

ИНФОРМАЦИОННЫЕ СООБЩЕНИЯ

А. И. Земсков

Создание новых информационных сервисов для цифровой эпохи – главная тема Билефельдского форума 2012 года

Обзор работы 10-й Международной конференции «Формирование инфраструктур будущего. Передача данных, создание информационно-коммуникационных технологий и сервисов для мира людей, родившихся в цифровую эпоху» (24–26 апр. 2012 г., Билефельд, Германия).

Ключевые слова: Билефельдский форум, Билефельд, 2012 г., информационные сервисы, информационно-коммуникационные технологии, семантические сети, открытый доступ.

В конце апреля 2012 г. в Билефельде состоялась 10-я Международная конференция «Формирование инфраструктур будущего. Передача данных, создание информационно-коммуникационных технологий и сервисов для мира людей, родившихся в цифровую эпоху» (Shaping Future INFO – Structures. Feeding data. Designing Information services and Building ICT-Infrastructure for a Digital Natives' World), в которой мне довелось принять участие. В этом обзоре представлены наиболее значимые и важные, на мой взгляд, аспекты этого форума.

Конференцию открыл Михаэль Хэппнер (Michael Hoerrner), директор библиотеки Университета г. Билефельд. (Он же произнес и заключительную речь, заявив, что уходит со своей должности, однако надеется, что в 2015 г. все участники этой конференции смогут снова приехать на очередной Билефельдский форум.)

Эта конференция проводится один раз в три года, и за время, прошедшее с предыдущей – в 2009 г. (в которой я также принял участие), между привокзальной гостиницей и конференц-комплексом сооружен новый огромный зал.

По мнению профессора Стефана Гредмена (Университет им. Гумбольдта, Берлин), эта конференция остается крупнейшим международным форумом, проводимом в Германии по данной тематике.

В конференции участвовали более 250 делегатов из 21 страны, правда, в основном – представители немецких университетов; от России (да и всех стран бывшего СССР) – один делегат.

Проведение конференции хорошо отработано. Было зачитано 4 приветствия по 15 минут, 2 ключевых доклада по 45 минут, 16 докладов по 30 минут (8 из них – в два параллельных потока), 2 публичные лекции приглашенных специалистов по 1,5 часа каждая; в перерывах между заседаниями (в отдельном помещении вблизи выставки) проведено 15 демонстраций информационных продуктов участников выставки (по 15 минут).

Для подготовки этого обзора получены разрешения на перевод текстов и работу с ними в рамках лицензии Creative Commons от пяти докладчиков (сетевой адрес материалов конференции: <http://conference.ub.uni-bielefeld.de/programme/>).

Ключевой доклад «От оболочки к содержанию и далее к контексту: изменяющаяся роль библиотек в исследованиях и в eScience» (From Containers to Content to Context: the Changing Role of Libraries in eScience and eResearch) представил профессор Стефан Гредмен (Stefan Gradmann) из Берлинской школы библиотечно-информационных наук Университета им. А. Гумбольдта (Humboldt University, Berlin / Berlin School of Library and Information Science).

Как отметил профессор Гредмен, каталогизация уделяет основное внимание оболочке (контейнеру) документа – это нашло отражение в правилах каталогизации. Даже новейшая система *Europeana* работает на принципах каталогизации. Библиотеки служат посредником в доступе к информационным объектам, что достигается путём использования каталогов. При этом посреднические связи осуществляются от метаданных к самим документам, которые входят в состав библиотечных фондов.

Профессор Гредмен считает, что сегодняшние базовые принципы, на которых построена обработка документа в библиотеке: это – объект как контейнер, содержащий информацию; наполнению контейнера не уделяется особого внимания; правила каталогизации концентрируются на признаках (атрибутах) контейнера; функциональная цепочка включает в себя получение, хранение, описание и поиск контейнеров с информацией.

