

Вызовы технологий «больших данных» для современных социогуманитарных наук*

Е.Ю. Журавлева

Научный интерес к появлению новых техник работы с интенсивными данными, или технологий «больших данных», генерирует широкий спектр новых исследовательских практик в почти каждой современной дисциплине. Статья нацелена на анализ изменений, связанных с освоением подобных исследовательских практик в социогуманитарных науках. Отмечаются два взаимосвязанных контекста рассмотрения технологий «больших данных»: индустриальный и научный. Предложены различные виды определений технологий «больших данных» с количественным фокусом (размеры, свойства, объемы, структура или композиция данных). Особо выделено определение технологий «больших данных» в перспективе изучения социального движения, поведения людей и общественного характера событий. В этой группе определений важная роль отводится производному от «больших данных» понятию «даннофикация». Даннофикация представлена как современный технологический макротренд по трансформации социальных действий в онлайн-количественные данные, взаимосвязанный с такими современными технологическими макротрендами, как дигитализация, сенсоризация и софтверизация. Основные изменения выявлены для современных социогуманитарных наук на когнитивном, эпистемологическом, методологическом, институциональном и этическом уровне познания. Особое внимание уделено таким проявлениям данноизма в научной деятельности, как разведывательная наука, поисковая наука, наука, свободная от гипотез, наука, движимая данными, данноцентричная наука, эмпирическая эпистемология и наука «больших данных».

КЛЮЧЕВЫЕ СЛОВА: технологии «больших данных», даннофикация, сенсоризация, софтверизация, данноизм, разведывательная наука, поисковая наука, наука, свободная от гипотез, наука, движимая данными, данноцентричная наука, эмпирическая эпистемология, наука «больших данных».

ЖУРАВЛЕВА Елена Юрьевна – кандидат философских наук, доцент кафедры общественных дисциплин Вологодского филиала РАНХиГС при Президенте РФ, г. Вологда.

eresearch7@gmail.com

Рукопись поступила в редакцию 21 августа 2016 г.

Цитирование: *Журавлева Е.Ю.* Вызовы технологий «больших данных» для современных социогуманитарных наук // *Вопросы философии.* 2018. № 9. С. 50–59.

* Исследование выполнено при финансовой поддержке РФФИ, проект «Софтверизация современной научно-исследовательской деятельности: эпистемологические основания» № 16-03-50064.

© Журавлева Е.Ю., 2018 г.

По мнению Л. Флориди, человечество накопило 180 экзабайт данных и информации от момента возникновения письменности до 2006 г. С 2006 до 2011 гг. уже произведено 1600 Экзабайт. Эта цифра увеличивается четырехкратно приблизительно каждые три года [Floridi 2012]. Каждый день создается 2,5 нониллиона (10^{30}) байт данных, это приводит к тому, что 90% данных в мире создается за последние два года.

Термин «большие данные» относится к огромным, различным, сложным, продолжным и/или распределенным наборам данных, сгенерированным и/или полученным от всех иных цифровых источников, доступных в настоящем и будущем времени. Обычно выделяют три основных источника «больших данных» [Kitchin 2014^b]. Во-первых, сбор данных от управляемого наблюдения. Во-вторых, автоматизированные формы производства данных, включающих в себя полученные в результате работы цифровых устройств, сенсоров, сканеров [Журавлева 2012]. И, в-третьих, генерация волонтерских данных, включающих в себя транзакции, социальные медиа (посты, фото, видео), «обратное наблюдение»¹, проекты краудсорсинга² и т.п.

Традиционно технологии «больших данных» рассматриваются в двух взаимосвязанных контекстах: индустриальном и научном. В первом случае как результат огромных инвестиций, которые потребуются для их полезного использования. Международная корпорация данных IDC опубликовала исследование, согласно которому рынок технологий «больших данных» и относящейся к ним аналитики в 2016 г. составит \$23,8 млрд, а к 2019 г. достигнет \$48,6 млрд [Foster 2016]. А во втором случае «большие данные» являются источником потенциального интереса всех ученых, работающих над одним и тем же феноменом. При этом научные институты увеличивают попытки по совершенствованию «практик памяти» — техник и технологий, направленных на «сохранение и восстановление фактов» [Bowker 2006].

Существует предположение, что истоки технологий «больших данных» прослеживаются в проектах «большой науки», которые имели огромный недостаток целевых теоретических драйверов. Этот фокус был замещен различными мотивациями по сбору данных о мире. Несомненно, ученые используют теорию, которая формирует их методологию и интерпретивные структуры, но тем не менее мотивацией для объединенного проекта было не тестирование гипотез или теории, а сбор данных об изучаемом явлении [Aronova... 2010].

