



Программа оцифровки и портал "История МСЭ"

Цель и концепция программы

В целях популяризации истории МСЭ и предоставления широкого доступа к его уникальным архивам для исследователей, а также сотрудников и делегатов Союза Библиотечно-архивная служба МСЭ уже несколько лет работает над программой оцифровки. Цель этой программы состоит в организации, каталогизации, оцифровке и публикации подборки исторических документов на веб-сайте МСЭ.

Приоритеты и принципы отбора: оцифровка важнейших материалов

Чтобы не допустить нарушения авторских прав, Библиотечно-архивная служба МСЭ осуществляет оцифровку исключительно материалов МСЭ. Сканированию каждой партии материалов предшествуют тщательное исследование и оценка библиотечных собраний Союза. Наши приоритеты – это качество и уникальность, а не количество и полнота. Наш принцип отбора обусловлен характером использования и востребованностью.

Ключевые этапы: от проекта к программе

То, что начиналось всего лишь как проект по сканированию документов, трансформировалось в полномасштабную программу оцифровки и в веб-портал со множеством документов и информации, по которым можно отследить путь развития МСЭ как международной организации. Сегодня Библиотечно-архивная служба МСЭ играет активную роль в передаче знаний об истории МСЭ.

❖ 2005–2007 годы: Медленный старт: реакция на запросы

К 2005 году на объединенную справочную стойку Библиотечно-архивной службы стали очень часто приходиться запросы на выдачу информации, содержащейся в документах полномочных конференций МСЭ.

Учитывая эти частые запросы, Библиотечно-архивная служба МСЭ решила действовать на упреждение и провести оцифровку всех изданий этой широко используемой и очень полезной серии документов. Был установлен профессиональный книжный сканер. Так начался наш первый проект по оцифровке.



❖ 2007–2009 годы: Пробы и ошибки и первый веб-портал

Простое сканирование всех документов подряд без тщательного анализа их содержания или состояния оказалось неправильным подходом. Стало ясно, что для создания высококачественных и авторитетных цифровых копий необходимо потратить больше времени и ресурсов на то, чтобы обеспечить правильную идентификацию, оценку, каталогизацию, отбор, подготовку и сканирование оригиналов.



В этот период было принято ключевое решение: документы будут группироваться по тем конференциями, на которых они были подготовлены. Объединяя документы в их контексте, мы даем пользователям возможность лучше понять их содержание.

После того, как документы прошли отбор, оценку и сканирование, нужно было найти способ их передачи нашим пользователям. Было решено создать веб-портал. По итогам этого первого этапа реализации программы в феврале 2010 года был запущен портал "История МСЭ". Отсканированные копии ключевых документов и публикаций, относящихся к полномочным конференциям МСЭ в период с 1865 года по настоящее время, а также к административным конференциям по телеграфии и телефонии 1879–1988 годов, были впервые бесплатно предоставлены всем, у кого есть доступ в интернет дома или на рабочем месте.

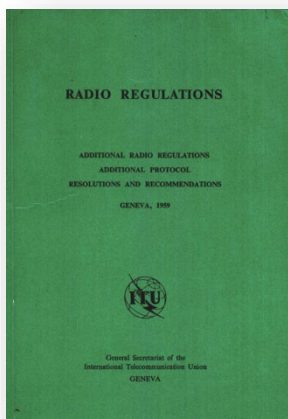
❖ 2010 год: Радио

В 2010 году на втором этапе реализации программы были добавлены ключевые документы и публикации первых международных радиотелеграфных конференций 1903–1947 годов, а также последних всемирных административных радиоконференций (ВАРК) и всемирных конференций радиосвязи (ВКР) 1979–2007 годов. "Сборник документов по радиосвязи", размещенный на портале "История МСЭ", был официально представлен 6 декабря 2010 года во время Всемирного семинара по радиосвязи (ВСП-10).

Новый персонал и новое оборудование позволили нам нарастить объемы обрабатываемых документов более чем в два раза.



❖ 2011 год: Регламент



Третий этап реализации программы в 2011 году был посвящен исторической серии изданий Регламента радиосвязи в период с 1906 года по настоящее время. Регламент радиосвязи имеет сложную историю создания; зачастую как сторонние исследователи, так и персонал МСЭ и делегаты просят отследить его различные редакции. Опять же, принцип отбора материалов для сканирования был обусловлен потребностями наших пользователей. Официальное представление этого сборника состоялось 18 октября 2011 года во время сессии Совета МСЭ 2011 года.

