

---

---

## ПЕРЕВОДЫ

Р. Китчин  
(Мейнута, Ирландия)

### БОЛЬШИЕ ДАННЫЕ, НОВЫЕ ЭПИСТЕМОЛОГИИ И СМЕНА ПАРАДИГМ<sup>1</sup>

Рассматриваются вопросы: как доступность больших данных в сочетании с новыми подходами к анализу данных ставят под сомнение традиционную эпистемологию применительно к естественным, социальным и гуманитарным наукам? В какой степени данные изменения выступают источником смены парадигм в различных научных дисциплинах? В частности, критически проанализированы новые формы эмпиризма, в рамках которых провозглашаются «конец теории», развитие науки, «идущей от данных», а не «от знания», а также цифровых гуманитарных и вычислительных социальных наук, предлагающих качественно новые способы осмысления культуры, истории, экономики и общества. Показано, во-первых, что большие данные и новые подходы к анализу данных представляют собой инновации, во многих отношениях перестраивающие привычный порядок проведения научного исследования. Во-вторых, существует острая необходимость в масштабном критическом осмыслении научным сообществом последствий разворачивающейся революции данных для эпистемологии. Едва ли можно отметить попытки решения данной задачи, несмотря на бурные изменения, происходящие в исследовательской практике. На ос-

---

**Роб Китчин** – PhD, профессор, Ирландский национальный университет в Мейнута. E-mail: Rob.Kitchin@nuim.ie.

*Переводчик:* **Ольга Николаевна Шаева** – кандидат социологических наук, старший преподаватель кафедры социологии Московского государственного лингвистического университета. E-mail: shaeva.olga@gmail.com.

<sup>1</sup> Перевод выполнен по: *Kitchin R. Big Data, New Epistemologies and Paradigm Shifts // Big Data & Society. 2014. Vol. 1. Iss. 1. P. 1–12*, и представлен на конкурсе, объявленный журналом «Социология: методология, методы, математическое моделирование» в 2016 г.

нове критического обзора возникающих в настоящее время эпистемологических оснований науки утверждается, что потенциально продуктивным подходом должно стать развитие эпистемологии, учитывающей специфику рассматриваемого явления в его контексте.

*Ключевые слова:* большие данные, анализ данных, эпистемология, парадигмы, конец теории, наука, «идущая от данных», цифровые гуманитарные науки, вычислительные социальные науки.

## *Введение*

Научным революциям часто предшествовали революции в методах измерения. Синан Эрал [15]

Большие данные приводят к радикальному сдвигу в нашем понимании научных исследований... [Они вызывают] качественные изменения на уровнях эпистемологии и этики. Большие данные по-новому ставят вопрос, что такое знание, об исследовательском процессе, как мы должны работать с информацией, а также о природе и способах упорядочения реальности... Большие данные устанавливают новые области объектов исследования, новые методы познания и новые определения социальной реальности [7].

Как это часто случается с быстро развивающимися понятиями, существует множество определений и подходов к операционализации больших данных. Они варьируются от банального объяснения больших данных как массивов данных, объем которых превышает возможности хранения на листе Excel или на одном устройстве [58], до более тщательно проработанных оценок сущности понятия, выделяющих его имманентные свойства [7; 41]. Опираясь на анализ значительного объема литературы по теме, автор [29] выделяет следующие свойства больших данных:

– огромный *объем* – они включают терабайты и петабайты данных;

- высокая *скорость* производства – они создаются в режиме реального времени или с незначительной задержкой;
- внутреннее *многообразие* – они по своей природе могут быть как структурированными, так и неструктурированными;
- *исчерпывающая* полнота – целью является охват всей генеральной совокупности или всего объема объекта (*n* – все наблюдения);
- высокий уровень *детализации* данных и уникальная *индексичность* при идентификации;
- *взаимосвязь* с другими массивами данных – они включают одинаковые поля, делающие возможным объединение различных наборов данных;
- *гибкость* – они обладают свойством *расширяемости* (т.е. к ним с легкостью могут быть добавлены новые поля) и *масштабируемости* (т.е. они могут быстро увеличиваться в объеме) [7; 17; 32; 40; 41; 61].

Иными словами, большие данные определяются не только объемом. Промышленность, органы власти и научное сообщество уже длительное время производят значительные по объему базы данных, например, содержащие результаты проведения переписей населения. Однако с учетом того, что создание, обработка, анализ и хранение таких данных требуют больших затрат и сопряжены с рядом сложностей, их сбор осуществлялся строго определенным образом с применением выборочного метода, который ограничивал объем, охватываемый период времени и размер базы данных [42]. Чтобы сделать процесс сбора данных переписи управляемым, их собирают раз в пять или десять лет, в них включают ответы на ограниченное количество вопросов – от 30 до 40, а их результаты имеют достаточно низкую разрешающую способность (касаются, например, районов или других единиц административно-территориального деления, а не отдельных людей или домохозяйств). Более того, методы, используемые при сборе данных, не подразумевают возмож-

ности какой-либо корректировки. Например, если перепись уже проводится по стандартной методике, практически невозможно изменять, добавлять или убирать из нее вопросы. В то время как при проведении переписей стремятся к всеохватности, учету всех людей, живущих в стране, при использовании опросного и других методов сбора данных используется выборочный метод, цель которого – репрезентация генеральной совокупности.

В отличие от этого, большие данные характеризуются тем, что они непрерывно собираются, стремлением к всеохватности и высокой детализации данных, а также гибкостью и возможностью изменения характеристик базы в процессе сбора данных. Примерами средств сбора подобных данных выступают цифровые системы видеонаблюдения, регистрация покупок в розничных сетях, цифровые устройства, регистрирующие и сохраняющие историю их использования (например, мобильные телефоны), регистрация транзакций и интеракций в цифровых сетях связи (например, электронной почте или интернет-банкинге), история посещений сайта или использования приложения, данные датчиков в различных предметах и объектах инфраструктуры, сканирование машиночитаемых объектов, таких как транспортные билеты или штрих-коды, фиксация активности пользователей в социальных сетях [30]. Все это создает крупные и при этом динамичные потоки разнообразных детализированных данных, между которыми можно установить взаимосвязь. Так, например, в 2012 г. компания Wal-Mart произвела более 2,5 петабайт ( $2^{50}$  байт) данных, содержащих информацию о более чем 1 миллионе операций покупателей *в час* [44], а Facebook сообщает, что *каждый день* в этой социальной сети обрабатываются 2,5 миллиарда объектов (ссылок, комментариев и т.д.), 2,7 миллиардов «лайков» и 300 миллионов загружаемых фотографий [12]. Обработка и анализ подобных данных представляют собой задачу, кардинально отличающуюся от работы с данными переписи, проводимой раз в 10 лет, или результатами опроса нескольких сотен респондентов.

Большие данные и ранее уже производились в некоторых сферах, таких как дистанционный сбор данных, прогнозирование погоды и финансовые рынки. Тем не менее такие технологические прорывы как повсеместное проникновение компьютерных технологий, широкое распространение межсетевое взаимодействия, новые структуры баз данных и способы хранения данных стали переломным моментом при переходе к непрерывному производству и анализу больших данных, что не в последнюю очередь подразумевает развитие новых способов анализа данных в условиях их обилия [30]. Традиционные методики анализа данных предназначены для получения выводов на основе редких, статичных, прошедших предварительную обработку и практически не связанных друг с другом баз данных, полученных в результате проведения выборочных исследований с конкретной целью и строго соответствующих определенным требованиям, таким как независимость, стационарность и нормальность распределения. Сбор и анализ этих данных проводился с оглядкой на специфический вопрос [42]. Основным вызов при анализе больших данных – это необходимость справиться с избыточным объемом, всеохватностью и многообразием, непрерывностью и динамизмом, неупорядоченностью и неопределенностью, высокой связанностью с другими данными. Кроме того, следует учитывать, что зачастую их производство не связано с каким-то вопросом, или оно оказывается побочным продуктом иной деятельности. До недавнего времени задача работы с такими данными представлялась слишком сложной и не могла быть реализована; сегодня она стала осуществимой благодаря росту вычислительной мощности компьютеров и появлению новых аналитических методик. Эти новые методики разрабатываются на основе исследований в сфере искусственного интеллекта и экспертных систем, направленных на развитие машинного обучения, позволяющего на основе автоматических вычислительных процедур производить интеллектуальный анализ данных и распознавать в них закономерности, строить прогнозные модели и

повышать их точность [21; 24]. Отметим, что различные модели обладают собственными достоинствами и недостатками, поэтому часто сложно предположить, какая модель окажется наиболее адекватной для имеющейся базы данных. В связи с этим может быть применен комплексный подход, позволяющий получить множественные решения [54]. Для получения лучшей модели или наиболее удачного объяснения при анализе базы данных могут быть применены буквально сотни различных алгоритмов [55]. Такой подход качественно отличается от традиционного, когда исследователь выбирает подходящий метод исходя из своего понимания различных подходов и информации о базе данных. Иными словами, анализ больших данных задействует совершенно новую эпистемологическую установку в познании мира: новый подход к анализу данных ориентирован на получение выводов, «проистекающих из данных», а не на проверку теории на основе анализа подходящих для этого данных.