В целом, проанализировав источники по интересующему вопросу, можно выделить несколько групп определений технологий «больших данных». Слово «data» в переводе с латинского языка означает данность, т.е. факт или то, что дано. По латинскому смыслу слова «данные» являются объектами, данными в противовес «фактам», объектам созданным. «Мы столкнулись тем не менее, — замечает Х.Д. Райнбергер, — с затруднительным положением относительно того, что в дискурсе современных наук смысл этих слов становится диаметрально противоположным. Данные рассматриваются как нечто, созданное инструментами и произведенные создателем. Данные допускают генерирование, они могут быть сохранены и восстановлены. В противоположность данным, «факты» допускаются и берутся как нечто данное. Они являются твердыми объектами. Они просто есть» [Rheinberger 2011]. Также термин «данные» означает точное и объективное документирование для начала изучения возможности какой-либо сущности или процесса.

Самыми распространенными являются определения с количественным фокусом на размерах, свойствах, объемах, структуре или композиции данных. Данные в этих определениях измеряют на трех уровнях: объем, скорость и сложность. Большие данные предполагают огромный объем цифрового контента, который генерируется или онлайн, или оффлайн в социальных, коммерческих, научных и правительственных базах данных. Но термин означает не просто увеличивающийся объем, но также быстроту коллекционирования и увеличивающуюся сложность источников и форматов данных [Ruppert... 2013].

Большие данные — термин, который, как правило, применяется для описания огромного массива данных, которые накапливаются со временем; данные становятся

«большими» благодаря размеру и скорости поступления. Они не могут быть обработаны принятыми для малых данных методами: требуются специализированные методы (например, параллельные вычисления тщательно отобранной значимой информации) [Clough et al, 2014].

Р. Китчин детализировал количественные свойства «Больших данных»: гигантский объем (измеряемый в терабайтах, петабайтах, и возможно, в дальнейшем в экзабайтах или йотабайтах), высокая скорость обработки (работа в реальном времени), разнообразие, расширение масштаба, мелкозернистое разрешение и уникальная индексированность; относительность, гибкость и др. [Kitchin 2014a].

Далее можно выделить определения термина с акцентом на процессе, который включает в себя сбор, хранение, распространение и использование данных. В целом, процесс-ориентированные определения рассматривают перспективы продвижения фронта вычислительных технологий в области управлении сложностью. В связи с чем Д. Вайнбергер предположил, что традиционная цель науки – конструирование теории, которая поддерживает и изучает факты – начинает постепенно трансформироваться в производство сложных моделей, которые объясняют сложные отношения [Weinberger 2011].

Следующая группа определений подчеркивает роль технологий больших данных в перспективах изучения социального движения, поведения людей и общественного характера событий. Большие данные – это изначально не систематизированная информация о социально-экономическом развитии и окружающей среде, генерируемая на основе информационно-коммуникационных технологий вне официальной статистики. Примером использования «больших данных» в этом ракурсе является проект ООН «Глобальный пульс». Его цель – создание сети инновационных лабораторий, координирующих исследования по применению технологий больших данных в магистральных направлениях деятельности ООН: климат и устойчивость, защита конфиденциальности данных, экономическое благополучие, продукты питания и сельское хозяйство, гендерные и гуманитарные вопросы, здоровье общества, цели устойчивого развития после 2015 г., оценка развития в режиме реального времени.

В этой группе определений особая роль отводится понятию «даннофикация». Даннофикация, по мнению В. Майер-Шенбергер и К. Кукьера, – это процесс представления явлений в количественном формате для дальнейшего сведения их в таблицу для анализа. Даннофикация несравнима с оцифровкой, так как при оцифровке аналоговая информация преобразуется в двоичный код, считываемый компьютером [Майер-Шенбергер, Кукьер 2014, 84]. Даннофикация, в свою очередь, – это процесс сбора, обработки, распределения, хранения данных полученных в результате работы цифровых устройств.