Технология библиотечной работы исчерпала себя, поэтому её время заканчивается. Тим Бернерс Ли, создатель Сети в её современном виде, ввёл понятие web of documents и предложил пользоваться форматом описания документов rdf (Resource Description Framework). rdf – это модель для представления данных, особенно метаданных, содержит сведения о ресурсах в виде, пригодном для машинной обработки.

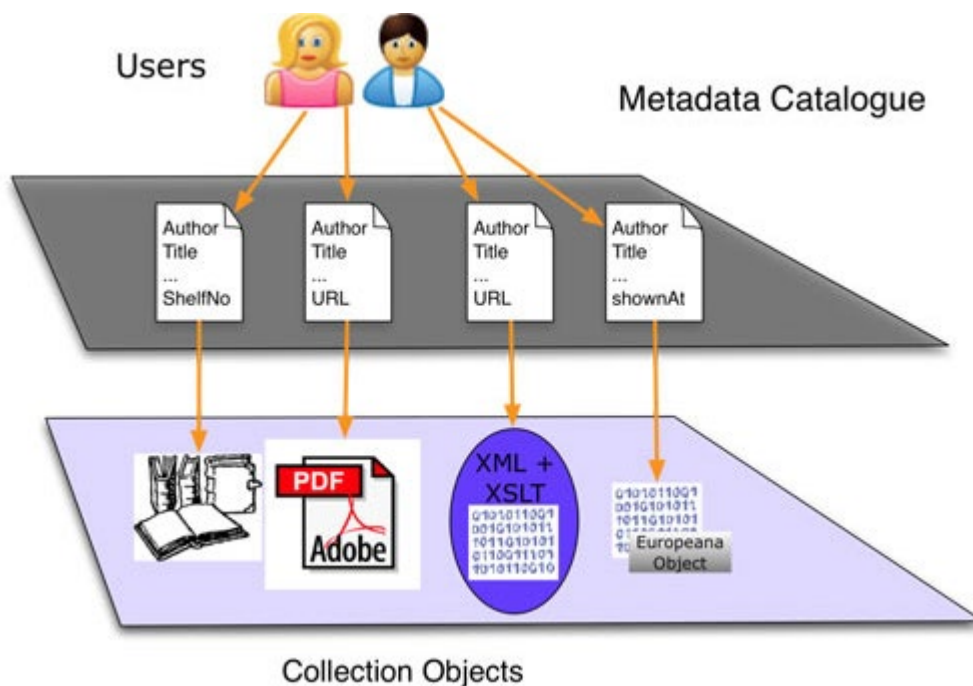


Рис. 1. Типовая схема работы традиционной библиотеки, представленная в докладе С. Гредмена: от пользователей (*users*) через посреднический слой метаданных каталога (автор, наименование, шифр, сетевой адрес, семантическая отсылка) осуществляется выход на различные коллекции документов в печатном или электронном формате.

Публикация документов (не только текстовых или изображений, но и любых иных – программ, аудио, видео, цифровых карт и др.) в Сети с простановкой смысловых меток (*semantic publishing*) позволяет формировать новый виртуальный мир, в котором группы сходных или связанных по смыслу документов образуют «кластеры».

Такого рода *semantic publishing* будет быстрее развиваться в естественно-научных и медицинских публикациях, поскольку в этих дисциплинах легче найти точные определения, чем в общественных науках и культуре. В качестве примера профессор Гредмен привёл две системы: материалы о научном сообществе 1500–1800 гг. и общении ученых «Mapping the Republic of Letters» (<https://republicofletters.stanford.edu/#maps>) и «Карельская культура» (Kulttuurisampo – <http://www.kulttuurisampo.fi/kulsa/historiailisetKartat.shtml>).

Семантическую публикацию можно определить как публикацию, включающую в себя то, что обогащает значение опубликованной журнальной статьи, упрощает её автоматизированный поиск, обеспечивает связь этой статьи с семантически смежными статьями, предоставляет доступ к данным, содержащимся в статье, действенным образом обеспечивает интеграцию данных из различных статей.

По мнению профессора Гредмена, постепенно произойдут следующие перемены: уйдут, т.е. станут ненужными, библиотечные каталоги, библиографические МАРК-записи, книжные хранилища.