Взрывной рост производства больших данных в сочетании с развитием новой эпистемологии привели к распространению точки зрения, согласно которой в настоящее время происходит революция в анализе данных, которая будет иметь далеко идущие последствия для производства знания, организации бизнеса и управления [2; 6; 19; 41]. В том, что касается производства знания, утверждается, что большие данные делают возможным развитие новой исследовательской парадигмы, охватывающей ряд дисциплин. Как указывает Кун [31], парадигма представляет собой способ постановки исследовательских вопросов и производства знания, разделяемый значительной частью исследователей в рамках какой-либо дисциплины в определенный период времени. Как утверждает Кун, периодически возникает новый способ мышления, который бросает вызов принятым теориям и подходам. Например, теория эволюции Дарвина вызвала качественные изменения в концептуальной основе биологии, а также поставила под сомнение религиозную доктрину креационизма. Джим Грей

Таблица 1

ЧЕТЫРЕ ПАРАДИГМЫ НАУКИ

Парадигма	Сущность	Форма	Период
Первая	Экспериментальная наука	Эмпиризм; описание природных явлений	Вплоть до Ренессанса
Вторая	Теоретическая наука	Построение объяснительных моделей и обобщение	До появления компьютеров
Третья	Вычислительная наука	Имитационное моделирование сложных явлений	До появления больших данных
Четвертая	Разведывательная наука	Переработка большого объема данных; статистический и интеллектуальный анализ данных	В настоящее время

Источник: составлено по работе Т. Хей [25]

[25] представляет эволюцию науки как смену четырех базовых парадигм (*табл. 1*).

Кун полагает, что смены парадигм происходят тогда, когда наука в ее текущем состоянии не может объяснить определенные явления или ответить на наиболее значимые вопросы, что создает потребность в формулировании новых идей. Грей, в отличие от него, связывает смены парадигмы с появлением новых форм данных и методик их анализа. Таким образом, он указывает на переход науки к четвертой парадигме в связи с возрастающей доступностью больших данных и развитием новых методов анализа.

Предложенный Куном подход подвергается значительной критике, не в самую последнюю очередь за то, что в некоторых областях научного знания не обнаруживаются признаков существования парадигм. Это особенно характерно для некоторых областей социальных наук, где принят спектр разнообразных философских оснований, например, в социально-экономической географии и социологии. В то же время в других областях научного знания, таких как естественные науки, сложилось более выраженное эпистемологическое единство относительно того, как проводятся научные исследования – с использованием четко определенного научного метода, на основе проверки гипотез с целью подтверждения или опровержения теории. Кроме того, объяснение через парадигмы дает упрощенную и линейную картину развития научных дисциплин, очищенную от всех проявлений неупорядоченности, многообразия, внутренних противоречий и несогласия, характерных для реального развития науки. Хотя само понятие парадигмы неоднозначно, оно хорошо подходит для терминологического структурирования современных дискуссий, касающихся развития больших данных и его последствий. Это связано с тем, что многие из высказываемых о производстве знания суждений основываются на том, что в настоящее время формируется принципиально новая эпистемология, что мы на-



ходимся на этапе перехода к новой парадигме. Однако вопрос, какую форму примет эта новая эпистемология, вызывает споры. Далее в статье производится критический обзор развивающейся в настоящее время четвертой научной парадигмы и ее форм, а также исследуется вопрос, в какой мере революция данных приводит к новой, отличающейся от традиционной, эпистемологии в гуманитарных и социальных науках и изменению практик проведения исследований.

### *Четвертая парадигма в науке?*

Джим Грей предполагает, что четвертая парадигма науки будет связана с интенсивным анализом большого объема данных и появлением принципиально новых форм традиционного научного метода. Другие авторы, однако, утверждают, что большие данные открывают новую эпоху эмпиризма, где большие объемы данных в сочетании с методиками, позволяющими обнаружить содержащиеся в них знания, приводят к тому, что сами данные начнут «говорить» о себе без какой-либо теории. Идеи эмпиризма завоевали признание за пределами научного сообщества, особенно в бизнесе, а также укоренились в новых областях «науки о данных» (*data science*) и других дисциплинах. Одновременно в традиционных научных дисциплинах возникает новая форма исследований – наука, «идушая от данных». В этой части статьи приведен критический анализ эпистемологических оснований обоих направлений; при этом учтены различия источников и целей научных исследований в бизнесе и научном сообществе. Первые заинтересованы в применении анализа данных в целях разработки новых продуктов, анализа рынков и поиска возможностей продвижения, а не в расширении знания как таковом, в то время как вторые нацелены на достижение лучшего понимания мира и поиск объяснений явлений и процессов.

Конец теории: новое рождение эмпиризма

Некоторые авторы, такие как Крис Андерсон, бывший главный редактор журнала *Wired*, считают, что большие данные и новые комплексные подходы к анализу данных указывают на наступление новой эпохи в производстве знания, характеризующейся «концом теории». В своей вызвавшей много споров работе 2008 г. Андерсон утверждает, что «в условиях мощного потока данных существующий научный метод становится неактуальным», что закономерности и взаимосвязи, содержащиеся в больших данных, сами по себе дают глубокие и практически полезные знания о сложных явлениях. Указывая на принципиальную способность больших данных привести к эмпиристскому производству знания, он утверждает:

Теперь появилась новая возможность. Петабайты данных позволяют нам сказать: «Корреляции достаточно»... Мы можем анализировать данные, не формулируя гипотезы о том, что в них содержится. Мы можем загрузить цифры в самые большие компьютерные кластеры, которые когда-либо существовали, и позволить статистическим алгоритмам обнаружить закономерности, которые не может выявить теория... Корреляция вытесняет каузальность, и наука может идти вперед даже без целостных моделей, общих теорий, вообще какого-либо механистического объяснения. Нет никаких оснований придерживаться старых подходов.

Пренски [47] утверждает практически то же самое:

...ученым больше нет необходимости делать обоснованные предположения, формулировать гипотезы и конструировать модели, а затем проверять их на основе экспериментальных данных или примеров. Вместо этого они могут углубляться в полный объем данных для того, чтобы обнаруживать там закономерности, лежащие в основе наблюдаемых явлений, и получать научно обоснованные выводы без необходимости проводить экспериментальные исследования.

Дюше [18] утверждает, что «добыча данных (*data mining*) обнаруживает такие закономерности и взаимосвязи, которые мы и не подумали бы искать». Стедмен [56] также указывает:

Характерный для больших данных подход к поиску знаний позволяет аналитику получить картину происходящего в мире «в высоком разрешении». Ничего не будет упущено, если он сосредоточится на каких-либо конкретных деталях в данных. Он также ничего не потеряет, если попытается охватить ситуацию в самой широкой перспективе, в связи с чем детали не будут учтены... Аналитику даже больше не нужно давать себе труд формулировать гипотезы.

Примеры, иллюстрирующие данную точку зрения, как правило, происходят из сфер маркетинга и розничной торговли. Скажем, Дюше [18] подробно рассматривает пример анализа данных о покупках, совершенных в определенной торговой сети за 12 лет. Анализ был направлен на выявление неизвестных взаимосвязей, определяющих набор товаров в корзине покупателя. Обнаружение корреляций между определенными товарами привело к разработке новой схемы выкладки товаров и увеличению дохода с каждой корзины на 16% в течение первого месяца. Изначально не было выдвинуто гипотез о том, что продукт А часто покупают вместе с продуктом Б, которые затем могли бы быть проверены. Был проведен анализ данных, направленный на поиск существующих взаимосвязей, некоторые из которых ранее могли остаться незамеченными. Схожим образом при использовании сайта Amazon.com система генерирует для покупателя рекомендации других товаров, в которых он может быть заинтересован. Она не использует информацию о культуре и нормах чтения книг, а просто выявляет закономерности в поведении покупателей и определяет, что если покупателю А нравится книга X, то ему понравится книга Y, поскольку таков паттерн, характерный для него и других покупателей. Хотя объяснение причин обнаруженных в данных взаимосвязей и их значения и может оказаться полезным, подобные объяснения отбрасываются по большей части за ненадобностью. В связи с этим Сигель [55, р. 90] утверждает, говоря о прогнозной аналитике: «Нам обычно не известны причинно-следственные связи, и обычно они не

важны для нас... Целью скорее выступает прогнозирование, а не объяснение мира... Важен практический результат; прогноз имеет преимущество перед объяснением».

Именно такая аргументация используется при продаже некоторых видов программного обеспечения для анализа данных. Например, программное обеспечение для глубинного анализа данных (*data mining*) и визуализации Ayasdi, как заявляют разработчики, способно:

...автоматически обнаруживать знания – независимо от уровня сложности – без необходимости задавать вопросы. Пользователи Ayasdi наконец-то могут получить ответы на вопросы, которые им бы не пришло в голову задать. Проще говоря, Ayasdi – это счастливая способность к случайным открытиям в цифровой форме [10].

Кроме того, разработчики утверждают, что они смогли полностью исключить:

...человеческий элемент из анализа данных, а вместе с ним и сопутствующую необъективность. Система не ждет, пока ей зададут вопрос или нацелят ее на анализ конкретных взаимосвязей в данных, а вместо этого самостоятельно выявит закономерности, которые пользователь мог и не подумать искать [10].

Лежащая в основе эмпиризма эпистемология опирается на комплекс следующих популярных и достаточно убедительных идей, вступающих в противоречие с дедуктивным подходом, господствующим в современной науке:

– большие данные могут охватить всю интересующую исследователя область и при этом дать ее детальную картину;

– отсутствует необходимость в априорной теории, моделях или гипотезах;

– применение методов, свободных от исходных предположений, позволяет данным самим «говорить» о себе, не завися от предубеждений и моделей мышления человека. Любые закономерности и взаимосвязи в больших данных по сути своей значимы, содержательны и достоверны;

– содержащиеся в данных знания выходят за пределы конкретного контекста или специфики какой-либо области и поэтому могут быть интерпретированы любым человеком, который в состоянии разобраться в статистических показателях или визуализациях.

В совокупности все это позволяет выдвинуть предположение, что формируется новый образ науки, *modus operandi* которой чисто индуктивен по своей природе.

Хотя эпистемология эмпиризма может быть привлекательной, все перечисленные идеи, служащие ее основанием, ошибочны. Во-первых, хотя большие данные и стремятся исчерпывающе охватить всю интересующую область и при этом дать ее детальную картину, они все же дают ее выборочное отображение. Это определяется используемыми технологиями и платформами сбора данных, онтологическими установками в отношении данных и нормативно-правовой средой их сбора, в связи с чем и в данном отображении могут проявляться ошибки выборки [14; 29]. Действительно, любые данные предлагают избирательный взгляд на мир – взгляд с определенного ракурса, полученный с использованием определенных инструментов, а не всевидящий непогрешимый взгляд Бога [1; 22]. По своей сути данные – не естественные и сущностные элементы мира, извлекаемые из него нейтральным и объективным способом, и поэтому принятые «как есть». Данные создаются внутри сложных ансамблей, чьи элементы и определяют содержание данных [49].

