороны, в ряде случаев происходит реальное обогащение доступом к документам и данным для других, уже существующих фондов. Например, структурированный в рамках ОА фонд чл.-корр. Алексея Андреевича Ляпунова дополняется документами из архива А.П. Ершова, которых не было в Открытом архиве. На Рис. 7 показан фрагмент списка писем и документов А.А. Ляпунова из Открытого архива, к которому добавились документы из архива А.П. Ершова (выделенные позиции). Копии данных писем в архиве А.А. Ляпунова не сохранились, но они сохранились у адресата, т. е. у А.П. Ершова.

Отзыв А.А. Ляпунов 06.12.1956 о работе А.П. Ершова "Программирующая программа для БЭСМ".	1969-12-21	Письмо Автор рассказывает об издании сборников "Исследование систем" и "Пути автоматизации научно-исследовательских работ", о планах проведения симпозиумов.	получатель
Отзыв А.А. Ляпунов 07.12.1956 о работе А.П. Ершова "Программирующая программа для БЭСМ".	1969-12-22	Письмо Рассказывается о работе семинара на физфаке МГУ по логическим и автоматным моделям в биологии и технике.	получатель
Отзыв А.А. Ляпунов 01.01.1958 о работе А.П. Ершова "Некоторые вопросы теории алгоритмов, св	1969-12-25 17:00:00	Представление к присвоению ученого звания профессора А.П. Ершову.	автор
Статья А.А. Ляпунов 27.06.1967 "О формализации понятия программы". Написана совместно с А.А.	1970-11-24	Постановление Общего собрания АН СССР о просьбе обсудить предложение, высказанное академиком П.С. Александровым. Приложена стенограмма его выступления.	получатель
Статья А.А. Ляпунов 09.07.1967 "О формализации понятия программы"; написана совместно с А.А.	1970-11-26	Письмо Автор рассказывает о своих делах и планах.	получатель
Представление А.А. Ляпунов > Л.В. Ореховичев 20.05.1968 о назначении А.П. Ершова на должность профессора по кафедре те	1970-11-28	Письмо с конвертом Автор просит прислать текст доклада для конгресса, чтобы можно было опубликовать до его начала.	получатель
Представление А.А. Ляпунов > А.П. Ершов 25.12.1969 к присвоению ученого звания профессора А.П. Ершову.	1970-11-28 17:00:00	Телеграмма Поздравление с избранием членом-корреспондентом АН СССР.	автор
Телеграмма А.А. Ляпунов > А.П. Ершов 28.11.1970 Поздравление с избранием членом-корреспондентом АН СССР.	1970-11-30	Письмо Автор спрашивает, не нужен ли А.А. Ляпунову лингвист для работы в отделе.	получатель
Отзыв А.А. Ляпунов 01.12.1970 о научной деятельности А.П. Ершова.	1970-12	Письмо Поздравление с Новым годом. Сообщает, что утверждение диссертации в ВАКе затянulo по неизвестной причине.	получатель
	1970-12-01 00:00:00	Отзыв о научной деятельности А.П. Ершова.	автор
	1970-12-06	Письмо Автор извещает об издании или запланирован издании книг И.А. С. Бабкина	получатель

Рис. 7. Слева фрагмент базы данных архива А.П. Ершова с письмами от А.А. Ляпунова. Справа – те же письма, проявившиеся в документах А.А. Ляпунова в Открытом архиве СО РАН

Одной из наиболее существенных проблем практического применения технологии Semantic Web является тот факт, что она не позволяет создать общую систему идентификации объектов реального мира, значит, нет единой системы идентификации для записей базы данных. На практике это означает, что каждая база данных присваивает идентификационные коды самостоятельно, фактически замыкая идентификационное пространство своими рамками. Интеграция таких баз данных требует отождествления идентификаторов записей, соответствующих одним и тем же сущностям. Процесс отождествления можно экспери-

ментально осуществить на сопоставлении (полных) имен сущностей одинаковых или родственных классов. Такой подход обладает определенной эффективностью, особенно когда имена зафиксированы в полном «официальном» варианте. Например, при обработке списка выпускников ММФ НГУ за несколько лет (более 5000 персон) было выявлено несколько вариантов полного совпадения фамилии, имени и отчества у разных людей. Эффективность отождествления быстро теряется при использовании неполных вариантов имен, инициалов, разных вариантов перевода и т. д. В таких случаях надо применять более сложные методы анализа не только имени, но и контекста [15].

