

ПЕРСПЕКТИВЫ РАЗВИТИЯ НАУЧНО-ОБРАЗОВАТЕЛЬНЫХ ИНФОРМАЦИОННЫХ РЕСУРСОВ НА ОСНОВЕ ОТКРЫТЫХ ТЕХНОЛОГИЙ

DEVELOPMENT PERSPECTIVES OF THE SCIENTIFIC AND EDUCATIONAL INFORMATION
RESOURCES BASED ON OPEN TECHNOLOGY

Антопольский А.Б.

Заведующий лабораторией ИНИПИ РАО,
доктор технических наук
E-mail: ale5695@yandex.ru

Antopolsky A.B.

Head of the laboratory at the Institute of
Scientific and Pedagogical Information of the
Russian Academy of Education, Doctor of
Science (Engineering).

E-mail: ale5695@yandex.ru

Аннотация. В статье характеризуются
возможности использования открытых
информационных ресурсов, в том числе в сфере
образования.

Annotation. The article characterizes the
possibilities of using open information
resources in the field of education.

Ключевые слова: открытое программное
обеспечение, открытые стандарты, открытые
архивы, открытое образование, открытая наука.

Keywords: open software, open standards,
open archives, open education, open science.

Об общественном движении открытых технологий

В последнее время стали быстро распространяться идеи и методы, связанные с открытыми технологиями. Назовем среди них такие направления, как открытое программное обеспечение, открытые стандарты, открытые архивы, открытое образование, открытая наука и др. Важным шагом по развитию открытых технологий стало распространение открытых данных и, как очередной этап технологического развития, связанных открытых данных. Общим для всех открытых технологий является то, что их использование предпо-

лагают отказ от всех или большинства ограничений, налагаемых на использование результатов исследований и разработок.

Эти ограничения могут иметь правовой, экономический или социальный характер. Например, множество ограничений накладывает современное авторское право, в котором нарушен баланс интересов потребителей и правообладателей информации в сторону интересов правообладателей. На снятие этих ограничений направлены различные общественные движения, для которых в качестве общего названия используется понятие *copyleft*. Наиболее популярным инструментом, используемым для ликвидации или хотя бы сокращения сферы этих ограничений, являются открытые лицензии *Common Creative*. Необходимо отметить, что недавно принятые изменения в часть четвертую Гражданского кодекса РФ наконец делают открытые лицензии легальными и в России.

Другой класс ограничений связан с рыночным характером некоторых научных и образовательных коммуникаций, например издательских. Платный доступ к знаниям, представленным в научных и учебных произведениях, несомненно, противоречит идее всеобщего образования и равенства образовательных прав. Кроме того, весьма распространено мнение, что продукты, создаваемые за счет бюджетных средств, должны быть бесплатно доступны для налогоплательщиков. Однако до сих пор не разработаны или не внедрены приемлемые бизнес-модели бесплатного доступа к информационным ресурсам, при которых были бы гарантированы экономические условия нормального функционирования для всех участников коммуникаций.

Ограничения также могут быть связаны с различием прав некоторых категорий физических лиц, например граждан данной страны и иностранцев или учащихся данного учебного заведения и других граждан.

Важное свойство открытых технологий – это предоставление пользователям права применять полученные результаты и продукты интеллектуальной деятельности различными способами, как некоммерческими, так и коммерческими.

Если формулировать в общем виде, движение открытых технологий – это не столько технологическое, сколько общественное движение тех, кто не удовлетворен нынешним рыночным креном в научных и образовательных коммуникациях.

Причем данное движение не предполагает массовое огосударствление названных коммуникаций, как это было реализовано в СССР. Речь идет именно об общественном движении, иницируемом и поддерживаемом участниками добровольно, подобно многим аналогичным инициативам информационного общества.

Открытые данные

Обычно под открытыми данными понимаются просто данные, открыто публикуемые в Интернете, однако это представляется упрощением. Чтобы открытые данные можно было свободно использовать, они должны удовлетворять определенным технологическим

требованиям. Прежде всего, это относится к связанным открытым данным, использующим технологии семантического веба.

Среди открытых данных, являющихся предметом настоящей статьи, важнейшее место занимают две категории – правительственные данные и научно-образовательные данные. За рубежом, особенно в США, эти понятия пересекаются: в систему открытого правительства включены и государственные структуры, осуществляющие или заказывающие научную деятельность.