Во-вторых, большие данные не появляются из ниоткуда, независимо от «регулирующего воздействия» философии [4, р. 8]. Напротив, системы сбора данных спроектированы так, чтобы улавливать определенные виды данных, а методы и алгоритмы анализа построены на основе принятой научной логики и прошли апробацию в научных исследованиях. Индуктивная стратегия поиска закономерностей в данных не существует сама по себе, не реализуется в научном вакууме, а уточняется в контексте результатов проведенных ранее исследований и существующих теорий,

на основе рассуждений, проистекающих из имеющегося опыта и знаний [36]. Новые методы анализа данных могут создавать впечатление автоматического получения знаний без предварительного формулирования вопросов исследования, но используемые при этом алгоритмы, несомненно, научно обоснованы и прошли проверку на валидность и достоверность.

В-третьих, данные не просто не могут быть собраны независимо от теории, но они также не способны и «говорить» сами за себя, без влияния предубеждений и моделей мышления человека. Как отмечает Гоулд, «неодушевленные данные не могут говорить сами за себя, мы всегда привносим в исследование, анализ и интерпретацию данных какую-либо концептуальную модель, будь она интуитивной и слабо проработанной или четко структурированной и формализованной» [20, р. 166]. Осмысление данных всегда происходит в рамках какой-либо системы координат – исследователь смотрит на данные с определенного ракурса, и это влияет на то, как они интерпретируются. Даже если процесс автоматизирован, алгоритмы обработки данных насыщены определенными ценностями и реализуются в контексте какого-либо научного подхода. Далее, обнаруженные в данных закономерности не имеют внутренне присущего им смысла. Корреляции между переменными в базе данных могут иметь случайную природу и не подразумевать причинно-следственных связей, и подобная интерпретация корреляций может приводить к серьезным угрозам экологической валидности. В случае больших данных указанные сложности могут усугубиться, так как принципы эмпиризма, по-видимому, поощряют практику «зачерпывания» данных без разбора – поиска каждой существующей взаимосвязи или модели.

В-четвертых, положение о том, что данные могут «говорить» сами за себя, подразумевает, что любой человек с достаточным знанием статистики должен быть способен интерпретировать данные без понимания контекста и без наличия специфических для конкретной области знаний. Подобные самонадеянные за-

явления делают некоторые специалисты по анализу данных и компьютерным наукам, а также другие исследователи, например, физики, активно включившиеся в исследования в сфере социальных и гуманитарных наук. Так, несколько физиков обратили свое внимание на города и провели анализ больших данных с целью моделирования социальных и пространственных процессов в них, а также выявления предполагаемых законов, определяющих их формирование и функционирование [5; 35]. Эти исследования часто умышленно игнорируют два столетия развития социальных наук, включая почти век количественного анализа и построения моделей. Результатом становится редукционистский и функционалистский анализ, упрощающий исследуемое явление и не принимающий во внимание влияние культуры, политических событий, общей стратегии управления и капиталов. Оказываются воспроизведенными все те ограничения, которые были характерны для количественного позитивистского подхода в социальных науках в середине XX в. Похожие опасения высказывают и представители естественных наук. Страссер [57], например, отмечает применительно к биологическим наукам, что биоинформатики, имеющие очень узкое и специфическое понимание биологии, заявляют права на сферы, традиционно занятые клиницистами, молекулярными биологами и биологами, занимающимися экспериментальными исследованиями. Эти биоинформатики несомненно игнорируют наблюдения, сделанные Порвеем [46]:

Если у вас нет возможности обратиться к специалистам в конкретной области, чтобы они предварительно сформулировали проблемы исследования, вы получаете неудовлетворительные результаты... Профильные специалисты вдвойне необходимы при оценке результатов работы, особенно если вы работаете со специфическими данными – данными о человеческом поведении, содержащими сведения личного характера. Как специалисты по обработке и анализу данных, мы обладаем достаточной квалификацией, чтобы ответить на вопросы «Что?», «Как?» по отношению

к данным, но нам не следует касаться вопроса «Почему?» в тех областях, в которых мы не специалисты.

Проще говоря, если данные и могут быть интерпретированы вне контекста и без наличия специфических для конкретной области знаний, построенная на подобной эпистемологии интерпретация, скорее всего будет неполной или не проясняющей сути дела, так как она остается в стороне от контекста текущих дискуссий и накопленных знаний.

Указанные ошибочные представления получили некоторое признание, особенно в бизнес-кругах, так как они предлагают удобный нарратив для устремлений бизнес-организаций, продающих услуги в сфере производства знания, например, компаний, занимающиеся сбором и продажей персональных данных, поставщиков услуг в сфере анализа данных, разработчиков программного обеспечения, консалтинговых компаний. В рамках эмпиристского подхода анализ данных дает возможность получения значимых, достоверных и имеющих практическую ценность знаний вне науки и без участия ученых, а значит, и сопутствующих излишних трат, непредвиденных расходов и необходимости формулирования объяснений и поиска истины. В этом смысле методики обработки и анализа данных могут действительно иметь высокую значимость для практиков, однако артикуляция принципов нового эмпиризма становится дискурсивным риторическим устройством, позволяющим упростить более сложный эпистемологический подход и убедить организации на рынке в значимости и практической полезности анализа больших данных.

### *Наука. «идущая от данных»*

В отличие от новых форм эмпиризма, наука, «идущая от данных», ориентирована на необходимость придерживаться основ существующей научной методологии. При этом она более открыта по отношению к гибридным сочетаниям абдуктивного,



индуктивного и дедуктивного подходов с целью углубления понимания исследуемого явления. Она отличается от традиционного, экспериментального и дедуктивного формата исследования тем, что выдвижение гипотез и получение выводов осуществляется исходя из данных, а не теории [27, р. 613]. Иными словами, она стремится интегрировать индуктивную логику в формат исследования, хотя индуктивное объяснение не ставится конечной целью, как в случае эмпиризма. Вместо этого она создает новый способ разработки гипотез, предшествующий применению дедуктивного подхода. Процесс индукции в исследовании также не возникает из ниоткуда, он локализован и контекстуализируется в рамках тщательно разработанной теоретической предметной области. Таким образом, эпистемологическая стратегия в науке, «идущей от данных», подразумевает применение направленных методик поиска знаний для определения возможных вопросов (гипотез), представляющих интерес для дальнейшей разработки и проверки.

Процесс обозначен как направленный в том смысле, что существующая теория управляет процессом поиска и обнаружения знаний. Речь не идет о простой надежде выявить все существующие взаимосвязи в базе данных, предполагая, что они в каком-либо смысле значимы. В сущности, то, как данные создаются или используются в новых целях, определяется некоторыми допущениями, основанными на теоретическом знании и практическом опыте. Эти допущения помогают понять, позволят ли используемые технологии и их сочетания получить соответствующий цели исследования и полезный материал. Сбор данных не осуществляется всеми возможными способами, с использованием каждой доступной технологии или всех возможных вариантов выборки. Напротив, способ сбора данных и их адаптации к различным исследовательским задачам тщательно продумывается; принимаются стратегические решения собирать специфические данные и игнорировать другие. Аналогичным образом обработка, использование и анализ полученных данных определяются представлением о том, какие

именно методики позволят получить значимые и содержательные выводы. К данным не применяют каждый возможный способ их осмысления, обработки и анализа с надеждой, что они раскроют какую-либо скрытую в них истину. Напротив, исследователи принимают теоретически обоснованные решения о том, какой способ работы с базой данных позволит выявить информацию, представляющую интерес для дальнейших исследований. Вместо того чтобы определять, какую истину несет каждая из обнаруженных взаимосвязей, исследователи фокусируют внимание на некоторых из них, отобранных по определенным критериям, как предполагается, наиболее значимых для продвижения исследования. Действительно, многие предполагаемые взаимосвязи в базе данных могут сразу же быть отброшены специалистом в конкретной области как не представляющие большого интереса или бессмысленные, в то время как другие будут отмечены как заслуживающие внимания [42].

Подобный способ принятия решений относительно методов сбора и анализа данных основан на абдуктивной логике. Абдукция – способ построения логических умозаключений, введенный в активное использование Ч.С. Пирсом (1839–1914) [42]. Ее цель – построение обоснованного и разумного заключения, которое не обязательно окончательно и не подлежит сомнению. Например, предпринимается попытка не вывести лучший способ сбора данных, а скорее найти такой подход, который представляется разумным и обоснованным в свете уже известного о способах сбора данных. Абдукция очень часто применяется в науке, особенно при выдвижении гипотез, хотя это и редко признается. В таком случае взаимосвязи, выявленные в данных, не возникают из ниоткуда и не «говорят» сами за себя. Процесс индукции, при котором догадки возникают из данных, осуществляется в контексте. Но эти догадки – не конечный пункт исследования, будучи объединенными и выстроенными в теорию. Они скорее выступают основой для формулирования гипотез, которые затем будут дедуктивно

проверены. Иными словами, наука, «идущая от данных» – модификация традиционного научного метода, вводящая новый способ построения теорий. Тем не менее при этом отмечается значительное изменение эпистемологии.

