

ПРАКТИКИ ГОВОРЯТ

DOI: 10.14515/monitoring.2017.3.04

Правильная ссылка на статью:

Одинцов А. В. Социология общественного мнения и вызов Big Data // Мониторинг общественного мнения: Экономические и социальные перемены. 2017. № 3. С. 30—43. DOI: 10.14515/monitoring.2017.3.04.

For citation:

Odintsov A. V. Sociology of public opinion and the Big Data challenge. *Monitoring of Public Opinion: Economic and Social Changes*. 2017. № 3. P. 30—43. DOI: 10.14515/monitoring.2017.3.04.

А. В. Одинцов

СОЦИОЛОГИЯ ОБЩЕСТВЕННОГО МНЕНИЯ И ВЫЗОВ BIG DATA

СОЦИОЛОГИЯ ОБЩЕСТВЕННОГО МНЕНИЯ И ВЫЗОВ BIG DATA

SOCIOLOGY OF PUBLIC OPINION AND THE BIG DATA CHALLENGE

ОДИНЦОВ Александр Владимирович — кандидат социологических наук, Волгоградский институт управления — филиал РАНХиГС, Волгоград, Россия.
E-MAIL: yugrasil@yandex.ru
ORCID: 0000-0002-9558-5296

Alexander V. ODINTSOV¹ — *Cand. Sci (Sociol.)*
E-MAIL: yugrasil@yandex.ru
ORCID: 0000-0002-9558-5296

¹ *Volgograd Institute of Management — branch of the Russian Presidential Academy of National Economy and Public Administration, Volgograd, Russia*

Аннотация. Big Data — один из факторов, поколебавший легитимность опросных методик как доминирующих в изучении социальной реальности. Появление Big Data поставило перед исследователями вопрос о природе тех данных, с которыми они работали прежде. Сущностные характеристики Big Data (3V — volume, velocity, variety) дают им возможность эффективнее создавать прогностические модели массового поведения, чем социология общественного мнения. Это, учитывая постоянную критику прогнозирования, базирующегося на массовых опросах, делает Big Data все более привлекательным для исследования общества.

Abstract. Big Data has become one of the factors that shook the legitimacy of the polling techniques being dominant in studying social reality. Facing the big data challenge, scholars have to answer the question about the nature of the data they used before. The basic characteristics of big data (3V — volume, velocity, variety) allow drawing up the mass behavior forecasting models more efficiently than the public opinion sociology does. Taken into account constant criticism over the forecasting based on public surveys, this feature makes Big Data more attractive for the research community. The article considers two strategies that the public

В статье рассматриваются две стратегии, которые может использовать социология общественного мнения в качестве ответа на появление Big Data: 1. критика и поиск уязвимостей (Д. Бойд, К. Кроуфорд, Дж. Фан, Ф. Хан, Х. Лю, Д. Лазер, М. Вилсоном, Л. Манович); 2. коллаборация (М. Косински, Х. Форд, С. Боуман).

Ключевые слова: Big Data, small data, большие данные, малые данные, социология общественного мнения, объединение данных, критика больших данных

Благодарность. Работа выполнена при поддержке гранта РФФИ № 15-06-02758.

opinion sociology can use in response to Big Data: criticism and search for vulnerabilities (D. Boyd, K. Crawford, J. Fan, H. Liu, D. Lazer, M. Wilson, L. Manovich), and collaboration (M. Kosinski, H. Ford, S. Bowman).

Keywords: Big Data, small data, public opinion sociology, data aggregation, critics of big data

Acknowledgment. The paper is supported by the Russian Foundation for Basic Research (no. 15-06-02758).

На различных конференциях и круглых столах, посвященных вопросам методологии и методики исследования общества, регулярно возникает тема «Big Data». Ее появление, как правило, сопровождается такими коннотациями, как «вызов», «проблема», «угроза» традиционным метода сбора данных в социологии. Показательно в этой связи название секции VI международной социологической Грушинской конференции — «Вызов Big Data. Регистрация поведения vs опросные декларации». Но Big Data все еще остаются «скрытой угрозой» («The Phantom Menace») по отношению к традиционным опросным методам — у профессионального сообщества нарастают опасения, но повод для них пока не очевиден. В данной статье мы попытаемся разобраться, в чем сущность брошенного опросным методикам вызова и каким может быть ответ на него.

Социологию общественного мнения нередко считают самой объективной частью современной исследовательской традиции во многом благодаря тому, что в отличие от других дисциплин, изучающих общество, она пользуется достаточно развитым и широко применяемым математическим аппаратом. Кроме того, до последнего времени социология общественного мнения наиболее соответствовала критериям индуктивной эмпирической науки.

