

ОТКРЫТЫЕ ДАННЫЕ В РОССИЙСКОЙ СИСТЕМЕ ОБРАЗОВАНИЯ

OPEN DATA IN THE RUSSIAN EDUCATION SYSTEM

**D. Kuzin
I. Danilenko**

Summary. The article discusses the concept of open data, their impact on the development of society and the government. The most famous Russian and international organizations, public and industry initiatives in the field of publishing open data, their presentation, standardization and monitoring are described. Six basic principles are formulated according to which data disclosure work should be published. A brief analysis of the current situation is given and the main problems in the field of publishing open data in the education system are described. Based on the results of the analysis, a proposal is formulated for the development of a universal semantic model of open data based on the Semantic Web technology, which has the properties of universality, distribution, extensibility and ensures the semantic connectivity of published open data within the entire industry. A list of tasks to be solved to build an effective system for publishing open data in the education sector is given.

Keywords: open data, semantic model, semantic web.

Кузин Дмитрий Александрович

*К.т.н., доцент, Сургутский государственный университет, г. Сургут
kuzin_da@surgu.ru*

Даниленко Иван Николаевич

*К.т.н., доцент, Сургутский государственный университет, г. Сургут
vice1@surgu.ru*

Аннотация. В статье рассматривается понятие открытых данных, их влияние на развитие общества и государства. Описываются наиболее известные российские и международные организации, общественные и отраслевые инициативы в области публикации открытых данных, их представления, стандартизации и мониторинга. Формулируются шесть основных принципов, в соответствии с которыми должна проводиться работа по раскрытию данных. Дается краткий анализ текущего состояния работ и описываются основные проблемы в области публикации открытых данных системы образования. По итогам анализа формулируется предложение по разработке универсальной семантической модели открытых данных, основывающейся на технологии Semantic Web, обладающей свойствами универсальности, распределенности, расширяемости и обеспечивающей семантическую связность публикуемых открытых данных в пределах всей отрасли. Приводится перечень задач, которые предстоит решить для построения эффективной системы публикации открытых данных сферы образования.

Ключевые слова: открытые данные, семантическая модель, семантический веб.

Обеспечение свободы доступа к информации является глобальным вызовом на пути построения развитого общества. Одним из направлений деятельности в данной области является раскрытие информации в формате т.н. «открытых данных» (open data). Под этим термином понимают информацию, которую кто угодно может свободно использовать и распространять [1]. Как правило, открытые данные представляют собой хорошо структурированную информацию в форме таблиц или наборов данных, возможно, связанных между собой.

Чаще всего понятие открытых данных связывают с обеспечением прозрачности и открытости деятельности государственных институтов, как необходимое условие функционирования демократических процедур. Необходимость раскрытия информации о деятельности государства осознается правительствами многих стран. В России действует федеральный закон, призванный обеспечить прозрачность действий органов власти и местного самоуправления [2]. Публи-

кация открытых данных государственными органами и подведомственными им организациями способствует подконтрольности государства обществу, развитию механизмов обратной связи, снижению уровня коррупции. Кроме того, открытые данные ведут к появлению разнообразных «вторичных» информационных сервисов, выполняющих аналитические функции, в том числе, и в интересах бизнеса. Создаются некоммерческие проекты, осуществляющие агрегацию и обработку открытых данных, и предоставляющие интерфейсные средства для работы с ними.

Существует больше количество общественных инициатив, изучающих методологические, технические, юридические и другие аспекты публикации открытых данных. Одной из наиболее известных международных инициатив является Open Knowledge Foundation [3]. Всемирный Банк развивает собственный проект в области открытых данных [4]. В России активно работает над продвижением культуры открытых данных АНО «Информационная культура» [5]. На федеральном

портале открытых данных опубликованы методические рекомендации по публикации открытых данных [6]. Существуют примеры публикации открытых данных отраслевого значения. Одним из примеров является международная инициатива прозрачности в добывающих отраслях [7], политика в отношении раскрытия данных которой изложена в [8].

Обобщая рекомендации указанных выше организаций и инициатив, можно сформулировать следующие общие принципы публикации открытых данных:

1. Открытость «по умолчанию». Раскрытию подлежат все данные, представляющие потенциальный интерес для общества, граждан, бизнеса, государства и способствующие экономическому и социальному развитию, при условии, что публикация этих данных не нарушает чьих-либо прав и не противоречит действующему законодательству, законодательству в области защиты персональных данных, а также если эти данные не относятся к информации ограниченного доступа.
2. Полнота и детализация. Следует предоставлять данные, по возможности, в «сыром» виде, без их статистической обработки или агрегации. Возможность обработки данных следует предоставить пользователю, чтобы не лишать его важных сведений.
3. Техническая открытость. Данные публикуются в открытом машиночитаемом формате. Наиболее простым вариантом является табличный формат (XLS, CSV), более технологичным — формат XML, JSON. Предпочтительным является онтологический формат (RDF, OWL), обеспечивающий интероперабельность и семантическую связность данных путем использования принципов, стандартов и спецификаций технологического стека Semantic Web.
4. Доступность. Публикуемые данные должны быть доступны в сети без процедур аутентификации, в том числе для поисковых машин. Сервер может предоставлять доступ к данным посредством выполнения запросов к API, но одновременно должны быть доступны полные наборы данных для скачивания. На сайте организации должен быть доступен каталог наборов данных с описанием структур наборов данных.
5. Актуальность. Открытые данные должны публиковаться с явным указанием момента или периода времени, на который они актуальны. Данные, обновляемые периодически, должны публиковаться регулярно и своевременно, по мере их обновления, с сохранением доступа к ранее опубликованным наборам данных.
6. Юридическая открытость. Поскольку законодательство запрещает использование объек-

