

## Zpráva ze zahraniční služební cesty

Jméno a příjmení účastníka cesty	<b>Mgr. Adolf Knoll</b>
Pracoviště – dle organizační struktury	Tajemník pro vědu, výzkum a mezinárodní vztahy
Pracoviště – zařazení	KGŘ
Důvod cesty	<b>Účast na konferenci Asia 2015</b>
Místo – město	<b>Taškent</b>
Místo – země	<b>Uzbekistán</b>
Datum (od-do)	<b>19. – 25. dubna 2015</b>
Podrobný časový harmonogram	19. dubna 2015 – odlet z Prahy 20. dubna 2015 – přílet do Taškentu 21. dubna 2015 – registrace a organizační záležitosti 22. – 24. dubna 2015 - konference 25. února 2015 – odlet a přílet do Prahy
Spolucestující z NK	xxx
Finanční zajištění	Národní knihovna Uzbekistánu (ubytování 21. – 25. 4., transfery), NK VaV 0135 (letenka, stravné + ubytování 2 noci)
Cíle cesty	Účast na konferenci na základě pozvání Národní knihovny Uzbekistánu, přednesení referátu, napsání odborného článku; pokračování dialogu o budoucí spolupráci
Plnění cílů cesty (konkrétně)	<p>Konference Asia 2015 je zaměřena na knihovnická témata a má již v Uzbekistánu tradici. Je zaměřena především na oblast střední Asie, nicméně zahraniční účastníci byli i z dalších zemí bývalého SSSR, Německa, Velké Británie, USA a několika dalších asijských zemí.</p> <p>V rámci konference dochází také tradičně k setkání ředitelů (příp. jejich zástupců) Knihovnického sdružení Euroasie (Библиотечная ассамблея Евразии – БАЕ/BAE - <a href="http://bae.rsl.ru/">http://bae.rsl.ru/</a>), takže přítomni byli z tohoto titulu ředitelé/zástupci takových velkých knihoven, jako jsou Ruská národní knihovna v Petrohradě, Ruská státní knihovna v Moskvě, Prezidentská knihovna B. N. Jelcina v Petrohradě, Národní knihovna Běloruska, Národní knihovna Ázerbájdžánu, Národní knihovna Kazachstánu a další. Referátům knihoven BAE bylo věnováno plenární zasedání druhý den konference. První den jednání probíhala v sekcích, přednesl jsem referát o bezešvé agregaci dat v prostředí digitální knihovny Manuscriptorium. Manuscriptorium agreguje data z téměř 70 zahraničních institucí a vidí své budoucí partnery i ve střední Asii – Národní knihovna Kazachstánu v Almaty již partnerem řadu let je.</p> <p>Automatizovaná agregace dat tohoto typu je výjimečná a klade určité technické nároky i na partnery. Výhodou pro ně je virtuální pospojování orientálních rukopisů z různých zdrojů. V Národní knihovně Kazachstánu došlo k obměně personálu, který zajišťoval kontakt s Manuscriptoriem, s novou paní náměstkyní ředitele Ažar</p>

Zpráva je pracovníkem do mezinárodního oddělení předložena nejpozději při vyúčtování cesty do 2 týdnů po jejím ukončení. Bez cestovní zprávy nebude provedeno vyúčtování. Při výjezdu více pracovníků na tutéž služební cestu s týmž programem lze odevzdat společnou cestovní zprávu.

	<p>Jusupovovou jsme si slíbili na dobrou spolupráci navázat. O zapojení jsme začali jednat i s Národní knihovnou Uzbekistánu, příslušný bod je i v připravované smlouvě o spolupráci, jejímuž uzavření brání pouze několikero změn (noví lidé) v nadřízených orgánech NK Uzbekistánu. NK Uzbekistánu podléhá přímo úřadu vlády země.</p> <p>Bylo mi umožněno prohlédnout si budovu NK Uzbekistánu, jde o novou moderní budovu, celkem nedávno dokončenou (NK ji však nevyužívá celou). Mohu konstatovat vysokou úroveň technického vybavení pracoviště digitalizace vč. robotického skeneru a specializovaných skenerů pro různé formáty a účely; digitalizován byl již překvapivě značný počet dokumentů. Pro zpřístupňování digitalizovaných moderních dokumentů, vydaných za podpory státu, mají v Uzbekistánu značně rozvázané ruce z pohledu autorských práv. Velmi vysokou úroveň mělo i moderní vybavení polygrafie, na ni bude navazovat vybudování pracoviště pro ochranu fondů a pořízení příslušného moderního vybavení. Negativem je vcelku pomalá rychlost internetového připojení, z toho důvodu knihovna poskytuje zdarma svým uživatelům pouze 45 minut denně, delší čas za úhradu.</p> <p>NK Uzbekistánu velmi stojí o naše zkušenosti z budování elektronického národního souborného katalogu. O návštěvě jejich odborníků se předběžně začalo jednat.</p> <p>Z příspěvků uživatelů a rozhovorů s kolegy z dalších knihoven lze také konstatovat, že velká kapacita digitalizace a moderní špičkově vybavená pracoviště jsou i v Národní knihovně Kazachstánu, Ázerbájdžánu, Ruské státní knihovně a samozřejmě Národní knihovně Běloruska (ta je též dlouholetým spolehlivým partnerem Manuscriptoria), takže potenciál pro spolupráci je značný.</p> <p>Pro plnohodnotná jednání s uzbeckými partnery, ale i s dalšími z regionu je takřka nezbytná dobrá znalost ruského jazyka, i když někteří (ale často ne ti klíčoví) mluví anglicky dobře.</p>
Program a další podrobnější informace	<p>Program konference: <a href="http://www.natlib.uz/ru/article/1495/">http://www.natlib.uz/ru/article/1495/</a> Zpráva o konferenci na webu NK Uzbekistánu: <a href="http://www.natlib.uz/ru/news/1507.html">http://www.natlib.uz/ru/news/1507.html</a></p>
Přivezené materiály	<p>Materiály vydané NK Uzbekistánu, konferenční sborník vč. CD, které též obsahuje veškeré zákony (v ruštině a uzbecky) týkající se knihovnictví v Uzbekistánu a příslušné výnosy prezidenta a vlády. Sborník a vydání příspěvků bohužel neodpovídá standardům, kladeným na výsledky RIV v ČR (absence ISBN, abstraktu a bibliografií – přestože to bylo požadováno a autoři je dodali). Můj článek je v příloze.</p>
Datum předložení zprávy	29. dubna 2015
Podpis předkladatele zprávy	
Podpis nadřízeného	Datum: Podpis:
Vloženo na Intranet	Datum: 29. dubna 2015 Podpis:
Přijato v mezinárodním oddělení	Datum: Podpis:

