

TECHNOLOGIE ZAPEWNIENIA CYKLU ŻYCIA ZASOBÓW CYFROWYCH BIBLIOTEKI NAUKOWEJ

Zwrócono uwagę na aspekt technologiczny cyklu życia tworzenia się zasobów cyfrowych biblioteki. Określone zostały podsystemy kompleksu cyfrowego biblioteki naukowej. Uzasadniony został skład i sekwencyjność technologicznych procesów digitalizacji zasobów bibliotecznych.

Słowa kluczowe: biblioteka cyfrowa, zasoby cyfrowe biblioteki, cykl życiowy zasobów cyfrowych, kompleks cyfrowy biblioteki naukowej.

TECHNOLOGY LIFECYCLE SUPPORTING OF RESEARCH LIBRARY DIGITAL RESOURCES

The technological level of the life cycle of library digital resources forming are considered. The subsystems of scientific library digital complex are defined. Composition and sequence of the technological processes of library documents digitization are grounded.

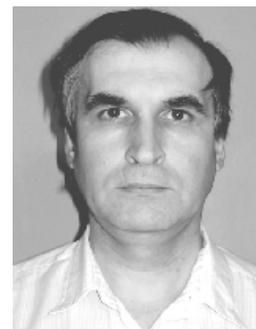
Key words: digital library, library digital resource, life cycle of digital resource, digital complex of research library.

ТЕХНОЛОГІЯ ЗАБЕЗПЕЧЕННЯ ЖИТТЄВОГО ЦИКЛУ ЦИФРОВИХ РЕСУРСІВ НАУКОВОЇ БІБЛІОТЕКИ

Розглянуто технологічні ланки життєвого циклу формування бібліотечних цифрових ресурсів. Визначено підсистеми цифрового комплексу наукової бібліотеки. Обґрунтовано склад і послідовність технологічних процесів оцифрування бібліотечних документів.

Ключові слова: цифрова бібліотека, бібліотечний цифровий ресурс, життєвий цикл цифрового ресурсу, цифровий комплекс наукової бібліотеки.

Главным современным направлением сохранения документального национального достояния является организация оцифровки фондов библиотек. Следствием осуществления различных программ и проектов по оцифровке историко-культурного наследия стало осознание того, что необходимо кооперативное взаимодействие на национальном и межгосударственном уровнях для создания полноценных цифровых коллекций, которые будут отражать как национальную самобытность отдельных регионов так и все разнообразие мирового научного и культурного опыта. Такая кооперация усилий поможет виртуальной реконструкции рассеянных между различными государствами коллекций, предотвратит излишнее дублирование информации, предоставит единую точку доступа пользователям к необходимой им информации. Осознание всех этих реалий привело к развертыванию международных проектов по сохранению и предоставлению доступа к



I. Lobuzyn
pracownik naukowy
wydziału
oprogramowania i
wsparcia
technologicznego
sieci komputerowych
Narodowej biblioteki
Ukrainy
im. W. Wernadskoho
(m. Kijów, Ukraina)

объектам национального достояния Мировая Цифровая Библиотека (World Digital Library, <http://www.wdl.org>), Европейская Электронная Библиотека (The European Library, www.theeuropeanlibrary.org), Европеана (Europeana, <http://www.europeana.eu>). Целью этих проектов является представление на единой технологической основе всего этнического и национального многообразия мировой культуры. Однако, здесь следует отметить, что несмотря на достаточно детально разработанные стратегии оцифровки фондов библиотек, можно констатировать отсутствие единого системного подхода к организации самой технологии формирования библиотечных цифровых ресурсов.

Для решения задач организации цифровых ресурсов библиотек современный информационный рынок предлагает разнообразные программно-технологические средства. Каждая библиотека должна определить приоритеты и выбрать из этого разнообразия наиболее рациональные модели и средства интеграции собственных информационных ресурсов в мировое информационное пространство. Как показывает мировая практика здесь есть несколько основных подходов: использовать уже готовые решения ведущих мировых дистрибьюторов, воспользоваться бесплатными программными платформами, разработанными в рамках инициативы архивов открытого доступа, или адаптировать доступные программные решения на платформе имеющейся в библиотеке автоматизированной библиотечной информационной системы (АБИС).

Независимо от того на какой из базовых платформ остановиться библиотека (с нашей точки зрения вариант АБИС будет наиболее рациональным по соотношению цены и качества), в процессе создания цифровых ресурсов возникает необходимость создать библиотечный цифровой комплекс, обосновать его структурные элементы и технологические звенья, наладить их функциональное взаимодействие.

Ведущие специалисты по управлению цифровыми ресурсами (digital curation) определяют следующие главные звенья организации жизненного цикла полноценного функционирования цифрового ресурса: концептуализация (планирование отбора и создания цифровых материалов); описание и отбор (введение метаданных, оценка цифрового материала и отбор для долгосрочного хранения и восстановления); управления (оценка цифрового материала и отбор для долгосрочного хранения и восстановления, соблюдения правовых требований, удаление материалов, которые не были выбраны для долгосрочного хранения и восстановления, передача материалов в архив, хранилище, центр обработки данных); сохранение (проведение мероприятий по обеспечению долгосрочного хранения: проверка метаданных, проверка целостности цифровых данных, запись данных с соблюдением соответствующих стандартов); доступ и использование (предоставление активного доступа пользователям к цифровому ресурсу: общедоступного, авторизованного или платного) [12, 13]. Именно по этим принципам построены комплексные решения управления цифровыми архивами (например, DocuShare XEROX).

Таким образом с технологической точки зрения, среди основных модулей (подсистем) информационной системы цифрового фонда, которые будут обеспечивать его стабильное и эффективное функционирование, целесообразно выделить следующие:

- создание цифровых копий;
- описания и учет цифровых ресурсов;
- обработка и управление цифровыми ресурсами;
- хранение цифрового фонда;
- поиск и публикация цифровых ресурсов.

