

Аналитический доклад

**Работа с открытыми данными:
особенности публикации
и использования
в российском правовом поле**

Доклад предназначен для всех публикаторов и пользователей открытых данных и ориентирован на распространение лучших практик при подготовке и использовании наборов данных без нарушения существующих норм.

Доклад подготовлен



**Информационная
культура**



**ЦЕНТР
ЦИФРОВЫХ
ПРАВ**

При поддержке



Microsoft

Оглавление

От авторов	4
Глоссарий	5
Введение и контекст	8
Открытые данные и государство	8
Проекты на основе государственных открытых данных.	9
Общественные открытые данные	11
Стандарты для открытых данных	12
Основные барьеры развития открытых данных	12
Политические барьеры	12
Технические барьеры	13
Культурные барьеры	14
Правовое поле открытых данных	15
Государственные открытые данные	15
Открытые данные и искусственный интеллект	17
Чувствительная информация.	19
Лицензии	20
Практические рекомендации по работе с открытыми данными	23
Паспортизация наборов данных	23
Выбор лицензии	24
Анонимизация данных.	28
Технические аспекты публикации.	30
Метаданные	30
Форматы файлов/публикаций	31
Типичные ошибки при публикации материалов	32
Использование открытых данных	33
Агрегация данных	33
Аналитика на основе данных	34
Повышение качества данных.	34
Примеры проектов по публикации данных	36
Рекомендации по развитию регулированию и внедрению открытых данных на государственном уровне	38

От авторов

Феномен открытых данных сегодня стал широко распространен, и рассматривается не только как один из аспектов развития прозрачности государства – в связи с чем чаще всего и употребляется это понятие, – но и в целом как один из драйверов развития. Открытые данные, как и многие другие концепции открытости: открытый код, открытый доступ, свободные лицензии, имеют множество практических областей применения как в общественной жизни, так и в коммерческих продуктах.

Открытые данные – это общедоступная информация, размещенная в Интернете для неоднократного, свободного и бесплатного использования в виде машиночитаемых систематизированных данных, в формате, позволяющем их автоматизированную обработку.

Пока ещё далеко от того, чтобы можно было говорить о полностью сформировавшейся культуре публикации и использования открытых данных, но Россия, как и весь мир в целом, двигаются к этому. Важным шагом для нашей страны стало начало со стороны государства работы в области открытых данных, появление соответствующего пласта нормативных актов, определяющих необходимость ведомствам публиковать открытые данные, и специализированных ресурсов как на уровне отдельных ведомств, так и интегральных порталов – общественного «Хаба открытых данных»¹ или официального портала государственных открытых данных². Хотя эту деятельность сложно назвать в полной мере успешной^{3,4}, но она дала важный толчок для развития направления и его восприятия в обществе. Например, сегодня открытые данные ФНС⁵ помогают в проверке контрагентов и позволяют разработчикам встраивать в свои приложения и веб-сервисы официальные адресные справочники.

Всплеск интереса к тематике открытых данных подстегивает развитие технологий искусственного интеллекта: для того, чтобы использовать технологии машинного обучения, необходимы массивы данных – так называемые «обучающие выборки», которые в идеале должны быть общедоступными для того, чтобы можно было получить сравнимость и верифицируемость результатов различных решений. Использование открытых данных способствует появлению таких инициатив, как World AI&Data Challenge⁶ – конкурс, привлекающий к решению социальных задач широкий круг специалистов в области искусственного интеллекта.

В этом докладе мы постарались отразить имеющуюся ситуацию и дать навигацию на более подробные материалы, а также предложить основные рекомендации как по использованию конкретных практик, так и по совершенствованию нормативной базы с тем, чтобы стимулировать развитие данной области. Не претендуя на исчерпывающую полноту, мы надеемся, что эта работа принесет пользу не только советами и ссылками, но и послужит материалом для дальнейшей дискуссии.

1 hubofdata.ru

2 data.gov.ru

3 [Открытость государства в России 2019](#)

4 [Открытость государства в России 2020](#)

5 www.nalog.ru/opendata

6 datamasters.ru

Глоссарий

Агрегированные данные – данные, сведенные вместе по тому или иному признаку непосредственно или с использованием определенного алгоритма обработки. Допустим, есть реестр контрактов, где каждая запись содержит сведения о сумме контракта, заказчике и поставщиках. Из такого реестра можно получить, например, список всех заказчиков, которые встречаются в контрактах, и для каждого заказчика посчитать, сколько контрактов он заключил и какова сумма заключенных им контрактов. Полученные цифры будут агрегированными данными.

База данных – система хранения комплексной информации, описывающей ту или иную предметную область. Допускает сложную иерархию данных и групп данных, а также позволяет выстраивать связи между различными группами данных. Также обладает инструментами фильтрации для быстрого поиска необходимых данных по заданным признакам.

Данные – поддающееся многократной интерпретации представление информации в формализованном виде, пригодном для передачи, связи, или обработки. Синонимом понятия данные является структурированная информация.

Документ – в данном случае файл немашиночитаемого формата, содержащий неструктурированную информацию. Это может быть, например, текст или изображение в форматах DOC(X), PDF, JPG, PNG и т.п. Мы специально добавили это понятие в наш небольшой глоссарий, чтобы подчеркнуть разницу. При этом довольно часто открытые данные могут ссылаться на те или иные документы, использование которых может быть неотъемлемой частью работы с данными.

Машиночитаемые форматы данных – форматы представления и публикации данных, предназначенные для обработки машинными методами. К ним относятся такие форматы, как: CSV (TSV), JSON, XML, XLSX, XLS и другие.

Метаданные – данные, описывающие набор данных или документ. Например, даты его создания и обновления; автора; публикатора; контактную и лицензионную информацию и т.п.

Набор данных – совокупность данных, представленных в машиночитаемом формате и описывающих те или иные процессы, явления или предметную область. В отличие от базы данных, обычно обладают плоской или несложной иерархической структурой и сравнительно небольшим объемом, поэтому для их публикации часто используются табличные форматы (CSV, XLS/XLSX).

Неструктурированная информация – информация, представленная в форме, пригодной, прежде всего, для человеческого восприятия. Например, текст, изображение, видео- или аудиозапись. В ряде случаев неструктурированную информацию возможно и даже целесообразно преобразовывать в структурированную.

Обладатель данных – в контексте открытых данных под этим подразумевается лицо, получившее доступ к данным. Однако тут есть тонкость – схожее понятие присутствует

в законе «Об информации»⁷, где «обладатель информации» определяется как «лицо, самостоятельно создавшее информацию либо получившее на основании закона или договора право разрешать или ограничивать доступ к информации, определяемой по каким-либо признакам».

Обезличенные данные – данные, которые нельзя соотнести с каким-либо конкретным субъектом. Федеральное законодательство не поясняет, что такое обезличенные данные, однако дает понятие «обезличивания персональных данных»: действия, в результате которых становится невозможным без использования дополнительной информации определить принадлежность персональных данных их конкретному субъекту. Также в законодательстве нет понятия анонимизированных данных, но это понятие есть в европейском регламенте GDPR (General Data Protection Regulation). По сути, это окончательно обезличенные данные – то есть данные, которые нельзя соотнести с их конкретным владельцем, проведя «реидентификацию». Если же такое возможно, то такой вид обезличенных данных называют «псевдоанонимизированными».

Открытые данные – данные, опубликованные в публичном доступе в сети Интернет в машиночитаемом формате и под свободной (открытой) лицензией, допускающей бесплатное повторное использование этих данных, их переработку и распространение. Приведенное каноническое определение затрагивает только машиночитаемые данные, однако практически все (кроме узко технических) рекомендации и нормы по работе с открытыми данными затрагивают и публикацию данных в не машиночитаемых форматах (например, просто текстовом).

Персональные данные – любая информация, относящаяся к определенному или определяемому на ее основании физическому лицу (субъекту персональных данных). Например, фамилия, имя, отчество в комбинации с датой и местом рождения или адресом проживания уже позволяют это сделать.

Повторное использование – практика многократного использования (в исходном или переработанном виде) одних и тех же данных или документов.

Свободная (открытая) лицензия – лицензия, дающая максимально широкие права на использование данных и документов, в том числе право на бесплатное повторное использование, переработку и распространение. Лицензии, требующие при повторной публикации ссылки на источник или указания авторства, обычно относят к открытым. С оговорками к ним иногда относят лицензии, требующие, чтобы на продукты, созданные с использованием лицензируемых материалов, распространялись те же права, что и на исходный материал. Лицензии, которые добавляют к этому запрет на коммерческое использование, к открытым не относятся.

Сырые данные – данные, не прошедшие какую-либо обработку и представленные в своем первичном виде. Пример не сырых (обработанных) данных это статистика, обобщающая показатели исходного набора, или агрегированные данные. То, на основе чего рассчитана эта статистика или произведена агрегация, и есть сырые данные.

⁷ Федеральный закон от 27.07.2006 N 149-ФЗ (ред. от 03.04.2020) «Об информации, информационных технологиях и о защите информации»

Человекочитаемые форматы данных – форматы представления и публикации данных, предназначенные для восприятия человеком. Например, DOC/DOCX, PDF, PPT/PPTX, JPEG, PNG, TIFF, GIF, HTML и т.п. Хотя некоторые из них относительно легко поддаются машинной обработке (как, например, HTML), их главная задача в том, чтобы упростить человеческое зрительное восприятие (зачастую в ущерб структурированности и последовательности, необходимым для корректной машинной обработки).

Чувствительные данные – данные, раскрытие которых может нести риски для субъекта, к которому они относятся. Для физического лица к ним можно отнести, например, информацию об этническом или расовом происхождении, состоянии здоровья, политических и религиозных убеждениях, а для коммерческого – информацию о производственных процессах и особенностях сделок. Этот термин не имеет сейчас определения в российском законодательстве, но уже определен в рамках европейского регламента по защите данных GDPR (General Data Protection Regulation).

Правительства по всему миру предпринимают существенные усилия в публикации открытых данных, собираемых в государственных информационных системах – и здесь Россия полноценный участник процесса. По некоторым аспектам наша страна находится даже среди лидеров открытости, обладая по мнению экспертов самой открытой системой государственных закупок, а также входя в лидеры по открытости бюджетного процесса¹⁰. Хотя при этом открытость государственных информационных систем находится в целом на недостаточном уровне, как показало недавнее исследование Счетной палаты¹¹.