Даннофикация является современным технологическим макротрендом по трансформации социальных действий в онлайн-количественные данные, таким образом предоставляя трекинг в реальном времени и осуществляя предиктивный анализ. Бизнес и правительственные агентства используют экспоненциально увеличивающиеся метаданные, собранные через социальные медиа и платформы коммуникации, таких как Facebook, Twitter, LinkedIn, Tumblr, iTunes, Skype, WhatsApp, YouTube и свободных сервисов электронной почты, таких как gmail и hotmail в порядке трекинга информации о поведении людей. Даннофикация рассматривается как легальный способ доступа к мониторингу и пониманию поведения людей, и становится управляющим принципом не только среди техноадептов, но и среди ученых, которые рассматривают даннофикацию как революционные исследования, способствующие изучению поведения людей. При этом даннофикация укореняется в проблематичных онтологических и эпистемологических требованиях [Van Dijck 2014].

Даннофикация тесно связана с тремя важными современными технологическими макротрендами: дигитализацией, сенсоризацией и софтверизацией. Даннофикация относительно новый феномен, сравнимый с дигитализацией, характеризуемый взаимодействием между цифровыми и физическими объектами и массовым выполнением по индивидуальному заказу продуктов и сервисов для и при конечном пользователе, скорее, чем просто процесс автоматизации или повышения эффективности, хотя техники даннофикации могут быть использованы, конечно, для всего этого. Процесс

улучшения является фундаментально различным для даннофикации, тем не менее предполагает различные стратегические ландшафты, а также необходимость новых административных качеств. Процесс даннофикации опирается на новые формы квантификации и техник, связанных с получением новых данных.

Даннофикация одновременно содействует сенсоризации и софтверизации общества и наоборот. Она проявляется в обществе сама по себе в разнообразии форм и часто, но не всегда, связана с сенсорами/приводами и возможностями интернет-вещей. Сенсоры становятся более проникающими и доступными по цене. Например, смартфоны содержат в себе различные типы сенсоров: движение, локализация и т.п. Эта сенсоризация общества открывает, впервые в человеческой истории, потенциал сбора в режиме реального времени огромного количества данных и деталей почти обо всем [Lingen 2013 web].

Софтверизация управляет формацией нового «метамедиума», в котором становится объединенным то, что было отдельным, что всегда существовало. Этот метамедиум состоит из смеси алгоритмов, структур данных и техник, которые генерируют цели и то, что является медиаспецифичным, комбинирует производство различных форм «гибридности» и «глубокое проникновение способностей» [Kitchin 2013 web].

Завершающей группой определений больших данных являются когнито-ориентированные определения, главной особенностью которых является внимание к особым когнитивным способностям человека.

Феномен больших данных – это переход количества накопленной человечеством информации в качество решения задач, стоящих перед ним. Поразительнее всего то, что обществу придется отказаться от понимания причинности в пользу простых корреляций: променять знание «почему» на «что именно». Это переворачивает установленный веками порядок вещей и ставит под сомнение наши фундаментальные знания о том, как принимать решения и постигать действительность [Майер-Шенбергер, Кукьер 2014].

Когнитивный вызов технологий больших данных. «Мы начинаем осознавать, что технологии ‘больших данных’ могут выявить проблемы в макромасштабе, хотя, с другой стороны, при осознании теории о том, как работает система или феномен, нужно учитывать, что их взаимодействие так комплексно и масштабно, что человеческий ум просто не в состоянии этого охватить», – отмечает Д. Вайнбергер, тем самым формулируя суть когнитивного вызова технологий больших данных [Weinberger 2011]. В. Пич замечает, что пока люди были более эффективными в определении того, какие данные сохранять и какие забывать, компьютеры, а точнее программное обеспечение, в свою очередь, часто сохраняет и управляет всеми собранными данными [Pietsch 2013 web].

Итак, большие данные представляют то, что превосходит человеческую способность к постижению, и, следовательно, возрастают потребности опосредования этих процессов через трансдисциплинарную работу, технологические инфраструктуры, статистический анализ, техники визуализации для увеличения возможностей их интерпретации.

Методологический вызов технологий больших данных. Производство знания во многих дисциплинах, в том числе социогуманитарных, в настоящее время опосредовано кодами программного обеспечения, алгоритмами и данными. Для некоторых энтузиастов триумф компьютеров, данных и квантификации означает замену дисциплинарной экспертизы. К. Андерсон [Anderson 2008] высказал весьма категоричную мысль о том, что наука больше не нуждается в теориях, моделях, методологиях, онтологиях. «Научный метод» анализа «потока данных» может заменить все перечисленное. Хотя позднее К. Андерсон постарался смягчить это утверждение, объяснив свой поступок тем, что он как редактор журнала иногда накачивает эмоциями заголовки, создавая небольшое преувеличение для большего эффекта.