Каждая из этих подсистем имеет свои принципы функционирования и требует отдельного исследования теоретических и технологических требований в процессе

обработки и упорядочивания цифровых ресурсов. Рассмотрим более подробно эти основные структурные блоки информационной системы цифрового комплекса научной библиотеки.

Подсистема создания цифровых копий

Эта подсистема является аналогом технологии комплектования цифровой библиотеки. Первым вопросом которую решает эта подсистема это отбор документов для оцифровки. Критериями такого отбора могут быть разнообразные факторы, основными из них являются:

- уникальные, редкие или ценные документы, существующие в одном или в ограниченном количестве экземпляров, в том числе рукописи и фотографии.
- поврежденные, хрупкие или крупногабаритные документы, нуждающихся в особых условиях доступа к ним. Оцифровка таких документов способствует их сохранению, при условии предоставления доступа пользователям к их цифровым копиям;
- документы, относящиеся к объектам национального достояния, введение которых в широкий культурный оборот имеет большое значение для восстановления национальной памяти, истории и культуры государства.

Для цифровых архивов пополнения информации происходит за счет перевода бумажных документов в электронный вид. Для этого используются, в зависимости от типа и формата документов, сканеры, цифровое фотооборудование и программное обеспечение цифровой обработки изображений.

Для подбора сканирующего комплекса современной научной библиотеки, в соответствии с практическими рекомендациями Администрации национальных архивов и документации США, необходимо определить характеристики и основные параметры оригиналов документов, цифровые копии которых необходимо будет изготовить [18]. К таким основным параметрам оригиналов относятся: тип бумаги и способы изготовления, локальные недостатки, требующие предварительной реставрации или ретуширования, тональность, резкость (четкость), габаритные размеры. Необходимо оценить размеры листов бумаги для рукописных и печатных документов, количество документов каждого формата. От этой оценки будет зависеть подбор аппаратного комплекса для оцифровки конкретного библиотечного фонда.

После сканирования полученная цифровая информация требует обработки для доведения отсканированного изображения до необходимого пользовательского вида. Создано много приложений с помощью которых можно корректировать полученные при сканировании изображения. Развитые возможности имеет популярный программный пакет Adobe Photoshop, который имеет огромное количество различных инструментов для работы с изображениями. Среди программ, которые позволяют осуществлять быстрое пакетное обработки полученных цифровых изображений следует отметить ACDSee, Microsoft PfotoEditor (встроенная в Microsoft Office), Raster Desk, Raster ID, Spotlight. Для преобразования отсканированных бумажных документов в электронные форматы для последующего редактирования – системы распознавания текста: ABBYY FineReader, OCR CuneiForm, Readiris, Microsoft Office Document Imaging. Для редактирования объемных изображений используются программы StudioMAX, Maya, Rhinoceros, SolidWorks и другие пакеты 3D-моделирования.

Следующим этапом создания эталонной цифровой копии документов является оценка полноты и качества полученных цифровых изображений. К полноте относится проверка наличия всех листов или изображений всех сторон оригинала. К понятию качества относятся в первую очередь чёткость полученных изображений и цветная

соответствие оригиналу. Для съема страховых копий редких и старинных документов с цветными изображениями, оцифровки документов происходит вместе с эталонной шкале цветов, которая часто прилагается к комплекту сопроводительной документации профессиональных сканеров.

Подсистема описания и учета цифровых ресурсов

Учет и инвентаризация цифровых массивов, в том числе в сфере образования, науки и культуры, является предметом ряда национальных и международных проектов: создание Проект Минерва (Европейский союз), <http://www.minervaeurope.org/>; Цифровые библиотеки для образования, науки и культуры (ЮНЕСКО), www.wdl.org.

Система учета цифровых ресурсов обеспечивает в цифровом фонде те же функции, что и система учета традиционной библиотеки – сохранность фонда. Именно эти вопросы сохранности цифровых ресурсов, объемы которых лавинообразно растут во всем мире, вызвали необходимость описания (создания метаданных) цифровых ресурсов для обеспечения полноценного управления цифровыми ресурсами и сохранения их в будущем. Очень важным компонентом стратегий сохранения цифрового наследия является метаданные, содержащие информацию, необходимую для документирования процесса хранения. Метаданные сохранности представляют собой информацию о формате, структуре и использовании цифрового ресурса; историю всех операций, проведенных над объектом хранения, в том числе любые изменения; аутентичность; техническую историю; историю хранения; ответственность; права, связанные с хранением и т.д. [16].

Таким образом, система учета цифровых ресурсов тесно связана с выбором метаданных для описания объектов создаваемого цифрового фонда. Основными требованиями к метаописаниям объектов являются: полнота представления знаний об объектах; возможность использования метаописаний объектов в других системах; использования общепринятого стандарта, который совместим с другими системами.

Добавление метаданных к электронным ресурсам системы предоставляет возможность более точного позиционирования информации об объектах, улучшает механизм фильтрации и отбора знаний, упрощает и ускоряет процессы доступа к необходимым программам, серверам, ресурсам.

Однако таких систем метаданных, как универсальных, так и ориентированных на отдельные виды цифровых массивов, насчитывается сейчас уже более 50 [6]. Наиболее перспективными и распространенным является стандарт Дублинского ядра метаданных (Dublin Core Metadata Initiative, DCMI) [10]. В среде библиотечного сообщества более популярный MARC форматы, в образовательной среде – LOM, в музейной среде – CIMI, архивного-EAD и т.д. Кроме того, даже в рамках Дублинского ядра возможны различные версии, обусловлены разным набором квалификаторов, различными методическими принципами. Однозначного решения этой проблемы не существует, хотя наиболее приемлемым выходом из положения является применение иерархии систем метаданных, в которой вершину образует Дублинское ядро, как наиболее общая система. Другие системы метаданных, могут быть иерархическим развитием Дублинского ядра, что обеспечит однозначное преобразование метаданных.