Продолжая развиваться как со стороны государственных органов, так и силами общественных организаций, работа с открытыми данными внутри государства имеет ряд проблем и противоречий:

1. **Технологии:** Довольно большой массив данных уже открыт, но с точки зрения технических форматов публикации представлен в виде, который затрудняет анализ и сопоставление данных, в том числе использование алгоритмов для анализа.
2. **Регулирование:** Отсутствие уважаемого и весомого координатора открытости, который мог бы задавать стандарты открытости, убеждать ведомства открываться по единым правилам.
3. **Противодействие:** Наличие мощных стейкхолдеров, не заинтересованных в раскрытии некоторой информации, в основном связанными с ведомственными и личными интересами.
4. **Спецрежимы:** Большое количество данных имеют гриф, указывающий на наличие коммерческой тайны или личных данных. Это не всегда может быть оправданно реальным содержанием информации.
5. **Достоверность данных:** Значительная часть содержащихся в открытом доступе данных не являются достоверными.

Системная скоординированная работа по установлению новых стандартов и внедрению технологических решений, наращиванию компетенций госслужащих, преодолению инерции и противодействия привыкших к закрытости ведомств позволяет обеспечить высокую доступность данных в случае восстановления повестки открытости в деятельности российского правительства.

Проекты на основе государственных открытых данных

Существует множество проектов, созданных на базе открытых государственных данных, которые могут касаться как самого государства (например, повышая его прозрачность и давая инструменты для общественного контроля), так и самых различных сфер жизни. Вот несколько наиболее ярких примеров отечественных и зарубежных проектов:

Системы проверки контрагентов

- ~ Зарубежные – craft.co, factset.com и другие используют открытые источники такие как налоговая служба США, комиссия по ценным бумагам США и иные официальные открытые данные для предоставления расширенных справок по компаниям.
- ~ Российские – Контур.Фокус (focus.kontur.ru), СПАРК Интерфакс (spark-interfax.ru), ФИРА

¹⁰ [Open Budget Index 2019](#)

¹¹ [Оценка открытости государственных информационных систем в России](#)

(fra.ru), Коммерсантъ Картоотека (kartoteka.ru) и другие используют опубликованные открытые данные о юридических лицах для создания сервисов проверки благонадёжности контрагентов

Порталы открытости госзакупок

- ~ Зарубежные – согласно ежегодному рейтингу Outsell (outsellinc.com/outsells-research), ведущие компании в сфере проверки контрагентов это Dun & Bradstreet (dnb.com), Bureau Van Dijk (bvdinfo.com), Bisnode (bisnode.com), Experian (experianplc.com), Crif (crif.com), Data Axle (в прошлом Infogroup, data-axle.com), Creditreform (creditreform.com), а также бесплатная Open Corporates (opencorporates.com), используют открытые источники, такие как налоговая служба США, комиссия по ценным бумагам США и иные официальные открытые данные, для предоставления расширенных справок по компаниям.
- ~ Российские – проект Госзатраты (clearspending.ru) обеспечивает открытость государственных расходов для граждан в всем государственным и муниципальным контрактам.

Доходы чиновников

- ~ Зарубежные – проект Little Sis (littlesis.org) позволяет осуществлять мониторинг деятельности лоббистов на основании публично опубликованных сведений.
- ~ Российские – проект Декларатор (declarator.org) позволяет осуществлять общественный контроль финансовых деклараций публичных лиц

Поисковые системы

- ~ Зарубежные – самые популярные поисковики google.com, bing.com и др. используют данные Википедии, официальные государственные открытые данные и иные источники открытых данных. Так, для обеспечения этого Google купил компанию Metaweb, занимавшуюся интеграцией открытых данных в единую базу знаний.
- ~ Российские – yandex.ru использует открытые данные для интеграции в сервисы карт и корректировки поиска, а также построения добавляемых в выдачу подсказок–«колдунчиков».

Другие применения:

- ~ Российские:
 - ~ ЦИАН (cian.ru) – коммерческий сервис поиска недвижимости включает данные о качестве школ, уровне преступности, качестве зданий полученные из открытых официальных источников (открытых данных) для большего удобства его пользователей и возможности принимать осознанные решения.
 - ~ Проект «Если быть точным» (tochno.st) предоставляет детальную статистику и визуализацию по проблемам социальной политики и здравоохранения, помогая принимать более ответственные решения официальными лицами.
 - ~ «Университетская информационная система Россия» (uisrussia.msu.ru) агрегирует официальную государственную статистику для её использования в академических исследованиях.

- ~ Проект «Российские школы» (russianschools.ru) агрегирует все доступные сведения о школах, помогая исследователям и официальным лицам сферы образования, а также делающим выбор родителям.
- ~ Система раскрытия информации о некоммерческих организациях «Открытые НКО» (openngo.ru) позволяет донорам получить профиль деятельности различных некоммерческих организаций на основе полученных ей государственных субсидий, грантов и контрактов.
- ~ Зарубежные:
 - ~ Предсказание лихорадки денге в Парагвае (odimply.org/case-paraguays-dengue-prediction.html) на основе открытых данных.
 - ~ Проект Humanitarian Data Exchange (data.humdata.org) по обмену данными между организациями, осуществляющими помощь при чрезвычайных ситуациях в целях повышения координации и качества прилагаемых ими усилий.

Общественные открытые данные

Помимо государственных открытых данных заметным является и сектор общественных открытых данных. Наборы данных в нем могут быть собраны самими разными субъектами – от учёных до межправительственных организаций. Наиболее яркими примерами являются:

- ~ Экосистема Фонда Викимедия (Wikimedia Foundation) – DBPedia (wiki.dbpedia.org), WikiData (www.wikidata.org/wiki/Wikidata:Main_Page)
- ~ Порталы научных данных, хорошим поисковым ресурсом по репозиториям которых является проект www.re3data.org
- ~ Хаб открытых данных от OKF и Datopian – datahub.io
- ~ Агрегатор данных Data.world – data.world
- ~ Проект Dataverse Гарвардского университета – dataverse.harvard.edu
- ~ Портал раскрытия научных данных EC Zenodo – zenodo.org
- ~ Наборы открытых данных на платформе Azure – azure.microsoft.com/en-us/services/open-datasets
- ~ Порталы данных межгосударственных структур: Всемирный банк (data.worldbank.org), ООН (data.un.org), ВОЗ (www.who.int/data/gho) и др.

Обращает на себя внимание то, что с российскими примерами в этой области дело обстоит заметно хуже. Как пример можно привести «Портал открытых данных НКО» ngodata.ru, но он пока он довольно слабо наполнен: некоммерческие организации предпочитают генерируемые в рамках своих проектов данные оставлять себе, либо вообще не осознавая их значимости, либо считая собственным активом. Редким примером научных открытых данных в России является «Национальный корпус русского языка» ruscorpora.ru. Постепенно круг таких проектов расширяется. Дополнительным стимулом к этому стало развитие технологий машинного обучения, подчеркнувшее важную инфраструктурную роль наличия общедоступных наборов данных.

Однако развитие такого рода проектов в России сдерживается не только отсутствием традиций открытых публикаций научных данных, но и отсутствием отработанных механизмов и традиций общественного финансирования для них – а поддержание проектов публикации открытых данных требует определенных затрат.

Стандарты для открытых данных

В мире уже сложились основные принципы и стандарты в области публикации открытых данных. Они инвариантны для разных стран и могут успешно применяться, не требуя содержательной переработки, так как в целом отвечают и общечеловеческим ценностям, и современному состоянию технологий:

- ~ Принципы Хартии открытых данных – opendatacharter.net/principles
- ~ Стандарт публикации открытых данных Frictionless Data – frictionlessdata.io
- ~ Открытые стандарты публикации данных в Project Open Data – project-open-data.cio.gov/open-standards
- ~ Открытые стандарты для данных от The ODI – standards.theodi.org
- ~ Открытые стандарты для данных от Университета Джона Хопкинса – datastandards.directory

С ориентацией на этот набор стандартов проектируются и платформы, предназначенные для размещения данных, поэтому можно рекомендовать следовать им там, где возможно, – в том числе и для облегчения решения технических вопросов.

Основные барьеры развития открытых данных

Политические барьеры

Нежелание чиновниками повышать открытость

Открытость данных, как и другие формы открытости деятельности государства, повышают активность общественного контроля. Вслед за этим растет и нагрузка на чиновников – от необходимости дополнительных работ по обеспечению открытости до увеличения объема взаимодействия с гражданами и общественными структурами, не связанными прямо с основным функционалом. Поэтому даже в случаях, когда нет рисков проведения политических расследований или выявления коррупции, во всех странах процесс повышения открытости вызывает естественное сопротивление со стороны чиновников.

Способы преодоления барьера:

1. Последовательная политика усиления влияния представительной власти и установление справедливых законов, формирующих открытость государства и общества.
2. Принятие международных соглашений и вступление в международные организации, такие как Open Data Charter (Хартия Открытых Данных) участие в которых предполагает обязательства по обеспечению открытости государства, включая открытость данных.

Ориентация государства на монополию в защите интересов граждан и оказании услуг

Государство часто стремится предоставлять услуги гражданам, заменяя ими рыночные отношения. В вопросах открытости данных представители государственного аппарата ссылаются на то что все те услуги которые могли бы быть созданы коммерческим сектором на основе открытых данных, в том или ином виде создаются самими органами власти и государственными учреждениями. Такая позиция особенно сильно выражена в нашей стране, но аналогичное поведение наблюдается и во многих других странах, включая развитые демократические.

Способы преодоления барьера:

1. Проведение и публикация исследований об экономическом эффекте открытости данных, сбор примеров использования открытых данных в задачах, приносящих общественную пользу, создающих рабочие места и дающих вклад в рост экономики.
2. Постоянный диалог с лицами, принимающими решения в государственном аппарате, по теме значимости экономического эффекта от сохранения и развития политики открытости данных.

Технические барьеры

Дефицит квалифицированных кадров по работе с данными

Государственная служба часто проигрывает конкуренцию за квалифицированные кадры коммерческим компаниям, особенно в технологических секторах экономики. Дефицит ИТ-сотрудников, в особенности специалистов по работе с данными^{12,13}, вызывает значительные сложности не только в работе органов власти с открытыми данными, но и в выстраивании политики работы с данными внутри органа власти в целом.