По мнению Э. Рупперта, понимание технологий больших данных требует и служит благоприятной возможностью для междисциплинарного развития методов, которые инновационно, критически и рефлексивно занимаются с новыми формами данных [Ruppert 2013].

Эпистемический вызов технологий больших данных. Еще в 2006 г. была предложена периодизация эволюции научных методов в контексте их соотношения с объемами используемых данных [Szalay, Gray 2006]. Первый период начался две с половиной тысячи лет

назад, когда данных было совсем мало и наука была эмпирической. В античные времена она ограничивалась описанием наблюдаемых феноменов и логическими выводами, сделанными на основе наблюдений. Второй период охарактеризовался появлением большего количества данных, позволяющих создавать теории, использующие в качестве доказательств те или иные аналитические модели. Затем наука стала вычислительной; компьютеры позволили использовать методы численного моделирования, что ознаменовало появление третьей парадигмы в последние 60 лет. Этому периоду свойствен «вычислительный поворот» в мышлении, и соответственно, в исследовании. Такой поворот ориентирует любую дисциплину в направлении использования вычислительных методов, трансформируя ее исследовательское поле аналогично тому, как методы, ориентированные на изучение языка, реконструируют поле в так называемом «лингвистическом повороте» англо-американской философии XX в. Например, в философии «вычислительный поворот» описывает сложные методы, благодаря которым увеличивается доступность информационно-коммуникационных технологий. Это предоставляет философам возможность исследовать спектр традиционных философских интересов, например, в логике, искусственном интеллекте, философии математики, этике, политической философии, эпистемологии, онтологии новыми методами, в новых ракурсах и выявить новые исследовательские модели и методы в цифровой среде, достойные внимания философа [Brey, Sørgaker 2009].

Четвертому периоду свойственна возможность обработки огромных объемов экспериментальных данных, появление новых научных методов, основанных на их анализе, и доминирование синтезирующих теорий. Данные содержат в себе много полезной информации, но их анализ невозможен ни в каком другом виде, кроме как автоматизированного [Heu... 2009]. Философию четвертой парадигмы, по мнению Д. Брукса, можно обозначить как «данноизм», который заключается в попытке собрать все данные и извлечь некоторую степень понимания [Brooks 2013]. Предполагается, что в основе современных социальных наук находится «даннологический поворот», который осуществляет не только серьезные изменения социологической методологии, но и более полную теоретизацию категории «социальное». Пришествие «даннологического» означает перераспределение технологий сбора и анализа социальных данных, изменяя тем самым эмпирическую социологию [Clough... 2014]. В принципе, этот «даннологический поворот» можно распространить почти на все науки.

Проявления данноизма обозначаются новыми понятиями: разведывательная наука (exploratory science), поисковая наука (Discovery Science aka "discovery-based science"), наука, свободная от гипотез (hypothesis-free science), наука, движимая данными (data driver science), данноцентричная наука (datacentric science), эмпирическая эпистемология (empiricist epistemology), наука больших данных (science Big Data).

В основе *разведывательной науки* находится метод разведки, который означает наблюдение, описание и картирование неизученной территории, а не тестирование теорий или моделей [Hallam 2013]. Философ науки Л. Франклин указывает на эксперименты, использующие «узкие» инструменты, с помощью которых получают группы измерений в ходе тестирования специфических гипотез, и «широкие» инструменты, позволяющие создавать тысячи и даже сотни тысяч измерений. С её точки зрения, целью разведывательной науки является максимизация эффективности научного поиска. Обычно «узкий» инструмент диктует, что возможно только одно или два измерения за один раз, что заставляет бережно относиться к каждому полученному результату. Иными словами, это эффективно при использовании инструмента по тестированию весьма специфических гипотез. С «широким» инструментом эти ограничения отпадают, эффективность поиска может достичь максимума при создании возможности множества измерений. Тем не менее эта экономия поиска может быть реализована, только если испытатель имеет несколько способов воспроизведения (процесс, анализ, сохранение, распределение) огромного количества данных, сгенерированных инструментом [Franklin 2005].

Поисковая наука или наука, основанная на поиске, — это научная методология, которая особое внимание уделяет анализу огромных объемов экспериментальных данных с целью найти новые паттерны или корреляции для формулировки гипотез и новых

научных результатов. Отличительный признак поисковой науки в том, что гипотезы быстро генерируются и тестируются на данных, которые коллекционируются потому, что имеют потенциал ответить на какие-то вопросы. Дж. Аллен считает, что поисковая наука равнозначна предоставлению компьютеру права «размышлять для нас» [Allen 2001, 542]. В плане развития поисковой науки интересен американский проект под названием «Центр больших данных для поисковой науки», который продвигает новую парадигму для взаимодействия с огромным количеством данных.

