

Культурное наследие Сибири в электронных коллекциях

© Бредихин С.В. Институт вычислительной математики и математической геофизики СО РАН, Новосибирск bred@nsc.ru	Деев В.Г. Сибирское музейное агентство, Новосибирск vlad@opensib .ru	Ткаченко Т.А. Агентство «Открытая Сибирь», Новосибирск tatyana@opensib .ru	Шабальников И.В. Институт вычислительных технологий СО РАН, Новосибирск igor@nsc.ru	Шаповалов А.В. Агентство «Открытая Сибирь», Новосибирск shap@opensib.ru
---	--	--	--	--

Аннотация

В докладе представлены результаты работы Агентства «Открытая Сибирь» (АОС) по созданию цифрового хранилища культурного наследия Сибири, в котором в настоящее время собрано свыше 20.000 файлов. Основное внимание сосредоточено на программе «Встречи на границах» и проектах «Открытый архив», «Электронный раритет», «Сибирская периодика». Приведены сведения о методах, технологиях и оборудовании, использовавшихся для создания электронных коллекций фотографий, рукописей, живописи, карт и пр. В заключение обсуждаются проблемы и перспективы деятельности АОС.

1 Программа «Встречи на границах»

Деятельность по созданию цифрового хранилища началась в рамках программы Библиотеки Конгресса США «Встречи на границах» [1]. Эта программа проводилась на территории Сибири и Дальнего Востока при поддержке Института «Открытое общество» (Фонд Сороса) в 2000-2004 гг. В сибирских музеях, архивах и библиотеках эксперты из Института истории СО РАН отбирали наиболее интересные коллекции, иллюстрирующие процесс русского освоения Сибири. Главное требование к отбираемым материалам – живость и непосредственность передачи впечатлений

участников социальных, политических, культурных процессов на территории Сибири в период ее освоения. Отобранные материалы оцифровывались в специально созданной в Новосибирске мобильной цифровой лаборатории.

За время реализации программы в ней приняло участие 40 организаций из различных регионов Сибири и Дальнего Востока. Территориальный охват участников – от Тобольска до Петропавловска-Камчатского. Крупные учреждения, например Новосибирский, Красноярский, Томский краеведческие музеи, представили для оцифровки несколько коллекций. Большим достоинством программы мы считаем то, что в ней смогли принять участие и малые провинциальные музеи, такие как Сузунский краеведческий музей, Игарский Музей вечной мерзлоты, Музей истории г. Иркутска и другие подобные учреждения.

В рамках программы было оцифровано 85 уникальных коллекций, содержащих более 9000 единиц хранения. Многие оцифрованные материалы ранее вообще не публиковались и были неизвестны широкой публике. Например, коллекция графических работ художника Смирнова, который в 1903-1904 гг., отбывая сибирскую ссылку, в путешествии от Урала до Иркутска создал портретную галерею сибирских персонажей. Оцифрованы бесценные старообрядческие рукописи из собрания сектора археографии и источниковедения Института истории СО РАН, редчайшие фотографии начала XX века, иллюстрирующие жизнь малых сибирских городов.

Оцифрованные по программе «Встречи на границах» коллекции стали основой Цифрового хранилища культурного наследия Сибири, созданного в Новосибирске Агентством «Открытая Сибирь». Материалы хранятся в специальном технологическом депозитарии на CD и DVD, и

сопровожаются подробными базами данных. Значительная часть материалов размещена в Интернете на сайте «Освоение Сибири» [2].

Для сибирских музеев, архивов и библиотек участие в этой программе стало возможностью сделать качественные страховые копии уникальных материалов культурно-исторического наследия региона для будущих поколений на самом современном технологическом уровне, увеличить их доступность для широких слоев населения, представить их мировой общественности.

2 Проекты Агентства «Открытая Сибирь»

Накопив значительный опыт в ходе осуществления программы «Встречи на границах», АОС в 2003 г. приступило к реализации собственных проектов по созданию электронных коллекций культурного наследия Сибирского региона. Главные задачи, которые АОС ставит перед собой, продолжая это направление своей деятельности, – это спасение раритетных коллекций музеев, библиотек и архивов Сибири и Дальнего Востока путем создания их высококачественных цифровых копий, пополнение Цифрового хранилища культурного наследия Сибири, расширение доступности культурно-исторических ценностей для населения.

В настоящее время реализуются следующие проекты:

2.1 «Открытый архив»

Проект направлен на работу с государственными архивами. Его цель – помочь архивам в создании электронных коллекций для обеспечения сохранности ценного архивного контента и быстрого свободного доступа пользователей к нему. В рамках проекта проведена оцифровка архивного фонда. В Государственном архиве Новосибирской области был выбран фонд XVIII в. «Умревинский острог и судная контора», состоящий из трех дел, включающих в себя 611 документов (свыше 1000 листов). Фактически на этой коллекции был пройден весь цикл работ: от переговоров с партнером до итогового продукта. Результатом этого проекта стало создание страховых копий ценного фонда (почти 1800 файлов) и электронной книги, в которой сохранена важная для библиографического описания архивная нумерация листов и документов, и значительно облегчен поиск информации внутри фонда для пользователя (исследователя, студента и др.).

