



## Наведение мостов: от пректа Europeana Libraries к проекту Europeana Newspapers

**Susan K. Reilly**  
LIBER  
The Hague, Netherlands  
E-mail: [susan.reilly@kb.nl](mailto:susan.reilly@kb.nl)

Session:

**119 – Users and portals: digital newspapers, usability, and genealogy – Newspapers Section with Genealogy and Local History Section**

### Краткое описание:

*Исследования показывают, что легкость доступа и особенно принцип органа комплексного обслуживания, являются любимыми для исследователей, поскольку это четкий и эффективный способ доступа к цифровому содержанию. Как и легкость доступа, качественно заверенное содержание также исключительно важно.*

*Проект Europeana Libraries<sup>1</sup> предназначается и для выпуска из 5,1 миллионов изображений, книг, видео и очерков, и для статей непосредственно из 19 европейских ведущих исследовательских библиотек. Источник данных означает, что в метаданные может быть помещена конфиденциальность, и качество изображений благодаря легкости доступа будет гарантирована через отдельный поиск всех объектов на веб-сайтах The European Library<sup>2</sup> и Europeana<sup>3</sup>.*

*Сотрудничество было положено в основу содействия совместной работе и инновациям. Устойчивое объединение инфраструктуры и возможности поиска по полному тексту, созданные в проекте Europeana Libraries, уже позволяют применять новую форму содержания через проект Europeana Newspapers<sup>4</sup>. Этот проект создаст 29 миллионов страниц газетного содержания по всей Европе, доступного через платформы The European Library и Europeana.*

### Введение:

Исследования показывают, что легкость доступа и особенно принцип органа комплексного обслуживания, являются любимыми для исследователей, поскольку это четкий и эффективный способ доступа к цифровому

содержанию. Как и легкость доступа, качественно заверенное содержание также исключительно важно.

Данная статья обрисовывает некоторую работу проекта Europeana Libraries<sup>5</sup>, который создаст полнотекстовое содержание национальной и исследовательской библиотеки, доступное для поиска на единственном портале. Статья иллюстрирует мотивацию и пользу совместной работы организаций для получения общего представления. Она также кратко передает мотивацию позади и то, чего мы достигнем в проекте Europeana Newspapers -- проекте, чье преимущество от портала и возможности поиска полного текста развиваются посредством Europeana Libraries.

### **Первый камень закладываемого фундамента:**

Europeana Libraries -- это двухлетний проект, начавшийся в январе 2011 года. Идея проекта пришла на стадии определения потребности в отдельном сборнике для европейских исследовательских библиотек, как национальных, так и университетских. Такие библиотеки в прошлом работали вместе, чтобы обеспечить тематическое содержание для Europeana, портала культурного наследия, посредством проекта Europeana Travel<sup>6</sup>. В рамках этого проекта было объединено содержание на тему путешествий или туризма через два отдельных сборника: для национальных библиотек и другого, учрежденного специально для собирания содержания из других библиотек проекта. Проект имел большой успех и с точки зрения обеспечения высокого качества цифровому содержанию для Europeana, и с точки зрения установления совместной работы между национальными и другими исследовательскими библиотеками, но он поставил знак вопроса над способностью длительного использования двух отдельных сборников.

Проект Europeana Libraries путем открытия национальной библиотечной службы сбора, The European Library, адресует этот вопрос выживаемости исследовательским библиотекам. Он использует эту службу чтобы собрать критическую массу ценного содержания из Европейских исследовательских библиотек. К концу проекта в декабре 2012 года больше 5,1 миллионов объектов, включая 1 200 фильмов и видеоклипов, 850 000 изображений и 4,3 миллионов текстов (книг, журнальных статей, очерков, писем) будет поглощено из исследовательских библиотек. Большая часть этого содержания -- это полный текст редкой ценности для исследователей. Для того, чтобы максимизировать потенциал этого содержания, проект также начинает развивать возможности полнотекстового поиска и поиска по portalу, который обеспечит определенным для исследования инструментом.

### **Сотрудничество:**

Соединить вместе содержание исследования и национальных библиотек также способствует соединению ключевых сетей European library, называемых CENL, CERL и LIBER.

Конференция европейских национальных библиотек (CENL) представляет 48 национальных библиотек и в настоящий момент владеет сборником The

European Library, который является единственным сборщиком домена European library. До сих пор он имеет собранное содержание только из национальных библиотек.

Внимание Консорциума европейских исследовательских библиотек (CERL) сосредоточено на усовершенствовании доступа к напечатанному европейскому наследию и его эксплуатации. Консорциум имеет 33 полноправных члена из исследовательских и национальных библиотек. CERL испытывает особый интерес и имеет особый опыт в индексировании и метаданных.