❖ 2012 год: Реорганизация и перепроектирование

На четвертом этапе реализации программы основное внимание уделялось ряду менее объемных серий ключевых публикаций, включая Ежегодный отчет МСЭ и различные публикации со статистическими данными исторического характера (по телеграфии, телефонии и радио). Более того, архитектура и серверная часть портала "История МСЭ" были полностью пересмотрены и перепроектированы в целях повышения качества предоставляемых на портале услуг и расширения его функционала.

❖ 2013 год и далее

В 2015 году МСЭ отметит свое 150-летие. В преддверии этого знакового события мы будем добавлять больше страниц с информационными материалами и расширять нашу коллекцию оцифрованных документов. Одной из приоритетных задач для нас является оцифровка журнала "Электросвязь" МСЭ. Мы также планируем изучать новые способы представления исторической информации, в том числе в виде карт и хронологических схем.

Персонал

За годы своего существования программа МСЭ по оцифровке исторических материалов, которая в начале сводилась лишь к периодическому сканированию документов в условиях очень ограниченных кадровых ресурсов, постепенно трансформировалась в полноценную программу, опирающуюся на настоящую управленческую команду. На сегодняшний день группа по оцифровке состоит из шести человек (на неполной ставке): двух библиотекарей и одного архивиста, которые занимаются отбором материалов, принимают решения и осуществляют интеллектуальный контроль (проводят оценку и каталогизацию, а также указывают метаданные, контекст и страницы веб-сайта для размещения информации). Три технических помощника отвечают за подготовку, сканирование, контроль качества и административную составляющую проекта. На эту группу приходится 75% от общего числа сотрудников нашей Библиотечно-архивной службы.

В 2012 году с привлечением нанятого по контракту персонала была осуществлена разработка нового веб-портала, на которую ушло семь месяцев работы.

Оборудование

До сегодняшнего дня все сканирование выполнялось собственными силами организации. Это является логическим следствием нашего принципа отбора важнейших материалов для оцифровки, для реализации которого на этапе сканирования необходим всесторонний интеллектуальный контроль. В Библиотечной службе используются различные цифровые устройства захвата изображения, в том

числе книжный сканер (черно-белый/оттенки серого), сканер планшетного типа (черно-белый/оттенки серого/ цветной), сканер с автоматической подачей документов (ADF) (черно-белый/оттенки серого/цветной) и сканер пленок/слайдов (негативы/диапозитивы).

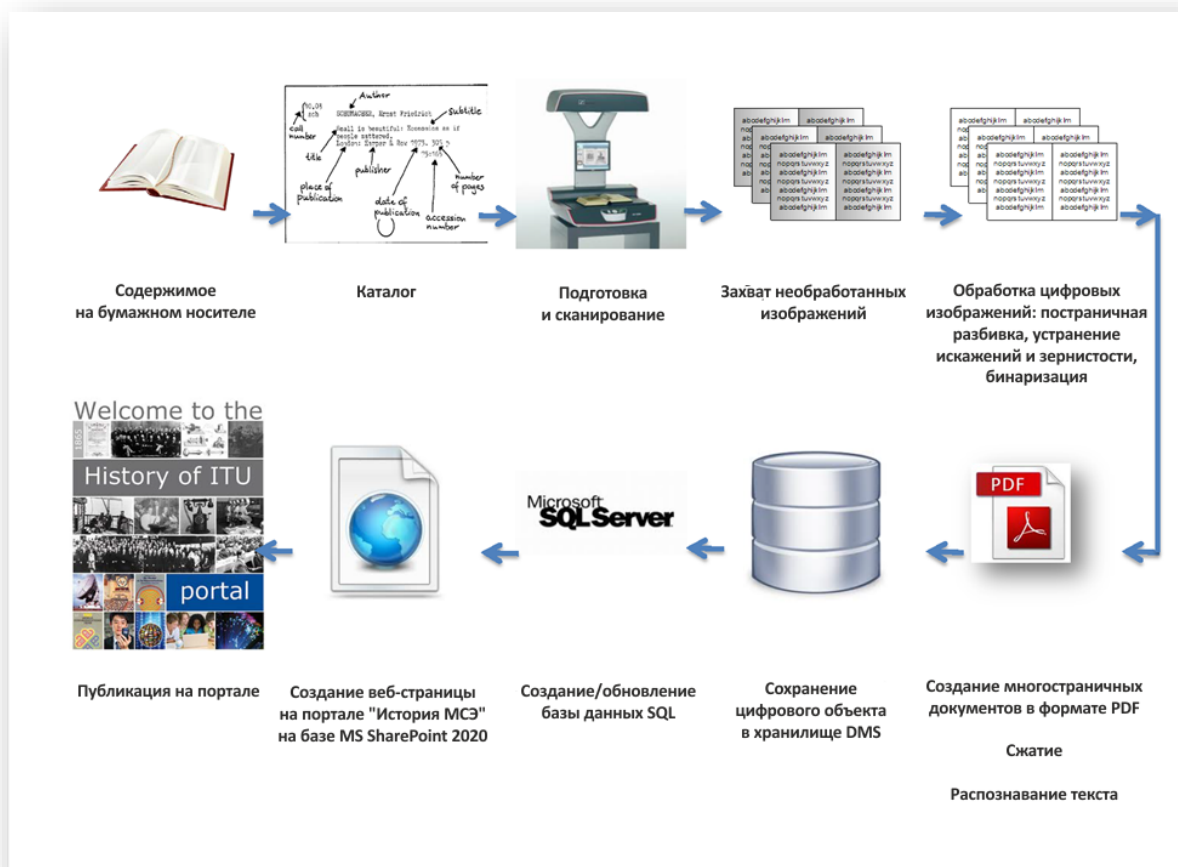
До 2007 года для простого сканирования документов использовалось различное офисное оборудование, включая копировальные аппараты и многофункциональные офисные устройства. Мы также получали поддержку от Службы репрографии, оснащенной профессиональными сканирующими устройствами.