Способы преодоления барьера

1. Формирование модели ИТ-компетенций¹⁴ и включение соответствующих курсов в обучение государственных гражданских служащих на уровне бакалавриата и магистерских программ в профильных высших учебных заведениях.
2. Создание центров переподготовки и дополнительного образования для подготовки государственных служащих включая высший руководящий состав органов власти.
3. Оформление ИТ-специалистов и формирование ИТ-команд в органах власти внутри государственных учреждений обеспечивая возможность выплачивать конкурентные зарплаты.
4. Обеспечение открытости кода государственных информационных систем для возможности повторного использования кода, в том числе применяемого для проектов по работе с открытыми данными. Пример: открытый код Счетной палаты РФ¹⁵

12 [Развитие цифровых компетенций государственных гражданских служащих Российской Федерации](#)

13 [Государство как платформа: люди и технологии](#)

14 [Модель компетенций команды цифровой трансформации в системе государственного управления](#)

15 code.ach.gov.ru/public

Культурные барьеры

Отсутствие культуры открытости в государственном аппарате

В бюрократических системах часто складываются разные практики восприятия инноваций. Открытость государства – это безусловная инновация, а открытые данные – это одна из форм этой открытости. Культура открытости может отсутствовать в государственном аппарате, не только по причине протекционизма государственной политики, но и в силу исторических и культурных причин.

Способы преодоления барьера

1. Принятие федеральных законов с чётко прописанными требованиями открытости.
2. Публичная демонстрация и акцент на открытости таких высших органов государственной власти как: правительство, парламент, министерства и иные органы власти.
3. Непрерывное просвещение, обучение и демонстрация пользы открытости данных для сотрудников государственного аппарата.

Отсутствие понимания пользы от открытости данных в практике госуправления

Сопrotивление государственных служащих открытости данных может быть следствием и практического непонимания применения открытых данных в их работе и, соответственно, снижение приоритета задачам связанных с открытостью данных по сравнению с другими задачами в государственном управлении.

Способы преодоления барьера

1. Приоритет открытым данным может быть обеспечен наличием постоянно действующего активного консультативного или исполнительного органа при Правительстве обеспечивающим формирование национального плана открытости и мониторинг его исполнения.
2. Подготовка методических материалов, объясняющих не только подходы, но и практику обеспечения открытости данных в работе органов власти.

Правовое поле открытых данных

Правовое поле открытых данных определяется тремя основными направлениями:

1. Нормативная база, обеспечивающая публикацию открытых данных государством. Помимо очевидного аспекта обеспечения подконтрольности и прозрачности государственного управления, это критически важный аспект распространения культуры открытых данных в обществе – особенно в условиях государственно-центричной ситуации в России. Закрепленная обязанность государственных органов публиковать такие данные и появление различных коммерческих и общественных проектов, базирующихся на них, создает важные прецеденты и позволяет отработать необходимые практики.
2. Правовые вопросы, связанные с самими публикуемыми данными. Важнейшим аспектом тут является регулирование публикации чувствительных данных и, в первую очередь, соблюдение приватности граждан. К этому же направлению относится обеспечение коммерческой тайны, авторских прав и т.д.
3. Лицензии, регламентирующие использование данных – и определяющих правовые отношения пользователя данных и их публикатора.

Государственные открытые данные

На государственном уровне в России понятие «открытые данные» и общие правила работы с ними, включая необходимость публикации, определяет Постановление Правительства РФ от 10.07.2013 г. № 583 «Об обеспечении доступа к общедоступной информации о деятельности государственных органов и органов местного самоуправления в информационно-телекоммуникационной сети «Интернет» в форме открытых данных»¹⁶. Требования этого документа выполняются не всегда, особенно на местном уровне. Тем не менее он является хорошим ориентиром, задающим некоторый минимальный уровень «хорошей практики» в этой области.

Существующее нормативно-правовое обеспечение задает различные типы доступа к данным, определяющие, в том числе, возможность их публикации в качестве открытых: полностью свободный (любой пользователь может получить данные без какой-либо аутентификации), частично свободный (данные можно получить после аутентификации), полностью закрытый (касается персональных данных, государственной тайны и т.п.). В зависимости от этого есть различные документы, регулирующие публикацию данных:

1. Федеральный закон от 27 июля 2006 г. №149-ФЗ «Об информации, информационных технологиях и о защите информации»;
2. Федеральный закон от 9 февраля 2009 г. №8-ФЗ «Об обеспечении доступа к информации о деятельности государственных органов и органов местного самоуправления»;
3. Федеральный закон от 7 июня 2013 г. №112-ФЗ «О внесении изменений в Федеральный закон «Об информации, информационных технологиях и о защите информации» и Фе-

¹⁶ [Постановление Правительства РФ от 10.07.2013 г. № 583 «Об обеспечении доступа к общедоступной информации о деятельности государственных органов и органов местного самоуправления в информационно-телекоммуникационной сети «Интернет» в форме открытых данных»](#)

деральный закон «Об обеспечении доступа к информации о деятельности государственных органов и органов местного самоуправления»;

4. Постановление Правительства Российской Федерации от 24 ноября 2009 г. №953 «Об обеспечении доступа к информации о деятельности Правительства Российской Федерации и федеральных органов исполнительной власти»;
5. Распоряжение Правительства Российской Федерации от 10 июля 2013 г. №1187-р «О перечнях информации о деятельности государственных органов, органов местного самоуправления, размещаемой в сети «Интернет» в форме открытых данных»;
6. Распоряжение Правительства Российской Федерации от 30 января 2014 г. №93-р «Об утверждении Концепции открытости федеральных органов исполнительной власти»;
7. Приказ Минкомсвязи России от 27 июня 2013 г. №149 «Об утверждении Требований к технологическим, программным и лингвистическим средствам, необходимым для размещения информации государственными органами и органами местного самоуправления в сети «Интернет» в форме открытых данных, а также для обеспечения ее использования»;
8. Закон РФ № 5485-1 от 21.07.1993 «О государственной тайне»;
9. Федеральный закон от 27.07.2006 N 152-ФЗ (ред. от 21.07.2014) «О персональных данных» (с изм. и доп., вступ. в силу с 01.09.2015);
10. Федеральный закон от 24.04.2020 № 124-ФЗ «О внесении изменений в отдельные законодательные акты Российской Федерации по вопросам обеспечения устойчивого развития экономики в условиях ухудшения ситуации в связи с распространением новой коронавирусной инфекции» (вносящий, в том числе, изменения по закрытию публикации конкурсов оборонной тематики).

Кроме того, необходимо учитывать различные нормативные акты, определяющие ограничения, касающиеся отдельных видов деятельности – например, таких как профессиональная тайна врачей (Федеральный закон от 21.11.2011 N 323-ФЗ «Об основах охраны здоровья граждан в Российской Федерации»). Это достаточно обширный пласт нормативных актов, приведение полного списка заняло бы не одну страницу, поэтому рекомендуем познакомиться с регулированием интересующей деятельности в специализированных источниках¹⁷.

Помимо нормативно-правовых актов большое влияние на открытость данных имеют установленные правительством планы:

- ~ планы открытости;
- ~ планы информатизации;
- ~ планы развития официальных сайтов;
- ~ планы мероприятий отдельных программ и ведомств.

17 www.consultant.ru/document/cons_doc_LAW_93980

Принципы открытости сформулированы в документах 1-8 приведенного выше перечня. Они определяют необходимые условия достижения поставленных целей от открытости данных, таких как: повышение прозрачности государства, поддержка бизнеса, вовлечение граждан в участие в государственном управлении. Фактически, в них отражены семь принципов хартии открытых данных: полнота, первичность, своевременность, открытость и доступность, пригодность к машинной обработке, свобода поиска, получения и распространения информации, соблюдение прав граждан и организаций

Регулярное соблюдение этих условий позволит в полной мере использовать предоставляемую информацию для улучшения как социального, так и экономического климата страны. Несмотря на достаточно развитую нормативную базу, некоторые положения нуждаются в совершенствовании и дополнении. Это важно для расширения круга органов, публикующих открытые данные в необходимом количестве и формате, а также обеспечения стандартизации.

Отчасти такая работа была проделана за время существования «Открытого правительства» когда органы власти получили обязанность публикации планов открытости, планов деятельности и раскрытия открытых данных на официальных сайтах. Однако после ликвидации «Открытого правительства» в 2018 году, контроль за соблюдением требований и рекомендаций по открытости прекратился и вопросы открытости отошли на второй план, а на вышедшие первый план инициативы по цифровизации государственного управления практически не учитывают этого аспекта.

Российской особенностью является то, что основные данные, как по объёму, так и по значению, накапливаются в государственных информационных системах. В докладе «Оценка открытости государственных информационных систем 2020 г»¹⁸ приведены примеры таких систем и оценена их текущая открытость. Из доклада следует, что текущее государственное регулирование не обеспечивает открытости данных в государственных информационных системах по умолчанию. Таких требований нет ни в конкурсной документации на разработку информационных систем, ни в методических рекомендациях, ни в типовых положениях для информационных систем.

В мировой практике важным направлением развития нормативной базы является внедрение принципа «открытость по умолчанию» (open by default). Текущее отечественное регулирование пока не использует этот принцип и, как следствие, в большинстве примеров государственных информационных систем недоступными остаются большая часть хранимых в них данных.

Открытые данные и искусственный интеллект

Открытые данные – это важнейший компонент многих общедоступных работ, связанных с искусственным интеллектом (ИИ). Например, без открытости наборов данных, были бы значительно ограничены возможности платформы Kaggle, предоставляющей доступ к большому числу наборов данных, используемых для машинного обучения и регулярно проводящей связанные с ними конкурсы¹⁹.

¹⁸ [Оценка открытости государственных информационных систем 2020](#)

¹⁹ www.kaggle.com/datasets

В Европейском союзе проводится мониторинг использования открытых данных для систем ИИ²⁰, а международная НКО Web Foundation проводит исследования по использованию открытых данных в ИИ, а также моральных и этических ограничений в применении новых технологий²¹.

Российская Национальная стратегия развития искусственного интеллекта на период до 2030 года²² включает положения о необходимости обеспечения доступности данных, в пункте V. Цели и основные задачи развития искусственного интеллекта указано в пункте 24:

в) повышение доступности и качества данных, необходимых для развития технологий искусственного интеллекта;

А также в пунктах 37 и 38:

37. Основными факторами развития технологий искусственного интеллекта являются увеличение объема доступных данных, в том числе данных, прошедших разметку и структурирование, и развитие информационно-коммуникационной инфраструктуры для обеспечения доступа к наборам таких данных.