Некоторые авторы утверждают, что скорее наука, «идущая от данных», а не новый эмпиризм или конец теории, станет новой парадигмой научного метода в эпоху больших данных, так как характерная для нее эпистемология позволяет получить новые ценные выводы, которые не могли бы быть получены традиционной наукой, «идущей от знания» [27; 37; 42]. Наука, «идущая от знания», использующая прямой дедуктивный подход, хорошо позволяет понять и объяснить мир в условиях малого объема данных и малых вычислительных мощностей компьютеров. Однако неразумно придерживаться такого подхода в условиях, когда развитие технологий и методологии дает возможность проводить более глубокий анализ данных, а именно применять новые методики анализа данных, объединять большие и разнородные массивы данных так, как ранее не было возможно. Это позволяет получать новые ценные данные, ставить новые интересные вопросы и решать их необычными способами. Более того, сторонники науки, «идущей от данных», утверждают, что она более адекватна задаче изучения, анализа и извлечения смысла из больших, связанных друг с другом баз данных. Она также способствует развитию междисциплинарных исследований, интегрирующих знания из различных предметных областей, так как она в меньшей степени ограничена исходными теоретическими основами исследования. Это может привести к разработке более целостных и масштабных моделей и теорий, а не их отдельных элементов [27].

Так, например, высказывается мнение о том, что наука, «идущая от данных», изменит наше понимание окружающей среды [9; 34]. Станет возможным получение детальных данных, зачастую в режиме реального времени, из разнообразных источников, таких как традиционные и мобильные метеостанции, спутниковая и

аэро съемка, метеорологические радиолокаторы и гидрометрические станции, наблюдения, сделанные населением, наземные лидары и лидарная аэро съемка, исследования качества воды, газа и почвы, а также распределенные датчики, измеряющие отдельные показатели, такие как температура и влажность воздуха. Эти данные смогут быть объединены в рамках создания детализированных моделей окружающей среды в ее динамике, в отличие от точечных измерений в конкретном месте и времени. Они также позволят выявлять взаимосвязи между конкретными явлениями и процессами, формулировать на этой основе новые гипотезы и теории, достоверность которых затем может быть проверена. Это также поможет выявить и лучше понять взаимосвязи между различными сферами окружающей среды, такими как атмосфера (воздух), биосфера (экосистемы), гидросфера (водные системы), литосфера (твердая оболочка Земли) и педосфера (почвы), и интегрировать существующие теории в рамках целостной теоретической конструкции. Это позволит лучше понять функционирование разнообразных взаимосвязанных процессов и их связь с человеческими системами, а также поможет создавать объяснительные и имитационные модели в целях прогнозирования долгосрочных тенденций и поиска возможных стратегий адаптации к ним.

### *Вычислительные социальные науки и цифровые гуманитарные науки*

Эпистемология, связанная с эмпиризмом в анализе больших данных и наукой, «идущей от данных», кажется, начинает трансформировать подходы к исследованиям, принятым в естественных науках, биологии, физических и технических науках. В то же время ее влияние на социальные и гуманитарные науки не так однозначно. Эти области исследований характеризуются разнообразием философских оснований, при этом только некоторые ученые придерживаются эпистемологии, характерной для естественных

наук. Те ученые, которые применяют метод естественных наук для объяснения и моделирования социальных явлений, в общем и целом опираются на идеи позитивизма, хотя они могут и не пользоваться таким обозначением [28]. Такие исследования, как правило, фокусируются на фактической и количественно измеримой информации, т. е. эмпирически фиксируемых явлениях, которые могут быть наблюдаемы и выражены в таких количественных показателях как количество, разница, стоимость, время, а не на более трудноуловимых аспектах жизни людей, таких как убеждения или идеология. При этом используется статистический анализ для установления взаимосвязей, построения теорий, объяснительных и имитационных моделей. Позитивистский подход имеет давнюю традицию в экономике, политологии, социально-экономической географии и социологии, но редко встречается в гуманитарных науках. Тем не менее в перечисленных выше дисциплинах, особенно в социально-экономической географии и социологии, в последние пятьдесят лет отмечается выраженный переход к постпозитивистским подходам.

Для позитивистски настроенных представителей социальных наук большие данные открывают возможность создавать более сложные, масштабные и детализированные модели человеческой жизни. Высказываются опасения, касающиеся доступа к большим данным по социальным и экономическим вопросам, значительная часть которых создается в частных интересах, и их качества. Тем не менее большие данные дают возможность перейти «от исследований, построенных на ограниченном объеме данных, к исследованиям, которым доступен значительный объем данных, от статических срезных измерений к исследованию динамики развития, от грубых обобщений к высокой детализации, от относительно простых объяснительных моделей к более сложным, тщательно проработанным имитационным моделям» [30, р. 3]. Становится возможным переход к эпохе вычислительных социальных наук, с исследованиями гораздо большего масштаба, глубины, охвата,

отражающими актуальные явления и процессы, исследованиями по природе своей лонгитюдными, что контрастирует с текущими практиками в социальных науках [3; 33]. Более того, многообразие, полнота, высокая детализация и взаимосвязанность данных в сочетании с растущей вычислительной мощностью компьютеров и развитием новых методик анализа данных позволяют снять некоторые пункты критики позитивистского подхода в исследованиях, прежде всего касающиеся его склонности к редукционизму и универсализму. Это возможно благодаря проведению более подробного, точного и чувствительного к деталям анализа, принимающего во внимание контекст и возможность непредвиденных событий. Это может способствовать уточнению и расширению теорий, объясняющих социальный и физический мир [29]. Кроме того, учитывая полноту данных, становится возможной проверка правдоподобия теорий в различных условиях и ситуациях. Можно утверждать, что в таких условиях знания об индивидах, сообществах, обществах и среде, в которой они существуют, станут более точными, глубокими и полезными с точки зрения разработки стратегии решения многочисленных проблем, с которыми сталкивается человечество.

Для постпозитивистски ориентированных ученых большие данные становятся источником как новых возможностей, так и трудностей. К возможностям относятся, во-первых, распространение, перевод в цифровой формат и установление взаимосвязей между значительным спектром разнообразных неструктурированных данных. Это – новые виды данных (например, по социальным сетям), к которым до настоящего времени было трудно получить доступ, скажем, миллионы книг, документов, газет, фотографий, произведений искусства, физических объектов, относящихся к различным историческим периодам. Уже около двух десятилетий их переводят в цифровой формат различные организации [11]. Во-вторых, существует возможность развития новых инструментов обработки, хранения и управления данными и их анализа, которые

можно применять к значительным объемам данных. Соответственно, становится возможным производить поиск и устанавливать взаимосвязи между значительным количеством работ, а не ограничиваться несколькими романами или фотографиями, или парой художников и их произведениями; возможно исследовать сотни тысяч медиа в Интернете, а не сосредоточиваться только на нескольких сайтах, чатах, видео или газетах [38]. Эти возможности наиболее активно осваиваются в развивающейся сфере цифровых гуманитарных наук.

Изначально цифровые гуманитарные науки включали в себя деятельность по обработке и анализу данных имеющих цифровой формат изначально, а также проекты по переводу в цифровой формат и созданию архивов. Цель этих проектов состояла в том, чтобы перевести аналоговые тексты и материальные объекты в цифровой формат, позволяющий систематизировать их и производить по ним поиск, а также осуществлять с ними обобщающий, автоматизированный или управляемый анализ, например, производить быструю визуализацию содержания [53]. Позднее представители цифровых гуманитарных наук разделились на два лагеря. Исследователи, относящиеся к первому из них, считают, что новые методики в цифровых гуманитарных науках – использование количественных показателей, графов, построение карт и сбор обобщенной информации о текстах без их чтения, – приносят методологическую выверенность и объективность в дисциплины, для которых до этого были характерны несистематичность и случайность в подходе и выборе предмета исследования [43; 48]. В противоположность им вторая группа исследователей утверждают, что новые методики скорее не заменяют традиционные методы и не приносят эмпирический или позитивистский подход в исследования в гуманитарных науках, а дополняют и усиливают существующие в гуманитарных науках методы, выступают как вспомогательное средство при использовании традиционных способов интерпретации и построения теории. Это позволяет проводить исследования большего масштаба и отвечать

на вопросы, на которые было бы практически невозможно найти ответы без использования вычислительных мощностей [4; 38].

Цифровые гуманитарные науки встречают не только одобрение, их критики утверждают, что использование компьютеров как «читающих устройств» [48], выполняющих задачи сбора обобщенной информации о текстах без их чтения исследователем [43], подрывает традиционные методы анализа, связанные с внимательным изучением содержания текста. Каллер [16, р. 22] отмечает, что вдумчивое чтение текста подразумевает, что «внимание уделяется тому, как конструируется и выражается смысл, к каким литературным и риторическим стратегиям и приемам прибегает автор, чтобы создать тот эффект, который работа или ее часть производят на читателя». Все это невозможно в случае сбора обобщенной информации о текстах без их чтения. Указанный автор также высказывает опасения относительно того, что характерный для цифровых гуманитарных наук подход способствует распространению исследований в сфере литературы без непосредственно чтения. Схожим образом Трампернер [59, р. 164] утверждает, что «статистически ориентированная модель истории литературы... вероятно, не может существовать без невидимой руки безличного исследователя», и продолжает: «Любая попытка увидеть общую картину требует обширных знаний, проницательности и исследовательской интуиции, позволяющей увидеть в целом, как литературные жанры и институты развиваются в исторической перспективе, а также развитых навыков точной и корректной интерпретации» [59, р. 170–171]. Схожим образом Марш [39] указывает, что артефакты культуры, такие как литература, не могут рассматриваться как просто набор данных. Литературный текст – не просто совокупность букв и слов, он имеет контекст, выражает какой-либо смысл и обладает качествами, которые сложно однозначно выразить. Аналитические алгоритмы не в состоянии адекватно уловить и расшифровать смысл или контекст, и, как утверждает Марш, рассматривают « всю литературу как однообразную». Далее он отмечает:



Построенный на алгоритмах анализ романов и газетных статей с необходимостью приводит к границе необоснованного упрощения. Процесс перевода литературы в данные стирает границу между ними. Он уничтожает уникальность. Он уничтожает усовершенствование, достигаемое благодаря критике. Он приводит к исчезновению истории восприятия и осмысления произведений.