Так, NASA имеет большой опыт хранения научных данных, выкладываемых в свободном доступе в Интернете, при этом надо особо отметить, что агентством накоплено огромное количество научных данных. Например, в рамках проекта «Системные данные NASA и информационная система наблюдения Земли» (NASA's Earth Observing System Data and Information System, EOSDIS) [1] начиная с 2005 г. было получено свыше 5 Пб данных, хранящихся в распределенных хранилищах данных. Для сравнения, приблизительно такой же объем данных имеет Библиотека Конгресса США. Агентство NASA уже выложило в Сеть значительную часть своих данных, полученных в ходе выполнения различных космических программ, включая данные в области наук, изучающих Землю (это свыше 4 Тб научных данных, полученных только за один день проведенных исследований).

В рамках инициативы открытого правительства (Open Government Initiative) агентство NASA работает над улучшением качества доступа к этим данным. В агентстве также проводятся исследовательские работы в области открытых государственных данных. В 2010 г. NASA добавило три массива научных данных и 18 инструментов для работы с ними на сайт открытого правительства США (<http://data.gov/>). Концепция открытого правительства и, следовательно, концепция использования открытых данных предполагает соблюдение следующих принципов [2]:

- увеличение прозрачности и подотчетности по отношению к гражданам;
- увеличение уровня вовлеченности граждан в деятельность NASA;
- повышение качества внутреннего сотрудничества между исследователями NASA;
- большее поощрение совместной работы, по сравнению с экономическим сотрудничеством;
- применение доктрины открытого правительства в NASA.

В 2004 г. Организацией экономического развития и сотрудничества (ОЭСР, Organisation for Economic Co-operation and Development, OECD) было подписано коммюнике, определяющее доступность для граждан всех архивов, финансируемых за счет государственных средств [3].

В 2007 г. были опубликованы Принципы и положения обеспечения доступа к научным данным исследований, финансируемых государством (OECD Principles and Guidelines for Access to Research Data from Public Funding) [4]. В них говорится о том, что в связи с высокой скоростью развития Интернета и технологий, связанных с передачей, хранением и обработкой информации, появились новые возможности совместного использования научных данных мировым сообществом. В качестве наглядного примера

в них приводится международный проект «Геном человека» (The International Human Genome Project).

В июле 2009 г. был опубликован первый вариант Пантонских принципов (Panton Principles), которые представляют собой свод рекомендаций по опубликованию научных данных, доступных для их последующего использования всеми заинтересованными лицами. В 2010 г. Пантонские принципы были уточнены и доработаны при участии членов рабочей группы «Открытые данные в науке» (Open Data in Science Working Group) фонда Открытых знаний (Open Knowledge Foundation) [5].

В Пантонских принципах, в частности, сказано о том, что научные данные должны представлять собой общественное достояние и публиковаться при наличии соответствующей лицензии таким образом, чтобы была возможность использовать их повторно.

В 2009 г. Европейская организация по ядерным исследованиям (CERN) опубликовала книжный каталог открытых данных в рамках проекта «Открытая библиотека» (Open Library) [6]. CERN входит в консорциум SCOAP (Sponsoring Consortium for Open Access Publishing in Particle Physics) и поддерживает инициативу открытого доступа (Open Access), поэтому неудивительно, что эта организация заключила соглашение с Американским физическим обществом, Институтом Сисса, издательскими домами «Эльзевир» и «Шпрингер», по которому научные данные, полученные при помощи Большого адронного коллайдера, публикуются в открытом доступе под лицензией Creative Commons [7].

Понятие связанных данных

Связанные данные (Linked Data) – это набор данных, опубликованных в RDF-формате в соответствии с концепцией семантического веба при использовании унифицированного идентификатора ресурсов URI для отождествления и связывания между собой элементов, которые они содержат. Связываемость данных – одна из основных возможностей, предоставляемых семантической паутиной. Семантические связи между данными повышают их ценность и предоставляют дополнительные возможности для информационного поиска. Таким образом, данные, интегрированные в единую семантическую паутину, представляют собой пространство знаний о некоторой предметной области.

Возможность и простота связывания данных в семантической паутине – результат использования стандартного формата их представления Resource Description Framework (RDF). RDF – это простой способ описания данных в формате «субъект – отношение – объект», в котором в качестве любого элемента RDF-триплета используются идентификаторы ресурсов. В виде RDF-триплетов может быть представлена любая информация, любые информационные объекты, которые можно выделить и идентифицировать в Интернете.

