

**РОССИЙСКАЯ АКАДЕМИЯ НАУК  
ИНСТИТУТ НАУЧНОЙ ИНФОРМАЦИИ  
ПО ОБЩЕСТВЕННЫМ НАУКАМ**

**БОЛЬШИЕ ДАННЫЕ  
В СОЦИАЛЬНЫХ  
И ГУМАНИТАРНЫХ НАУКАХ**

**СБОРНИК ОБЗОРОВ И РЕФЕРАТОВ**

**МОСКВА  
2019**

ББК 32.97+60+72

Б 79

Серия

*«Наука, образование и технологии»*

*Центр научно-информационных исследований  
по науке, образованию и технологиям*

Редакционная коллегия серии:

*Е.Г. Гребенщикова* – д-р филос. наук, главный редактор,  
*М.П. Булавинова* – ответственный секретарь,  
*Е.В. Алферова* – канд. юр. наук, *В.И. Аршинов* –  
д-р филос. наук, *Т.В. Виноградова* – канд. психол. наук,  
*И.Т. Касавин* – член-корреспондент РАН,  
*В.А. Лекторский* – академик РАН, *С.М. Пястолов* –  
д-р эконом. наук, *А.И. Ракитов* – д-р филос. наук,  
*А.И. Селиванов* – д-р филос. наук.

**Б 79 Большие данные в социальных и гуманитарных науках:** Сб. обзоров и рефератов / РАН. ИНИОН. Центр науч.-информ. исслед. по науке, образованию и технологиям; отв. ред. – Гребенщикова Е.Г. – М., 2019. – 193 с. – (Сер.: Наука, образование и технологии).

**ISBN 978-5-248-00912-1**

Сборник включает обзоры и рефераты работ, посвященных актуальным проблемам больших данных в социальном и гуманитарном знании.

Для научных работников, преподавателей вузов, аспирантов и студентов.

ББК 32.97+60+72

ISBN 978-5-248-00912-1

© ИНИОН РАН, 2019

## СОДЕРЖАНИЕ

Предисловие .....	6
-------------------	---

### ФИЛОСОФСКИЕ И МЕТОДОЛОГИЧЕСКИЕ ПРОБЛЕМЫ БОЛЬШИХ ДАННЫХ

<i>Тёрнберг П., Тёрнберг А.</i> Пределы вычисления: Философская критика современных исследований в области больших данных. (Реферат) .....	10
Общественные науки и веб-ресурсы: От «скрытного» к междисциплинарному исследованию больших данных / Боун Дж., Эмиль Ч., Абдул А., Когхилл Дж., Пан Уэй. (Реферат) .....	19
<i>Лоури А.</i> Алгоритмическая рациональность: Эпистемология и эффективность в информационных науках. (Реферат) .....	25
Методы качественного анализа и данные в условиях цифрового общества / Хаусли У., Дикс Б., Хенвуд К., Смит Р. (Реферат) .....	30
<i>Джуитт К., Прайс С., Седо А.</i> Исследуя человека в цифровых контекстах: К новым методологическим дискуссиям – от методов в практических областях к методологии общественных наук. (Реферат) .....	32
<i>Чёрчленд П.С., Сейновский Т.Дж.</i> Соединяя вычислительную и экспериментальную нейронауку. (Реферат) .....	38
<i>Милковский М.</i> Механистическая концепция вычислительных объяснений в когнитивной науке и вычислительной нейронауке. (Реферат) .....	41
Большие данные в орнитологии: Перспективы и вызовы / Ла Сорте Ф.А., Лепчик К.А., Барнетт Дж.Л., Хурлберт А.Х., Тингли М.В., Цукерберг Б. (Реферат) .....	46

## БОЛЬШИЕ ДАННЫЕ В СОЦИАЛЬНЫХ НАУКАХ

<i>Бласкес Д., Доменеч Х.</i> Источники и методы изучения больших данных для социального и экономического анализа. (Реферат) .....	53
<i>Райбс Д.</i> STS встречается с наукой о данных. (Реферат) .....	59
Культуры данных, содержащиеся в мобильных приложениях для знакомств: Вопросы для критического социологического исследования / Албари К., Бургесс Дж., Лайт Б., Рэйс К., Уилкен Р. (Реферат) .....	65
Полная картина вечеринки: Научный эксперимент по взаимодополнению социальных данных / Блок А., Карлсен Х.Б., Йёргенсен Т.Б., Мадсен М.М., Ралунд С., Педерсен М.А. (Реферат) .....	69

## БОЛЬШИЕ ДАННЫЕ В ЭКОНОМИКЕ И ОРГАНИЗАЦИИ НАУКИ

<i>Садовский Ж.</i> Данные как капитал: Датафикация, аккумуляирование и извлечение. (Реферат) .....	71
<i>Кёниг П.Д.</i> Место обусловленности и индивидуальной ответственности в «экономике больших данных». (Реферат) ....	79
<i>С.М. Пястолов.</i> Экономические и социальные ценности, риски и организационные парадоксы больших данных. (Обзор) .....	85
<i>Ки К.Ф., Шрок А.Р.</i> Лучшие социальные и организационные практики успешных научных интернет-порталов и киберинфраструктурных проектов. (Реферат) .....	95
<i>Масланов Е.</i> Краудсорсинг в науке: Новый элемент научной инфраструктуры. (Реферат) .....	98
<i>Тхашил Н.</i> Построение глобальных данных: Автоматизированные методы экологического мониторинга, осторожного отношения к рискам и их наглядного представления. (Реферат) .....	103
<i>Бёрнер К.</i> Научная политика, основанная на управлении данными. (Реферат) .....	112

## ПРАВОВОЕ РЕГУЛИРОВАНИЕ БОЛЬШИХ ДАННЫХ

<i>Т.В. Захаров.</i> Правовые проблемы оборота больших данных в условиях цифровой экономики. (Статья) .....	116
<i>Д.В. Красиков.</i> Сбор и использование больших данных в правозащитной деятельности: Потенциал и риски с позиции международного права прав человека. (Обзор) .....	128
<i>Галиндо Ф., Марко Х.Г.</i> Свобода и Интернет: Наделение граждан полномочиями и решение проблемы дефицита прозрачности поисковых систем. (Реферат) .....	135
<i>Кокколи Я.</i> Вызовы новых технологий в области реализации прав человека: Анализ некоторых критических вопросов в цифровую эпоху. (Реферат) .....	144
<i>Хинг А.</i> Некоторые размышления о достоинстве как альтернативной правовой концепции в регулировании защиты данных. (Реферат) .....	151
<i>Роен М., Фэн Ц.И.</i> Почему «компьютер говорит нет»: Иллюстрация риска дискриминации при использовании больших данных с помощью науки о сложных системах. (Реферат) .....	158

## ВЗАИМООТНОШЕНИЯ НАУКИ И ОБЩЕСТВА В ЭПОХУ БОЛЬШИХ ДАННЫХ: ЭТИЧЕСКИЕ, ЭПИСТЕМОЛОГИЧЕСКИЕ И ОРГАНИЗАЦИОННЫЕ ИЗМЕРЕНИЯ

<i>Михаэль М., Лаптон Д.</i> К созданию манифеста для изучения проблемы «понимания публикой больших данных». (Реферат) .....	164
<i>Сивков Д.Ю.</i> Большие данные в этнографии: Вызовы и возможности. (Реферат) .....	172
<i>Флориди Л.</i> Большие данные и их эпистемологический вызов. (Реферат) .....	176
<i>Прайс У.Н., Козн И.Г.</i> Частная жизнь в эпоху больших медицинских данных. (Реферат) .....	181
<i>Эдвардс С.</i> Постоянное согласие или осторожный контроль за использованием цифровых данных? (Реферат) .....	189

## ПРЕДИСЛОВИЕ

В сборнике рассматривается социогуманитарная проблематика больших данных, ее этико-философские, эпистемологические, экономические, правовые и методологические измерения, влияние на организационные структуры науки и специфику взаимоотношений между наукой и обществом. Особенность проблематики заключается в том, что становление больших данных происходит параллельно с их теоретическим осмыслением в общественных науках, поиском этико-правовых регулятивов и формированием новых способов производства знания.

В первом разделе раскрываются философские и методологические проблемы больших данных, становление новых подходов к пониманию взаимосвязи общества, природы и культуры, а также гуманитарных аспектов цифровых технологий, переопределяющих современный социальный опыт и способы социокультурной идентификации человека.

Второй раздел посвящен правовым аспектам развития новых технологий в эпоху больших данных, в частности их потенциалу в предотвращении нарушений прав человека благодаря открывшимся возможностям прогнозирования в этой сфере. Однако развитие подхода к правам человека, ориентированного на использование больших данных, сопряжено с определенными рисками. Во-первых, такой подход пригоден лишь для предотвращения отдельных нарушений прав человека; во-вторых, он сам может породить дополнительные нарушения прав человека; в-третьих, может возникнуть режим «приватизации» прав человека на международной арене за счет активного привлечения в дело защиты прав человека корпораций, а не государственных структур. Для снижения этих рисков, как показано в одном из материалов этого раздела, необходимо налаживать диалог о более

ответственном, этичном и легальном подходе к использованию больших данных в сфере международного обеспечения прав человека.

Еще один аспект рассматриваемой проблематики связан с исследованием потенциала и рисков использования больших данных в контексте продолжающегося глобального миграционного кризиса. Вопреки общепринятому восприятию больших данных исключительно в качестве инструмента слежения и контроля, эти данные могут и должны использоваться как средство защиты прав, в том числе мигрантов: обращение к ним может усовершенствовать процесс принятия решений и принятия мер для предотвращения смертей на море, жестокого обращения с мигрантами и торговли ими.

Возможности больших данных в социальных исследованиях раскрываются в третьем разделе. В частности, фиксируется переход от 3V-модели – объем (volume), скорость (velocity) и вариативность (variety) – к модели 4V, к которой добавлена ценность (value), а в последнее время – модели 5V, в которой значимое место занимает достоверность (veracity) – параметр, означающий адекватное управление данными и соблюдение права на частную жизнь. Последняя модель представляет интерес с учетом особого внимания специалистов разного профиля к вопросам права на частную жизнь и связанную с ней «культуру данных». «Культура данных» может пониматься в четырех разных смыслах: институционализированные правила, привычки и практики; различные способы, с помощью которых данные обрабатываются; процессы датафикации культуры посредством алгоритмической логики цифровых медиа; культура использования данных, представленных в мобильных приложениях смартфонов, например в мобильных приложениях для знакомств. Для пользователей всех мобильных приложений сегодня проблема безопасности весьма актуальна, а в случае платформ знакомств и подобных им программных решений она приобретает особую значимость.

Экономическим аспектам проблематики больших данных, как и вопросам ее влияния на организационные структуры науки, посвящен четвертый раздел сборника. Так, в одном из рефератов раскрывается проблема сбора и обмена базами данных как новый центральный компонент в системе функционирования современного капитализма. Хотя совсем недавно базы данных воспринимались исключительно как объект, интересный для научных исследе-

дований, и как побочный продукт различных производственных процессов. До самого последнего времени компании просто с легкостью удаляли накопившиеся данные, теперь они тщательно их собирают и накапливают по мере возможности. Более того, формирование и использование огромных баз данных все больше становятся объектом целенаправленного инвестирования средств и, следовательно, подчиняются общим принципам ориентации на получение прибыли. Большие данные теперь – необходимый компонент при принятии важнейших решений как на уровне крупных компаний (например, приобретение компании WholeFoods компанией Amazon), так и на уровне государственного управления, например решений городских властей об инвестировании в создание беспроводных сетей. Анализ больших данных не как одного из видов новых товаров, а как формы капитала, безусловно, способствует углублению анализа современной системы «цифрового капитализма». Его суть, как правило, описывается всеми или одним из тезисов: 1) большие данные обладают стоимостью и способны создавать новую стоимость; 2) большие данные оказывают мощное и всеобъемлющее влияние на процессы развития современного бизнеса и государственного управления; 3) распространение больших данных часто связано с такими явлениями, как неравенство, несправедливость и эксплуатация. В то же время объем и стоимость баз данных непрерывно возрастают, что хорошо видно на примере таких богатейших компаний, как Facebook и Google.

Пятый раздел охватывает разнообразные направления исследований, связанные с влиянием больших данных на взаимоотношения науки и общества. При этом, как отмечено в одном из реферируемых материалов, возникают разные «эпистемные лица»: «непрофессиональные пользователи» (пользователи, или потребители); «непрофессиональные вкладчики» (те, кто генерирует и поставляет большие данные); «непрофессиональные эксперты» (те, кто владеет определенными компьютерными знаниями и занимается процессами, качеством и политикой в этой области); «эксперты-непрофессионалы» (те, кто пишет программы и изобретает технологии). Эти категории не являются стабильными или взаимоисключающими, но они дают представление о том, насколько сложны проблемы вовлеченности публики в динамику больших данных. По сути, речь идет о новом подходе, предполагающем изменение привычного стиля мышления, что уже нашло отражение в таких областях, как цифровая социология, цифровые культуры, исследования программного обеспечения и исследования Интернета.

Изучение проблем «понимания науки обществом» и «вовлечения общественности в науку» неизбежно ведет к формированию новых теоретических моделей производства знания и его распространения, к дальнейшему переосмыслению концепта «большие данные» в исследованиях науки и технологий и в социально-гуманитарном знании в целом.

*Е.Г. Гребенщикова*

# ФИЛОСОФСКИЕ И МЕТОДОЛОГИЧЕСКИЕ ПРОБЛЕМЫ БОЛЬШИХ ДАННЫХ

Тёрнберг П., Тёрнберг А.

## ПРЕДЕЛЫ ВЫЧИСЛЕНИЯ: ФИЛОСОФСКАЯ КРИТИКА СОВРЕМЕННЫХ ИССЛЕДОВАНИЙ В ОБЛАСТИ БОЛЬШИХ ДАННЫХ (Реферат)

Ref. ad op.: Törnberg P., Törnberg A. The limits of computation: A philosophical critique of contemporary Big Data research // Big data & society. – 2018. – Vol. 5, N 2. – P. 1–12. – Mode of access: <https://doi.org/10.1177/2053951718811843>

Петер Тёрнберг – сотрудник Университета Амстердама (Нидерланды), Антон Тёрнберг работает в Университете Гётеборга (Швеция). Статья посвящена современным дискуссиям, связанным с возникновением новой вычислительной парадигмы в социальных науках, в частности эпистемологическим и онтологическим эффектам больших данных.

Термин «большие данные» (БД) используется для описания объема информации, полученной с помощью новых коммуникационных и компьютерных технологий – мобильных устройств, систем позиционирования, онлайн-сервисов и т.д. Сегодня человек живет в оцифрованном мире. Каждый день люди оставляют огромное количество следов – данных, – занимаясь покупками, обменом, просмотром, поиском, общаясь с друзьями, родственниками, коллегами и т.д., – которые дают беспрецедентные возможности социологам исследовать не только общество в целом, но и жизнь отдельных людей, их мысли, желания и чувства. «Чем больше данных, тем больше открытий можно сделать» [цит. по: с. 1]. Данные постоянно накапливаются, и происходит это с огромной скоростью. По мнению специалистов, это может привести к кардинальной трансформации самого исследовательского процесса. Многие ученые уже говорят о «четвертой на-

учной парадигме», поскольку новые алгоритмические, вычислительные и аналитические инструменты способны добывать «золото» из огромного ресурса данных.

На этом пути, однако, встают большие препятствия: БД связаны с рядом научных проблем и опасностей. Данные могут стать для ученых одновременно и «соблазнительными, и досадными»: они часто заставляют исследователей заниматься изучением явлений, идентифицируемых в растущем количестве БД, и игнорировать важные явления, не вписанные в них. Ученые также встречаются множество трудностей, связанных с освоением новых технологий, а также с такой проблемой, как различие данных по количеству и качеству. В связи с этим на первый план выходят фундаментальные онтологические проблемы: спорность эпистемологических обоснований; валидность разделения науки на области; критика методологического монизма; отказ от упрощений.

В отличие от подходов, применяемых в естественных науках, которые успешно справляются с проблемой, вызванной новыми формами данных, традиционные подходы в общественных науках испытывают большие трудности. Центральным среди новых подходов авторы называют «вычислительную социальную науку», которая находится на стыке социальных и вычислительных наук и включает в себя такие методы, как анализ наблюдения данных веб-масштаба, виртуальные лабораторные эксперименты и вычислительное моделирование. Эти методики позволяют решать вопросы, связанные с БД и огромным множеством и сложностью их различных взаимодействий.

Пересмотр границ между естественными и социальными науками, вызванный БД, ведет, по мнению авторов, к возвращению натурализма – «обновленного натурализма» – и «концу теории» [с. 2]. Общество теперь рассматривается не с позиций традиционной картезианской – ньютоновской парадигмы, а как совокупность частиц и потоков по аналогии «с лавинами и гранулированными потоками, стаями птиц и рыб, сетями взаимодействия в неврологии и т.д.» [с. 2]. Ученые говорят о рождении новой, цифровой парадигмы.

Опираясь на прошлые теоретические основания, авторы статьи пытаются выявить эпистемологические пределы цифровой парадигмы. Прежде всего они анализируют дискуссии, доминировавшие в социальных науках в первые дни цифровой эры, когда оцифровка рассматривалась не как нечто, что сделает общество более поддающимся формальным методам исследования, а скорее как полная противоположность – как неотъемлемая часть постмодернизма.

**Современные цифровые исследования: вычисления и отношения.** Несмотря на название «большие данные», размер, пожалуй, не является определяющей особенностью БД: этот термин скорее описывает ряд параллельных явлений в различных дисциплинах, которые объединяет общий знаменатель – распространение огромного количества данных, с трудом вписываемых в существующие парадигмы. В общественных науках проблемы, связанные с БД, отличаются от таковых в естественных науках, в которых это давно не вызывает трудностей: даже огромные количества структурированных данных переписи относительно легко обрабатывать с помощью традиционных инструментов. Главное отличие заключается в *качестве* данных. Эффект БД оказался не методологическим, а связанным с глубокими изменениями на уровне эпистемологии.

Что касается методологии, то этот переход, по мнению авторов, лучше всего обозначить как переход от *математически организованных данных* к *алгоритмически организованным данным*. Извлеченные данные структурируются с помощью алгоритмической обработки, подразумевающей индексированные структуры данных и сети профилей.

Социальная онтология, отмечают авторы, нацелена не на суммирование населения в фиксированных категориях, а на выявление связей и взаимодействий отдельных людей, которые находятся в постоянной динамике. Это подразумевает, что данные производят отталкиваясь не от целого как суммы его частей, а наоборот – от частей и их расположения в структуре данных. Таким образом, БД – возникший естественным путем субпродукт, составленный из отпечатков непрерывного социального процесса. Незавершенность означает, что онтология, выявленная БД, и есть настоящая природа социального мира, которая ранее была скрыта.

Эта идея необработанности, незавершенности ярко проявляется в вычислительных социальных науках как основа нового подхода к изучению социального мира, который, по мнению экспертов, поможет решить многие глубоко укоренившиеся проблемы социологии. Алекс Пентланд называет этот подход «социологией XXI в.»: цифровые данные «дают нам шанс рассматривать общество во всей его сложности, через миллионы сетей личных обменов» [цит. по: с. 3]. Эта новая наука – ответ на кризис старых подходов эмпирической социологии; методы опросов и интервью заменяются новыми методами – интеллектуальным анализом данных и геоинформационным анализом (GIS analysis).