Однако, для цифровых библиотечных ресурсов все же определенные преимущества имеют решения, основанные на семье MARC-форматов, так как это дает возможность использовать уже готовые описания библиотечных документов с электронных каталогов и библиографических баз данных [14].

Ведущие специалисты по созданию электронных архивов, такие как Администрация национальных архивов США [18], рекомендуют при формировании фонда цифровых копий

вводить не только описательные метаданные документов, а обязательно предусмотреть технические и правовые метаданные. Технические метаданные должны отразить основные характеристики изображения (цвет, разрешение, размер, объем), оборудование и программное обеспечение средствами которых созданы изображения и электронная версия документа. Правовые метаданные особенно важны для документов архивного фонда, где по правилам на копии распространяются те же ограничения, которые касаются их бумажных аналогов. Одной из существенных проблем, которая связана с регламентацией доступа к информации, с которой сталкиваются библиотеки при обслуживании электронными информационными ресурсами является соблюдение авторских прав и правил архивного обслуживания.

Подсистема обработки и управления цифровыми ресурсами

После создания оцифрованных копий библиотечных документов полученный цифровой фонд требует дальнейшего упорядочения. Основные процессы, которые необходимо осуществить на этом этапе это создание электронных версий документов для полноценного их представления пользователям и формирования цифровых коллекций с соответствующим поисковым аппаратом. Рассмотрим более подробно эти процессы с точки зрения обработки цифровых ресурсов научной библиотеки.

Требования к электронным версиям документов должны в первую очередь отвечать критерию достоверности, а именно документ в электронном виде должен максимально воспроизводить оригинал. Кроме того, информация в электронном виде предоставляет дополнительные возможности со стороны удобства манипулирования информацией: удобная навигация в материале и поисковые индексы.

Для представления документов в электронных библиотеках могут использоваться различные форматы, в том числе наиболее распространенными являются: PDF, DjVu, DOC, TXT, TIFF, JPEG, HTML, XML, FB (FictionBook), CHM, SWF (FLASH-документы).

Могут быть также предусмотрены особые форматы для представления аудиовидеоинформации, геоданных, трехмерных или анимационных объектов, экспериментальных данных и других особых видов информации.

Форматы изображений для хранения страховых копий документов должны соответствовать следующим основным техническим требованиям: поддержка битовой глубины более 24 для однотоновых документов; схемы сжатия данных без потерь; поддержка различных цветовых моделей и возможность управления цветом, высокое разрешение мастер копий; открытость и подробное документирование, широкая поддержка и кроссплатформенная совместимость.

Проблему защиты авторских прав в библиотеках успешно решают системы удаленного защищенного просмотра документов такие как «DefView», «Vivaldi», «SDViewer» и другие. Реализация средств защищенного просмотра позволяет библиотекам организовать обслуживание удаленных пользователей документами защищенными авторским правом в виртуальном читальном зале.

Еще одним важным вопросом является распознавание документов и хранение их в символьных форматах. Распознавание текста (Optical character recognition, OCR) может быть полным целого документа, возможны также промежуточные решения – распознавание некоторых частей документов (содержание, аннотация, титульный лист) или так называемое «черное» распознавания, т.е. распознавание без корректуры. Принятие решения в данном случае оказывается весьма непростой задачей. Распознавание позволяет посимвольную обработку текста, следовательно, различных способов работы с текстами (поиск, редактирование, экспорт, импорт и т.д.). В то же время, оцифровывание документов

с переводом в символьные форматы значительно дороже [17]. Программы OCR работают не идеально, допуская определенное количество ошибок, после их работы для качественной подготовки текста нужна еще и ручная корректура. В процессе перевода качественного четкого текста количество ошибок составляет 1% и меньше. В то же время в процессе оцифровки старинных книг количество ошибок доходит до 50% [4].

Для текстов «неоптимальных» для распознавания, например рукописных текстов и текстов с математической нотацией, нетрадиционных алфавитов и т.п., проблема вообще не может быть решена в ближайшем будущем. Альтернативой технологии «сканирование - распознавание - корректура» является технология ручного ввода (перепечатки) текстов, которая во многих случаях оказывается дешевле [1].

В процессе подготовки цифровых копий уникальных и ценных печатных и рукописных объектов, в соответствии с современными требованиями профессионального кураторства цифровым ресурсом и обеспечения его сохранности в будущем, необходимо соблюдать технологическую политику, которая обеспечивает возможность сохранения электронных копий в течение длительного времени и основанной на следующих принципах: создание базовой (депозитарной) цифровой копии объекта; формирование на основе депозитарной копии необходимых для пользователя представлений цифрового объекта в различных удобных для работы форматах; возможность итеративного возврата с целью модификации и развития цифровой копии объекта в процессе его эксплуатации.

Подсистема хранения цифрового фонда

С точки зрения обеспечения сохранности больших объемов цифровых данных, в последние годы широкое распространение получила концепция управления жизненным циклом информации, в основе которой лежит принцип разделения общего массива данных на классы в зависимости от содержания, частоты обращений и срока хранения. В соответствии с этим подходом выделяются три ключевые задачи хранения цифровых данных: оперативный доступ к информации, резервное копирование и архивное хранение. Для решения каждой из них применяется различное оборудование – по специфическим требованиям к хранению и доступу [2].

Оперативный архив реализуют размещением на сервере в специальном разделе жесткого диска (для небольших объемов информации), на отдельном жестком диске или RAID-массиве. Для повышения надежности функционирования подобных систем используют средства резервного копирования, дублирования информации и «зеркальных дисков». RAID-массивы также используют для хранения метаданных и учетной базы данных архива. В практике американских электронных архивов эти данные называются Master References File (эталонный файл) [18]. Наиболее перспективным решением, одновременно удовлетворяет требованиям скорости и надежности является RAID-5 или RAID 6.

