

ПРОИЗВОДСТВО НАУЧНОГО ЗНАНИЯ

ДЕНИС ЮРЬЕВИЧ СИВКОВ

кандидат философских наук,
доцент кафедры философии и социологии
Волгоградского института управления РАНХиГС
Волгоград, Россия;
e-mail: d.y.sivkov@gmail.com



УДК 303.04

Большие данные в этнографии: вызовы и возможности¹

Появление феномена больших данных в маркетинге и сфере управления стало серьезным вызовом для социальных наук и заставило социальные науки переоценить свои возможности. В «эмпирической эпистемологии» считается, что «механическая объективность» устраняет субъекта-исследователя и оставляет принятие решений техническим специалистам. В этом смысле этнография и качественные исследования, кажется, вступают в противоречие с методами разработки больших данных. В корпоративных исследованиях этнографический маркетинг показал ограниченность подходов, основанных на машинном обучении. Помимо противопоставления этнографии и методов машинного обучения, более перспективной видится коллаборативная работа, в которой возможно реляционное объединение и корреляция больших и малых данных, поиск теоретических рамок. Большим данным необходимы «маленькие паттерны», определяющие место, объект и способ поиска данных. Малые данные этнографии также уточняют определение и применение больших данных. С одной стороны, этнография позволяет быть более внимательным к малым и исключенным группам пользователей, которые оставляют специфические цифровые следы, но в общем контексте остаются незамеченными (Б. Уеллес). С другой стороны, исследования больших данных помогают уточнить и проблематизировать традиционные качественные понятия, например, понятие контекста (Н. Сивер).

Ключевые слова: данные, большие данные, этнография, антропология, качественные исследования, количественные исследования, коллаборация, контекст, непредставленные сообщества.

¹ Статья подготовлена при поддержке гранта РФФИ № 15-06-02758 «Реляционное объединение данных разнородных социологических исследований в единый массив и интеллектуальный анализ с применением технологий машинного обучения». Текст был представлен в виде доклада на воркшопе «Большие данные в социальных науках» в рамках Волгоградского проекта по большим данным.

Введение

В конце первого десятилетия XXI в. в корпоративном секторе появился интерес к информации, оставляемой пользователями в виде цифровых следов в социальных сетях, при посещении сайтов, активности в социальных сетях и т. п., которые хранятся в виде данных на сервисах компаний и учреждений. Оказалось, что эти данные можно более или менее успешно использовать для анализа и предсказания поведения пользователей. Объемы, выборка ($n=all$), непосредственность доступа и сбора, скорость получения и обработки обозначили эти данные как большие данные (big data).

Появление больших данных в корпоративном секторе также привело к изменениям в социальных науках. Например, успехи в анализе и прогнозировании социальных действий с помощью машинного обучения поставили под вопрос применимость традиционных методов в социологии и необходимость обращения к проблематике больших данных и машинного обучения [Burrows and Savage, 2014]. В широком контексте большие данные — это своего рода фон, на котором все социальные науки пытаются проблематизировать собственные основания: «Большие данные переосмысливают ключевые вопросы о конституции знания, процессах исследования, о том, как мы должны взаимодействовать с информацией, о природе и о категоризации реальности» [Boyd and Crawford, 2012, p. 665].

Большие данные заставили социальных ученых поставить вопрос: «что такое данные?», поскольку в течение длительного времени *data* были фигурой умолчания и считались чем-то само собой разумеющимся. В этой связи на фоне больших данных через различные метафоры уточняется природа качественных данных: «малые данные» (small data) [Kitchin, 2015], «насыщенные данные» (thick data) [Wang, 2013], «глубокие данные» (deep data) [Brock, 2015], «культурные данные (cultural data) [Curran, 2013]. Как отмечает Э. Рупперт, «так как эмпирические социальные науки переживают кризис, большие данные представляют и вызов, и возможность» [Ruppert, 2013, p. 273]. В статье предполагается выявить и обсудить коллаборативный потенциал этнографии (и шире — качественных исследований вообще) и больших данных, понять, какие вызовы исходят от больших данных и какие возможности они могут предоставить этнографии. Под этнографией в данном случае понимается обозначение междисциплинарных подходов, берущих начало в социальной и культурной антропологии.

Этнография и квантификация

Общим местом в антропологии и качественных исследованиях является акцент на «насыщенном описании» (thick description), данные которого не квантифицируются вообще или квантифицируются с большим трудом. С. Э. Мерри показывает, что так называемые объективные числовые показатели, используемые различными организациями, например Комитетом ООН по правам человека, существенно расходятся с опытом женщин, пострадавших от домашнего насилия, выраженным в этнографических интервью [Merry, 2016].