Таким образом, новые данные способствуют сближению социальных и естественных наук в рамках нового подхода и новой онтологии. Различие в социальной онтологии, пользующейся старыми данными, и социальной онтологии, пользующейся новыми данными, во многом соответствует различию между «сложными» (complicated) и «комплексными» (complex) системами. Согласно науке о сложных системах, в их (сложных систем) состав входит множество сложных компонентов с простыми и ограниченными взаимодействиями, в то время как в состав комплексных систем – множество простых компонентов с очень сложными взаимодействиями. Этот подход доказал свою пригодность для анализа многих типов систем, которые были недоступны для формальных подходов. Наука о поведении сложных систем использует формальные модели массовых взаимодействий и сосредоточена не столько на социальных фактах и интерпретации, сколько на выявлении общих образцов и моделей. То, что появляется, рождается в процессе, – полная противоположность тому, что получилось, – агрегации. Целое отличается от суммы его частей. В науках, изучающих сложное коллективное поведение, широко применяется этот принцип сложности, поскольку известно, что существуют заметные различия между индивидуальным и совокупным уровнями. В социальных науках традиционно анализ ведется на разных уровнях, и основной вопрос обычно заключается в том, какой уровень – микро- или макро- – подходит для того или иного исследования.

Вычислительная социальная наука предыдущего десятилетия была частью науки о сложности и никогда не имела дела с БД, скорее она исследовала общество в основном с помощью методов симуляции и агентного моделирования. С появлением БД наука о сложности испытывает своего рода «уничтожение путем инкорпорации», она переходит к строительству неявного основания для многих инструментов и подходов, используемых в общественных науках. Например, в мейнстриме социальных наук широко используется анализ сложных сетей, который хорошо показывает, как отношения и взаимодействия на микроуровне приводят к образованию социальных моделей высших уровней.

Широкое применение подходов, заимствованных из науки о сложности, отражает текущие изменения в характере социального взаимодействия. Они заключаются в том, что цифровая социальная жизнь кажется более количественной, регулярной, предсказуемой (как видно из успехов компаний, занимающихся анализом платформ

и данных), что мотивирует применять более естественный научный подход к данным. Можно выделить два аргумента в пользу этого.

Во-первых, социальные медиа значительно увеличили возможности социального взаимодействия и подсчета числа связей. Количество лайков, полученных постом на Фейсбуке, не требует никаких дополнительных операций для подсчета. Социальная жизнь, таким образом, стала более естественной, более упорядоченной и структурированной. Все чаще используются цифры и кодированные данные для навигации в социальном мире, обмен информацией упрощается с помощью чисел, слов и смайликов.

Во-вторых, социальные медиа повысили распространенность синхронного поведения – в форме каскадов «вирусных» действий. Технологическая сложность вывела людей из рефлексивного состояния ближе к реактивному, анималистическому и инстинктивному поведению, подобному, например, поведению стай птиц и рыб.

Это сочетание социальной жизни – более реактивной, инстинктивной и квантифицированной – и понимания БД как чего-то принципиально нового, сырого и натурального, снова возрождает к жизни наивный натурализм. Благодаря методам и алгоритмам, сформировавшим новые цифровые данные, натуралистическая онтология «принимает форму сложности» [С. 5]. Эта идея размытия границ между естественным и социальным мирами не всегда явно и открыто формулируется, тем не менее она очевидна для большинства ученых, применяющих вычислительные подходы в анализе социального мира.

**Ранние исследования цифровой эры: текучесть и постмодернизм.** В ранних исследованиях цифровые технологии в первую очередь рассматривались сквозь призму дематериализации – перехода технологии от атомов к битам. Основное внимание в литературе уделялось социальным последствиям возможностей быстрых изменений, вызванных цифровыми технологиями: через Интернет технологические изменения могут распространяться на миллиарды пользователей в течение нескольких секунд, и реакции пользователей могут быть мгновенно оценены. Утверждалось, что технологии сократили свою функцию стабилизатора социальных структур: социальный контекст и основа для интерпретации становятся более текучими, нестабильными. Понятия оцифровывания (преобразования данных в цифровую форму) и дематериализации были тесно связаны с горячими дискуссиями вокруг таких терминов, как «постмодерн», «текущая современность» (З. Бауман), «поздний капитализм» и «ускорение». Ученые рассматривали цифровые технологии как часть поздней

современности – неуправляемой и калейдоскопической по форме. Предполагалось, что оцифровывание и дематериализация – это составные части процессов постмодернизма как устраняющего препятствия на пути изменений. Это часть большого процесса современности, в котором «вместо того чтобы жить в стабильном мире долговечных предметов, люди оказались втянутыми в ускоренный процесс производства и потребления» [с. 5].

По мнению многих ученых, цифровизация может быть интерпретирована как еще один переходный этап, на котором капитализм достиг своей чистой формы: «Процесс окончательно растопил саму материальность технологии, и все твердое расплавилось в воздухе» [с. 5]. С этой точки зрения стабильность социального мира тесно связана с материальностью технологии: поскольку материальные изменения, как правило, проходят медленно, технологии обеспечивают прочную основу для социальных моделей, на которую можно опереться. Таким образом, именно эта стабилизация подрывается дематериализацией, принесенной цифровыми технологиями.

Точка зрения, предложенная ранними исследователями информационных процессов, заключается в том, что цифровизация, стимулирующая технологические изменения, приводит к открытости общества как системы, что, в свою очередь, приносит то, что Д.А. Лейн и Р.Р. Максфилд<sup>1</sup> называют «онтологической неопределенностью» – повышенной склонностью к качественным изменениям. Эта склонность иллюстрируется понятием «веб-время», которое описывает увеличение темпов социотехнических изменений.

Таким образом, ранние исследования описывали влияние цифровизации на социальную жизнь в терминах текучести и нестабильности значений и структур, постоянно «кипящих» под поверхностью фиксированных чисел и символов. Однако с тех пор цифровые технологии развивались в непредвиденных, новых в то время направлениях, и это означало, что текучесть смысла и структур стала направляться по неожиданным руслам.

**Текущие технологии в эпоху платформ.** Следует выделить два фактора, характеризующих современное развитие цифровых технологий. Во-первых, сегодня Интернет становится местом крайней централизации: информационные системы создали хорошие условия для крупных платформенных компаний. Сформировавшиеся структуры все в большей степени становятся похожими

---

<sup>1</sup> Lane D.A., Maxfield R.R. Ontological uncertainty and innovation // Journal of evolutionary economics. – 2005. – Vol. 15, N 1. – P. 3–50.

на частные правительства, которые обладают властью контролировать потоки информации и даже брать плату (облагать налогом) с пользовательской базы.

Во-вторых, цифровизация создала условия для более широких обратных процессов инноваций, в частности оценки того, как инновации влияют на социальную сеть, частью которой они и становятся. Различные способы оценки социальных практик – сложный анализ данных, А/В-тестирование, – развивающиеся на цифровых платформах, позволяют владельцам платформ формировать поведение своих пользователей с беспрецедентной точностью и под строгим контролем. Временной цикл между оценкой и инновациями все время сокращается, так как владельцы технологий постоянно получают точные и подробные данные о том, как потребляется их продукция.

Эти два фактора означают, что текучесть и способность к быстрой смене дематериализованной технологии не только играют свою роль в постмодернистской культуре позднего капитализма, но и вылились в новые формы власти для владельцев технологий. Дематериализация технологий означает, что право собственности даже на потребительские товары стало возможным централизовать. Программное обеспечение с нулевой предельной стоимостью привело не к концу капитализма, как предсказывали некоторые социологи, а скорее к бизнес-модели – от продажи к аренде. Другими словами, вместо того чтобы подорвать режим капитала – частную собственность, эффект проявился в разрушении незначительной собственности потребителей. Вместо того чтобы работники приобретали в собственность средства производства, они все чаще теряют владение даже своими товарами потребления.

Таким образом, растущая фрагментация и текучесть, следующие из дематериализации, подразумевают усиление централизации контроля, поскольку они позволяют владельцам технологий властвовать и контролировать, осторожно формируя смыслы и структуры. Контроль перемещается в нижние онтологические слои, формируя результаты скорее посредством основных правил взаимодействия, чем с помощью явного регулирования. В этой дематериализованной современности текучесть смыслов и структур позволяет использовать парадоксальную форму контроля – снизу вверх. Власть в эпоху цифровых платформ осуществляется не сверху вниз, а через невидимое и осторожное подталкивание потребителей и формирование их поведения, формирование социальных правил и практики, следовательно, контроль заложен в самих правилах взаимодействия. Этот переход от технологии в качестве слепого инструмента для социального кон-

троля к технологии в качестве «виртуального социального скальпеля» подразумевает, что цифровизация принесла новую эру – эру мощи платформ, в которую технологии обеспечивают новый уровень «управления стадом» [с. 7].

**Природа цифровых данных.** Революция БД описывает мир как мир невообразимого ранее количества данных, потоков следов жизни, снов и чувств сотен миллионов людей. Цифровая социальная наука получила беспрецедентную возможность раскрытия истинного характера социальной жизни, которую можно теперь не только измерить, но и прогнозировать. Появилась возможность наблюдать, как производятся данные, отбираются и предоставляются владельцами платформ, преследующими собственные интересы. Однако многие аспекты исключены из этих данных.

Между тем внимание исследователей сосредоточено на моделях взаимодействия, которые, как только теряют свою естественную обстановку, натурализуются и деконтекстуализуются. Когда БД рассматриваются как просто закодированные, измеримые версии социальной реальности, сложные социальные и технологические силы, которые их произвели, «сплющиваются»: данные сделаны так, чтобы казаться скорее естественными и неизбежными, чем случайными и оспариваемыми; они сделаны скорее предметом овеществления, чем предметом критики. Вместо простого одностороннего кодирования «производство данных осуществляется с помощью цифровых платформ, которые направляют и ограничивают процесс с помощью “грамматики действий”, делающей определенные действия выполнимыми, что позволяет измерять, анализировать, коммодифицировать социальные действия и манипулировать ими»<sup>1</sup>.

Поведение и содержание в Интернете – следствие того, как работают цифровые технологии и что люди делают с ними. Содержание понимается как результат взаимодействия (спутанности) двух факторов – человеческого поведения и технологической составляющей, которые и представляют социально-техническую систему.

Данные, таким образом, делают более заметной «игру в игру», скрывая «правила игры» и интересы, их сформировавшие. Таким образом, данные идеально подходят для натурализующей науки, которая стремится видеть правила как универсальные и результаты их работы – как неизбежные. Интерпретация данных не стала менее важной в исследовательском процессе; ее локус

---

<sup>1</sup> Van Dijck J. The culture of connectivity: A critical history of social media. – Oxford: Oxford univ. press, 2013. – 33 p.

просто сдвинулся, поскольку взаимодействие является одновременно и более количественным, и более фрагментированным и «мерцающим». Большие данные парадоксальным образом связаны с обеими этими позициями: они одновременно и более текучи, и более квантифицированы; и более открыты, и более измеримы; они образуются в большей степени «снизу вверх» и больше поддаются контролю. Иными словами, их становится легче подсчитать, но намного труднее интерпретировать.

**Заключение.** Таким образом, заключают авторы, общество не является ни сложной, ни комплексной системой, скорее оно отображает свойства обеих систем, которые делают его качественно отличным от них. Технологическая власть владельцев платформ в большой степени обеспечивается теми же новыми инструментами для анализа данных, которые используют социологи, и частный сектор часто является движущей силой для развития этих инструментов. Но цели корпораций всегда значительно отличаются от целей исследователей: корпорации стремятся получить возможности прогнозирования и контроля, а исследователи стараются понять и объяснить мир.

Исследователи должны не просто изучать процессы возникновения данных, но изучать их, имея в виду, что нет ничего естественного в поведении человека и что нет такой вещи, как «необработанные данные». Возможно, им следует думать о социальных сетях не как о «саваннах свободно бегущих стад людей, а как о зоопарках, в которых пользователи в клетках вынуждены танцевать под звуки капитала: “Нет данных, больших или малых, которые могут быть интерпретированы без понимания процесса, который их породил”» [цит. по: с. 10]. И эти процессы запутались в интересах капитала.

Назрела острая необходимость в критической цифровой социальной науке, которая не жертвует контекстом, ясностью и критикой ради автоматической идентификации крупномасштабных моделей, основанных на представлении о том, что ширина может заменить глубину и контекст как основу для интерпретации. Лишь имея стабильные онтологические позиции, исследователи смогут увидеть не только то, что данные говорят о социальном мире, но и те вещи, о которых «они так странно молчат».

*М.П. Булавинова*

**ОБЩЕСТВЕННЫЕ НАУКИ И ВЕБ-РЕСУРСЫ:  
ОТ «СКРЫТНОГО» К МЕЖДИСЦИПЛИНАРНОМУ  
ИССЛЕДОВАНИЮ БОЛЬШИХ ДАННЫХ / БОУН Дж.,  
ЭМИЛЬ Ч., АБДУЛ А., КОГХИЛЛ Дж., ПАН УЭЙ  
(Реферат)**

Ref. ad op.: The social sciences and the web: From «lurking» to interdisciplinary «big data» research / Bone J., Emele Ch., Abdul A., Coghill G., Pang Wei // Methodological innovations. – 2016. – Vol. 9. – P. 1–14. – DOI: 10.1177/2059799116630665

Авторы из Великобритании поднимают ряд проблем, вставших перед социологической наукой в связи с появлением чрезвычайно интересной для социологии области больших данных (БД) – огромных массивов социологических данных, размещаемых на сайтах в Интернете и легкодоступных для исследователя. С этим связаны этические проблемы, касающиеся законности доступа к обширной персональной информации, а также проблемы сбора и анализа широчайшей и бесконечно разнообразной информации, представленной на всевозможных сайтах, форумах, в блогах и т.д. Поднимая такие проблемы, авторы ставят своей целью понять и описать новый инструментарий социологического исследования – инструментарий исследования онлайн.

В результате развития компьютерных технологий в их многообразном приложении и связанного с этим экспоненциального роста социальной коммуникации использование данных с сайтов Интернета практикуется в широком спектре областей научного исследования. Сегодня у социологической науки есть прямой интерес к онлайн-исследованию, и прежде всего интерес к позднему типу онлайн-исследования в форме так называемых больших данных [с. 1]. Такое развитие представляется неизбежным, если видеть рост масштабов хранения информации, что становится

главной особенностью Интернета и ключевым интересом его пользователей. Также признано, что этот новый и бурно растущий ресурс создает для социологов как определенные вызовы, так и хорошие исследовательские возможности. Прежде всего ясно, что положение традиционного социологического исследования «пошатнулось» из-за возрастающей аналитической активности в отношении социологических данных со стороны коммерческих организаций. Кроме того, неуклонное возрастание сегодня значения Интернета как средства массовой коммуникации, содержащего громадное количество мнений, комментариев и индивидуальных самопрезентаций, уже радикально реформировало социальное взаимодействие / социальный обмен – базу общественной жизни, общественной динамики. Все это открывает огромные возможности для исследователей-обществоведов, которые вынуждены теперь переоценивать не только свои методологические подходы к пониманию процессов социальной коммуникации и социального взаимодействия, но и сами теории общества.

В настоящее время происходит стремительная эволюция и самого цифрового ландшафта, и исследовательского инструментария, призванного в этот ландшафт вписываться. Исследователями уже пройден определенный путь от робкого вхождения в этот огромный виртуальный мир к полномасштабному сотрудничеству с учеными-компьютерщиками с целью более полной реализации исследовательского потенциала богатого нового ресурса. Примерно 10 лет назад все начиналось с освоения таких ключевых платформ, притягивающих исследовательское внимание, как Фейсбук, Инстаграм и Твиттер, тогда они находились еще в «детском» возрасте. Но уже и тогда эти платформы давали гораздо больше информации и гораздо меньшей ценой, чем было в эпоху до Интернета, причем даже при использовании традиционных (старых) исследовательских методов. Сегодня в распоряжении социолога, даже не имеющего финансирования и раньше с борьбой добивавшегося проведения интервью с маленькой выборкой респондентов, находятся сотни, тысячи, а то и миллионы (в случае с Твиттером) «респондентов» при явно меньших затратах усилий и времени. Также признано, что скрытное присутствие исследователя на форумах в Интернете позволяет избежать проблем, связанных с эффектом «галло», неизбежно возникающим, когда респонденты знают, что они респонденты – участники исследования. Но и наоборот – в новой исследовательской реальности существует вопрос «честности» онлайн-респондентов, которые становятся для исследователя респондентами не по его выборке, а сами собой, и которые

поэтому вполне могут создавать ложные о себе впечатления без какого-либо исследовательского контроля. Это действительно большая проблема для работающего в виртуальном мире социолога. Самоотбирающиеся (self-selecting) группы респондентов могут формировать искаженное представление о проблемах, что сильно затруднит оценку исследовательских данных, если вообще не сделает ее невозможной. Вместе с тем, напротив, некоторые комментаторы развития реальности БД предполагают, что само по себе разрастание данных, их многообразия и будет сводить указанные искажения к минимуму, делать эти искажения хотя и неизбежной, но несущественной статистической величиной.

Как ясно из дискуссий вокруг онлайн-этики, наряду с возникновением новых дилемм для онлайн-исследования остаются и некоторые вечные вопросы традиционного социологического исследования. Но не только цифровой ландшафт вместе с соответствующими исследовательской техникой и исследовательскими умениями претерпевает сегодня громадную метаморфозу, трансформируются и этические проблемы исследования. Возникает важный вопрос – обеспечит ли появление этого нового ресурса путь к лучшему пониманию общества и более рациональному принятию решений или приведет к эксплуатации и контролю со стороны властных интересов, которые станут использовать инструмент «фетишизации и диктатуры данных» [с. 4]. Мир онлайн-данных рассматривается некоторыми экспертами по аналогии с «массовым обществом» середины XX в., когда его высокая атомизация, т.е. высокая степень отчуждения людей друг от друга – антипод общественных отношений доверия, – давала элитам хорошую возможность манипулировать людьми, превращая индивида в «массового человека», человека толпы.

Наряду с тем что необходимо заботиться о возможном этическом (или неэтическом) использовании БД, в той же мере важно видеть способность этого инструмента открывать новые горизонты в общественном развитии – решении управленческих, исследовательских проблем. Человек, общество всегда и по самой своей природе связаны с этическими проблемами. Любое технологическое преобразование общества имеет этическую составляющую, иначе человек перестал бы быть человеком, и вхождение современного общества в мир БД – не исключение. Например, с политической точки зрения исследовательская техника сбора и анализа БД несет в себе разный потенциал. С одной стороны, открывая широкие пути для проявления общественного мнения, она вынуждает политиков к

большей ответственности в отношении нужд и ожиданий избирателей, а с другой – напротив, может стать в руках тех же политиков средством манипулирования общественным мнением. Поэтому ничего нового в этике исследовательской работы с БД нет: «Любого масштаба использование исследовательской техники в конечном счете зависит от тех, кто ее использует, – от их этики. Они обязаны постоянно взвешивать свои действия с точки зрения последствий этих действий для онлайн-респондентов и общества в целом» [с. 4].