В XVI—XVII веках в Европе сформировалось специфическое видение эмпирической науки, своеобразным идеалом которого стал метод «элиминативной индукции», подробно описанный Ф. Бэконом в «Новом органоне». Квинтэссенция сбора, хранения и упорядочивания данных этого метода — «табличный метод» [Бэкон, 1978]. Его сущность состояла в том, что эмпирическая наука — это индуктивный путь познания, который возможен только в том случае, если у исследователя есть полное описание всех проявлений исследуемого свойства (явления).

Благодаря успешному опыту применения теории выборки и статистических форм анализа данных, социология общественного мнения всерьез претендовала на то, чтобы быть единственным максимально исчерпывающим способом описания массового поведения людей. Иначе говоря, теория выборки, широко примененная к изучению общества Дж. Гэллапом, предложила вариант, при котором для полного описания реальности (генеральной совокупности) можно обойтись неполной регистрацией случаев, проводя измерение только выборочной совокупности. В полной мере этот факт прослеживается в гэллаповской традиции массовых опросов [Докторов, 2013].

Появление новых средств регистрации и хранения результатов наблюдений позволило не только собирать петабайты данных, но и анализировать их с помощью автоматизированных средств. Именно этот факт стал широко известен как Big Data. Примененная к изучению общества, эта форма организации наблюдения и анализа обладает двумя преимуществами — оперированием популяциями, а не выборками ($n=All$) и автоматизированной регистрацией поведения, а не опосредованного опросным инструментарием мнения. «С большим объемом потенциально доступных данных не придется жертвовать размером для лучшего качества набора данных и наоборот»¹.

Таким образом, Big Data стали одним из факторов, поколебавших легитимность опросных методик как доминирующих в изучении социальной реальности. Кроме того, если прежде, как подчеркивает С. Лионелли, исследователи и те, кто пользуются их результатами, больше внимания уделяли получаемому продукту, то теперь они вынуждены учитывать и процессы, посредством которых эти результаты были получены [Leonelli, 2014].

Сущностные характеристики Big Data

Что же делает так называемые Big Data настолько уникальными? Р. Китчин утверждает, что как феномен Big Data появились в рамках 3D-моделирования, их основные характеристики (3V) были выделены Д. Лейни в 2001 г.²:

- volume (размер): огромные по объему, состоят из терабайтов и петабайтов данных;
- velocity (скорость): высоко динамичные, создаются в реальном времени или близком к нему;
- variety (разнообразии): различающиеся по типам данных, структурированные и неструктурированные, часто связаны с временем и местом (объединены метаданными).

С более широким распространением Big Data, в том числе в области генетики, здравоохранения и исследования общества, ученые стали добавлять новые сущностные характеристики Big Data:

¹ Manovich L. Trending: The Promises and the Challenges of Big Social Data [Электронный ресурс]. 2011. URL: manovich.net/content/04-projects/067-trending-the-promises-and-the-challenges-of-big-social-data/64-article-2011.pdf (дата обращения: 16.12.2016).

² Laney D. 3D data management: Controlling datavolume, velocity and variety. In: Meta Group. [Электронный ресурс]. 2001. URL: <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf> (дата обращения: 16.12.2016).

- достоверность и ценность³;
- изменчивость⁴, которая предполагает, что данные изменчивы и их интерпретация существенно зависит от контекста;
- исчерпывающая полнота (exhaustivity) — $n = \text{all}$ [Майер-Шенбергер, Кукьер, 2014];
- «высокое разрешение» (fine-grained in resolution) и однозначная индексация идентификаторами (uniquely indexical identification) [Dodge, Kitchin, 2005], то есть все данные обладают уникальными идентификаторами, позволяющими приписать их конкретным объектам наблюдения (например, ID пользователей), что позволяет масштабировать и реляционно объединять данные;
- реляционность [Boyd, Crawford, 2012], то есть наличие у различных баз данных общих полей, которые позволяют их объединять в единые массивы;
- расширяемость и масштабируемость [Marz, Warren, 2015] (scaleability) предполагают, что в зависимости от задач исследования массив данных можно расширять за счет присоединения новых массивов, с другой стороны, это дает возможность сузить фокус исследовательского внимания до малой социальной группы и индивида.

Отсутствие выборок и непосредственная работа с генеральными совокупностями или очень крупными их частями, масштабируемость данных, постоянный автоматизированный сбор данных в архивы и возможность их быстрой обработки приводят в конечном итоге к высокой достоверности и востребованности прогнозирования «в реальном времени» на основании Big Data. Именно это, учитывая постоянную критику прогнозирования, базирующегося на массовых опросах, делает Big Data все более привлекательными для исследования общества.

