

## ПЕРСПЕКТИВЫ ПРИМЕНЕНИЯ СВЯЗАННЫХ ОТКРЫТЫХ ДАННЫХ В СФЕРЕ ОБРАЗОВАНИЯ

### OUTLOOK FOR RELATED OPEN DATA IN EDUCATION

**O. Grishmanovskaya  
P. Grishmanovsky**

*Summary.* Opportunities Of Linked Open Data In The Education Field. Grishmanovskaya O.N., Grishmanovskiy P.V. Surgut State University

The current trend towards increasing the information openness of various spheres of human activity contributes to the spread of open data technologies, the potential of which lies in the possibility of creating new socially significant services that provide citizens with relevant and objective information.

The use of linked open data is associated with solving a number of problems on the way to the dissemination of open data technologies in general and in the field of education in particular. The solution of a number of technological, social and legal issues is necessary for the creation and use of open data of educational organizations and the construction of information services based on them.

*Keywords:* open data, linked open data, education, information service.

**Гришмановская Ольга Николаевна**

Аспирант, Сургутский государственный университет

*grishmanovskaya\_on@surgu.ru*

**Гришмановский Павел Валерьевич**

К.т.н., доцент, Сургутский государственный университет

*grishmanovskiy\_pv@surgu.ru*

*Аннотация.* Существующая в настоящее время тенденция к повышению информационной открытости различных сфер человеческой деятельности способствует распространению технологий открытых данных, потенциал которых заключается в возможности создания новых общественно-значимых сервисов, предоставляющих гражданам актуальную и объективную информацию.

Применение связанных открытых данных сопряжено с решением ряда задач на пути распространения технологий открытых данных в целом и в сфере образования в частности. Решение ряда технологических, социальных и правовых вопросов необходимо для создания и использования открытых данных образовательных организаций и построения информационных сервисов на их основе.

*Ключевые слова:* открытые данные, связанные открытые данные, образование, информационный сервис.

Современное общество характеризуется высокой степенью зависимости социальных, политических, экономических, деловых и др. процессов от наличия, достоверности и актуальности информации, необходимой для осуществления соответствующей деятельности. С другой стороны, стремительное увеличение объемов данных, представленных в цифровом виде, оказывает влияние на процессы и изменяет их, переводя из плоскости выполняемых человеком операций с данными, такими как поиск, запрос, отбор, обработка и представление информации, в разряд информационных сервисов, предоставляющих актуальную информацию на основе потребностей человека в решении тех или иных задач. Одной из таких областей является сфера образования, однако, наличие множества разрозненных или слабо связанных источников информации в этой сфере не позволяет построить полноценные сервисы, так как адекватность консолидированных данных от таких источников не может быть гарантирована. Решением может являться вовлечение участников образовательной деятельности в процесс

организации информации в виде системы наборов связанных открытых данных.

Открытые данные (ОД) — это технологии, которые являются сугубо прикладными и ориентированными на потребителя. Для создания общественно-значимых сервисов на основе данной технологии может быть использована практически любая информация в форме открытых данных, кроме информации, относящейся к государственной, коммерческой тайне или к персональным данным.

Открытые данные можно условно разделить на четыре категории по характеру доступа к предоставляемой информации [1]:

- ◆ государственная информация ограниченного доступа;
- ◆ официальная государственная открытая информация, предоставляемая для общего пользования, т.е. доступная обществу;
- ◆ коммерческая (деловая) информация, в т.ч. информация о деятельности предприятий и орга-

низаций, нацеленная на улучшение количества и качества предоставляемых услуг;

- ◆ бесплатная (свободно-распространяемая) информация.

Первая категория — это информация, которая не является распространяемой и к ней также можно отнести и другие защищаемые данные, такие как персональные данные, поэтому технологии открытых данных не применимы к этой категории информации. Относительно следующих трех категорий технологии открытых данных могут применяться, однако, под ОД подразумевают, как правило, официальную государственную информацию для общего доступа, например, ОД муниципалитетов [1].