Дальнейшим развитием решений по хранению больших объемов данных являются специализированные системы хранения данных (СХД). Современная система хранения данных должна включать следующие основные компоненты: устройства хранения данных (жесткие диски, компакт диски и т.д.); инфраструктуру доступа серверов к устройствам хранения данных; систему архивирования и резервного копирования данных; программное обеспечение управления хранением данных. Основными задачами, которые выполняет СЗД является надежное хранение данных, а также отказоустойчивость устройств и высокопроизводительный доступ серверов к средствам хранения данных. Основные методы, используемые сегодня для реализации этих задач, можно охарактеризовать одним словом – дублирование (дублирование данных и каналов доступа к ним) [5]. Большинство

современных систем хранения данных осуществляют процессы репликации и резервного копирования в автоматизированном режиме.

Для организации области долгосрочного (архивного) хранения используются сменные носители информации CD / DVD / UDO и роботизированные библиотеки на их основе. Наиболее совершенные решения, поддерживают сверхплотный оптический формат (Ultra Density Optical, UDO) и технологии Blue-Ray. Популярность технологии UDO основывается на возможности записать на один носитель 30-60 Гб информации и сроком хранения информации 50 лет. Роботизированная библиотека является следующим этапом создания электронного архива длительного хранения, она представляет собой массив переменных дисков, размещенный в специальном корпусе. Решение на основе роботизированной библиотеки со сменными носителями информации также позволяет достаточно легко решить вопросы технической защиты информации и обеспечить ее надежное хранение [7].

С точки зрения хранения самих файлов электронных документов есть также два основных подхода: при использовании первого из них электронные документы хранятся в бинарном виде в соответствующих полях самой базы данных архива, второй подход базируется на том, что файлы могут храниться отдельно, а в базе данных есть только электронные адреса файлов (указывается путь доступа к файлу). Первый из способов быстро приводит к целому ряду проблем: осложнения резервирования данных, переход к другой информационной платформы, снижение быстродействия системы. Он может быть рекомендован только для архивов с небольшим физическим объемом информации.

Кроме надежности и производительности электронного архива или хранилища данных важна также логическая структура хранилища. За базовую структуру рекомендуется принять иерархию папок с соответствующими документами, которая будет отображать логическую связь основных узлов системы.

Подсистема поиска и публикации цифровых ресурсов

Окончательной целью любого проекта по оцифровке историко-культурных и научных библиотечных информационных ресурсов является реализация доступа к созданному информационному массиву для широких слоев населения.

Любая цифровая библиотека, предназначенная для определенного круга пользователей и должна предоставлять поддержку конкретных видов деятельности. Таким образом, для разработки качественного пользовательского интерфейса, функционирования цифровые библиотеки должна быть оценена в контексте их целевой аудитории и сферы деятельности [17]. Развитые функциональные и поисковые возможности системы с точки зрения эргономичного интерфейса научной библиотеки, должны включать: навигацию во всем доступном информационном пространстве, лексический поиск, символьный поиск, атрибутивный поиск (автор, название, место издания, дата и т.п.); просмотр содержания цифрового объекта и его структуры; многооконный режим работы; поддержка аппарата гипертекстовых и гипермедийных связей; протоколирование сеанса работы пользователя; экспорт информации из системы с указанием на источник информации [1].

Относительно дальнейшего развития когнитивных возможностей интерфейса цифровой библиотеки следует рассмотреть круг вопросов, относящихся к организации интеллектуального поиска информационных ресурсов библиотеки. Разработку такого интеллектуального инструментария обеспечивают библиотечные технологии семантической (научной) обработки информации, когда при описании ресурсов учитываются не только формальные описательные признаки документов, а в первую очередь раскрывается их содержание. Дополнительными средствами раскрытия

содержания цифровых ресурсов, особенно для неструктурированной графической информации оцифрованных страниц, является введение содержания документов и аннотаций. Текстовая информация этих дополнительных точек доступа автоматически включается в поисковые индексы, что значительно расширяет поисковый потенциал, созданного информационного массива.

Использование устоявшихся и хорошо разработанных семантических библиотечных технологий значительно оптимизирует усилия на усовершенствование поискового аппарата цифровой библиотеки. При формировании цифрового массива есть возможность воспользоваться уже готовыми средствами тематического поиска: рубрикаторами, классификаторами, тезауруса, авторитетными файлами; авторитетными файлами: лиц и организаций; научными описаниями рукописных и архивных фондов научной библиотеки. Все эти средства значительно улучшают поисковый аппарат цифрового фонда, предоставляют пользователю в удобном визуализированном гипертекстовой виде дополнительную справочно-библиографическую информацию о его информационного запроса, позволяют реализовывать наукоемкие средства информационного поиска и получить полную и достоверную информацию.

Выводы

Предложенный модульный цифровой комплекс был сформирован и апробирован в течение четырех лет в рамках цифрового проекта Национальной библиотеки Украины имени В. И. Вернадского. Как показал накопленный опыт, каждая из подсистем (модулей): создание; описания и учет; обработка и управление; хранение; поиск и публикация цифровых ресурсов соответствует определенному этапу жизненного цикла формирования цифрового фонда научной библиотеки. Разработка отдельных модулей с определенной степенью автономными задачами создало предпосылки для детального исследования особенностей каждого технологического звена, более рациональной организации и распределения профессиональных обязанностей специалистов библиотеки. Окончательно, созданный цифровой комплекс и последовательность модулей с определенным кругом технологических процессов обеспечили:

- решение двух главных задач цифрового проекта научной библиотеки: 1) создание цифровой библиотеки открытого доступа и 2) организацию страхового фонда цифровых копий документов;
- логическую последовательность выполнения операций технологического цикла;
- необходимую технологическую инфраструктуру для выполнения всех задач, которые предусматривает цифровой проект научной библиотеки;
- рациональное использование метаданных цифровых объектов на всех технологических этапах: создание - сохранение - учет - публикация цифровых ресурсов;
- обоснованное распределение функциональных обязанностей и ответственности сотрудников, принимающих участие в формировании цифрового ресурса;
- развитую систему поиска информации в цифровой библиотеке, ориентированную на проведение научных исследований;
- совместимость с общепринятыми стандартами и возможность интеграции созданных цифровых ресурсов к международные цифровые проекты.