По мере развития программы потребность в более сложной обработке на последующем за сканированием этапе и оптическом распознавании текста стала стимулом к приобретению в 2007 году профессионального книжного сканера (Zeutschel). Качество отсканированных материалов многократно повысилось. В середине 2010-х годов к нашему оборудованию добавился быстрый сканер планшетного типа с ADF (Kodak): объем отсканированного материала вырос вдвое. На сегодняшний день бóльшая часть наших оцифрованных копий (<80%) производится с помощью сканера с ADF.

Тип цифрового оборудования, используемого для захвата изображения, определяется характером подлежащего оцифровке материала, в частности его состоянием и конструктивными особенностями. Перед оцифровкой все сборники просматриваются, и предпочтение отдается тем, которые требуют минимальной обработки в процессе захвата изображения с помощью цифровых устройств. При выборе цифрового устройства захвата изображения используются следующие общие руководящие принципы:

- ❖ большинство документов проходят оцифровку в черно-белом формате. Если на страницах встречаются фотографии или графические изображения, они могут быть отсканированы в оттенках серого и в редких случаях в цвете;
- ❖ на книжном сканере сканируются переплетенные тома, издания очень большого формата или особо хрупкие документы;
- ❖ при возможности – в случае наличия нескольких копий оригинала – мы расшиваем тома с помощью резальной машины и сканируем отдельные страницы на сканере с ADF;
- ❖ ленточные пленки, негативы, смонтированные слайды оцифровываются с помощью сканера негативов.

Рабочий процесс



- ❖ **Отбор:** сотрудники руководящего уровня отбирают материалы для оцифровки.
- ❖ **Библиографическое описание:** описание всех материалов представлено в каталоге библиотеки (Voyager/Marc21/AACR2).
- ❖ **Оценка и подготовка:** материалы проверяются на предмет их состояния, наличия сфальцованных вклеек, графических изображений, требующих сканирования в оттенках серого/цвете; скобы удаляются.
- ❖ **Оцифровка и обработка изображений:** публикации/документы сканируются с разрешением 400 точек на дюйм. Программное обеспечение для обработки изображений (Kodak Capture Pro/Omniscan) осуществляет их захват и обработку: изображения разделяются, выравниваются, устраняются искажения и зернистость; затем изображение преобразуется в двухтональную страницу]. В конце изображение объединяется с документом в формате PDF.
- ❖ **Контроль качества:** полученные файлы в формате PDF проверяются на предмет полноты и точности. В целях повышения качества изображения для лучшей читаемости может выполняться последующая обработка. Как основные, так и справочные копии документов представляют собой многостраничные файлы в формате PDF (изображения в формате TIFF, созданные в процессе оцифровки, сохранению не подлежат).