2.2 «Электронный раритет»

Проект направлен на построение системы цифрового копирования раритетных коллекций письменных памятников, хранящихся в библиотеках и музеях Сибири, и обеспечение к ним доступа широкой публики. В ходе «пилотных» работ по данному проекту проведена оцифровка 3 редких

изданий из библиотек Новосибирска и Иркутска: «Путешествие флота капитана Сарычева по Северовосточной части Сибири, Ледовитому морю и Восточному Океану...» – СПб, 1802 г. (общий объем оцифрованного материала – 460 файлов); Межов В.И. «Сибирская библиография. Указатель книг и статей о Сибири на русском языке и одних только книг на иностранных языках за весь период книгопечатания» – СПб, 1891. – Т.1-4. (1500 файлов); Крашенинников С.П. «Описание земли Камчатки». – СПб, 1755. – Т.1-2. (800 файлов). Разработаны макеты электронных книг, обеспечивающие удобную навигацию и полнотекстовый поиск.

2.3 «Цифровая коллекция сибирской периодики»

В рамках этого проекта ведется оцифровка раритетных журналов и газет из фондов сибирских библиотек. Как правило, такие газеты и журналы находятся в крайне ветхом состоянии, но их востребованность разными пользователями весьма высока. Уже проведена оцифровка журналов «СССР на стройке» за 1937 г. (250 файлов) и «Красная Сибирячка» за 1928 г. (330 файлов), газеты «Томский листок» за 1896 г. (1000 файлов). «СССР на стройке» был элитным советским журналом 1930-1940-х гг., в котором работали классики советской фотографии, сформировавшие каноны жанра мировой фотожурналистики. «Красная Сибирячка» – яркий пример советских пропагандистских журналов 1920-1930-х гг. для женщин, с помощью которых советская власть стремилась максимально вовлечь женское население в общественный труд. Был самым массовым женским журналом Западной Сибири. Газета «Томский листок» – одно из немногих сохранившихся сибирских периодических изданий XIX века.



В настоящее время завершена оцифровка газеты «Советская Сибирь» за 1941-1945 гг. (более 3400 файлов). Эта общественно-политическая газета, старейшее печатное издание Новосибирской

области, выходит с 1919 года. В годы Великой Отечественной войны была официальным органом Новосибирского обкома и горкома ВКП(б), областного и городского Советов депутатов трудящихся, выходила 5-6 раз в неделю. Электронное издание «Советской Сибири» выполнено на DVD и содержит материалы военного времени, ранее доступные только в библиотеке или архиве. На диске в формате PDF собрано 1123 номера газеты с 21 июня 1941 года по 30 июня 1945 года. Предусмотрен полнотекстовый поиск стандартными средствами системы Adobe Reader 6.0.

Все оцифрованные по проектам АОС коллекции пополняют Цифровое хранилище культурного наследия Сибири.

3 Процесс оцифровки коллекции

3.1 Цифровая лаборатория

Работы по оцифровке отобранных материалов проводит специально созданная в Новосибирске мобильная цифровая лаборатория. Персонал лаборатории в настоящее время состоит из 5 человек, прошедших обучение у специалистов Библиотеки Конгресса США и накопивших значительный практический опыт по оцифровке документов различного типа.

3.2 Техническое оснащение

Среднеформатная камера Hasselblad 501 cm с объективами Carl Zeiss Planar C 2.8/80, Carl Zeiss Macro-Planar CFi 4/120; цифровой задник: Phase One StudioKit с эффективным диапазоном – 3500x3500 пикселей, АЦП – 12 бит на цвет, два планшетных сканера UMAX PL2100 XL с разрешением 800x1600 пикселей, осветительная аппаратура, необходимое количество вычислительной техники, объединенной локальной сетью с выходом в Интернет, система стабилизированного электропитания и вспомогательное оборудование.

3.3 Сканирование изображений

Все документы (карты, фотографии, рукописные и печатные листы) оцифровываются с применением планшетного сканера либо цифровой камеры с разрешением 3000 пикселей по длинной стороне листа документа. При сканировании фотоальбомов и книг для аккуратного разворота этих документов используется специальная «колыбель» (book cradle) и люминесцентные осветители, дающие равномерное мягкое освещение. Освещение «холодное», что позволяет избежать теплового воздействия на документы и эффекта мерцания во время сканирования.

Отдельные листы документов формата более А2 (например, географические карты) сканируются также с помощью цифровой камеры по фрагментам, которые затем склеиваются программным путем.