Zpráva je pracovníkem do mezinárodního oddělení předložena nejpozději při vyúčtování cesty do 2 týdnů po jejím ukončení. Bez cestovní zprávy nebude provedeno vyúčtování. Při výjezdu více pracovníků na tutéž služební cestu s týmž programem lze odevzdat společnou cestovní zprávu.

## МЕЖДУНАРОДНАЯ ЦИФРОВАЯ БИБЛИОТЕКА РУКОПИСЕЙ: БЕСШОВНАЯ АГРЕГАЦИЯ ДАННЫХ

Adolf Knoll

Секретарь по научно-исследовательской работе и международному сотрудничеству,  
Национальная библиотека Чешской Республики

Abstract:

The Manuscriptorium Digital Library is Europe's largest digital library of manuscripts providing access to data from more than 120 institutions across Europe incl. ca. 55 most important Czech collections. In March 2015, it contained 43,372 fully digitized documents and more than 337,000 bibliographic records. The documents are not only of European origin (Western, Slavonic, Hebrew, Armenian ...), but there are also data representing Arabic, Persian, Ottoman, Indian and other manuscript traditions. The digital library was launched in 2003 by the National Library of the Czech Republic, in partnership with the Czech AiP Beroun Company. In comparison and in contrast with other aggregation services, it provides seamless access to full digital documents on the fly, reusing thus repositories of partner digital libraries. Users do not need to travel or to navigate from one physical place or digital library to another one to access needed documents. Manuscriptorium provides uniform interface to any data from any digital library that signed the partnership agreement. The internal metadata solution is based on the robust TEI P5 approach and suits both library and researchers' requirements for description and structuring. It aggregates not only fully digitized documents, but also the partially digitized ones and bibliographic records. It contains manuscripts as well as old printed books, historical maps, and historical posters. The documents are mostly represented by digital images and in certain extent also by full texts, TEI-structured and unstructured. The Manuscriptorium larger environment provides also online support to TEI authoring and online metadata validation and upload. For users, it offers interesting personalization tools working across all the collections independently of the physical location of data. Participation in Manuscriptorium opens possibilities for indexation of metadata in large European portal services, both general and manuscript-related, and in the world most-known professional scientific resource discovery services.

Начало оцифровки документов в Национальной библиотеке Чешской Республики тесно связано с первыми годами существования программы ЮНЕСКО *Память мира*. Еще в 1992 году библиотека заключила договор с ЮНЕСКО о разработке одного из пилотных проектов этой программы, результатом которой стал диск со списком самых ценных документов из наших фондов - их образцы были представлены на диске в цифровом виде. Диск вышел в апреле 1993 года [1], и, благодаря скорости своего изготовления, стал первым пилотным проектом всей программы, известным как *Пражский проект*. До 1995 года мы издали на компактных дисках еще две полные рукописи в цифровом виде, которые также стали пилотными проектами программы ЮНЕСКО.

В ходе работы над полными цифровыми изданиями мы стали себе отдавать отчет в том, что отдельные презентации рукописей на дисках перспективно не являются надёжным

решением, так как в этом случае существует очень сильная зависимость между содержанием, т. е. данными и метаданными, и программным обеспечением, необходимым для доступа к данным.

Для нас было ясно, что самый ценный результат оцифровки это данные, т. е. цифровые копии отдельных страниц документов (в нашем случае старых рукописей и редких печатных изданий), для которых нужно сделать все необходимое, чтобы доступ к ним был как можно более открытым и не подлежал частым изменениям в течение времени. Однако, как обеспечить высокую независимость данных в цифровом мире? Это довольно сложный вопрос; для его решения необходимо найти в цифровом мире самые стабильные элементы, которые предположительно не будут резко изменяться и по истечении определенного времени.

Для данных, т. е. для копий страниц документов, таким элементом, вероятно, является такой формат цифрового изображения, который наиболее распространен, но в 90-е годы прошлого века только начинается внедрение Интернета в библиотеки. Это еще далеко не начало бытовой цифровой фотографии, это время, когда жесткий диск (внешняя память) мощностью в один гигабайт был редким исключением, когда частота процессоров компьютеров достигала десятка мегагерц, а объем внутренней памяти постепенно рос от 8, 16, 32 и т. д. мегабайт. Интересно это сравнить, например, с обычным смартфоном сегодня (внешняя память 8 + 16/32 гигабайт, частота процессора, имеющего четыре ядра, 1,3 гигагерц, а внутренняя память - один гигабайт).