Ясно, что технологический прогресс, подведший социологов к сбору и анализу БД, огромен до такой степени, что многое в социологической науке переходит в компетенцию компьютерных наук. Таким образом, формируется новое междисциплинарное поле исследований в области общественных наук, основой которого выступают именно компьютерные науки. Исследовательский союз ученых-обществоведов и ученых-компьютерщиков представляется скорее трансдисциплинарным, нежели междисциплинарным, поскольку речь идет о сотрудничестве разного рода наук.

Что касается методологии исследования в трансдисциплинарной области, то есть основания сфокусироваться на данных Твиттера – платформы, признаваемой социологами. Социальная сеть Твиттер – гибкая коммуникативная структура: пользователи могут включать хештеги (ключевые слова с прибавлением символа «#») в свои твиты, чтобы сделать сообщения видимыми для тех, кто следит за хештегом. Решающим в платформе Твиттера является то, что хештеги обеспечивают механизм общения пользователей, связи между ними, даже если эти пользователи уже не «следят» друг за другом в социальной сети. Главная проблема исследователя коммуникации в Твиттере – охватить исчерпывающую или по крайней мере репрезентативную выборку твитов, связанную с исследуемым событием или сообществом. Эта проблема может быть решена довольно просто – с помощью выборки твитов, содержащих хештеги на тему исследования. Твиттер обеспечивает доступ к твитам посредством двух ключевых элементов своего «приложения, программирующего интерфейс» (application programming interface) (ППИ): ППИ-поиска и ППИ-потока. Из этих двух элементов первый используется для нахождения прошлых твитов (по ключевым словам / хештегам, отправителям и т.д.). Причем ППИ-поиск будет находить лишь ограниченное число твитов, он не даст весь их архив, например твиты, содержащие специфические хештеги. К тому же существуют запрограммированные ограничения на количество ключевых слов или запросов пользователей в единицу времени. Если ППИ-поиск фокусируется на прошлом

контенте, то ППИ-поток, напротив, используется для подписки на непрерывный поток новых твитов, содержащих специфические ключевые слова. Но и в случае ППИ-потока имеются существенные ограничения на число пользователей или ключевых слов.

При этих ограничениях исследователю, стремящемуся установить разумно широкий ряд данных в отношении исследуемого события, будет нужно проследить за событием с того момента, когда оно произошло, иначе исследователь, пользующийся ППИ-поиском, рискует упустить важную раннюю информацию о событии, которую ППИ-поиск может не дать. Должна быть охвачена вся последовательность твитов либо с помощью ППИ-потока – подписки на идущий поток твитов, либо с помощью программы ППИ-поиск, используемой в режиме постоянного поиска позднейших прошлых твитов. Но даже и такой метод использования ППИ-поиска не может гарантировать исчерпывающий охват данных Твиттера. Кроме того, существует очень мало надежных средств для проверки данных на достоверность, поскольку Твиттер не имеет такой программы. Тем не менее, несмотря на все эти ограничения, такое исследование остается весьма важным, поскольку в любом случае охватывает огромную выборку твитов – информацию гораздо более содержательную, чем те данные, которые способен получить исследователь, применяющий стандартные методы социологического анализа.

Есть мнение, что поскольку Твиттер и прочие платформы для социальной коммуникации онлайн – относительно новые и междисциплинарные подходы к качественному и количественному исследованию данных с этих платформ не стандартизированы, это не дает возможности воспроизводить такие исследования в схожих контекстах. Поэтому онлайн-методология социологического исследования нуждается в детализации, чтобы показать ее применимость во всем спектре исследований в междисциплинарном / трансдисциплинарном поле сотрудничества общественных и компьютерных наук.

Требуемая детализация методологии должна основываться на том факте, что расшифровка, «добывание» мнений, высказываемых в текстах пользователей Интернета, – базовая задача аналитики в отношении сетей социальной коммуникации. Этот материал дают большинство платформ, использующих инструмент блогов, что позволяет пользователям выражать свои взгляды и мнения в единстве текста, образов и смысловых «картинок». Это и есть данные, собираемые аналитиком, который пропускает их через сито техник

статистического и текстового анализа для выявления и сравнения между собой образцов коммуникации. Такой анализ обеспечивает инструменты описания активности в Твиттере, с помощью которых можно, например, определить частоту употребления искомым терминов / ключевых слов. Далее идентифицируются темы и создаются тематические модели, позволяющие вероятностное моделирование частот употребления значимых терминов. Можно также использовать техники анализа тональности текста (*sentiment analysis techniques*) для оценки позиций по определенным темам, связанным с ключевыми словами.

Нужно просто хорошо осознавать, подчеркивают авторы, и растущие возможности, и сложности этих междисциплинарных / трансдисциплинарных исследований нового типа, в которых стираются границы между дисциплинами, весьма далекими друг от друга по своим предметам, объединяются радикально разные научные интересы, методологические подходы, исследовательские умения, техники, дисциплинарные лексиконы. Надо понимать, что Интернет с его сетями социальной коммуникации поставил общественные науки перед необходимостью радикального сдвига в исследовательском инструментарии, в самой исследовательской коммуникации, со всеми вытекающими последствиями, главное из которых – формирование трансдисциплинарного исследовательского поля (объединения общественных и компьютерных наук) с ломкой многих элементов в традиционной исследовательской практике, в том числе в сфере этики науки.

Действительно, пишут авторы в заключение, «сама пропасть между дисциплинарными практиками ученых-обществоведов, с одной стороны, и ученых-компьютерщиков – с другой, должна фрустрировать тех и других вследствие необходимости сотрудничать в сложившихся условиях, мотивируя к такому “странному” – трансдисциплинарному – сотрудничеству. Вместе с тем эти фрустрирующие вызовы для ученых, которые привыкли существовать в комфортной зоне своих дисциплин, автономно друг от друга, сохраняются, и есть соблазн вернуться с новой исследовательской территории к испытанным и проверенным методам. Однако такое возвращение вполне может оказаться возвращением в никуда – порог невозврата для общественных наук уже пройден» [с. 13].

*А.А. Али-заде*

**Лоури А.**

**АЛГОРИТМИЧЕСКАЯ РАЦИОНАЛЬНОСТЬ:  
ЭПИСТЕМОЛОГИЯ И ЭФФЕКТИВНОСТЬ  
В ИНФОРМАЦИОННЫХ НАУКАХ  
(Реферат)**

Ref. ad op.: Lowrie I. Algorithmic rationality: Epistemology and efficiency in the data sciences // Big data & society. – 2017. – Vol. 4, N 1. – P. 1–13. – DOI: 10.1177/2053951717700925

Автор из США, основываясь на своем опыте 12 месяцев полевого исследовательского (этнографического) общения с российскими учеными-информационщиками, моделирует подсказанную этими специалистами новую эпистемологическую реальность, которую можно назвать эпистемологией алгоритмов и которая в связи со вторжением в общественные науки феномена больших данных (БД), большой информации стала философско-научным сопровождением прочного альянса обществознания и оперирующих алгоритмами информационных наук.

Современные информационные данные представляют собой огромный объем информации, пишет автор. Их сбор производится постоянно и повсеместно, а их хранение и анализ вызывают большие проблемы для всех, кто этим занимается. Сбор данных – главная забота государственного управления, по крайней мере начиная с развития статистики в конце XVIII в. За последние же два десятилетия произошел взрывной рост объема информации, который связан с радикальным умножением центров ее производства, сбора и анализа. В этот же период параллельно с ростом объема информационных данных возникли новые формы интеллектуального и прагматического отношения к информации. «В то время как каждая научная дисциплина должна работать с БД своими методами, существует и

некое единство методов и теорий работы с большой информацией – единая интеллектуальная практика ученых-информационщиков, производящих согласованную форму алгоритмической рациональности, не сводимую к сумме практик или теоретическим “лоскутам”, надерганным из математики или компьютерной науки. Эту рациональность невозможно вывести из появления инфраструктуры больших данных и дальнейших “напластований” компьютерной реальности. Исследование того, как возникают, распространяются и взаимодействуют подобные формы научной экспертизы, представляется важнейшей задачей социологии современного знания» [с. 1].

Для решения этой задачи автор предпринял собственное исследование этнографического характера. В 2014/2015 acad. г. он провел 12 месяцев в России, общаясь с российскими учеными-информационщиками. Он хотел понять не интеллектуальный, а социально-культурный механизм формирования феномена алгоритмической рациональности. В своем исследовании автор использовал свободно котируемые респондентские выборки среди примерно 80 ученых-информационщиков преимущественно из Яндекса и Высшей школы экономики, а также из других университетов, институтов и корпораций, имеющих активные исследовательские программы в области информационной науки. Проводимые им интервью были справедливо разделены между академическими респондентами и респондентами-производственниками. Они продолжались от одного до двух часов и дополнялись информацией, почерпнутой на научных семинарах, деловых встречах, информацией о производственных событиях, дискурсным анализом научных статей и т.д. Цель всего этого «гранулированного» исследовательского наблюдения состояла в том, чтобы получить ясную картину интеллектуальной жизни, карьерных траекторий и повседневных рабочих будней российских ученых-информационщиков.

В своих исследованиях антропологи обычно ищут радикальную специфику, поэтому и антропология интеллектуальной работы главным образом сосредоточивается на том, чтобы понять, как возникает не знание в целом, а специфическое, «местное» (здесь и теперь) знание. Она фокусируется не на экологии знания в целом, а на местной (в данном случае российской) экологии знания. Построение в России институтов и инфраструктуры экономики знаний – невероятно трудная задача. Причина отчасти заключается в традиционном в стране верховенстве теоретического знания над знанием прикладным и наличии относительно высокого барьера между академической наукой и индустрией. В таких об-

стоятельствах, возможно, развитие информационных наук подтолкнет процесс трансформации «людских ресурсов» в «человеческий капитал», трансформации, необходимой для перехода от экономической архаики (добывающей экономики) к экономической современности (экономике знаний).

Интересно, что когда автор спрашивал российских специалистов-информационщиков, что есть интеллектуально уникального в российской информационной науке, они удивлялись такому вопросу и отвечали, что они не уникальны, они – часть мирового научного сообщества (хотя и, самокритично замечая, – маргинальная часть). Разумеется, можно согласиться с тем, что операционное ядро информационных наук не подвержено заметным географическим вариациям, но важно иметь в виду институциональный и культурный контексты конкретизации этого ядра – собственно российские респонденты заявили вполне личностное отношение к своей науке, как и вполне «местное» свое видение того, что означает быть ученым. В связи с этим автор предупреждает, что его исследование «должно рассматриваться не как исчерпывающая разработка эпистемологической динамики информационных наук, не как нечто готовое, а скорее как “сочувственный” анализ интеллектуального проекта ученых-информационщиков, пытающихся выстроить целостную и эпистемологически изошренную модель алгоритмической рациональности» [с. 2].

Существует научная литература, посвященная техническим изменениям в области анализа и хранения БД, рассматривающая процессы построения технологической инфраструктуры работы с ними и ее взаимодействия с широкими инфраструктурными изменениями в государственном управлении, экономике потребления и науке. Есть исследования, изучающие работу профессиональных и непрофессиональных пользователей с БД. Однако мало внимания уделяется вопросам построения эпистемологических рамок, внутри которых технологии БД осуществляются и развиваются. Автор считает, что его исследование устраняет этот пробел: это – «исследование, которое входит в тесный диалог с теми, кто стремится понять эпистемологию, порожденную БД, исследование, анализирующее ту научную работу, которая в качестве своего объекта берет алгоритмы – математически формализуемые процедуры оперирования с информационными данными» [с. 2–3].

Респонденты, участвующие в исследовании, уверенно считали себя учеными, а свою работу – скорее наукой, чем инженерным делом. Эта твердая уверенность в том, что они занимаются

именно «наукой» об алгоритмах, отчасти связана со стремлением установить интеллектуальную основу институционализации новой исследовательской области, отличной от уже существующих областей математики и софтверной инженерии. Другая причина убежденности в том, что они добывают теоретическое знание об алгоритмах, обязана особенностям российской интеллектуальной среды, отдающей предпочтение абстрактному теоретическому мышлению и фундаментальным исследованиям в ущерб прикладной науке и инженерии. Хотя стремление российских специалистов-информационщиков очистить информационные науки от прикладных и технологических измерений, несомненно, происходит из российской институциональной и социально-культурной среды, нет оснований сомневаться в искренности их стараний достичь глубокого теоретического понимания объективных характеристик алгоритмов. Тем не менее в своих попытках отделить науку от технологии они не могут отрицать критерий эффективности, поскольку, в отличие от многих других объектов научного исследования, алгоритмы не имеют очевидной «объектности» и с ними приходится работать как с математическими абстракциями, оценивая их функционирование внутри вычислительных полей. По сути дела, «российские специалисты в области информационных наук пересмотрели критерий технологической эффективности, поставив его в качестве эпистемологического стандарта на службу новой форме научного исследования» [с. 3].

За малыми исключениями, продолжает автор, корневые операции научного исследования феномена БД заимствованы. Это не должно удивлять, поскольку информационная наука находится в междисциплинарном поле статистики, компьютерных технологий и математики и вполне предсказуемо пользуется многими методами этих дисциплин. Она их преобразует и по-новому сочетает, так что на новой почве они оказываются методическими инновациями. Структура информационной науки характеризуется двумя сквозными различиями: между специфическими аналитическими подходами и между специфическими областями проблем. Существуют два уровня исследований в этой области, которые выстраивают науку БД: 1) разработка конкретных областей прикладных задач; 2) эпистемологическое исследование природы и функционирования алгоритмических ансамблей, которые в таких разработках используются. Автор отмечает, что его российские информаторы последовательно подчеркивали приоритет исследовательского уровня (2). «Однако такой выбор есть некое паразитирование на уровне (1): в конце кон-

цов, сами по себе алгоритмы – инертные, лишённые смысла механизмы, пока они не работают на базы данных» [с. 10]. Базы данных должны откуда-то прийти, и было заметно в отношении российских ученых-информационщиков, что они предпочитают работать с теми рядами данных, которые уже существуют в публичном пространстве, свойства которых уже установлены. Между тем с эпистемологической точки зрения новые ряды данных – это возможность строгого тестирования работы наличных алгоритмов и эмпирическое основание для возможного появления эпистемологической инновации. В общем встреча между алгоритмами и новыми данными обещает многое, поэтому не удивительно, что ученые-информационщики и их исследовательские системы – «жадные» потребители данных. Сама природа исследований в области науки БД принуждает к связи с тем, что российские информационщики называют «миром прикладных задач» [с. 11].

Тем не менее, пишет автор в заключение, «судя по высказанной российскими учеными-информационщиками консолидированной позиции, они явно предпочитают теоретическую работу решению конкретных проблем. Такая позиция, однако, – исключительное заблуждение: тот, кто ее исповедует, соглашается, что критерий эффективности, а это главный критерий работы алгоритмов, никогда не может быть внутренним критерием находящейся на такой позиции исследовательской системы, скорее он должен фигурировать только в рамках диалога со специалистами в прикладных областях. Но представляется невероятным, что те, кто стоит на позиции теоретического пуризма, когда-нибудь сумеют полностью очистить свою исследовательскую практику от технологических / прикладных материй» [с. 11].

*А.А. Али-заде*

**МЕТОДЫ КАЧЕСТВЕННОГО АНАЛИЗА  
И ДАННЫЕ В УСЛОВИЯХ ЦИФРОВОГО ОБЩЕСТВА /  
ХАУСЛИ У., ДИКС Б., ХЕНВУД К., СМИТ Р.  
(Реферат)**

Ref. ad op.: Qualitative methods and data in digital societies /  
Housley W., Dicks B., Henwood K., Smith R. //  
Qualitative research. – 2017. – Vol. 17, N 6. – P. 607–609. –  
DOI: 10.1177/1468794117730936

Взрывной поток социологических данных, возникший в результате развития информационных и коммуникационных технологий, пишут авторы, стал вызовом для общественных и информационных наук, потребовав от них ответа на наступающий кризис эмпирической социологии. Этот «потоп данных» создал для исследователей во всех областях общественных наук, с одной стороны, массу возможностей, а с другой – массу проблем, поселив в исследовательском сообществе разные настроения – от энтузиазма до скепсиса и консерватизма. В связи со столь радикальным изменением социальной – коммуникационной и информационной – практики человека крайне важное значение приобретает органичный для познания человеческой / социальной реальности качественный анализ, методические инновации которого предоставляют новые возможности и помогают справляться с проблемами эры больших социологических данных (БСД).

Например, есть опыт применения принципов философии Л. Витгенштейна к анализу информации в социальных сетях (social media). Суть этой методики состоит в том, что «пользователи социальных сетей рассматриваются как те, кто уже погружен в анализ социально-сетевой информации, сам выступает аналитиком, непосредственно развивающим “аналитику больших социологических данных” (big social data analytics)» [с. 607]. Вместе с тем использо-

вание цифровых методов, определяющее социальные сети в качестве БСД, часто упускает онтологические и эпистемологические аспекты социально-сетевого жизненного мира. Кроме того, формальные попытки подобного использования для «дистанционного» (thousand foot view) тестирования теорий в общественных науках несут опасность игнорирования центральных для социально-сетевой практики интерактивных и лингвистических эффектов.

Существуют также разработки, исследующие социально-сетевую коммуникацию через такие популярные интернет-платформы, как Твиттер и Фейсбук, методами онлайн- и офлайн-наблюдения в свете этнометодологических принципов. Неизвестный ранее феномен сетевых сообществ вызвал необходимость разработки этнографического подхода, поскольку сущность всякого сообщества, всякой социальной группы в принципе является этнографической. Этот подход предполагает изучение социально связанных между собой людей по картине поведенческих и коммуникационных образцов в данной субкультуре. Кроме того, взаимоотношения в сетевых сообществах неиерархические, горизонтальные, и сетевые сообщества могут дать ключ к пониманию механизмов социального сотрудничества, актуального в настоящее время.

Особый исследовательский интерес во времена социальных сетей представляют сообщества, склонные к криминальной деятельности. Члены таких сообществ используют вымышленные имена и прочий «камуфляж», включая мобильные коммуникационные платформы. В изучении подобных скрытых онлайн-сообществ методический инструмент цифровой этнографии (digital ethnography) имеет ключевое значение.

«Цифровые коммуникационные технологии, – пишут авторы в заключение, – предельно демократизировавшие и сделавшие глобальной коммуникацию, создали не только феномен “больших социологических данных”, но и социальную основу БСД – беспрецедентно многосубъектное общество, в котором понятие “Я” стало важным для всех, кто широко вовлечен в производство БСД. Концептуализация этого феномена многосубъектности внутри цифровой социальной среды в качестве фактора производства БСД – новая и интересная исследовательская задача, решение которой создаст теоретико-методологическую рамку выстраивания общественных наук, адекватных времени БСД» [с. 608–609].

*А.А. Али-заде*

**Джуитт К., Прайс С., Седо А.**

**ИССЛЕДУЯ ЧЕЛОВЕКА В ЦИФРОВЫХ КОНТЕКСТАХ:  
К НОВЫМ МЕТОДОЛОГИЧЕСКИМ ДИСКУССИЯМ –  
ОТ МЕТОДОВ В ПРАКТИЧЕСКИХ ОБЛАСТЯХ  
К МЕТОДОЛОГИИ ОБЩЕСТВЕННЫХ НАУК  
(Реферат)**

Ref. ad op.: Jewitt C., Price S., Sedo A. Conceptualising and researching the body in digital contexts: Towards new methodological conversations across the arts and social sciences // Qualitative research. – 2017. – Vol. 17, N 1. – P. 37–53. – DOI: 10.1177/1468794116656036

Авторы из Великобритании (К. Джуитт, С. Прайс) и США (А. Седо) полагают, что качественное исследование (qualitative research), органичное для общественных наук, могло бы развивать свою методологию, все более расширяя и углубляя ее, на основе освоения так называемых методов-умений (arts-methods), т.е. эмпирических (не опосредованных теорией) методов. По их мнению, необходимо пересмотреть образ человеческого (body) – его фрагментированность и зональность, сенсорные и материальные качества, границы – человеческого, помещенного именно в цифровые контексты.