Позже исследователи стали выделять их негативные черты, в том числе с этической точки зрения. Так, И. Апричард указывает на бесполезность, вульгарность и «нарушающий» (violating) характер Big Data⁵. А Д. Лаптон⁶ добавляет такие характеристики, как порочность, провокативность, включенность в частную жизнь и т. д.

Тем не менее Big Data стали позиционироваться как идеал получения, хранения и обработки информации об обществе, поставив перед опросами общественного мнения задачу соотносить себя с ними. Кроме того, их появление в социальных науках стало тем водоразделом, за которым все традиционные методики получения и обработки информации об общественном мнении стали именоваться не иначе как small data. Различия между small data и Big Data на примере данных исследований, административной статистики и собственно «больших данных» проанализировали Р. Китчин и Г. МакАрдл [Kitchin, 2014; Kitchin R., McArdle, 2016].

³ Marr B. Big data: The 5 vs everyone must know. March 6. [Электронный ресурс]. URL: <https://www.linkedin.com/pulse/20140306073407-64875646-big-data-the-5-vs-everyone-must-know> (дата обращения: 16.12.2016).

⁴ McNulty E. Understanding Big Data: The seven V's. 2014. 22May. [Электронный ресурс]. URL: <http://dataconomy.com/seven-vs-big-data/> (дата обращения: 04.09.2016).

⁵ Uprichard E. Big data, little questions. Discover Society, 2013. 1 October. [Электронный ресурс]. URL: <http://discoversociety.org/2013/10/01/focus-big-data-little-questions/> (дата обращения: 17.09.2016).

⁶ Lupton D The thirteen Ps of big data. The Sociological Life, 2015. 13 May. [Электронный ресурс]. URL: <https://simplysociology.wordpress.com/2015/05/11/the-thirteen-ps-of-big-data/> (дата обращения: 16.12.2016).

В данный момент условно можно выделить четыре стратегии ответа социальных наук на появление Big Data, которые могут быть использованы и социологией общественного мнения:

1. критика и поиск уязвимостей с Big Data;
2. сотрудничество с Big Data;
3. изучение Big Data как особой техники;
4. игнорирование Big Data.

Первая и одна из самых очевидных стратегий реакции на любое нововведение, подрывающее устоявшийся status quo, является агрессия. Отчасти популярность данной позиции в отечественной социологии может также объясняться «методологическими травмами социолога», на которые обратила внимание Г. Г. Татарова [Татарова, 2006], в том числе игнорированием значительного математического аппарата и замена его анализом простых распределений.

Критика и поиск уязвимостей

Д. Бойд и К. Кроуфорд выявили несколько значимых позиций для критики Big Data, связанных с их сущностными характеристиками [boyd, Crawford, 2012]. Специфика регистрации данных, на первый взгляд не связанной с деятельностью наблюдателя, но совершаемой в автоматическом режиме, а также их исчерпывающий объем ($n = All$), создают иллюзию объективности получаемого знания. В действительности в центре обработки полученных данных находится не технологизированная «добыча данных» (data-mining), а интерпретация найденных закономерностей (собственно, data-analysis), субъективная по своей природе и не зависящая от объема обработанных данных. Кроме того, как отмечают Д. Бойд и Кроуфорд, больше данных не всегда значит лучше — Big Data вовсе не обязательно являются всеми данными о предмете изучения. К этому следует добавить и замечание Р. Китчина о том, что исчерпывающий объем как требование Big Data вынуждает собирать огромное количество информации, в которой вообще может не оказаться искомым корреляций [Kitchin, 2014: 73].

Последний критический аргумент, который приводят Д. Бойд и К. Кроуфорд, — контекстуальность как сбора, так и анализа Big Data. Вне контекста собранные данные попросту теряют свой смысл.

Дж. Фан, Ф. Хан и Х. Лю также обращают внимание на проблемы Big Data, связанные с их размерами и большим количеством измерений. Отдельно ими были выделены три типа проблем [Fan, Han, Liu, 2014; Fan, Liao, 2014]:

1. накопление шумов (Noise Accumulation);
2. ложные корреляции (Spurious Correlation);
3. побочная эндогенность (Incidental Endogeneity).

Первый тип проблем связан с тем, что при большом количестве измерений, применяемых в больших данных, наличие шума даже в одной из переменных существенно искажает результаты обработки данных. Как убедительно показывают авторы, такое накопление шумов заметно влияет на результаты даже в случае ошибки в простой бинарной классификационной (номинальной) переменной. Второй тип проблем также обусловлен большим количеством переменных, используемых в Big Data. Дж. Фан, Ф. Хан и Х. Лю поставили математический экс-

перимент, который убедительно продемонстрировал, что вычисление корреляций при большом количестве переменных может подвергаться существенным искажениям. Побочная эндогенность является обратной стороной ложных корреляций и не позволяет подтвердить реальную связь между признаками. Она возникает по двум причинам:

1. по мере накопления максимального количества случаев измерения увеличивается и возможность того, что часть из них связана с остаточным, случайным шумом (ошибками);
2. «Big Data как правило агрегируют из различных источников с потенциально разными схемами генерирования данных. Это повышает возможность смещения выборки и ошибок измерения, которые также повышают потенциал побочной эндогенности» [Fan, Han, Liu, 2014]. Действие указанных проблем ярко демонстрирует ставшая известной общественности ошибка Google в прогнозировании эпидемии гриппа (Google Flu Trends). Д. Лазер с коллегами нашли два основных типа ошибок, которые подвели прогноз Google [Lazer et al. 2014]. Как показала практика, при поисковых запросах по ключевым словам (например, «кашель», «лихорадка») поисковик Google перенаправлял пользователя к результатам запросов по симптомам гриппа и их лечения. То есть Google сам увеличивал вероятность получения прогноза о возникновении эпидемии гриппа. Успешные прогнозы 2009—2011 гг. фактически сформировали у поисковых алгоритмов Google ошибку, которая в классических опросах общественного мнения считается систематической и связывается с личностью интервьюера. Это указывает и на еще одну техническую проблему Big Data — проблему измерения и репрезентации. Алгоритм Google не только собирал и анализировал Big Data, но и непосредственно конструировал их. Этот кейс позволил утверждать авторам доклада AAPOR, что «большое количество данных не должно приводить к игнорированию фундаментальных вопросов измерения и конструированию надежности и валидности...» [Джапек и др. 2015: 28—29]. Таким образом, большие объемы не способны компенсировать ошибки, отклонения и недостаток отдельных параметров данных.

Сходная проблема обнаруживается М. Вилсоном в сборе Big Data в социальных сетях. Он отмечает, что социальные медиа не обязательно свидетельствуют о социальных феноменах, протекающих вне цифровой реальности, социальные медиа сами являются феноменами [Wilson, 2015: 348].

Таким образом, на данный момент критику технической составляющей Big Data можно связать с проблемами:

1. сбора и регистрации данных (в том числе связанными с различиями в схемах конструирования данных и самих процедур измерения, контекстуальной зависимости данных, наличием в них шумов и ошибок);
2. первичного анализа и поиска корреляций (data-mining), алгоритмы которого при большом количестве переменных могут давать заметные искажения;
3. анализа данных (data-analysis), который продолжает зависеть не только от контекста, но и от интерпретации data-аналитика.

В сущности, это демонстрирует то, что Big Data отягощены теми же проблемами, что и массовые опросы. Как отмечают эксперты AAPOR, «эти вопросы долж-

ны заботить тех, кто занимается анализом больших данных даже в том случае, если данные можно считать достоверными. Ошибки, не связанные с выборкой (nonresponse error), только усугубляют эти проблемы» [Джапек и др. 2015: 30—31]. Однако эти вопросы являются техническими и ответ на многие из них наверняка будет дан раньше или позже. Напротив, этические вопросы, связанные с Big Data, представляются наиболее уязвимым местом для внешней критики, защищающей сложившийся в изучении общества status quo.

Все сомнения в этичности Big Data можно сгруппировать в следующие блоки:

1. нарушение приватности,
2. отслеживание поведения по «цифровым следам» и в реальном времени,
3. закрытость Big Data и «цифровое неравенство».

Е. В. Карчагин на основании анализа широкого круга критических статей выделяет следующие этические вопросы к Big Data: приватность, слежение, упреждающее слежение, открытость, справедливость, цифровое неравенство [Карчагин, 2016].

Проблема нарушения приватности кроется в базовых свойствах Big Data — «высоком разрешении», которое выделяет Р. Китчин, и реляционности. Дело в том, что базы данных в Big Data объединяются преимущественно с использованием тех или иных форм персональных идентификаторов (ID). При этом считается, что субъекты, обладающие такими данными, не станут пользоваться ими для нарушения вашей приватности и слежки, ограничивая себя исключительно аналитикой поведения больших групп населения. Хотя никакого ограничения кроме доброй воли исследователя, имеющего возможности Big Data, фактически не существует. В действительности действие Big Data — это в некотором роде «вглядывание» в поведение каждого человека, который оставил свои «цифровые следы» в больших массивах данных. Как справедливо отметили Д. Бойд и К. Кроуфорд, «существует значительная разница между тем, чтобы быть на публике (то есть сидеть в парке) и быть публичным (то есть активно привлекать внимание)» [Boyd, Crawford, 2012].