Мы тогда решили, что, возможно, именно Интернет сможет стать таким средством, которое будет играть в будущем решающую роль. Надо снова учесть, что это было тогда, когда скорость Интернета измерялась в килобитах за секунду, а не в мегабитах или гигабитах, как это происходит сегодня, а между тем мы уже стали работать с довольно объемными файлами с точки зрения их реального переноса через компьютерную сеть.

Надо сказать, что наше тогдашнее решение идти путем, связанным с Интернетом, был риск, но время показало, что это был правильный выбор. Что нам оно практически дало?

Прежде всего это была ставка на:

- на форматы изображений, рекомендуемые мировым веб-консорциумом, т.е. прежде всего JPEG и как вспомогательные GIF и PNG;
- унификацию форматов цифровых документов как документов комплексных, состоящих из данных (в нашем случае изображений, но это могли быть, например, и звуковые файлы, и файлы полных текстов) и из метаданных, т.е. данных особого типа, вторичных, которые сами по себе являются не целью доступа к цифровым копиям документов, а сопровождением, необходимым для их идентификации (описания документов, библиографические или любые другие), для построения их структуры (структурные метаданные), для их верного воспроизведения во времени (технические, административные метаданные)

Основной вопрос тогда был, как это оформить, чтобы таким образом построенный документ был стабильный в течение долгого времени и чтобы к нему был более-менее открытый доступ без специального программного обеспечения?

Ответ мы нашли на платформе так называемых языков разметки, общим знаменателем которых является SGML. В 1995-96 гг. у нас возникает первый формат комплексного цифрового документа, соединяющий в себе содержательную и формальную разметку, совместимость которых была решена несколько лет позднее на той же общей платформе разработкой языка XML, к нему относящихся инструментов и их общим принятием в компьютерной среде.

Своим форматом документа мы стали пользоваться для регулярной оцифровки рукописей в 1996 г., а периодики в 1999 году. Наш подход был признан рекомендуемым и образцовым подходом ЮНЕСКО для программы *Память мира* в 1999 году [2].

## Рождение цифровой библиотеки

С 1996 года происходила регулярная оцифровка рукописей и редких печатных изданий из фондов нашей библиотеки, но очень скоро стало ясно, что оцифровка может приблизить отдаленные документы пользователю, так что ему уже не надо будет странствовать по различным библиотекам в разных странах, чтобы изучить все рукописи, необходимые для его научной работы. На протяжении веков в рукописном мире, а также иногда и в области очень редких старых изданий случалось, что из-за военных действий или в процессе перепродажи рукописи из одной скриптории или одной библиотеки попадали в разные собрания, которые сегодня географически очень далеки друг от друга.

Прежде всего для нас было важно включить в процесс оцифровки старинных рукописей и другие организации нашей страны, а также попробовать работать на основе нами предложенного стандарта, который между тем позитивно проявил себя, так что в ходе регулярной оцифровки полностью отвечал нашим потребностям. На основе нашего опыта в 1997 году произошло его усовершенствование.

Благодаря Министерству культуры Чешской Республики в конце 90-ых гг. была разработана грантовая программа для поддержки публичных информационных служб библиотек с девятью подпрограммами, из которых одна касалась проектов оцифровки рукописей и очень редких печатных изданий. Подпрограмма была названа в честь программы ЮНЕСКО *Память мира* MEMORIAE MUNDI SERIES BOHEMICA. До сих пор она дает возможность публичным библиотекам различных организаций финансировать до 70% расходов оцифровки путем ежегодных проектов. Благодаря этому, единый стандарт распространился в более, чем 50 библиотеках. Таким образом в подпрограмме приняли участие не только научные и публичные библиотеки, а также библиотеки различных музеев, архивов, монастырей, замков, вузов и др., т. е. большинство важнейших владельцев ценных собраний рукописей и старых печатных изданий.

Очень быстро накапливается то, что мы назвали *критическая масса оцифрованных рукописей*, одновременно появляется стремление получить доступ к их копиям. Сначала это было возможно только в читальном зале Отдела рукописей и старых печатных изданий на определенных компьютерах, но появляется также интерес к доступу в сети Интернет. Сначала мы не предполагали, что это будет необходимо, так как изучение рукописей предполагает специальные знания, и специалистов, ими обладающих, не уж так много.

Параллельно в рамках европейского проекта MASTER шли работы над форматом электронного описания рукописей на платформе TEI. Одним из партнеров этого проекта была также наша библиотека. Когда в начале нулевых новый формат под названием MASTER был опубликован [3], было принято решение сделать его основой нашего нового стандарта для оцифровки рукописей. Для этого надо было к схеме MASTER добавить ещё элементы, позволяющие выразить структуру документа и технические метаданные. Для технических метаданных мы внедрили стандарт, известный сегодня как Z39.87 [4] и рекомендации Color Management Consortium, касающиеся ICC профилей [5] цифровых изображений. Все эти стандарты и решения стали составными частями нового формата (новой схемы/определения типа документа DTD) masterx.xsd/masterx.dtd [6].