Человеческое, пишут авторы, тесно связано с социальным – отношением людей к себе, друг к другу и окружающей действительности. Цифровые технологии – это фундаментальная часть современного социального опыта, требующая пересмотра его человеческого измерения. Вследствие развития технологий человек становится субъектом интенсификации развития, генерирования идей, бесконечных реконструкций. Развитие цифровых технологий ставит вопросы о взаимосвязи общества, природы и культуры, оспаривая все прежние

идеи об этой взаимосвязи. Поворотным моментом стала манифестация новой человеческой реальности в виде киборгов – машинно-организменных гибридов (hybrid machine-organism), интерфейсов, не обремененных физической телесностью. «Наступившую вследствие развития цифровых технологий новую социальную и человеческую реальность можно квалифицировать как реальность постгуманистическую, а постгуманизм – как междисциплинарную платформу для радикального переосмысления того, что такое человек и каковы могут быть цели этого “нового человека”» [с. 39]. Исследователи, работающие в постгуманистической парадигме, отвергают дуализмы человеческого / нечеловеческого, технологического / человеческого «Я», природного / научного. Постгуманизм привлекает критическое внимание к тому, что составляет и не составляет понятия «человек» и «человеческая сложность», предлагая изменчивый, эфемерный, множественный, текучий образ «человеческого», переопределяющий место в мире человечества как одной из многих форм жизни.

Теория постгуманизма породила теории нового материализма, призванные ответить на поставленный современным развитием вопрос о материальном измерении культуры. Исследователи – представители школы нового материализма – изучают проблему вещественности / материальности в сочетании с научным и технологическим развитием, рассматривая материю как то, что бесконечно рождается, формируется и перестраивается, а не то, что существует априори, статично, фиксированно и пассивно. С позиции нового материализма невозможно отделить материальные объекты от их жизненных контекстов, от их практик. С разрушением в новом материализме дуализма субъекта и объекта природа перестала рассматриваться как просто нейтральный пейзаж, в который агрессивно и авантюрно вторгается человеческая культура, и человек оказался не чуждым природе существом, а всего лишь одним из многих природных «тел» / материальностей.

Новый материализм, не противопоставляющий человека природе, субъекта объекту, материальное нематериальному, ставит тем самым интересные вопросы в отношении человеческой материальности. Человек уже не дуалистичен, как нематериально-материальное / духовно-телесное существо, но представляет новую материальность / телесность, органичную в среде цифровых технологий. Цифровая среда действительно создала «нового человека», который от этой среды-интерфейса неотделим, составляя особую ее телесность – «цифровую телесность» (digital body). Исследованием границ цифровой телесности, т.е. человека цифровой социальной

среды, и должны заниматься общественные науки, для них это методологический вызов. «Переосмысление, вследствие развития цифровых технологий, что такое человек, требует радикального изменения методов качественного исследования. Требуется новая методология для новых времен – методология XXI в.» [с. 40].

Авторы полагают, что построение искомой методологии могло бы основываться на результатах реализации проекта «Методологические инновации в цифровой практике и общественных науках» (Methodological innovations in digital arts and social sciences) (МИЦПОН), а также на результатах этнографического исследования в высокотехнологичных практических областях (дизайн, мода и др.) и общественных науках (образование, психология, социальное взаимодействие). Этнография здесь важна, поскольку помогает понять, как происходит совместное становление «нового человека» и формирующей его цифровой среды и как производятся соответствующие смыслы. Этнографический подход полезен потому, что подключает респондентов к пониманию исследовательских задач и методов, превращая их из объектов в субъектов исследования. В сформированных таким образом (благодаря этнографическому подходу) исследовательских группах обсуждается весь ход исследования в мультимодальном формате, рождающем смыслы из разных смысловых источников, каждый из которых обеспечивает свои отдельные возможности и ограничения для формируемого смыслового целого. Сочетание мультимодальности и этнографии оптимально, поскольку исследовательский подход, использующий только инструмент беседы, не способен зондировать среды, где главным являются не слова, а дела и свершения. «Подобному комбинированному подходу становится доступной, хотя бы отчасти, сложность исследуемой области: этнографические данные, собранные по такой (комбинированной) методологии, помогают поместить эту сложность в конкретный социальный, культурный и исторический контексты» [с. 41].

Используя такую комбинированную методологию, авторы провели собственное исследование, в котором участвовали следующие шесть исследовательских групп.

1. Группа дизайна (design group) сосредоточилась на данных, полученных от преподавателей в дизайнерских колледжах, на данных о том, как дизайнеры получают доступ к нужной им информации, отбирают (составляют) и интерпретируют ее, используя эклектические методы практического свойства (когда практика и есть метод) для проведения междисциплинарных исследований, которые

сочетают теоретические извлечения из дизайна продукции и графического дизайна с концепциями человека из философии, социологии, антропологии, географии культуры, физиологии и биологии.

2. Группа моды (fashion group) работала с Центром технологий и инноваций в области моды в колледже моды, изучая опыт Центра по экспериментированию в области моды, инновационному использованию цифровых технологий и профессиональный опыт человеческой коммуникации вместе с соответствующими методами и теоретическим багажом, включающим феминистские и гендерные теории, идеи из материальной антропологии, культурологических и медиаисследований.

3. Группа акционизма (performance group) изучала работу университетской студии цифрового акционизма, сочетающую художественный акционизм, театр, живое искусство (live art) и новые медийные технологии, а также исследовательский интерес к боди-акционизму (body-centered performance), нацеленный на создание инновационного и экспериментального акционизма с помощью новых технологий; изучался и метод этой студийной работы – «практика как исследование» (practice as research), теоретическое ее сопровождение, взятое из философии, феноменологии, социологии, антропологии, культурологических и медиаисследований, философии науки и технологии, нейронауки и биологии.

4. Группа образования (education group) занималась университетской исследовательской командой, проводящей социологические исследования по социальному взаимодействию в цифровых средах обучения (школах, музеях).

5. Группа психологии (psychology group) следила за исследовательской работой в области возрастной психологии – за изучением проблемы индивидуального восприятия физического пространства и пространственного ориентирования с помощью методов наблюдения, психометрии и интервью.

6. Группа социального взаимодействия (social interaction group) курировала небольшую академическую команду, исследующую взаимодействие и коммуникацию в хирургических операциях с целью выстроить концепцию человека в реальной среде взаимодействия.

Указанные шесть групп были подобраны таким образом, чтобы они могли обеспечить доступ исследования к методологиям человеческой практики в цифровой среде, т.е. в качестве респондентов – членов исследовательских групп отбирались непосредственные практиканты в цифровой среде, методологи этой практики.

Шесть исследовательских групп представляли образцы для изучения человека, погруженного в цифровую среду, и возможности подобной исследовательской работы в высокотехнологичных практических областях и общественных науках. В каждой из исследовательских групп в течение восьми-десяти месяцев были проведены полевые исследования для создания детальной картины «исследовательской экологии» (research ecology – термин, относящийся к представленным в группах культурным и социальным ценностям, методам, теориям, концепциям человека и цифровой среды) каждой группы.

Человек как объект исследования (как «физический человек» – body) подвергся фрагментации для шести исследовательских групп. Он был разделен на разные физические зоны исследовательского внимания – на физического, цифрового и виртуального человека. Такая фрагментация способна пролить свет на процесс формирования концепции человека в дисциплинарном, онтологическом и методологическом контекстах. Размышление над тем, как фрагментируется «физический человек» в шести исследовательских группах, может быть полезной аналитикой, которая отвечает на вопросы: каким образом исследование «дробит» человека, почему и что из этого следует? Благодаря этому открываются более широкие методологические возможности, для того чтобы внести вклад в дискуссии вокруг человека как важного объекта исследования в эпоху постгуманизма, нового материализма и методологического постмодернизма (неограниченных методологических поисков). Человек этой эпохи может быть понят как «зонированный» внутри самых разных коммуникационных контекстов, которые вместе «складывают» человека в целостное социальное существо.

«Зонирование» человека было использовано в шести исследовательских группах как метод упорядочивания и ограничения «беспорядочной» и «безграничной» человеческой реальности. Например, в исследовательской группе образования «зоны» рассматривались как «строительные камни для возведения здания целостного понимания человеческой коммуникации в сфере обучения» [с. 45]. Подобным образом рассматривались «зоны» и в исследовательской группе моды, где каждая «зона» связывалась со «своим» знанием и технологией моды вместе с темой соответствующего обучения. Исследователи группы дизайна осуществляли подход к человеку, открывающий для них доступ к социальному и психологическому опыту людей, информацию о котором можно использовать для лучшей организации и поддержки дизайнерской практики. Участники этой исследова-

тельской группы часто высказывались о человеке – объекте своего исследования как о «сенсорном органе» (sensory organ) и «живом исследовательском инструменте» (living research tool), который они использовали для анализа экспериментальных данных. «Зонированный человек» в этой исследовательской группе состоял из чувственных зон – тактильной, обонятельной, зрительной и моторной. Участники исследовательской группы психологии фрагментировали «своего» человека на когнитивную / интеллектуальную и телесную зоны, определяя человеческий образ именно через интеллект. Напротив, в исследовательской группе акционизма человек трактовался не как рефлексирующее (когнитивное – интеллектуальное) существо, а как «телесная реализация образа, воплощающегося через телесную кинетику» [с. 45].

«Исследовательское фрагментирование человека, – пишут авторы в заключение, – полезный аналитический инструмент качественного исследования в цифровых контекстах, поскольку адресуется к сложной, эфемерной и беспорядочной человеческой природе, помогает “увидеть”, как человеческий опыт “картографируется в среде цифровых технологий”. А переключение исследовательского внимания между “фрагментированным человеком” и “целостным человеком” дало бы возможность прийти к новым идеям о человеческой / социальной реальности и новым методам ее познания» [с. 46–47].

*А.А. Али-заде*

**Чёрчленд П.С., Сейновский Т.Дж.  
СОЕДИНЯЯ ВЫЧИСЛИТЕЛЬНУЮ  
И ЭКСПЕРИМЕНТАЛЬНУЮ НЕЙРОНАУКУ  
(Реферат)**

Ref. ad op.: Churchland P.S., Sejnowski T.J.  
Blending computational and experimental neuroscience //  
Nature reviews neuroscience. – 2016. – Vol. 17, N 11. –  
P. 667–668. – DOI: 10.1038/nrn.2016.114. – Mode of access:  
<http://www.nature.com/nrn/journal/v17/n11/abs/nrn.2016.114.html>

Заслуженный профессор философии Калифорнийского университета в Сан-Диего П.С. Чёрчленд и нейрочученый, руководитель Лаборатории вычислительной нейробиологии Института биологических исследований Солка Т. Сейновский в небольшом комментарии рассуждают о теоретических основаниях современных масштабных (large-scale) проектов в нейронауке и вызовах, бросаемых их развитием.

Так, часто приходится слышать мнение о том, что нейронаука богата на эмпирические данные, однако бедна в теоретическом отношении, что она не располагает программой, способной организовать эти данные в единое целое. Тем не менее даже это утверждение не является вполне корректным: старожилам этого поля исследований, отмечают авторы, прекрасно известно, что на уровне сетей нейронаука бедна также и на эмпирические свидетельства. Данное обстоятельство, в свою очередь, сдерживает развитие теоретических инноваций: теории, направленные на установление связей поведения отдельных нейронов с высокоуровневыми системами, оказываются недостаточно определенными (under-constrained).

Запущенный в 2013 г. в США амбициозный проект The BRAIN Initiative («Исследование мозга через передовые инновационные ней-

ротехнологии» – Brain Research Through Advanced Innovative Neuro-technologies) – был направлен на поиск способа понимания работы сетей мозга на микро- и макроуровнях и влияния на них других сетей. Прогресс в решении этой проблемы в значительной степени зависит от развития новых технологий (позволяющих получать новые данные и обрабатывать их), и проект The BRAIN Initiative ориентирован на их разработку.

Фактически одновременно с запуском этого проекта появились инновационные методы исследования мозга (например, оптогенетика), позволяющие получить огромные массивы новых данных, обработка которых становится возможной благодаря все возрастающим вычислительным мощностям современных компьютеров.

В течение более чем 50 лет едва ли не главным двигателем прогресса в нейронауке оставался метод регистрации активности отдельных нейронов. И хотя данный метод был наилучшим из доступных в свое время, его, по замечанию авторов, можно уподобить попытке наблюдения за какой-либо сценой через соломинку, «исследуя один пиксель за единицу времени из произвольного места. Без понимания высокоуровневых отношений между пикселями (или нейронами), – пишут они, – трудно извлечь их более общие свойства и конфигурацию» [с. 667].

Однако сегодня, когда значительные технологические прорывы сделали рутинной единовременную регистрацию активности сотен нейронов, необходимо, подчеркивают авторы, задаться вопросом о том, какие концептуальные средства способны представить получаемые данные в терминах того, как сети мозга репрезентируют мир и принимают решения о дальнейших действиях.

Так, авторы отсылают к своей вышедшей в свет в 1992 г. и получившей немало откликов книге «Вычислительный мозг»<sup>1</sup>. В этой совместной работе, основываясь на принципах применявшихся в то время искусственных нейронных сетей (ИНС), они попытались выйти за пределы парадигмы одного нейрона к представлениям о репрезентации мира большими популяциями нейронов. Используя в те годы нейросетевые модели обучались при помощи изменений силы соединения между узлами сети в ответ на подаваемые на вход данные. Они могли решать нетривиальные вычислительные проблемы благодаря распределенным репрезентациям, находившимся в скрытом (или промежуточном) слое сети.

---

<sup>1</sup>Churchland P.S., Sejnowski T.J. The computational brain. – Cambridge, MA: The MIT press, 1992. – 560 p.

Однако пропорциональное увеличение размеров ранних ИНС было безнадежной задачей без значимого прогресса в увеличении производительности компьютерного оборудования. Современные сети индустриального масштаба состоят из выстроенных в десятки слоев миллионов искусственных нейронов и триллионов соединений между ними. Получившие не так давно небывалую известность далеко за пределами поля исследований искусственного интеллекта алгоритмы так называемого глубинного обучения (deep learning) позволили этим сетям впервые достичь невероятного успеха в решении задач распознавания объектов на изображениях или распознавания речи, с которыми ранее уверенно справлялся исключительно человеческий интеллект.

Так, например, недавно созданная ИНС AlphaGo смогла выиграть в го у корейского профессионала Ли Седоля со счетом 4:1, наглядно продемонстрировав силу метода «обучения с подкреплением в соединении с глубинным обучением» [с. 668]. Высокая продуктивность метода обучения с подкреплением для новых многоуровневых сетей идет вразрез с представлениями многих психологов (навеянными критикой бихевиоризма Н. Хомским) о его слабых возможностях в области моделирования высокоуровневых когнитивных способностей и превосходстве моделей, использующих правила логики.

Тем не менее даже современные мощные ИНС все еще не способны обладать многими фундаментальными функциями – мотивацией, желаниями, они не могут проявлять агрессию, любопытство и т.д. Чтобы продвинуться в понимании этих проблем, полагают авторы, для следующего поколения развивающихся в том числе в кооперации с проектом The BRAIN Initiative моделей будет важно постараться принять во внимание большее число систем мозга и способов их взаимодействия.

Серьезным вызовом для нейронауки сегодня является необходимость понимания того, как нервная система интегрирует текущие восприятия и уже имеющееся знание в контексте нужд и целей агентов. Прогрессирующее понимание того, как продвинутые ИНС справляются со сложными задачами, может, утверждают авторы, предоставить концептуальное основание для решения этой проблемы.

*М.А. Сущин*

**Милковский М.**

**МЕХАНИСТИЧЕСКАЯ КОНЦЕПЦИЯ  
ВЫЧИСЛИТЕЛЬНЫХ ОБЪЯСНЕНИЙ В КОГНИТИВНОЙ  
НАУКЕ И ВЫЧИСЛИТЕЛЬНОЙ НЕЙРОНАУКЕ  
(Реферат)**

Ref. ad op.: Miłkowski M. A mechanistic account of computational explanation in cognitive science and computational neuroscience // Computing and philosophy: Selected papers from IACAP 2014 / Ed. V.C. Müller. – Berlin: Springer, 2016. – P. 191–205. – Mode of access: <http://www.springer.com/br/book/9783319232904#>

Специалист в области философии когнитивной науки, научный сотрудник Института философии и социологии Польской академии наук М. Милковский предпринимает попытку представить целостную интерпретацию вычислительных объяснений в когнитивной науке и вычислительной нейронауке в рамках механистической объяснительной перспективы. Общеизвестно, что вычислительное моделирование является одним из краеугольных камней данных дисциплин. Тем не менее, несмотря на огромную значимость вычислительных объяснений, единого методологического представления о них до сих пор выработано не было.

Оказавшаяся в определенных отношениях продуктивной трехуровневая модель анализа сложных систем обработки информации Д. Марра также сталкивается с целым рядом теоретических затруднений. Например, отношения между постулированными в ее рамках уровнями могут пониматься по-разному, зависимость между способностью (competence) и ее физической реализацией также не ясна. Равно как и нет ответа на вопрос о том, следует ли понимать альтернативный этой модели восходящий (bottom-up) метод моделирования как всецело неверный, и т.д.

Как поясняет автор, с точки зрения сторонников применявшихся не так давно подходов в области теории механистического объяснения<sup>1</sup>, объяснение какого-либо феномена заключается в выявлении причинных оснований его функционирования, т.е. в создании модели причинной структуры механизма, лежащего в его основе. В общем виде механизмы понимаются как организованные системы, состоящие из релевантных в причинном отношении частей и активностей, вносящих, соответственно, вклад в набор способностей этих систем.

Развиваемый автором механистический подход к вычислительным объяснениям не предполагает каких-либо иных понятий (например, понятия «репрезентация»), кроме понятия «информация» (в информационно-теоретическом, а не в семантическом смысле). В своих построениях автор следует представлению об информации как о количественном структурно-информационном содержании, изложенному классиком науки Д.М. Маккеем: «Физический носитель, – пишет автор, – должен быть способен принимать по крайней мере два различных состояния, чтобы считаться несущим информацию» [цит. по: с. 193]. Соответственно, с точки зрения автора, вычисление и есть не что иное, как обработка информации.

Как и Д. Марр в рамках своей программы, автор подчеркивает необходимость четкого определения для механистических моделей того, что именно они собираются объяснить, т.е. четкого определения объясняемого феномена, или экспланандума (*explanandum phenomenon*). В то же время, замечает он, все охватываемые в рамках объяснения механизмы выделяются исследователями на основании их эпистемических интересов. Поэтому они могут быть рассмотрены под разными углами, в зависимости то того, что именно исследователь хочет объяснить.