Одним из активных сторонников идеологии открытых данных является знаменитый создатель веба и основатель Фонда открытого знания Тим Бернес-Ли. Он пред-

ложил идею связанных данных [8], в основе которой четыре принципа веба для доступа к данным:

- применение универсальных идентификаторов ресурсов (URI) в качестве имен объектов;
- применение идентификаторов HTTP URI для реализации возможности обращения по этим именам;
- предоставление полезной информации тому, кто обращается по URI, с помощью стандартов RDF, SPARQL;
- включение ссылок на другие универсальные идентификаторы ресурсов, позволяющих найти дополнительную информацию.

Распределенная открытая модель данных, используемая в концепции связанных данных, делает этот подход очень удобным для интеграции данных, хранящихся в различных базах данных и файловых системах, а также приложений, касающихся конкретных данных. Ниже перечислены некоторые возможности, благодаря которым связанные данные становятся очень удобным инструментом интеграции данных:

- единый интерфейс (определенный методами HTTP), универсально понимаемый и неизменный для всех приложений;
- универсальная схема адресации (благодаря URL) как для идентификации, так и для доступа ко всем объектам;
- простая и в то же время расширяемая модель данных для описания ресурсов, которая не требует предварительного изучения специальной терминологии.

Связанные данные опираются на существующую инфраструктуру Интернета и обладают такими важными характеристиками, как распределенность и масштабируемость.

Связанные открытые данные

Связанные открытые данные (Linked Open Data, LOD) можно определить как связанные наборы данных, опубликованные в RDF-формате и доступные для свободного использования всеми пользователями без каких-либо ограничений в виде авторских прав, патентов и других механизмов контроля.

В 1999 г. консорциум W3C опубликовал набор открытых стандартов семантической паутины, включающий в себя описание модели RDF. Этот стандарт и стал использоваться в проекте DBpedia [9], который составил ядро пространства наборов данных Linked Open Data, также называемое LOD-облаком (LOD-cloud). Проект DBpedia стартовал в 2007 г.

Тим Бернерс-Ли предложил также пятизвездочную шкалу классификации открытых данных [8]:

- одна звезда: данные доступны в вебе (в любом формате), но подпадают под лицензию открытых данных правительства Великобритании (Open Government Licence for public sector information);

- две звезды: открытые данные доступны в качестве машиночитаемых структурированных данных (например, в виде Excel-таблицы вместо отсканированного изображения таблицы);
- три звезды: открытые данные соответствуют двум звездам плюс представлены в непроприетарном формате (например, в формате CSV вместо Excel);
- четыре звезды: открытые данные соответствуют трем звездам плюс представлены в открытых стандартах консорциума W3C (RDF и SPARQL), предназначенных для идентификации данных;
- пять звезд: открытые данные соответствуют четырем звездам плюс связаны с другими данными с учетом использования открытых данных.

Из перечисленного следует, что наиболее предпочтительны и удобны в использовании связанные открытые данные, интегрированные в пространство LOD (имеющие пять звезд по представленной шкале классификации открытых данных), – это позволяет не только повторно их использовать, но и производить их машинную обработку, осуществлять поиск, разрабатывать аналитические инструменты и интегрировать их в различные системы.

Открытые данные в образовании

Основными участниками движения открытых технологий, в частности открытых данных, за рубежом стали университеты. Естественно, в пространство открытых данных, в том числе связанных открытых данных, стали активно вовлекаться различные образовательные ресурсы. Это тем более удобно, поскольку в последние годы стали активно распространяться открытые образовательные технологии, включая массовые открытые онлайн-курсы. В качестве примера можно привести проекты OLKOS (Open eLearning Content Observatory Services), Coursera, деятельность фонда Хьюлетт по открытым образовательным ресурсам и многие другие. Открытые образовательные ресурсы было естественно размещать в форматах открытых данных. В современном университете, ориентированном на мировые образовательные стандарты, очень востребованы, например, базы данных по экспертам, научным сотрудникам, научным публикациям, учебным курсам и другие ресурсы, относящиеся к процессам жизнедеятельности университета.

Существует большая проблема интеграции этих баз данных внутри университета. Данные из одной базы не всегда синхронизируются с данными из другой. Часто возникают ошибки при синхронизации баз, при выгрузке данных из одной базы в другую. Связанные данные на основе онтологий – это мощный инструментарий для решения подобных проблем. Университетов, занимающихся исследованиями в области связанных данных на Западе довольно много. Они выкладывают свои данные в виде связанных открытых данных и делают их «перевязку» в облаке связанных открытых данных.