Все это дало нам возможность создать цифровую библиотеку Манускрипториум, в которой мы соединили все оцифрованные рукописи и старые печатные издания всех наших чешских партнеров. С первых моментов ее открытия можно было сказать, что она встретила с позитивной оценкой ученых и других пользователей. Такой успех был возможен благодаря технической поддержке чешской компании AiP Veroun.

## На пути к международному сотрудничеству

Сразу же после появления нашей цифровой библиотеки проявились позитивные моменты единого интерфейса. Благодаря этому, а также благодаря факту, что в грантовой комиссии Министерства культуры приняли участие специалисты по рукописным и старым печатным книгам, появилась концепция виртуально объединить документы из различных чешских организаций, чтобы облегчить их сравнение и изучение. Речь шла о вторичных критериях отбора, дополняющих основной критерий, которым мы руководились с самого начала оцифровки, т. е. охране оригиналов путем предоставления пользователям их копий. Таким образом внутри цифровой библиотеки постепенно стали появляться отдельные виртуальные собрания, напр. средневековые музыкальные рукописи или рукописи, происходящие из определенного скриптория.

Интересным начинанием было также ограниченное по времени сотрудничество с Министерством образования в области использования Манускрипториума для изучения древней чешской истории и литературы в школах. Для этого специально были подготовлены программы обучения, затрагивающие отдельные известные произведения и документы..

Территория Центральной Европы и вообще всей Европы сравнительно невелика, поэтому история многих европейских народов и их культурное наследие между собой тесно взаимосвязаны. Результатом этого является факт, что стремление изучать историю и культуру своего народа предполагает использование ресурсов не только отечественных библиотек, но и фондов библиотек иностранных. Европейские рукописи можно разделить на отдельные группы, что касается использования определенных шрифтов; речь идет, например, о западноевропейских рукописях, рукописях славянских, греческих и еврейских, причем не надо забывать, что сведения о Европе находятся также в рукописях, написанных арабским шрифтом (прежде всего османских, но частично и арабских). С чешской точки зрения слово *славянский* здесь не касается славянских рукописей, написанных на чешском, польском, хорватском или словенском языках, т.е.

славянских народов, имеющих средневековую рукописную культуру на своих языках и пользующихся исключительно латинским шрифтом.

С этой точки зрения Манускрипториум, разумеется, является прежде всего цифровой библиотекой западноевропейской рукописной традиции, однако в настоящее время в него включены и тысячи оцифрованных славянских рукописей, а также рукописи других культурных традиций.

Естественный интерес вызывают, конечно, соседние страны, которые в определенные исторические периоды частично или полностью сосуществовали с нами в одном государстве или с нашей сегодняшней точки зрения являлись автономными частями, скажем, такой федерации, каковой была со времен Карла Великого по 1806 г. Священная Римская империя.

Так что неудивительно, что первым зарубежным партнером нашей цифровой библиотеки рукописей стала польская Библиотека Вроцлавского университета. Однако до 2007 года число зарубежных партнеров росло очень медленно. Их было не больше, чем 10 - из Польши, Словакии, Венгрии и Словении, но даже это показало глубокий смысл виртуальной интеграции ценных собраний.

Интеграция оцифрованных документов чешских партнеров и начало интеграции документов иностранных партнеров показали, что техническое решение распространения Манускрипториума за рубеж должно быть построено на других принципах, чем простое накопление данных на наших серверах.

Накопление полных цифровых документов у себя предполагает постоянную работу над внедрением новых и новых документов существующих партнеров. С технической точки зрения это несложно, но проблемы могут появиться в культурно-политической области, так как эта концепция предполагает предоставление своих данных, т.е. изображения оригиналов, чужому партнеру, в том случае, если получено согласие с интеграцией данных на одном месте. Однако практика показала, что именно это согласие зависит от многих субъективных факторов и что было бы легче, если бы изображения, т.е. сама цифровая репрезентация оригиналов, приходила прямо к пользователю из серверов их владельцев, минуя сервер интегратора как промежуточное звено.

Формат нами используемого комплексного цифрового документа делает это возможным при условии агрегации в одной базе данных библиографических описаний документов вместе с метаданными, выражающими структуру рукописей и местонахождение файлов, изображающих отдельные страницы, на серверах партнеров. Остальное потом остается делом конверсий и договоров.

Кроме того, практика также показала, что на первом этапе внедрения конкретного удаленного ресурса нужно активное техническое сотрудничество партнерской библиотеки. И здесь появились проблемы, поэтому мы решили попробовать идти путем европейского проекта, в котором бы приняли участие партнеры из нескольких стран. Это удалось реализовать в рамках финансируемого Европейской комиссией европейского проекта ENRICH (2007–2009), в котором приняло участие 18 партнеров из 11 стран.

## ENRICH: Манускрипториум совершенствуется и растет

Проект был построен на концепции дистрибуированного цифрового документа, который полностью реализуется только во время работы пользователя в интерфейсе цифровой библиотеки как результат его взаимодействия с ней.

С рукописями работают те субъекты, в собраниях которых они находятся, т. е. прежде всего библиотеки, и те, которые ими пользуются для своих исследований. Из этого исходит разница подходов к их описанию в электронной среде: библиотекам нужна регистрация и основная информация о физическом объекте, а для ученых - электронное пространство, в котором можно работать над содержанием рукописи. Если сравнить оба подхода на уровне описания рукописи, так библиотечное описание в принципе короткое и ему достаточно заполнить пространство форматов типа MARC, а описание научное гораздо богаче на всех уровнях и своим способом как бы приближается к транскрипции текста рукописи, поэтому ему близка среда языков разметки, а именно подход кодирования текстов TEI (Text Encoding Initiative). Эта разница вытекает из разного уровня стандартизации печатной книги как культурного объекта (прежде всего современной книги) и рукописи, где ее нет.

