

УДК 316

А.А. Гусева**A.A. Guseva**Пермский национальный исследовательский
политехнический университет

Perm National Research Polytechnic University

**«БОЛЬШИЕ ДАННЫЕ»: ПОНЯТИЕ,
ИСТОЧНИКИ, ВОЗМОЖНОСТИ****BIG DATA: CONCEPTS, SOURCES,
OPPORTUNITIES**

Дано определение понятию «Большие данные». Приведены примеры использования инновационного подхода Big Data в социальных и экономических исследованиях. Показана эмпирическая доступность и значимость данной технологии для социологов.

Ключевые слова: «Большие данные», технологии, источники данных, информация, система, парадигмальный сдвиг.

The concept of Big Data is revealed. Provides an overview of examples of the use of the innovative approach in social and economic research. Shown empirical accessibility and relevance of Big Data technologies for social scientists.

Keywords: Big Data, technology, Data sources, information, system, paradigm shift.

Как часто нам приходится в повседневной жизни сталкиваться с различными информационными технологиями? Многочисленные личные технические устройства, персональный компьютер, Интернет, электронные карты, личные счета, системы видеорегистрации, предприятия и организации, окутанные локальными сетями, кассовые аппараты, CRM-системы, различные системы учета – в городской среде технологии встречаются на каждом шагу.

Все эти системы и устройства ежедневно фиксируют нашу жизнь, причем фиксация может быть не явной, а латентной функцией информационных систем. Становится очевидным тот факт, что огромное количество информации, ежедневно воспроизводимое обществом в процессе жизнедеятельности, является кладезем реального, потенциально полезного, поддающегося различным видам анализа эмпирического материала. Информационные технологии, развиваясь и расширяясь вглубь ежедневных практик современного общества, предлагают нам качественно пересмотреть наше представление об эмпирических данных. Так появилась технология Big Data.

Термин Big Data (BD) имеет множество различных определений и остается весьма расплывчатым понятием. Возможно, ввиду отсутствия четкого представления о сущности BD, все возможности работы в данной сфере не оцениваются до конца должным образом.

По мнению научного редактора журнала «Открытые системы. СУБД» Л. Черняка, достоверной датой рождения Big Data можно считать 3 сентября 2008 года. Это день, когда «вышел специальный номер старейшего британского научного журнала Nature, посвященный поиску ответа на вопрос: как могут повлиять на будущее науки технологии, открывающие возможности работы с большими объемами данных? Специальный номер подытоживает предшествующие дискуссии о роли данных в науке вообще и в электронной науке (e-science) в частности» [1]. Подчеркнем для формирования собственного представления о «Больших данных» их главные свойства: e-science, большие объемы данных, технологии.

BD не просто технология, это парадигмальный сдвиг, новая эмпирическая культура, которая практически формируется в реальном времени, на наших глазах. К. Линч (редактор журнала Nature) «предложил для новой парадигмы специальное название “Большие данные”, выбранное им по аналогии с такими метафорами, как большая нефть, большая руда и т.п., отражающими не столько количество чего-то, сколько переход количества в качество» [1].

Американская ассоциация исследователей общественного мнения (AAPOR) в своем отчете о «Больших данных» [2] также рассматривает BD как парадигмальный сдвиг в обзорном/социальном исследовании: «За последние годы мы наблюдаем увеличение объема статистической информации, описывающей различные социальные феномены, основанные на так называемых “Больших данных”... Изменение в становлении новых типов данных, их доступность, способы сбора и распространения фундаментальны. Это изменение конституирует парадигмальный сдвиг в массовых опросах (survey researches)» [2, с. 9].