В России технологии открытых данных внедряются как на уровне государства, так и в регионах, и с этой целью в 2009 г. запущен некоммерческий проект OpenGovData.ru — хранилище открытых данных. На данном портале представлены государственные открытые данные по различным направлениям деятельности. В настоящее время в РФ идет интенсивное внедрение предоставления в электронном виде различных государственных услуг и в этом контексте ОД приобретают особое значение.

Об этом свидетельствуют также существующие проекты, созданные на базе открытых государственных данных, которые могут представлять информацию государственных структур и учреждений, относящимся к различным сферам жизнедеятельности. В качестве примеров таких проектов можно представить: «Университетская информационная система Россия» ([uisrussia.tsu.ru](http://uisrussia.tsu.ru)) — собирает официальную государственную статистику для её использования в академических исследованиях; ЦИАН ([cian.ru](http://cian.ru)) — коммерческий сервис поиска недвижимости, который включает данные о качестве школ и качестве зданий, полученные как открытые данные из официальных источников; «Если быть точным» ([tochno.st](http://tochno.st)) — предоставляет основанную на открытых данных статистику и ее визуализацию по проблемам социальной политики и здравоохранения; «Российские школы» ([russianschools.ru](http://russianschools.ru)) — агрегирует все доступные сведения о школах, помогая исследователям и официальным лицам сферы образования, а также родителям, делающим выбор образовательного учреждения; «Открытые НКО» ([openngo.ru](http://openngo.ru)) — система раскрытия информации о некоммерческих организациях, предоставляющая информацию о деятельности различных некоммерческих организаций на основе полученных ей государственных субсидий, грантов и контрактов [2]. Однако, в целом процесс внедрения открытых данных пока находится на начальном уровне, и информация зачастую представлена в недостаточно формализованном виде.

Наряду с правительственными данными важное место занимают научно-образовательные открытые данные. Это обусловлено тем, что образование является одной из важнейших потребностей граждан в современном обществе и одним из приоритетов государственной политики. Интерес к открытым данным образовательных организаций существует как у граждан, так и у государственных структур, осуществляющих управление и контроль в сфере науки и образования, и у самих образовательных и научных организаций, которые могут как формировать (генерировать) ОД, так и использовать их для обеспечения оптимальной работы образовательной организации, улучшении качества оказываемых услуг в области учебной, научно-исследовательской, административно-хозяйственной и др. деятельности.

В частности, такие данные вузов, как проходной балл при поступлении, включая проходной балл прошлых лет, перечень специальностей и направлений подготовки и реализуемых образовательных программ, трудоустройство выпускников и т.д. интересны, в первую очередь, абитуриентам и их родителям. Что касается данных, в которых заинтересованы сами вузы, то это востребованность выпускников тех или иных направлений и специальностей подготовки на рынке труда, направления подготовки, востребованные у абитуриентов, и их ожидания в части содержания образовательных программ, приоритеты государственной политики в части получения специалистов тех или иных направлений подготовки для нужд национальной и региональных экономик и др. Могут использоваться ОД из разных сфер деятельности, и возможность их генерации, сбора и анализа позволяют провести исследования и спрогнозировать ситуацию трудоустройства и спроса на подготовку по различным специальностям и направлениям [3].

При этом информация в виде уже полученных результатов исследований и выводов не совсем интересна для дальнейшей обработки и анализа, необходимы исходные данные, которые можно было бы собирать и анализировать как в масштабах страны, так и отдельных регионов или субъектов и т.п. Это даст любым заинтересованным сторонам возможность увидеть картину в целом по вузам РФ: номенклатура и спрос на те или иные специальности и направления подготовки, баллы ЕГЭ по годам (проходной балл, минимальный балл), трудоустройство выпускников, административно-хозяйственная деятельность организации и т.д.