тов интеллектуальной собственности без явного разрешения, открытые данные должны распространяться с открытой лицензией Creative Commons [9] или Open Data Commons [10].

Система образования в Российской Федерации состоит преимущественно из государственных образовательных учреждений, финансируемых из бюджета. Образовательные программы лицензируются и аккредитуются соответствующим государственным органом. Устанавливаются единые требования к абитуриентам вузов, ключевым из которых является сдача единого государственного экзамена. Также жестко регламентированы управленческие процедуры, отчетность и документирование деятельности образовательных учреждений. В таких условиях практически полного контроля ответственность за повышение прозрачности и доступности информации сферы образования несет государство. Среди прочих многочисленных регламентирующих документов необходимо выделить приказ Рособрнадзора от 14 августа 2020 года N831 «Об утверждении Требований к структуре официального сайта образовательной организации в информационно-телекоммуникационной сети Интернет и формату представления информации». В данном документе приведен обширный перечень данных, которые должны публиковаться на сайте образовательной организации, например, сведения о реализуемых образовательных программах, численности обучающихся, педагогических работниках, материальном обеспечении и многое другое. Оценивая данную инициативу Рособрнадзора как безусловно позитивную, следует отметить, что практически все перечисленные в упомянутом приказе сведения вполне можно отнести к открытым данным. Однако, по причине отсутствия явного указания на это, образовательные учреждения при публикации этих сведений на сайте не придерживаются принципов работы с открытыми данными. В частности, информация размещается преимущественно в форматах, не являющихся машиночитаемыми, а там, где существует возможность скачать данные в табличном формате, отсутствует стандартизованное описание набора данных в соответствии с [6].

С другой стороны, Минобрнауки РФ, публикуя открытые данные на соответствующем портале <http://открытые-данные.минобрнауки.рф>, делает это согласно методическим рекомендациям [6], однако перечень имеющихся на портале наборов данных ограничен. Например, здесь можно найти список подведомственных Минобрнауки РФ образовательных организаций, однако более подробные сведения, например, о реализуемых ими образовательных программах, отсутствуют. Причиной этого, очевидно, является невозможность агрегировать на портале данные с более высокой сте-

пенью детализации по причине их большого объема и частого обновления. Последнее обстоятельство позволяет сделать важный вывод — открытые данные должны публиковаться непосредственно их источником, то есть образовательной организацией, а не вышестоящим госорганом.

Другой серьезной проблемой в области открытых данных системы образования является отсутствие какой-либо связности между публикуемыми наборами данных. Это означает, что каждый публикуемый набор данных представляет собой либо «плоскую» таблицу, либо структуру с ограниченным количеством уровней вложенности, элементы которой никак не связаны с другими наборами данных. Сопоставление элементов одного набора данных элементам другого возможно только «вручную» специалистом-экспертом предметной области. Это создает непреодолимые препятствия на пути построения эффективных сервисов обработки открытых данных и сдерживает процесс их публикации, негативно влияя на степень открытости системы образования.

О низком качестве управления данными в сфере образования говорится в Стратегии цифровой трансформации отрасли науки и высшего образования [11]. Один из проектов Стратегии («Датахаб») призван систематизировать процесс управления данными в сфере образования и науки и сделать его более эффективным.

Как было отмечено выше, российская система образования управляется централизованно и пронизана общими регулирующими нормами, стандартами и отчетностью. В этих условиях вполне естественным была бы разработка единого регламента публикации открытых данных всеми образовательными учреждениями. Такой регламент, по мнению авторов, должен основываться на универсальной семантической модели открытых данных, разработка которой является важнейшей перспективной задачей в области работы с открытыми данными системы образования. Главной задачей, которую должна решать универсальная семантическая модель открытых данных системы образования является обеспечение семантической связности данных. Это качество модели может быть достигнуто при переходе к «знаниевой» модели открытых данных, в которой будут определены универсальные способы представления семантики, а сами открытые данные будут от этой семантики неотделимы. В основе такого подхода, называемого «онтологическим», лежит представление любой информации в виде семантического графа произвольной структуры, узлами которого являются концепты, а ребра — отношениями. Наиболее известным применением онтологического подхода является проект Semantic Web — семантический web

[12]. Это концепция развития web-среды по внедрению в информацию метаданных, которые позволят хранить и обрабатывать семантику документов. Кроме того, использование специальных инструментов для извлечения данных позволит осуществлять не контекстный, а «семантический поиск», то есть получать информацию с использованием логического вывода.

