

# Открытые данные, доступные данные и перспективы территориального анализа

Ф.Б. КАЦ (Университет ИТМО, Санкт-Петербург)



Кац Филипп Борисович в 2009 г. окончил Казанский государственный архитектурно-строительный университет (магистр архитектуры). В 2011-2012 гг. прошел обучение в Институте Стрелка по постдипломной программе (совместно с Массачусетским технологическим институтом) под кураторством Карло Ратти. Преподаватель курса городской информатики магистерской программы международной лаборатории «Дизайн городских экосистем» Национального Исследовательского Университета Информационных технологий, Механики и Оптики.



Предсказанная полвека назад Маршаллом Маклюэном информационная революция совершилась; каждый из нас производит и отправляет в сеть информацию ежечасно и ежесекундно. Решения, принимаемые на основе этой информации, и постоянное взаимодействие людей определяют изменения, сила и скоротечность которых значительно превышает традиционные возможности прогнозирования и целеполагания на уровне обществ и государств. Сосредоточения интересов, ресурсов и технологий города в широком смысле стали наибольшими бенефициарами и первыми жертвами этой революции.

Бурные изменения – цена и следствие бума информационных технологий. Результат – увеличение рисков и сложность прогнозирования, взаимозависимость множества факторов, как следствие, увеличение шансов кумулятивного эффекта, того, что иначе называют «Идеальным штормом» или «Чёрным лебедем». С другой стороны, новые возможности усилили и механизмы самонастройки общества. Так, вооружённые информацией и технологией коммерческие платформы, организации активистов и локальные сообщества рачительных горожан ведут экспансию в сферах, традиционно делегируемых авторитетным институциям, правительственным и городским администрациям. Ни «лебедей», ни процессы самонастройки фактически невозможно заметить на стадии появления и развития традиционными методами. В то же время их потенциал – опасный или полезный может быть раскрыт с помощью современных методов анализа данных. Информационные технологии переживают период пространственного осмысления. Существующие решения и разработки адаптируются к текущим запросам городов. С другой стороны, и сами города создают процессы, формирующие запрос на технологию и информатизацию. В этом двунаправленном процессе города постепенно входят в эпоху больших данных и автоматизации управления.

## Умные города и приборные панели

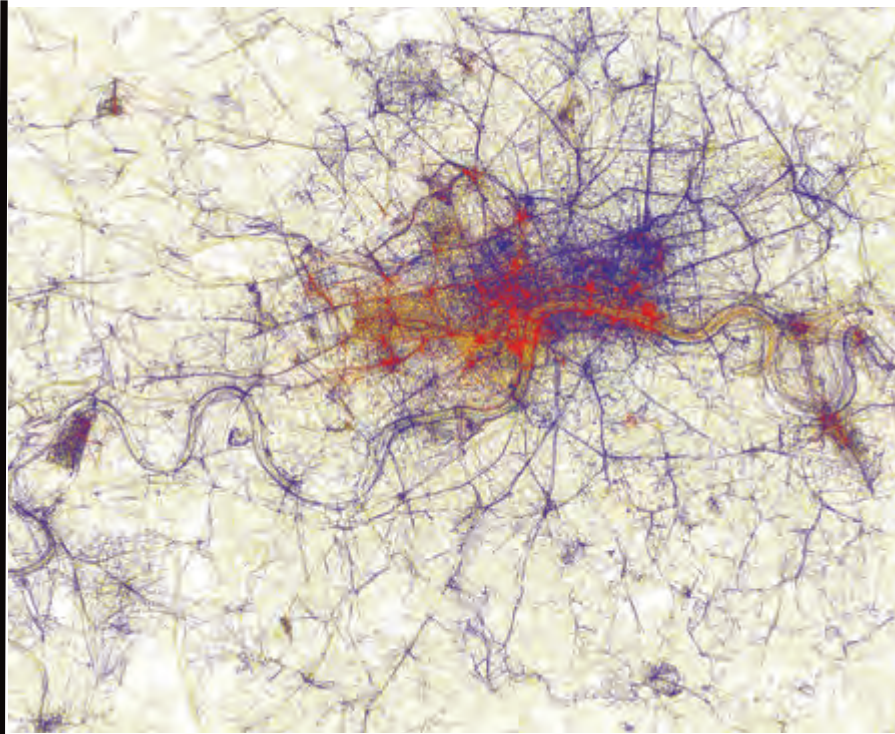
Впрочем, ещё лет десять назад, задолго до появления дисциплины городской информатики и термина «Большие данные» идея автоматизации городского управления нашла себя в модернистской концепции «Умного города» (Smart city), предлагающей сделать города машинами для жилья в самом что ни на есть прямом смысле. Использование вычислительных машин в управлении городами активно продвигается большими корпорациями, такими как CISCO, IBM и др. Как правило, проекты предполагают интеграцию существующих датчиков и отчётности в централизованные магистральные потоки, позволяющие отслеживать ситуацию и автоматизировать, а значит, и оптимизировать часть городских процессов.