Как пишет автор: «В некоторой степени спецификация экспланандума соответствует определению когнитивной компетенции (понимаемой в целом как способность механизма)» [цит. по: с. 194]. Однако же разительное отличие механистической объяснительной перспективы от традиционных вычислительных концепций когнитивных компетенций заключается в том, что в ее рамках

---

<sup>1</sup> Bechtel W. Mental mechanisms: Philosophical perspectives on cognitive neuroscience. – N.Y.: Routledge (Taylor & Francis Group), 2008. – 322 p.; Craver C. Explaining the brain: Mechanisms and the mosaic unity of neuroscience. – Oxford: Oxford univ. press, 2007. – 328 p.

экспланандум объясняется вместе с его физической реализацией находящимися в основании уровнями механизма. Для механистической точки зрения объяснение реализации представляет собой такую же релевантную задачу, как и объяснение компетенции (или способности). В общем виде экспланандумом для вычислительных объяснений служит «способность вычислять, или обрабатывать информацию» [с. 194].

В рамках исходных положений своей концепции автор проводит еще одно важное методологическое различие. Дело в том, что не все части используемых моделей (будь они чисто компьютерными моделями или же, например, воплощенными физически робототехническими моделями изучаемых животных) непосредственно в точности соответствуют своему экспланандуму. Например, робототехническая модель фонотаксиса у кузнечиков в силу технических соображений может включать в себя использование плат, которым, конечно же, ничего не соответствует в реальных кузнечиках. Поэтому, утверждает автор, необходимо различать части модели, прямо соответствующие объясняемому феномену (объяснительный фокус), и ее поддерживающие элементы (scaffolding), выполняющие вспомогательные функции.

В случае если объясняемый феномен полностью охватывается объяснительным фокусом модели, данная модель, в терминологии автора, будет «механистическим объяснением как-в-действительности (how-actual explanation); если модель включает в себя определенные темные пятна (black boxes), функция которых более или менее хорошо определена, она является схемой механизма; в противном случае она остается эскизом механизма» [с. 195].

Конститутивная форма механистического объяснения предполагает анализ феномена на трех уровнях: 1) на нижнем (-1) уровне описываются внутреннее устройство частей механизма и их взаимодействие; 2) на изолированном (0) уровне дается описание взаимодействия частей механизма (активностей и операций); 3) на контекстуальном (+1) уровне функции механизма рассматриваются в более широком контексте окружающей механизм среды [с. 195].

По отношению к объясняемому феномену модели могут быть либо слабо эквивалентными, если они описывают только подаваемую на вход информацию и результаты вычислений (т.е. выход), либо строго эквивалентными, если помимо входной и выходной информации также описывается сам внутренний процесс получения выходной информации. Использование обоих видов моделей практиковалось когнитивистскими подходами начиная с 1960-х годов, однако только

строго эквивалентные модели являются подлинно объяснительными в рамках механистической объяснительной перспективы.

Надлежащая механистическая модель вычисления должна состоять из двух компонентов. Во-первых, она должна включать в себя абстрактное описание вычисления, охватывающее все релевантные в причинном отношении переменные. Во-вторых, она должна содержать детальный план (blue print) механизма на всех уровнях организации. Так, чтобы починить сломанный ноутбук, недостаточно знать, что в формальном отношении он эквивалентен универсальной машине Тьюринга. Так же как и абстрактное описание когнитивной способности само по себе не говорит о том, как могут проявиться психические расстройства. Здесь, очевидно, потребовалось бы знать детали физической реализации вычислительных описаний. В вычислительном моделировании обе части модели нуждаются в сложной процедуре объединения, однако абстрактному описанию вычисления, по утверждению автора, может соответствовать только изолированный (0) уровень детального плана реализации: «Формальная модель вычислений (обычно) не включает в себя среду, в которой вычисления имеют место, так же как и лежащие в основе физические механизмы» [с. 197–198].

Для иллюстрации работы своей схемы автор использует заимствованный у Д. Марра известный пример кассового аппарата. В данном примере экспланандумом служит способность системы подсчитывать сумму к оплате. При этом на контекстуальном уровне рассматривается функция, которую кассовый аппарат выполняет в супермаркете, обеспечивая быстрый подсчет сумм и, соответственно, облегчая задачу продавцов и т.д. На изолированном уровне рассматриваются как программное обеспечение, позволяющее аппарату производить вычислительные операции, так и его аппаратное обеспечение («железо»), реализующее программный вычислительный код. Наконец, на нижнем уровне объясняются особенности операций непосредственно электронных частей машины.

Как утверждает автор, механистические объяснительные модели отличает акцент на анализе уровней в более тесном сопряжении, нежели в рамках программы Д. Марра. Кроме того, механистические объяснения допускают вопрос «Почему?» (который в схеме Марра относился к самому верхнему уровню) на каждом из уровней (например, почему дизайн аппарата удобен для пользователя и т.д.).

Автор также вкратце демонстрирует возможности применения механистической объяснительной перспективы для анализа современных нейровычислительных моделей действия, в частно-

сти модели действия в рамках программы предсказывающего кодирования<sup>1</sup>. Данный предсказывающий байесовский алгоритм предположительно реализуется системой зеркальных нейронов. Автор стремится показать, как механистический подход может быть использован для изучения предсказывающей модели действия на нескольких уровнях исследования, что помогает, по его мнению, оценить ее зрелость с точки зрения «завершенности и эмпирической адекватности» [с. 201].

Механистические методологические принципы своим возникновением обязаны исследованиям в науках о жизни, нейронауке и когнитивной науке. Механистическая объяснительная перспектива, заявляет автор, допускает исследовательский плюрализм, поскольку требует только того, чтобы модели объясняемых феноменов включали в себя «релевантные в причинном отношении факторы» [с. 203].

*М.А. Суцзин*

---

<sup>1</sup> Kilner J.M., Friston K.J., Frith C.D. Predictive coding: An account of the mirror-neuron system // Cognitive processing. – 2007. – Vol. 8, N 3. – P. 159–166.

**БОЛЬШИЕ ДАННЫЕ В ОРНИТОЛОГИИ:  
ПЕРСПЕКТИВЫ И ВЫЗОВЫ /  
ЛА СОРТЕ Ф.А., ЛЕПЧИК К.А., БАРНЕТТ Дж.Л.,  
ХУРЛБЕРТ А.Х., ТИНГЛИ М.В., ЦУКЕРБЕРГ Б.  
(Реферат)**

Ref. ad op.: Opportunities and challenges for Big Data ornithology /  
La Sorte F.A., Lepczyk C.A., Burnett J.L., Hurlbert A.H.,  
Tingley M.W., Zuckerberg B. // *The condor*. – 2018. – Vol. 120,  
N 2. – P. 414–426. – DOI: 10.1650/CONDOR-17-206.1

Последние достижения в области информационных технологий и сбора данных не только создали новые исследовательские возможности для использования больших данных (БД) в орнитологии, но и продемонстрировали новые вызовы и проблемы. Авторы реферируемой статьи, сотрудники ряда университетов США, дают обширный обзор предыдущего опыта и перспектив применения БД в орнитологии.

Для популяционной экологии птиц по-прежнему важны традиционные структурированные ресурсы, однако активно применяются и новые ресурсы – частично структурированные (например, eBird) и даже неструктурированные данные (например, «поточные» данные погодных радаров). Авторы пытаются выяснить, как используются БД в исследовании и моделировании разнообразия птиц и стратегий их миграции.

**Определение больших данных.** Термин «большие данные» используется для описания больших наборов цифровых данных, возникших в результате применения последних достижений в области информационных технологий и сбора данных<sup>1</sup>. БД отличаются от

---

<sup>1</sup> Tien J.M. Big data: Unleashing information // *Journal of systems science and systems engineering*. – 2013. – Vol. 22. – P. 127–151.

традиционных источников данных: 1) многочисленностью; 2) быстротой формирования и усвоения; 3) невозможностью организации в традиционных реляционных базах данных [с. 415]. Концепция БД вышла за рамки простого определения больших и сложных наборов данных – немалый интерес представляют культурные, технологические и научные последствия роста объема этих уникальных ресурсов.

**БД в естественных науках.** Как правило, БД делят на две категории: 1) физического или природного происхождения, полученные с помощью разнообразных датчиков; 2) полученные в результате социальной или экономической деятельности.

Примеры неструктурированных БД, которые относятся к природным явлениям, – это спутниковые снимки, материалы сейсмической съемки, астрономические изображения и т.д. Их поставщиками могут быть платформы дистанционного зондирования, сеть датчиков, встроенных в окружающую среду, и другие источники. Извлечение научных знаний об окружающем мире из массива БД – сложная задача. Для достижения надежных результатов требуется разработка новых аналитических подходов. Так, появились содержательные области пересечения биоинформатики и экоинформатики, позволяющие извлечь и применить такие типы данных в естественных науках.

**БД в орнитологии** – результат многолетнего развития систем сбора и вывода информации. На протяжении всей истории орнитологии натуралисты и орнитологи-профессионалы используют постоянно расширяющийся набор инструментов. Ружье уступило место биноклю, сетям, следовым полосам и геолокаторам. Кроме того, стабильные изотопы позволили по-новому взглянуть на диету птиц, а секвенирование ДНК помогло глубже понять структуры популяций и их эволюционную историю. Вначале к прекурсорам БД прибегали для достижения конкретных и специфических целей, но со временем их применение расширилось в связи с увеличением числа экологических проблем. Некоторые источники БД были задуманы именно такими с самого начала, например ресурс проекта науки граждан eBird. К еще одному источнику БД относятся потоки данных общего характера, которые, как оказалось, дают уникальные возможности для проведения орнитологических исследований. БД позволяют глубже понять географию распределения и экологию птиц.

**Структурированные данные** – это первые в истории наборы данных во всех естественных науках. Как правило, они характеризуются малым объемом, низкой скоростью поступления в базу и

малым разнообразием. В орнитологии имеются ограничения по сбору структурированных данных: любой процесс их сбора согласуется с определенным протоколом. Структурированные данные часто собираются в контексте заданного вопроса, и возможные отклонения от заданных параметров тщательно контролируются исследователями. В орнитологии их сбор приводит к наполнению баз данных разумного размера и сложности. Примерами могут служить базы, полученные в рамках программ North American Breeding Bird Survey (BBS), The British and French Breeding Bird Survey и Audubon's Christmas Bird Count (CBC). Эти «официальные» массивы данных можно определить как прекурсоры по отношению к новым слабоструктурированным и неструктурированным орнитологическим данным, поскольку они в течение нескольких десятилетий каждый год стабильно дают новые данные [с. 416].

**Слабоструктурированные данные** характеризуются огромным объемом массива, высокой скоростью сбора и большим разнообразием. Для них характерно отсутствие строгих стандартов и методологий сбора. Однако в отличие от неструктурированных эти данные имеют некоторую организацию, что значительно упрощает их анализ. Слабоструктурированные данные часто рассматривают как форму данных структурированных.

Спрос на такие массивы данных в орнитологии возник в результате многочисленных попыток увеличить гибкость традиционных массивов в связи с возникновением новых задач. Развитие слабоструктурированных массивов данных совпало с появлением программ науки граждан в орнитологии, которые привели к росту разнообразия массивов, подчиняющихся более категоричным протоколам.

Самым известным примером больших слабоструктурированных данных в орнитологии является массив eBird. Сайт eBird – это глобальный проект науки граждан по мониторингу. Он помогает добровольцам вносить данные своих наблюдений из любого места и в любое время. С использованием нескольких базовых протоколов данные наблюдений организуют в формате списка и вводят в онлайн-хранилище данных. Сайт eBird является полуструктурированным в том смысле, что наблюдатели должны придерживаться формата контрольного перечня при сборе данных, но при этом они свободны собирать данные с любым пространственным и временным разрешением. Таким образом, наблюдения могут колебаться по качеству и объему в широком диапазоне. До сих пор eBird удержи-

вает первенство по объему слабоструктурированного массива данных с непревзойденной скоростью обновления [с. 417].

**Неструктурированные данные** тоже характеризуются чрезвычайно большими массивами, скоростью поступления и разнообразием. Для их хранения не предусмотрена какая-либо структура или матрица. Их источником часто выступают пассивные датчики, которые постоянно поставляют аудио-, видео-, текстовые или изобразительные файлы. В некоторых случаях данные не предназначены для орнитологов, но могут быть ими использованы. Характерным примером является массив неструктурированных данных, поставляемых погодным радаром, которым научились пользоваться и орнитологи, в частности при мониторинге миграций птиц.

Извлечение орнитологической информации из выходных массивов погодных радаров – сложная задача. Одна из причин – наличие больших помех, связанных с атмосферной пылью, насекомыми и летучими мышами. Тем не менее с помощью методов машинного обучения и других технологий анализа БД изображения, выдаваемые погодными радаром, успешно применяются для получения высотных профилей плотности птичьих стай, распределения скорости и направления миграции.

Специализированным источником неструктурированных данных для орнитологов является неинвазивный акустический мониторинг. Птицы относятся к вокальным таксонам, поэтому была сформирована коллекция акустических образцов поведения птиц с использованием микрофонных антенн [с. 417].

**Вклад данных нового типа в орнитологию.** Гибкая природа полуструктурированных и неструктурированных данных позволила орнитологам решать новые задачи и проверять старые гипотезы. В отличие от традиционных ежегодных программ мониторинга, приуроченных к определенным периодам времени, например к периодам размножения, более гибкие слабоструктурированные или неструктурированные потоки дают возможность собрать больше разнообразных данных о популяции птиц.

Понимание закономерностей развития биоразнообразия на больших географических масштабах – одно из нескольких направлений исследований, которые могут быть развиты только с помощью массивов БД.

**Моделирование распределения видов.** Динамический характер распределения птиц требует применения подходов, учитывающих различные масштабы пространственно-временной вариации видов. Для птиц эта динамика зависит от конкретной фазы годового

цикла, например стационарного периода (гнездование) или более динамичного периода (миграция). Соответственно, подход, обеспеченный такими массивами данных, позволяет оценить распределение птиц даже без традиционного моделирования динамических орнитологических процессов.

Альтернативой моделированию распределения видов с применением только полуструктурированных данных является объединение структурированных и полуструктурированных данных с целью балансировки соотношения между количеством и качеством данных. Такие комбинированные подходы особенно важны при оценке распределений очень редких видов или видов, которые трудно обнаружить, потому что в таких случаях традиционные структурированные данные часто отсутствуют.

**Мониторинг миграции на больших масштабах.** С появлением БД возникли уникальные возможности для изучения миграции птиц на макроэкологических масштабах. Так, было выявлено, что многие мигрирующие популяции птиц придерживаются «петляющих» миграционных траекторий. Исследователи теперь могут формулировать вопросы, касающиеся устранения источников вариации в пределах этих широких общностей, таких как миграция расстояния или наличие экологических барьеров для миграции [с. 419].

**Миграция, размножение, фенология.** БД позволяют решать орнитологические проблемы, связанные с явлениями, происходящими в течение короткого периода времени и плохо поддающимися прогнозам. Так, было значительно улучшено понимание фенологии птиц, в частности были уточнены сроки кратковременных фаз миграции и размножения.

**Проблемы использования БД в орнитологии.** Большой объем и высокая размерность массивов БД способствовали разработке новых статистических и вычислительных методов, однако их применение вызвало новые проблемы с точки зрения анализа и интерпретации данных. Как правило, они связаны с накоплением шумов, ложными корреляциями и случайной эндогенностью. Общей чертой БД является их высокая размерность – наличие большого числа переменных, оцениваемых на большом количестве образцов. Применение традиционных параметрических подходов к этим массивам данных оказалось неэффективным. В таких обстоятельствах оценки коэффициентов могут дать существенные ошибки. К таким ошибкам приводят даже незначительные изменения в данных или в параметрах статистической модели. Методы, обеспечивающие надежные статистические выводы, начинают рабо-

тать, когда объем данных очень велик. Многие такие методы доступны или находятся в стадии разработки, но исследователи должны четко понимать существующие ограничения.

**Достоверность данных.** В связи с повышенной гибкостью и неявной структурой массива данных на первое место выходит проблема надежности данных. Для решения этой проблемы были разработаны научные программы повышения надежности БД в орнитологии, включающие сложные алгоритмы проверки данных. Так, в проектах Feeder Watch и eBird используется комбинация метода автоматизированных фильтров<sup>1</sup> и методов экспертной оценки<sup>2</sup>. Несмотря на эффективность применения новых подходов, вопросы доверия к данным, скорее всего, так и останутся проблемой для ученых, использующих ресурсы такого рода.

**Перспективы БД в орнитологии.** Сведения о птицах могут поступать из нетрадиционных источников, которые наполняются пользователями случайным образом, например поставляются краудсорсерами. Примером слабоструктурированных данных по вокализации птиц является база сайта Xeno-canto<sup>3</sup>, в которой содержится более 350 тыс. записей. Эти данные загружаются посетителями сайта. Эти записи предоставляют такие возможности для изучения фенологии птиц и их поведения, которые недоступны для традиционных, структурированных наборов данных вокализации. Еще более обещающим является потенциал вкладов неструктурированного орнитологического краудсорсинга, содержащихся в интернет-репозиториях фотографий или в обычных социальных сетях [с. 420].

**Мониторинг перелетных птиц.** Технологии мониторинга отдельных перелетных птиц совершенствовались на протяжении последних нескольких десятилетий. Более старые геолокаторы и более новые спутниковые передатчики предоставляют все более качественную подробную потоковую информацию о траекториях, скорости полета и текущих высотах отдельных птиц. Для динами-

---

<sup>1</sup> Bonter D.N., Cooper C.B. Data validation in citizen science: A case study from Project FeederWatch // *Frontiers in ecology and the environment*. – 2012. – Vol. 10. – P. 305–307.

<sup>2</sup> eBird: A human/computer learning network for biodiversity conservation and research: Proceedings of the twenty-sixth AAAI conference on Artificial Intelligence / Kelling S., Gerbracht J., Fink D., Lagoze C., Wong W.-K., Yu J., Damoulas T., Gomes C.P. – Toronto, ON, Canada: AAAI press, 2013. – P. 2229–2236.

<sup>3</sup> Xeno-canto: Голоса птиц по всему миру. – Режим доступа: [www.xeno-canto.org](http://www.xeno-canto.org)

ческого хранения таких массивов данных созданы весьма перспективные базы данных, например Movebank<sup>1</sup>.

**Автономные записывающие устройства.** Традиционная полевая орнитология опирается на прямой подсчет особей. Развитие и внедрение автономных записывающих устройств позволяет заменить традиционные методы на отбор образцов акустических сигналов сообществ птиц. Особенно эффективны автономные устройства непрерывной записи тогда, когда речь идет об исследованиях в удаленных местностях в отношении редких, проблемных с точки зрения существования видов.

**Экологические оценки в реальном времени.** Несколько ресурсов БД обеспечивают в реальном времени информацию об эволюционных и экологических последствиях быстрых изменений окружающей среды. Ключевую роль в этом процессе играют представители науки граждан<sup>2</sup>. Например, быстрые оценки появления птиц в том или ином количестве позволяют определить, не страдают ли популяции птиц от экстремальных погодных явлений или антропогенных стихийных бедствий.

В заключение авторы делают несколько важных выводов. Ресурсы БД и соответствующие расчетно-аналитические методы предоставляют исследователям уникальные возможности для описания различных свойств сложных природных систем. В некоторых случаях они могут дополнять выводы, полученные с помощью традиционных подходов. Однако применение БД имеет ряд проблем: большой объем и их размерность могут привести к накоплению шумов, ложных корреляций и случайных эндогенных факторов. Использование этих ресурсов должно сопровождаться четким пониманием присущих им ограничений. Так, в дополнение к разработке орнитологических приложений необходимо усовершенствовать аналитические методы для смягчения ограничений с учетом накопленного опыта.