Литература:

1. Антопольский А. Б. Правовые и технологические проблемы создания и функционирования электронных библиотек / А. Б. Антопольский, Е. А. Данилина, Т. С. Маркова. – М. : ПАТЕНТ, 2008. – 207 с.

2. Корепанов И. Как сохранить архив на десятилетия? [Электронный ресурс] / И. Корепанов // «Журнал сетевых решений (LAN)». – 2008. – № 3. – URL: <http://www.osp.ru/lan/2008/03/4899898/>.
3. Лобузін І. В. Створення бази даних цифрового фонду бібліотеки: основні проблеми та технологічні рішення / І. В. Лобузін // Науково-технічна інформація (НТІ). – 2013. – № 1. – С. 67–62.
4. Масевич А. Ц. К созданию электронных коллекций старопечатных книг в библиотеке Российской академии наук: на примере работы над двумя проектами / А. Ц. Масевич, Е. А. Савельев, А. К. Багажков // Новые технологии в информац. обесп. науки : тр. конф. — М. : Биоинформсервис, 2001. — С.132–140.
5. Меленець А. В. Обрання типу сховища для зберігання документів страхового фонду в електронному вигляді // Системи обробки інформації. – Харків, ХУПС. – 2010. – № 6(87). – С. 199 – 203.
6. Обзор форматов метаданных / укл.: Ю.Е. Хохлов, С.А. Арнаутов ; Институт развития информационного общества [Электронный ресурс] // Российские электронные библиотеки. – URL: http://www.elbib.ru/index.phtml?page=elbib/rus/methodology/md_rev.
7. Петров І. В. Обґрунтування вибору програмно-апаратних засобів архівів електронних документів довготермінового збереження / І. В. Петров, А. М. Стеценко, Н. В. Солоніна // Реєстрація, зберігання і обробка даних. – 2010. – Т. 12. – № 1. – С. 79 – 88.
8. Best Practice Guidelines for Digital Collections [Electronic resource] / edited by Susan Schreibman // University of Maryland Libraries. - 2006.- URL: <http://www.lib.umd.edu/digital/>.
9. Chowdhury S. Usability and impact of digital libraries: a review / Sudatta Chowdhury, Monica Landoni and Forbes Gibb // Online Information Review. – 2006. – № 6. – P. 656 – 680.
10. DDCMI (Dublin Core Metadata Initiative) [Electronic resource]. – URL: <http://dublincore.org>.
11. Guidelines for digitization projects for collections and holdings in the public domain, particularly those held by libraries and archives [Electronic resource] / IFLA ; UNESCO // IFLA. – 2003. – URL: <http://archive.ifla.org/VII/s19/pubs/digit-guide.pdf>.
12. Harvey R. Digital Curation : A How-To-Do-It Manual / R. Harvey. – New York ; London : Neal-Schuman Publishers, Inc., 2010. – 17 p. – (How-to-do-it manuals ; № 170).
13. Higgins S. Digital Curation: The Emergence of a New Discipline // The International Journal of Digital Curation. – 2011. – Vol. 6, № 2. – P. 78 – 88.
14. MARC 21 Format for Bibliographic Data [Electronic resource] // Library of Congress. – URL: <http://www.loc.gov/marc/>.
15. Open Archives Initiative (OAI) [Electronic resource]. – 2007. – URL: www.openarchives.org/.
16. PREMIS Data Dictionary for Preservation Metadata [Electronic resource] : Final Report of the PREMIS Working Group. – 2005. – URL: www.loc.gov/standards/premis/v2/premis-2-0.pdf.
17. Smith A. Strategies for Building Digitized Collections. / Digital Library Federation ; Council on Library and Information Resources. – Washington (DC), 2011. – 35 p.
18. Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Image [Электронный ресурс] / S. T. Puglia, J. Reed, E. Rhodes; U.S. National Archives and Records Administration (NARA). – 2004. – 87 p. – URL: <http://www.archives.gov/preservation/technical/guidelines.html>.

TECHNOLOGY LIFECYCLE SUPPORTING OF RESEARCH LIBRARY DIGITAL RESOURCES

I. LOBUZIN

The main modern direction of preservation of documentary national heritage is the organization digitization of libraries. A consequence of the implementation of various programmes and projects on digitization of cultural heritage was the realization that it is necessary cooperative interaction at the national and international levels to create a complete digital collections, which will reflect both the national identity of the individual regions and the diversity of the world's scientific and cultural experience. This cooperative effort will help the virtual reconstruction of the scattered between different States collections, will prevent unnecessary duplication of information, will provide a single point of access for users to the information they need. Awareness of these realities led to the deployment of international projects for the preservation and provision of access to national heritage sites the World Digital Library World Digital Library, <http://www.wdl.org>), the European digital Library (The European Library, www.theeuropeanlibrary.org), Europeana (Europeana, <http://www.europeana.eu>). The aim of these projects is the view on a single technological basis of all ethnic and national diversity of world culture. However, it should be noted that despite the fairly elaborate strategy digitization of libraries, it is possible to ascertain the lack of a unified system approach to the technology of forming a library of digital resources.

For solving problems of organization of digital resources of libraries of modern information market offers a variety of software and technological tools. Each library must prioritize and choose from this variety the most rational models and tools integration of information resources into the global information space. As world practice shows, there are several basic approaches: use ready-made solutions to the world's leading distributors of free software platforms developed in the framework of the initiative archives open access or adapt the available software solutions on the platform available in the library of automated library information system (ALIS).

Regardless, any of the basic platforms to stop the library from our point of view option ALIS will be the most efficient in terms of price and quality), in the process of creating digital resources necessary to create digital library complex, to justify its structural elements and technological links, to establish their functional interaction.