Кажется, что большие данные и этнография находятся на разных полюсах, и этнографы только могут изучать, как появляются большие данные в контексте

деятельности программистов и маркетологов. Однако все не так однозначно. В начале антропологии у Тайлора и особенно у Малиновского стоял вопрос об использовании статистических данных в этнографии. Этот вопрос тесно связан с требованием научной объективности антропологических изысканий.

В ставшем классическим введении к «Аргонавтам западной части Тихого океана» Малиновский для построения наиболее полной картины примитивного общества предлагал использовать разнородные данные: статистические таблицы, дневниковые записи наблюдений, а также записи мифов, магических формул и высказываний информантов [Малиновский, 2004, с. 42]. Методологический проект Малиновского предполагает объединение массивов разнородных данных: «важно, что Малиновский не пытался создать только этнографические описания, его главной задачей было развитие жизнеспособной модели общества» [Curran, 2013, p. 69].

В этом смысле этнография — это всегда напряжение между поэзией насыщенного описания и квантификацией строгой науки. Антропология балансирует между не-репрезентативными техниками и «доверием к числам» (trust in numbers). Доверие к числам появляется как стремление избежать индивидуальности и субъективности в исследовании [Porter, 1995], а не-репрезентативная этнография, напротив, является субъективной и не заботится об «обезличенной, нейтральной и надежной манере» исследования [Vannini, 2015, p. 318].

Также очевидно, что чем более «насыщенным» является описание какой-либо ситуации, тем больше данных оно будет содержать (хотя бы текстовых результатов транскрибирования, дневниковых записей и т. п.). В нулевых годах наметилась тенденция квантификации качественных данных и использования дополнительных количественных данных для анализа результатов полевых исследований [Angrosino, 2007]. Более того, глобальные процессы дигитализации превращают всех качественников в цифровых гуманитариев, которые «собирают, находят, организуют и анализируют; подробно описывают, аннотируют, цитируют, размышляют о, представляют друг другу и архивируют исследовательские данные и материалы» [Antonijevic, 2015, p. 37]. Активное использование цифровых методов и инструментов существенно изменило исследовательскую интенцию: теперь даже на персональном уровне данных больше и масштаб их анализа изменился. «Дигитализация процесса исследования в социальных и гуманитарных науках приводит к тому, что ученые в разных областях все больше приходят к осознанию, что они имеют дело с данными» [ibid, p. 125].

Конфликт: big versus small

При каталогизации отношении *big data* и этнографии, пожалуй, первое, что приходит на ум, это противопоставление их методологий и эпистемологий. Большие данные делают то, что не могут «схватить» этнографы, и наоборот: этнография способна видеть вещи, недоступные алгоритмам машинного обучения [Curran, 2013].

Появление аналитики больших данных было отмечено оптимизмом технических специалистов и корпоративного сектора, так как отпадала необходимость в социальных и иных исследователях для понимания тех или иных процессов и явлений: «Так, поисковая система Google определяет распространение гриппа, не хуже,

чем официальная статистика, основанная на реальных визитах врачей к пациентам» [Майер-Шенбергер и Кукьер, 2014, с. 18]. Теперь информационные технологии не только не нуждаются в услугах науки, но и наоборот: «самое время спросить: чему наука может научиться у Google» [Anderson, 2008].

На стыке социальных и технических наук появляется специфическая «эмпирическая методология», согласно которой «большие данные могут охватить целую область и обеспечить высокое разрешение; нет необходимости в теории, модели или гипотезе a priori; в использовании агностической аналитики данные могут говорить сами за себя без какого-либо человеческого влияния или конструирования, любые паттерны и отношения внутри больших данных по существу наполнены смыслом и истиной; смысл выходит за пределы контекста знания, характерного для данной области и поэтому могут быть проинтерпретированы тем, кто сможет декодировать статистическую или данных визуализацию» [Kitchin, 2014, p. 4].

Эмпирическая методология также называется датаизмом (dataism) — это «тенденция высоко оценивать технически генерированные и отобранные данные, рассматривать их как более объективные и поэтому более надежные, делая теорию устаревшей» [Lohmeier, 2014, p. 79]. Таким образом, с подачи программистов появляется идея упразднения всех социальных наук и программ исследования, в том числе и этнографии. В «эмпирической эпистемологии» (Р. Китчин) считается, что «механическая объективность» устраняет субъекта-исследователя и оставляет принятие решений техническим специалистам. В этом смысле этнография и — шире — качественные исследования с акцентом на малых данных и контекстуальности, кажется, вступают в противоречие с методами разработки больших данных.