Многие библиотеки понимали эту разницу, поэтому для каталогов и цифровых библиотек рукописей с самого начала использовали подход на основе TEI, а именно вышеуказанный формат MASTER (напр. наши партнеры - национальные библиотеки Исландии и Чехии, а также Университетская библиотека в немецком Гейдельберге), в то время как некоторые библиотеки (напр. национальные библиотеки Испании, Румынии, Италии и др.) и в этой области пользовались форматом MARC. Поэтому для агрегации данных нужно было разработать такой подход, который бы внутри агрегирующей библиотеки сохранил все метаданные партнеров. Проблема заключалась в том, что подходы MARC и MASTER были трудно совместимы, так что во время конверсии из одного формата в другой и в обратном направлении терялись бы данные или их гранулярность.

Цифровая библиотека пользуется метаданными в области описания документа для индексации и для изображения контента, поэтому совмещение разных подходов всегда означает на определенном уровне проведение конверсии/трансформации метаданных таким образом, чтобы они были совместимы со способами индексации и изображения контента агрегирующей цифровой библиотеки. Наш первый замысел был воспользоваться контейнером METS, а внутри агрегирующего Манускрипториума все метаданные оставить так (в таких же форматах), как они были получены, причем для индексации и изображения воспользоваться конверсионными стереотипами, характерными для отдельных агрегируемых собраний рукописей.

Однако во время дискуссии с партнерами проекта было решено усовершенствовать формат MASTER таким образом, чтобы он смог дополнительно отобразить структуру документа и содержание изображений, а также, чтобы он стал достаточно гибким для внедрения подхода форматов типа MARC. Таким образом на новой платформе №5 (P5) языка TEI в рамках проекта был создан новый международный формат для описания и структурирования рукописей, спецификация которого носит название проекта, т. е. ENRICH (enrich.dtd) [7].



С точки зрения комплексности речь идет о весьма богатом обменном формате оцифрованных рукописей, который, по решению консорциума партнеров проекта ENRICH, стал внутренним форматом цифровой библиотеки Манускрипториум. Этот формат как таковой входит и в нашу новейшую спецификацию информационных пакетов, которые являются продуктом оцифровки рукописей и старых печатных книг чешских библиотек, пользующихся грантовой системой Министерства культуры Чешской Республики. Содержание этого информационного пакета описано в специальном документе [8], где, кроме описания и структурирования документов, говорится и о конструкции наименований единых идентификаторов рукописей и названий компьютерных файлов с содержанием рукописи, а также о технических метаданных, включая метаданные о настройке сканеров и о характере цифровых изображений (профили ICC, сканированные калибровочные таблицы и их математическое выражение и др.).

После внедрения этого нового формата открылся путь для агрегации данных партнеров европейского проекта и также данных сопутствующих партнеров вне проекта. Для каждого партнера были разработаны специальные правила трансформации метаданных их собственного формата во внутренний формат Манускрипториума. Таким образом для каждого партнера у нас существует отлаженный коннектор, с помощью которого в нашу цифровую библиотеку смогут входить все будущие пополнения цифровых библиотек партнеров. При приеме метаданных партнера мы отдаем предпочтение протоколу OAI-PMH. Если партнер регулярно следит за своим (с нами согласованном) профилем OAI, мы способны с помощью периодического автоматизированного харвестинга комплектовать метаданные обо всех, в том числе и о новых, поступлениях в его цифровую библиотеку. Профиль OAI должен содержать описание и структуру цифрового документа с ссылками на цифровые изображения отдельных страниц на его сервере. Эти изображения являются частью его цифровой библиотеки, а мы с помощью нашей программы ими просто вторично пользуемся. Изображения всегда остаются на серверах партнера, с которым заключен письменный договор. Этот договор дает нам право использовать их для презентации материалов партнера в цифровой библиотеке Манускрипториум.

Если партнер не способен вести с нами коммуникацию с помощью OAI, мы ищем другие решения путем индивидуального подхода к данной проблематике. Практика показывает, что проблемы чаще появляются у партнеров с коммерческими библиотечными системами или с весьма закрытыми и не вполне гибкими системами цифровых библиотек (напр. Digitool), которые не позволяют оптимальную автоматизацию коммуникации.

Наш общий принцип всегда остается одинаковым; он заключается в том, что метаданные мы принимаем в таком виде, в каком они существуют у партнера, а конверсией занимается Манускрипториум, конечно, при условии, что метаданные содержат вышеуказанную информацию и что файлы изображений для презентации в Манускрипториуме имеют форматы, рекомендованные веб-консорциумом, т. е. JPEG, PNG или GIF.