Leading experts on digital asset management (digital curation) define the following main parts of the organization life cycle full functioning digital resource: conceptualization (planning selection and digital content creation); description and selection (introduction metadata, evaluation of digital material and selection for long-term storage and recovery); management (evaluation of digital material and selection for long-term storage and recovery, compliance with legal requirements, the removal of materials that were not selected for long-term storage and recovery, transfer of materials in the archive, store, data center); save (implementation of measures to ensure long-term storage: check the metadata, the integrity of the digital data, recording data with respect relevant standards); access and use (providing active user access to the digital resource: public, authorized or paid) [12, 13]. According to these principles is a comprehensive solution for managing digital archives (e.g., XEROX DocuShare).

So from a technological point of view, among the main modules (subsystems) of the information system digital Foundation, which will ensure stable and efficient operation, it is appropriate to highlight the following:

- creating digital copies;

- description and accounting of digital resources;
- processing and management of digital resources;
- digital storage of the Fund;
- search and publication of digital resources.

Each of these subsystems has its own principles of operation and requires a separate study of theoretical and technological requirements in the process of processing and organizing digital resources. Let us consider in more detail these key building blocks of the information system digital research library.

Subsystem create digital copies

This subsystem is similar to the technologies of acquisition of the digital library. The first questions which solves this subsystem is the selection of documents for digitization. The criteria for this selection can be a variety of factors, the main ones are:

- unique, rare, or valuable documents, your existing one or a limited number of instances, including manuscripts and pictures.
- damaged, fragile or bulky documents requiring special conditions of access to them. The digitization of these documents contribute to their conservation, provided that user access to their digital copies;
- documents relating to national heritage sites, the introduction of which in a broad cultural circulation is of great importance for the recovery of national remembrance, history and culture of the state.

For digital archives of new information occurs due to the conversion of paper documents into electronic form. For this purpose, depending on the type and format documents, scanners, digital photographic equipment and software for digital image processing.

For the selection of the scanning range of modern research library, in accordance with the practical recommendations of the Administration of the national archives and records USA, you must define the characteristics and main parameters of the original documents, digital copies of which will need to be made [18]. These main parameters of the originals: paper type and methods of making, local deficiencies that require restoration or retouching, tonality, sharpness (sharpness), dimensions. It is necessary to estimate the size of the sheets of paper for handwritten and printed documents, number of documents of each format. From this assessment will depend on the selection of hardware to digitize specific library.

After scanning the received digital information requires processing to bring the scanned images to the appropriate user type. There are many applications which allow you to adjust the scanned image. Mature has a popular software package Adobe Photoshop, which has a huge number of different tools for working with images. Among programs that allow for fast batch processing of digital images, it should be noted ACDSee, Microsoft PfotoEditor (built-in Microsoft Office), Raster Desh, Raster ID, Spotlight. To convert scanned paper documents into electronic formats for further editing system OCR: ABBYY FineReader OCR CuneiForm, Readiris, Microsoft Office Document Imaging. To edit the volume images are programs StudioMAX, Maya, Rhinoceros, SolidWorks and other 3D software packages modeling.

The next step of creating the reference digital copies of documents is to assess the completeness and quality of the received digital images. To the fullness is checking if all the sheets or images of all sides of the original. The concept of quality are first and foremost the sharpness of the obtained image and the color matches the original. For removal of insurance copies of rare and antique documents with color images, digitization of documents occurs with the reference color scale, which is often attached to the kit documentation of professional scanners.

Subsystem description and registration of digital resources

Accounting and inventory digital arrays, including in the sphere of education, science and culture, is the subject of a number of national and international projects: creating a Project Minerva (European Union), <http://www.minervaeurope.org/>; Digital libraries for education, science and culture organization (UNESCO), www.wdl.org.

The accounting system of digital resources provides digital Fund the same functions as the traditional accounting system library preservation Fund. These issues of preservation of digital resources which avalanche grow around the world, has caused the need to describe (create metadata) digital resursu to ensure proper management of digital resources and store them in the future. A very important component of strategies for the preservation of digital heritage is metadata that contains information necessary to document storage process. The preservation metadata is information about the format, structure and use of digital resources; the history of all operations performed on the object storage, including any changes; authenticity; technical history; a history storage; responsibility; rights associated with the storage and so on [16].

Thus, the accounting system of digital resources is closely linked with the choice of metadata to describe the objects created digital Fund. The main requirements of meta descriptions of objects are: completeness of representation of knowledge about the objects; use meta descriptions of objects in other systems; the use of the accepted standard, which is compatible with other systems.

Adding metadata to electronic resources system provides more accurate positioning information about the objects, improves the filtering and selection of knowledge, simplifies and accelerates the processes of access to the necessary programs, servers, and resources.

However, such metadata systems, both universal and targeted for certain types of digital arrays, there are now more than 50 [6]. The most promising and widely used is the standard Dublin core metadata (Dublin Core Metadata Initiative, DCMI) [10]. In the environment of the library community more popular MARC formats in the educational environment - LOM, in the Museum environment - CIMI archive-EAD, etc. in Addition, even within the Dublin core available in different versions, due to different set of qualifiers, the various methodological principles. One solution to this problem does not exist, although the most acceptable solution is the use of a hierarchy of metadata systems, in which the top forms of the Dublin core as the most common system. Other metadata systems can be hierarchical development of the Dublin core, which will provide an unambiguous translation metadata.

However, for digital library resources still certain advantages have solutions based on the family of MARC formats, as it gives an opportunity to use a ready-made descriptions of library documents with electronic catalogs and bibliographic databases [14].

Leading specialists in the creation of electronic archives, such as the administration of the national archives of the United States [18], is recommended for the formation of the digital copies to enter not only descriptive metadata documents, as necessary to provide technical and legal metadata. Technical metadata should reflect the basic characteristics of the image (color, resolution, size, volume), the hardware and software tools which created images and the electronic version of the document. Legal metadata is particularly important for the documents of the archive Fund, where the rules on the copy subject to the same limitations that affect their paper counterparts. One of the significant problems associated with the regulation of access to information faced by libraries in the service of electronic information resources is copyright compliance and rules of the archival service.