ЗАКЛЮЧЕНИЕ

В результате достаточно длительной эволюции сложился подход к обеспечению совершенствования, в первую очередь технологического, долго живущих информационных систем архивной направленности. На повестке дня – выработка принципов сосуществования разных информационных систем, вопросов их интеграции и дезинтеграции. Были проведены содержательные эксперименты по частичному или полному включению ресурсов одного электронного архива в состав другого. Показано, что такое включение может осуществляться без разрушения целостности систем, представляющих авторскую композицию.

СПИСОК ЛИТЕРАТУРЫ

1. Центральный фондовый каталог Архивного фонда РФ
URL: <http://cfc.rusarchives.ru>
2. *Kornienko S., Gagarina D.* Information systems: new methods of Russian history sources study // International Multidisciplinary Scientific Conferences Social Sciences & Arts SGEM 2015. Conference Proceedings. Anthropology, Archaeology, History and Philosophy. Sofia, 2015.
3. Фотоархив СО РАН. URL: <http://www.soran1957.ru/>
4. Открытый архив СО РАН. URL: <http://odasib.ru/>
5. Древнерусский берестяные грамоты (Рукописные памятники Древней Руси). URL: <http://gramoty.ru/birchbark/>
6. Подвиг народа (Электронный банк документов участников Великой отечественной войны 1941–1945). URL: <http://podvignaroda.ru/>

7. Архив академика В.И. Вернадского.
URL: <http://www.ras.ru/vivernadskyarchive/>
 8. Бессмертный барак (Имена и краткие сведения о более чем 1,9 млн. репрессированных граждан СССР). URL: <https://bessmertnybarak.ru/>
 9. Исторические материалы. URL: <http://istmat.info/>
 10. Прожито (корпус личных дневников и воспоминаний).
URL: <https://prozhito.org/>
 11. Электронный архив Ольги Михайловны Фрейденберг (1890–1955).
URL: <http://freidenberg.ru/>
 12. Трошков С.Н. Об опыте миграции приложений на свободно распространяемое программное обеспечение с открытым кодом // Вестник НГУ Серия: Информационные технологии. 2018. Том 16. Вып. 2. С. 86–94.
<http://doi.org/10.25205/1818-7900-2018-16-2-86-94>. ISSN 1818-7900.
 13. Антюфеев С.В., Марчук А.Г., Немов А.Н., Филиппов В.Э., Черемных Н.А. Электронный архив академика А.П. Ершова // Электронные библиотеки. 2004. Том 7. № 5. С. 1–13.
 14. Berners-Lee Tim, Hender James, Lassila Ora. The Semantic Web// Scientific American. 2001. Vol. 284. No. 5. P. 34–43.
 15. Марчук А.Г., Марчук П.А. Базовая онтология неспецифических сущностей BONE и её использование для построения информационных систем // Вестник СибГУТИ. 2014. № 4. С. 118–128.
-