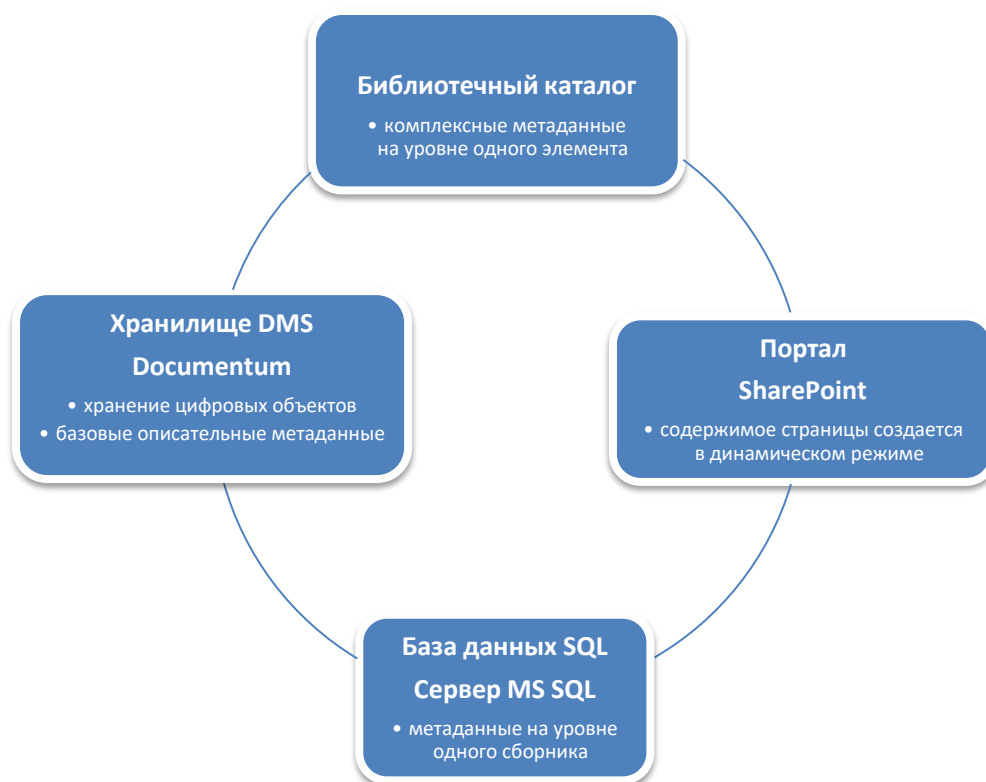
О программе оцифровки и портале "История МСЭ"

- ❖ **Сжатие и распознавание текста:** в целях повышения скорости передачи справочные копии сжимаются; текст в них распознается, и к ним добавляется страница с указанием происхождения цифровой копии. Эти цифровые объекты регистрируются во внутренней системе управления документами МСЭ.
- ❖ **Загрузка файла в хранилище документов (DMS):** все цифровые объекты (основная копия + справочные копии) регистрируются в хранилище документов МСЭ с некоторыми базовыми метаданными.
- ❖ **Отображение на веб-странице:** оцифрованные сборники предоставляются через портал "История МСЭ", разработанный на платформе SharePoint. Веб-страницы создаются динамически, и все их содержимое хранится в базе данных.

Контроль качества осуществляется на нескольких этапах данного процесса: во время захвата изображения, сжатия и распознавания текста в документе формата PDF, загрузки файла и публикации изображений в интернете.

Архитектура

Архитектура цифровых объектов, метаданных и веб-контента имеет достаточно сложную структуру. Объекты и данные регистрируются и хранятся на нескольких платформах и в различных системах.



Многосоставные описания всех материалов (оцифрованных или неоцифрованных) на уровне одного элемента (одно описание на документ) регистрируются в библиотечной системе, созданной на основе программного обеспечения Voyager. Описания увязаны с учетными записями, содержащими названия конференций, имена авторов и темы.

О программе оцифровки и портале "История МСЭ"

Цифровые объекты загружаются в хранилище документов МСЭ, построенное на платформе Documentum. При регистрации в хранилище объектам присваивается ряд самых базовых метаданных.

Веб-сайт МСЭ разработан на платформе SharePoint Server 2010. Веб-страницы портала "История МСЭ" по большей части построены динамически на лежащей в их основе базе данных. Это позволяет нам создавать страницы со сложными списками, которые могут быть отсортированы или отфильтрованы конечным пользователем. База данных SQL содержит метаданные на уровне сборника документов.

