

This is a preprint of a conference paper presented at Румянцевские чтения [Rumyantsev Readings] 2010, held April 20–22, 2010, in Moscow, Russia.

**Коллективный репозиторий Hathitrust как хранитель и  
распространитель всемирной книжной культуры**

Кевин Хокинс (Мичиганский университет, г. Энн-Арбор;

Королевская ирландская академия, г. Дублин)

Джереми Йорк (Мичиганский университет, г. Энн-Арбор)

Hathitrust - это электронный репозиторий совместного пользования, созданный в результате сотрудничества ряда исследовательских библиотек для хранения и обеспечения доступа к их электронным коллекциям.<sup>i</sup>

Проект был запущен в октябре 2008 года 25-ю американскими исследовательскими библиотеками, которые использовали свой совместный опыт в создании электронных фондов для организации устойчивой и расширяемой инфраструктуры, способной обеспечить хранение, управление и доступ к их коллекциям. Беспрецедентный по своим масштабам этот репозиторий на декабрь 2009 года включает пять миллионов оцифрованных томов, а в перспективе объём увеличится до 8 миллионов в 2010 году.<sup>ii</sup> Созданный как альтернатива Google Books, Hathitrust благодаря своей управляемости и схемам финансирования является образцом для создания и поддержки хранилища продуктов массовой оцифровки. Входящая в него система наблюдения за правами демонстрирует, как частное предприятие и государственное учреждение смогли бы работать в рамках закона об авторских правах, защищая при этом интересы и владельцев авторских прав и исследователей. Но наибольшую пользу принесут исследователям по всему миру печатные коллекции организаций-партнёров — в дополнение к Hathitrust в цифровой форме по мере оцифровки — выполняя задачу Hathitrust - «способствовать общей выгоде при помощи

сбора, организации, сохранения, передачи и совместного использования зафиксированных человечеством знаний».<sup>iii</sup>

### **Что такое HathiTrust?**

HathiTrust — это партнёрство исследовательских библиотек, создавших репозитарий с одноимённым названием, которое является крупнейшей инициативой хранения цифровых данных в библиотечной истории, направленная на сохранение и предоставление доступа к миллионам томов из печатных коллекций этих библиотек.<sup>iv</sup> Репозитарий находится в совместном владении и под совместным управлением организаций-партнёров, которые оплачивают поддержку инфраструктуры в пропорции к объёму их контента в репозитарии.<sup>v</sup> HathiTrust использует в качестве основы репозитарий, разработанный в библиотеке Мичиганского университета (прежде именовавшийся MBooks) для хранения изображений страниц, получаемых благодаря сотрудничеству с компанией Google, и предусматривается продолжение развития репозитария для достижения соответствия со стандартами цифрового хранения и процессов проверки.<sup>vi</sup>

Несмотря на то, что больше количество контента HathiTrust защищено авторским правом и следовательно не общедоступна, значительная часть материалов всё же находится в свободном доступе, а небольшая, но увеличивающаяся часть материалов предоставлена в общий доступ по соглашению с владельцами прав.<sup>vii</sup>

Большая часть контента, загруженного в HathiTrust на данный момент, получена от компании Google в рамках библиотечного проекта Google Books; также, однако, разрабатываются средства для интегрирования в HathiTrust контента, исходящего не от Google.<sup>viii</sup>