LONG-TERM ELECTRONIC ARCHIVES: MODERNIZATION AND INTEGRATION

A. G. Marchuk¹ [0000-0001-8455-725X], S. N. Troshkov² [0000-0003-2952-9509],

I. A. Krayneva³ [0000-0002-0601-9795]

¹⁻³A.P. Ershov Institute of Informatics Systems SB RAS

¹mag@iis.nsk, ²kamronis@yandex.ru, ³cora@iis.nsk

Abstract

Over a period of about twenty years, the IIS SB RAS has created information systems of historical orientation: the Electronic Archive of Academician A.P. Ershov, the Photo Archive of the Siberian Branch of the RAS, the Archive of the newspaper "Science in Siberia", the Open Archive of the SB RAS, etc. Each of the resources has its own specifics but in general their content is based on the general social and territorial basis of scientific and public activities of the SB AS USSR/RAS and the Novosibirsk Akademgorodok. In this report we will look at some of the problems of integrating/disintegrating disparate electronic resources into a common platform using existing and emerging tools.

Keywords: *integration of electronic resources, high-quality information, open archives, interdisciplinarity, history of science, Siberian Branch of the RAS, proprietary software, Semantic Web, Drupal.*

REFERENCES

1. Central Fund Catalog of the Archive Fund of the Russian Federation. URL: <http://cfc.rusarchives.ru>
2. Kornienko S., Gagarina D. Information systems: new methods of Russian history sources study // International Multidisciplinary Scientific Conferences Social Sciences & Arts SGEM 2015. Conference Proceedings. Anthropology, Archaeology, History and Philosophy. Sofia, 2015.
3. Photoarchive of the SB RAS. URL: <http://www.soran1957.ru/>
4. Open archive of the SB RAS. URL: <http://odasib.ru/>
5. Old Russian birch bark letters (Manuscripts of Ancient Russia). URL:

<http://gramoty.ru/birchbark/>

6. The feat of the people (Electronic bank of documents of participants in the Great Patriotic War 1941–1945). URL: <http://podvignaroda.ru/>

7. Archive of Academician V.I. Vernadsky.
URL: <http://www.ras.ru/vivernadskyarchive/>

8. Immortal barrack (Names and brief information about more than 1.9 million repressed citizens of the USSR). URL: <https://bessmertnybarak.ru/>

9. Historical materials. URL: <http://istmat.info/>

10. Lived (corpus of personal diaries and memoirs).
URL: <https://prozhito.org/>

11. Electronic archive of Olga Mikhailovna Freidenberg (1890–1955).
URL: <http://freidenberg.ru/>

12. *Troshkov S.N.* On Experience in Migrating Applications to the Freely Distributable Open Source Software // *Vestnik NSU. Series: Information Technologies.* 2018. Vol. 16. No. 2. P. 86–94.

13. *Antufeev S.V., Marchuk A.G., Nemov A.N., Phillipov V.E., Cheremnikh N.A.* Academician A.P. Ershov Electronic Archive // *Russian Digital Libraries.* 2004. Vol. 7. No. 5. P. 1–13.

14. *Berners-Lee Tim, Hendler James, Lassila Ora.* The Semantic Web // *Scientific American.* 2001. Vol. 284. No. 5. P. 34–43.

15. *Marchuk A.G., Marchuk P.A.* BONE base ontology for unspecific entities and its use for construction of informatics systems // *Vestnik SibSUTIS.* 2014. No. 4 (28). P. 118–128.

СВЕДЕНИЯ ОБ АВТОРАХ



МАРЧУК Александр Гурьевич – д. ф.-м. н., зав. лабораторией.

Alexander Gurevich MARCHUK – Doctor of Physical and Mathematical Sciences, Head of the laboratory.

email: mag@iis.nsk.su

ORCID: 0000-0001-8455-725X



ТРОШКОВ Сергей Николаевич – программист 2 кат.

Sergey Nikolaevich TROSHKOV – programmer

email: kamronis@xtech.ru

ORCID: 0000-0003-2952-9509



КРАЙНЕВА Ирина Александровна – д. и. н., с. н. с.

Irina Alexandrovna KRAYNEVA – Doctor of Historical Sciences, Senior Research Officer.

email: cora@iis.nsk.su,

ORCID: 0000-0002-0601-9795

Материал поступил в редакцию 24 октября 2021 года