Несмотря на эпистемологический оптимизм *big data*, в бизнес-этнографии, в первую очередь в маркетинговых исследованиях, заговорили о недостатках больших данных и преимуществе этнографии. Общеизвестно, что корпорации активно нанимают антропологов для проведения качественных маркетинговых исследований. При этом считается, что этнография способна провести более глубокие точечные замеры социальных потребностей, чем *big data*. Биг-аналитики «не имеют доступа к обоснованию через абстракцию и софистику своих статистических методов и уделяют внимание только тому, что может быть измерено» [Fisher, 2015, p. 85].

Так, в маркетинговых исследованиях компании *Adidas* возник вопрос: является ли йога спортом? [Madsbjerg, Rasmussen, 2014]. Вопрос появился в связи с тем, что появлялось много новых потребителей спортивной одежды, не принадлежащих к традиционным видам спорта. Эти потребители не были охвачены маркетинговым вниманием. Йога очень популярна среди приверженцев здорового образа жизни. При этом адепты разнообразных йогических практик ни с кем не состояются, даже с собой, и ничего не выигрывают, также как и большое количество участников неагрессивных видов спорта. Собственно, не ясно, является ли это спортом: «традиционное мышление *Adidas* состояло в том, что покупатели приобретают свою одежду для того, чтобы она помогала им побеждать», но «покупатели хотят продукты, которые помогали бы им вести здоровый образ жизни, а не выигрывать в соревнованиях» [Baer, 2014]. Понять все это и изменить маркетинговую стратегию помогли антропологи с их наблюдениями и глубинными интервью, а не большие данные. Этнография «была способна увидеть то, что продукт *означает* для потребителей тем способом, которым большие данные не владеют» [ibid].

Однако эксплицированное здесь противопоставление оказывается менее эффективным, чем совместная коллаборативная работа над решением проблем, когда «антропологу и кодеру нужно сидеть за одним эпистемологическим столом» [Curran, 2013, p. 71].

Коллаборации и рамки

Кроме противопоставления больших данных и этнографии, можно отметить такую тенденцию, как дополнение или взаимное использование возможностей друг друга в междисциплинарной перспективе. Анализ и результаты этого анализа больших данных зависят от ряда существенных факторов: технологий и платформ, на которых они получены, рамок выборки и рассмотрения, контекста генерации данных (события и обстоятельства), онтологии данных (принципы классификации и калибровки), доступа, связанного с защитой и безопасностью [Kitchin, 2015, p. 465].

Поскольку «сырые данные — это оксюморон» [Gitelman, 2013], а «интерпретация является центром анализа данных» [Brock, 2015, p. 1085], постольку при «приготовлении» больших данных нужна некоторая рамка или перспектива. Этнография может предложить рамки для интерпретации больших данных и результатов машинного анализа. Данные не существуют сами по себе, независимо от исследователя, это всегда результат или процесс сложной работы, в которой задействовано множество сил: «базы больших данных репрезентируют культурные, моральные и социально обусловленные выборы используемой технологии в противоположность технологически определенному инструментальному детерминизму» [ibid].

Кроме того, объемы данных требуют от исследователя определенной фокусировки на конкретных проблемах или сегментах. Большим данным требуются «*маленькие паттерны* (small patterns)», которые говорят, где искать данные, что именно следует исследовать и как это использовать [Floridi, 2012, p. 436]. Вместе с рамками для анализа данных этнографы могли бы составить коллаборации со специалистами по большим данным. Большие данные и этнография могут совместно реализовываться как коллаборативный проект. Этнография — это, по сути, сотрудничество, так как, например, включенное наблюдение всегда предполагает участие других, работа с данными — также недостаток технических компетенций или переговоры по доступу в поле и к данным².

Как отмечает американский антрополог Дж. Маркус, «коллаборация всегда была измерением, присущим этнографии, более или менее осознанным. Но сегодня это также эксплицитная идеология и распространенная форма для бизнеса всех видов — научного, корпоративного, инфраструктурного» [Marcus, 2015, p. 39].

В качественных исследованиях производятся и используются большие массивы разнородных данных: полевые заметки, аудио и видеофайлы, фотографии, транскрипты интервью, данные переписей и опросов. Сюда же следует добавить метаданные и параданные, содержащиеся в корпусах качественных и количественных

² Маргарет Мид в своей автобиографии «Иней на цветах ежевики» (“Blackburry Winter”) хорошо иллюстрирует коллаборативность этнографии через описание трудностей вхождения в поле и ведения переговоров с широким набором акторов — с представителями колониальной администрации и местными жителями [Mid, 1988, с. 13–20].