38. Основными направлениями повышения доступности и качества данных, необходимых для развития технологий искусственного интеллекта, в соответствии с требованиями законодательства Российской Федерации являются:

а) разработка унифицированных и обновляемых методологий описания, сбора и разметки данных, а также механизма контроля за соблюдением указанных методологий;

б) создание и развитие информационно-коммуникационной инфраструктуры для обеспечения доступа к наборам данных посредством:

- ~ создания (модернизации) общедоступных платформ для хранения наборов данных, соответствующих методологиям описания, сбора и разметки данных;*
- ~ хранения наборов данных (в том числе звуковых, речевых, медицинских, метеорологических, промышленных данных и данных систем видеонаблюдения) на общедоступных платформах для обеспечения потребностей организаций - разработчиков в области искусственного интеллекта;*
- ~ установления приоритетного доступа российских государственных органов и организаций к общедоступным платформам.*

Отсутствие единой нормативной терминологии мешает употреблению определения «открытые данные» напрямую, а сама национальная стратегия развития искусственного интеллекта разрабатывалась и принималась в отсутствие национальной стратегии работы с данными, создание которой необходимо по мнению многих экспертов чтобы закрыть пробелы в нормативном регулировании в этой области.

Запущенный в настоящее время московский регуляторный эксперимент по развитию технологий искусственного интеллекта потребовал принятия специального федерально-

20 www.europeandataportal.eu/en/highlights/ai-and-open-data-crucial-combination

21 webfoundation.org/research/white-paper-series-opportunities-and-risks-in-emerging-technologies

22 www.garant.ru/products/ipo/prime/doc/72738946

го закона²³. В его рамках с 1 июля 2020 года сроком на 5 лет предусматривается введение специального правового режима «в целях создания необходимых условий для разработки и внедрения технологий ИИ, а также последующего использования результатов его применения». Одним из важных моментов в нем является возможность установления московским правительством по согласованию с Роскомнадзором специальных – более облегченных – условий по обработке обезличенных персональных данных. Но пока рано говорить о том, что именно будет происходить: никаких нормативных актов Правительства Москвы, определяющих уже конкретику реализации, на момент подготовки этого документа не появилось.

Чувствительная информация

При публикации открытых данных необходимо учитывать, что существует достаточно обширный пласт информации, раскрытие которой может нести риски для субъекта, к которому она относится. Хотя в целом понятие «чувствительных данных» или «чувствительной информации» в российском законодательстве не зафиксировано, существует ряд конкретных направлений, получивших своё нормативное регулирование.

В первую очередь это относится к персональным данным физических лиц, работа с которыми регулируется отдельным законом²⁴. Хотя многими экспертами отмечается его ориентация на реалии «доцифровой» эпохи, он формирует достаточно большой объем норм для сферы данных, которые требуют соблюдения.

Помимо персональных данных, существуют и другие виды информации, использование которой при формировании и публикации наборов открытых данных может стать препятствием из-за существующего режима защиты такой информации. Мы уже говорили об этом выше, в разделе, посвященном государственным открытым данным – нормы, касающиеся оборота информации с ограниченным доступом, актуальны не только для государственных данных.

Подробное рассмотрение вопросов работы с чувствительными данными выходит за рамки настоящего доклада. Но необходимо отметить, что так как глобальность интернета дает возможность использования данных вне страны, планируя ориентированную на это публикацию данных необходимо принимать во внимание и нормы международного регулирования.

Первым обязывающим инструментом, защищающим физических лиц от злоупотреблений, которые могут иметь место при сборе и обработке данных, и ставящий в то же время задачу регулировать трансграничный поток персональных данных была Конвенция о защите частных лиц в отношении автоматизированной обработки данных личного характера №108 Совета Европы²⁵, вступившая в действие ещё в 1985 году, а последние изменения,

²³ [Федеральный закон от 24.04.2020 № 123-ФЗ «О проведении эксперимента по установлению специального регулирования в целях создания необходимых условий для разработки и внедрения технологий искусственного интеллекта в субъекте Российской Федерации - городе федерального значения Москве и внесении изменений в статьи 6 и 10 Федерального закона “О персональных данных”»](#)

²⁴ [Федеральный закон РФ от 27 июля 2006 года № 152-ФЗ «О персональных данных»](#)

²⁵ [Конвенция о защите частных лиц в отношении автоматизированной обработки данных личного характера](#)

связанные с необходимостью актуализации, получившая 1999 году. К ней присоединилась и Россия, ратифицировав²⁶ в 2005 году.

Конвенция не только дает гарантии применительно к сбору и обработке персональных данных, но и запрещает, если национальное право не обеспечивает надлежащих гарантий, обработку чувствительных данных относительно расовой принадлежности лица, его политических взглядов, здоровья, религии, сексуальной жизни, уголовного прошлого и т.п. Конвенция также дает лицу право знать, что данные о нем собраны, и в случае необходимости иметь возможность их исправить. Ограничения изложенных в Конвенции прав возможны только в случае, когда под угрозой оказываются высшие интересы (безопасность государства, интересы обороны и т.д.).

Конвенция также предписывает определенные ограничения применительно к трансграничным потокам личных данных в те государства, где правовое регулирование не обеспечивает их должной защиты.

В Европейском союзе пошли дальше, приняв регламент General Data Protection Regulation (GDPR) – европейское законодательство о защите личных данных²⁷. Положения данного Регламента носят экстерриториальный характер: Европейский Союз распространяет его действие на любые организации, предоставляющие сервис, товары или услуги потребителям на его территории. При этом не важно, где зарегистрирована компания – также в ЕС или за его пределами, в том числе и на территории Российской Федерации.

Лицензии

Лицензия – это договор, по которому правообладатель произведения (лицензиар) предоставляет другой стороне (лицензиату) право использования произведения в установленных договором пределах. Применительно к данным – это правила, определяющие то, как именно в дальнейшем смогут использоваться опубликованные вами материалы.

Все лицензии можно разделить на коммерческие (несвободные) и некоммерческие (свободные). Первый тип лицензий используется с целью заработать на своем информационном продукте деньги, а второй – с целью дать другим возможность безвозмездно его использовать. Соответственно, публикация открытых данных требует использование лицензий, которые обеспечат правовую основу этой открытости.

Некоммерческие лицензии условно также можно разделить на два вида: открытые и свободные. Несмотря на, казалось бы, схожие названия, эти лицензии действуют по разным принципам. Основное их различие состоит как раз в правах на создание производных произведений (в нашем случае это набор открытых данных, или, по российскому законодательству, база данных).

Свободные лицензии – лицензии с разрешением использовать произведение в любых целях, изучать его, создавать и распространять копии произведения, вносить в произведение изменения, публиковать и распространять такие изменённые производные произведения. Самый широкий вариант свободной лицензии – открытая лицензия. Она дает максималь-

²⁶ [Федеральный закон РФ от 19 декабря 2005 года N 160-ФЗ](#)

²⁷ [Общий регламент защиты персональных данных \(GDPR\) Европейского союза](#)

но широкие права на использование данных и документов, в том числе право на бесплатное повторное использование, переработку и распространение. Лицензии, требующие при повторной публикации ссылки на источник или указания авторства, обычно тоже относят к открытым. Лицензии, которые добавляют к этому запрет на коммерческое использование, к открытым не относятся.

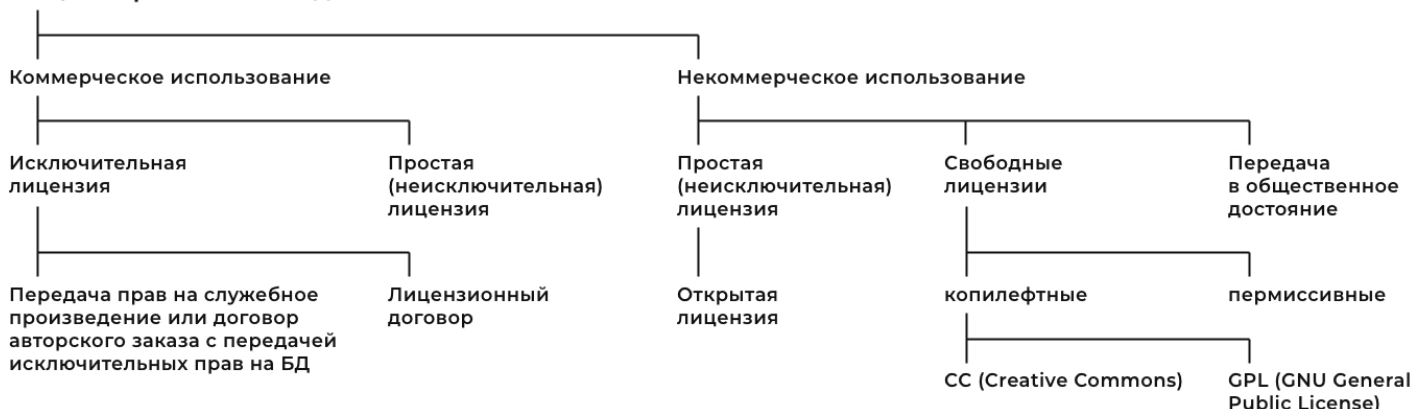
Статья 1286.1 ГК РФ²⁸, регулирующая открытые лицензии, говорит об этом достаточно пространно: «лицензиар может предоставить лицензиату право на использование принадлежащего ему произведения для создания нового результата интеллектуальной деятельности. В данном случае, если иное не предусмотрено открытой лицензией, считается, что лицензиар сделал предложение заключить договор об использовании принадлежащего ему произведения любым лицам, желающим использовать новый результат интеллектуальной деятельности, созданный лицензиатом на основе этого произведения, в пределах и на условиях, которые предусмотрены открытой лицензией. Акцепт такого предложения считается также акцептом предложения лицензиара заключить лицензионный договор в отношении этого произведения».

В предложенной законом процедуре не совсем ясно, как лицензиар, то есть автор или владелец набора данных сможет узнать, какие производные произведения были созданы с использованием его набора данных. То есть для автора будет затруднительно понять с кем у него заключены лицензионные договоры. Более понятным представляется механизм свободных лицензий, в которых механизм создания производных произведений, а также другие способы использования прав на произведение четко структурированы.

Плюсом использования открытой лицензии является более сформированная судебная практика ее защиты. То есть, при обращении в суд защитить свою позицию будет легче.

Применение свободных лицензий предполагает, что одна сторона договора определяет условия в стандартной форме, а другая присоединяется к этому договору не иначе как с принятием правил всего договора в целом.