Ситуация в России

В России движение открытых данных еще не приобрело массового характера. Тем не менее можно отметить несколько инициатив, представляющих интерес. Исторически первая из них – это открытая социальная сеть для науки Соционет [10], созданная и руководимая С.И. Париновым, подлинным энтузиастом открытой науки. Соционет включает открытый архив и целый ряд сервисов, обеспечивающих реализацию открытых технологий в науке. С.И. Паринов является также автором многих публикаций, содержащих интересные идеи по созданию онлайн-инфраструктуры открытой науки.

Можно отметить цикл публикаций А.И. Земскова и Я.Л. Шрайберга в журнале «Научные и технические библиотеки» за 2010–2011 гг., в которых пропагандируются идеи открытого доступа – одного из центральных направлений движения открытых технологий.

Определенные результаты по открытию данных в сфере науки и образования получены и в рамках открытого правительства, хотя в основном программа электронного правительства, равно как и законодательные инициативы, направлена на открытие данных органов государственной власти и местного самоуправления. Подробности о деятельности правительства в области открытых данных, а также полуофициальный реестр открытых правительственных данных можно найти на сайте НП «Информационная культура», возглавляемом Иваном Бегтиным [11].

В целом идеология открытых данных еще не стала базовой для деятельности в области научно-образовательных информационных ресурсов, хотя применение открытых программных средств для учреждений среднего образования вошло в государственную программу информатизации образования.

Что касается собственно научно-образовательных информационных ресурсов, то в настоящее время можно констатировать отсутствие государственной идеологии в этой сфере, как и вообще государственной информационной политики в области науки и образования. Соответственно не определено и отношение государства к движению открытых данных в этой сфере.

Важным событием стало создание российского отделения Фонда открытого знания, которое также возглавил И. Бегтин. Заметим, что почетным председателем этого фонда является Тим Бернерс-Ли.

В рамках данной общественной организации создана секция «Открытая наука и образование», которая начала работать на базе Научной педагогической библиотеки им. К.Д. Ушинского. Душой секции является И.А. Радченко, научный сотрудник НИУ ВШЭ, которая ведет информационный ресурс по данной проблеме [12] и является автором множества инициатив в области открытых данных. Одной из таких инициатив стала организация в Высшей школе экономики Школы открытых данных [13], где дается дополнительное образование по данной проблеме для всех желающих. Кроме того, И.А. Радченко ведет учебные курсы по журналистике данных – новой дисциплине, возникшей на основе движения открытых данных. В деятельности секции «Открытая наука и образование» принимает участие и автор данной статьи.

Мы надеемся, что российское отделение Фонда открытого знания станет центром кристаллизации идей и инициатив в сфере открытой науки и образования и приглашаем к участию в работе секции всех заинтересованных лиц, особенно сотрудников научных и образовательных учреждений, связанных с проблематикой информационных ресурсов.

Автор выражает благодарность И.А. Радченко за предоставленные материалы, которые были использованы при написании данной статьи.

Список литературы:

1. Earth Observing System Data and Information System (EOSDIS) // Data.NASA [Электронный ресурс] : project of the NASA Open Government. – Mode of access: <http://data.nasa.gov/earth-observing-system-data-and-information-system-eosdis>.
2. NASA Open Government Plan // Open Government at NASA [Электронный ресурс] : сайт открытого правительства США. – Mode of access: <http://www.nasa.gov/open>.
3. Science, Technology and Innovation for the 21st Century. Meeting of the OECD Committee for Scientific and Technological Policy at Ministerial Level. 29–30 January 2004. Final Communique // OECD [Электронный ресурс] : [сайт Организации экономического развития и сотрудничества]. – Mode of access: <http://www.oecd.org/science/scienceandtechnologypolicy/sciencetechnologyandinnovationforthet21stcenturymeetingoftheoecdcommitteeforscientificandtechnologicalpolicyatministeriallevel29-30january2004-finalcommunique.htm>.
4. OECD Principles and Guidelines for Access to Research Data from Public Funding // OECD [Электронный ресурс] : [сайт Организации экономического развития и сотрудничества]. – Mode of access: <http://www.oecd.org/science/scienceandtechnologypolicy/oecdprinciplesandguidelinesforaccesstoresearchdatafrompublicfunding.htm>.
5. Open Knowledge Foundation [Электронный ресурс] : [сайт фонда Открытых знаний]. – Mode of access: <http://okfn.org>.
6. Open Library = Открытая библиотека [Электронный ресурс] / НКО “The Internet Archive”. – Mode of access: <http://openlibrary.org>.
7. Supporting Open Access Publishing // Scientific Information Service [Электронный ресурс] : [сайт Научно-информационной службы]. – Mode of access: <http://library.web.cern.ch/library/OpenAccess/OpenAccessPolicy.html>.
8. Berners-Lee, T. Linked Data / Tim Berners-Lee // The World Wide Web Consortium (W3C) [Электронный ресурс] : [сайт Консорциума Всемирной паутины – Международного сообщества по содействию развитию Интернета]. – Mode of access: <http://www.w3.org/DesignIssues/LinkedData.html>.
9. DBpedia [Электронный ресурс] / Universität Leipzig, Freie Universität Berlin, OpenLink Software. – Mode of access: <http://dbpedia.org/About>.
10. Соционет [Электронный ресурс] : науч.-образоват. социальн. сеть / ЦЭМИ РАН ; ИЭОПП СО РАН. – Режим доступа: socionet.ru.