Subsystem processing and digital asset management

After creating digitized copies of library documents received digital Fund requires further improvement. The main processes that need to be performed at this stage is the creation of

electronic versions of documents for a full presentation to users and forming digital collections with the appropriate search tools. Let's examine these processes from the point of view of the digital resources of the research library.

Requirements for electronic versions of documents must first meet the criteria of reliability, namely the document in electronic form should reproduce the original. In addition, information in electronic form provides additional functionality from the convenience of manipulation of information: convenient navigation in material and search indexes.

For representing documents in digital libraries can be used in a variety of formats, including the most common are: PDF, DjVu, DOC, TXT, TIFF, JPEG, HTML, XML, FB (FictionBook), CHM, SWF (FLASH document).

May also include special formats for submission of audiovideoinformatsiynny, geodatabase, three-dimensional or animated objects, experimental data and other special types of information.

The image formats for storing copies of insurance documents must meet the following basic technical requirements: support bit depths greater than 24 for monochrome documents; schemes for lossless data compression; support for different color models and the ability to control color, high resolution master copies; openness and detailed documentation, broad support and cross-platform compatibility.

The problem of protection of copyright in libraries successfully solve the system secure remote document viewing such "DefView", "Vivaldi", "SDViewer" and others. Implementation funds protected view allows the libraries to provide support to remote users documents are protected by copyright virtual reading room.

Another important issue is the recognition of documents and storing them in a symbolic format. OCR (Optical character recognition, OCR) can be full of the whole document, it is also possible interim solution - the recognition of certain parts of documents (contents, abstract, title page) or the so-called "black" detection, i.e. detection without correction. The decision in this case turns out to be a very difficult task. Recognition allows character-by-character processing of the text, therefore, different ways of working with texts (search, edit, export, import, and so on). At the same time, digitization of documents translated into a character formats are much more expensive [17]. The OCR program not work perfectly, allowing a certain number of errors, after their work is to produce high quality text and still need a manual proofreading. In the process of translation quality clear text the number of errors is 1% or less. At the same time, in the process of digitizing old books the number of errors reaches 50% [4].

Lyrics for "sub-optimal" for recognition, such as manuscripts and texts with mathematical notation, alternative alphabets and so on, the problem cannot be solved in the near future. Alternative technologies "scan - detection - correction" technology is a manual input (reprint) texts, which in many cases turns out to be cheaper [1].

In the process of preparing digital copies of unique and valuable printed and handwritten objects, in accordance with the requirements of modern professional curating digital resource and protect it in the future you must comply with the technological policy, which provides the ability to store electronic copies for a long time and is based on the following principles: creating a basic (Depository) digital copy of the object; forming on the basis of Depository copies of the required user views the digital object in a variety of convenient formats; the ability iterative return for modification and development of the digital copy of the object in the course of its operation.

The storage subsystem digital Fund

From the point of view of securing large amounts of digital data in recent years, the widespread concept of lifecycle management of information, which is based on the principle of

sharing a common data set into classes depending on the content, frequency and retention period. In accordance with this approach there are three key tasks of storing digital data: on-line access to information backup and archival storage. For each of them used different equipment - specific requirements for storage and access [2].

Online archive implement hosted on a server in a special partition on your hard drive (for small amounts of data), on a separate hard drive or RAID array. To improve the reliability of such systems use backup tools, backup, and mirror disk. RAID arrays are also used to store metadata and user database archive. In the practice of American electronic archives these data are called Master reference File (reference file) [18]. The most promising solution that simultaneously satisfies the requirements of speed and reliability is RAID-5 or RAID 6.

Further development of solutions for the storage of large amounts of data are specialized data storage system (DSS). Modern data storage system should include the following main components: data storage devices (hard drives, CDs and so on); access infrastructure servers to data storage devices; system backup and data backup; software storage management. The main tasks performed by the TAS is a reliable data storage and availability of devices and high-performance servers access to the means of data storage. The main methods used today to implement these tasks can be described in one word - duplication (duplication of data and access to them) [5]. Most modern storage systems carry out the processes of replication and backup automatically.

For organizations with long-term (archival) storage used removable media CD / DVD / UDO and robotic libraries based on them. The most advanced solutions, support superdense optical format (Ultra Density Optical, UDO) and technology Blu-Ray. The popularity of UDO technology relies on the ability to record on one medium 30-60 GB and shelf information 50 years. Robotic library is the next step of creating an electronic archive for long term storage, it is an array of variables discs placed in a special case. The decision on the basis of a robotic library with removable media also makes it easy to solve problems of technical protection of information and to ensure its safe storage [7].

From the point of view of storage of files, electronic documents, there are also two main approaches: in the first of these electronic documents are stored in binary form in the appropriate fields of the database archive, the second approach is based on the fact that the files can be stored separately, and in the database there is only the email address file (specify the path to the file). The first way quickly leads to a number of problems: complications data backup, move to other information platform, the system performance. It can be recommended only for archives with a small physical volume of information.

In addition to the reliability and performance of electronic archive or data warehouse is also important logical storage structure. For the basic structure, it is recommended to take the folder hierarchy with the relevant documents, which will display the logical relationship of the main components of the system.

The search engine and publishing digital resources

The ultimate goal of any project for the digitization of historical, cultural and scientific library and information resources is to implement access to information array for the General population.

Any digital library, designed for a specific range of users and should provide support for particular activities. Thus, to develop high-quality user interface of the digital library should be assessed in the context of their target audience and activities [17]. Developed functional and search capabilities of the system from the point of view of the ergonomic interface of the scientific library should include: navigating all of the available information space, the lexical search, string

search, attribute search (author, title, place of publication, date, etc); the contents of the digital object and its structure; multi mode; support apparatus hypertext and hypermedia links; logging a user session; export information from the system indicating the source of information [1].