Что придет взамен? Библиотеки будут обслуживать науку в качестве «Сети связанных открытых данных» (*Linked Open Data web*) или вообще упразднятся.

Бирте Кристенсен-Дальсгаард (Birte Christensen-Dalsgaard), представлявшая на конференции Королевскую библиотеку Дании (The Royal Library Copenhagen, Denmark), в ключевом докладе «Проблема данных – для

библиотек» (The Data Challenge – For Libraries) подчеркнула, что следует переопределить роль библиотек. Традиционные библиотечные функции быстро меняются из-за процессов оцифровки и глобализации. Нужно заново определить роль библиотек, исходя из разных целей: библиотека как хранилище документов; библиотека как публичный центр обслуживания; формирование новых типов читальных залов; создание новых сервисов для студентов; интеграция библиотек в систему образования. Новые сервисы для учёных включают в себя также библиометрию и создание репозитариев.

Группа специалистов European Research Library Organisation (LIBER) из библиотек университетов Тилбурга, Оксфорда, Хельсинки, Утрехта, Британской библиотеки и Королевской библиотеки Дании сформулировала рекомендации по перспективной политике научных библиотек. Основные направления развития научных и университетских библиотек: партнерство, повышение квалификации и переподготовка персонала, формирование метаданных для повторного использования данных.

В заявлении Технологического университета Квинсленда (The Queensland University of Technology) «Управление научными данными» (Management of research data) отмечено: научные данные могут иметь форму фактов, наблюдений, изображений, компьютерных программ, записей, измерений или экспериментов, на основе которых строятся доказательства, теории, гипотезы либо иные виды научной продукции. Данные могут быть числовыми, описательными, визуальными или тактильными, а также сырыми, очищенными или обработанными, и храниться в любых форматах, на любых носителях. Данные служат исходным материалом для работы, но зачастую пользователю предоставляется только конечный продукт.

Как считает Б. Кристенсен-Дальсгаард, с функциональной точки зрения историю технологического развития можно разделить на три этапа (сейчас мы находимся где-то на втором этапе): на первом – новые технологии вписываются в существующие процессы (т.е. делается в основном то же самое, но быстрее или дешевле); на втором – новые технологии интегрируются в существующие процессы (т.е. системы совершенствуются); на третьем – новые технологии проникают и распространяются, создавая совершенно новые процессы и системы.

Для организации работы с массивами данных, в первую очередь для обеспечения их сохранности и повторного использования, предлагается: обеспечить постоянный доступ силами самой библиотеки или в сотрудничестве с центрами данных; установить связи и партнерство с архивами и центрами данных, чтобы укрепить сетевую и взаимно-совместимую инфраструктуру для доступа к данным, их обнаружения и совместного использования; наладить кооперацию со специалистами по информационным технологиям для обеспечения надежного хранения массивов данных; подготовить инструкции по работе с данными (при самостоятельной работе пользователей, а также и с помощью библиотекарей) и процедуры работы с данными для местных групп учёных; оказывать учёным помощь в составлении заявок на гранты по тематике работы с данными.

Особенно важным для понимания трендов развития информационно-библиотечных технологий в Германии стало выступление представителя Немецкого научного фонда DFG Анне Липп (Anne Lipp) «Формирование цифровой трансформации – вклад DFG в инфраструктуру инновационной научной информации» (Shaping the Digital Transformation – The Contribution of the German Research Foundation to an Innovative Information Infrastructure for Research).

Бюджет DFG на 2012 г. – 2,5 млрд евро (около 100 млрд р.), в том числе на формирование информационной инфраструктуры – 60 млн евро. (Для справки: бюджетом 2012 года на научные исследования Российскому фонду фундаментальных исследований выделено 6 млрд р., Российскому государственному научному фонду – 1 млрд р. В 2011 г. финансирование обоих фондов было точно таким же; с учетом инфляции это означает, что бюджеты РФФИ и РГНФ сократились примерно на 10%.)