В каких областях управления городом работают Smart cities?

Среди основных интересов: транспорт, энергетика, преступность и экология. Такая приоритизация имеет свои причины. С одной стороны, городские власти хотят получать данные в реальном времени там, где они могут в реальном же времени реагировать – вызвать наряды полиции, переключить светофоры, задействовать резервные электроподстанции.

С другой стороны, все три первостепенных аспекта – энергетика, транспорт, безопасность – изначально обладают частью инфраструктуры и в принципе поддаются контролю. В то же время для централизованного мониторинга такого комплексного понятия, как городская среда требуются действительно серьёзные вложения и тяжёлая работа по поддержанию. Сам принцип централизованной системы датчиков едва ли масштабируем, а значит, большие средовые исследования делать с его помощью невозможно.

В результате жители города, их ощущения и действия из системы, по сути дела, изъяты – универсальный умный город, доводя идею до предела, будет функционировать и даже, весьма вероятно, покажет более высокие показатели



**1** Карта геолокаций постов сетей Твиттер и Инстаграм на карте Москвы, Эрик Фишер, 2010 (<https://www.flickr.com/photos/walkingst/4671589629/in/set-72157624209158632>)

## Данные

С развитием технологий и бумом интернет-сервисов общественная и коммерческая ценность данных раскрывается максимально. Идея открытых данных – машиночитаемых, свободных от каких-либо ограничений и авторских прав – официально появилась в 2007 году и быстро получила признание по всему миру. Активисты считают, что все данные, собираемые и хранимые государством, по возможности должны быть опубликованы ради общественной пользы. Действительно, на основе публикуемых данных работают коммерческие и бесплатные сервисы, проводятся научные исследования, осуществляется общественный контроль. Все это, в свой черёд, позволяет оптимизировать процессы на аутсорсе за счёт независимых научных исследований и народной смекалки. Не в последнюю очередь открытость данных увеличивает взаимодействие и внутри самого государственного аппарата, позволяя «правой руке ведать, что творит левая», принимать более продуманные решения.

Открытые данные имеют свои ограничения и болевые точки. В первую очередь до публикации данные нужно ещё и собрать, что составляет 90% работы и требует постоянного поддержания качества и чистоты отчётности. Для этого необходима отдельная развитая структура, защищённая от влияния со стороны заинтересованных лиц. Часть работы приходится на волонтеров и активистов: зачастую наборы данных собираются волонтерами, государство лишь верифицирует и публикует их.

Помимо этого, структура и содержание данных сильно зависят от сложившейся практики и законодательства, технической инфраструктуры. К примеру, практически единственный постоянно обновляющийся набор данных, опубликованный правительством Москвы, – база лицензий на торговлю спиртным, которая, по всей видимости, обладает достаточной динамикой, чтобы техническое обновление требовалось самим чиновникам.

Впрочем, официальные данные, открытые или нет, – лишь малая толика информации, которая может быть использована для исследований и мониторинга, и уж точно не самая ценная, если ваш предмет интереса – общество или городская среда. Настоящая революция произошла с распространением смартфонов. Смартфон произвёл с обществом удивительную трансформацию: модная дива или молодая студент-

ли эффективности, если останется без жителей.

За всю свою историю концепция умного города реализовалась лишь в небольшом количестве пилотных проектов. Основным тормозящим фактором является цена, однако и воплощённые решения оставляют вопросы эффективности каждого проекта в целом. К примеру, экогород Тяньцзинь так и остаётся полуселённым по экономическим соображениям, заложенный девять лет назад Масдар пропал с новостных радаров. Что характерно, почти все города в семёрке самых умных городов мира по версии ICF (Intelligence Community Forum) в 2015 году – города сравнительно небольшие (как, например, округ Арлингтон в Вирджинии, город Колумбус в Огайо и Ипсвич в Австралии, попавшие в первую тройку). Их умная составляющая предполагает значительно более скромные и практичные решения, нежели тот же шумевший Масдар.