Дженкинс [26] приходит к следующему выводу:

Ценность искусства, качество игры на сцене или картины не могут быть количественно измерены. Можно ввести в машину все возможные данные: даты, цвета, изображения, кассовые сборы, но ничто из этого не сможет объяснить, что есть конкретное произведение искусства, что оно значит, почему оно оказывает такое воздействие на людей. Для этого необходим человек [sic], а не машина.

Многие авторы, таким образом, считают, что цифровые гуманитарные науки способствуют распространению поверхностного анализа, обладающего рядом недостатков, а не глубинного, проникающего в суть анализа. Их методики грубы и приводят к излишним упрощениям, жертвуют сложностью, спецификой, контекстом, глубиной и критическим подходом ради достижения масштаба и широты охвата, автоматизации, описания закономерностей, чтобы произвести впечатление, что интерпретация не требует глубоких знаний по теме и понимания контекста.

Подобная критика может быть направлена и на вычислительные социальные науки. Так, например, карта языка твитов может помочь обнаружить паттерн распределения различных этнических сообществ в географическом пространстве города [50], но скорее важны другие вопросы – что создает его, почему он существует, какие процессы лежат в основе его формирования и воспроизведения, и каковы его социальные и экономические последствия. Одно дело – обнаружить закономерности, и другое дело – объяснить их. Последнее требует социологической теории и глубоких знаний в предметной области. Закономерности сами по себе выступают не конечной, а скорее на-

чальной точкой для дальнейшего анализа, который почти наверняка потребует использования других наборов данных.

Как и в случае с описанной выше критикой количественного и позитивистского подхода в социальных науках, вычислительные социальные критикуются постпозитивистами за механистичность, атомизм и ограниченность, сведение разнообразных индивидов и сложных, многоуровневых социальных структур всего лишь к отдельным значениям данных [60]. Более того, такой анализ пронизан положениями социального детерминизма, как показывает Пентланд [45]: «То, какой вы человек, главным образом определяется социальной средой, в которой вы находитесь, поэтому, если я вижу какие-либо проявления вашего поведения, я могу вывести из этого все остальное, просто сопоставляя вас с людьми в вашем окружении». Противоположная точка зрения заключается в том, что человеческие сообщества слишком сложны, зависимы от обстоятельств и неупорядочены, чтобы быть сведенными к формулам и законам. Получаемые количественные модели мало способствуют пониманию таких явлений, как войны, геноцид, домашнее насилие и расизм и могут дать только ограниченное понимание других созданных человеком систем, таких как экономика, а также недостаточно учитывают роль политики, идеологии, социальных структур и культуры [23]. Люди не поступают рациональным, заранее определенным образом, их жизни скорее полны противоречий, парадоксов и непредвиденных событий. Формы организации и способы функционирования обществ варьируются в пространстве и времени, и не существует их оптимальной или идеальной формы или универсальных свойств. Действительно, в мире имеется необыкновенное разнообразие индивидов, культур и способов жизни. Сводя эту сложность к неким абстрактным субъектам, из которых построены обобщающие модели, мы совершаем символическое насилие по отношению к производству знания. Более того, позитивистский подход сознательно не принимает во внимание метафизические аспекты человеческой жизни,

связанные со смыслами, верованиями и жизненным опытом, а также нормативные вопросы – этические и моральные дилеммы, касающиеся должного, а не сущего [28]. Иными словами, позитивистский подход сосредоточен на определенных типах вопросов, на которые он отвечает редуционистски, при этом игнорируя вопросы, что значит быть человеком и жить в многообразных обществах и территориях. Это не означает, что количественные подходы не имеют практической ценности – они очевидно имеют ее. Но следует осознавать их ограниченную способность объяснить человеческую жизнь и дополнять их другими подходами.

Брукс [8] в соответствии с этим утверждает, что для аналитики больших данных трудности представляют: социальное (люди нерациональны и ведут себя непредсказуемым образом, созданные человеком системы необыкновенно сложны, находятся друг с другом в противоречивых и сложных взаимоотношениях) и контекст (данные лишены социального, политического, экономического и исторического контекста), гигантский объем «стогов» создаваемых данных (включающих в себя ложные корреляции, в связи с чем затруднительно выявлять значимые для цели исследования «иголки»), работа с масштабными проблемами (особенно социальными и экономическими), благоприятствование мемам, а не шедеврам (они выявляют тенденции, но не всегда значимые явления, которые могут дать начало тенденции), затемнение ценностей и целей (тех, кто собирает данные, и тех, кто анализирует их). Иными словами, хотя анализ больших данных и может дать некоторые выводы, необходимо признать, что они охватывают ограниченный объем явлений, позволяют получить лишь специфические виды знания, и, наконец, требуют осмысления в контексте другой информации по теме, например, в рамках существующих теорий, официальных документов, исследований, проведенных на меньшем объеме данных, исторических данных, что позволит понять выявленные закономерности [13].

Помимо вопросов эпистемологии и методологии, часть трудностей связана с тем, что сбор и анализ больших данных не направлен

конкретным исследовательским вопросом и может проистекать скорее из желания применить какой-либо метод или использовать содержание данных. Кроме того, базы данных могут использоваться для решения иных задач, чем те, для которых они предназначены в первую очередь. Так, например, данные геотега в Твиттере не были изначально предназначены для получения информации о пространственном распределении языковых групп в городе и выявления процессов, определяющих развитие таких пространственных характеристик во времени. Вероятно, не следует удивляться тому, что эти данные дают лишь поверхностное представление, хотя это по-своему и интересно, а не глубокое понимание географии расы и языка, агломераций и сегрегаций в определенной местности.

Тогда как многие представители цифровых гуманитарных наук признают ценность внимательного чтения текста и подчеркивают, что сбор обобщенной информации о текстах выступает их дополнением, придавая им глубину и предоставляя контекст, позитивистские формы социальных наук противопоставлены постпозитивистским подходам. Различие между гуманитарными и социальными науками в данном аспекте объясняется тем, что статистика, используемая в цифровых гуманитарных науках, по большей части является описательной и служит выявлению закономерностей и определению их характеристик. Вычислительные социальные науки применяют научный метод, дополняя описательную статистику статистическим выводом, предназначенным для определения взаимосвязей и причинно-следственных связей. Иными словами, вычислительные социальные науки построены на эпистемологии, ориентирующей на создание сложных статистических моделей, объясняющих, воспроизводящих в виде моделей и прогнозирующих явления человеческой жизни. Это гораздо труднее сочетать с постпозитивистским подходом. Защита подхода выстраивается вокруг того, насколько хорошо работают методы и модели, какова их практическая ценность, а не самооценность использования в исследовании большей по масштабу и охвату базы данных.

Существует и иная потенциально продуктивная альтернатива изложенной позиции. Она связана с принятием и развитием эпистемологии, лежащей в основе критических исследований с использованием геоинформационных систем и радикальной статистики (*radical statistics*). Данные исследования включают применение количественных методик, статистического вывода, объяснительных и имитационных моделей, вместе с тем оставаясь чувствительными и открытыми к собственным эпистемологическим ограничениям. Эти исследования также опираются на критическую традицию в социологической теории в том, что касается проведения исследования, осмысления выводов и применения полученных знаний. Здесь признается, что исследование – это не нейтральная объективная деятельность, дающая непредвзятые выводы; напротив, используемые базы данных, сам процесс проведения исследования и интерпретация его результатов всегда внутренне насыщены актуальными политическими вопросами и проблемами [22; 51]. Таким образом, признается, что исследователь имеет определенную позицию (которая проявляется в его знаниях, опыте, убеждениях, целях и т.д.), что исследование осуществляется в определенном контексте (внутридисциплинарных научных дискуссий, вопросов финансирования, социальной политики в широком смысле и т.д.), что данные отражают методику их сбора и в связи с этим обладают определенными особенностями (связанными с характеристиками выборки, онтологическими предпосылками, «чистотой» данных, их полнотой, непротиворечивостью, достоверностью и точностью) и, наконец, что используемые методы анализа оказывают определенное воздействие на полученные результаты и их интерпретацию. Более того, признается, что исследования не осуществляются идеологически нейтрально, что на них явно и неявно оказывают влияние намерения и цели исследователей и спонсоров, источники финансирования, а также те, кто переводит результаты исследований на язык политических программ, инструментов и действий. Иными словами, в рамках такой эпистемологии исследование

обладает свойствами рефлексивности и открытости в том, что касается процесса исследования, признавая наличие непредвиденных обстоятельств и относительности применяемого подхода, что позволяет получить подробные и точные выводы, встроенные в определенный контекст. Такая эпистемология также не отрицает возможности дополнения вычислительных социальных наук исследованиями, использующими малый объем данных в целях получения дополнительных углубляющих исследование выводов [13]. Иными словами, можно представить себе новые эпистемологии, не отрицающие анализ больших данных и не отказывающихся от него, а скорее подразумевающую применение методологического подхода, характерного для науки, «идущей от данных», в контексте иной эпистемологии, позволяющей представителям социальных наук делать на основе больших данных глубокие выводы, осмысленные в определенном контексте и сформулированные с учетом существующих ограничений метода.

### *Заключение*

Едва ли можно сомневаться, что развитие больших данных и новых методов анализа данных делает возможным трансформацию эпистемологии естественных, социальных и гуманитарных наук, и этот процесс уже в полной мере выражен в различных дисциплинах. Большие данные и новые методы анализа данных создают основу для развития новых подходов к сбору и анализу данных, позволяющих по-новому ставить исследовательские вопросы и отвечать на них. Вместо задачи формулирования выводов на основе баз данных ограниченного масштаба, временного охвата и объема, большие данные ставят сложную задачу иного рода – обработать и проанализировать динамично изменяющиеся и разнообразные базы данных огромного размера. В качестве решения выступает разработка новых способов управления данными и аналитических методик, основанных на машинном обучении и новых способах визуализации.