Основные направления финансовых вложений Немецкого научного фонда:

1. Лицензирование доступа к электронным документам в национальном масштабе. От издателей получено разрешение размещать статьи в Сети на личном веб-сайте автора или в репозитарии. Чаще всего используется модель «подвижной стенки» (*moving wall*) с периодом эмбарго (задержки между публикацией и размещением в Сети) 1 год.

2. Индексирование и оцифровка специальных коллекций (архивных записей, средневековых рукописей, газет и др.).

3. Поддержка систем открытого доступа.

4. Управление информационными потоками.

Постоянный и активный участник Билефельдских конференций – бывший директор библиотек Университета г. Тилбург, президент (2006–2010 гг.) Ассоциации европейских научных библиотек LIBER Ханс Гельинзе (Hans Geleijnse). Его доклад «Библиотеки и библиотечные сети на перепутье» (Libraries and Library Networks at a Crossroads) содержит предложения по направлениям библиотечного сотрудничества и краткий обзор состояния дел в некоторых странах Европы.

Прежде всего Х. Гельинзе подчеркнул: библиотеки нуждаются друг в друге для подготовки эффективных соглашений о лицензировании, формирования и поддержки региональной, национальной и международной информационной инфраструктуры, решения крупных проблем (в том числе архивирования резервных коллекций методических материалов, обеспечения сохранности цифровых документов), придания политического веса своей позиции по тем или иным вопросам (авторского права, налогов, бюджета).

Х. Гельинзе отметил, что сегодня востребованы определённые новые направления кооперации библиотек, обусловленные снижением спроса на традиционные библиотечные услуги (полагаясь только на свои силы, библиотеки не могут выдержать конкуренции и оправдать ожидания пользователей, когда существуют системы и устройства Google, iPads и т.п.), давлением на библиотеки со стороны университетов и финансирующих организаций (это касается снижения расходов, эффективности, экономичности), доступностью «облачных» сервисов.

«Облачные» сервисы осуществляются путем сетевого доступа к программным приложениям и организацией обслуживания на удаленных серверах. Разработаны «облачные» версии библиотечных систем автоматизации и поисковых средств. То, что в настоящее время управляется локально, может управляться посредством сотрудничества учреждений или через услуги сторонних организаций. Программные приложения могут добавляться библиотеками и другими организациями. При этом создается возможность комбинировать сервисы на различных платформах и от различных поставщиков, что существенно расширит ассортимент библиотечных услуг.

Говоря о снижении расходов, эффективности, экономичности, Х. Гельинзе отметил позитивный опыт, накопленный в кооперативной каталогизации и подготовке лицензионных соглашений силами консорциумов или на национальном уровне: «Мы можем сделать больше за счёт снижения безумных цен на подписные коммерческие журналы посредством национальной и международной кооперации и пропаганды, более активного стимулирования открытого доступа и репозитариев, комбинирования процессов лицензирования с требованием обеспечения открытого доступа, совместного приобретения библиотечных систем и программ, аутсорсинга.»

Эволюция библиотечных сетей. Сети, создававшиеся для совместной каталогизации (*OhioLink, GBV, Pica, Bibsys, ABES и др.*) и совместного лицензирования (*Fin-Elib, Heallink, ANKOS, CBUC и др.*), расширили направления деятельности.

Положение дел в некоторых странах

Норвегия. Библиотечная сеть Bibsys в течение многих лет поставляла в библиотеки интегрированную систему автоматизации библиотек собственной разработки. Однако традиционные системы управления библиотекой уже не вписываются в современные условия. Планируется к 2013 г. перейти на «облачную» систему Web-scale Management Services (WMS), разработанную OCLC, которой будут пользоваться более 100 библиотек.

Финляндия. Сетевая служба библиотек National Library Network Services располагает различного типа программным обеспечением для библиотек. Публичный интерфейс Национальной электронной библиотеки (National Digital Library) будет основан не на коммерческом продукте, а на программе открытого доступа VuFind.