В этом же отчете приводится довольно общее определение BD: «Термин “Большие данные” – это своего рода описание больших по объему и разнообразных по составу характеристик, практик, технических приемов, этических проблем и последствий, которые связаны с данными» [3, с. 14]. Со ссылкой на отчет компании «Гартнер»¹ авторы рассматривают основные характеристики «Больших данных» [2, с. 18]: для описания основных характеристик BD в английском языке используют три буквы V: Volume (объем), Velocity (скорость) и Variety (многообразие). Иногда такое определение дополняют характеристиками: изменяемость, достоверность, комплексность (возможность состыковки источников данных). Таким образом, перед нами раскрывается понятие BD как данных, обладающих набором определенных свойств.

Отечественный социолог Ю.Н. Толстова исследует возможности и проблемы «Больших данных». Ей близко представление о BD как о технологии: «ядром этой технологии является набор достижений компьютерной науки,

¹ «Гартнер» (Gartner) – одна из самых влиятельных компаний в исследовательской и консалтинговой сферах с более чем миллиардным оборотом, в основном специализируется на исследовании рынка информационных технологий.

позволяющих искать и формировать массивы больших, разбросанных, изменяющихся, структурированных и неструктурированных данных» [4, с. 6]. BD представляет собой совокупность информационных технологий – способов и средств сбора и анализа информационных потоков.

Итак, становится очевидно, что Big Data – это комплексное понятие, сочетающее в себе:

- 1) непосредственно данные (множество закодированной информации);
- 2) совокупность технологий работы с этими данными;
- 3) новый взгляд, новую парадигму в науке о данных (data science).

Нужно заметить, что новая парадигма возникает не только в рамках data science, но и проявляется в каждой науке, имеющей дело со статистическим анализом, в том числе и в социологии. Социолог РАНХиГС Д. Рогозин связывает обращение к BD с «формированием новой культуры анализа социальных данных» [2, с. 6].

BD – многоуровневое понятие, каждый из его аспектов имеет свое значение и свой интерес для дальнейшего изучения. Далее сконцентрируемся на практической составляющей – источниках BD и примерах использования данной технологии в социальных, экономических и других исследованиях. Источники Big Data «стали обрабатываться с целью нахождения взаимосвязей в экономических и социальных системах, где ранее при помощи опросов, экспериментов и этнографических наблюдений выводились заключения и строились прогнозы» [2, с. 14].

В отечественной научной литературе часто Big Data упоминаются в контексте противопоставления данным, полученным традиционными методами. Так или иначе, но «Большие данные» имеют ключевое отличие: «чаще всего они являются вторичными данными, первоначально предназначенными для других целей. Это означает, что “Большие данные”, как правило, изначально не имеют прямого отношения к исследовательским задачам и лишь повторно используются исследователями для социологических наблюдений» [2, с. 19]. Для лучшего понимания разницы между «Большими» и обычными данными можно обратиться к сравнению органических и сконструированных данных Роберта Гроувза или полученных и созданных данных Ш. Тейлора.

Говоря о доступности такого вида эмпирического материала, можно заметить, что BD – это клад, до ценности которого еще нужно докопаться. Поскольку «Большие данные» коренятся в среде информационных технологий (ИТ), то вместе с развитием и распространением ИТ-источники «Больших данных» также становятся доступнее. Данные социальных сетей и других интернет-сервисов, информация с регистрирующих устройств, данные транзакций, административные сведения – это наиболее распространенные источники BD.

Технические средства и программные продукты для работы с «Большими» данными возникают и совершенствуются в реальном времени. Подроб-

нее о некоторых доступных сервисах пишет Н.В. Корытникова [3, с. 20–21]. Существует также большое количество сервисов, практически предоставляющих доступ к «Большим данным» или работающих с ними, ориентированных на удовлетворение коммерческих, а не исследовательских запросов. Речь идет, например, о компаниях, занимающихся таргетингом – подбором целевой аудитории для рекламных кампаний по заданным характеристикам профиля какой-либо социальной сети. Теоретически это является примером технологии работы с «Большими данными» социальных сетей, примером их вторичного использования. Таким образом, мы понимаем, что возможности для работы с BD есть практически у каждого. Западные социологи уже оценили открывающиеся возможности и претворили в жизнь целый ряд проектов, базирующихся на технологиях Big Data.