*С.В. Егоров*

---

<sup>1</sup> Movebank. – Mode of access: <http://www.movebank.org/>

<sup>2</sup> Dickinson J.L. Citizen science as an ecological research tool: Challenges and benefits // Annual review of ecology, evolution, and systematics. – 2010. – Vol. 41. – P. 149–172.

# БОЛЬШИЕ ДАННЫЕ В СОЦИАЛЬНЫХ НАУКАХ

**Бласкес Д., Доменеч Х.**

## **ИСТОЧНИКИ И МЕТОДЫ ИЗУЧЕНИЯ БОЛЬШИХ ДАННЫХ ДЛЯ СОЦИАЛЬНОГО И ЭКОНОМИЧЕСКОГО АНАЛИЗА (Реферат)**

Ref. ad op.: Blazquez D., Domenech J. Big Data sources and methods for social and economic analyses // Technological forecasting & social change. – 2018. – Vol. 130. – P. 99–113. – Mode of access: <http://dx.doi.org/10.1016/j.techfore.2017.07.027>

Свою цель авторы, сотрудники Политехнического университета Валенсии (Испания), видят в том, чтобы наметить пути и способы, которые позволили бы использовать большие данные в социэкономической сфере.

Концепт «большие данные» появился в конце 1990-х годов и в начале 2000-х годов стал определяться как модель 3V: объем (volume), скорость (velocity) и вариативность (variety). Эта модель эволюционировала, приспособливаясь к новой цифровой реальности, и превратилась в модель 4V: был добавлен такой параметр, как ценность (value) – процесс извлечения ценной информации из набора данных, известный как аналитика больших данных. В настоящее время большие данные чаще определяют уже в терминах модели 5V, в которую добавлен параметр достоверности (veracity), означающий адекватное управление данными и соблюдение права на частную жизнь [с. 99].

Предполагается, что феномен больших данных внесет радикальные изменения в формирование социэкономической политики, управление и исследования в этой области. Выявление того, какие источники данных доступны, какой тип данных они поставляют и как с ними обращаться, служит основой, позволяющей

генерировать максимальное количество ценной информации для компании или организации. В этом контексте архитектура больших данных, адаптированная к специфической области и цели организации, вносит свой вклад в систематизацию процесса генерирования ценного знания. Эта архитектура должна быть способной управлять всем жизненным циклом данных в организации, включая их усвоение, анализ и хранение.

Для использования больших данных в социальной или экономической областях до сих пор не было предложено специфической архитектуры. Авторы статьи предлагают свой вариант архитектуры больших данных, который позволил бы прогнозировать текущие и будущие изменения в этой сфере.

Цифровые следы, которые оставляют люди, вызвали экспоненциальный рост количества источников данных (помимо традиционных опросов и официальных отчетов), доступных для социального и экономического анализа. Хотя причин создания этих новых данных множество, способы, с помощью которых они генерируются, имеют важные этические и юридические импликации. Например, персональные данные, касающиеся осуществляемых покупок, не могут использоваться для тех же целей, что и данные из профиля человека, представленного в Твиттере. В определенной степени использование данных ограничивается тем, как они создаются. Это заставило авторов рассмотреть и классифицировать нетрадиционные источники социальных и экономических данных в зависимости от цели пользователя, который их генерировал [рис. на с. 102].

Первый уровень таксономии включает пять категорий: 1) причина для поиска информации; 2) цель осуществления транзакции, которая может быть финансовой или не финансовой; 3) цель распространения информации; 4) цель осуществления социальной интеракции; 5) отсутствие определенной цели [с. 101]. Данные, которые попали в последнюю категорию, были разделены на три типа: используемые данные, данные о локации и персональные данные. Описание каждой категории и примеры источников, задействованных в каждом процессе генерирования данных, авторы представили в виде таблицы на с. 102.

Большинство нетрадиционных источников социальных и экономических данных предполагают использование Интернета. Огромное количество людей, компаний и организаций ищут, передают и генерируют ежедневно «тонны» информации. Эти цифровые следы, помещенные в рамки адекватной архитектуры больших данных, могут помочь в описании поведения, решений и интенций

тех, кто их оставил, и на основе этого можно провести мониторинг ключевых экономических и социальных изменений и трендов. Авторы приводят перечень и краткое описание этих нетрадиционных источников данных.

**Гугл-тренды: мощьность поисковых систем.** Гугл-тренды (ГТ) – это базирующейся на Интернете девайс, который впервые был использован в мае 2006 г. С его помощью можно получать сведения об объеме поисковых запросов начиная с января 2006 г. по специфическому ключевому слову или тексту. Таким образом можно узнать, как потребность в информации по определенной теме менялась со временем, что служит полезным источником данных для выявления возникающих тенденций и лежащих за ними интересов общества. С его помощью также можно получить информацию о поведении потребителей в разных секторах, о некоторых макроэкономических переменных (объеме продаж домов и автомобилей, въездном туризме и заявках безработных) и пр. [с. 101].

**Сайты и блоги в социальных сетях.** Социальные сети специально предназначены для того, чтобы пользователи имели возможность выразить свои чувства и мнения по любому вопросу. Поэтому информация, которую они содержат, в некоторой степени отражает то, что происходит в обществе. Однако разнообразие и сложность формата содержащейся в них информации приводят к тому, что они остаются малоизученными.

**Веб-сайты и приложения: транзакционные платформы и распространение информации.** В цифровую эру фирмы формируют свой официальный публичный образ в Интернете путем создания корпоративных веб-сайтов. Примечательно, что веб-сайты имеют сложную структуру, которая отличается в каждом случае, так что стандартизация, хранение и анализ их информации требуют специфичной архитектуры больших данных. Эта сложность стала причиной того, что корпоративные веб-сайты остаются практически неизученным источником данных.

Выдающаяся роль Интернета в современной экономике и жизни общества привела к появлению электронных бизнес-сервисов, которые могут использовать фирмы для продажи своих продуктов, для взаимодействия онлайн со своими заказчиками (э-коммерции), для подбора кандидатов на замещение вакантных должностей (э-рекрутинга) или предложения других услуг онлайн (например, э-банкинга). Многие из них предоставляют возможность высказать свое мнение по очень широкому кругу вопросов, от конкретных товаров до пребывания в отеле или посещения ресторана.

Мобильные приложения дают доступ к информации и услугам, которые могут или не могут быть предоставлены другими способами (например, на веб-сайтах). Поскольку использование приложений становится широко распространенным в повседневной жизни людей и организаций, они стали источником данных с большим потенциалом для прогнозирования социальных и экономических тенденций.

**Городские и мобильные сенсоры и датчики.** Успехи компьютерных технологий привели к созданию беспроводных, незаметных и недорогих сенсоров для сбора информации о повседневной жизни граждан. Среди городских сенсоров самый распространенный и широко используемый – это ридер кредитных карточек. Богатым источником информации о предпочтениях покупателей также служат сканеры, используемые в розничной торговле. Эти данные уже доказали свою полезность с точки зрения прогнозирования поведения потребителей, объема продаж и изменения цен.

Некоторые датчики, встроенные в мобильные телефоны, также служат потенциальным источником социальных данных: GSM, GPS, Bluetooth, акселерометр или сенсоры для соединения с телефонной сетью через Base transceiver stations. Данные с этих сенсоров полезны для идентификации мест, которые интересны людям, а также для определения личностных характеристик, которые компании могут использовать для персонализации своих услуг.

Данные, полученные из нетрадиционных социоэкономических источников, объемны, разнообразны и не структурированы или полуструктурированы, что порождает серьезные трудности, когда дело доходит до их извлечения, обработки, анализа и хранения. К настоящему времени разработано много методов и техник, связанных с машинным обучением и большими данными. Многие из этих методов широко применяются в других областях, таких как инженерные науки, медицина и биостатистика. Несмотря на их потенциал с точки зрения работы с социоэкономическими данными, их использование в данной области пока находится в зачаточном состоянии.

Авторы перечисляют и описывают методы, которые подошли бы для работы с большими социоэкономическими данными. Они подразделяют их на следующие группы: методы для структурирования данных; методы для моделирования данных; методы для оценки поведения и мощности модели [с. 104–106]. Таксономия нетрадиционных методов работы с социальными и экономическими данными представлена авторами в виде рисунка на с. 105.

Цифровые данные имеют множество преимуществ, прежде всего это простота их обмена, воспроизведения и рекомбинации, что позволяет использовать их многократно. Но для того чтобы можно было воспользоваться этим преимуществом, данные должны быть правильно собраны, переработаны и сохранены. Потеря данных или их повреждение могут дорого стоить. Поэтому важно определить, какие фазы и процессы составляют жизненный цикл данных, чтобы применить надежную и гибкую архитектуру для управления ими.

Жизненный цикл данных – это последовательность стадий, которые проходят данные от момента, когда они входят в систему, до момента, когда они из нее извлекаются или отправляются на хранение. Обзор работ по этой теме позволил авторам выделить и описать стадии полного жизненного цикла данных в контексте экономического и социального анализа. Всего они выделили девять таких стадий: изучение и планирование; сбор данных; документирование данных и обеспечение их качества; интеграция данных; приведение данных в нужный формат; анализ; публикация и шеринг; хранение; повторное использование данных [с. 108].

Отталкиваясь от описания жизненного цикла данных в организациях, авторы предлагают свой вариант архитектуры больших данных для прогнозирования текущих социальных и экономических изменений, который отображен на рисунке 4 [с. 109]. Эта архитектура организована в виде трех слоев. Слой анализа данных содержит основные процессы генерирования знания: от извлечения данных из множественных источников до их публикации или представления доклада. Он состоит из шести модулей, которые работают последовательно, от рецепции данных до публикации результатов (модули: получения данных; предварительной обработки данных; интеграции данных; приведения данных в нужный формат; анализа данных; публикации результатов) [с. 108–10].

Наряду с этим слоем существуют еще два слоя, функция которых состоит в поддержке анализа данных. Слой управления расположен горизонтально по отношению к остальной системе. В задачи управления входят приложение существующих норм и регулятивных правил ко всему жизненному циклу данных, а также решение вопросов лицензирования. Он состоит из пяти модулей, четыре из которых относятся к жизненному циклу данных, а еще один используется в целях аудита. В их число входят следующие модули: сбора данных; переработки; результатов; архивирования и извлечения данных; аудита [с. 111].

Персистентный слой обеспечивает другие слои, занимаясь всеми вопросами, связанными с потребностями хранения. Он охватывает не только сами данные, но и процедуры хранения, которые используются в разных модулях для получения доступа и преобразования данных.

В заключение авторы отмечают, что главный вклад их статьи состоит в предложении архитектуры больших данных, которая адаптирована к особенностям экономического и социального анализа и опирается на подход, исходящий из жизненного цикла данных в организациях [с. 111].

*Т.В. Виноградова*

**Райбс Д.**  
**STS ВСТРЕЧАЕТСЯ С НАУКОЙ О ДАННЫХ**  
**(Реферат)**

Ref. ad op.: Ribes D. STS, meet data science, once again //  
Science, technology & human values. – 2018. – In press. –  
DOI: 10.1177/0162243918798899

Исследования науки и технологии (STS) и недавно возникшая область – наука о данных, пишет автор – социолог, сотрудник Вашингтонского университета (США), имеют, как это ни удивительно, «избирательное сродство»<sup>1</sup>. На все расширяющемся пересечении этих областей представителей STS ждут не только большие возможности, но и серьезные трудности.

Посещая разного рода конференции и семинары, автор обратил внимание на то, как много специалистов в области STS обсуждают науку о данных (под разными названиями: большие данные, машинное обучение и пр.). В настоящее время ведутся серьезные дискуссии о том, что такое наука о данных и какие ученые работают в этой области. В качестве отправной точки автор взял определение, данное Ассоциацией компьютерных исследований США: «Мы используем термин “наука о данных” в самом широком смысле, включая сбор, обработку, аналитику и архитектуру данных». В настоящее время наука о данных переживает период бурного роста. «Она позиционирует себя в качестве *универсальной науки*, имеющей основополагающее значение как для всех буду-

---

<sup>1</sup> Термин «избирательное сродство» восходит к научному термину, обозначающему в химии XVIII в. способность химических веществ сочетаться с определенными веществами или их соединениями, отдавая им предпочтение перед другими. – *Прим. реф.*

щих наук, так и за их пределами, которая тесно связана с промышленностью и государственными структурами».

«Избирательное средство» описывает отношения, в которых некоторое количество акторов, разделяющих определенные аналогии и значения (религиозные, интеллектуальные или, как в данном случае, социотехнические), включаются в отношения реципрокного притяжения и влияния, взаимной селекции, активной конвергенции и взаимного усиления. Этот термин в социальный теоретический дискурс вошел благодаря Максус Веберу. Как известно, он доказывал, что определенная работа и основанная на общественном доверии этика протестантизма, способствовавшие первоначальному накоплению капитала, тем самым дали толчок развитию капитализма. «Избирательное средство стало важнейшей концепцией, с помощью которой М. Вебер связал идеи и интересы».

Это означает, что протестантизм и капитализм не обязательно должны были разделять одну и ту же логику, интересы или даже иметь одних и тех же акторов, чтобы взаимно поддерживать друг друга. Аналогичным образом, хотя STS и наука о данных, изучая науку, экспертизу и практики, идут разными путями и преследуют разные цели, они имеют избирательное средство, которое взаимно укрепляет каждую из сторон.

Корни избирательного средства STS и науки о данных достаточно многочисленны. Роли социальных ученых и гуманитариев разной специализации хорошо документированы в компьютерной области: например, участие антропологов и психоаналитиков в Масы Group<sup>1</sup>, сыгравшей важную роль в развитии кибернетики, или прямое участие таких областей социологии, как символический интеракционизм или этнометодология, в дизайне интерактивных систем, решении проблемы взаимодействия человек – компьютер. В своей статье, которая, как подчеркивает автор, носит программный характер и скорее ставит вопросы, чем на них отвечает, он прослеживает лишь два аспекта избирательного средства STS и науки о данных.

По словам автора, не так давно к нему обратились из ННФ с просьбой провести исследование «Big data regional hubs and spokes» (BDHubs), нового национального зонтика для американской науки о данных. Подчеркивалось, что это «уникальная возможность посмотреть, как развиваются структуры научного со-

---

<sup>1</sup> Масы Group – кибернетическая группа, спонсировавшаяся Фондом Джошуа Мейси. – *Прим. реф.*

трудничества. Четыре хаба получили финансирование с одной и той же целью, но каждый имеет свой характер, внутреннюю организацию и пр.». Более 10 лет назад научный руководитель автора Дж. Боукер (G. Bowker) провел аналогичную работу по изучению GEON (geosciences network) – «Cyber infrastructure for the earth science». Этот проект стал докторской диссертацией автора.

Почему сторонникам качественного подхода, социологам и специалистам в области STS, каковым является и сам автор, было предложено заняться КИ (киберинфраструктурой), а теперь – практиками науки о данных? Если коротко, то один из ответов состоит в том, что «социальное» является предметом озабоченности в КИ, а сегодня, еще в большей степени, – в кругах, занимающихся наукой о данных.

В данной статье автор трактует понятие «социальное» не как всегда присутствующую, фундаментальную черту интерактивного человеческого мира. Вслед за М. Фуко, Н. Роузом и К. Кнорр-Цетиной он рассматривает природу «социального» как результат определенного историко-институционального устройства. В широком смысле «социальное» есть продукт пограничной работы, которая в данном случае отделяет социальное от технического, а иногда и от научного. Эта граница – один из центральных предметов изучения в STS. Автора интересует, как различные политические парадигмы, прежде всего в области КИ и науки о данных, определяют полезность социального и иногда социальных ученых для успешного развития новых технологий и управления вызовами, связанными с организацией сотрудничества, междисциплинарностью, работой крупномасштабных коллективов и т.д.

Он обнаружил, что прагматические или «функциональные» вопросы сотрудничества, координации и социальной организации, которые стали главными в исследовании КИ, более или менее сохранили свою актуальность и в отношении современной науки о данных. Однако возникла и гораздо более широкая повестка дня, что автор иллюстрирует на примере Института больших данных, существующего в рамках его родного Вашингтонского университета.

Сегодня наука о данных «натолкнулась» на общество. Если проект по созданию КИ в основном был нацелен на естественные науки, наука о данных не видит этих границ. «Наука о данных» – это термин, который в Кремниевой долине появился раньше, чем в академических кругах, а сегодня он широко используется в государственных агентствах и неправительственных организациях. «В последнее десятилетие можно наблюдать, как проблема полу-

чения и обработки данных стала касаться практически каждого аспекта частной и общественной жизни. Это привело к актуализации значения социальных факторов в науке о данных, а соответственно, появилось относительно широкое поле деятельности для специалистов в STS».

На протяжении многих лет в ходе своих «полевых» исследований автор пользовался категорией «домен»<sup>1</sup> (domain) и понятием «ученый домена» (domain scientist). Когда автор начинал свои полевые исследования в 2003 г., термин «домен» постоянно использовался исследователями, которые совместно занимались разработкой КИ. Домен – это ключевое слово для науки о данных и смежных с ней областей.

Сегодня использование термина «домен» носит скорее договорной, чем формальный характер. Если упростить, домены относятся к тем областям (часто научным, но не исключительно), которые связаны с определенными и специфическими материями, например для лингвистов домен – это язык, для биологов – сфера органической жизни и т.д. Логика доменов делит мир на две главные категории – тех, кто «в домене», и тех, кто работает «независимо» от какого-либо домена.

Во время полевых исследований, проводившихся автором в 2000-е годы, категория «ученого домена» была вполне определенной. В рамках изучения GEON сказать «доменный ученый» было то же самое, что сказать «геоученый», а любой другой был компьютерщиком или технологом. Сегодня различие не столь однозначно. Многие «доменные ученые» идентифицируют себя как тех, кто занимается данными, а таковым может быть и геофизик. Но термин «домен» по-прежнему работает, определяя или конституируя область между и за пределами индивидуальных доменов как конечную цель науки о данных.

Существуют также не доменные задачи, модели экспертизы и технологии. Но эта позиция не имеет названия и не маркируется, часто определяется от обратного, например «независимая от домена». Еще пример – инструменты анализа данных, не зависящие от домена, могут оказаться полезными внутри геонауки и биологии. Можно ли сказать, что наука о данных, компьютерная наука или информатика также являются доменами со своей собственной экспертизой и практикой, задается вопросом автор. В принципе можно, считает он, но

---

<sup>1</sup> Домен (область) – исторически владение короля или какого-нибудь феодала в Средние века. – *Прим. редф.*

это выражение редко используется таким образом. В рамках принятого жаргона наука о данных, компьютерная наука и информатика – «другие» по отношению к «доменам».

Важно, что логика доменов не в состоянии сказать ничего специфического относительно определенного домена или того, как он может отличаться от других. Этот вопрос носит целиком эмпирический характер, ответ на него может быть получен путем изучения документов, данных или в процессе сотрудничества с членами данного домена.