Сеть ABES (Франция). Сеть ABES, подчиненная Министерству образования и науки Франции, ставит на 2012–2015 гг. новые цели: служить узлом метаданных (*metadata hub*) для электронной научной информации, осуществлять лицензирование на национальном уровне в интересах современной науки, помогать библиотекам в приобретении новых библиотечных систем.

Нидерланды. Научные библиотеки отлично сотрудничают в формировании Национального каталога (*National Catalogue – GGC*), МБА, лицензировании, обмене текущей научной информацией.

Планы на ближайшее будущее: создать инфраструктуру национального репозитория, провести совместную разработку сетевых сервисов (управление, курирование и хранение научных данных).

Германия. В настоящее время шесть библиотечных сетей предлагают ряд проектов. Однако гармонизация усилий и сотрудничество пока ограничены. Инновации в значительной мере стимулируются Немецким научным фондом (*DFG*).

Совместные рекомендации специальной рабочей группы и DFG (03.02.2011): следует шире использовать огромные возможности сетей; нужна большая координация, интеграция каталогов и сервисов, совместная разработка сервисов на основе открытой архитектуры в интересах всей страны.

Альма Сван (*Alma Swan*) в докладе «Кто выигрывает и сколько? Поддержка развития открытых систем и работа компании *Key Perspectives Ltd., UK*» (*Whose Value, What Value? Enabling Open Scholarship and Key Perspectives Ltd., UK*) постаралась ответить на вопрос, вынесенный в название доклада. Итак, от открытого доступа (ОД) выигрывают: учёные, учреждения, национальные экономики, наука и общество в целом, причём учёные выигрывают за счет большей «заметности» созданных ими публикаций, их использования и воздействия.

Цитирование публикаций, размещённых в ОД, увеличивается по различным отраслям науки на 36–200%.

Выгода для учреждения при использовании ОД – это расширение сферы его воздействия (т.е. его «заметности»), маркетинг и профилирование учреждения.

Значение ОД для науки и для общества в целом – повышение ценности знания за счёт более широкого его использования. Открытый доступ – более дешевая система: наука продвигается быстрее и эффективнее.

Пример системы ОД – *PubMed Central*: 2 млн полнотекстовых статей, 420 тыс. уникальных пользователей ежедневно (25% – университеты, 17% – компании, 18% – правительственные и иные учреждения, 40% – граждане).

Эффективность систем открытого доступа (*Houghton et al, 2009, 2010*)

Варианты ОД	Годовая экономия от перехода к ОД, млн евро			Правительственные агентства США
	Великобритания	Нидерланды	Дания	
Журналы ОД («золотой» вариант ОД)	480	133	70	Выгода за 30 лет составила 1 млрд долларов – в 6 раз больше, чем стоимость архивирования
Репозитории ОД с подпиской («зеленый» вариант ОД)	125	50	30	
Репозитории ОД с наложенными сервисами	около 480	около 133	около 70	

В качестве примера можно сравнить уровень и динамику цитирования статей из журналов «European Physical Journal», «Journal of High Energy Physics», «Nuclear Physics», «Physical Review», «Physics Letters», которые были одновременно переданы для традиционной публикации и в репозиторий ОД ArXiv: к моменту выхода в свет журнала уровень цитирования статей, поступивших в репозиторий, уже достиг 20% и фактически сравнялся с тем абсолютным уровнем цитирования, который печатные журналы достигнут лишь через два года!

Для многих научных библиотек весьма актуальна фактическая динамика использования электронных журналов и электронных книг. В качестве иллюстрации приведём данные из выступления Хайнца Вайнхаймера (Heinz Weinheimer, Business Modells in STM Publishing Springer Science + Business Media, Heidelberg).