LIBER, ассоциация для европейских исследовательских библиотек, насчитывает свыше 420 членов (национальных, университетских и других исследовательских библиотек) из Европы и ее соседей. 19 членов LIBER обеспечивают содержанием проект и, в конечном счете, служба развития внутри Europeana Libraries будет распространяться на всех членов LIBER.

Europeana Libraries -- это первая возможность, которую использовали эти три организации, чтобы работать вместе над одной сильной общностью, и это то, что все их члены имеют содержание, которое является ценным исследовательским материалом, и хотят создать свое содержание, доступное и используемое исследовательскими сообществами. Исход выживаемости проекта будет гарантирован посредством применения рассечения этой общности на действующие и согласованные ценные предложения и построение отношений между сетями.

### **Определение ценности сотрудничества:**

Ценность проекта заключается не только в создании отдельной службы сбора для библиотек, хотя это и значительный аспект, но также и в том, что он потенциально предлагает донести исследовательское содержание из библиотек до исследователей по всему миру. Потенциально он расширяет богатство коллекций и национальных, и исследовательских библиотек за пределы границ, установленных исследовательскими сообществами, и регионов. Проект использует общую репутацию библиотек, как доверенных поставщиков качественной информации и хороших метаданных. Хорошо установленным фактом является то, что библиотеки непосредственно связаны с книгами<sup>7</sup>. Обеспечение полнотекстового содержания коллекций оцифрованных книг параллельно другим цифровым содержанием, таким как изображения, видео и аудио файлы, не говоря уже о научном содержании, таком как статьи и очерки, означает, что исследователи могут получить более ценные результаты поиска. Расширяя обзор этого содержания, библиотеки могут повысить эффективность своих значительных инвестиций в оцифровку<sup>8</sup>.

Для исследователя ценность заключается в возможности доступа и поиска критической массы культурного наследия и связи результатов исследования в одном месте. Изучение показывает, что исследователи проводят более комплексное исследование, а также располагают меньшим количеством времени для своей исследовательской деятельности<sup>9</sup>, следовательно, принцип органа комплексного обслуживания является привлекательным предложением.

## **Designing the Portal:**

### **Проектирование портала:**

Большая часть содержания, собранная посредством проекта Europeana Libraries, будет доступна на портале Europeana, но учитывая происхождение собранного содержания, портал The European Library был реконструирован с учетом гуманитарных исследователей, как конечных пользователей. Такое предложение чрезвычайно важно сообществу гуманитарных исследований, для кого определение исследовательской информации является комплексом: *«Сообщество гуманитариев нуждается в критической массе цифровых ресурсов и в общих инструментах, службах и хранилищах, если они идут дальше 'проектов бутиков' для твердой основы теории и метода»*<sup>10</sup>.

Несколько циклов семинаров и тестирований конечных пользователей, а также разработка статистической документации попали в проектирование нового портала, имеющего сейчас следующие особенности:

- возможность полнотекстового поиска
- возможность просмотра записи необработанных метаданных индивидуальных объектов с целью получить в конечном итоге доступ к большим базам данных для исследований.
- увеличение широко использовало предметные заголовки CERIF. Это сделало исследование возможным через собрание объектов, которые связаны общей темой или временным интервалом.
- развитие общеевропейской коллекции в сотрудничестве с национальными, исследовательскими и университетскими библиотеками в Европе. Это распространение действующих виртуальных выставок<sup>11</sup>, которые отображают содержание ряда источников всей Европы.
- временная последовательность показывает наличие поиска по определенному периоду на протяжении столетий.
- программный интерфейс приложения (API), который позволит проанализировать содержание базы данных и отобразить ее в контекстах вне портала The European Library. Это означает, что исследователи могут переносить их содержание в собственную среду исследования и изучать новые способы использования содержания.
- прямой экспорт записей в службы управления популярными ссылками, такими как Mendeley и Zotero.

Запущенный однажды, портал будет постоянно реконструироваться в соответствии с развивающейся исследовательской практикой в гуманитарных науках и цифровых гуманитарных науках. Дальнейшее изучение будет проводиться в том, как исследователи могут эксплуатировать и использовать это уникальное содержание, а стратегия содержания будет развиваться в соответствии с этим.

### **Наведение мостов:**

Так же как и служба сбора, портал и возможности полнотекстового поиска, развивающиеся в проекте Europeana Libraries, используются сейчас для того, чтобы показать новый тип содержания: газеты. Недавние улучшения в OCR,

произведенные в проекте IMPACT<sup>12</sup>, сейчас применяются к содержанию газет из 12 национальных и исследовательских библиотек Европы. 18 миллионов страниц газет будет усовершенствовано и станет доступно на порталах Europeana and The European Library. Проект The Europeana Newspapers начинает обращаться к особо специфическим трудностям, которые создают полнотекстовые документы старых газет, доступные для поиска. Это сделает использование методов усовершенствования для сегментации OCR, OLR/статьи, извлечение сущностей и классовое распознавание страницы. Большинство из того, что было улучшено и изучено посредством Europeana Libraries, будет сейчас применяться к Europeana Newspapers:

### 1. Собираение:

Может быть идентифицировано четыре типа существующих цифровых коллекций газет:

- а) Изображения с только структурными метаданными
- б) Изображения со структурными метаданными и полным текстом для поиска (OCR)
- в) Изображения со структурными метаданными, распознаванием статей (OLR/отслеживание статей) и OCR
- г) Изображения со структурными метаданными, OLR, OCR и семантическим обогащением.