Еще одним направлением применения технологии ОД в сфере образования является предоставление образовательного контента и информации и нем. Основными участниками использования технологии

открытых данных в этой области являются зарубежные университеты. В последние годы активно распространяются открытые образовательные технологии, например, массовые открытые онлайн-курсы (MOOC, OLCOS — Open eLearning Content Observatory Services), курсы на различных платформах, таких как Открытое образование, Coursera и другие. Размещение образовательных ресурсов в формате открытых данных является очень удобным способом, так как в вузах востребована информация по учебным курсам, научным публикациям, рабочим программам дисциплин, преподавателям, исследователям и экспертам и т.д. Даже в рамках одного вуза могут возникать проблемы внедрения соответствующих баз данных из-за сложностей с синхронизацией информации, в то время как создание наборов открытых данных в едином формате позволяют интегрировать ресурсы образовательных организаций для создания востребованных учебных курсов и программ.

Однако, существует ряд препятствий на пути распространения открытых данных в сфере образования [2, 4]:

1. Отсутствие единого формата открытых данных. Данные выкладываются на сайте вуза не в машиночитаемом формате, а в форматах, требующих дополнительной обработки, поэтому использовать такие данные весьма сложно, а порой практически невозможно в силу большого количества объектов, форматов их представления и различного качества. Часто документы размещаются в проприетарных форматах pdf, doc (MS Word), в виде отсканированных изображений и др. Все это не имеет отношения к ОД и машиночитаемым форматам.
2. Неинформативные для конечного потребителя наборы данных. Наборы данных, представленные на портале Рособнадзора, зачастую не представляют большого интереса с точки зрения их анализа, как правило, это списки организаций (школы, дошкольные образовательные учреждения), даты и данные проверок, отчетность. Существует обширный перечень наборов ОД, но нет четкой структуры, что усложняет их использование.
3. Нет понимания сущности открытых данных и того, как с ними работать, у участников образовательного процесса, не хватает специалистов по работе с данными, вследствие чего ОД вузов представлены очень скудно.
4. Не разрешены многие вопросы безопасности личной информации и правовые аспекты публикации контента в форме ОД.

Для преодоления этих препятствий необходимо, в первую очередь, учитывать, что под открытыми дан-

ными понимаются данные или информация, которые представлены в машинно-читаемом формате, доступны для использования и последующей обработки без ограничений, непроприетарны и др.

Существуют основополагающие классификации, связанные с открытыми данными, которые позволят оценить степень реализации этой технологии. Известную классификацию по доступности и удобству применения ОД предложил один из авторов основных концепций «Всемирной Паутины» (World Wide Web) Тим Бернерс-Ли [1, 4]. Это пятизвездочная шкала классификации открытых данных [5, 6], в которой каждый уровень включает предыдущий и дополняет его новым требованием:

- ◆ одна звезда: данные доступны в сети Интернет в любом формате, но обязательно под открытой лицензией;
- ◆ две звезды: открытые данные доступны в качестве машиночитаемых структурированных данных в проприетарном формате (например, файл в формате Excel);
- ◆ три звезды: открытые данные соответствуют двум звездам и представлены в структурированном машиночитаемом виде в непроприетарном формате (например, в формате CSV);
- ◆ четыре звезды: открытые данные соответствуют трем звездам и представлены с использованием открытых стандартов консорциума W3C (RDF и SPARQL), предназначенных для идентификации объектов и их элементов таким образом, чтобы можно было использовать URI ссылки на них;
- ◆ пять звезд: открытые данные соответствуют четырем звездам и связаны с другими наборами открытых данными (например, структурированные в виде взаимоувязанных таблиц), что позволяет определять семантику данных посредством создания общего контекста [1, 7].

Способом обращения к ОД при поиске и использовании необходимой информации являются web-сервисы. Если используются единые формат и структура ОД, представленных разными поставщиками, то взаимодействие с ними реализуется технически просто. Для этого достаточно реализации на уровне трех звезд. Однако, построение сервиса, например, национального масштаба, кроме предоставления данных в унифицированном формате, требует наличия четких семантических связей, которые должны быть проработаны в виде ER-диаграмм и отражать связи между сущностями, что соответствует пятизвездочной модели. В настоящее время существует ряд исследований как в России, так и за рубежом, посвященных разработке онтологий образовательной системы и отражающих различные ее аспекты [5].