Concerning the further development of cognitive capabilities of the interface of the digital library should be considered a range of issues related to the organization of intelligent search, information resource library. The development of such intellectual Toolkit provides a library of semantic technologies (scientific) information processing, when resources are taken into account not only the formal descriptive characteristics of the documents, and first of all their contents. Additional means of disclosure of the content of digital resources, especially for unstructured graphical information digitized pages, is the introduction of the content of documents and annotations. Text information of these additional access points automatically included in search indexes, which significantly extends the search capabilities, the information array.

The use of well-established and well-developed semantic library technologies significantly optimizes efforts to improve search apparatus of a digital library. When forming the digital array is able to use a ready means of thematic search: headings, classifications, thesauri, authority files; authority files: individuals and organizations; scientific descriptions of the manuscripts and archives research library. These tools significantly improve search apparatus and digital Fund provide the user a convenient visualized hypertext additional bibliographic information about its information request, allow the implementation of high-tech means of information search and obtain complete and accurate information.

Conclusions

The proposed modular digital complex was formed and tested in the course of four years within a digital project of the National library of Ukraine named C. I. Vernadsky. As shown by experience, each of the subsystems (modules): creation; descriptions and records; processing and management; storage; search and publish digital resources corresponds to a particular stage of the life cycle of the development of digital scientific library. Development of individual modules with a certain degree of Autonomous tasks created prerequisites for a detailed study of the characteristics of each process stage, a more rational organization and distribution of the professional responsibilities of librarians. Finally, we have created a digital complex and the sequence of modules with a certain number of technological processes provided:

- two main tasks of digital project library science: 1) the creation of digital libraries, open access, and 2) the organization of the insurance Fund of digital copies of documents;
- logical the sequence of operations of the technological cycle;
- the necessary technological infrastructure to perform all of the tasks, which provides digital project scientific library;
- rational use of metadata of digital objects at all technological stages: creation - preservation - records - publishing digital resources;
- reasonable distribution of functional duties and responsibilities of staff involved in the formation of the digital resource;
- developed system of information retrieval in digital library-oriented research;
- compatible with generally accepted standards and integration of digital resources to the international digital projects.

References:

1. Антопольский А. Б. Правовые и технологические проблемы создания и функционирования электронных библиотек / А. Б. Антопольский, Е. А. Данилина, Т. С. Маркова. – М. : ПАТЕНТ, 2008. – 207 с.

2. Корепанов И. Как сохранить архив на десятилетия? [Электронный ресурс] / И. Корепанов // «Журнал сетевых решений (LAN)». – 2008. – № 3. – URL: <http://www.osp.ru/lan/2008/03/4899898/>.
3. Лобузін І. В. Створення бази даних цифрового фонду бібліотеки: основні проблеми та технологічні рішення / І. В. Лобузін // Науково-технічна інформація (НТІ). – 2013. – № 1. – С. 67–62.
4. Масевич А. Ц. К созданию электронных коллекций старопечатных книг в библиотеке Российской академии наук: на примере работы над двумя проектами / А. Ц. Масевич, Е. А. Савельев, А. К. Багажков // Новые технологии в информац. обесп. науки : тр. конф. — М. : Биоинформсервис, 2001. — С.132–140.
5. Меленець А. В. Обрання типу сховища для зберігання документів страхового фонду в електронному вигляді // Системи обробки інформації. – Харків, ХУПС. – 2010. – № 6(87). – С. 199 – 203.
6. Обзор форматов метаданных / укл.: Ю.Е. Хохлов, С.А. Арнаутов ; Институт развития информационного общества [Электронный ресурс] // Российские электронные библиотеки. – URL: http://www.elbib.ru/index.phtml?page=elbib/rus/methodology/md_rev.
7. Петров І. В. Обґрунтування вибору програмно-апаратних засобів архівів електронних документів довготермінового збереження / І. В. Петров, А. М. Стеценко, Н. В. Солоніна // Реєстрація, зберігання і обробка даних. – 2010. – Т. 12. – № 1. – С. 79 – 88.
8. Best Practice Guidelines for Digital Collections [Electronic resource] / edited by Susan Schreibman // University of Maryland Libraries. - 2006.- URL: <http://www.lib.umd.edu/digital/>.
9. Chowdhury S. Usability and impact of digital libraries: a review / Sudatta Chowdhury, Monica Landoni and Forbes Gibb // Online Information Review. – 2006. – № 6. – P. 656 – 680.
10. DDCMI (Dublin Core Metadata Initiative) [Electronic resource]. – URL: <http://dublincore.org>.
11. Guidelines for digitization projects for collections and holdings in the public domain, particularly those held by libraries and archives [Electronic resource] / IFLA ; UNESCO // IFLA. – 2003. – URL: <http://archive.ifla.org/VII/s19/pubs/digit-guide.pdf>.
12. Harvey R. Digital Curation : A How-To-Do-It Manual / R. Harvey. – New York ; London : Neal-Schuman Publishers, Inc., 2010. – 17 p. – (How-to-do-it manuals ; № 170).
13. Higgins S. Digital Curation: The Emergence of a New Discipline // The International Journal of Digital Curation. – 2011. – Vol. 6, № 2. – P. 78 – 88.
14. MARC 21 Format for Bibliographic Data [Electronic resource] // Library of Congress. – URL: <http://www.loc.gov/marc/>.
15. Open Archives Initiative (OAI) [Electronic resource]. – 2007. – URL: www.openarchives.org/.
16. PREMIS Data Dictionary for Preservation Metadata [Electronic resource] : Final Report of the PREMIS Working Group. – 2005. – URL: www.loc.gov/standards/premis/v2/premis-2-0.pdf.
17. Smith A. Strategies for Building Digitized Collections. / Digital Library Federation ; Council on Library and Information Resources. – Washington (DC), 2011. – 35 p.
18. Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Image [Электронный ресурс] / S. T. Puglia, J. Reed, E. Rhodes; U.S. National Archives and Records Administration (NARA). – 2004. – 87 p. – URL: <http://www.archives.gov/preservation/technical/guidelines.html>.