Истоки понятия «науки, свободные от гипотез» прослеживаются главным образом в биоинформатике. У «науки, свободной от гипотез» есть два преимущества. Первое заключается в том, что она фактически свободна от человеческих предубеждений, мнений или неустойчивых интерпретаций. Это основано на экстремально точных измерениях, часто использующих огромное количество образцов, собранных через интернациональные коллаборации, которые прилагают одинаковые измерительные методы и общераспространенный план анализа, так что результаты точно сопоставимы и не подвергнуты упрощенным вариациям, лежащим в основе небольшого количества моделей [Ioannidis... 2001, 306]. Во-вторых, исследования, «свободные от гипотез», не зависят от предшествующего знания, предоставляя огромный прорыв в научном поиске, пробуждающий новые открытия в понимании предмета или явления [Hunter, Kraft 2007, 436].

Но у такой науки есть не только сторонники, но и оппоненты, которые встревожены тем, что «наука, движимая гипотезой» изображается как «наука малого объема» и «узкосфокусированная». Критики считают, что истинная наука невозможна без гипотез, что уменьшение гипотез сигнализирует об отсутствии ясного понимания проблемы и что данные, полученные в рамках свободной от гипотез науки, не могут быть истолкованы и, следовательно, бесполезны.

Множество ученых способствует определению науки, движимой данными. Этот подход применяют в качестве способа извлечения данных при использовании установленных знаний и абстракции в исследовательской аналитике. Его сторонники вдохновляются идеей, что компьютерное программное обеспечение может играть известную роль в извлечении научно значимых структур (паттернов) из данных через статистический анализ или через поисковые механизмы в базах данных [Leonelli 2014].

В плане изучения философии науки интенсивных данных интересен европейский проект «Эпистемология “больших данных”», целью которого является познание того, как изменяются исследовательские практики в цифровой век и как это влияет на современное понимание научной эпистемологии в и за пределами философии науки. В особенности участники проекта фокусируются внутри и за пределами философии науки на мерах, которые увеличивают доверие к онлайн-базам данных.

С. Леонелли обозначает термином «наука, движимая данными» любое предприятие, в котором основное внимание уделяется генерированию, распределению, анализу и/или интерпретации данных. Отличительной особенностью такого исследования является высокая степень внимания к «ручным» практикам сбора данных (которые обязательно исключают теорию, модели, инструменты, программное обеспечение. Но для описания этого научного подхода Леонелли предпочитает термин «данноцентричный», а не «движимый данными», так как ведь практики данных почти постоянно переплетаются с иными компонентами исследования). В то же время существующие проявления *данноцентричной науки* имеют различные особенности. Например, обычно они связаны с возникновением масштабных, мультинациональных сетей ученых, которые настаивают на важности распределения данных и ценности исследовательских результатов самих по себе, с институционализацией процедур и норм для данных, распространяемых через движения «Открытой науки» и «Открытых данных».

По мнению Р. Китчина, технологии больших данных вводят многие научные дисциплины в новую эпоху эмпиризма и в связи с чем образуется раздел науки – *эмпирическая эпистемология* [Kitchin 2014⁶]. Во-первых, через большие данные можно исчерпывающе рассмотреть целую область науки и дать полную её характеристику. Во-вторых, технологии или системы программного обеспечения ориентируются на сбор опре-

деленных видов данных, а аналитики и алгоритмисты³ тестируют обоснованные научные рассуждения. В-третьих, данные не генерируются свободными от человеческих предубеждений. П. Гулд пишет по этому поводу: «Скучные данные не могут ничего сказать сами за себя, и мы всегда привносим некоторую концептуальную структуру, занимаясь их анализом или интерпретацией. Создание смысла данных всегда структурировано, они изучаются через особенные линзы, которые меняют то, что они интерпретируют. Даже если процесс является автоматическим, используемый алгоритм был создан в рамках особых ценностей и определённого научного подхода» [Gould 1981]. В-четвертых, если поддерживается идея, согласно которой данные говорят сами за себя, то предполагается, что кто-то, с пониманием статистики, может способствовать их интерпретации без контекста или доменно-специфичного знания.