## Дэшборды (или «Приборные панели»)

Другой предшествующий тренд – увлечение приборными панелями города (city dashboards). Суть идеи – создание «терминалов», или информационных панелей, позволяющих ответственным лицам видеть все необходимые показатели в реальном времени. Образ терминалов, очевидно, создан приборными

панелями самолётов, фантастической литературой и залами управления полётов NASA. В то же время опыт практического применения панелей пришёл из бизнеса. Наибольшую известность получили терминалы Блумберг, но на самом деле свои терминалы существуют в большинстве современных корпораций.

Наиболее известной реализацией приборной панели города является, пожалуй, Операционный Центр IBM в Рио-де-Жанейро, впечатляющий своими бесчисленными экранами и постоянно агрегирующий информацию 30 городских департаментов. Аналогичные решения получили и другие города, в том числе Балтимор и Лондон. С развитием сети Интернет терминалы получили веб-версии, а с какого-то момента стали и вовсе публичными, предоставляя данные любому желающему, впоследствии же многие из них переродились в платформы открытых городских данных.

Идея приборной панели, красивая в теории, на поверку не всегда оказывалась эффективной. Во-первых, получение и обновление данных – это достаточно ресурсоёмкий процесс, а большие временные перерывы между отчётами делают всю идею бессмысленной. Во-вторых, обобщённые показатели позволяли лишь оценить общие тренды развития города или районов, служба показателями эффективности руководства, но не помогая в принятии решений.



**2** Карта заведений Ростова-на-Дону на Foursquare, 2012.  
Александр Музыченко, Филипп Кац (<http://www.4sqstat.ru/>)

ка, бизнесмен или школьник, все те, кто ещё вчера казались безумно далёкими от технологий и компьютеров, стали активными пользователями разнообразных сервисов, приложений и гаджетов. Смартфон стал настоящим пультом управления жизнью – одно нажатие может обеспечить вас пищей, маршрутом или ожидающим у порога такси.

Телефоны оказались удивительно богатым источником информации для исследователя. С их помощью становится возможным отследить перемещение человека, узнать его пол и возраст, стиль коммуникации, при желании – классифицировать его деятельность, активность в течение суток, множество других показателей. Что немаловажно, такая сеть автоматически масштабируется, другими словами, стоимость исследования практически неизменна вне зависимости от того, получаете ли вы данные со ста или ста тысяч устройств.

Безусловно, огромное поле возможностей таит в себе информация от компаний-провайдеров. Данные о перемещениях людей, поисковые запросы, записи транзакций кредитных карт. Эти данные уже сейчас внимательно изучаются маркетологами и, безусловно, являются ценным источником для городских исследователей. Ограничениями служат, как правило, цена и вопросы приватности.

Однако не менее широкие возможности исследования существуют и при отсутствии больших бюджетов.

Источником данных могут стать всевозможные сервисы и социальные сети. Спортивные маршруты можно найти на спортивных сервисах типа Strava, заведения района находятся на Foursquare, достопримечательности и удачные ракурсы – на Flickr, жалобы и настроения – на Twitter, бесплатная открытая картография – на OSM, а такой сложно оценимый параметр, как загруженность улиц

и качество дорог, как выяснилось, гораздо проще доверить навигаторам в телефонах автомобилистов, нежели централизованной системе камер.

Отдельно стоит упомянуть сервисы для активистов, создаваемые самими городами, такие как SeeClickFix, Citizens Connect, Город.Мос и другие, позволяющие горожанам напрямую передавать властям сообщения о поломках, недостатках и прочих проблемах городской инфраструктуры напрямую.

Практически каждая современная платформа или сервис позволяют получить доступ к своим базам, причём в подавляющем большинстве случаев бесплатно или за очень скромную плату.

Там, где стандартного доступа к данным не предусмотрено, в первую очередь на сайтах, изначально не предназначенных для сбора информации, сбор также может быть проведён, отличие лишь в инструментах. При определённой изобретательности такие «посторонние» источники могут оказаться наиболее доступным и простым источником, предоставляя неожиданно богатые и интересные результаты. К примеру, коммерческие базы ресторанов и кафе часто классифицируют заведения по ближайшей станции метро, позволяя проанализировать городское пространство без какой-либо сложной обработки данных.

