

информационные технологии

Открытые двери

— кооперация —

Звон монет

Сергей Сапонов, директор по развитию бизнеса ООО «Потенциал», говорит, что принципы использования открытых данных на российском рынке отличаются от зарубежных. «Москва, как пионер в этой сфере, ориентируется не на последующую реализацию продуктов на основе данных, а на прямое потребление данных пользователем», — утверждает он. — Безусловно, на нашем рынке имеются интересные продукты на основе опубликованных данных, но их уровень монетизации либо остается достаточно непрозрачным, неочевидным, либо оставляет желать лучшего. Это связано с психологией российского пользователя, который не привык платить за контент. А создавать интересные и актуальные продукты для бесплатного потребления могут либо энтузиасты, либо профессиональные разработчики по заказу различных структур». Видимо, по этой причине на портале открытых данных data.gov.ru перечислено всего 242 готовых приложения, которые могут скачать пользователи. Это мало по сравнению с тем, сколько наборов данных доступно — 7060 по самым разным темам и регионам. В разделе «Экология» доступно 228 наборов данных, а приложений всего 2: для мониторинга чистоты воздуха в Москве («ЭкоМониторинг») и для получения сведений о качестве водопроводной воды по районам столицы («Вода в Москве»). В разделе «Здоровье» 320 наборов и 11 приложений, включая такие, которые подсказывают месторасположение туалетов



Раскрытие данных госорганами — необходимый процесс для построения информационного общества

в Воронеже, помогают записаться к врачу в Санкт-Петербурге и так далее. Привлечь внимание стартаперов, а также вовлечь людей в активную государственную деятельность пока не особенно удалось. Виктор Гриднев, руководитель Центра технологий управления данными компании «АйТи», говорит, что придумать, как использовать большую часть размещенных на data.gov.ru данных, — занятие достаточно сложное. «С другой стороны, и у наших западных коллег, на которых у нас принято ссылаться, большая часть размещенных данных посвящена мировому океану и погодным условиям», — говорит он. — Какую пользу несет «Перечень учреждений культуры Боровичского муниципального района» (из российского

data.gov.ru) или «Анализ данных водного столба гидролокатора» (из американского data.gov) очень непросто оценить — наверное, найдутся люди, кому это нужно и важно».

Александр Чулапов, консультант по работе с государственными организациями Teqdata, рекомендует при раскрытии информации ставить себя на место потребителей этих данных. «Если открытость информации нивелируется отсутствием возможности связать данные из различных источников между собой или для этого требуются значительные усилия, то какой-либо практический смысл от подобной открытости отсутствует». По его мнению, раскрываемые данные должны содержать ключевые поля, которые бы позволили связать различные данные между собой и сопоставить. Из других недочетов, которые сегодня наблюдаются при раскрытии данных, «Инфокультура» отмечает то,

что госучреждения часто указывают условия их использования, не обозначая тип лицензии, по которой их можно потреблять. Это затрудняет работу с такими данными. Также некоторые наборы приходится скачивать кусками, а то и вовсе они могут быть недоступны для скачивания, как, например, информация о движении поездов РЖД. Некоторые сведения, перечисленные в Хартии G8, в России доступны лишь за деньги, например информация о юридических лицах (ЕФРЮЛ и ЕФРИП). «Говоря о содержательности, можно отметить, что наиболее полно относительно прочих сейчас представлены открытые данные по финансам (публикуемые Казначейством и отчеты Министерства финансов) и данные по статистике (ЕМИСС)», — говорится в аналитическом отчете «Инфокультуры».

Василий Пушкин в той же записи в блоге подтверждает, что недочеты еще имеются: «В текущем году государству необходимо выработать подходы к единым стандартам публикации и связывания данных, обеспечить доступ к данным через единый портал, доработать нормативно-правовую документацию, а также продолжить обучение основам работы с данными, в том числе открытыми».

В качестве хорошего примера практического использования открытых данных господин Гриднев приводит «Яндекс.Транспорт» и другие инициативы этой компании, а также проекты, связанные с путешествиями и культурой, которые разрабатывали субъекты РФ. Успешно монетизируются проекты «Гарант» и «Консультант+», сегодня наблюдаются при раскрытии данных, «Инфокультура» отмечает то,

на их основе. Правда, начали они это делать задолго до появления термина «открытые данные».