Именно повышение заметности делает Big Data ценным инструментом наблюдения, особенно для институтов национальной безопасности и власти вообще. Самый шумевший опыт разоблачения использования Big Data для слежки в интересах властных структур связан с Э. Сноуденом. «То, что пользователи по неведению раскрывают на таких платформах, как Facebook и Twitter, а также используя свои телефоны, дает ценные данные для «национальной безопасности» или полицейских задач», — отмечает Д. Лион [Lyon, 2014: 3]. В российской практике слежения и отслеживания существует Система оперативно-розыскных мероприятий (СОРМ). Если СОРМ-2 отслеживала трафик основных провайдеров, то, по предположению разработчиков, СОРМ-3 будет иметь возможность сохранять трехлетний архив всего биллинга и всех логов соединений пользователей на территории Российской Федерации.

Как считают критики, приоритет национальной безопасности над приватностью приведет к значимым социально-политическим последствиям [Lyon, 2014: 4]. Воздействие практик Big Data на плебисцитарную демократию продемонстрировал журнал Das Magazin, изучивший роль компании Cambridge

Analytica в Брекзите и выборах Д. Трампа в 2016 г.⁷ Журналистское расследование показало, что применение психометрических шкал к цифровым следам, регистрируемым Big Data, способно быть продуктивным при манипулировании общественным мнением.

Важнейшей проблемой использования Big Data является цифровое неравенство, которое кроется как в неравном доступе к данным, так и в различии знаний и возможностей информационных технологий [Жарчагин, 2016: 56].

Л. Манович, говоря об отношении субъектов к Big Data, выделил три «дата-класса»: создатели данных (как правило, простые пользователи, которые генерируют данные сознательно и как «цифровые следы»); те, кто имеет средства собирать данные; те, кто имеет опыт анализа Big Data⁸. Преимущества от использования Big Data неравномерно распределены между этими тремя классами. «Создатели данных» могут получать все более и более совершенные услуги, ориентированные на их собственные предпочтения, но у них нет ни технических, ни технологических возможностей двух других классов. В результате Big Data разделяют общество на тех, кто способен полностью пользоваться их результатами (включая прогностический потенциал и манипулирование поведением больших групп населения), и тех, кто вынужден оставаться объектом этого пристального взглядывания со стороны коммерческих и государственных «больших братьев».

Именно поэтому часть исследователей утверждает, что Big Data должны существовать только в случае всеобщего равного доступа к их потенциалу, что предполагает также широкое распространение информационных компетенций среди исследователей и населения. Так, Н. Кшетри полагает, что «равенство доступа является необходимым, но не достаточным условием, чтобы привести к цифровому равенству. Такие факторы, как наличие навыков и возможностей, социальные и культурные установки в отношении технологии, институциональная среда и социальные преобразования, тесно связаны с возможностью использовать технологию эффективно и правильно» [Kshetri, 2014].

Сотрудничество

Коммерческий опыт использования Big Data и последние случаи их применения к решению политических задач побудили исследователей искать точки соприкосновения с ними традиционных методик изучения общества. На данный момент эффективные практики такого взаимодействия прослеживаются по трем основным направлениям:

1. Использование Big Data для исследования традиционных областей интересов социальных наук.
2. Дополнение результатов применения Big Data традиционными социологическими методиками (small data).

⁷ Grassegger V. H., Krogerus M. Ich habe nur gezeigt, dass es die Bombe gibt [Электронный ресурс] // Das Magazin №48-3. Dezember 2016. URL: <https://www.dasmagazin.ch/2016/12/03/ich-habe-nur-gezeigt-dass-es-die-bombe-gibt/> (дата обращения: 16.12.2016).

⁸ Manovich L. Trending: The Promises and the Challenges of Big Social Data [Электронный ресурс]. 2011. URL: manovich.net/index.php/projects/trending-the-promises-and-the-challenges-of-big-social-data (дата обращения: 16.12.2016).

3. Применение механизмов Big Data к собранным традиционными социологическими методиками данным.

Первый путь взаимодействия очевиден и, учитывая «исчерпывающую природу» Big Data, является более эффективным основанием для реализации индуктивного способа познания общества. Он состоит в том, чтобы, как предлагает вице-президент Google и профессор Стэнфорда П. Рагхаван, взять «большие проблемы» социологии и решить их с помощью «больших данных», чтобы получить действительно интересные социологические идеи, которые были бы статистически надежными⁹. Однако для социологии общественного мнения это решение фактически означает капитуляцию и сдачу предметного поля. На данный момент наиболее продуктивно именно взаимодействие «качественных» социологических методов и Big Data.