Пространство между доменами и не доменами не остается пустым, его населяют так называемые «люди, имеющие форму П», т.е. стоящие на двух ногах. Они обладают экспертными знаниями как в области науки о данных, так и в доменной области (например, в геонауке). Противоположность им составляют так называемые «ученые, имеющие форму Т», т.е. стоящие на одной ноге, имеющие лишь одну специализацию.

Одна из целей науки о данных состоит в создании горизонтальных отношений между разными научными областями, а также с государством и промышленностью (все они обозначаются как домены). Она не пытается разрушить границы доменов, а скорее – облегчить и создать средства для налаживания взаимодействия поверх них. Пока автор не может описать логику и практику доменов, но предлагает три аналитические отправные точки.

1. Домены – это итог пограничной работы, которая проводится не ради разрушения этих границ (являются ли они техническими, культурными или практическими), а с целью упростить их пересечение.

2. Логика доменов отличается от логики междисциплинарности, хотя часто и действует в тандеме с последней. Например, конечная цель GEON состояла в разработке путей, которые позволили бы, например геофизикам, опираться на данные сейсмологов без изучения специфических категорий (незнакомых и эзотерических), с помощью которых генерируются и организуются сейсмологические данные. Междисциплинарный подход пытается смягчить или устранить границы (разделяя общий язык, методы, объекты или исследовательские проблемы); логика домена пытается сохранить различие, одновременно налаживая связи поверх границ.

3. Наконец, примечательная черта доменов в том, что они носят голографический или фрактальный характер: их свойства обнаруживаются везде, даже внутри них самих, и в любом масштабе.

Социальные ученые помогали в решении задач, которые стояли перед двумя другими категориями ученых, занимавшихся созданием КИ, – доменных ученых и компьютерных ученых. В этой триаде специалисты по компьютерным технологиям были разработчиками новых ресурсов; «доменные ученые» – теми, кто будет пользоваться этими ресурсами, а социальные ученые – теми, для кого этот социотехнологический процесс служит предметом изучения, а иногда и объектом обслуживания. Одна из ключевых исследовательских задач, утверждает автор, состоит в изучении и описании доменов.

Помимо прямого участия специалистов STS в проектах, связанных с развитием науки о данных, например в проведении опросов или «полевых» исследований или заимствовании концептов и методологии STS, их роль также важна при составлении программ обучения и учете так называемого человеческого фактора. «Изучение пользователей и взаимодействие с ними, людей, контекстов или социотехнологических систем доказало, что они поддаются исследовательским методам социальной науки, хотя и бросают ей серьезные вызовы».

*Т.В. Виноградова*

**КУЛЬТУРЫ ДАННЫХ, СОДЕРЖАЩИЕСЯ  
В МОБИЛЬНЫХ ПРИЛОЖЕНИЯХ ДЛЯ ЗНАКОМСТВ:  
ВОПРОСЫ ДЛЯ КРИТИЧЕСКОГО СОЦИОЛОГИЧЕСКОГО  
ИССЛЕДОВАНИЯ / АЛБАРИ К., БУРГЕСС Дж.,  
ЛАЙТ Б., РЭЙС К., УИЛКЕН Р.  
(Реферат)**

Ref. ad op.: Data cultures of mobile dating and hook-up apps:  
Emerging issues for critical social science research / Albury K.,  
Burgess J., Light B., Race K., Wilken R. // Big data & society. –  
2017. – Vol. 4, N 2. – P. 1–11. – Mode of access:  
<https://doi.org/10.1177/2053951717720950>

Социальные и этические последствия интеллектуального анализа данных, алгоритмической курации и автоматизации в контексте социальных медиа, по словам группы авторов из Австралии и Великобритании, стали очень важной темой для широкого круга специалистов. И особенно актуальным является вопрос о сохранении права на частную жизнь в этих новых условиях. Несмотря на широкое распространение и экономическую выгоду, мобильные приложения для знакомств еще не стали предметом серьезного анализа. Авторы намерены восполнить этот пробел.

Сайты знакомств и приложения, помогающие искать романтических и сексуальных партнеров, активно разрабатываются для мобильных девайсов. Сайты и приложения для знакомств содержат большое количество данных, которые служат посредниками, формируют и сами формируются под влиянием культур гендера и сексуальности. Это делает их особенно интересными с точки зрения изучения того, как персональные и социальные данные анализируются и используются корпорациями и как с ними поступают пользователи, т.е. в рамках разных и пересекающихся культур данных (навыков и манеры работы с данными) [с. 2].

Авторы используют термин «культура данных» в четырех разных смыслах. 1. Это понятие можно использовать для обозначения того, что можно назвать культурой производства приложений для знакомств, а именно институционализированных правил, привычек и практики их разработчиков. 2. «Культура данных» может относиться к различным способам обработки данных. Как известно, не существует «сырых данных», которые затем могут быть интеллектуально проанализированы. «Сырые данные» – это оксюморон. 3. Этот термин можно использовать для описания датафикации культуры посредством алгоритмической логики цифровых медиа, таких как мобильные приложения для знакомств, и их интеграции в более широкую логику социальных медиа, которые формируют общество. 4. Наконец, можно говорить о культуре использования данных, представленных в приложениях для знакомств [с. 2]. В своей статье авторы анализируют эти виды культуры данных, которые сложились в сфере мобильных приложений для знакомств.

Для романтических и сексуальных знакомств и в докомпьютерную эру использовались опосредованные, доступные на тот момент технологии. Появление компьютеров открыло новые широкие возможности и в этой области. В конце 1990-х годов появился такой веб-сайт, как GaydarMatch.com. Компании типа eHarmony стали использовать психологические алгоритмы для создания профилей пользователей на основе проведения опросов, повторяя тактику служб знакомств, которые они вытеснили. По понятным причинам сведения о месте жительства пользователей очень важны для таких онлайн-систем знакомств.

Наряду с большинством других видов использования Интернета и социальных медиа онлайн-знакомства мигрировали в мобильные устройства. В мобильных приложениях для знакомств появилось больше возможностей для использования GPS и других «пассивных» геолокационных технологий. Разнообразные типы и форматы данных пользователей мобильных приложений для знакомств собираются и интегрируются большим количеством корпоративных и индивидуальных акторов. В частности, как показывают авторы, эти данные используются разработчиками приложений или издателями, для того чтобы усовершенствовать пользование приложениями и, следовательно, увеличить прибыль. Поэтому вопрос об охране частной жизни, особенно когда включается обмен данными между платформами, становится все более актуальным. Учитывая озабоченность этим вопросом в обществе, авторы проанализировали политику в области охраны частной

жизни двух ведущих приложений Grindr и Tinder с точки зрения охраны частной жизни. Они установили, что ни одна из этих служб не снабжает пользователей детальной инструкцией относительно того, как добиться безопасности [с. 4].

Наиболее известный пример развернутого использования данных с сайтов знакомств – это работа, проделанная создателем самого известного сайта знакомств OKCupid К. Раддером<sup>1</sup>. Благодаря характерной для таких сайтов комбинации данных, внесенных самим пользователем, и сведений о его геолокации OKCupid может рассматриваться, считает К. Раддер, как «новая демография». Он характеризует этот подход как централизованный на человеке или даже как гуманитарный, противоположный подходу к использованию больших данных правительством и корпорациями.

Обсуждение больших данных по-прежнему предполагает односторонние отношения между корпоративными и институционализированными большими данными и индивидуальными пользователями. Но в случае мобильных приложений для знакомств некоторые пользователи сумели воздействовать на большие данные. Существует ряд примеров того, как пользователи «побеждали» OKCupid с помощью аналитики данных и даже путем создания побочных бизнесов типа TinderHacks [с. 5]. Эта субкультура присутствует в Сети, и даже выпущена электронная книга, которая была написана и издана за свой счет бывшим «обычным пользователем» К. Макинли<sup>2</sup>.

Беспокойство общества и технологическая неопределенность, связанные с мобильными приложениями для знакомств, привели к тому, что проблемой их использования тинейджерами заинтересовались законодательные органы и органы здравоохранения. Помимо этого возникла необходимость дать образовательные, правовые и политические ответы на проблему технологически опосредованного сексуального насилия, домогательства и шантажа. Ни в Австралии, ни в Британии на общенациональном уровне пока нет унифицированных правовых или образовательных ответов на эти вызовы.

Некоторые приложения (Grindr прежде всего) предприняли ряд публичных шагов к тому, чтобы взять на себя ответственность за безопасность пользователей, остальные – не всегда хотят отве-

---

<sup>1</sup> Rudder C. Dataclism: Who we are? When we think no-one looking. – N.Y., 2014. – 300 p.

<sup>2</sup> McKinlay C. Optimal Cupid: Mastering the hidden logic of OK Cupid. – [S.l]: Create space independent publishing platform, 2013. – 64 p.

чать за утечку данных или угрожающее поведение пользователей [с. 7]. Это объясняет, почему этими вопросами занялись активисты и сообщества пользователей. Например, Coding Rights Network, международный коллектив женщин разных профессий и национальностей, возглавляемый бразильским правоведем Джоанной Варон (J. Varon), издал специальное руководство<sup>1</sup>.

Широкое распространение приложений для знакомств, помимо прочего, поднимает вопросы, касающиеся здоровья и благополучия пользователей. В частности, администраторы заинтересовались тем, как настроить эти технологии для предотвращения распространения ВИЧ. Во многих странах по-прежнему сокрытие ВИЧ-статуса перед сексуальным партнером или использование нездоровых лекарств считаются преступлением. В США был проведен ряд совещаний на высшем уровне, задача которых состояла в поиске технологических решений проблемы здоровья нации. В частности, был поставлен вопрос о том, как заставить технологические компании осуществлять интервенции через приложения для знакомств, направленные на профилактику ВИЧ [с. 8]. В свою очередь, цифровые компании не часто хотят обсуждать детали законодательства и допускать секретные службы к базе данных их пользователей.

По словам авторов, в своей статье они впервые попытались описать разные варианты культуры данных мобильных приложений для знакомств. Свою работу они считают предварительной и полагают, что нужны дальнейшие серьезные исследования этой проблемы [с. 9].

*Т.В. Виноградова*

---

<sup>1</sup> Felizi N., Varon J. Safer nudes: A sexy guide to digital security. – 2015. – Mode of access: [http://www.codingrights.org/wp-content/uploads/2015/11/zine\\_ingles\\_lado2.pdf](http://www.codingrights.org/wp-content/uploads/2015/11/zine_ingles_lado2.pdf) (Accessed: 4 December 2015)

**ПОЛНАЯ КАРТИНА ВЕЧЕРИНКИ:  
НАУЧНЫЙ ЭКСПЕРИМЕНТ  
ПО ВЗАИМОДОПОЛНЕНИЮ СОЦИАЛЬНЫХ ДАННЫХ /  
БЛОК А., КАРЛСЕН Х.Б., ЙЁРГЕНСЕН Т.Б.,  
МАДСЕН М.М., РАЛУНД С., ПЕДЕРСЕН М.А.  
(Реферат)**

Ref. ad op.: *Stitching together the heterogeneous party: A complementary social data science experiment* / Blok A., Carlsen H.B., Jørgensen T.B., Madsen M.M., Ralund S., Pedersen M.A. // *Big data & society*. – 2017. – Vol. 4, N 2. – P. 1–15. – Mode of access: <https://doi.org/10.1177/2053951717736337>

Эра исследования больших данных (БД) и цифровой социальной науки, по словам авторов, сотрудников Копенгагенского университета (Дания), породила ряд изменений внутри и вне социальных наук. Одно из них касается того, что границы между ранее самостоятельными форматами данных (цифровыми, численными, этнографическими, визуальными и пр.) стали размываться и эти форматы начали иным образом взаимодействовать друг с другом.

Большие социальные данные и этнография, как доказывает ряд исследователей, имеют схожие черты. Их интересуют повседневная жизнь, паттерны социальности, интеракции людей в пространстве, а также целостный подход к анализу современного мира. В своей статье авторы приводят пример сотрудничества шести этнографов и социологов (авторов этой статьи), которые работали вместе в течение недели, проводя эксперимент по объединению форматов «больших» транзакционных данных и «малых» этнографических наблюдений.

Объектом изучения стала вечеринка для первокурсников, которая проходила в Датском техническом университете. В ходе этого эксперимента была предпринята попытка объединить «большие»

цифровые данные, представленные в форме записей, и «малые» этнографические данные в виде результатов включенного наблюдения. «Задача состояла в том, чтобы исследовать шаги, которые нужны для того, чтобы сшить гетерогенные миры данных этнографии и компьютерной науки новым, комплементарным и неиерархическим способом» [с. 2]. Решая эту задачу, авторы обратились к анализу «отливов» и «приливов» в социальной интенсивности в атмосфере вечеринки, сопоставляя этнографические полевые заметки, которые вела одна из авторов, и сигналы Bluetooth. Результаты этого анализа оказались вполне успешными.

Используя методы этномайнинга и качественно-количественный подход Б. Латура, авторы показали, как на практике этнографические и цифровые транзакционные данные могут быть объединены и взаимно дополнять друг друга. Их подход «одновременно подтвердил и расширил уже существующие попытки практического изучения интерфейса больших данных и этнографии, добавляя к ним реальное описание и методологический язык, который может быть использован при работе с этими двумя мирами данных, продемонстрировав возможности их одновременного и взаимодополняющего использования» [с. 12].

*Т.В. Виноградова*

# **БОЛЬШИЕ ДАННЫЕ В ЭКОНОМИКЕ И ОРГАНИЗАЦИИ НАУКИ**

**Садовский Ж.**

## **ДАнные КАК КАПИТАЛ: ДАТАФИКАЦИЯ, АККУМУЛИРОВАНИЕ И ИЗВЛЕЧЕНИЕ (Реферат)**

Ref. ad op.: Sadowski J. When data is capital:  
Datafication, accumulation, and extraction // Big data & society. –  
2018. – Vol. 6, N 1. – P. 1–12. – DOI: 10.1177/2053951718820549

Автор статьи, сотрудник Сиднейского университета (Австралия), анализирует проблемы, связанные с большими данными, в частности сбор и обмен данными как новый центральный компонент в системе функционирования современного капитализма.

В современной системе производства, в отраслях инфраструктуры, энергетики, финансов, страхования в настоящее время базы данных рассматриваются как одна из форм капитала, хотя совсем недавно они воспринимались исключительно как объект, интересный для научных исследований, и как побочный продукт различных производственных процессов. До самого последнего времени компании просто с легкостью удаляли накопившиеся данные, теперь они тщательно их собирают и накапливают по мере возможности.

Формирование и использование огромных баз данных – больших данных (БД) – все больше становятся объектом целенаправленного инвестирования средств и, следовательно, подчиняются общим принципам ориентации на получение прибыли. БД теперь – необходимый компонент при принятии важнейших решений как на уровне крупных компаний (например, приобретение компании Whole Foods компанией Amazon), так и на уровне государственного управления (например, решение городских властей об инвестировании в создание беспроводных сетей). В 2017 г. в журнале The Economist отмечалось, что «...индустриальные ги-

ганты, подобные “Дженерал электрик” и “Симменс”, преподносят себя на рынке как компании, обладающие солидными базами данных»<sup>1</sup>. Таким образом, делает вывод автор статьи, формирование и использование БД становится ключевым компонентом политической экономики XXI в.

Парадигма анализа БД как капитала охватывает самые разные сферы и аспекты современной экономической реальности. Стремительное развитие цифровой экономики и расширение спектра специфических интернет-продуктов и услуг приводят к тому, что использование БД становится необходимым компонентом и базовой формой капитала для создания различных современных систем, включая «умные дома» и «умные городские пространства», электронные системы управления государственными финансами, производством и распределением продукции, маркетингом и многими другими процессами и т.д., которые просто невозможны сегодня без применения масштабных баз данных. Современная литература, как правило, трактует БД всего лишь как один из видов новых товаров; автор реферируемой статьи ставит задачу проанализировать БД как форму капитала. По его мнению, это будет способствовать развитию и углублению анализа современной системы «цифрового капитализма».

В научной литературе на сегодняшний день сформировалось немалое количество терминов, концептуализирующих взаимоотношения, сложившиеся между БД и современным капитализмом: «надзирающий капитализм», «информационный капитализм», «коммуникативный капитализм» и др. Однако за их множеством, по мнению автора, скрывается одна и та же суть, которая может быть выражена в виде трех следующих тезисов: 1) БД обладают стоимостью и способны создавать новую стоимость; 2) БД оказывают мощное и всеобъемлющее влияние на процессы развития современного бизнеса и государственного управления; 3) распространение БД часто связано с такими явлениями, как неравенство, несправедливость и эксплуатация [с. 2].

Большинство современных компаний и организаций все в большей степени подчиняют свою деятельность так называемому императиву данных, который требует собирать данные из всех возможных источников всеми возможными средствами [с. 2]. Это

---

<sup>1</sup> The world’s most valuable resource is no longer oil, but data // The Economist. – 2017. – 6 May. – Mode of access: <https://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rulesworlds-most-valuable-resource> (Accessed: 9 October 2017.)

ведет к тотальной «датафикации»: наблюдению и сбору данных о происходящих процессах, поведении людей и взаимосвязях, возникающих между ними. Практически любая активность в Интернете в настоящее время – это основа для формирования баз данных, которые не просто являются информацией, отражающей реальный мир, а используются для того, чтобы воздействовать на него, изменять его в интересах тех, кто владеет этой информацией. Таким образом, информация, содержащаяся в базах данных, дает ее обладателям определенную власть, позволяющую воздействовать на ситуацию. Затем эта же власть помогает им аккумулировать новые объемы информации, на основе которой они отслеживают характер протекающих изменений и в значительной мере определяют их направленность. Поэтому, по мнению автора, неправильно говорить только о собирании или извлечении данных, правильнее говорить об их одновременном *производстве*, поскольку данные возникают в результате сложного взаимодействия одних субъектов информационной среды с другими. Сбор все новых и новых данных делает информационную среду все более контролируемой и управляемой, что, в свою очередь, создает возможность для сбора еще большего количества данных. Этот поток новых данных непосредственно и связан с увеличением рыночной власти и прибыли.

По мнению автора, чтобы рассматривать БД в качестве капитала, необходимо прежде всего определить их соответствие традиционным критериям, согласно которым в политической экономике нечто определяется как капитал. Для этой цели он приводит две трактовки понятия «капитал» – Карла Маркса (классическую) и Пьера Бурдьё.

К. Маркс описывает процесс формирования капитала, начиная с анализа взаимоотношений между товаром и деньгами в процессе обращения, в конечном итоге выделяя добавочную стоимость как конечную цель и движущую силу всего процесса. Согласно его определению, капитал существует только в процессе непрерывного обращения, на каждом витке которого увеличивается его стоимость. Это увеличение происходит в результате производственного процесса, в котором участвуют две формы реального капитала – постоянный и переменный. Переменный капитал, авансируемый на оплату рабочей силы, является источником прибавочной стоимости, которая формирует прибыль и в конечном счете делает капитал капиталом.

П. Бурдьё выделил другие универсальные формы капитала:  
1) капитал, который может быть быстро конвертирован в денежную

форму и который может быть объектом четко специфицированных прав собственности; 2) культурный и социальный капитал, который может конвертироваться в экономический только при наличии определенных специфических условий, поскольку он облегчает своему обладателю взаимодействие с соответствующей социальной средой<sup>1</sup>.