В том, что касается естественных наук, доступность больших данных и новых исследовательских практик привела к тому, что некоторые авторы объявили о возникновении новой, четвертой парадигмы науки, основанной на интенсивном анализе большого объема данных и оспаривающей укоренившийся в науке дедуктивный подход. Сегодня очевидно, что большие данные представляют собой инновацию, изменяющую научное исследование, создавая возможность нового подхода в науке. Он еще не принял конкретную форму, но выделяются два возможных направления, характеризующиеся различной эпистемологией, – эмпиризм, в рамках которого данные говорят сами за себя вне теории, и наука, «идущая от данных», которая трансформирует существующий научный метод, интегрируя элементы абдукции, индукции и дедукции. Учитывая недостатки в основных положениях аргументации эмпиризма, можно с высокой вероятностью предполагать, что с течением времени в связи с распространением больших данных и совершенствованием новых методик анализа данных победит наука, «идущая от данных», что станет серьезным вызовом привычному научному методу, «идущему от знания». Данным изменениям, однако, должна сопутствовать проработка и обсуждение философских оснований науки, «идущей от данных», сосредоточенная на базовых положениях ее эпистемологии, принципах и методологии с тем, чтобы была создана устойчивая теоретическая основа новой парадигмы.

Ситуация в гуманитарных и социальных науках осложнена разнообразием принятых в них философских оснований, в связи с чем большие данные и новые методы анализа данных вряд ли могут привести к формированию новых парадигм. Большие данные скорее увеличат спектр данных, доступных для анализа, и создадут возможность развития новых подходов и методик анализа, чем полностью заменят традиционные исследования на ограниченных объемах данных. В определенной степени причиной этого выступают существующие философские основания науки. Кроме того, маловероятно, что будут созданы такие большие данные, которые

могут быть использованы для поиска ответов на конкретные исследовательские вопросы, что актуализирует потребность в более узконаправленных исследованиях со специфическими целями. Тем не менее, как утверждают Китчин [29] и Рупперт [52], большие данные предоставляют исследователям в области социальных и гуманитарных наук ряд возможностей, прежде всего использовать значительные объемы ценных данных по социальным, культурным, экономическим, политическим и историческим вопросам. Большие данные также создают и ряд сложностей, включая нехватку навыков анализа и осмысления таких данных, а также необходимость создания эпистемологического подхода, делающего возможным развитие постпозитивистских форм вычислительных социальных наук. Одним из возможных вариантов решения проблемы является эпистемология, вдохновляемая критическими исследованиями геоинформационных систем, а также работами, основанными на радикальной статистике. Здесь количественные методы и модели используются рефлексивно, с пониманием того, что контекст, позиция и проводимая политика имеют значение, вместо того чтобы немедленно отказаться от такого подхода. Потенциальные преимущества данной эпистемологии также связаны с тем, что она признает и применяет абдукцию и способствует развитию более рефлексивной науки, «идущей от данных». Как показывает наше предварительное обсуждение, существует острая потребность в осмыслении последствий больших данных и новых методов анализа данных для эпистемологии в широком плане. Несмотря на динамичные изменения в сфере обработки и анализа данных, едва ли можно сказать, что мы начали решать эту задачу.

### *Благодарности*

Автор выражает признательность Эвелин Рупперт и Марку Бойлу за высказанные ими ценные замечания по первоначальному варианту статьи.



Источником финансирования исследований, лежащих в основе этой статьи, выступил грант Европейского исследовательского совета по проекту «The Programmable City» (ERC-2012- AdG-323636).

#### ЛИТЕРАТУРА

1. *Amin A., Thrift N.* Cities: Reimagining the Urban. London: Polity, 2002.
2. *Anderson C.* The End of Theory: The Data Deluge Makes the Scientific Method Obsolete // *Wired*. 2008. 23 June. URL: [http://www.wired.com/science/discoveries/magazine/16-07/pb\\_theory](http://www.wired.com/science/discoveries/magazine/16-07/pb_theory) (data of access: 12.10.2012).
3. *Batty M., Axhausen K.W., Giannotti F., et al.* Smart Cities of the Future // *European Physical Journal Special Topics*. 2012. Vol. 214. Is. 1. P. 481–518.
4. *Berry D.* The Computational Turn: Thinking about the Digital Humanities // *Culture Machine*. 2011. 12. URL: <http://www.culturemachine.net/index.php/cm/article/view/440/470> (data of access: 03.12.2012).
5. *Bettencourt L.M.A., Lobo J., Helbing D., et al.* Growth, Innovation, Scaling, and the Pace of Life in Cities // *Proceedings of the National Academy of Sciences*. 2007. Vol. 104. No. 17. P. 7301–7306.
6. *Bollier D.* The Promise and Peril of Big Data // *The Aspen Institute*. 2010. URL: [http://www.aspeninstitute.org/sites/default/files/content/docs/pubs/The\\_Promise\\_and\\_Peril\\_of\\_Big\\_Data.pdf](http://www.aspeninstitute.org/sites/default/files/content/docs/pubs/The_Promise_and_Peril_of_Big_Data.pdf) (data of access: 01.10.2012).
7. *Boyd D., Crawford K.* Critical Questions for Big Data // *Information, Communication and Society*. 2012. Vol. 15. Is. 5. P. 662–679.
8. *Brooks D.* What Data Can't Do // *New York Times*. 2013. 18 February. URL: <http://www.nytimes.com/2013/02/19/opinion/brooks-what-data-cant-do.html> (data of access: 18.02.2013).
9. *Bryant R., Katz R.H., Lazowska E.D.* (2008) Big-data Computing: Creating Revolutionary Breakthroughs in Commerce, Science and Society // *Computing Research Initiatives for the 21st Century*, Computing Research Association, 2008, 8. URL: [http://www.cra.org/ccc/docs/init/Big\\_Data.pdf](http://www.cra.org/ccc/docs/init/Big_Data.pdf) (data of access: 12.10.2012).
10. *Clark L.* No Questions Asked: Big Data Firm Maps Solutions without Human Input // *Wired*. 2013. 16 January 2013. URL: <http://www.wired.co.uk/news/archive/2013-01/16/ayasdi-big-data-launch> (data of access: 28.01.2013).
11. *Cohen D.* Contribution to: The Promise of Digital History (roundtable discussion) // *Journal of American History*. 2008. Vol. 95. Is. 2. P. 452–491.
12. *Constine J.* How Big is Facebook's Data? 2.5 Billion Pieces of Content and 500 Terabytes Ingested Every Day. 2012. 22 August. URL: <http://techcrunch.com/2012/08/22/how-big-is-facebooks-data-2-5-billion-pieces-of-content-and-500-terabytes-ingested-every-day/> (data of access: 28.01.2013).

13. *Crampton J., Graham M., Poorthuis A., et al.* Beyond the Geotag? Deconstructing 'Big Data' and Leveraging the Potential of the Geoweb. 2012. URL: [http://www.uky.edu/tmute2/geography\\_methods/readingPDFs/2012-Beyond-the-Geotag-2012.10.01.pdf](http://www.uky.edu/tmute2/geography_methods/readingPDFs/2012-Beyond-the-Geotag-2012.10.01.pdf) (data of access: 21.02.2013).
14. *Crawford K.* The Hidden Biases of Big Data // Harvard Business Review Blog. 2013. 1 April. URL: <http://blogs.hbr.org/2013/04/the-hidden-biases-in-big-data/> (data of access: 18.09.2013).
15. *Cukier K.* Data, Data Everywhere // The Economist. 2010. 25 February.
16. *Culler J.* The Closeness of Close Reading // ADE Bulletin. 2010. Vol. 149. P. 20–25.
17. *Dodge M., Kitchin R.* Codes of Life: Identification Codes and the Machine-readable World // Environment and Planning D: Society and Space. 2005. Vol. 23. Is. 6. P. 851–881.
18. *Dyche J.* Big data 'Eurekas!' Don't Just Happen // Harvard Business Review Blog. 2012. 20 November. URL: [http://blogs.hbr.org/cs/2012/11/eureka\\_doesnt\\_just\\_happen.html](http://blogs.hbr.org/cs/2012/11/eureka_doesnt_just_happen.html) (data of access: 23.11.2012).
19. *Floridi L.* Big Data and Their Epistemological Challenge // Philosophy and Technology. 2012. Vol. 25. Is. 4. P. 435–437.
20. *Gould P.* Letting the Data Speak for Themselves // Annals of the Association of American Geographers. 1981. Vol. 71. Is. 2. P. 166–176.
21. *Han J., Kamber M., Pei.* Data Mining: Concepts and Techniques, 3rd ed. Waltham: Morgan Kaufmann, 2011.
22. *Haraway D.* Simians, Cyborgs and Women: The Reinvention of Nature. New York: Routledge, 1991.
23. *Harvey D.* Social Justice and the City. Oxford: Blackwell, 1972.
24. *Hastie T., Tibshirani R., Friedman J.* The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. New York: Springer, 2009.
25. *Hey T., Tansley S., Tolle K.* Jim Grey on eScience: A Transformed Scientific Method // The Fourth Paradigm: Data-Intensive Scientific Discovery / Eds: Hey T., Tansley S., Tolle K. Redmond: Microsoft Research, 2009. P. XVII–XXXI.
26. *Jenkins T.* Don't Count on Big Data for Answers // The Scotsman. 2013. 12 February. URL: <http://www.scotsman.com/the-scotsman/opinion/comment/tiffanyjenkins-don-t-count-on-big-data-for-answers-1-2785890> (data of access: 11.03.2013).
27. *Kelling S., Hochachka W., Fink D., et al.* Data-intensive Science: A New Paradigm for Biodiversity Studies // BioScience. 2009. Vol. 59. Is. 7. P. 613–620.
28. *Kitchin R.* Positivist Geography and Spatial Science // Approaches in Human Geography / Eds: Aitken S., Valentine G. London: Sage, 2006. P. 20–29.
29. *Kitchin R.* Big Data and Human Geography: Opportunities, Challenges and Risks // Dialogues in Human Geography. 2013. Vol. 3. Is. 3. P. 262–267.
30. *Kitchin R.* The Real-time City? Big Data and Smart Urbanism // GeoJournal. 2014. Vol. 79. P. 1–14.