Big Data представляет собой средство поиска значимых корреляций признаков без привлечения к этому процессу процедуры выдвижения гипотез. При этом обнаруженные корреляции могут быть достаточно неожиданными как в случае с обнаружением связи между лайками постов особого вида картошки фри (curly fries) и более высокого IQ, которые обнаружили исследователи Кембриджского университетского проекта во главе с М. Косински. Проблемой статистической корреляции признаков является то, что она не указывает на детерминацию этих признаков. То есть, используя исключительно статистику, мы не можем сказать, повлияла ли картошка на высокий IQ или, напротив, высокий IQ на выбор картошки, либо существует третий признак, который обуславливает и выбор картошки, и высокий IQ. Так или иначе, для прояснения детерминации необходимо привлечение стороннего опыта.

Х. Форд, рассматривая Big Data в разрезе социальных сетей, утверждает, что этнографы могут получать достаточно интересные данные о том, каким образом люди взаимодействуют в сложных медиа-платформах. «Исследователи данных, в свою очередь, могут извлечь выгоду из более качественного понимания последствий отсутствия данных, неполноты данных и социальных значений, присваиваемых данным. Работая вместе, этнографы и специалисты по обработке данных могут не только производить строгие исследования, но также могут найти способы диверсификации своих исследовательских навыков» [Ford, 2014: 3]. Достаточно известны примеры применения этнографических методов в качестве дополнения Big Data в компании Adidas¹⁰, а также использования психологических измерений в Cambridge Analytica. Д. Мерфи и С. Боуман предлагают предоставлять услуги анализа данных тем компаниям, которые уже работают с Big Data, но ограничены во времени и в экспертных возможностях [Murthy, Bowman, 2014]. Но этот путь коллаборации в сущности не охватывает традиционные практики изучения общественного мнения.

Идея объединения получаемых социологией данных в более крупные выборки достаточно очевидна. Отличным примером того, как масштабы влияют на глубину

⁹ Mann R. Five minutes with Prabhakar Raghavan: Big data and social science at Google. Impact of Social Sciences, London School of Economics and Political Science [Электронный ресурс]. 2012. URL: <http://eprints.lse.ac.uk/52128/> (дата обращения: 16.12.2016).

¹⁰ Baer D. Here's Why Companies are Desperate to Hire Anthropologists [Электронный ресурс] // Business Insider. 2014. URL: <http://www.businessinsider.fr/us/heres-why-companies-aredesperateto-hireanthropologists-2014-3/> (дата обращения: 16.12.2016).

анализа получаемых данных, служат крупные кросс-культурные проекты изучения ценностей World Values Survey и European Values Survey. Следует помнить, что они объединены единой методикой и методологией исследования, в том числе интерпретацией и операционализацией основных понятий (насколько это возможно в рамках перевода вопросов анкеты). Р. Китчин указывает на то, что более ценными, чем чистые Big Data могут быть интуиции (insights), полученные из small data путем увеличения их масштаба до более крупных наборов данных, что сделает их более доступными для повторного использования с помощью инфраструктуры цифровых данных [Kitchin, 2014: 46].

В современной биологии такое «укрупнение» small data происходит при расшировке генома человека. Механизм этой интеграции для решения более широкого спектра задач фундаментально раскрыт Д. Лионелли [Leonelli, 2014]. Решения Big Data достаточно широко применяются в корпусных исследованиях в лингвистике (в особенности в таких крупных коллекциях текстов, как Google Books). Также в пример можно привести использование новых методов интеграции и обработки массивных наборов данных о культуре в лаборатории Л. Мановича¹¹.

Объединение small data в массивы больших размеров, хотя бы частично соответствующие критериям Big Data, обуславливается стремлением повторно вовлечь прежде собранные данные в научный оборот, а также получить новые, не определяемые в каждом отдельном массиве корреляции.

Значимым на пути интеграции данных в социологии общественного мнения является проект, осуществленный Д. Рогозиным и Д. Сапоновым [Рогозин, Сапонов, 2014]. В рамках указанного проекта базы данных телефонных опросов, проведенных ФОМ, были реляционно интегрированы для поиска влияния на результаты опросов важнейшей латентной переменной — ошибки. То есть была осуществлена попытка решить сразу две методологические сверхзадачи — объединить массивы в более крупный и единый набор данных и обнаружить ошибку, которая не локализована ни в одном из массивов. Фактически примененные авторами механизмы позволили проверить согласованность данных множества массивов — в перспективе этот механизм можно использовать для машинной проверки достоверности социологических исследований.

Объединением разнородных баз данных социологических опросов также занимается коллектив исследователей «Волгоградского проекта». В рамках его работы был обнаружен ряд проблем интеграции классических социологических small data в большие объемы данных [Одинцов, 2015]. Их можно объединить в несколько групп: проблемы, связанные с использованием выборочного метода и анонимностью опроса; проблемы, обусловленные неоднозначностью операционализации понятий и измерения исследуемых свойств. Однако, как показывает практика реализации проекта, они имеют решение.