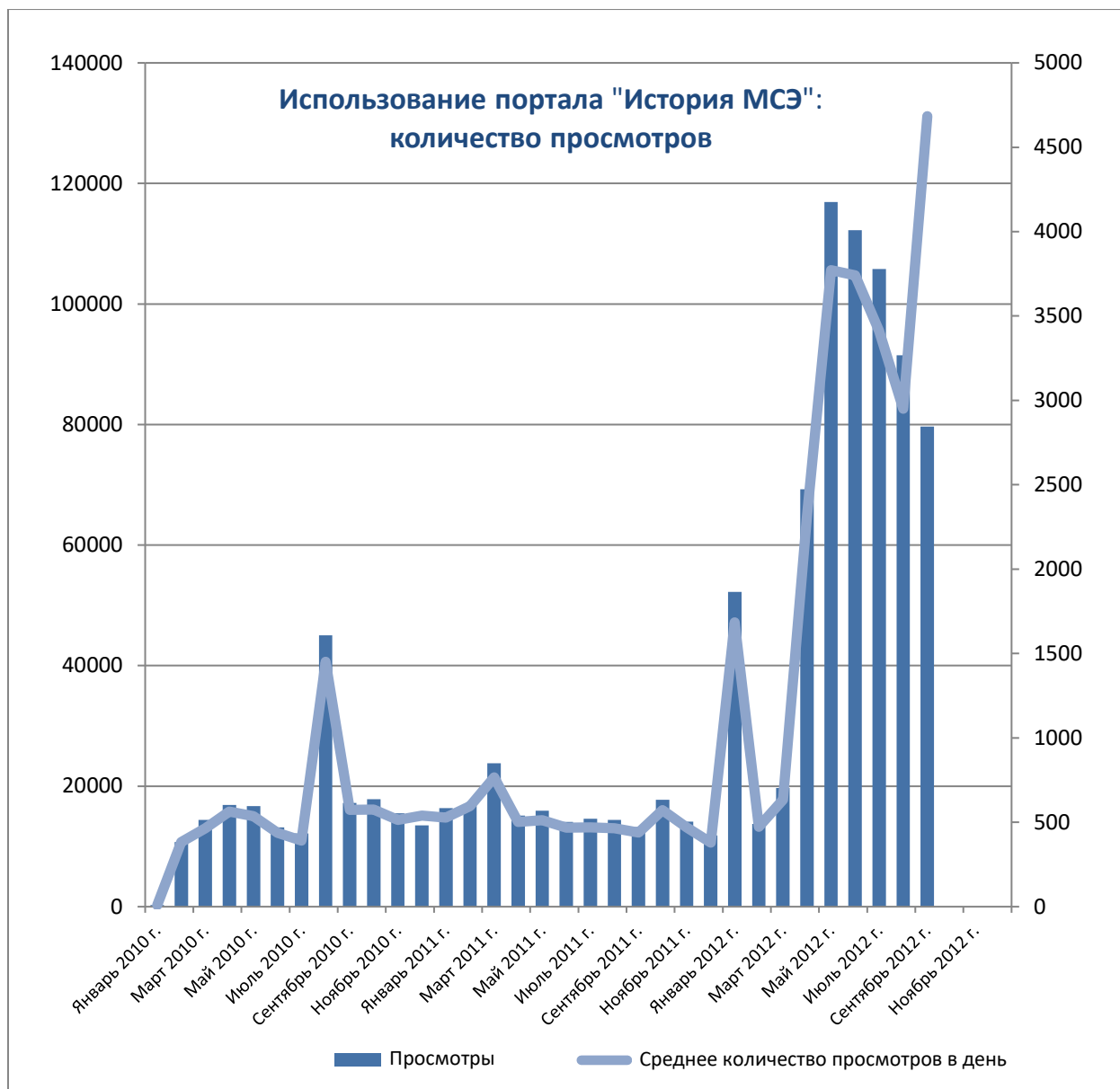
Немного статистики



На сегодняшний день (ноябрь 2012 г.) мы оцифровали около 500 000 страниц, распределенных по различным сборникам. Половина оцифрованных документов относится к сборнику "Радиосвязь".

Посещаемость

Первая версия портала "История МСЭ" была запущена в феврале 2010 года. С тех пор количество просмотров неуклонно растет каждый месяц – с примерно 400 просмотров в день в феврале 2010 года до почти 5000 просмотров в день в сентябре 2012 года. Мы ожидаем, что новая версия портала "История МСЭ", представленная в ноябре 2012 года, будет значительно способствовать удобству пользования веб-сайтом и, соответственно, повысит его посещаемость.



Подготовлено Библиотечно-архивной службой МСЭ

Октябрь 2012 года