данных. П. Васамба задает резонный вопрос: «Что происходит после анализа данных и написания отчета?» [Wasamba, 2012, p. 5]. Зачастую у этнографов «большая часть материалов утрачена или находится у отдельных исследователей, которые считают их своей собственностью» [ibid, p. 9]. Действительно, существует проблема уничтожения данных в этнографии и социальных науках вообще [Antonijevic, 2015, p. 63, 66]. Например, транскрипты сохраняются, а из-за недостатка хранилищ исходные аудиофайлы уничтожаются. Архивирование и доступ к архивам делает возможным работу с «сырыми данными», возможность повторного исследования этнографических коллекций данных, создания корпусов данных, извлечения дополнительной информации в виде метаданных и анализ «цифровых следов».

Кроме того, разрыв между архивистами и полевыми связан с тем, что в музеях и архивах хранят, но не используют данные, а в университетах и лабораториях исследуют, но не хранят данные. Решение проблемы использования П. Васамба видит в «коллаборации между производителями информации, исследователями, хранителями, распространителями и конечными пользователями» [Wasamba, 2012, p. 10]. При этом предлагается общее цифровое решение: интернет мог бы быть общим для различных групп интерфейсом для обмена, анализа и хранения данных. «Дигитизация делает возможным преодоление разрыва между коллекциями данных, хранением и простым использованием. Она объединяет работников поля, архивистов и пользователей в комплементарном партнерстве» [ibid, p. 11].

Этнограф Х. Форд рассказала о конкретной работе в рамках коллаборации со специалистами-аналитиками *big data*. Форд профессионально интересовалась Википедией и на конференции *WikiSym* в Маунтин-Дью (Калифорния) в 2011 г. встретила с двумя компьютерными специалистами Д. Музикантом и Ш. Сеносом. В 2013 г. они представили совместный доклад на той же конференции *WikiSym*. Форд обращает внимание на то, что специалисты, работающие с данными, и этнографы имеют много общего, их навыки являются комплементарными друг другу. Совместные открытия более продуктивны, чем изолированная работа. Коллаборация показала, что большие данные предлагают исследователям охват, этнография — необходимую для интерпретации данных глубину. Кроме того, коллаборация предлагает этнографу доступ к большим данным [Ford, 2014]. Сам по себе этот доступ является серьезной исследовательской проблемой и есть только у корпоративных этнографов [Lohmeier, 2014, p. 79]. Совместная работа в рамках коллаборации позволяет ученым расширить свои навыки и учиться друг у друга: «В случае коллаборации, связанной с источниками Википедии, Дейв [Музикант] и Шилад [Сенос] имели необходимые навыки и ресурсы, чтобы выделить 67 миллионов источников, размещенных в 3,5 миллионах статей в Википедии. Основываясь на интервью, которые я делала о способах выбора и записи источников в энциклопедии, я могла предложить идеи для различных способов разделения данных — для получения новых открытий. Дейв и Шилад имели доступ к сложному программному обеспечению и устройствам обработки информации для управления большими объемами данных, а у меня было знание о практиках Википедии» [Ford, 2014, p. 3]. Таким образом, коллаборативная работа специалистов в области больших данных и этнографов способствует совместному решению конкретных задач.

Взаимные уточнения

Коллаборативный потенциал этнографии и больших данных раскрывается во взаимной проблематизации неких узловых моментов, концептов, методов и даже фоновых оснований.

Цифровой аналитик Б. Ф. Уэллс исследовала пожилых женщин, которые проводили достаточно много времени в онлайн играх. Перед ней стояла задача отыскать это сообщество. Для поиска использовалась статистика игры *Second Life*. У этого симулятора примерно 10 млн пользователей, треть из них женщины, один из двадцати игроков — 50 лет и выше, из них только 1% играет 1000 часов. В итоге у Б. Ф. Уэллс сообщество пожилых женщин-игроков составило примерно 1500 из 10 млн пользователей.

Автор делает из результатов этого поиска ряд любопытных выводов. При анализе больших социальных групп существует серьезный риск не заметить такие сообщества, поэтому «обращение к статистическим меньшинствам представляет одну из наиболее волнующих возможностей больших данных» [Welles, 2014, p. 2]. Часто женщины и другие меньшинства вообще исключаются из исследования. Пожилые женщины-игроки представляют собой недостаточно представленную (underrepresented) группу. Такие группы не только не заметны на общем фоне анализа больших данных, зачастую доступ к ним затруднен. И малый фокус или малый паттерн больших данных существенно облегчают доступ к таким меньшинствам.