Лицензирование баз данных



²⁸ [Гражданский кодекс Российской Федерации \(часть четвертая\), Статья 1286.1](#). Открытая лицензия на использование произведения науки, литературы или искусства (введена Федеральным законом от 12.03.2014 N 35-ФЗ)

Существует два основных типа свободных лицензий: вирусные и пермиссивные.

- ~ Пермиссивные, например лицензия BSD (Berkeley Software Distribution license), изначально были написаны для программного обеспечения. Суть этих лицензий заключается в следующем: объект можно использовать как угодно (в том числе, как составную часть коммерческого продукта), но необходимо указывать автора оригинального произведения. Такой тип лицензии не накладывает никаких других ограничений.
- ~ Вирусные лицензии, придуманные Ричардом Столлманом. Альтернативное название — «копилефт-лицензии» (copyleft license). У таких лицензий есть отличительная черта: если берется произведение, которое распространяется по такой лицензии, то все производные от него тоже должны распространяться по той же лицензии. Если, например, опубликованные под такое лицензией данные используются для создания нового набора данных, то в целом должна распространяться на условиях этой свободной лицензии.

В случае, если условия лицензии не соблюдены, правообладатель может предъявить недобросовестному пользователю претензию о незаконном использовании интеллектуальной собственности.

Фактически описанные подходы базируются в первую очередь на международной практике, но российское законодательство в этой части позволяет в прямую использовать эти подходы. Единственное, что приходится учитывать – это то, что правоприменительная практика развита слабо, а у судов и следственных органов отсутствуют необходимые компетенции и опыт, чтобы эффективно работать в данной сфере.

Для непрограммного контента (тексты, фотографии, музыка, видео и т. п.), к которому относятся и данные, чаще всего в мире используются лицензии семейства Creative Commons. Постепенно это становится хорошей практикой и в России. Такая стандартизация позволяет практически «автоматом» решать вопрос с подготовкой лицензии при публикации данных. Однако многочисленность семейства таких лицензий требует понять, какую «широту» использования она может предоставить в отношении тех или иных данных. Например, лицензия CC BY-SA подразумевает, что все производные вашего произведения должны распространяться также свободно, как и исходное. Лицензия CC BY требует только атрибуцию, что означает, что производное вашего произведения может распространяться под любой другой лицензией. Лицензия CC BY-NC предполагает, что произведение используется исключительно в некоммерческих целях, а CC BY-NC-ND – самая несвободная из всех лицензий – что произведение можно только скачивать и распространять, внесение изменений или использование в коммерческих целях запрещено²⁹. При этом стоит отметить, что если набор данных не является общественным достоянием или не связан с лицензией, дающей право на свободное повторное использование, то такой набор данных нельзя считать открытым, даже если он выложен в машиночитаемом виде в сеть. Подобрать подходящую лицензию из группы Creative Commons можно на специальном сайте³⁰. Отметим, что если организация публикует несколько типов контента, не соотносящихся друг с другом, то, скорее всего, на каждый из них необходимо будет определить свой тип лицензии.

Выбранную лицензию необходимо указать в паспорте набора открытых данных, проставив ссылку на текст лицензии. Естественно, что при публикации в России нужно, чтобы текст был представлен на русском языке – в случае использования лицензий группы Creative Commons с этим проблем не возникнет, все тексты уже существуют на русском.

²⁹ creativecommons.org/licenses/?lang=ru

³⁰ creativecommons.org/choose

Практические рекомендации по работе с открытыми данными

Паспортизация наборов данных

При публикации данных важно сопроводить набор данных информацией, которая обоснует легитимность создания набора открытых данных и позволит пользователям определить круг действий, на которые они имеют право при его использовании, а также предоставить всю необходимую техническую информацию чтобы обеспечить пользователя информации о правильной интерпретации сведений в наборе данных.

В русскоязычной терминологической практике такая информация получила название «паспорта набора данных».

Паспорт набора открытых данных относят к метайнформации набора открытых данных. Для описания этой метайнформации были разработаны официальные международные стандарты и стандарты *де-факто*, рекомендуемые к использованию в порталах и каталогах открытых данных. К таким стандартам относятся: DCAT v2³¹, Dublin Core³², [Schema.org](https://schema.org) Dataset type³³. Существует также множество более специализированных стандартов описания метаданных, применяемых, например, применительно к географическим сведениям, статистическим показателям и иным данным с дополнительными характеристиками.

Обычно паспорт набора содержит следующие сведения: название набора, использованные алгоритмы преобразования, формат данных, тип хранилища, дата, авторы или владельцы, структура набора данных, ключевые слова, ссылка на набор, дата актуальности, контакты, обзор данных на выходе, источники, правила использования и распространения.

Эти сведения можно назвать минимальными и достаточными для большинства наборов данных, в том числе публикуемых государственными и муниципальными органами.

На основе этих стандартов при поддержке Открытого Правительства в Российской Федерации В 2013 были разработаны методические рекомендации³⁴ по публикации наборов открытых данных государственными органами и органами местного самоуправления. Несмотря на то, что документ это не обновлялся уже более семи лет, изложенные в нем рекомендации, в том числе, по паспортизации наборов данных, остаются актуальными. Предложенный в нем подход к паспортизации может быть использован не только органами государственной и муниципальной власти, но и частными компаниями, некоммерческими организациями, исследователями.

31 Data Catalog Vocabulary (DCAT) - Version 2, www.w3.org/TR/vocab-dcat, 2020

32 Dublin Core Metadata Initiative, www.dublincore.org, 2020

33 Dataset type, schema.org/Dataset, 2015

34 [Методические рекомендации по публикации открытых данных версия 3.0](#)

Рекомендации

1. Паспорт должен содержать четкую структуру:
 - ~ Наименование набора открытых данных
 - ~ Описание набора открытых данных
 - ~ Владелец набора открытых данных
 - ~ Ответственное лицо
 - ~ Телефон ответственного лица
 - ~ Адрес электронной почты ответственного лица
 - ~ Гиперссылка (URL) на открытые данные
 - ~ Формат набора открытых данных
 - ~ Описание структуры набора открытых данных
 - ~ Дата первой публикации набора открытых данных
 - ~ Дата последнего внесения изменений
 - ~ Содержание последнего изменения
 - ~ Дата актуальности набора данных
 - ~ Ключевые слова, соответствующие содержанию набора данных
 - ~ Гиперссылки (URL) на версии открытых данных
 - ~ Гиперссылки (URL) на версии структуры набора данных
2. Рекомендуется размещать паспорт набора открытых данных как в человекочитаемом, так и в машиночитаемом видах одновременно.
3. Наиболее простой способ паспортизации наборов данных – это использование бесплатного ПО с открытым кодом для каталогизации данных. Наиболее популярные инструменты: SKAN³⁵, DKAN³⁶, JKAN³⁷, Dataverse³⁸, все они включают необходимый инструментарий для этого.

Выбор лицензии

Для выбора оптимальной лицензии при публикации данных, можно использовать следующую несложную инструкцию:

Перед публикацией любых материалов в открытом доступе необходимо предварительно проверить, не существует ли на публикацию имеющейся информации законодательных ограничений, иначе может возникнуть ответственность за распространение информации, доступ к которой, согласно законодательству, ограничен (например, персональные данные).

35 SKAN, ckan.org

36 DKAN, dkan.org

37 JKAN, github.com/timwis/jkan

38 The Dataverse Project, dataverse.org/

Рекомендации

1. Если есть уверенность, что собранные вами данные и созданные на их основе материалы являются общедоступными, а также не имеют никаких ограничений на распространению, то можно просто перейти к выбору лицензии.
2. Если есть сомнения, надо рассмотреть публикуемые данные с точки зрения трех ключевых критериев публикации данных и проверить, не попадает ли публикуемая вами информация под какой-либо из этих критериев – так как ответственность за распространение информации, доступ к которой, согласно законодательству, ограничен, будет лежать на распространителе.:
 - a. категория доступа к информации;
 - b. порядок распространения информации;
 - c. гражданско-правовой режим.

Категория доступа подразделяет информацию на различные категории в зависимости от типа доступа к ней и закладывает основы обращения различных лиц с такой информацией. Ограничение доступа к информации производится на основании федеральных законов (см. раздел Государственные открытые данные), при этом соблюдение конфиденциальности такой информации является обязательным.

Порядок распространения определяет какая информация может распространяться свободно, какая имеет особые режимы и правила распространения, а какая полностью запрещена к распространению на территории Российской Федерации.

1. Информация, распространяемая свободно – общеизвестные сведения и иная информация, доступ к которой не ограничен.
2. Информация, распространение которой ограничено:
 - a. продукция эротического характера;
 - b. реклама;
 - c. предвыборная агитация;
 - d. нецензурная брань;
 - e. сведения и материалы, порочащие репутацию граждан и юридических лиц;
 - f. изображения граждан (в том числе несовершеннолетних);
 - g. охраняемые результаты интеллектуальной деятельности.
3. Информация, распространение которой запрещено.
 - a. информация, признанная имеющей экстремистский характер;
 - b. пропаганда наркотиков;
 - c. пропаганда суицида;
 - d. продукция порнографического характера, в том числе детская порнография.

4. Информация, распространение которой специально маркируется. Например, к данной категории относится информация, созданная некоммерческими организациями, внесенными в реестр Министерства юстиции - «реестр ОРГАНИЗАЦИЯ, выполняющих функции иностранного агента».

Гражданско-правовой режим

Критерий гражданско-правового режима относится к подразделению информации на категории относительно совокупности правил, регулирующих определенную сферу деятельности и отношения между людьми по поводу различных объектов гражданского права.

1. Неохраняемые объекты

- a. Неохраноспособные объекты. К ним относятся товарные знаки, которые вошли во всеобщее употребление для обозначения товаров определенного вида. Также объектами, не попадающими под его охрану являются результаты интеллектуальной деятельности, противоречащие принципам гуманности и морали, общественным интересам. Не получают охрану в качестве изобретений идеи, научные теории и математические выводы, открытия.
- b. Объекты, охрана которых прекратилась в связи с истечением установленных сроков или по воле правообладателя.

2. Охраняемые объекты

- a. Объекты авторского права (произведения).
- b. Объекты смежных прав (исполнения, фонограммы, сообщение в эфир или по кабелю теле- и радиопередач; содержание баз данных; произведения в рамках публикаторского права).
- c. Ноу-хау (секреты производства, имеющие действительную или потенциальную коммерческую ценность вследствие неизвестности их третьим лицам).

Помните, что при публикации материалов должно соблюдаться российское законодательство относительно охраняемых объектов.