Результаты сканирования сохраняются в формате TIFF на CD. Среднестатистический размер одного кадра составляет 30 Мб. Таким образом, каждый CD содержит порядка 20 изображений. Ведется журнал сканирования.

3.4 Сканирование аудио и видео материалов

Одна из задач проекта «Освоение Сибири» была связана с оцифровкой коллекций грамзаписи – раритетных виниловых дисков. Для ее решения был построен специальный технический стенд, в состав которого входит:

- самодельный проигрыватель дисков на скоростях 78/45/33 оборотов в минуту с приводом типа «direct drive»; стробоскопическое устройство управления и контроля скоростью вращения диска; тонарм, допускающий установку звукозаписывающих устройств различного типа и регулировку давления иглы;
- 2-х канальный 12-ти полосный фильтр частот с аналоговым шумоподавителем и корректором электромагнитных звукозаписывающих устройств;
- ПК с АЦП различного типа;
- многоканальный широкополосный усилитель со звуковыми мониторами.

Результатом сканирования является аудио файл в формате MP3.

Сканирование видео материалов осуществляется в два этапа. На первом этапе средствами нелинейного монтажа выделяются отдельные фрагменты (кадры), которые описываются и хронометрируются. На втором этапе осуществляется оцифровка выбранного фрагмента, который сохраняется в формате DivX. Выделение кадров и их ввод осуществлялся с применением оборудования «Форвард» [3].

Следует отметить, что работа с аудио и видео материалами изначально носила экспериментальный характер и, по ряду причин, так и не переросла его.

3.5 База данных

Одновременно со сканированием ведется заполнение базы данных. Каждому документу соответствует одна запись в базе. В рамках проекта «Встречи на границах» были определены следующие элементы (поля) базы, требующие обязательного заполнения [4]. Следуя рекомендациям «Дублинского ядра» [5] все элементы описания коллекций и документов условно разбиты на три группы. Первая группа содержит элементы, описывающие коллекцию в целом; сюда относится название коллекции и ее описание. Вторая группа содержит информация об учреждении – хранителе коллекции, обладающем правом интеллектуальной собственности на эту коллекцию. В третьей группе содержатся элементы, описывающие каждый документ коллекции: название документа, его инвентарный (учетный) номер, автор документа, дата создания, информация о носителе (размер, материал, техника), место создания, ключевые слова и примечания.

Первоначально база данных ведется в формате Microsoft Access 2000, затем она конвертируется в форматах Oracle либо MySQL.

4 Методы и технологии представления коллекций

4.1 Сайт «Освоение Сибири. Информационные ресурсы»

Одной из важных, конечных целей создания электронных коллекций культурно-исторического наследия является увеличение его доступности для широкой публики. На первоначальном этапе, в рамках программы «Встречи на границах», основной формой представления оцифрованных коллекций было выставление их на сайте «Освоение Сибири».



Один из принципов показа оцифрованных материалов – разделение по типу исходного материала. Выделены следующие категории: фотографии, фотоальбомы, рукописи, документы, графика, живопись, карты, атласы, периодика, аудио и видеоматериалы. Для каждого типа есть свои особенности представления.

Другой основополагающий принцип, которому мы следовали, – это коллекционный подход. Все документы сгруппированы в коллекции. Показ фотографий, рукописей, карт и других источников сопровождается дополнительным описанием коллекции, в которую они входят. Название коллекций и их описания сформулированы нами в согласовании с учреждениями – хранителями.

4.2 Представление изображений

Фотографии и фотоальбомы, рукописи, графика и живопись, карты и атласы на сайте «Освоение Сибири. Информационные ресурсы» представлены в виде изображений различного размера. Все документы сопровождается информацией о коллекции, к которой они принадлежат.

Показ какого-либо документа сопровождается выводом на экран его изображения размером 140 пикселей по длинной стороне документа и информации из базы данных. Для детального рассмотрения доступно увеличенное изображение документа – 640 пикселей. Размер карт – 800 пикселей.

Листы фотоальбомов и атласов разбиты на группы по восемь страниц в каждой. При выборе конкретной страницы на экран выводится ее увеличенное изображение, номер и подписи к странице, если они есть. Фотоальбомы снабжены списками фотографий и подписями к ним. Существует возможность просмотра не только изображения страницы в целом, но и каждой конкретной фотографии на этой странице в отдельности.

При работе с рукописями читатель видит изображение ее первой страницы, описание рукописи, ссылку на группы страниц и ссылки на страницы, которые имеют иллюстрации. Для детального изучения отдельной страницы возможен ее вывод в увеличенном формате.