Интересна идея о том, что вычисления и цифровые технологии управления данными создают возможности для автоматизации не только их производства, но и интерпретации [Allen 2001, 542]. В продолжение этой идеи В. Пич считает, что центральная особенность технологий больших данных связана с полной автоматизацией циклов научного процесса от сбора данных через работу над данными и построение моделей для новых прогнозов. *Наука больших данных* является первой подлинной машиной науки, в которой существенные процессы могут быть автоматизированы [Pietsch 2013].

Таким образом, применение технологий больших данных в качестве исследовательских практик предоставляет возможность реализации новой исследовательской парадигмы и разновидностей новых современных структур для множества дисциплин.

Институциональный вызов технологий больших данных. Среди вызовов развития технологий больших данных можно назвать институциональный, заключающийся в перемещении социальных исследований из одного типа института (к примеру, от департамента социальных наук) к другому (корпоративной исследовательской лаборатории). Об этом пишут в своих статьях М. Севидж, Р. Берроуз, Л. Адкинс, С. Лури, Р. Роджерс, рассуждая о том, к чему приведет дигитализация, приобщение социальной жизни к категории «цифровое», смещение возможностей исследования от департаментов социальных и гуманитарных наук к информационной индустрии [Savage, Burrows 2007; Adkins, Lury 2009; Rodgers 2009].

Сетевое цифровое пространство испытывает на себе влияние стэков (корпоративных технологических вертикально интегрированных социальных сетей, под которыми подразумевают компании, подобные Google, Apple, Facebook, Amazon, Microsoft). Все стэки создают подразделения, лаборатории, целью которых является анализ доступных им по роду деятельности больших объемов социальных цифровых данных. С одной стороны, они (стэки) предоставляют ученым больше эмпирических, аналитических, логических возможностей, но с другой, серьезно угрожают традиционным социальным исследованиям, концентрируя производство и анализ данных в нескольких исследовательских центрах лабораторий технологических компаний. Ситуация постепенно приводит к тому, что к настоящему времени вместо социологов новыми социальными экспертами чаще всего являются алгоритмисты и аналитики стэков, которые рассматривают пользователей Интернета как потенциальных покупателей (присутствует коммерческий интерес). Остается под вопросом экспертиза и критика подобных исследований. Хотя стэки поддерживают открытость, содействуя получению нового знания, но повторить эти исследования с некоммерческим интересом вне технологической платформы стэка весьма проблематично. Стэки оказывают влияние на то, как ученые создают, преобразуют и распространяют исследования и знания (устанавливают своеобразные правила игры).

Этический вызов технологий больших данных. Наука предполагает огромное количество возможностей по улучшению частной и публичной жизни. Тем не менее такое потенциально высокое позитивное влияние связано с важными этическими переменными. Экстенсивное использование все увеличивающегося количества данных (технологий их получения) увеличивает доверие к алгоритмам анализа и принятия решений (машинное обучение), а также постепенное уменьшение присмотра за многими авто-

матическими процессами поднимает важные вопросы о справедливости, ответственности и уважении к правам человека [The Ethics... web]. Идеология «данноизма» включает в себя доверие к (институциональным) агентам, которые собирают, интерпретируют и распределяют данные, полученные благодаря посредничеству интернет-платформ и иных коммуникационных технологий [Van Dijck 2014].

По некоторым прогнозам в ближайшей перспективе около половины всех этических нарушений в бизнесе будет связано с большими данными. Одной из проблем является прозрачность больших данных. Так как сенсоры становятся всепроникающими, многие защитники технологий больших данных считают, что коллекции данных служат документом того, что мир становится всё более прозрачным. Тем не менее значительная часть персональных данных собирается в агрегированные базы «за сценой», в основном без ведома пользователя. Эти базы данных могут быть сохранены в неизвестных и отдаленных облачных сервисах, где скрытые алгоритмы добывают информацию для заинтересованных клиентов [Richards, Jonathan 2013]. Детализация перспектив этики больших данных возможна через рассмотрение вопросов о соблюдении основных ценностей человека: приватности, конфиденциальности, прозрачности, идентичности, свободы.

Таким образом, активное использование технологий больших данных создаёт глубокие изменения на когнитивном, методологическом, эпистемологическом, институциональном и этическом уровнях, что проявляется в соответствующих вызовах, требующих для их решения участия ученых, представляющих различные области знания.

Примечания

¹ Обратное наблюдение — процедура контроля рядовых граждан за коммерческими и правительственными структурами с помощью современных информационно-коммуникационных технологий (в основном, видеокамер, подключенных к Интернету).