### **Чем HathiTrust отличается от World Digital Library, Open Library и Google Books?**

HathiTrust, World Digital Library и Open Library — это проекты с глобальными амбициями, но они значительно отличаются друг от друга по масштабам и целям. Тогда как HathiTrust нацелен на сохранение коллекций исследовательских библиотек и предоставление доступа к ним<sup>ix</sup> (включая ограниченный доступ к материалам, защищённым авторским правом), в World Digital Library хранятся только избранные материалы-первоисточники из национальных библиотек и других организаций-партнёров. HathiTrust в большей степени похожа на Open Library, которая была запущена в 2007 году, потому что оба эти хранилища включают сотни тысяч полнотекстовых материалов в открытом доступе. Однако цель Open Library – «отображать в Интернете запись о каждой книге, которая когда-либо была опубликована», создавая не «каталог для общего межбиблиотечного использования» (как WorldCat), но «каталог для общественного использования».<sup>x</sup> Страничка каждой книги может включать ссылки на внешние источники информации, оцифрованные копии, найденные в Интернете, и ссылки, по которым можно купить книгу.<sup>xi</sup> HathiTrust в чём-то схож с порталом Google Books, с помощью которого пользователи могут осуществлять полнотекстовый поиск книг, предоставленных партнёрами Google из числа издателей и библиотек, и — за исключением ограничений по авторским правам — получать доступ к ним. Корпус HathiTrust во многом совпадает с корпусом Google, но совпадения, скорее всего, со временем будут уменьшаться по мере того, как партнёры Google и HathiTrust будут независимо друг от друга добавлять в свои репозитории контент, оцифрованный не в рамках библиотечного проекта Google Books. Тогда как цель Google при оцифровке заключается в выполнении его задачи «организовать мировую информацию и сделать её доступной

и полезной во всём мире»<sup>xii</sup>, Hathitrust сфокусирован на сохранении, а доступ является значительной частью его стратегии сохранения и более широкой задачи по обеспечению общественного блага, доступного каждому.<sup>xiii</sup>

Hathitrust и Google Books используют различные методы отображения результатов поиска для работ, защищённых авторским правом,<sup>xiv</sup> и Hathitrust отстаивает права согласно закону США об авторском праве, чего не делает Google, например, предоставляя полный доступ к материалам, разработанным федеральным правительством США, и позволяющим пользователям по распечатке материалов, которые выписывают экземпляр книги, также получить доступ к полному тексту.<sup>xv</sup> Сотрудники в университете Мичигана осуществляют мониторинг материалов, опубликованных в США с 1923 по 1964 год с тем, чтобы установить, подпадают ли они под защиту авторского права,<sup>xvi</sup> и Hathitrust предоставляет определённые материалы по согласованию с владельцами прав.<sup>xvii</sup> Дополнительные преимущества Hathitrust перед Google Books — это его интегрированность с библиотечными каталогами, что позволяет пользователям находить отдельные выпуски серийных изданий, пользоваться предметными указателями и другой метаинформацией каталога, которая зачастую недостаточно корректно указывается в Google Books.<sup>xviii</sup> Hathitrust также предоставляет метаинформацию по книгам, находящимся в общем доступе в форматах MARC21 и Dublin Core через OAI-PMH с тем, чтобы любая библиотека могла добавить её в свой каталог.<sup>xix</sup>

### **Управление и схема финансирования**

В то время как культурные учреждения и учебные заведения часто желают сотрудничать в рамках созданных ими сервисов, зачастую затруднительно достичь эффективного сотрудничества. Барбара

МакФадден Аллен описывает два типа неэффективного сотрудничества: сотрудничество, при котором происходит «смещение цели» (уход от изначально поставленных целей) и сотрудничество, при котором нет содействия от административных органов, что приводит к недостаточности финансирования и поддержки.<sup>xx</sup> Она также описывает два типа затруднений, которые могут возникать при сотрудничестве — опасения участников по поводу потери независимости и опасения участников по поводу того, что сотрудничество затруднит принятие решений — и предлагает преодолеть эти затруднения путём организации добровольного и гибкого сотрудничества, «позволяющего участникам делать вложения на разных уровнях, согласно предоставленным гарантиям», при структурах принятия решений, «которые позволяют широкие вложения без резкого замедления принятия решений».<sup>xxi</sup>

NathiTrust предпринимает попытки наладить такое успешное сотрудничество. Партнёры NathiTrust платят за гигабайт находящихся на хранении данных, плюс годовой сбор за каждый год в течение которого на хранение помещается новый контент.<sup>xxii</sup> Такая модель позволяет учреждениям вносить оплату в пропорции к их объёму использования репозитория, что обеспечивает низкий барьер допуска для новых партнёров. В настоящее время проект управляется представителями организаций-учредителей (исполнительный комитет), при руководящем участии по распространению и планированию представителей всех организаций-членов (стратегический консультативный совет). Управление и схема финансирования будут пересматриваться по мере роста NathiTrust.<sup>xxiii</sup>