31. *Kuhn T.* The Structure of Scientific Revolutions. Chicago: Univ. of Chicago Press, 1962.
32. *Laney D.* 3D Data Management: Controlling Data Volume, Velocity and Variety // Meta Group. 2001. URL: <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-VolumeVelocity-and-Variety.pdf> (data of access: 16.01.2013).
33. *Lazer D., Pentland A., Adamic L., et al.* Computational Social Science // Science. 2009. Vol. 323. P. 721–733.
34. *Lehning M., Dawes N., Bavay M., et al.* Instrumenting the Earth: Next-generation Sensor Networks and Environmental Science // The Fourth Paradigm: Data-Intensive Scientific Discovery / Eds: Hey T., Tansley S., Tolle K. Redmond: Microsoft Research, 2009. P. 45–51.
35. *Lehrer J.* A Physicist Solves the City // New York Times. 2010. 17 December. URL: [http://www.nytimes.com/2010/12/19/magazine/19Urban\\_West-t.html](http://www.nytimes.com/2010/12/19/magazine/19Urban_West-t.html) (data of access: 23.12.2013).
36. *Leonelli S.* Introduction: Making Sense of Data-driven Research in the Biological and Biomedical Sciences // Studies in History and Philosophy of Biological and Biomedical Sciences. 2012. Vol. 43. Is. 1. P. 1–3.
37. *Loukides M.* What is Data Science? // O'Reilly Radar. 2010. 2 June. URL: <http://radar.oreilly.com/2010/06/what-is-data-science.html> (data of access: 28.01.2013).
38. *Manovich L.* Trending: The Promises and the Challenges of Big Social Data. 2011. URL: [http://www.manovich.net/DOCS/Manovich\\_trending\\_paper.pdf](http://www.manovich.net/DOCS/Manovich_trending_paper.pdf) (data of access: 09.11.2012).
39. *Marche S.* Literature is not Data: Against Digital Humanities // Los Angeles Review of Books. 2012. 28 October. URL: <http://lareviewofbooks.org/article.php?id=1040&fulltext=1> (data of access: 04.04.2013).
40. *Marz N., Warren J.* MEAP (ed.), Big Data: Principles and Best Practices of Scalable Realtime Data Systems. Westhampton: Manning, 2012.
41. *Mayer-Schonberger V., Cukier K.* Big Data: A Revolution that Will Change How We Live, Work and Think. London: John Murray, 2013.
42. *Miller H.J.* The Data Avalanche is Here. Shouldn't We Be Digging? // Journal of Regional Science. 2010. Vol. 50. Is. 1. P. 181–201.
43. *Moretti F.* Graphs, Maps, Trees: Abstract Models for a Literary History. London: Verso, 2005.
44. Open Data Center Alliance. Big Data Consumer Guide. Open Data Center Alliance. 2012. URL: [http://www.opendatacenteralliance.org/docs/Big\\_Data\\_Consumer\\_Guide\\_Rev1.0.pdf](http://www.opendatacenteralliance.org/docs/Big_Data_Consumer_Guide_Rev1.0.pdf) (data of access: 11.02.2013).
45. *Pentland A.* Reinventing Society in the Wake of Big Data // Edge. 2012. 30 August. URL: <http://www.edge.org/conversation/reinventing-society-in-the-wake-of-big-data> (data of access: 28.01.2013).

46. *Porway J.* You Can't Just Hack Your Way to Social Change // Harvard Business Review Blog. 2013. 7 March. URL: [http://blogs.hbr.org/cs/2013/03/you\\_cant\\_just\\_hack\\_your\\_way\\_to.html](http://blogs.hbr.org/cs/2013/03/you_cant_just_hack_your_way_to.html) (data of access: 09.03.2013).

47. *Prensky M.H.* Sapiens Digital: From Digital Immigrants and Digital Natives to Digital Wisdom // Innovate. 2009. Vol. 5(3). URL: <http://www.innovateonline.info/index.php?view=article&id=705> (data of access: 12.10.2012).

48. *Ramsay S.* Reading Machines: Towards an Algorithmic Criticism. Champaign: University of Illinois Press, 2010.

49. *Ribes D., Jackson S.J.* Data Bite Man: The Work of Sustaining Long-term Study // 'Raw Data' is an Oxymoron / Ed.: L. Gitelman. Cambridge, MA: MIT Press, 2013. P. 147–166.

50. *Rogers S.* Twitter's Languages of New York Mapped // The Guardian. 2013. 21 February. URL: <http://www.guardian.co.uk/news/datablog/interactive/2013/feb/21/twitter-languages-new-york-mapped> (data of access: 03.04.2013).

51. *Rose G.* Situating Knowledges: Positionality, Reflexivities and Other Tactics // Progress in Human Geography. 1997. Vol. 21. Is. 3. P. 305–320.

52. *Ruppert E.* Rethinking Empirical Social Sciences // Dialogues in Human Geography. 2013. Vol. 3. Is. 3. P. 268–273.

53. *Schnapp J., Presner P.* Digital Humanities Manifesto 2.0. 2009. URL: [http://www.humanitiesblast.com/manifesto/Manifesto\\_V2.pdf](http://www.humanitiesblast.com/manifesto/Manifesto_V2.pdf) (data of access: 13.03.2013).

54. *Seni G., Elder J.* Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions. San Rafael: Morgan and Claypool, 2010.

55. *Siegel E.* Predictive Analytics. Hoboken: Wiley, 2013.

56. *Steadman I.* Big Data and the Death of the Theorist // Wired. 2013. 25 January. URL: <http://www.wired.co.uk/news/archive/2013-01/25/big-data-end-of-theory> (data of access: 30.01.2013).

57. *Strasser B.J.* Data-driven Sciences: From Wonder Cabinets to Electronic Databases // Studies in History and Philosophy of Biological and Biomedical Sciences. 2012. Vol. 43. P. 85–87.

58. *Strom D.* Big Data Makes Things Better // Slashdot. 2012. 3 August. URL: <http://slashdot.org/topic/bi/bigdata-makes-things-better/> (data of access: 24.10.2013).

59. *Trumpener K.* Critical Response I. Paratext and Genre System: A Response to Franco Moretti // Critical Inquiry. 2009. Vol. 36. Is. 1. P. 159–171.

60. *Wyly E.* Automated (Post)positivism. Urban Geography. (In press)

61. *Zikopoulos P.C., Eaton C., DeRoos D., et al.* Understanding Big Data. New York: McGraw Hill, 2012.

**Kitchin Rob**

*National Institute for Regional and Spatial Analysis, National University of Ireland Maynooth, County Kildare, Ireland*

Translated by:

**Shaeva Olga**

*Moscow State Linguistic University, Moscow, shaeva.olga@gmail.com*

### **Big Data, new epistemologies and paradigm shifts**

This article examines how the availability of Big Data, coupled with new data analytics, challenges established epistemologies across the sciences, social sciences and humanities, and assesses the extent to which they are engendering paradigm shifts across multiple disciplines. In particular, it critically explores new forms of empiricism that declare ‘the end of theory’, the creation of data-driven rather than knowledge-driven science, and the development of digital humanities and computational social sciences that propose radically different ways to make sense of culture, history, economy and society. It is argued that: (1) Big Data and new data analytics are disruptive innovations which are reconfiguring in many instances how research is conducted; and (2) there is an urgent need for wider critical reflection within the academy on the epistemological implications of the unfolding data revolution, a task that has barely begun to be tackled despite the rapid changes in research practices presently taking place. After critically reviewing emerging epistemological positions, it is contended that a potentially fruitful approach would be the development of a situated, reflexive and contextually nuanced epistemology. *Keywords:* Big Data, data analytics, epistemology, paradigms, end of theory, data-driven science, digital humanities, computational social sciences

### **References**

1. Amin A., Thrift N. *Cities: Reimagining the Urban*. London: Polity, 2002.
2. Anderson C. The end of theory: The data deluge makes the scientific method obsolete, *Wired*, 23 June 2008. [on-line] URL: [http://www.wired.com/science/discoveries/magazine/16-07/pb\\_theory](http://www.wired.com/science/discoveries/magazine/16-07/pb_theory) (date of access: 12.10.2012).
3. Batty M., Axhausen K.W., Giannotti F., et al. Smart cities of the future, *European Physical Journal Special Topics*, 2012, 214 (1), 481–518.

4. Berry D. The computational turn: Thinking about the digital humanities, *Culture Machine* 12, 2011. URL: <http://www.culturemachine.net/index.php/cm/article/view/440/470> (date of access: 03.12.2012).
5. Bettencourt L.M.A., Lobo J., Helbing D., et al. Growth, innovation, scaling, and the pace of life in cities, *Proceedings of the National Academy of Sciences*, 2007, 104 (17), 7301–7306.
6. Bollier D. *The Promise and Peril of Big Data*. The Aspen Institute. 2010. URL: [http://www.aspeninstitute.org/sites/default/files/content/docs/pubs/The\\_Promise\\_and\\_Peril\\_of\\_Big\\_Data.pdf](http://www.aspeninstitute.org/sites/default/files/content/docs/pubs/The_Promise_and_Peril_of_Big_Data.pdf) (date of access: 01.10.2012).
7. Boyd D., Crawford K. Critical questions for big data, *Information, Communication and Society*, 2012, 15 (5), 662–679.
8. Brooks D. What data can't do, *New York Times*. 18 February 2013. URL: <http://www.nytimes.com/2013/02/19/opinion/brooks-what-data-cant-do.html> (date of access: 18.02.2013).
9. Bryant R., Katz R.H., Lazowska E.D. “Big-data computing: Creating revolutionary breakthroughs in commerce, science and society”, in: *Computing Research Initiatives for the 21st Century*, Computing Research Association, 2008, 8. URL: [http://www.cra.org/ccc/docs/init/Big\\_Data.pdf](http://www.cra.org/ccc/docs/init/Big_Data.pdf) (date of access: 12.10.2012).
10. Clark L, Questions asked: Big data firm maps solutions without human input. *Wired*, 16 January 2013. [on-line] URL: <http://www.wired.co.uk/news/archive/2013-01/16/ayasdi-big-data-launch> (date of access: 28.01.2013).
11. Cohen D. Contribution to: The promise of digital history (roundtable discussion), *Journal of American History*, 2008, 95 (2), 452–491.
12. Constine J. *How big is Facebook's data? 2.5 billion pieces of content and 500p terabytes ingested every day*, 22 August 2012. URL: <http://techcrunch.com/2012/08/22/how-big-is-facebooks-data-2-5-billion-pieces-ofcontent-and-500-terabytes-ingested-every-day/> (date of access: 28.01.2013).
13. Crampton J., Graham M., Poorthuis A., et al. *Beyond the Geotag? Deconstructing 'Big Data' and Leveraging the Potential of the Geoweb*. 2012. URL: [http://www.uky.edu/tmute2/geography\\_methods/readingPDFs/2012-Beyond-the-Geotag-2012.10.01.pdf](http://www.uky.edu/tmute2/geography_methods/readingPDFs/2012-Beyond-the-Geotag-2012.10.01.pdf) (date of access: 21.02.2013).
14. Crawford K. The hidden biases of big data, *Harvard Business Review Blog*. 1 April 2013. [on-line] URL: <http://blogs.hbr.org/2013/04/the-hidden-biases-in-big-data/> (date of access: 18.09.2013).