По мнению авторов, технологии Semantic Web должны стать основой для работы с открытыми данными. Отличительным качеством семантической модели открытых данных является ее расширяемость «снизу», то есть по инициативе конечных пользователей. Для этого субъекты, публикующие открытые данные, могут самостоятельно дополнять модель собственными элементами, определяя семантику их связей с существующими элементами без внесения изменений в «ядро» модели. Учитывая это, можно определить следующий перечень задач, которые предстоит решить для построения эффективной системы публикации открытых данных сферы образования:

1. Разработка универсальной семантической модели (онтологии) открытых данных сферы образования.
2. Размещение модели в виде спецификации и методических рекомендаций к ней в открытом доступе.
3. Создание информационных ресурсов, содержащих общедоступные классификаторы и справочники, интегрированные в модель на технологическом уровне.
4. Организация постоянно действующей площадки для обсуждения модели и обратной связи пользователей с экспертами — разработчиками модели.
5. Внедрение программных инструментальных средств для подготовки и публикации открытых данных в соответствии с разработанной моделью.
6. Разработка методики и процедур мониторинга и оценки субъектов, публикующих открытые данные с точки зрения их полноты и актуальности.

Выводы

В настоящее время в российской системе образования процесс публикации открытых данных представляется малоэффективным. К причинам этого относятся слабая вовлеченность в процесс образовательных учреждений, отсутствие отраслевых стандартов и методик по работе с открытыми данными, а также отсутствие семантической связности между теми наборами данных, которые все же публикуются. Существенным шагом к улучшению ситуации может стать разработка универсальной семантической модели открытых дан-

ных и ее реализация с использованием технологий Semantic Web. Разработка и внедрение модели в практику работы образовательных учреждений будет наи-

более эффективной, если будет поддержана государственными органами управления и контроля в сфере образования.

ЛИТЕРАТУРА

1. Гордеев М.О., Гришмановский П.В. Роль открытых данных // Вестник кибернетики, No. 3, 2017. pp. 26–30.
2. Федеральный закон от 9 февраля 2009 г. N8-ФЗ Об обеспечении доступа к информации о деятельности государственных органов и органов местного самоуправления // Открытые данные России. URL: <https://data.gov.ru/federalnyy-zakon-ob-obespechenii-dostupa-k-informacii-o-deyatelnosti-gosudarstvennyh-organov-i>
3. Open Knowledge Foundation. Руководство по открытым данным [Электронный ресурс] // Open Data Handbook: [сайт]. URL: <http://opendatahandbook.org/guide/ru/>
4. Open Government Data Toolkit [Электронный ресурс] // The World Bank Group: [сайт]. [2019]. URL: <https://opendatatoolkit.worldbank.org/en/>
5. АНО «Информационная культура». // Информационная культура: [сайт]. URL: <https://www.infoculture.ru/>
6. Минэкономразвития России. Методические рекомендации по публикации открытых данных версия 3.0 // Открытые данные России. URL: <https://data.gov.ru/metodicheskie-rekomendacii-po-publikacii-otkrytyh-dannyh-versiya-3-0>
7. Инициатива прозрачности в добывающих отраслях. Глобальный стандарт для эффективного управления нефтью, газом и минеральными ресурсами [Электронный ресурс] URL: <https://eiti.org/>
8. Инициатива прозрачности в добывающих отраслях. Международный секретариат. Политика в отношении открытых данных и раскрытие данных. Руководящая записка // Глобальный стандарт для эффективного управления нефтью, газом и минеральными ресурсами. URL: https://eiti.org/files/documents/guidance_note_19_on_open_data_clean_ru.pdf
9. Creative Commons Corporation. License Chooser [Электронный ресурс] // Creative Commons: [сайт]. URL: <https://creativecommons.org/choose/>
10. Open Knowledge Foundation. Conformant Licenses [Электронный ресурс] // Open Definition: [сайт]. URL: <https://opendefinition.org/licenses/>
11. Минобрнауки РФ. Стратегия цифровой трансформации отрасли науки и высшего образования 2021. URL: https://www.minobrnauki.gov.ru/documents/?ELEMENT_ID=36749&fbclid=IwAR3k4F7qUFLW0EFZdRjM8aDFrn73PpqrhemtNqBU883pHwsaHDhgoE9YkU
12. The World Wide Web Consortium (W3C). Semantic Web [Электронный ресурс] URL: <https://www.w3.org/standards/semanticweb/>

© Кузин Дмитрий Александрович (kuzin_da@surgu.ru), Даниленко Иван Николаевич (vice1@surgu.ru).

Журнал «Современная наука: актуальные проблемы теории и практики»