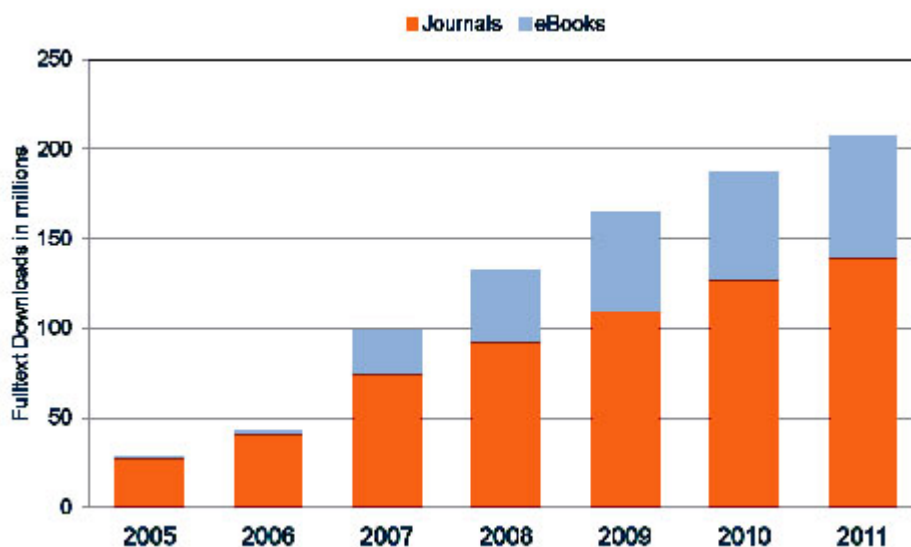


Рис. 2. На диаграмме отображена динамика (миллионы полнотекстовых выгрузок) использования электронных журналов (жёлтые) и электронных книг (голубые) системы *SpringerLink* за 2005–2011 гг.

В заключение – несколько слов о новейших тенденциях. Проект и компания Mendeley созданы три года назад в Нью-Йорке. Система предназначена для формирования личных научных библиотек; программа Mendeley desktop извлекает научные данные и агрегирует их в «облако». Таким образом, проведение научных исследований становится более прозрачным, коллективным. Система связывает: документы между собой (семантика), документы и их потребителей (т.е. тех, кто читает эту статью) и просто пользователей между собой (общие интересы и подходы).

К настоящему времени в систему загружено 160 млн документов от 1 600 тыс. индивидуальных и коллективных пользователей, среди которых крупнейшие университеты – Кембридж, Стэнфорд, Оксфорд, Гарвард и др.

Распространению системы помогают 1 200 волонтеров по всему миру. Доктор Вернер Фогельс (Werner Vogels) из компании Amazon.com заявил: «Я твердо верю, что Mendeley может изменить лицо науки».

Подводя итог всего услышанного и увиденного на очередном Билефельдском форуме, можно сделать следующие выводы и предложения:

1. Количество информации, и в частности научной, растёт слишком быстро, поэтому традиционными библиотечными методами невозможно её собрать, обработать и ввести в процессы обслуживания.
2. Обсуждаемые на конференции предложения по сбору и вторичному использованию так называемых сырых данных (*raw data*), на мой взгляд, нуждаются в доработке и оценке. Если не установить жестких критериев селекции, то поток поступающей на хранение информации (массивов сырых данных) приведет к «информационному загрязнению» (*information pollution*). Очень важно убедиться в фактической востребованности сырых данных, набрать полноценную статистику повторного использования. (Пока что шум вокруг этого направления очень напоминает ситуацию с «проблемой тысячелетия» Y2K – много слов,

мало фактов.)

3. Проработка вопросов развития семантической сети (*semantic web*) и технологии связи данных (*linked data*), развития форматов *rdf* в принципе может быть достаточно интересной и перспективной задачей для крупных библиотек, поскольку у них имеются и хорошие заделы и специалисты по индексированию. Могут лишь повторить вывод профессора Гредмена: «Библиотеки будут обслуживать науку в качестве “Сети связанных данных” (*Linked Open Data web*) или вообще станут ненужными».

4. Формирование систем открытого доступа и поддержка существующих репозитариев может быть выходом из назревающего кризиса перепроизводства научной информации. Дело не столько в ожидаемой выгоде от применения ОД, сколько в том, что жить по-старому научным и научно-техническим библиотекам невозможно.