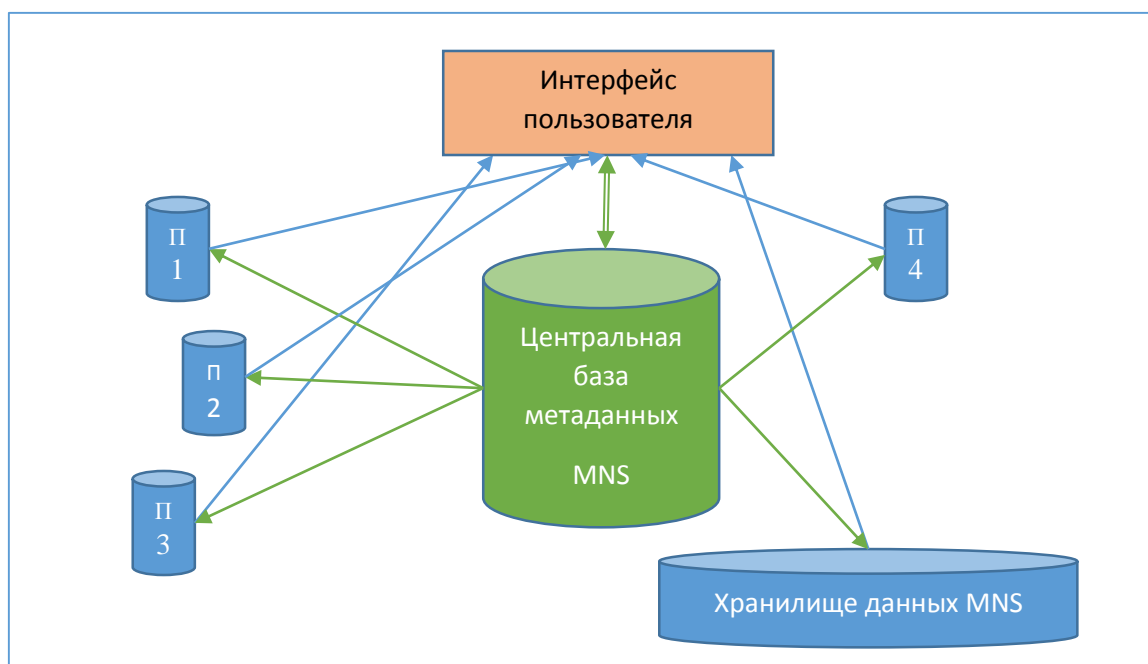


Рис. № 1. Коммуникация партнерских библиотек с Манускрипториумом с точки зрения пользователя

На рисунке выше, где MNS обозначает Манускрипториум, а П1, П2, П3 и П4 несколько партнерских цифровых библиотек, голубым цветом обозначены хранилища данных (изображений) и трансфер данных, в то время как зеленый цвет представляет базу метаданных и трансфер метаданных. Из указанного становится ясно, что данные (изображения страниц рукописей) из всех хранилищ собираются только в интерфейсе пользователя на основе его запросов, минуя систему Манускрипториума. Какие бы не были системы партнеров и их поведение по отношению к пользователю, в Манускрипториуме все документы из разных гетерогенных систем ведут себя определенным унифицированным способом, свойственным цифровой библиотеке Манускрипториум. Такая агрегация данных называется *бесшовная виртуальная агрегация*.

## Манускрипториум сегодня

С точки зрения оцифровки библиотечных документов, Манускрипториум предлагает доступ к рукописям и старым печатным изданиям более, чем пятидесяти чешских владельцев, из которых самым большим является Национальная библиотека Чешской Республики. Все эти организации дигитализируют в соответствии с программой Министерства культуры в рамках годовых проектов. Кроме этого, Национальная библиотека ЧР работает над проектом оцифровки старых книг с компанией Google. Целью этого проекта является массовая оцифровка книг, изданных в 17-ом и 18-ом веках, а также книг 19-ого века, находящихся в нашей Славянской библиотеке. Самые ценные старинные книги и рукописи дигитализируются в рамках программы Министерства культуры. В 2014 году, когда оцифровка только началась, Google дигитализировал около 39 тысяч книг, из которых около 85% составляли книги профиля, совпадающего с Манускрипториумом. Доступ ко всем книгам нашего проекта с компанией Google

открыт сразу же после оцифровки в цифровой библиотеке Google Books, а старые печатные издания (до года издания 1800) доступны также в Манускрипториуме.

Кроме этого, Национальная библиотека работала над проектом оцифровки современных книг (года издания после 1800) и периодики. Этот проект назывался *Национальная цифровая библиотека* и был финансируван из европейских структуральных фондов. Одновременно с книгами, оцифрованными перед началом работы над проектом (проект длился до прошлого года), в цифровой библиотеке Крамериус [9] доступно около 110 тысяч названий книг и 2 тысячи названий периодических изданий - прежде всего из фондов нашей библиотеки и Моравской библиотеки в г. Брно. Многие другие библиотеки, как напр. Библиотека Академии наук и региональные научные библиотеки, работают над своими проектами оцифровки современных изданий, но их рукописи и книги, появившиеся до 1800 года, дигитализируются, как указано выше, и доступны через Манускрипториум.

С помощью Манускрипториума доступны и данные более, чем 70 иностранных библиотек. Не все участники посылают полные тексты оцифрованных документов, некоторые из них предоставляют только библиографические данные. На начало марта 2015 у нас было зарегистрировано 367 тысяч библиографических записей, причем 43 372 документа было полностью оцифровано, из них более 18 тысяч изданий из наших фондов обработал Google. Из 25 тысяч документов, дигитализированных не Google, а другими субъектами, около 80% составляют документы наших иностранных партнеров. Кроме изображений, некоторые представлены также полными текстами.