Принцип *big data* «чем больше, тем правдивее», как оказывается, работает далеко не всегда. Уэллс отмечает: «делая большие данные малыми, мы исправляем исторические пропуски и предвзятости в социальных науках и выстраиваем лучшее, более осознанное, *большее* понимание человеческого поведения» [ibid].

Здесь обнаруживается одна существенная проблема для качественных исследований, а именно проблема локализации сообществ. Если в классический век этнографии они были локализованы пространственно в границах острова, деревни, гетто или перекрестка, то в больших городах, а тем более в виртуальном пространстве, выявить сообщество достаточно сложно. В первом десятилетии XXI в. одной из важных антропологических проблем стала проблема экспликации скрытых и труднодоступных сообществ или популяций, размываемых большими популяциями. Для поиска и локализации таких сообществ используются: а) количественные данные о мексиканских женщинах в «черной» Атлантике; б) глубинное интервью с бродягами и алкоголиками на улицах Анкориджа или в) структурный анализ для наркоманов, обменивающихся шприцами [Angrosio, 2007]. Таким образом, большие данные с малым фокусом — это еще один способ решения важной этнографической проблемы, а именно проблемы выявления границ некоторого непредставленного или плохо представленного сообщества, не имеющего пространственной локализации.

Итак, в первом случае мы видим, как, с одной стороны, этнографические рамки позволяют с помощью методов машинного обучения выявлять и анализировать недостаточно представленные меньшинства. В другом случае, речь идет о том, что технологии больших данных проблематизируют, казалось бы, незыблемое для качественных исследований понятие контекста.

Контекстуальность анализа больших данных — это одно из главных требований со стороны социальных и гуманитарных наук. Необходимо учитывать и делать поправку на контекст «приготовления» и интерпретации массивов. В известной кри-

тической статье о *big data* Бойд и Кроуфорд один из ключевых тезисов звучит так: «Без контекста большие данные теряют свое значение» [Boyd and Crawford, 2012, p. 670]. Антрополог, изучающий алгоритмы и машинное обучение, Н. Сивер как-то вечером слушал музыку, используя контекстное музыкальное приложение в телефоне *Songza*, которое предлагало ему различные темы. Выбирая списки и рубрики, Ник вспомнил критический тезис Бойд и Кроуфорд о контексте. Дело в том, что «среди разработчиков коммерческих систем рекомендованной музыки сегодня популярно предположение, что слушатели действительно хотят зависеть от своего контекста» [Seaver, 2015, p. 1002]. Подобные музыкальные приложения подбирают музыку в соответствии с геотегами, прогнозом погоды, временем суток и другими персональными параметрами телефона. Здесь появляется интересное напряжение: с одной стороны, контекст — это основание этнографии и требование по отношению к данным вообще и большим данным, в частности. С другой стороны, получается, что контексты создаются с помощью данных различных сенсоров, хранящихся на серверах и анализируемых с помощью алгоритмов машинного обучения. Сивер задается вопросом: «Является ли локация вашего смартфона контекстом, или это данные, нуждающиеся в контекстуализации?» [ibid, p. 1005].

Действительно, когда мы говорим: «Это зависит от контекста», имеем ли мы в виду, что контекст есть некий стабильный фон или контейнер, в который мы помещаем конкретные случаи? Большие данные музыкальных приложений показывают, что контексты могут быть созданы с помощью технических средств. Соответственно «контекстуальная революция» в рекомендациях предлагает перспективу того, как по-разному контекст придумывается и управляется в различных группах людей, а «конструкции контекста сами по себе могут быть контекстуально случайными». Н. Сивер полагает, что исследование музыкальных приложений позволяет «понять, как практики больших данных сами производят контекст разнообразными способами. Нам следует помнить, что зачастую наши споры не о преимуществах контекста, а о том, каким его следует сделать» [ibid, p. 1007].

Большие данные показывают, что контекст может быть сконструирован с помощью техники. Таким образом, обращение к работе технических устройств, программному обеспечению и методам анализа данных может прояснить базовые этнографические концепты.

Заключение

Итак, этнография и большие данные, несмотря на различие подходов и методов, несмотря на противопоставление, могут совместно решать исследовательские задачи и быть полезными друг для друга. Есть даже точка зрения, что и большие данные, и этнография решают одну задачу, а именно «как большие данные создают культурные профили того, чем мы являемся». Поскольку «каждый аспект нашей жизни: политическая диспозиция, которую мы занимаем, то, что мы приобретаем как потребители, медиа, к которым подключаемся, что нам нравится и кому мы поклоняемся», плотно связан с большими данными, постольку большие данные — это «культурные данные» [Curran, 2013, p. 64].