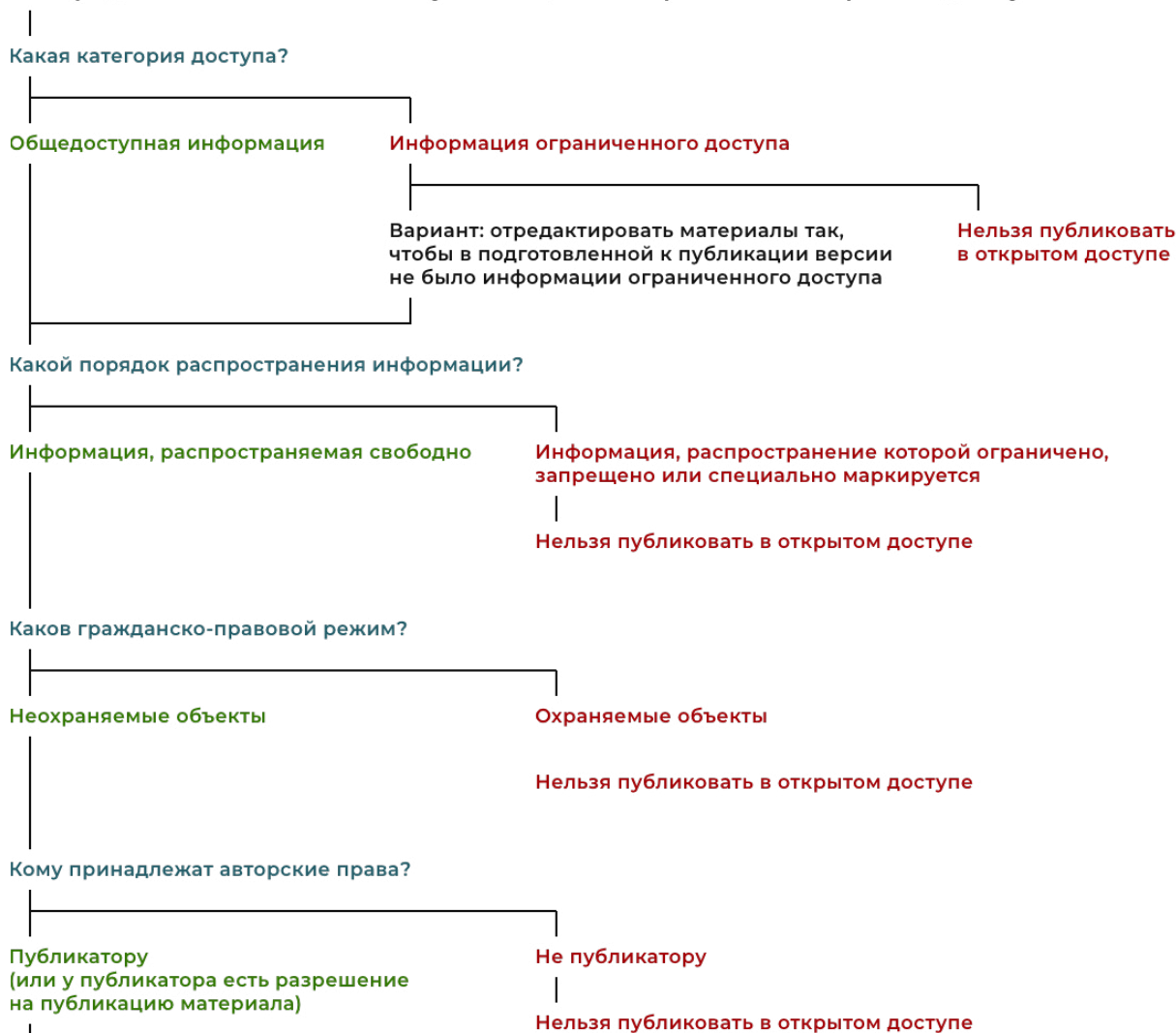
Первоначальным обладателем авторских прав является автор, т. е. физическое лицо, творческим трудом которого создано произведение. В соответствии с установленными правилами, права на объекты, созданные сотрудниками в рамках служебных обязанностей, принадлежат работодателю.

Стоит отметить, что есть разные варианты работы с авторами. Как правило, принадлежность того или иного произведения организации фиксируется в договоре. Например, если речь идет об исключительных правах на произведение сотрудников организации, то данное положение должно быть закреплено в трудовом договоре/приказе. Если организация работает с подрядчиком — в договоре подряда, а если с волонтером — в волонтерском договоре. Для того чтобы избежать спорных случаев, разумнее всего оговорить и зафиксировать условия принадлежности прав на произведение с исполнителем работ.

Кроме этого, если производство контента (проведение исследований, создание методических пособий и т. п.) оплатил заказчик или на эту работу был получен грант, то необходимо

обращать внимание на прописанные в договоре на финансирование условия в части авторских прав на интеллектуальную собственность. Например, в случае грантов подобные произведения могут принадлежать непосредственно фонду, а могут принадлежать грантополучателю с определёнными ограничениями на публикацию и распространение или, напротив, обязательную публикацию таких произведений под открытыми лицензиями.

1. Определить возможность публикации материалов в открытом доступе



2. Выбрать лицензию



3. Указать лицензию при публикации

Дать ссылку на страницу с указанием условий лицензии

Анонимизация данных

Анонимизацию данных можно считать одним из способов защиты приватности и иной чувствительной информации. В результате этого процесса данные становятся невозможно соотносить с каким-либо конкретным лицом или организацией. При этом российское законодательство не содержит понятия анонимизации персональных данных, однако содержит более общее понятие – обезличивание данных.

Под обезличиванием персональных данных понимается один из способов обработки персональных данных, в результате которого становится невозможно без использования дополнительной информации определить принадлежность персональных данных конкретному субъекту. Пример использования обезличивания при публикации открытых данных – публикация судебных решений,

Требования и методы обезличивания персональных данных утверждены приказом Роскомнадзора от 5 сентября 2013 года № 996³⁹. Так как это единственный источник, фактически Роскомнадзор применяет перечисленные в нем требования к частным компаниям, хотя де-юре приказ написан для государственных и муниципальных органов.

В приказе перечислены следующие методы обезличивания:

- ~ введение идентификаторов (замена части сведений, составляющих персональные данные, идентификаторами с созданием таблицы соответствия таких идентификаторов исходным данным);
- ~ изменение состава или семантики (обобщение, изменение значений атрибутов персональных данных или удаление части сведений, позволяющих идентифицировать субъекта);
- ~ декомпозиция (разделение массива персональных данных на несколько составляющих частей с последующим их раздельным хранением);
- ~ перемешивание (перестановка отдельных значений атрибутов персональных данных в массиве).

При выборе метода обезличивания мы рекомендуем ознакомиться с положениями этого приказа, поскольку методы применяются в зависимости от целей и задач обработки персональных данных.

В этом же приказе приводится понятие деобезличивания – действий, в результате которых обезличенные данные принимают вид, позволяющий определить их принадлежность конкретному субъекту персональных данных.

На первый взгляд, это говорит о том, что обезличенные данные не являются персональными данными, этот вывод подтверждается в российской судебной практике⁴⁰. Однако существует и противоположная позиция, заключающаяся в том, что информация, не позволяющая однозначно идентифицировать лицо, признавалась персональными данными. Она

³⁹ [Приказ Федеральной службы по надзору в сфере связи, информационных технологий и массовых коммуникаций \(Роскомнадзор\) от 5 сентября 2013 г. N 996](#) г. Москва “Об утверждении требований и методов по обезличиванию персональных данных”, 2013

⁴⁰ Решение Верховного Суда РФ от 26 января 2011 г. № ГКПИ10-1510; решение Арбитражного суда Удмуртской области по делу № А71-6910/2013 от 25 сентября 2013 г.

исходит из анализа норма закона, из которых исходит, что обработка обезличенных данных является обработкой персональных данных. Эта позиция подтверждается самим определением обезличивания, которое подразумевает, что при обезличивании информация приводится к тому виду, при котором идентифицировать субъекта персональных данных все же возможно посредством привлечения дополнительной информации.

Таким образом, при подготовке к публикации наборов открытых данных, содержащих персональные данные, нужно либо заручиться согласием на обработку персональных данных на все действия, которые планируете с ними совершать, у лиц, информация о которых содержится в этих данных, либо прибегнуть к такому методу обезличивания, при котором косвенная идентификация лица с использованием дополнительной информации невозможна. При этом обращаем внимание, что даже в этом случае закон допускает обработку обезличенных данных без согласия лица только в статистических или иных исследовательских целях.

В этой связи в европейском законодательстве утвердился подход, в котором разделяют псевдоанонимизацию и анонимизированность. Первая предполагает, что в результате использования дополнительной информации все же возможно определить принадлежность данных определенному лицу. Вторая подразумевает такую степень обезличенности, при которой данные не относятся к определенному или определяемому лицу, и, соответственно, не являются персональными.

В последнее время в России можно наблюдать тенденцию усиления регулирования анонимизированных данных. Так, например, идет работа над законопроектом Министерства цифрового развития РФ, в котором даются определения обезличенных данных и обезличенных персональных данных. Подход к этим понятиям повторяет европейское законодательство. В законопроекте предусмотрены и другие изменения в части целей, способов обработки данных, обезличивания, а также способов выражения согласия субъекта данных на обработку.

Рекомендации

1. Главная рекомендация в работе с данными которые могут содержать персональные данные – это, если эти данные не являются значимыми для повторного использования и анализа, то удалять их из публикуемых наборов данных.
2. Если персональные данные имеют значение, и Вы уверены что хотите опубликовать этот набор данных, то воспользуйтесь методическими рекомендациями⁴¹ Роскомнадзора утвержденными в дополнение к их приказу N966

⁴¹ [Методические рекомендации по применению приказа Роскомнадзора от 5 сентября 2013 г. N 996 «Об утверждении требований и методов по обезличиванию персональных данных»](#) (утв. Роскомнадзором 13.12.2013), 2013

Технические аспекты публикации

Ещё до того, как открытые данные стали темой общественного внимания, разработки стандартов и международной деятельности, они активно публиковались научными организациями, как для обеспечения открытости, так и для обмена данными с другими исследовательскими центрами.

Основными способами публикации данных в то время было размещение их в каталогах FTP и HTTP серверов по некоторым заданным правилам, зная которые разработчики программ-потребителей данных выгружали их на регулярной основе.

С развитием технологий и появлением специальных продуктов, помогающих в подготовке и публикации данных, появилось гораздо больше возможностей делать это быстро и охватывать самые разные аудитории, от программистов до исследователей.