4.3 Электронные книги

При создании электронных книг мы обеспечиваем возможность полнотекстового поиска. В качестве примера представим перечень работ, выполненных в процессе создания электронной книги 4-х томного издания Межов В.И. «Сибирская библиография»:

- выбор максимального разрешения сканирования с условием размещения всех четырех томов книги на одном CD размером 700 Мб (в данном случае выбрано разрешение 300 dpi);
- с помощью системы FineReader создается 4 (по количеству томов) «пакета», каждый пакет содержит все изображения страниц данного тома;
- запускается процесс распознавания текста на каждом пакете. Язык распознавания «Русский (старая орфография)». В некоторых местах (иностраный текст) ставятся языки «Английский, Немецкий»;
- изображения страниц и распознанный текст сохранялся в формате PDF, в режиме «текст под изображением»;
- для каждого из томов были созданы закладки и расставлены ссылки на соответствующие страницы;
- все тома были проиндексированы для ускорения полнотекстового поиска; PDF файлы были записаны на CD в формате Adobe AcrobatReader;
- изготовление макетов этикетки диска и вкладыша в пластиковую коробку.

5 Проблемы: взгляд из региона

Как и всякое новое начинание, применение цифровых технологий в деле сохранения культурно-исторического наследия сталкивается с целым

рядом трудностей и проблем технологического, финансового, организационного характера.

Сохраняется настороженное отношение руководства и рядовых работников библиотек, архивов и музеев к новым технологиям, особенно консервативны в этом отношении архивы. В данном случае сказывается и вполне понятный консерватизм, обуславливаемый самим характером их деятельности. Главная задача для таких учреждений – сохранить культурное наследие, музейные, архивные работники опасаются лишней раз использовать ценные и, как правило, ветхие материалы. Но присутствует и элементарное незнание возможностей, которые могут дать эти технологии учреждениям культуры.

Другая проблемная ситуация – последующее использование созданных электронных коллекций. Оцифровка культурного наследия расширяет возможности хранящих его учреждений – появляются новые формы представления культурно-исторических памятников, возможности привлечения новых категорий посетителей, электронные коллекции могут быть использованы в коммерческих целях. Но эффективная реализация этих возможностей требует применения современных маркетинговых и менеджерских технологий, которыми многие провинциальные учреждения культуры не владеют.

Существующие федеральные программы создания электронных библиотек практически не затрагивают регионы, в них участвуют преимущественно московские, Санкт-Петербургские и близкие им учреждения. В результате отсутствия устойчивого финансирования оцифровка в регионах проводится от случая к случаю по мере нахождения денег под конкретный проект. Такая ситуация приводит к тому, что в региональных библиотеках, музеях, архивах скапливаются разрозненные и плохо стыкуемые друг с другом электронные коллекции, наличие которых отнюдь не равнозначно формированию единой электронной библиотеки.

Это лишь некоторые из проблем, с которыми приходится сталкиваться в регионах, при переводе культурного наследия в электронный вид.

Ресурсы Интернет

- [1] Встречи на границах. Веб-сайт библиотеки конгресса США, 2005
frontiers.loc.gov
- [2] Освоение Сибири. Информационные ресурсы. Веб-сайт Агентства «Открытая Сибирь», 2005
www.sibheritage.nsc.ru
- [3] Веб-сайт компании «СофтЛаб-НСК», 2005
www.softlab-nsk.com/rus/
- [4] Рекомендации по созданию цифровых коллекций проекта «Встречи на границах». Веб-сайт библиотеки конгресса США, 2005
frontiers.loc.gov/intldl/mtfhtml/mfabout/mfdigitz.html

[5] DCMI Metadata Terms. Веб-сайт «Dublin Core Metadata Initiative», 2005
dublincore.org/documents/dcmi-terms/#A1

Cultural Heritage of Siberia in Electronic Collections

Bredikhin S.V.
Institute of Computational Mathematics and
Mathematical Geophysics, Siberian Branch of the
Russian Academy of Sciences, Novosibirsk
bred@nsc.ru

Deev V.G.
Siberian museum agency, Novosibirsk
vlad@opensib.ru

Shabalnikov I.V.
Institute of Computational Technologies, Siberian
Branch of the Russian Academy of Sciences,
Novosibirsk
igor@nsc.ru

Shapovalov A.V.
“Open Siberia” Agency, Novosibirsk
shap@opensib.ru

Tkachenko T.A.
“Open Siberia” Agency, Novosibirsk
tatyana@opensib.ru

In the report there are presented the results of the “Open Siberia” Agency (OSA) activity on creating of the digital depository of Siberian cultural heritage, which contains more than 20000 files now. The primary attention is focused on the “Meeting of frontiers” program and “Open archives”, “Electronic rarity”, “Siberian periodicals” projects. The information is provided on the methods, technologies and equipment used for the creation of the electronic collections of photographs, manuscripts, paintings, maps, etc. In conclusion the problems and perspectives of the OSA activity are discussed.