Несмотря на шум вокруг больших данных и несмотря на то, что *big data* становится зонтичным понятием, обозначающим все и ничего, появление феномена и методов ставит под вопрос основания социальных наук и проблематизирует концепты. Необходимо очертить контуры перспективных исследований в области взаимодействия этнографии и *big data*.

Важным в этнографии больших данных также является вопрос, что такое данные в качественных исследованиях. Для ответа необходимо распаковать «черный ящик» и развернуть практики «приготовления» данных, а также практики упаковки, когда данные превращаются в нечто естественное, нечто само собой разумеющееся. Необходимо проверить, как в исторической и теоретической перспективе работает напряжение между требованием быть строгой наукой и использовать принципы квантификации и не-репрезентативным потенциалом этнографии, отрицающим классические требования объективности.

Кроме того перспективным направлением является анализ «цифровых следов» в качественных данных. Речь идет о том, чтобы с помощью цифровых методов и инструментов анализировать дополнительную информацию, содержащуюся в качественных данных для получения более обстоятельного знания.

Литература

- Майер-Шенбергер В., Кукье К. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим. М.: Манн, Иванов и Фербер, 2014.
- Малиновский Б. Избранное: Аргонавты западной части Тихого океана. М.: РОССПЭН, 2004.
- Мид М. Культура и мир детства: Избранные произведения. М.: Главная редакция восточной литературы издательства «Наука», 1988.
- Anderson C. The End of Theory: The Data Deluge Makes the Scientific Method Obsolete? // *Wired*. 2008. [Электронный ресурс]. URL: <https://www.wired.com/2008/06/pb-theory/> (дата обращения 16.12.2016).
- Angrosino M. *Doing Cultural Anthropology: Projects for Ethnographic Data Collection*. Second Edition. Long Grove: Waveland Press, 2007.
- Antonijević S. *Amongst Digital Humanists. An Ethnographic Study of Digital Knowledge Production*. New York: Palgrave Macmillan, 2015.
- Baer D. Here's Why Companies are Desperate to Hire Anthropologists // *Business Insider*. 2014. [Электронный ресурс]. URL: <http://www.businessinsider.com/heres-why-companies-are-desperate-to-hire-anthropologists-2014-3#ixzz3VER2SwFi> (дата обращения 16.12.2016).
- Boyd D., Crawford K. Critical Questions for Big Data. Provocations for the Cultural, Technological and Scholarly Phenomenon // *Information, Communication & Society*. № 5 (15). 2012. P. 662–679.
- Burrows R., Savage M. After the Crisis? Big Data and the Methodological Challenges of Empirical Sociology // *Big Data & Society*. April–June. 2014. P. 1–6.
- Brock A. Deeper Data: a Response to Boyd and Crawford // *Media, Culture & Society*. 2015. Vol. 37(7). P. 1084–1088.
- Curran J. Big Data or “Big Ethnographic Data”? Positioning Big Data within the ethnographic space // EPIC. Ethnographic Praxis Industry Conference. 2013. P. 62–73.
- Fisher M. Time, Camera and the (Digital) Pen. *Writing Culture Operating Systems 1.0–3.0* // *Starn O. Writing Culture and the Life of Anthropology*. Durham; London: Duke University Press, 2015. P. 72–104.
- Floridi L. Big Data and their Epistemological Challenge // *Philosophy & Technology*. 2012. № 25 (4). P. 435–437.