Технические аспекты публикации включают:

- ~ Метаданные – данные о данных, структура и описание наборов данных
- ~ Форматы файлов/публикации – форматы файлов
- ~ Инструменты – с помощью каких инструментов готовят данные
- ~ Типичные ошибки при публикации данных – то, чего лучше избегать

Метаданные

Публикация как наборов данных, так и документов должна сопровождаться внесением метаданных – информации о публикуемых материалах, позволяющей пользователям получить точное представление об их содержании, условиях создания и правилах распространения. Метаданные должны публиковаться одновременно с набором данных и составляют его паспорт.

Метаданные публикуются в соответствии с набором стандартов, подробнее о них рассказывается в разделе [Паспортизация наборов данных](#).

Рекомендации

1. Одна из наиболее частых ошибок – забывать публиковать описания наборов данных и иных метаданных. Помните, что, если Вы не указали, например, какие годы охватывает набор данных или когда он создан, это может создать проблемы в его использовании другим пользователям.
2. Обязательно убедитесь, что инструменты, которые Вы используете для публикации наборов данных соответствуют Вашим потребностям в их описании. Если, например, Вы работаете с данными показателей то выбор и стандарта публикации метаданных и самих данных должен быть, например, стандарт SDMX⁴² используемый в статистической работе и также по аналогии с другими областями применения данных.

42 SDMX Community, sdmx.org, 2020

Форматы файлов/публикаций

В отличие от документов, которые используют люди, наборы данных создаются для их использования программными продуктами и поэтому данные должны быть представлены в так называемых машиночитаемых форматах файлов данных.

Важными характеристиками форматов файлов данных являются:

- ~ простота машинной обработки
- ~ распространённость/стандартизованность
- ~ возможность проверки структуры (валидация и наличие схемы структуры данных)
- ~ пригодность для публикации данных большого объёма

Таблица 1. Форматы файлов

Форматы данных	Описание	Критерии			
		Машиночитаемость	Стандартизованность	Схема	Публикации больших данных
XLS, XLSX	Форматы файлов, используемые табличными редакторами. Ограниченно машиночитаемы и ограничены по объёму	Ограниченно	Ограниченно	Нет	Нет
CSV	Наиболее распространённый формат публикации табличных данных. Позволяет публиковать условно неограниченные объёмы данных, но слабо поддерживает их валидацию.	Да	Да	Частично, Frictionless Data ⁴³	Да
JSON	Распространённый формат публикации данных популярных в веб-приложениях.	Да	Да	Да, JSON Schema ⁴⁴	Ограниченно
XML	Один из наиболее стандартизированных форматов файлов для обмена и публикации данных. Часто используется в государственных и корпоративных системах.	Да	Да	Да, XSD ⁴⁵	Ограниченно
SQL, BSON, JSON lines	Форматы файлов, используемых для создания дампов баз данных. Хорошо машиночитаемы и используются для публикации больших наборов данных	Да	Да	Да	Да
ZIP, GZ, 7z, RAR	Архивные форматы файлов, используемые как контейнеры для хранения других данных большого объёма.	Да	Частично	Зависит от формата	Да

43 Frictionless Data, frictionlessdata.io, 2020

44 JSON Schema, json-schema.org, 2020

45 XML schema, en.wikipedia.org/wiki/XML_schema, 2020

Рекомендации

1. Выберите формат для публикации для публикации данных в зависимости от того какого объёма и типа данные Вы публикуете.
 2. Данные большого объёма, от 100 мегабайт предпочтительнее публиковать после сжатия в архив, например, ZIP архив.
 3. Для данных, имеющих иерархию и вложенные объекты лучше использовать форматы JSON или XML.
 4. Если Вы можете опубликовать базу целиком, то лучше сделать её дамп в SQL или JSON lines (JSONL) формат.
 5. Для CSV файлов предпочтительно использовать стандарт Frictionless Data⁴⁶, например с помощью такого инструмента как Data Curator⁴⁷
-

Типичные ошибки при публикации материалов

Ошибка 1. Публикация в неадекватном формате. Например, нет смысла публиковать большую таблицу с данными в виде многостраничного файла PDF. Такой публикацией просто нельзя воспользоваться, потому что этот формат крайне затрудняет дальнейшую компьютерную обработку, в нем невозможны подсчеты и сортировка, а путем чтения из него нельзя вынести обобщенной информации.

Ошибка 2. Публикация только результатов анализа без исходных данных. Типичный пример – отчеты об исследованиях на основе данных. Разумно такие документы публиковать именно в виде документов (например, в формате PDF). Но при этом очень желательно сопровождать публикацию такого материала дополнительными файлами, содержащими сами данные, собранные или использовавшиеся в ходе исследования. Разумеется, такая сопутствующая публикация не всегда возможна (есть ограничения, связанные с персональными данными, а также с закрытыми данными, не предназначенными к публикации). Но во всех случаях, когда это возможно, «сырые» данные публиковать нужно, потому что они зачастую представляют собой и самостоятельный интерес, не говоря уже о том, что их публикация позволяет проверять и верифицировать обобщения, сделанные в ходе исследования. То же самое касается публикации инфографики и любых других материалов, в основе которых лежат структурированные данные.

Ошибка 3. Публикация с неинформативным описанием. Чтобы пользователь определил, насколько ему интересен опубликованный материал, у материала должно быть краткое и емкое описание. Кроме того, имеют значение даты его создания и последнего обновления, а также его источник (-и).

46 [Frictionless Data](#)

47 [Data Curator tool](#)

Ошибка 4. Отсутствие справочников для данных, содержащих шифры или условные обозначения. Чтобы данными можно было пользоваться, необходимы расшифровки и пояснения используемых сокращений, обозначений, единиц измерения и т.п. Это особенно актуально в случаях, когда публикуются таблицы или иные структурированные данные, в которых часто бывают сокращенными названия полей (столбцов) или приведены только величины без их размерности. Например, если столбец называется «сумма», то требуется расшифровка, в которой бы указывалось, в чем эта сумма измеряется. Справочники лучше всего делать в виде отдельных файлов — документов или таблиц.

Использование открытых данных

Агрегация данных

Важным ограничением в использовании открытых данных является то, что публикуемые данные изначально рассеяны по множеству источников, как правило, не связанных между собой. Интеграция данных из этих источников создают как дополнительную ценность для потребителей, так и дополнительные риски раскрытия коммерческой тайны, персональных данных и иных защищаемых законом сведений. То, что недоступно в разрозненной форме, приобретает иное качество после интеграции.

В мировой практике, в разведывательном сообществе, это называется работой с открытыми источниками (open source information) где на основе косвенных или интегрированных данных можно делать выводы, например, о направлениях инвестиционной деятельности крупной корпорации, её реальных, а не отчётных показателях, или о крупных государственных проектах и расходах непубличного характера.

Также сведение данных из открытых источников, как открытых данных, так и просто общедоступных данных позволяет провести деанонимизацию данных и провести повторную идентификацию физического лица (re-identification) и, как следствие, делает возможным злоупотребления собранными сведениями.

По этой причине проекты по агрегации и интеграции данных всегда находятся в балансе между соблюдением общественного интереса и не-раскрытия данных, ограниченных к публикации. В 2007 году в США ряд сведений о контрактах, публикуемых на общественном портале Fedspending.org⁴⁸ были удалены по причине раскрытия персональных данных пользователя⁴⁹. Аналогичные примеры существуют во многих странах, когда создатели цифровых сервисов вынуждены удалять или корректировать ранее опубликованные открытые данные.

Подобные риски необходимо учитывать при создании общественных и коммерческих цифровых продуктов, закладывая в их реализацию возможность исправления или удаления данных в тех случаях когда их публикация нарушает права гражданина, раскрывает коммерческую тайну или иным образом нарушает права.

48 [Fedspending.org](https://www.fedspending.org/) by OMB Watch

49 [Is it true an individual's sensitive personal information was discovered on FedSpending.org?](https://www.fedspending.org/)

Рекомендации

1. При использовании и обработке данных убедитесь, что в данных нет персональных данных и иной личной идентифицирующей информации. Если она всё же есть, постарайтесь связаться с владельцем набора данных и попросить удалить эти сведения.
 2. При сведении данных вместе, проконсультируйтесь с юристами не нарушает ли это каких-либо местных нормативных документов.
 3. Даже если первичные данные являются открытыми государственными данными в которых могут упоминаться конкретные физические лица, иногда может быть важным сокрытие полных ФИО физических лиц и раскрытие информации только в агрегированном виде или же заменив ФИО словами «Физическое лицо». Такой практики придерживаются, например, в Европейском союзе при публикации получателей грантов в системе FTS (Financial Transparency System)⁵⁰.
-

Аналитика на основе данных

При использовании опубликованного набора открытых данных для аналитики основная юридическая трудность – определить тот спектр действий, которые публикатор набора открытых данных разрешил производить с этим набором. В этой задаче обычно помогает лицензия, которая в упрощенном виде также может содержаться в паспорте набора открытых данных (см. раздел Лицензии).

Базовый критерий при определении прав на действия с набором данным – являются ли они безвозмездными, либо владелец набора данных установил вознаграждение за его использование. В зависимости от этого лицензия будет коммерческой или некоммерческой. Так как мы говорим об открытых данных, наиболее вероятно, что пользователь столкнется с некоммерческой лицензией.

Рекомендации

1. Обратите внимание на лицензию(-и) под которыми публикуются данные. Ваш аналитический продукт должен соответствовать их условиям.
 2. В сложных случаях, когда Вы интегрируете много разных данных, опубликованных под разными лицензиями, обратитесь к юристам чтобы избежать ненужных рисков.
-

Повышение качества данных

Помимо собственно присутствия данных в открытом доступе и снабжения их некоторой правовой основой, на возможностях использования также сказывается качество данных.

50 [Financial Transparency System](#)

Под качеством данных понимается совокупность характеристик, определяющих пригодность данных к использованию. В их число входят такие моменты, как:

- ~ Своевременное обновление (актуализация) данных;
- ~ Наличие метаданных, позволяющих судить о правах использования данных, их составителях, публикаторах и тематике;
- ~ Наличие описания структуры данных (их схемы, а также значения полей);
- ~ Валидность структур, содержащих данные (при наличии ошибок автоматическая обработка может оказаться невозможной);
- ~ Соответствие данных описанию (в том числе в части форматов значений; наличия/отсутствия значений null);
- ~ Отсутствие в данных фактических ошибок (неверных значений).