Александр Чулапов рассказывает о зарубежном опыте: «Среди наиболее заметных проектов по сбору данных государственными органами отмечу Австралию. Система стандартизированной бизнес-отчетности позволяет унифицировать технологию обработки и сбора данных, необходимых государству для осуществления своих функций. Открытость бизнес-гlossария и требований к форматам привлекает независимых разработчиков на рынок сбора отчетности и служит инструментом снижения затрат на информационные технологии в государственном секторе». Один из клиентов Teqdata в госсекторе — Австралийская налоговая служба — получает более 60 форм отчетности в стандартном формате. Затраты на разработку систем сбора отчетности и их сложность уменьшаются, освободившиеся средства перенаправляются на анализ собранной информации. Тенденция к открытости и прозрачности способствует доверию между гражданами, коммерческим сектором и государством. Господин Чулапов приводит в пример также проект, обобщающий мировой опыт, — сайт Sunlight Foundation, собирающий десятки историй использования Open Data на благо общества. В нашей стране «Инфокультура» развивает похожий проект HadofData.ru, где ведет каталог и хранит открытые данные из всех российских источников.

Андрей Свиришевский, руководитель направления углубленной аналитики «SAS Россия/СНГ», смотрит на вопрос работы с открытыми источ-

никами данных широко. Он отмечает, что сегодня практически вся личная информация, находящаяся в открытых источниках (соцсети, поисковые сайты, поисковые запросы и т. д.), собирается и используется с самыми разными целями. Ее также можно считать разновидностью открытых данных. Их могут использовать, например, для показа более релевантной рекламы. Либо для повышения уровня безопасности. Так, в Голландии таможенная служба с помощью средств текстовой аналитики проанализировала контент форумов и обнаружила поток наркотрафика, идущего через Канаду.

Степан Томянович рассказывает о решении, разработанном ФОРС, которое анализирует данные из социальных сетей (ForSMedia) и извлекает из открытых данных новые знания о существующих или потенциальных клиентах компании. По сути, это платформа для целевого онлайн-маркетинга. В РОСЭКО был реализован проект по созданию национальной системы стоимостной аналитики на базе когнитивных ситуационных центров. Частью этого проекта было создание системы для оценки недвижимости на основе данных, доступных в открытых интернет-источниках.

Можно считать, что основные положения Хартии G8 Россия реализовала на практике. Заложена основа для дальнейшего развития этой темы. На следующем этапе «Инфокультура» предлагает создать национальную инфраструктуру данных — платформу для каталогизации, сбора и интеграции данных, что облегчит доступ к ним и поможет обеспечить связность.

Светлана Рагимова

«Для урбанистов открывается масса возможностей прогнозирования»

— эксперт —

ВИКТОР СТРЕЛКОВ, первый вице-президент компании «Техносерв» по системной интеграции и программной инженерии, рассказывает о будущем российских городов, которые благодаря большим данным должны стать более безопасными и удобными. По его словам, системы класса «Безопасный город» уже функционируют или внедряются в более чем 50 населенных пунктах РФ.

— Какого рода данные анализируются сегодня в ведомственных системах и что с ними можно делать? — Сейчас «Техносерв» реализует проект АПК «Безопасный город» в Курске, Нижнем Тагиле, Нижнем Новгороде, Якутске, Мордовии и Великом Новгороде. По большому счету регионов мы работаем в тандеме с «Ростелекомом». Те системы, которые уже функционируют, собирают значительное количество данных. Пока их объемы не так велики, но спектр весьма широк. Это информация о поведении

людей (фиксация правонарушений, в том числе уголовных), о техногенных катастрофах, социальных мероприятиях, транспортных потоках, платежах за парковки и за общественный транспорт и т. д.

На следующем этапе будут разработаны модели, с помощью которых эти данные можно будет анализировать и выработать сценарии реагирования на различные ситуации. Эти модели затем можно будет обобщать и тиражировать в других регионах, разных отраслях и т. д. Одного года

сбора данных в принципе достаточно, чтобы выработать такие модели. Это позволит более эффективно реагировать на события, а также прогнозировать их. Например, можно легко предвидеть транспортный коллапс при возникновении нескольких событий, его предвещающих. — То есть сейчас уже возможно делать такие прогнозы?

— Для некоторых событий уже накоплено достаточно данных. Но по тем, которые, например, связаны с сезонностью, пока информации собрано мало. Если говорить о транспортных потоках, то дачный сезон случается раз в год и имеет свою специфику. Чтобы провести анализ, выявить закономерности того, как люди передвигаются по городу и пригородам в это время, нужно несколько лет.

— Могут ли «городские» данные пригодиться бизнесу?