## **Управление правами**

Как уже было сказано, NathiTrust предоставляет пользователям в полнотекстовом виде многие материалы, которых нет в Google Books. Каждому материалу в NathiTrust соответствует запись в базе данных прав, содержащая правовой атрибут (контролирующий то, кому может предоставляться полный текст), основание для этого атрибута и источник решения.<sup>xxiv</sup> Интерфейс этой базы для персонала, просматривающего правовой статус материалов, позволил предоставление в доступ многих материалов, правовой статус ранее был неизвестен.

### **Доступ к мировой культуре**

Когда оцифровка библиотечной коллекции начиналась, расходы на реформатизацию, хранение и предоставление доступа к контенту были столь высоки, что надо было тщательно отбирать для оцифровки только редчайшие материалы или материалы, пользующиеся наибольшим спросом. Тогда как американские библиотеки были среди первых библиотек, которые частично оцифровывали свои коллекции, эта вторая стадия отбора означала, что практически весь оцифровываемый материал был на английском языке. Однако благодаря снижению расходов на оцифровку и перехода к «массовой оцифровке», все типы контента, который до этого можно было найти лишь в нескольких исследовательских библиотеках, становятся доступными пользователям во всём мире. В NathiTrust хранятся оцифрованные материалы общего пользования на 190 языках (на основе языковых кодов MARC), и так как только около 14% материалов NathiTrust находится в общем доступе, без сомнения, имеются материалы на гораздо большем количестве языков, которые были оцифрованы, но ещё не помещены в общий доступ.<sup>xxv</sup>

HathiTrust обеспечивает инновационную модель совмещения опыта и финансовых ресурсов отдельных библиотек для создания цифрового репозитория общего пользования, который функционирует в рамках закона об авторском праве. В то время как основной целью этого проекта является сохранение, он значительно увеличивает доступ к мировому культурному наследию.

---

<sup>i</sup> Авторы благодарны Борису Орехову за перевод статьи с английского.

<sup>ii</sup> HathiTrust Reaches 5 Million Volumes [Электронный ресурс] / John Weise // [BLT] Blog for Library Technology. — 2009. — Режим доступа: [mblog.lib.umich.edu/blt/archives/2009/12/hathitrust\\_reac.html](http://mblog.lib.umich.edu/blt/archives/2009/12/hathitrust_reac.html).

<sup>iii</sup> Mission & goals [Электронный ресурс] // HathiTrust : a shared digital repository. — 2009. — Режим доступа: [http://www.hathitrust.org/mission\\_goals](http://www.hathitrust.org/mission_goals).

<sup>iv</sup> This library never forgets / Jeremy York // Archiving 2008 [2009] final program and proceedings. — Режим доступа: <http://www.hathitrust.org/documents/This-Library-Never-Forgets.pdf>.

<sup>v</sup> How to join [Электронный ресурс] // HathiTrust : a shared digital repository. — 2009. — Режим доступа: <http://www.hathitrust.org/join>.

<sup>vi</sup> Accountability [Электронный ресурс] // HathiTrust : a shared digital repository. — 2009. — Режим доступа: <http://www.hathitrust.org/accountability>.

<sup>vii</sup> Google & the future of books [Электронный ресурс] : an exchange / by Paul N. Courant [и др.] // The New York review of books. — Т. 57, № 1 (14-го янв. 2010). — Режим доступа: <http://www.nybooks.com/articles/23565>.

<sup>viii</sup> Update on November 2009 activities [Электронный ресурс] // HathiTrust : a shared digital repository. — 2009. — Режим доступа: [http://www.hathitrust.org/updates\\_november2009](http://www.hathitrust.org/updates_november2009).