За пределами статьи остаются две стратегии: игнорирование Big Data как феномена, угрожающего традиционным методам социологии; изучение Big Data в качестве еще одного предмета социологии.

¹¹ Manovich L. The Science of Culture? Social Computing, Digital Humanities and Cultural Analytics [Электронный ресурс]. 2016. URL: http://manovich.net/content/04-projects/088-cultural-analytics-social-computing/cultural_analytics_article_final.pdf (дата обращения: 16.12.2016)

Игнорирование как стратегия ответа на вызов больших данных, вполне возможно, действительно эффективно. Как показывает теория игр, замирание жертвы нередко является успешной тактикой в присутствии хищника. Однако само применение данной стратегии предполагает отсутствие документированных ответов в виде научных статей или выступлений, что делает рассмотрение этой стратегии непродуктивным.

Напротив, рассмотрение Big Data в качестве объекта социологии, в особенности подхода «наука — технология — общество» (Science-Technology-and-Society approach, STS), — крайне перспективная, но все еще недостаточно часто реализуемая стратегия. Как правило, сторонники ее применения ограничиваются критикой М. Фуко, пускай и вполне обоснованной, игнорируя обыденные практики, на которых строится данная индустрия, роли в ней конкретных исследователей. Чрезмерная уверенность в объективности Big Data исчезает, если принять во внимание то, что и измерение, и регистрация данных продолжают зависеть от схемы сбора данных. При этом на саму обработку данных существенно влияют аналитики, выбранные ими инструменты и контекст, в котором осуществляется анализ.

Выводы

Ранее у института науки не существовало достаточных мощностей для регистрации и обработки данных, необходимых для осуществления гносеологически оправданной индукции (по крайней мере, в идеале, заложенном Ф. Бэконом). В естественных науках этот факт компенсировался применением к эмпирическому наблюдению индукции, сосредоточенной в процедурах выдвижения научных гипотез.

В знании об обществе даже переписи населения при всей их затратности, можно было считать только отдельными срезами социальной реальности, с крайне ограниченным потенциалом обработки и дальнейшего анализа. Временным решением проблемы недостаточности данных при относительно эффективной обработке стало применение выборочного метода и механизмов статистического анализа. Именно это позволило социологии общественного мнения на достаточно долгое время стать (или просто временно выглядеть) наиболее эмпирически достоверным знанием об обществе, которое можно распространить на крупные группы населения.

Big Data как специфический набор регистрации, обработки и анализа данных претендует на то, чтобы перехватить пальму первенства у социологии общественного мнения на право конструировать максимально полное (на текущий момент) описание общественной реальности. Этому способствуют ключевые характеристики Big Data, например: работа непосредственно с генеральной совокупностью ($n=All$), автоматическая регистрация данных, регистрация поведения, а не мнения или установок общественного мнения. При этом Big Data — это не просто очередное проявление «веры в числа», свойственной и социологии общественного мнения, это вызов самой природе данных, которые прежде собирались социологией. И на этот вызов, как было указано выше, можно ответить четырьмя способами: критиковать, выявляя уязвимые моменты Big Data, пытаться задействовать их наряду с традиционными методами, изучать их, как особую технику (в рамках

STS, антропологии и этнографии), или игнорировать их существование. Стратегии изучения Big Data как особой техники и их игнорирования нельзя признать продуктивными, так как суть вызова Big Data состоит в претензии на более полное и достоверное описание социальной реальности, чем то, которым обладают традиционные социологические данные (small data).

Критика Big Data осуществляется по двум основным направлениям: техническому (критика процедур регистрации и сбора данных, изучение проблем первичного анализа и поиска корреляций, а также интерпретации найденных закономерностей) и этическому (проблема нарушения приватности, отслеживание поведения по параданным и «цифровым следам», закрытость процедур Big Data и «цифровое неравенство»).

Взаимодействие с Big Data наиболее продуктивно в использовании Big Data для решения традиционных проблем социологии, дополнении Big Data качественными социологическими методами, применении решений Big Data к тем данным, которые уже собраны и собираются традиционными средствами социологии. Наиболее интересны попытки объединить имеющиеся у социологии small data в крупные массивы и работа с ними средствами, используемыми «большими данными» (проекты Д. Рогозина и Д. Сапонова, а также «Волгоградский проект»).

Список литературы (References)

Бэкон Ф. Новый органон XII—XV. Соч. в двух томах. М. : Мысль. 1978. Том 2. [Bacon F. (1978) The new organon XII—XV. Collected Works in 2 volumes. Moscow, Izdatelstvo «Mysl'». Vol. 2. (In Russ.)]