Все доступные данные будут собраны в The European Library. Данные будут преобразованы в EDM -- модель данных Europeana -- и распределены для Europeana.

### 2. Стандартизация метаданных.

В настоящее время используется разнообразие стандартов метаданных. Чтобы улучшить доступ к цифровому содержанию, должны быть приняты общие стандарты. Все существующие форматы метаданных будут идентифицированы и сообществу будут обеспечены лучшие методы решений.

### 3. Возможности лучшего отображения.

Сделать газеты простыми для поиска и привлекательно представить их онлайн на данный момент представляется сложным. Проект The Newspapers Online оценит работу, проведенную проектом Europeana Libraries, и будет основываться на этой работе, чтобы определить соответствующие требования к поиску и представлению, которые могут быть использованы веб-сайтом Europeana.

Сейчас текст газет прошлого, еще 18-ого века, будет полностью доступным для поиска онлайн. И, более того, пользователи смогут просматривать эти документы в контексте, наряду с изображениями, фотографиями, существенными как книги и статьи. Посредством представления содержания газет таким образом, могут быть созданы новые коллекции, и могут быть открыты двери для новых видов исследования и совместной работы.

Существуют и другие преимущества работы этого проекта как процедур, с которыми будет обновлено содержание газет, включая OCR, OLR/отслеживание статей, NER и классовое распознавание страницы. Для каждой из этих технических задач лучшей практической рекомендацией будет распознанная и изданная. Это будет огромной пользой для широкой сети библиотек с сетями CERL, CENL и LIBER. Это поможет снизить стоимость

проектов по оцифровке газет и увеличить доступность коллекций цифровых газет сейчас и в будущем.

### **Заключение:**

Проект Europeana Libraries был лучшей практической сетью, которая предназначалась для весьма практических потребностей отдельного сборщика из European libraries. При этом он объединил ключевые библиотечные сети. Работая в направлении общего видения, эти сети создали ресурс, который мог иметь огромную потенциальную ценность для исследовательских сообществ. Он также создал условия для национальных и исследовательских библиотек работать вместе более плавно, основываясь на видении, чтобы соединить содержание и исследователей. Содержание европейских библиотек богатое и разнообразное. Это особенно касается активов газет. Ввод этих активов онлайн, в то время, когда технологии были усовершенствованы, чтобы значительно использовать полный текст, создает огромную возможность исследователям для привлечения новых коллекций и их взаимодействия между богатым материалом культурного наследия Европы. Это также олицетворяет то, как организации, сети и учреждения, работая вместе, могут добиться инновационных результатов, повысить эффективность и обеспечить доступность. Такая работа будет иметь простирающийся эффект, не только для библиотек или даже для доступности европейского культурного наследия, но и для каждой страны в мире с массой напечатанного культурного материала.

### **Литература:**

Bulger, Monica, et al., *Reinventing Research? Information Practices in the Humanities*. London: Research Information Network, 2011.

Christine L. Borgman. "The Digital Future is Now: A Call to Action for the Humanities" *Digital Humanities Quarterly* 4.1 (2010). Available at: <http://works.bepress.com/borgman/233>

De Rosa, Cathy et al. 2005. *Perceptions of Libraries and Information Resources: A Report to the OCLC Membership*. Dublin, Ohio: OCLC. <http://www.oclc.org/reports/2005perceptions.htm>.

European Commission, Maurice Lévy, Elisabeth Niggemann, and Jacques de Decker. *The New Renaissance*. Brussels: European Commission, 2011. [http://ec.europa.eu/information\\_society/activities/digital\\_libraries/doc/reflection\\_group/final\\_report\\_%20cds.pdf](http://ec.europa.eu/information_society/activities/digital_libraries/doc/reflection_group/final_report_%20cds.pdf)

Europeana (2012) Retrieved from <http://www.europeana.eu> on 19<sup>th</sup> April 2012.

Europeana Libraries (2012), Retrieved from <http://www.europeana-libraries.eu>, on 9<sup>th</sup> March 2012.

Europeana Travel (2012), Retrieved from <http://www.europeanatravel.eu/> on 19<sup>th</sup> April 2012.

IMPACT, Retrieved from <http://www.impact-project.eu/> on 19<sup>th</sup> April 2012.

The European Library(2012), Retrieved from <http://www.theeuropeanlibrary.org> on 9<sup>th</sup> March 2012.