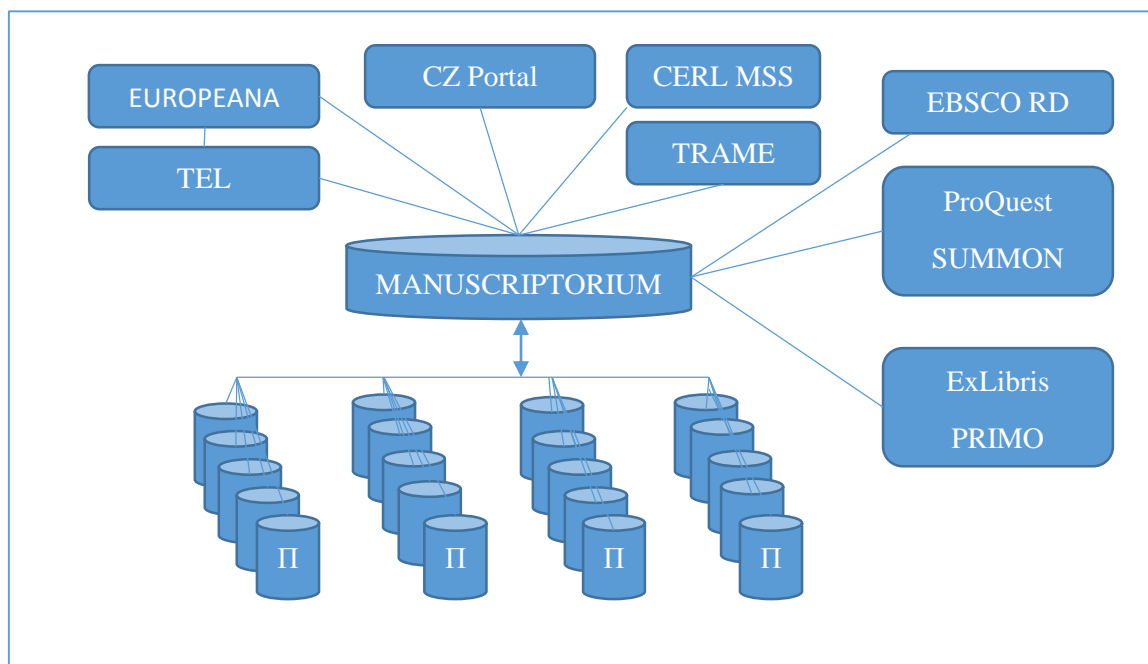
Бесшовная интеграция дает возможность работать в единой среде со всеми документами и их частями независимо от того, где их цифровые копии находятся. Таким образом, любой пользователь, у которого есть учет в цифровой библиотеке Манускрипториума, может в своем личном пространстве создавать свои собственные собрания документов из фондов наших партнеров, т.е. он может виртуально воссоздавать старинные, ныне в пространстве разбросанные, коллекции, которые когда-то принадлежали известным персонам, возникли на одном определенном месте или находились в фондах одной библиотеки.

Он может не только виртуально собирать отдельные названия, но и использовать любую аналитическую часть любой рукописи, которая представлена отдельным файлом. Это в большинстве случаев страница, точнее ее цифровое изображение. Таким образом, читатель имеет возможность из страниц разных документов создавать новые виртуальные материалы, страницы которых будут приходиться из географически отдаленных фондов: напр. первая страница из Университетской библиотеки в Лунде (Швеция), вторая из Музея Западной Чехии в г. Пльзень, третья и четвертая из материалов, хранящихся в фондах Национальной библиотеки Казахстана в Алматы и т. д. В данном случае речь идет о персидских рукописях, которые названные выше библиотеки представляют в Манускрипториуме. Мы думаем, что эта функция персонализации предоставляет интересные возможности напр. для изучения иллюминаций или похожих частей текстов.

Самостоятельной частью являются полные тексты. Они разного качества - начиная от транслитерированных текстов с разметкой по нашим схемам на основе TEI и кончая

текстами, которые являются продуктом автоматического распознавания старых печатных изданий в рамках оцифровки компанией Google.

Особую технологию изображения используют исторические географические карты, специальную разметку исторические музыкальные документы, есть пилотные испытания автоматизированного распознавания графических форм в иллюминациях, информационного поиска с употреблением тезаурусов, разработанных CERL (Консорциум европейских научных библиотек - <http://www.cerl.org/>), который занимается прежде всего созданием базы данных старинных печатных изданий в научных библиотеках Европы.



*Рис. № 2. Манускрипториум в мировом информационном пространстве, где делает доступными данные партнеров (П) европейским культурным порталам, национальному portalу, европейским порталам рукописей и профессиональным научным службам поиска контента*

В более широком (мировом) информационном пространстве Манускрипториум используют различные международные порталы и профессиональные службы, в том числе библиотечные и культурные, как напр. EUROPEANA, Европейская Библиотека TEL или рукописный портал Консорциума Европейских Научных Библиотек CERL-MSS, а также такие известные службы поиска научных информационных ресурсов, как напр. EBSCO Discovery Service или The Summon® Service. Практически это значит, что данные любого нашего партнера, включая самый маленький региональный музей, становятся автоматически составными частями важнейших научных и культурных мировых информационных ресурсов.