Сеть Интернет предоставляет большое разнообразие источников. Ограничением является в первую очередь находчивость исследователя в интерпретации существующих данных. Любопытный исследователь может найти удивительные инсайты в блогах энтузиастов, справочниках, онлайн-барахолках, базах недвижимости.

Данные, полученные с помощью телефонов или в сети Интернет, – те самые мистические «Большие данные», соот-

ветствующие всем трём классическим критериям – серьёзным объёмам, непрерывно обновляемому потоку (данные постоянно обновляются, и анализ самой динамики обновления зачастую даёт интересные инсайты) и разнообразию – информацию часто приходится чистить и трансформировать, интерпретировать. Для её сбора и обработки нужны определённые технические навыки и ресурсы. Впрочем, как правило, эти затраты в разы меньше, чем даже опросы населения.

Такой тип данных, безусловно, обладает своей спецификой. В первую очередь практически все они являются побочным продуктом, собраны не под конкретное исследование или запрос и в большой степени находятся под влиянием источника: например, смена интерфейса приложения может значительно изменить результаты. Соответственно, информация требует особенно аккуратной выверки и осторожности при интерпретации.

Большая проблема данных такого рода – неполнота и нерепрезентативность. Действительно, существующая информация в большинстве своём не является полной или репрезентативной, описывают только толику происходящего. Только часть фотографий города (самые красивые?) залита на Flickr, только часть жителей города (самые прогрессивные, молодые и богатые?) используют Foursquare. Оценить репрезентативность данных и нормализовать их – отнюдь не простая задача. В большинстве случаев границы интерпретации и логический аппарат приходится формулировать специально под задачи конкретного исследования.

Другой аспект – приватность информации и необходимость анонимизации данных.

Минусы такого подхода, как правило, нивелируются плюсами: сравнительно низкой стоимостью, возможностью масштабирования или переноса. Сбор и анализ фотографий в Рязани ничем не отличается от сбора и анализа в Суздале, может быть выполнен теми же инструментами и той же командой, часто требуя лишь нескольких дополнительных нажатий клавиш.

### Наука о данных

Наличие данных дополняется доступностью и активным развитием различных техник их обработки. Активное развитие происходит по всему спектру инструментария – это ПО для очистки, автоматизированной сборки и визуализации



зации больших объёмов информации, автоматизация краудсорсинга, инструменты машинного обучения и продукты на его основе. Сравнительно простые инструменты – распознаватели языка или эмоций по тексту, улыбок на фотографии – позволяют обрабатывать потоки информации под новыми углами, выуживая дополнительные смыслы и инсайты. За этим развитием происходит становление «Науки о данных» (Data science) – дисциплины, объединяющей в себе классическую статистику и навыки компьютерной обработки информации. В настоящее время дисциплина переживает бурный рост.

## Городская информатика

Новые возможности городских исследований, предоставленные недоступными ранее количеством данных и технологиями их анализа, ещё только раскрываются и осознаются. Исследовательское поле формируется и напоминает Клондайк – любая не использованная ранее технология, источник данных или оригинальная интерпретация могут создать впечатляющий результат даже без сложного и кропотливого анализа. Именно поэтому исследования зачастую носят поисковый характер. Большую роль в формировании дисциплины пока играют художники и энтузиасты от технологии. Так, широко известные карты геолокаций сообщений твиттера и фотографий Инстаграма или фотографий, загруженных на сервис Flickr для десятков разных городов мира, – дело рук Эрика Фишера, цифрового художни-

ка, и были сделаны без какой-либо исследовательской цели и последующей аналитики, в первую очередь как своего рода доказательство: города безошибочно угадываются в очертаниях, сформированных тысячами точек.

Безусловно, ситуация меняется, и на смену экспериментаторам приходят серьёзные исследовательские коллективы. К пространственному анализу данных подключается всё больше специалистов, исследования становятся более сложными, а их результаты – более практически применимыми.

Академическими исследованиями по городской информатике профильно занимаются всего несколько институций в мире. В то же время сам термин «Городская информатика», к примеру, получил известность благодаря Маркусу Фоссу, создателю лаборатории городской информатики при Технологическом Университете Квинсланда в Австралии. Такой «вброс» показателен – множество небольших групп и лабораторий занимаются темой эпизодически или работают в смежных областях. Это архитекторы и градостроители, географы и экономисты, не в последнюю очередь и журналисты.