Джапек, Л., Крейтер, Ф., Берг, Мю. и др. Отчет AAPOR о больших данных: 12 февраля 2015 / Американская ассоциация исследователей общественного мнения ; пер. с англ. Д. Рогозина, А. Ипатовой, Е. Вьюговской ; предисловие Д. Рогозина. М., 2015. [Japac L., Kreuter F., Berg M., and others. (2015) AAPOR Report on Big Data. February 12, 2015. American Association for Public Opinion Research. Moscow. (In Russ.)]

Докторов Б. З. От соломенных опросов к постгэллоповским опросным методам. М. : Радуга, 2013. [Doktorov B. Z. (2013) From straw polls to post-Gallup technologies. Moscow: Raduga. (In Russ.)]

Карчагин Е. В. Справедливость в информационной области: этический аспект Big Data // Вестник Волгоградского государственного университета. Серия 7. 2016. № 1. С. 50—58. [Karchagin E. V. (2016) Justice in the field of information: the ethical aspect of Big Data. *Bulletin of Volgograd State University*. Series 7. No1. P. 50—58. (In Russ.)]

Майер-Шенбергер В., Кукьер К. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим / пер. с англ. Инны Гайдюк. М. : Манн, Иванов и Фербер, 2014. [Mayer-Schonberger V., Cukier K. (2014) Big Data. A revolution that will transform how we live work and think. Moscow: Mann, Ivanov i Ferber. (In Russ.)]

Одинцов А. Проблемы объединения разнородных баз данных в социологии // Известия ВолГТУ. 2015. № 9. С. 72—77. [Odintsov A. (2015) Problems of data integration of heterogeneous sociological research. *Izvestia VSTU*. No9. P. 72—77. (In Russ.)]

Рогозин Д. М., Сапонов Д. И. Корпус ошибок автоматизированного телефонного опроса // Полития: Анализ. Хроник. Прогноз. 2014. № 4. С. 145—161. [Rogozin D. M., Saponov D. I. (2014) Errors in Automated telephone surveys. *Journal of Political Philosophy and Sociology of Politics*. No 4, P. 145—161. (In Russ.)]

Татарова Г. Г. Методологическая травма социолога. К вопросу интеграции знания // Социологические исследования. 2006. № 9. С. 3—12. [Tatarova G. G. (2006) Methodological trauma of the sociologist: on the integration of knowledge. *Sociological studies*. No 9. P. 3—12. (In Russ.)]

Boyd D., Crawford K. (2012) Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*. Vol. 15. No 5. P. 662—679.

Dodge M., Kitchin R. (2005) Codes of life: Identification codes and the machine-readable world. *Environment and Planning D: Society and Space*. Vol. 23. No 6. P. 851—881. DOI: 10.1068/d378t.

Fan J., Han F., Liu H. (2014) Challenges of Big Data analysis. *National Science Review*. Vol. 1. No 2. P. 293—314. ULR: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4236847/> (accessed 16.12.2016).

Fan J., Liao Y. (2014) Endogeneity in ultrahigh dimension. *Annals of Statistics*. Vol. 42. No 3. P. 872—917. DOI: 10.1214/13-AOS1202.

Ford H. (2014) Big Data and Small: Collaborations between ethnographers and data scientists. *Big Data & Society*. Vol. 1. No 2. P. 1—3. DOI: 10.1177/2053951714544337.

Kitchin R. (2014) The Data Revolution. Big Data, Open Data, Data Infrastructures & Their Consequenses. Los Angeles, London, Singapore, Washington DC: SAGE.

Kitchin R., McArdle G. (2016) What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*. Vol. 1. No 10. DOI: 10.1177/2053951716631130.

Kshetri N. (2014) The emerging role of Big Data in key development issues: Opportunities, challenges, and concerns. *Big Data & Society*. P. 1—20. DOI: 10.1177/2053951714564227.

Lazer D. M., Kennedy R., King G., Vespignani A. (2014) The parable of google flu: Traps in big data analysis. *Science*. No 343 (6176). P. 1203—1205.

Leonelli S. (2014) What difference does quantity make? On the epistemology of Big Data in biology. *Big Data & Society*. P. 1—11. DOI: 10.1177/2053951714534395.

Lyon D. (2014) Surveillance, Snowden, and Big Data: Capacities, consequences, critique. *Big Data & Society*. P. 1—13. DOI: 10.1177/2053951714541861.

Marz N., Warren J. (2015) Big Data: Principles and Best Practices of Scalable Realtime Data Systems. MEAP edition. Westhampton, NJ: Manning.

Murthy D., Bowman S. (2014) Big Data solutions on a small scale: Evaluating accessible high-performance computing for social research. *Big Data & Society*. P. 1—12. DOI: 10.1177/2053951714559105.

Wilson M. (2015) Morgan Freeman is dead and other big data stories. *Cultural geographies*. Vol. 22. No 2. 345—349. DOI: 10.1177/1474474014525055.