Иван Бегтин, директор НП «Информационная культура»:

Российские реалии, в которых у нас нет ни возможности системного открытия данных, ни качественного контроля за приватностью раскрываемых баз данных, безусловно, должны меняться.

Приватность в открытых данных

В последние годы открытые данные перестали быть чужеродной темой для российского государства и тесно вплелись в информационную жизнь не только тех, кто занимался ими очень давно, но и сотрудников органов власти, ведущих официальные сайты, специалистов по ИТ в госорганах и многочисленных потребителей данных.

Федеральный закон Российской Федерации от 7 июня 2013 г. N 112-ФЗ «О внесении изменений в Федеральный закон «Об информации, информационных технологиях и о защите информации» и Федеральный закон «Об обеспечении доступа к информации о деятельности государственных органов и органов местного самоуправления» закрепил открытые данные в нормативно-правовой базе и сформировал направление, в котором теперь органы власти действуют во всех вопросах работы с общедоступной информацией.

При этом открытых данных становится всё больше. Тысячи наборов данных публикуются на федеральных и региональных специализированных порталах, на сайтах государственных информационных систем, на сайтах органов власти. Эти данные доступны в разном качестве, в разном объёме и далеко не всегда удобны и пригодны для работы потребителей, но практически всегда их раскрытие — это решение ответственного за них органа власти.

Доступность данных — это хорошо или плохо? Для предпринимателей, для бизнеса, работающего с данными — безусловно, хорошо. Это возможность создавать новые коммерческие продукты на бесплатном «информационном топливе» открытых данных и, тем самым создавать рабочие места, выплачивать налоги и зарабатывать.

Однако данные публикуемые органами власти могут нести как пользу для общества, так и создавать новые риски отдельным его членам. Один из таких рисков — это риск «повторной идентификации» (reidentification). Повторная идентификация — это возможность реконструировать персональные данные о человеке, используя несколько наборов данных вместе. Этот риск возникает в том случае, когда набор данных покидает контролируемую среду внутри органа власти или иного его владельца и оказывается в открытом доступе.

Общества защиты персональных данных граждан, такие как Open Rights Group [1] в Великобритании, вырабатывают собственные правила и рекомендации, как именно открытые данные должны публиковаться.

В частности, в его кампании Open Data Privacy[2] были определены следующие правила:
- открытые данные должны поощряться, но приватность первостепенна;
- государство должно осознавать риски повторной идентификации с помощью открытых данных;
- должны проходить правильные общественные обсуждения до запуска любого плана по коммерциализации анонимизированных данных;
- технологии анонимизации должны быть предметом экспертной оценки;
- должен существовать ответственный механизм раскрытия информации для безопасных уведомлений в случаях нарушения приватности.

Сейчас, когда подобных анонимизированных данных в России всё ещё публикуется немного, эти рекомендации и правила являются пожеланиями, которые необходимо учесть в ближайшем будущем. При решениях об открытии данных, столь ожидаемых и востребованных бизнесом: — о качестве образования по школам, о криминальной статистике, анонимизированных данных переписи населения — нам необходимо помнить и о последствиях открытия данных и рассматривать все возможные пути их будущего использования.

Впрочем, многие вопросы актуальны прямо сейчас.

Например, несколько лет назад, мы создали проект *Госзатраты*, в котором автоматически собираем все данные о государственных закупках и контрактах. В России данные о госконтрактах не считаются ни служебной, ни коммерческой тайной. Они регулируются отдельным федеральным законом N 44-ФЗ и множеством сопутствующих ему нормативных документов. Однако за всё время существования нашего проекта каждый месяц как минимум один гражданин писал нам запрос на «удаление его персональных данных».

Почему так происходит? Потому что, с одной стороны, российское законодательство требует обеспечивать доступность данных о закупках; с другой стороны, данные о регистрации юридических лиц и индивидуальных предпринимателей также общедоступны, но если при ликвидации юридического лица его статус просто меняется в базе данных, то после ликвидации ИП граждане начинают искать себя в интернете и требовать удалять себя из всех баз данных, аргументируя тем, что теперь они просто физические лица.

---

[1] Open Rights Group https://www.openrightsgroup.org
[2] Open Data Privacy Campaign https://www.openrightsgroup.org/campaigns/opendata/open-data-privacy

И это лишь один из немногих примеров, не учитывающий многочисленных ошибок, совершаемых госслужащими при раскрытии информации о том, что раскрываться не должно. Одна из ключевых проблем работы с государственными данными — публикация их в PDF-форматах, является и ключевой причиной, почему так сложно выявить нарушения приватности и необоснованное раскрытие личных данных. Ведь эти данные можно найти только путем ручного просмотра значительного числа документов.

Российские реалии, в которых у нас нет ни возможности системного открытия данных, ни качественного контроля за приватностью раскрываемых баз данных, безусловно, должны меняться.

Создание единых правил работы с информацией, формирование технологических, нормативных и организационных решений в создании национальной государственной инфраструктуры по работе с данными всех видов — вот та задача, которую нам предстоит решить в ближайшие годы.