11. Информационная культура [Электронный ресурс] / НП «Инфокультура». – Режим доступа: <http://infoculture.ru>.
12. Радченко, И. Открытые данные, семантические технологии и светлое будущее [Электронный ресурс] / И. Радченко. – Режим доступа: <http://iradche.ru/category/открытая-наука-2>.
13. Школа открытых данных [Электронный ресурс] : [сайт]. – Режим доступа: <http://opendataschool.ru>.

Spisok literatury:

1. Earth Observing System Data and Information System (EOSDIS) // Data.NASA [Электронный ресурс] : project of the NASA Open Government. – Mode of access: <http://data.nasa.gov/earth-observing-system-data-and-information-system-eosdis>.
2. NASA Open Government Plan // Open Government at NASA [Электронный ресурс] : сайт открытого правительства США. – Mode of access: <http://www.nasa.gov/open>.
3. Science, Technology and Innovation for the 21st Century. Meeting of the OECD Committee for Scientific and Technological Policy at Ministerial Level. 29–30 January 2004. Final Communique // OECD [Электронный ресурс] : [сайт Организации экономического развития и сотрудничества]. – Mode of access: <http://www.oecd.org/science/scienceandtechnologypolicy/sciencetechnologyandinnovationforthet21stcenturymeetingoftheoecdcommitteeforscientificandtechnologicalpolicyatministeriallevel29-30january2004-finalcommunique.htm>.
4. OECD Principles and Guidelines for Access to Research Data from Public Funding // OECD [Электронный ресурс] : [сайт Организации экономического развития и сотрудничества]. – Mode of access: <http://www.oecd.org/science/scienceandtechnologypolicy/oecdprinciplesandguidelinesforaccesstoresearchdatafrompublicfunding.htm>.
5. Open Knowledge Foundation [Электронный ресурс] : [сайт фонда Открытых знаний]. – Mode of access: <http://okfn.org>.
6. Open Library = Открытая библиотека [Электронный ресурс] / NKO «The Internet Archive». – Mode of access: <http://openlibrary.org>.
7. Supporting Open Access Publishing // Scientific Information Service [Электронный ресурс] : [сайт Научно-информационной службы]. – Mode of access: <http://library.web.cern.ch/library/OpenAccess/OpenAccessPolicy.html>.
8. Berners-Lee, T. Linked Data / Tim Berners-Lee // The World Wide Web Consortium (W3C) [Электронный ресурс] : [сайт Консорциума Всемирной паутины – Международного сообщества по содействию развитию Интернета]. – Mode of access: <http://www.w3.org/DesignIssues/LinkedData.html>.
9. DBpedia [Электронный ресурс] / Universität Leipzig, Freie Universität Berlin, Open-Link Software. – Mode of access: <http://dbpedia.org/About>.
10. Sotsionet [Электронный ресурс] : науч.-образоват. социал'н. сет' / TsEMI RAN ; IEOPP SO RAN. – Режим доступа: socionet.ru.

11. Informatsionnaia kul'tura [Elektronnyĭ resurs] / NP «Infokul'tura». – Rezhim dostupa: <http://infoculture.ru>.
12. *Radchenko, I.* Otkrytye dannye, semanticheskie tekhnologii i svetloe budushchee [Elektronnyĭ resurs] / I. Radchenko. – Rezhim dostupa: <http://iradche.ru/category/otkrytaia-nauka-2>.
13. Shkola otkrytykh dannykh [Elektronnyĭ resurs] : [saĭt]. – Rezhim dostupa: <http://opendataschool.ru>.

Интернет-журнал
«Проблемы современного образования»
2014, № 3