Данные, качество которых ниже допустимого уровня, невозможно использовать, даже если они формально отвечают критериям открытости. То же самое можно сказать и о документах (помимо упомянутых характеристик, здесь можно также упомянуть качество графических изображений). Соответственно, повышение качества материалов — это одно из важных условий их централизованной публикации. Платформа, которая берет на себя такую функцию, должна располагать возможностью, с одной стороны, поощрять публикацию материалов высокого качества, а с другой стороны, сообщать пользователю, какого качества ему стоит ожидать от той или иной публикации.

Рекомендации

Основные способы повышения качества данных:

Пользовательские комментарии. Это самый простой способ передачи информации о качестве материалов. В комментариях (отзывах) пользователи могут высказывать свои претензии к качеству, делиться мнениями о том, что можно было бы улучшить и обсуждать эти мнения. Также пользователи могут оставлять одобрительные отзывы. Тем самым, с одной стороны, публикатор получает от пользователей обратную связь и, следовательно, возможность учитывать пожелания и совершенствовать свои материалы; а с другой стороны, пользователи могут ориентироваться на комментарии, чтобы составить себе представление о качестве публикации. Достоинство такого подхода в простоте реализации. Недостаток в том, что разрозненные комментарии плохо поддаются обобщению; их написание и чтение занимают время; наличие возможности комментировать требует модерации.

Рейтингование. Более совершенным способом оценивания публикаций представляется рейтинг. Рейтинг может строиться на основе пользовательских оценок, выносимых конкретным публикациям. Эти оценки могут использоваться как для присвоения оценки публикации (например, как среднее значение всех пользовательских оценок), так и для присвоения оценки организации-публикатору (уже не напрямую, а как среднее значение оценок, полученных ее материалами).

Рейтинг, с одной стороны, может служить стимулом для публикатора, чтобы повышать качество, а с другой стороны, индикатором качества для пользователей. Процедура вынесения пользовательской оценки может быть разной. Наиболее простым способом было бы предложить пользователям выносить общую оценку набору или документу, например, по пятибалльной шкале. Но при таком подходе значение рейтинга может оказаться больше статусным, чем информативным для пользователей. Другой вариант — это предлагать пользователю, желающему вынести оценку, небольшую форму с детализацией параметров оценки (таких, как качество описания данных, заполненность метаданных, актуальность набора данных, качество данных в наборе, корректность документа и т.п.). Общая оценка набора или документа в таком случае будет выводиться из оценок по каждому из предложенных параметров. Это усложнит процедуру вынесения оценки и, вероятно, демотивирует некоторых пользователей, но может способствовать более ответственному вынесению оценок и позволяет другим пользователям судить о том, насколько для них релевантны отмеченные недостатки.

Стандартизация. В отличие от двух предыдущих подходов, наличие стандарта дает возможность выносить не произвольную оценку, а оценку относительно соответствия этому стандарту. В то же время стандарт представляет собой готовый эталон, на который может ориентироваться публикатор при подготовке своих материалов. Разработка стандарта требует специальной целенаправленной работы, которая может оказаться долгой и ресурсозатратной. Однако стандартизация может оказаться наиболее эффективным инструментом повышения качества публикаций.

Примеры проектов по публикации данных

В качестве образцов использования лучших практик приведем несколько примеров порталов, на которых осуществляется публикация открытых данных.

Портал QRI

Qri (qri.cloud) – облачный портал открытых данных созданный одноименным стартапом и включающий инструменты версионирования, автоматизации и распространения наборов данных.

В частности, QRI обеспечивает доступ к данным и позволяет осуществлять их подготовку с помощью инструментов командной строки и Qri desktop (qri.io/download), настольного приложения для паспортизации данных и загрузки их в облако, а также получения их из облака.

Qri предоставляет API и широкий набор инструментов для работы с API через языки программирования и инструментов командной строки.

Портал Data.world

Data.world (data.world) – специальный портал для публикации данных, с акцентом на возможности загрузки данных для последующей визуализации и интеграции как с источниками данных, так и с инструментами аналитики и обработки данных.

С одной стороны, на портале предусмотрено заполнение лишь минимальных метаданных таких как: наименование набора данных, ключевые слова, лицензия и описание. С другой стороны, портал предоставляет гибкие инструменты ведения проектов, основанных на данных, в особенности – данных в табличных форматах CSV которые можно не только редактировать, но и осуществлять обработку и фильтрацию с помощью языка запросов схожего с языком программирования баз данных SQL.

Data.world включает набор библиотек и инструментов для работы разработчиков через их API.

Портал открытых данных Министерства культуры РФ

Министерство культуры РФ разработало специальный портал публикации данных (opendata.mkrf.ru), в рамках которого изначально реализован ряд важнейших функций, необходимых потребителям данных в сфере культуры.

Портал поддерживает такие функции как:

- ~ Связанные данные (Linked Data);
- ~ Подписку на обновление наборов данных;
- ~ Морфологический поиск;
- ~ Программный интерфейс доступа (REST API);
- ~ Удобное представление данных в табличной и картографической формах.

На 2020 год этот портал остаётся наиболее продуманным порталом открытых данных, эксплуатируемым органами власти в нашей стране. Он обеспечивает доступность 74 наборов данных, включая государственные каталог музейного фонда, и имеет значительное число частных и корпоративных пользователей, таких как Яндекс, 2do2go и других.

Портал данных НКО

В 2018 году АНО «Информационная культура» на средства президентского гранта был создан портал «Данные НКО» (ngodata.ru). Целью которого было создать площадку для раскрытия данных некоммерческими организациями для публикации данных, созданных в рамках своей деятельности.

Проект был создан на базе программного обеспечения для создания каталогов данных SKAN, которое было переведено на русский язык и доработано для возможности учёта регистрации данных от российских юридических лиц – включая коды ИНН/ОГРН организации.

Проект можно назвать частично успешным. Его запуск показал общую неготовность НКО публиковать данные о своей деятельности по причине общего непонимания принципов работы с данными и на сегодняшний день проект продолжается в форме обучения некоммерческого сектора работе с данными.

При создании проекта были разработаны методические рекомендации⁵¹ для НКО по публикации данных, которые включают подробные практические инструкции, использовать которые могут не только некоммерческие организации.

51 ngodata.ru/metod_web.pdf

Рекомендации по развитию регулированию и внедрению открытых данных на государственном уровне

Рекомендации по совершенствованию регулирования открытых данных основываются на лучших мировых практиках и учитывают особенности сбора данных органами власти в России, а также национального регулирования информационных технологий и государственных информационных систем.

1. В рамках общей государственной политики открытости данных:

1. Рекомендуется подписать Хартию Открытых данных (<https://opendatacharter.net>), ранее подписанную 74 национальными и местными правительствами по всему миру
Результаты: взаимодействие с международным сообществом и представителями стран, внедряющих открытость, данных даст возможность трансляции лучших международных практик в России и лучших российских практик в мировое сообщество.

2. Рекомендуется определить исполнительный или консультативный орган при Правительстве РФ (Правительственная комиссия или совет) ответственный за проведение политики открытости данных в РФ.

Результаты: наличие постоянно действующего исполнительного или консультативного органа создаст площадку коммуникации общественных организаций и бизнеса с представителями органов власти, а также обеспечит возможность создания национального плана по открытости, соответствующего современным реалиям.

3. Рекомендуется вернуться к разработке национального плана по открытости данных и планов по открытости властей в субъектах федерации.

Результаты: национальный план открытости позволит утвердить дорожную карту раскрытия ключевых данных, хранимых в органах власти, и внедрит мониторинг исполнения работ по этому плану в практику государственного управления.

2. В рамках государственной политики регулирования информационных технологий:

1. Рекомендуется разработать и принять на уровне федерального закона обязательные требования к раскрытию данных, затрагивающие все вновь разрабатываемые и уже созданные государственные информационные системы за исключением информационных систем, содержащих сведения ограниченного доступа и государственную тайну.

Результаты: основные данные органов власти в России содержаться именно в государственных информационных системах при том, что их регулирование, особенно в части открытости данных, является недостаточным. Введение обязательных требований к публикации данных позволит обеспечить долгосрочную открытость данных в государственных информационных системах.

2. Рекомендуется определить одним из направлений поддержки ИТ-компаний создание инструментов по работе с открытыми данными, доступные широкому числу пользователей, государственных и коммерческих ИТ-специалистов.

Результаты: данная мера стимулирует значительный рост числа компаний, использующих открытые данные и создающих продукты на их основе.

3. В рамках государственной политики поддержки научных исследований:
1. Рекомендуется сформировать инициативу открытости научных исследований (open access) на уровне Министерства высшего образования и науки Российской Федерации и включить создание инициатив по раскрытию данных, созданных по результатам исследований (по аналогии с проектом Zenodo в Европейском союзе).
Результаты: данная мера позволит обеспечить тесную кооперацию российских исследователей с исследователями из других стран, обеспечит значительную открытость российских исследований и повторное использование результатов исследований другими учеными и научными коллективами.
 2. Рекомендуется включить обязательные требования по публикации открытых данных по результатам научных исследований в конкурсы, проводимые государственными научными фондами, субсидиями и контрактами на исследования, а также в государственное задание университетов и научных организаций.
Результаты: данная мера внедрит стандарты Открытого доступа (Open Access) для исследований, проводимых за общественные средства (государственных бюджетов) и обеспечит прозрачность исследований в целом.
 3. Рекомендуется создание инициативы доступа исследователей к данным органов власти ограниченного доступа на специальных условиях, требующих аккредитации, подписания соглашений о неразглашении и иных требований (по аналогии с UK Data Service в Великобритании и Australia Research Data Commons в Австралии).
Результаты: данная мера позволит обеспечить баланс защиты данных рядовых граждан и потенциального значительного социального и экономического эффекта от инноваций и исследований в областях общественного здоровья, социально-демографической политики, государственного управления и т.д.
4. В рамках государственной политики поддержки общественных и некоммерческих инициатив:
1. Рекомендуется включить в направления поддержки Фонда президентских грантов поддержку проектов по созданию открытых данных.
Результаты: значительный рост числа общественных проектов основанных на открытых данных и создающих открытые данные.
 2. Рекомендуется разработать типовые требования по обязательному раскрытию данных некоммерческими организациями-получателями субсидий, грантов и иной финансовой формы поддержки от государства.
Результаты: повышение прозрачности некоммерческих организаций и результатов их деятельности, включая деятельность в целях общественного блага осуществляемую за общественные средства (государственные бюджеты).