² Краудсорсинг — процесс выполнения производственных задач большим количеством независимых друг от друга людей, координирующих свою деятельность с помощью информационно-коммуникационных технологий.

³ Алгоритмисты — представители профессии нового вида — являются экспертами в сфере компьютерных наук, математики, статистики, а также в аспектах права, экономики, социальных исследований, анализе больших данных и прогнозирования.

Ссылки — References in Russian

Журавлева 2012 — *Журавлева Е.Ю.* Эпистемический статус цифровых данных в современных научных исследованиях // Вопросы философии. 2012. № 2. С. 113–123.

Майер-Шенбергер, Кукьер 2014 — *Майер-Шенбергер В., Кукьер К.* Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим. М.: Манн, Иванов и Фербер, 2014.

Calls Technologies ‘Big Data’ for the Modern Social Sciences and Humanities*

Elena Yu. Zhuravleva

Scientific interest in the emergence of new techniques for working with intensive data aka «Big Data» technologies generates a wide range of new research practices in almost every modern scientific discipline. The purpose of the article is to analyze the changes associated with the development of similar research practices in the social sciences and humanities. In the article presented two interrelated contexts for considering the «Big Data» technologies: industrial and scientific. And also various kinds of definitions of "Big Data" technologies are proposed: with a quantitative focus (dimensions, properties, volumes, structure or composition of data); with an emphasis on the process and cognitive-oriented. The definition of the «Big Data» technologies in the perspectives of studying the social movement, the behavior of people and the public nature of events is highlighted. In this group of definitions the important role is assigned to the derivative of «Big Data» concept of «datafication». The datafication is presented as a modern technological macrotrend on transformation of social actions into online quantitative data, interconnected with such modern technological macrorends as digitalization, sensorization and softwarization.

The main changes were found in the form of calls «Big Data» technologies for modern social sciences and humanities at the cognitive, epistemological, methodological, institutional and ethical level of knowing. Special attention is paid to such manifestations dataism in scientific activities as exploratory science, discovery science, hypothesis-free science, data driver science, datacentric science, empiricist epistemology and science Big Data.

KEYWORDS: technology «Big Data», datafication, sensorization, softwarization, dataism, exploratory science, discovery science, hypothesis-free science, data driver science, datacentric science, empiricist epistemology, science Big Data.

ZHURAVLEVA Elena Yu. – CSc in Philosophy, associate professor, Russian Academy of National Economy and Public Administration under the President of the Russian Federation, Vologda, Russia.

ereseach7@gmail.com

Received at August, 21, 2016.

Citation: Zhuravleva, Elena Yu. (2018) “Calls Technologies ‘Big Data’ for the Modern Social Sciences and Humanities”, *Voprosy Filosofii*, Vol. 9 (2018), pp. 50–59.

DOI: 10.31857/S004287440001353-3

References

Adkins, Lisa, Lury, Celia (2009) ‘What is the empirical?’, *European Journal of Social Theory, special issue*, № 12, pp. 5–20.

Allen, John F. (2001) ‘In Silico Veritas. Data-Mining and Automated Discovery: The Truth Is in There’, *EMBO Reports*, № 2, pp. 542–544.

Anderson, Chris (2008) ‘The end of theory: The data deluge makes the scientific method obsolete’, *Wired*, Vol. 16 (7), pp. 108–109.

Aronova, Elena, Baker, Karen S., Oreskes, Naomi (2010) ‘Big science and big data in biology: From the International Geophysical Year through the International Biological Program to the Long Term Ecological Research (LTER) Network, 1957– Present’, *Historical Studies in the Natural Sciences*, 40, pp. 183–224.

* This research was supported by the Russian Foundation for Basic Research via grant № 16-03-50064 «Softwarization modern research activities: epistemological foundations».