В действительности можно наблюдать, что открытые данные в некоторых сервисах недостаточно структурированы, не всегда соблюдаются требования технологического аспекта, не учитывается нормативно-правовая база предоставления ОД.

Также следует отметить, что информация должна быть связанной. Так, если совместимость данных отсутствует, то это усложняет обработку информации в автоматизированных системах. Использование ОД позволяет устранить эти проблемы в системах автоматизации, в частности, обращаться в различные государственные или другие учреждения можно будет виртуально и многие вопросы можно будет решать дистанционно, так как уменьшится повторный ввод данных в связанные информационные системы, что позволит избежать ошибок, свойственных ручному вводу данных при их переносе из одной системы в другую. Связанность информации между автоматизированными системами позволит агрегировать, систематизировать, передавать и обрабатывать информацию в оперативном режиме, без ошибок и задержек [1].

Приведение информации к единообразной системе должно включать в себя следующее:

- ◆ четкое описание структуры ОД соответствующие нормативам;
- ◆ унифицированный формат представления ОД;
- ◆ четкий регламент обновления ОД, связанный с актуальностью информации, а также сохранение и ведение истории предыдущего состояния данных до внесения изменений и др.

Формирование и обновление ОД должно руководствоваться четким регламентом, при этом за актуальность информации отвечать должны организации или структуры, которые являются первоисточником этой информации. Сохранение истории состояния на определенные моменты времени позволяют различать актуальность ОД на момент запроса, а также иметь возможность обращаться к данным, актуальным в прошлые периоды времени, понимая, что они не являются актуальными в данный момент [1]. Кроме того, открытые данные могут быть связаны с информацией, которая не может распространяться свободно по тем или иным причинам.

С учетом всех перечисленных аспектов при создании подобной модели, можно подойти к «пяти звездам» по классификации Тима Бернерса-Ли, то есть к интеграции данных на всех уровнях и совместному использованию этих данных. Выполнить эту задачу можно на уровне государства.

Тим Бернерс-Ли ввел понятие связанных открытых данных и предложил четыре принципа связанных данных [6, 8]:

1. Следует применять URI — универсальные идентификаторы ресурсов — для идентификации любых объектов.
2. Использование HTTP URI для поиска и интерпретации объектов.
3. При обращении по URI, предоставлять полезную информацию, используя открытые стандарты, такие как RDF и SPARQL.
4. При публикации данных включать ссылки на другие URI для получения дополнительной связанной информации.

Таким образом, связанные открытые данные (Linked Open Data, LOD) — это наборы данных, опубликованные в RDF-формате с использованием унифицированного идентификатора ресурсов URI и доступные для свободного использования всеми пользователями без ограничений и механизмов контроля [9]. В совокупности они образуют семантическую сеть — информационную модель предметной области, включающую не только собственно данные, но и ссылки для анализа и использования связанных данных, т.е. представляют собой сети данных, объединяемых семантическим контекстом. При этом, использование стандартного формата RDF — это простой способ описать данные в виде «субъект — отношение — объект» или «предмет — вид взаимодействия — другой предмет», причем в качестве любого элемента URI идентифицирует любой объект, понятие или отношение. В формате RDF можно представить любые данные.

В части решения нормативно-правовых вопросов использования ОД существует три основных направления:

1. Нормативная база, обеспечивающая публикацию открытых данных государством.
2. Правовые вопросы, связанные с самими публикуемыми данными.
3. Лицензии, регламентирующие использование данных и определяющие правовые отношения пользователя данных и их публикатора [2].