Очень интересно посмотреть на статистические данные пользования Манускрипториумом: у него более, чем 110 тысяч индивидуальных пользователей в год, что гораздо больше, чем число пользователей всех наших лицензированных подписных ресурсов вместе взятых, среди которых, например, и *ebrary Academic Complete* с 120 тысячами названий прежде всего англоязычной научной литературы, а также многие

другие. Совершенно очевидно, насколько важно найти оптимальное место в широкой информационной среде: после подключения Манускрипториума к европейскому portalу культурного наследия EUROPEANA, около 27% конкретных запросов на просмотр индивидуальных рукописей и книг в Манускрипториуме генерируется этим порталом, причем половина наших пользователей приходит из-за рубежа (практически со всего мира).

Вышеупомянутое означает, что рукописи и старые печатные издания пользуются большим спросом. Этот факт становится еще очевиднее, если рассмотреть статистику цифровой библиотеки *Google Books*, в принципе самой большой цифровой библиотеки мира. Анализируя 50 названий из нашего фонда старых печатных книг, пользующихся наибольшей популярностью в этой цифровой библиотеке, можно сделать вывод, что только за последних шесть месяцев ими заинтересовалось более, чем 145 тысяч читателей. Это бесспорно подтверждает, что имеет глубокий смысл открывать онлайн доступ к старинным памятникам мировой культуры.

Все данные Манускрипториума находятся в свободном доступе; любой пользователь может в нем свободно зарегистрироваться и создавать здесь свои виртуальные коллекции и документы, доступ к которым он может разделять со своими коллегами по науке. К своим коллекциям и документам пользователь может присоединять собственные заметки и комментарии, причем все эти дополнения ведут себя в Манускрипториуме в соответствии с правилами этой библиотеки. Отдельные модули системы Манускрипториум предлагают также онлайн поддержку созданию описаний и структур документов на основе платформы TEI P5 (модуль *m-tool*) [10] и загрузке метаданных в так называемый Манускрипториум для кандидатов (модуль *m-can*), но спрос на пользование этих модулей в последнее время существенно понизился, так как агрегация данных из других библиотек работает с уже готовыми цифровыми документами, метаданные которых надо трансформировать во внутренний формат Манускрипториума, а показ данных гомогенизировать для единого интерфейса.

В настоящее время (март 2015) происходит миграция Манускрипториума (<http://www.manuscriptorium.eu> или <http://www.manuscriptorium.com>) из версии №2 на версию №3, которая сейчас работает параллельно на URL <http://v3.manuscriptorium.com>. Новая версия работала в режиме тестов почти целый год и реагирует на комментарии пользователей. Одновременно будет доступна и версия для мобильных устройств, а в регулярном режиме появятся также некоторые функциональности, которые до сих пор работали только в испытательном режиме.

- [1] Paměť světa - Memory of the World - Mémoire du Monde / Сценарий Adolf Knoll, Введение Alena Richterová and Adolf Knoll. Praha, Národní knihovna - Albertina icome Praha, 1993. 1 CD-ROM
- [2] Digitization of Rare Library Materials: Storage and Access to Data / Менеджмент проекта Adolf Knoll и Stanislav Psohlavec. Авторы: Adolf Knoll, Stanislav Psohlavec, Jan Mottl, Jan Vomlel, Tomáš Mayer, ... Prague, National Library - Albertina icome Praha, 1999. 1 CD-ROM, Memoriae Mundi Series Bohemica
- [3] Reference Manual for the MASTER Document Type Definition. Discussion Draft. [Электронный ресурс]. Edited by Lou Burnard for the MASTER Work Group, URL: [http://www.tei-c.org/About/Archive\\_new/Master/Reference/oldindex.html](http://www.tei-c.org/About/Archive_new/Master/Reference/oldindex.html) (дата обращения: 10. 03. 2015)
- [4] ANSI/NISO Z39.87-2006 (R2011) [Электронный ресурс]. Data Dictionary - Technical Metadata for Digital Still Images. Bethesda, NISO Press, 2006. 107 стр., URL: [http://www.niso.org/apps/group\\_public/download.php/6502/Data%20Dictionary%20-%20Technical%20Metadata%20for%20Digital%20Still%20Images.pdf](http://www.niso.org/apps/group_public/download.php/6502/Data%20Dictionary%20-%20Technical%20Metadata%20for%20Digital%20Still%20Images.pdf) (дата обращения: 10. 03. 2015)
- [5] см. <http://www.color.org/iccprofile.xalter> (дата обращения: 10. 03. 2015)
- [6] см. <http://digit.nkp.cz/MMSB/1.1/msnkaip.xsd> (дата обращения: 10. 03. 2015)
- [7] см. <http://projects.oucs.ox.ac.uk/ENRICH/> (дата обращения: 10. 03. 2015)
- [8] Definice digitálního dokumentu pro potřeby zpřístupnění a trvalého uložení v podprogramu VISK6. Verze 1.0 [Электронный ресурс] / Olga Číperová, Štěpán Černohorský, Tomáš Klimek, Tomáš Psohlavec. 117 стр. URL: [http://www.manuscriptorium.com/sites/default/files/docs/manuscriptorium\\_visk6\\_definice.pdf](http://www.manuscriptorium.com/sites/default/files/docs/manuscriptorium_visk6_definice.pdf) (дата обращения: 10. 03. 2015)
- [9] см. <http://kramerius4.nkp.cz/> (дата обращения: 10. 03. 2015)
- [10] M-Tool – user’s manual. Version 2.0 online [Электронный ресурс]. URL: [http://www.manuscriptorium.com/apps/m-tool/docs/m\\_tool\\_manual\\_en.pdf](http://www.manuscriptorium.com/apps/m-tool/docs/m_tool_manual_en.pdf) (дата обращения: 10. 03. 2015)