Технологии BD оказываются доступнее, чем кажется. Заинтересованные игроки уже имеют свою выгоду от вторичного использования накопленной информации. В коммерческой среде технологии Big Data используются, развиваются, хоть и не всегда позиционируются именно как «Большие данные». Исследовательская среда, особенно отечественная, не так активно подхватывает современные ноу-хау, и найти примеры использования BD возможно только в зарубежной литературе или переводах.

Среди наиболее известных примеров – разработка в 2009 году компанией Google системы, анализирующей популярность конкретных поисковых запросов с целью выявления новых очагов гриппа H1N1. Система была основана на данных эпидемиологов за 2007–2008 годы и собственных данных Google. «В итоге были выявлены 45 поисковых запросов, которые на 97 % (!) коррелировали с фактическими данными. Медики получили надежный, работающий в реальном времени инструмент выявления новых очагов распространения гриппа» [4, с. 7]. С его помощью была решена проблема выявления новых очагов и принятия оперативных мер органами здравоохранения.

Другой пример академической инициативы – проекты Массачусетского технологического института Billion Prices Project и PriceStats. В рамках этих проектов для проведения экономических исследований использовались цены, которые ежедневно собирались с сайтов интернет-магазинов по всему миру. В рамках проекта PriceStats велась оценка инфляции в США на основе все тех же интернет-цен. Изменение инфляционных тенденций оперативнее отражалось в PriceStats по сравнению с главным индикатором инфляции в стране – индексом потребительских цен [2, с. 17].

Богатый источник полезных данных – социальные сети. Так, Мичиганский университет создал индекс потери работы (University of Michigan Social Media Job Loss Index) на основе сообщений из «Твиттера». С помощью анализа сообщений о потере работы строятся прогнозы об исках о получении страховки по безработице. В Нидерландах обнаружили корреляцию между

мнениями, зафиксированными в социальной сети Facebook, и индексом потребительского доверия, построенным по данным опросов.

Несколько примеров применения технологий Big Data показывают нам, что потенциальные возможности «Больших данных» колоссальны, ведь сейчас специалисты всего мира только начинают работать в этом направлении.

К сожалению, отечественные примеры участия в подобных проектах или собственные разработки, научные диссертации по методологии и проблемам «Больших данных» полностью отсутствуют. Это свидетельствует о неоправданно низкой заинтересованности отечественной социологии в описанной нами технологии. В отдельно встречающихся публикациях на тему ВД встречаются призывы к введению отдельной дисциплины или курса по «Большим данным» для студентов-социологов. На наш взгляд, необходимо систематизировать и популяризовать имеющиеся знания и опыт по использованию технологий Big Data. Это может послужить мощным толчком к развитию методов ВД в прикладной социологии в целом.

Список литературы

1. Черняк Л. «Большие данные» – новая теория и практика // Открытые системы. – 2011. – № 10.
2. Отчет AAPOR о «Больших данных» / Л. Джапек, Ф. Крейгер, М. Берг [и др.] / Американская ассоциация исследователей общественного мнения. – М., 2015.
3. Корытникова Н.В. Online Big Data как источник аналитической информации в online-исследованиях // Социс. – 2015. – № 8. – С. 14–24.
4. Толстова Ю.Н. Социология и компьютерные технологии // Социс. – 2015. – № 8. – С. 3–13.

Получено 15.03.2016

Гусева Александра Александровна – магистрант кафедры «Социология и политология», гуманитарный факультет, Пермский национальный исследовательский политехнический университет.

Научный руководитель – **Колесниченко Милана Борисовна**, кандидат социологических наук, доцент кафедры «Социология и политология», гуманитарный факультет, Пермский национальный исследовательский политехнический университет.