В последние годы распространение направлений, связанных с открытыми технологиями, например, открытое образование, открытая наука и др., стало популярным и необходимым. Важную роль в этом представляют ОД и связанные ОД. Существуют различного рода ограничения на использование, распространение результатов исследований, технических и технологических разработок. Связано это может быть с социальной, экономической, правовой сферой, например, с авторским правом, правом патентообладателя. В то же время, открытые технологии предполагают снятие большинства подобных ограничений.

Использование открытых лицензий Creative Commons позволяют привести к сокращению данных ограничений. Например, лицензия CC Attribution-ShareAlike (сокращённо CC BY-SA) является самой распространённой, эта лицензия является копилефт-лицензией [7]. Копилефт (англ. copyleft) — концепция и практика использования законов авторского права для обеспечения невозможности ограничить любому человеку право использовать, изменять и распространять как исходное произведение, так и произведения, производные от него [10, 11]. Изменения, внесенные в Гражданский кодекс РФ, дают возможность использование открытых лицензий: Гражданский кодекс Российской Федерации (часть четвертая) от 18.12.2006 № 230-ФЗ (ред. от 30.04.2021), ГК РФ Статья 1286.1. Открытая лицензия на использование произведения науки, литературы или искусства (введена Федеральным законом от 12.03.2014 N35-ФЗ).

Таким образом общественное продвижение открытых технологий стремится предоставить пользователям возможность использовать результаты интеллектуальной деятельности для улучшения качества жизни граждан [9].

На сегодняшний момент тема ОД является актуальной и востребованной, но для полноценного использования этой технологии и получения от нее общественно-значимой пользы, необходимо решать проблемы технологические, нормативно-правовые и методические, нужны стандарты работы с ОД, которые позволят решить вопросы со структурой, обработкой ОД, внедрением систем автоматизации на их основе и исключением влияния человеческого фактора. Также необходимо учесть и защитить интересы участников создания, формирования и использования ОД.

#### ЛИТЕРАТУРА

1. Открытые данные: проблемы и решения [Электронный ресурс] — Режим доступа: <https://cyberleninka.ru/article/n/otkrytye-dannye-problemy-i-resheniya/viewer>
2. Работа с открытыми данными: особенности публикации и использования в российском правовом поле [Электронный ресурс] — Режим доступа: <https://www.infoculture.ru/wp-content/uploads/2020/11/OpenDataReview.pdf>
3. Закрытые открытые данные [Электронный ресурс] — Режим доступа: <http://libinform.ru/read/articles/Zakrytye-otkrytye-dannye/>
4. Революция открытых данных: ожидания и результаты [Электронный ресурс] — Режим доступа: <https://te-st.ru/2016/06/08/open-data-revolution/>
5. Шульга, Т.Э. О проблемах разработки открытых связанных данных в сфере высшего образования и науки РФ / Т.Э. Шульга, А.А. Сытник // Математические методы в технике и технологиях — 2019. — Т. 11. — С. 52–61.
6. Tim Berners-Lee. Linked Data [Электронный ресурс] — Режим доступа: <https://www.w3.org/DesignIssues/LinkedData.html>
7. Лицензии и инструменты Creative Commons [Электронный ресурс] — Режим доступа: [https://ru.wikipedia.org/wiki/Лицензии\\_и\\_инструменты\\_Creative\\_Commons](https://ru.wikipedia.org/wiki/Лицензии_и_инструменты_Creative_Commons)
8. Связанные данные [Электронный ресурс] — Режим доступа: [https://hmong.ru/wiki/Linked\\_open\\_data](https://hmong.ru/wiki/Linked_open_data)
9. Антопольский, А.Б. Перспективы развития научно-образовательных информационных ресурсов на основе открытых технологий / А.Б. Антопольский // Проблемы современного образования. — 2014. — № 3. — С. 129–138.
10. Копилефт [Электронный ресурс] — Режим доступа: <https://dic.academic.ru/dic.nsf/ruwiki/7954>
11. What is Copyleft? [Электронный ресурс] — Режим доступа: <https://www.gnu.org/copyleft/>