- Ford H. Big Data and Small: Collaborations between Ethnographers and Data Scientists // *Big Data & Society*. July–December. 2014. P. 1–3.
- Gitelman L. «Raw Data» is an Oxymoron. Cambridge (Mass.); London: The MIT Press, 2013.
- Kitchin R. Big Data. New Epistemologies and Paradigm Shift // *Big Data & Society*. April–June. 2014. P. 1–12.
- Kitchin R., Lauriault T. P. Small Data in the Era of Big data // *GeoJournal*. 2015. 80(4). P. 63–475.
- Lohmeier Ch. The Researcher and the Never-Ending field: Reconsidering Big Data and Digital Ethnography // Hand M., Hillyard S. (eds.) *Big data? Qualitative Approaches to Digital Research*. Bingley: Emerald, 2014. P. 77–90.
- Madsbjerg C., Rasmussen M. B. *The Moment of Clarity: Using the Human Sciences to Solve Your Hardest Business Problems*. Boston: Harvard Business Review Press, 2014.
- Marcus G. The Legacies of *Writing Culture* and the Near Future of the Ethnographic Form: A Sketch // *Starn O. Writing Culture and the Life of Anthropology*. Durham; London: Duke University Press, 2015. P. 35–51.
- Merry S. E. *The Seductions of Quantification: Measuring Human Rights, Gender Violence and Sex Trafficking*. Chicago; London: Chicago University Press, 2009.
- Porter Th. M. *Trust in Number. The Pursuit of Objectivity in Science and Public Life*. Princeton: Princeton University Press, 1995.
- Ruppert E. Rethinking Social Sciences // *Dialogues in Human Geography*. 2013. № 3. P. 268–273.
- Seaver N. The Nice Thing about Context is that Everyone has it // *Media, Culture and Society*. 2015. № 37. P. 1101–1109.
- Vannini P. Non-representational Ethnography: New Ways of Animating Lifeworlds // *Cultural Geographies*. 2015. № 22 (2). P. 317–327.
- Vidich A., Lyman S. Qualitative Methods: Their History in Sociology and Anthropology. Denzin N. K., Lincoln Y. S. (eds.) *Handbook of Qualitative Research*. Thousand Oaks: SAGE, 1994. P. 23–44.
- Wang T. Big Data Needs Thick Data // *Ethnography Matters*. 2014. [Электронный ресурс]. URL: <http://ethnographymatters.net/blog/2013/05/13/big-data-needs-thick-data/> (дата обращения 16.12.2016).
- Wasamba P. Going beyond Data Collection in Ethnography: Options for Bridging the Gap between Researchers and Archivists // *International Journal of African Renaissance Studies – Multi-, Inter- and Transdisciplinarity*. 2012. № 7 (2). P. 4–17.
- Welles B. F. The Minorities and Outliers: The Case for Making Big Data Small // *Big Data & Society*. April–July. 2014. P. 1–2.

Big Data and Ethnography: Challenges and Opportunities

DENIS YU. SIVKOV

professor, department of philosophy and sociology,
Russian Presidential Academy of Economics and Public Administration, Volgograd Branch,
Volgograd, Russia;
e-mail: d.y.sivkov@gmail.com

The emergence of the phenomenon of big data in marketing and management, the ability to explain and predict the behavior of large groups have become a serious challenge and forced the social sciences to overestimate their capabilities. Firstly, digitized data processing and machine conceive the necessity of the social sciences as the matter of concern. “Empirical epistemology” (R. Kitchin) considers “mechanical objectivity” addressing the subject-researcher and leaves decision-making to technical experts. In this sense, ethnography and wider qualitative research with a focus on small data, full-bodied with rich description and contextuality seem to be in contradiction with the methods of big data

development. In corporate ethnographic research, marketing has shown the limitations of approaches based on machine learning. Besides of the opposition of ethnography and machine learning methods, the collaborative work seems more perspective, it has ability to relational integration and correlation of big and small data, as well as search for theoretical frameworks (H. Ford). Big data needs "small patterns" that define place, object and method of data retrieval. Small ethnographical data also clarify definition and application of big data. On the one hand, ethnography allows us to be more attentive to the small and excluded groups of users, which leave specific digital trails, but generally go unnoticed (B. Welles). On the other hand, study of big data helps us to clarify and problematize traditional quality concepts such as the concept of "context" (N. Seaver).

Key words: data, big data, ethnography, anthropology, qualitative studies, quantitative studies, collaboration, underrepresented community, context, underrepresented community.