Спектр коммерческого применения дисциплины также возрос. Анализ среды перестал рассматриваться в плоскости интересов городской администрации, девелоперов и геомаркетологов. В исследованиях такого рода заинтересованы и сервисы, такие как Amazon и Google, Bloomberg и Palantir.

Несмотря на большой запрос и хорошие перспективы, развитие дисциплины все

ещё ограничено небольшим количеством специалистов и междисциплинарностью.

В 2013 году вместе с коллегами я начал вводный курс «Городская информатика» для первого курса магистратуры «Дизайн городских экосистем» Университета Информационных Технологий, Механики и Оптики.

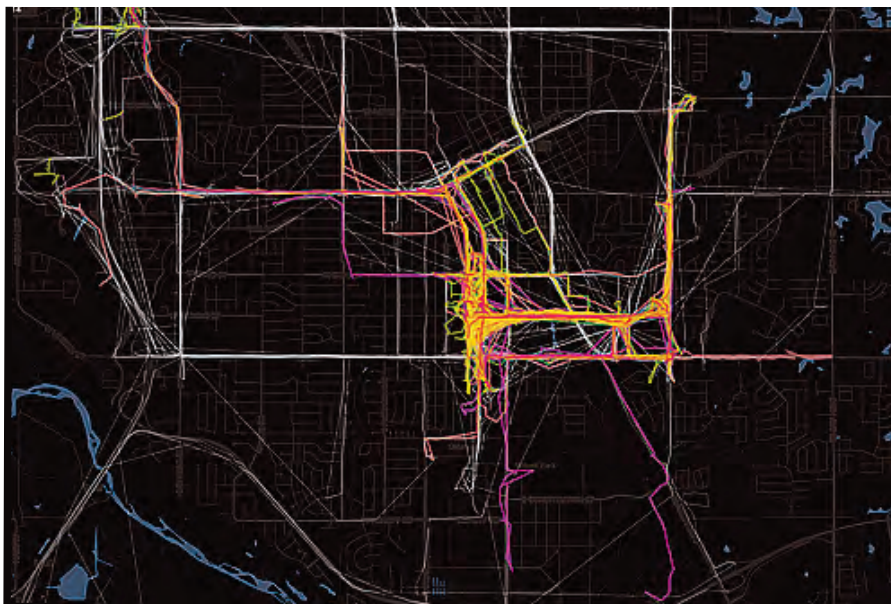
## Утопии и практические воплощения

Желание, а главное, возможность, информационного анализа, происходящего в городской среде в реальном времени, может напоминать антиутопию. Однако я склонен относиться к этим интенциям со сдержанным гражданским торжеством: в то время как умные города, предполагая механистическую концепцию управления, продолжают в определённом смысле идеи модернизма, городская информатика фактически признает несостоятельность подобного подхода в применении к обществу, предполагает вместо него вдумчивый анализ и интерпретацию происходящего. Иначе говоря, общество вновь оказывается сложнее моделей.

В условиях России такой подход оказывается едва ли не безальтернативным. К сожалению, мы практически не имеем доступа к подробной, качественной и регулярно обновляемой статистике по городам и отдельным кварталам. В этом вакууме информация, собираемая в сети Интернет и социальных сетях, с вышек сотовой связи и трудами активистов, становится единственным источником для исследования и принятия решений.

Городская информатика все ещё формируется, но формируется очень быстро. За последние годы в дисциплину вошли серьёзные академические институции и технологические гиганты. На смену поисковым и сравнительно простым исследованиям приходят исследования комплексные и использующие передовые алгоритмы.

Конечно, городская информатика не сделает города идеальными и даже, скорее всего, не сильно изменит обыденную жизнь горожан. Однако такая работа поможет совершать меньше ошибок, ослабить социальные напряжения, направлять необходимую помощь вовремя туда, где она действительно нужна. Городская информатика может помочь увидеть и понять процессы, самопроизвольно зарождающиеся в городском пространстве, и позволить вовремя на них реагировать, в конечном итоге действительно делая мир чуть лучше. ■



**3** Персональный архив перемещений по городу Норман США, 2014 – данные за 1 месяц, Филипп Кац