— Муниципальные власти имеют прекрасную возможность монетизировать накопленную информацию. Например, данные о грузоперевозках могут подсказать, где установить заправки для дальнотойщиков, автосервисы, кафе и мотели. Это идеи, которые на поверхности — можно уходить вглубь почти бесконечно. К примеру, если соединить данные о покупках жителей небольшого города (вытащить их через карты лояльности торговых сетей) и информацию об их передвижениях (из систем фото- и видеофиксации), можно понять, где построить продуктовый магазин и что именно там продавать.

— Могут ли злоумышленники использовать такие данные? Например, понять, когда, в каком районе лучше всего совершать грабежи?

— Не исключено, но в случае с грабежами это слишком затратный метод. Самый верный способ — иметь инсайтера в грузовой компании, который сообщит, когда фура с дорогими товарами проедет через определенный населенный пункт.

Кроме того, к счастью, большие данные уровня города всегда обезличены. Если речь идет об информации о правонарушениях конкретных лиц или базе автовладельцев, то они хорошо защищены и не могут использоваться за пределами контура информационной системы госучреждения. Системы, анализирующие связи между различными событиями и объектами, берут только обезличенные данные из разных не связанных между собой систем.

Конечно, любую систему могут взломать. Но в решения «Техносерв» мы встраиваем по умолчанию средства защиты нескольких уровней. К примеру, запрос о конкретном автомобиле сопровождается использованием электронной подписи конкретного сотрудника. Всегда можно выяснить, кто, когда и зачем обращался к системе.

— Как работает аналитика больших данных в системах «Безопасный город»?

— В комплексе «Безопасный город» помимо фотовидеонаблюдения входит мониторинг инфраструктуры и окружающей среды. Ведется наблюдение за объектами водоснабжения, сельского хозяйства, промышленными пред-



РЕКЛАМА

вместе с ним — схему того, как нужно действовать по регламенту.

Служба 112 — это первая линия работы с населением, фактически это контактный центр. ЕДДС и наша система стоят во многих случаях сразу за 112. Но не во всех городах уровень автоматизации одинаков. В некоторых случаях мы реализуем элементы 112, потому что они еще не созданы. Также интегрируем ее с Комплексной системой экстренного оповещения населения (КСЭОН).

— Расскажите о возможностях предиктивного (прогнозного) анализа в масштабах города.

— Для урбанистов открывается масса возможностей прогнозирования: системы анализа больших данных умеют искать закономерности и первопричины событий и предсказывать их. К примеру, из прогнозов метеорологов становится ясно, что лето будет очень жарким. Система может проанализировать расположение торговых точек и подсказать, где необходимо сделать санитарные вырубки или принять другие превентивные меры, чтобы избежать пожаров. Если возгорание все же возникло, система может выявить причины и в случае наличия признаков противоправных действий передать эту информацию в соответствующее ведомство. Предиктивный анализ может хорошо монетизироваться. Мы можем заранее понять, куда будет развиваться город и с какой скоростью. Если появилась застройка, появляются и магазины, за ними — транспортные магистрали, магазины с товарами непостоянно спроса (мебельные, например), заправки. Вдоль построенных магистралей застройка пойдет дальше.

— Есть ли уже примеры такого рода монетизации?

— Различная статистическая информация уже доступна, и ею можно пользоваться. Есть открытые кадастры и данные о самых разных событиях. Но пока эти данные не слишком удобны для использования. «Техносерв», кстати, такой же потребитель этих открытых данных, как и все. Мы используем их для создания моделей при проектировании систем «Безопасный город» для конкретных населенных пунктов. Модели для каждого города уникальны. Где-то есть река, а где-то гористая местность, где-то есть градообразующее промышленное предприятие, а где-то мосты разводятся ночью и т. д.

— С чего обычно муниципалитеты начинают строить «Безопасный город»?

— Как правило, с систем фотовидеонаблюдения, потому что понятно, как их монетизировать: регион получает в бюджет отчисления от штрафов за нарушения ПДД, и дополнительно идет снижение аварийности. После этого уже думают о социальных системах, приносящих пользу населению. Затем начинают внедрять комплексные системы обеспечения безопасности, систему «интеллектуального» видеонаблюдения, ту же КСЭОН, систему мониторинга экологической обстановки и т. д. — все зависит от конкретного региона.

Интервью взяла Мария Анастасьева



AT Consulting

КРУПНЕЙШИЙ ПОСТАВЩИК УСЛУГ
в сфере информационных технологий

Доверие. Энергия. Экспертиза.



/ Внедрение и сопровождение сложных информационных систем

/ Разработка ПО на заказ

/ Управленческий и операционный бизнес-консалтинг

/ Управление проектами

/ ИТ-аутсорсинг

www.at-consulting.ru