<sup>ix</sup> Welcome to the shared digital future [Электронный ресурс] // HathiTrust : a shared digital repository. — 2009. — Режим доступа: <http://www.hathitrust.org/about>.

<sup>x</sup> Frequently asked questions [Электронный ресурс] // OpenLibrary. — 2009. — Режим доступа: <http://openlibrary.org/about/faq>.

<sup>xi</sup> Open Library developer's meeting : one Web page for every book ever published // Musings on information and librarianship / Eric Lease Morgan. — 2009. — Режим доступа: <http://infomotions.com/musings/open-library/>.



- 
- xii About Google Books [Электронный ресурс] // Google. — 2009. — Режим доступа: <http://books.google.com/intl/en/googlebooks/history.html>.
- xiii This library never forgets / Jeremy York // Archiving 2008 [2009] final program and proceedings. — Режим доступа: <http://www.hathitrust.org/documents/This-Library-Never-Forgets.pdf>.
- xiv Universities add their own search of Google Books / Jeff Young // Wired campus / Chronicle of higher education. — 2009. — Режим доступа: [http://chronicle.com/blogPost/Universities-Add-Their-Own/8901/?sid=wc&utm\\_source=wc&utm\\_medium=en](http://chronicle.com/blogPost/Universities-Add-Their-Own/8901/?sid=wc&utm_source=wc&utm_medium=en).
- xv HathiTrust accessible interface [Электронный ресурс] // HathiTrust : a shared digital repository. — 2009. — Режим доступа: [http://mblog.lib.umich.edu/blt/archives/2009/10/hathitrust\\_acce.html](http://mblog.lib.umich.edu/blt/archives/2009/10/hathitrust_acce.html).
- xvi Copyright review management system – IMLS National Leadership Grant [Электронный ресурс] // MLibrary. — 2009. — Режим доступа: <http://www.lib.umich.edu/copyright-review-management-system>.
- xvii HathiTrust rights database [Электронный ресурс] // HathiTrust : a shared digital repository. — 2009. — Режим доступа: [http://www.hathitrust.org/rights\\_database](http://www.hathitrust.org/rights_database).
- xviii Google Books : a metadata train wreck / Geoff Nunberg // Language log. — 2009. — Режим доступа: <http://languagelog.ldc.upenn.edu/nll/?p=1701>.
- xix Data distribution & APIs [Электронный ресурс] // HathiTrust : a shared digital repository. — 2009. — Режим доступа: <http://www.hathitrust.org/data>.
- xx Come together right now / Barbara McFadden Allen // Digital libraries : a vision for the 21<sup>st</sup> century : a festschrift in honor of Wendy Lougee on the occasion of her departure from the University of Michigan. — Ann Arbor, MI : Scholarly Publishing Office, University of Michigan Library, [s.a.]. — Режим доступа: <http://hdl.handle.net/2027/spo.bbv9812.0001.001>.
- xxi Come together right now / Barbara McFadden Allen // Digital libraries : a vision for the 21<sup>st</sup> century : a festschrift in honor of Wendy Lougee on the occasion of her departure from the University of Michigan. — Ann Arbor, MI : Scholarly Publishing Office, University of Michigan Library, [s.a.]. — Режим доступа: <http://hdl.handle.net/2027/spo.bbv9812.0001.001>.
- xxii Cost calculator [Электронный ресурс] // HathiTrust : a shared digital repository. — 2009. — Режим доступа: <http://www.hathitrust.org/cost>.
- xxiii Governance [Электронный ресурс] // HathiTrust : a shared digital repository. — 2009. — Режим доступа: <http://www.hathitrust.org/governance>.
- xxiv HathiTrust rights database [Электронный ресурс] // HathiTrust : a shared digital repository. — 2009. — Режим доступа: [http://www.hathitrust.org/rights\\_database](http://www.hathitrust.org/rights_database).

---

<sup>xxv</sup> Public domain by language [Электронный ресурс] // HathiTrust : a shared digital repository. — 2009. — Режим доступа: [http://www.hathitrust.org/public\\_domain\\_language](http://www.hathitrust.org/public_domain_language).