15. Cukier K. Data, data everywhere, *The Economist*. 25 February 2010 (accessed 12 November 2012).
16. Culler J. The closeness of close reading, *ADE Bulletin*, 2010, 149, 20–25.
17. Dodge M., Kitchin R. Codes of life: Identification codes and the machine-readable world, *Environment and Planning D: Society and Space*, 2005, 23 (6), 851–881.
18. Dyche J. Big data ‘Eurekas!’ don’t just happen, *Harvard Business Review Blog*. 20 November 2012. URL: [http://blogs.hbr.org/cs/2012/11/eureka\\_doesnt\\_just\\_happen.html](http://blogs.hbr.org/cs/2012/11/eureka_doesnt_just_happen.html) (date of access: 23.11.2012).
19. Floridi L. Big data and their epistemological challenge, *Philosophy and Technology*. 2012, 25 (4), 435–437.
20. Gould P. Letting the data speak for themselves, *Annals of the Association of American Geographers*, 1981, 71 (2), 166–176.
21. Han J., Kamber M., Pei J. *Data Mining: Concepts and Techniques*, 3rd ed. Waltham: Morgan Kaufmann, 2011.
22. Haraway D. *Simians, Cyborgs and Women: The Reinvention of Nature*. New York: Routledge, 1991.
23. Harvey D. *Social Justice and the City*. Oxford: Blackwell, 1972.
24. Hastie T., Tibshirani R., Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York: Springer, 2009.
25. Hey T., Tansley S., Tolle K. “Jim Grey on eScience: A transformed scientific method”, in: Hey T, Tansley S and Tolle K (eds) *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmond: Microsoft Research, 2009. P. xvii–xxxii.
26. Jenkins T. Don’t count on big data for answers, *The Scotsman*, 12 February 2013. URL: <http://www.scotsman.com/the-scotsman/opinion/comment/tiffanyjenkins-don-t-count-on-big-data-for-answers-1-2785890> (date of access: 11.03.2013).
27. Kelling S., Hochachka W., Fink D., et al. Data-intensive Science: A new paradigm for biodiversity studies, *BioScience*. 2009, 59 (7), 613–620.
28. Kitchin R. “Positivistic geography and spatial science”, in: Aitken S and Valentine G (eds.) *Approaches in Human Geography*. London: Sage, 2006. P. 20–29.
29. Kitchin R. Big data and human geography: Opportunities, challenges and risks, *Dialogues in Human Geography*. 2013, 3 (3), 262–267.

30. Kitchin R. The real-time city? Big data and smart urbanism, *GeoJournal*, 2014, 79, 1–14.
31. Kuhn T. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press, 1962.
32. Laney D. *3D data management: Controlling data volume, velocity and variety*. Meta group. 2001. URL: <http://blogs.gartner.com/douglaney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-VolumeVelocity-and-Variety.pdf> (date of access: 16.01.2013).
33. Lazer D., Pentland A., Adamic L., et al. Computational social science, *Science*, 2009, 323, 721–733.
34. Lehning M., Dawes N., Bavay M., et al. “Instrumenting the earth: Next-generation sensor networks and environmental science”, in: Hey T, Tansley S and Tolle K (eds) *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmond: Microsoft Research, 2009. P. 45–51.
35. Lehrer J. A physicist solves the city, *New York Times*. 17 December 2010. URL: [http://www.nytimes.com/2010/12/19/magazine/19Urban\\_West-t.html](http://www.nytimes.com/2010/12/19/magazine/19Urban_West-t.html) (date of access: 23.12.2013).
36. Leonelli S. Introduction: Making sense of data-driven research in the biological and biomedical sciences, *Studies in History and Philosophy of Biological and Biomedical Sciences*, 2012, 43 (1), 1–3.
37. Loukides M. What is data science? *O'Reilly Radar*. 2 June 2010. URL: <http://radar.oreilly.com/2010/06/what-is-data-science.html> (date of access: 28.01.2013).
38. Manovich L. Trending: *The promises and the challenges of big social data*. 2011. URL: [http://www.manovich.net/DOCS/Manovich\\_trending\\_paper.pdf](http://www.manovich.net/DOCS/Manovich_trending_paper.pdf) (date of access: 09.11.2012).
39. Marche S. Literature is not data: Against digital humanities, *Los Angeles Review of Books*. 28 October 2012. URL: <http://lareviewofbooks.org/article.php?id=1040&fulltext=1> (date of access: 04.04.2013).
40. Marz N., Warren J. In: MEAP (ed.), *Big Data: Principles and Best Practices of Scalable Realtime Data Systems*. Westhampton: Manning, 2012.
41. Mayer-Schonberger V., Cukier K. *Big Data: A Revolution that Will Change How We Live, Work and Think*. London: John Murray, 2013.
42. Miller H.J. The data avalanche is here. Shouldn't we be digging? *Journal of Regional Science*, 2010, 50 (1), 181–201.
43. Moretti F. *Graphs, Maps, Trees: Abstract Models for a Literary History*. London: Verso, 2005.



44. *Open Data Center Alliance*. Big Data Consumer Guide. Open Data Center Alliance. 2012. URL: [http://www.opendatacenteralliance.org/docs/Big\\_Data\\_Consumer\\_Guide\\_Rev1.0.pdf](http://www.opendatacenteralliance.org/docs/Big_Data_Consumer_Guide_Rev1.0.pdf) (date of access: 11.02.2013).
45. Pentland A. Reinventing society in the wake of big data, *Edge*. 30 August 2012. URL: <http://www.edge.org/conversation/reinventing-society-in-the-wake-ofbig-data> (date of access: 28.01.2013).
46. Porway J. You can't just hack your way to social change, *Harvard Business Review Blog*. 7 March 2013. URL: [http://blogs.hbr.org/cs/2013/03/you\\_cant\\_just\\_hack\\_your\\_way\\_to.html](http://blogs.hbr.org/cs/2013/03/you_cant_just_hack_your_way_to.html) (date of access: 09.03.2013).
47. Prensky M.H. Sapiens digital: From digital immigrants and digital natives to digital wisdom, *Innovate*, 2009, 5(3). URL: <http://www.innovateonline.info/index.php?view=article&id=705> (date of access: 12.10.2012).
48. Ramsay S. *Reading Machines: Towards an Algorithmic Criticism*. Champaign: University of Illinois Press, 2010.
49. Ribes D., Jackson S.J. "Data bite man: The work of sustaining long-term study", in: Gitelman L (ed.) *'Raw Data' is an Oxymoron*. Cambridge, MA: MIT Press, 2013. P. 147–166.
50. Rogers S. Twitter's languages of New York mapped, *The Guardian*. 21 February 2013. URL: <http://www.guardian.co.uk/news/datablog/interactive/2013/feb/21/twitter-languages-new-york-mapped> (date of access: 03.04.2013).
51. Rose G. Situating knowledges: Positionality, reflexivities and other tactics, *Progress in Human Geography*, 1997, 21 (3), 305–320.
52. Ruppert E. Rethinking empirical social sciences, *Dialogues in Human Geography*, 2013, 3 (3), 268–273.
53. Schnapp J., Presner P. *Digital Humanities Manifesto 2.0*. 2009. URL: [http://www.humanitiesblast.com/manifesto/Manifesto\\_V2.pdf](http://www.humanitiesblast.com/manifesto/Manifesto_V2.pdf) (date of access: 13.03.2013).
54. Seni G., Elder J. *Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions*. San Rafael: Morgan and Claypool, 2010.
55. Siegel E. *Predictive Analytics*. Hoboken: Wiley, 2013.
56. Steadman I. Big data and the death of the theorist, *Wired*. 25 January 2013. URL: <http://www.wired.co.uk/news/archive/2013-01/25/big-data-end-of-theory> (date of access: 30.01.2013).
57. Strasser B.J. Data-driven sciences: From wonder cabinets to electronic databases, *Studies in History and Philosophy of Biological and Biomedical Sciences*, 2012, 43, 85–87.

58. Strom D. Big data makes things better, *Slashdot*. 3 August 2012. URL: <http://slashdot.org/topic/bi/bigdata-makes-things-better/> (date of access: 24.10.2013).
59. Trumpener K. Critical response I. Paratext and genre system: A response to Franco Moretti, *Critical Inquiry*. 2009, 36 (1), 159–171.
60. Wylly E. (in press) *Automated (post)positivism*. Urban Geography.
61. Zikopoulos P.C., Eaton C., DeRoos D., et al. *Understanding Big Data*. New York: McGraw Hill, 2012.