- Bowker, Geoffrey C. (2006) *Memory Practices in the Sciences*, MIT Press, Cambridge, MA.
- Brey, Philip, Søraaker, Johny H. (2009) 'Philosophy of Computing and Information Technology', *Philosophy of Technology and Engineering Sciences. Vol. 14 of the Handbook for Philosophy of Science*, ed. A. Meijer, gen. ed. D. Gabbay, P. Thagard and J. Woods, Elsevier.
- Brooks, David (2013) 'The philosophy of data', *New York Times*, 04.02.2013 <http://www.nytimes.com/2013/02/05/opinion/brooks-the-philosophy-of-data.html>
- Clough, Patricia, Ticineto, Karen, Gregory, Benjamin, Haber, Scannell, Joshua (2014) *The Datalogical Turn in Nonrepresentational Methodologies: Re-Envisioning Research*, ed. Phillip Vannini, Taylor & Francis, Oxford, pp. 182–206.
- Floridi, Luciano (2012) 'Big Data and their epistemological challenge', *Philosophy and Technology*, 25 (4), pp. 435–437.
- Foster, Allan (2016) 'Initiatives', *Business Information Review*, № 33, pp. 54–69.
- Franklin, Laura R. (2005) 'Exploratory Experiments', *Philosophy of Science*, № 72, pp. 888–899.
- Gould, Peter (1981) 'Letting the data speak for themselves', *Annals of the Association of American Geographers*, № 71 (2), pp. 166–76.
- Hallam, Stevens (2013) *Life Out of Sequence: A Data-Driven History of Bioinformatics*, The University of Chicago Press, Chicago and London.
- Hunter, David J., Kraft, Peter (2007) 'Drinking from the fire hose: Statistical issues in genome wide association studies', *N Engl J Med*, № 357, pp. 436–439.
- Ioannidis, John P.A., Ntzani, Evangelia E., Trikalinos, Thomas A., Contopoulos-Ioannidis, Despina G. (2001) 'Replication validity of genetic association studies', *Nat Genet*, № 29, pp. 306–309.
- Kitchin, Rob (2013) *Review of Software Takes Command by Lev Manovich (Bloomsbury, 2013)*, https://www.goodreads.com/author_blog_posts/6010103-review-of-software-takes-command-by-lev-manovich-bloomsbury-2013.
- Kitchin, Rob (2014^a) *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*, Sage, London.
- Kitchin, Rob (2014^b) 'Big data, new epistemologies and paradigm shifts', *Big Data and Society*, № 1, pp. 1–12.
- Leonelli, Sabina (2014) 'Data Interpretation in the Digital Age', *Perspectives on Science*, № 22 (3), pp. 397–417.
- Lingen, Frank Van (2013) *A Software Aware Society Driven by Sensors, Analytics and APIs*, 21.11.2013 <http://blogs.cisco.com/sp/a-software-aware-society-driven-by-sensors-analytics-and-apis-2/>.
- Mayer-Schönberger, Viktor, Cukier, Kenneth (2013) *Big Data. A Revolution That Will Transform How We Live, Work, and Think*, Houghton Mifflin, Harcourt (Russian Translation).
- Pietsch, Wolfgang (2013) 'Big Data – The New Science of Complexity', *6th Munich-Sydney-Tilburg Conference on Models and Decisions (Munich; 10-12 April 2013)*, <http://philsci-archival.pitt.edu/9944/>
- Rheinberger, Hans-Jörg (2011) 'Infra-experimentality: From Traces to Data, from Data to Patterning Facts', *History of Science*, № 49 (3), pp. 337–348.
- Richards, Neil M., Jonathan, H. King (2013) 'Three Paradoxes of Big Data', *Stanford Law Review Online*, № 41, pp. 41–46.
- Rodgers, Richard (2009) *The End of the Virtual: Digital Methods, Media*, Amsterdam University Press, pp. 1–25.
- Ruppert, Evelyn (2013) 'Rethinking Empirical Social Sciences', *Dialogues in Human Geography*, № 3 (3), pp. 268–273.
- Ruppert, Evelyn, Law, John, Savage, Mike (2013) 'Reassembling Social Science Methods: The Challenge of Digital Devices', *Theory, Culture & Society, Special Issue on The Social Life of Methods*, № 30 (4), pp. 22–46.
- Savage, Mike, Burrows, Roger (2007) 'The Coming Crisis of Empirical Sociology', *Sociology*, № 41, pp. 885–899.
- Szalay, Alexander, Gray, Jim (2006) 2020 'Computing: Science in an exponential world', *Nature*, Vol. 440, pp. 413–414.
- Hey, Tony, Tansley, Stewart, Tolle, Kristin ed. (2009) *The fourth paradigm: data-intensive scientific discover*, Microsoft Research, Redmond, Washington.
- Dijk, José van (2014) 'Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology', *Washington Historical Quarterly*, № 2 (12), pp. 197–208.
- Weinberger, David (2011) *Too big to know: Rethinking knowledge now that the facts aren't the facts, experts are everywhere, and the smartest person in the room is the room*, Basic Books, New York.
- Zhuravleva, Elena Yu. (2012) 'Epistemic status of digital data in modern scientific research', *Voprosy Filosofii*, Vol. 2 (2012), pp. 113–123 (in Russian).