References

- Anderson, C. (2008) The End of Theory: The Data Deluge Makes the Scientific Method Obsolete? // *Wired*. 2008. Available at: <https://www.wired.com/2008/06/pb-theory/> (date accessed 16.12.2016).
- Angrosino M. (2007) *Doing Cultural Anthropology: Projects for Ethnographic Data Collection*. Second Edition. Long Grove: Waveland Press.
- Antoničević S. (2015) *Amongst Digital Humanists. An Ethnographic Study of Digital Knowledge Production*. New York: Palgrave Macmillan.
- Baer D. Here's Why Companies are Desperate to Hire Anthropologists // *Business Insider*. 2014. Available at: <http://www.businessinsider.com/heres-why-companies-aredesperateto-hireanthropologists-2014-3#ixzz3VER2SwFi> (date accessed 16.12.2016).
- Boyd D., Crawford K. (2012) Critical Questions for Big Data. Provocations for the Cultural, Technological and Scholarly Phenomenon // *Information, Communication & Society*. № 5 (15). P. 662–679.
- Burrows R., Savage M. (2014) After the Crisis? Big Data and the Methodological Challenges of Empirical Sociology // *Big Data & Society*. April–June. P. 1–6.
- Brock A. (2015) Deeper Data: a Response to boyd and Crawford // *Media, Culture & Society*. Vol. 37(7). P. 1084–1088.
- Curran J. (2013) Big Data or 'Big Ethnographic Data'? Positioning Big Data within the ethnographic space // EPIC. Ethnographic Praxis Industry Conference. P. 62–73.
- Fisher M. (2015) Time, Camera and the (Digital) Pen. *Writing Culture Operating Systems 1.0–3.0* // Starn O. *Writing Culture and the Life of Anthropology*. Durham; London: Duke University Press. P. 72–104.
- Floridi L. (2012) Big Data and their Epistemological Challenge // *Philosophy & Technology*. № 25 (4). P. 435–437.
- Ford H. (2014) Big Data and Small: Collaborations between Ethnographers and Data Scientists // *Big Data & Society*. July–December. P. 1–3.
- Gitelman L. (2013) «Raw Data» is an Oxymoron. Cambridge (Mss.); London: The MIT Press.
- Kitchin R. (2014) Big Data. New Epistemologies and Paradigm Shift // *Big Data & Society*. April–June. P. 1–12.
- Kitchin R., Lauriault T. P. (2015) Small Data in the Era of Big Data // *GeoJournal* (2015) 80. P. 463–475.
- Lohmeier Ch. (2014) The Researcher and the Never-Ending field: Reconsidering Big Data and Digital Ethnography // Hand M., Hillyard S. (eds.) *Big data? Qualitative Approaches to Digital Research*. Bingley: Emerald. P. 77–90.
- Madsbjerg C., Rasmussen M. B. (2014) *The Moment of Clarity: Using the Human Sciences to Solve Your Hardest Business Problems*. Boston: Harvard Business Review Press.

- Majer-Shenberger V., Kuke K. (2014) Bolshie dannye. Revolyutsiya, kotoraya izmenit to, kak my zhivem, rabotaem i myslim [Big Data: A Revolution That Will Transform How We Live, Work, and Think]. M.: Mann, Ivanov i Ferber (in Russian).
- Malinovskij B. (2004) *Izbrannoe: Argonavty zapadnoy chasti Tikhogo okeana* [Argonauts of the Western Pacific]. M.: ROSSPEN (in Russian).
- Marcus G. *The Legacies of Writing Culture and the Near Future of the Ethnographic Form: A Sketch* // Starn O. *Writing Culture and the Life of Anthropology*. Durham; London: Duke University Press. P. 35–51.
- Merry S. E. (2016) *The Seductions of Quantification: Measuring Human Rights, Gender Violence and Sex Trafficking*. Chicago; London: Chicago University Press.
- Mid M. *Kultura i mir detstva: Izbrannye proizvedeniya* [Culture and World of Childhood]. M.: Glavnaya redakciya vostochnoy literatury izdatelstva «Nauka», 1988 (in Russian).
- Porter Th. M. (1995) *Trust in Number. The Pursuit of Objectivity in Science and Public Life*. Princeton: Princeton University Press.
- Ruppert E. (2013) Rethinking Social Sciences // *Dialogues in Human Geography*. № 3. P. 268–273.
- Seaver N. (2015) The Nice Thing about Context is that Everyone has it // *Media, Culture and Society*. 37. 1101–1109.
- Vannini P. (2015) Non-representational Ethnography: New Ways of Animating Lifeworlds // *Cultural Geographies*. № 22 (2). P. 317–327.
- Vidich A., Lyman S. (1994) *Qualitative Methods: Their History in Sociology and Anthropology*. Denzin N. K., Lincoln Y. S. (eds.) *Handbook of Qualitative Research*. Thousand Oaks: SAGE. P. 23–44.
- Wang T. Big Data Needs Thick Data // *Ethnography Matters*. 2014. Available at: <http://ethnographymatters.net/blog/2013/05/13/big-data-needs-thick-data/> (date accessed 16.12.2016).
- Wasamba P. (2012) Going beyond Data Collection in Ethnography: Options for Bridging the Gap between Researchers and Archivists // *International Journal of African Renaissance Studies – Multi-, Inter- and Transdisciplinarity*. № 7 (2). P. 4–17
- Welles B. F. (2014) The Minorities and Outliers: The Case for Making Big Data Small // *Big Data & Society*. April–July. P. 1–2.