Автор статьи считает понятие культурного и социального капитала, разработанное П. Бурдьё, более подходящим для анализа БД в качестве капитала. БД – это не просто информация о мире: подобно культурному и социальному капиталу, они упрощают взаимодействие с информационной средой, что создает условия для конвертации капитала в экономические блага, таким образом формируя чистый доход. Этот доход не всегда имеет денежную форму, а может выражаться в создании каких-либо иных преимуществ, увеличивающих конечную эффективность экономической деятельности. Однако то, что БД могут непрерывно аккумулироваться и обновляться, делает их похожими на капитал в трактовке К. Маркса, которая предполагает бесконечный и безграничный процесс обращения и накопления капитала. По мнению же автора статьи, в условиях цифрового капитализма БД подобны скорее не реальному капиталу, который приносит прибыль, непосредственно преобразуясь в товарную форму, а финансовому капиталу, который приносит своему владельцу прибыль без непосредственного соприкосновения с реальными производственными процессами. Поэтому он создает иллюзию самовозрастания капитала в чистом виде. Так же и данные имеют тенденцию возрастать в процессе своего использования и таким образом увеличивать свою ценность. Во всяком случае, подобно капиталу в трактовке К. Маркса, который рассматривается как инструмент непрерывного увеличения прибыли, БД никогда не трактуются как некий конечный результат, а только как средство получения все новых и новых данных. Крупные компании часто создают новые продукты не столько для получения прямой прибыли, сколько с целью накопления данных, которые затем капитализируются при продвижении и реализации каких-либо других продуктов. Таким образом, сбор данных часто становится важнее конкретного краткосрочного финансового результата, и даже наличие риска того, что условия для капитализации собираемых в данный момент данных могут никогда не наступить, не останавливает процесс сбора данных.

---

<sup>1</sup> См.: Bourdieu P. The forms of capital // Handbook of theory and research for the sociology of education / Richardson J. (ed.). – Westport, CT: Greenwood, 1986. – P. 241–258.

По мнению автора статьи, в современной экономике процесс сбора данных является таким же значимым, как и процесс движения товаров и капиталов. Либеральные экономисты и политики сегодня выступают за свободу сбора и обмена данными точно так же, как за свободу движения финансовых потоков и международной торговли. Более того, все эти процессы рассматриваются как взаимосвязанные. Карл Бильд, возглавляющий Глобальную комиссию по управлению Интернетом (Global Commission on Internet Governance – GCIG), в частности, заявил: «Барьеры против свободного движения данных в конечном итоге оборачиваются барьерами против торговли» [цит. по: с. 5]. По его мнению, любое ограничение международного обмена данными создает препятствия на пути распространения инноваций, экономического развития и экономического роста. Результатом такого подхода должно стать формирование мирового рынка баз данных, что предполагает их распространение за пределы национальных границ, так что данные об американских потребителях могут накапливаться, например, на Тайване, а затем быть проданы в Европу и наоборот.

БД, как и любой другой вид капитала, могут иметь самое разное применение и приносить прибыль самыми разными способами. Многообразные отрасли и сферы производственной и экономической деятельности используют БД для достижения самых разных целей. Внимание исследователей, как правило, фокусируется на так называемых пользовательских базах данных, которые содержат сведения о физической активности и состоянии здоровья больших групп населения, геолокации и т.д., но существует множество других применимых видов данных, например данные об использовании технических и транспортных средств и др. Существует несколько основных типов данных и способов их использования, позволяющих получать прибыль.

• **Данные для использования профилей и целей пользователей.** Множество бизнес-моделей базируются на предположении о том, что знания о потребностях потенциальных клиентов могут существенно увеличить прибыль и / или рыночную власть. Например, реклама компаний в Интернете будет более эффективной в том случае, если клиентам предоставлять персонализированные рекламные предложения. Посредники, специализирующиеся на сборе данных, аккумулируют информацию о потребителях в конкретных сегментах рынка. Бюро кредитных историй, например, собирают данные, позволяющие минимизировать риски при выдаче кредитов. Риелторы могут варьировать ценовые предложения

объектов недвижимости, ориентируясь на индивидуальные характеристики потенциальных потребителей. Политики формируют свои послания к потенциальным избирателям, исходя из имеющейся у них информации о политических предпочтениях избирателей и наиболее актуальных для них социальных и экономических проблемах, что повышает эффективность их воздействия на электорат.

- **Данные, используемые для оптимизации систем.** Эффективность многих процессов может быть повышена с помощью данных, позволяющих сократить потери времени, повысить производительность и снизить ресурсоемкость. В промышленном производстве это может достигаться за счет установления специального сенсорного оборудования для постоянного мониторинга производственных процессов. По сути, это развитие, на новой технологической базе, системы Тейлора, зародившейся еще в начале XX в.

- **Данные, используемые для управления и контроля за различными объектами.** Эти данные формируются исходя из представлений о том, что чем больше знаний о функционировании тех или иных объектов, тем больше возможностей эффективно управлять ими. Так, например, систематическое наблюдение и сбор данных о функционировании системы дорожного движения позволяют более эффективно управлять процессами перемещения людей и транспорта в городском пространстве. Такого рода базы данных помогают повысить эффективность соответствующих систем или дают необходимую информацию для функционирования автоматических систем, регулирующих различные ситуации в режиме реального времени.

- **Данные, используемые для вероятностного моделирования.** При наличии большого количества данных, охватывающих продолжительный временной период, многие компании осуществляют разработку алгоритмов, позволяющих с достаточной степенью уверенности предсказывать будущие события. Разумеется, эти предсказания носят вероятностный характер, однако рынок подобных прогнозных инструментов в самых разных сферах стремительно расширяется. Использование таких инструментов затрагивает деятельность не только частных компаний, но и государственных структур. В частности, органы правопорядка формируют специальный список «горячих мест», где наиболее вероятным является проявление какой-либо криминальной активности, и берут эти территории под особый контроль. В системах, определяющих вероятность криминальных проявлений, используются данные о раз-

ных факторах, таких как плотность населения, его половозрастной состав, социальный и имущественный статус, расположение общественных мест (баров, церквей, остановок общественного транспорта и др.) и т.п.

• **Данные, используемые для оказания услуг.** Значительная часть современных цифровых систем, предоставляющих различные услуги, не может функционировать без обширной и актуальной, т.е. постоянно обновляемой в режиме реального времени, базы данных. Например, знаменитая система Uber не может работать без постоянно обновляющейся базы данных о водителях и пассажирах. Модернизация значительного количества коммерческих и социальных услуг – результат использования значительного объема данных. Дальнейшее развитие технологий потребует еще большего объема и разнообразия данных о среде, в которой эти технологии будут использоваться.

• **Данные, применяемые для увеличения стоимости активов.** Материальные активы, такие как здания, инфраструктурные сооружения, машины и механизмы, имеют тенденцию обесцениваться, что является прямым следствием их физического износа. В финансовом плане он выражается в виде соответствующих амортизационных отчислений. Современные средства контроля за физическим состоянием объектов позволяют эффективно контролировать процессы физического износа, максимально замедлять его и нивелировать негативные последствия. В результате процесс обесценивания материальных активов может быть если не полностью остановлен, то, по крайней мере, значительно замедлен, что способствует поддержанию высокой рыночной стоимости этих активов.

Большая часть баз содержит данные об индивидуальном потребительском поведении, собранные в результате отслеживания активности пользователей в Интернете с применением «умных устройств». Интеграция соответствующих технологий в повседневную жизнь потенциальных потребителей позволяет собирать огромный объем информации об их потребительском поведении и предпочтениях и применять его для формирования персонализированной рекламы и предложения товаров и услуг. Однако автор статьи видит в таких способах извлечения необходимых данных элемент произвольного присвоения. Если данные представляют собой капитал, то очевидно, что он формируется как побочный продукт повседневной потребительской активности и присваивается безвозмездно. Так же как наемные работники, для того чтобы удовлетворить свои жизненно важные потребности, должны участвовать в создании приба-

вочной стоимости и накоплении капитала, современные потребители столь же безвозмездно должны участвовать в формировании баз данных. Как и в случае с наймом рабочей силы, который оформляется трудовым договором, изъятие данных в большинстве случаев оформляется пользовательскими соглашениями. Однако согласие пользователей с условиями этих соглашений зачастую является формальным и неосознанным выбором, и использование персональной информации пользователю никак не компенсируется. Вопрос о соответствующей компенсации является весьма сложным, поскольку стоимость персональной информации не может быть адекватно оценена. К тому же различные типы данных в разной степени востребованы и по-разному оцениваются в различных сферах бизнеса. Поэтому для индивидов, чья персональная информация является источником формирования баз данных, единственной компенсацией остается получение доступа к различным интернет-ресурсам: социальным сетям, обучающим и развлекательным ресурсам и другим онлайн-платформам, а также возможность использования различных высокотехнологичных продуктов.

В то же время объем и стоимость баз данных непрерывно возрастают. Богатейшие современные мировые компании типа Facebook и Google – главные производители обширных и дорогостоящих баз данных. Годовой оборот торговли базами данных в современной мировой экономике достигает 200 млрд долл.<sup>1</sup> Кроме того, значительная часть финансового сектора, сферы страхования, промышленного производства и других отраслей и сфер экономической деятельности активно используют БД и извлекают значительную часть своей прибыли. Поэтому, утверждает автор статьи, в современной цифровой экономике БД становятся ведущим видом капитала, и его анализ является приоритетной задачей современной экономической науки. Решение этой задачи открывает новый этап в развитии экономических исследований, соответствующих новому этапу развития капитализма и новым методам извлечения прибыли и накопления капитала. Таким образом, заключает автор, процесс «датафикации» – это новый этап развития современного капитализма.

*М.О. Лихачев*

---

<sup>1</sup> Crain M. The limits of transparency: Data brokers and commodification // New media & society. – 2018. – Vol. 20, N 1. – P. 88–104.

**Кёниг П.Д.**  
**МЕСТО ОБУСЛОВЛЕННОСТИ**  
**И ИНДИВИДУАЛЬНОЙ ОТВЕТСТВЕННОСТИ**  
**В «ЭКОНОМИКЕ БОЛЬШИХ ДАННЫХ»**  
**(Реферат)**

Ref. ad op.: König P.D. The place of conditionality and individual responsibility in a «data-driven economy» // Big data & society. – 2017. – Vol. 4, N 2. – P 1–14. – DOI: 10.1177/2053951717742419

Автор статьи, сотрудник департамента политических наук Университета им. И.В. Гёте (Франкфурт, Германия), рассматривает экономические и политические проблемы, возникающие при использовании больших данных (БД) в современной цифровой экономике. В качестве центральных автор выбрал два понятия: понятие обусловленности (conditionality), которая понимается как создание официальных прав на защиту и получение выгод в случае соблюдения определенных норм поведения, и понятие индивидуальной ответственности.

Принцип обусловленности играет огромную роль в процессах реформирования государства благосостояния, происходящих в настоящее время в таких странах, как Швеция, Нидерланды, Германия и Великобритания. Эти процессы, как правило, сопровождаются соответствующим изменением политической риторики, в которой центральное место занимает принцип индивидуальной ответственности. В основе этого принципа лежит представление о справедливости и заслуженности доступа к социальным благам. Это представление восходит еще к началу Нового времени, когда в английском законодательстве о бедных впервые были сформулированы понятия «достойной» и «недостойной» бедности. Согласно этому представлению, на помощь со стороны общества имеют

право только те бедные, кто стремится к труду либо не способен работать в силу своего физического состояния. Такой подход имеет две стороны: 1) он устанавливает определенную систему отношений между гражданином и государством, определяя права и обязанности каждой стороны; 2) он в значительной мере способствует переносу социальной ответственности с государства и общества на самого индивида, делая его ответственным за собственное благосостояние.

В рамках реформирования современного государства благосостояния такой подход используется для ограничения социальных расходов и ослабления системы социальной защиты со стороны государства. Благодаря этому государство имеет возможность сократить число получателей социальной помощи, предоставляя ее только тем, кто с точки зрения принятых нормативов ее заслуживает. Эта тенденция отчетливо просматривается в современной социальной политике таких стран, как Великобритания, США, Швеция, Дания, Нидерланды и Германия.

С развитием информационных технологий, создающих новые способы коммуникации и вызывающих развитие децентрализованной социальной активности, неизбежно возникает потребность в усовершенствовании методов управления социальными процессами. Это в равной мере касается проблем социальной политики, обусловленности социальной помощи и индивидуальной ответственности. Решение вопроса о том, заслуживает ли некто социальной помощи, на основе разработанных критериев обусловленности требует прежде всего сбора и анализа обширной информации об индивидуальном поведении.

В цифровой экономике БД становятся ценным объектом собственности и различных экономических операций. В эту категорию входят и так называемые личные данные. Современные технологии позволяют осуществлять сравнительно дешевый и практически незаметный сбор данных о самых разных аспектах индивидуальной деятельности. В сочетании с технологиями, позволяющими быстро извлечь необходимую информацию из гигантских БД, значимость этого ресурса с точки зрения возможностей его коммерциализации в будущем сопоставима с той значимостью, которую имела нефть в начале XX в. Сегодня уже предпринимаются попытки определить ценность персональных БД и сделать коммерциализацию этого ресурса основой для развития новейших бизнес-структур. Базы персональных данных все больше превращаются в исходный материал, используемый для самых разнообразных целей – чисто аналитических и

практических (оптимизация бизнес-планирования и бизнес-процессов, реализация программ персонализированного маркетинга и т.д.). В результате в современной экономике складывается ситуация, в которой процесс создания ценностей в значительной мере обусловлен наличием БД об индивидуальном поведении потенциальных потребителей. Согласно данным Европейской парламентской исследовательской службы, бизнес, основанный на БД, будет давать около 1,9% ВВП Евросоюза в 2014–2020 гг. [с. 5].

Многие правительства развитых государств рассматривают современный бизнес в сфере БД как весьма перспективный сектор национальной экономики, который в будущем будет активно развиваться. Так, правительство Великобритании в своих официальных документах именуется эту сферу деятельности важнейшим сектором и будущей «основой национальной экономики», которая определит перспективы ее развития и обеспечит процветание [цит. по: с. 5]. В целом «экономику БД» можно определить как новый, динамично растущий и развивающийся сектор национальных экономик развитых стран, который эффективно дополняет сформированную ими экономику знаний. Однако «экономика БД» в своем развитии существенно зависит от степени доступности индивидуальных БД. Поэтому правительство должно создать соответствующие правовые и институциональные условия, для того чтобы эта сфера могла развиваться в интересах всего общества.

Развитие «экономики БД» сталкивается с существенными затруднениями. Процесс создания стоимости требует беспрецедентной детализации данных об индивидуальном поведении, которая позволяет извлекать из них информацию о самых разных аспектах человеческой деятельности. Для многих компаний это становится источником конкурентных преимуществ. Информация об индивидуальном поведении позволяет бизнесу делать потребителям персонализированные предложения, направленные на удовлетворение их специфических потребностей. В то же время такая информация дает коммерческим организациям беспрецедентную рыночную власть. В результате потенциальные потребители сами превращаются в ресурс или товар, они могут утратить свой суверенитет, превратившись в объект манипуляций. Потребитель будет получать вознаграждение за поведение, которое желательно для тех, кто использует персональные данные о его поведении. Следовательно, проблема использования БД выходит за рамки чистой экономики. Она создает целый комплекс моральных, социальных и политических проблем, которые требуют ре-

шения. В число этих проблем входят также и проблемы обусловленности и индивидуальной ответственности.

В современном обществе забота о сохранности персональной информации в значительной мере возлагается непосредственно на самого индивида. Сложился институт информирования пользователей и получения их добровольного согласия на использование их персональных данных. Дав такое согласие, пользователь берет ответственность за возможные неблагоприятные последствия на себя. Поэтому необходимо создание соответствующих правовых условий и инфраструктуры, которая позволила бы индивидуальному пользователю стать реальным субъектом двухстороннего соглашения с субъектами, формирующими БД об индивидуальном поведении. В этих условиях он сможет контролировать доступность своих индивидуальных данных. Однако современное состояние дел далеко от этого. В настоящий момент, если пользователь сам не предпринимает мер по защите своих персональных данных, все возможные негативные последствия ложатся целиком на него самого. Это представляет собой специфическую форму реализации принципа обусловленности в современную эпоху. В соответствии с этим принципом, принятым на вооружение в государствах благосостояния, индивид имеет право на защиту только при условии соблюдения соответствующих норм поведения. Так же как и в случае с благосостоянием индивида, такой подход стимулирует его к тому, чтобы предпринимать активные меры по защите своей информации. Однако поскольку вероятность успешности этих усилий крайне низка, стимул к самостоятельной защите данных понижается. Таким образом, в современной «экономике БД» организации, осуществляющие сбор и коммерциализацию данных, занимают доминирующее положение по сравнению с рядовыми пользователями.

Отношение к этой проблеме на уровне государственной политики различается в странах Евросоюза. Так, правительство Германии придерживается более жесткой ограничительной политики в отношении возможностей сбора и коммерциализации БД, содержащих информацию об индивидуальном поведении. Власти Великобритании отстаивают более свободный, рыночно ориентированный подход, перекладывая ответственность за безопасность использования индивидуальной информации на самих пользователей. Правительство рассматривает «экономику БД» как одну из самых перспективных сфер увеличения национального производства и повышения конкурентоспособности национальной экономики. В одном из программных документов британского правительства говорится буквально

следующее: «Конечно, новые технологии приносят с собой новые риски, и мы нуждаемся в более широкой системе защиты данных, обеспечивающей безопасность наших граждан. Но распространение данных неизбежно. Если мы не создадим благоприятный климат для использования этих новых возможностей, можно быть уверенным в том, что инновации, связанные с БД, будут распространяться в других странах, и нам придется приобретать права на их использование у других» [цит. по: с. 8].

Исходя из этих соображений, правительство поставило цель вывести Великобританию в лидеры в плане развития экономики на базе использования БД, что станет гарантией поддержания устойчивого экономического роста и процветания в будущем. Такая стратегия предполагает, что в будущем БД об индивидуальном поведении будут все более активно использоваться для получения экономического эффекта, и государство намерено создавать для этого все необходимые условия. Согласно этой стратегии, БД – это «глобальные товары», а это предполагает минимизацию ограничений на операции с ними. Кроме того, стремительное снятие ограничений на использование данных сопровождается существенным изменением статуса индивидов в создании БД и связанной с ними новой рыночной стоимости.

Стратегия правительства содержит в себе раздел, посвященный правам граждан и ответственности за использование данных. Прежде всего упор делается на индивидуальную ответственность пользователя за сохранность его личной информации. Это явно соответствует принципу индивидуальной ответственности, на котором строится современное государство благосостояния. В скрытой форме здесь также присутствует принцип обусловленности. Ведь возможности защиты индивидуальных данных доступны только людям, обладающим достаточно высокой степенью компетентности в области информационных технологий. В результате степень защиты для пользователя оказывается обусловленной степенью его технической компетентности.

Подобная тенденция наблюдается и в Германии, правительство которой придает все возрастающее значение развитию «экономики БД». Оно также заявляет о необходимости развития экономики на базе современных информационных технологий, и в частности – активного использования современных БД [с. 9]. Такие заявления – это открытый сигнал о необходимости создания благоприятных правовых и институциональных условий для развития бизнеса, использующего БД в качестве товара и производственного

ресурса. При этом все более очевидным становится тот факт, что существующие ограничения использования БД и меры по защите личных данных пользователей вступают в существенное противоречие с заявленными целями, направленными на развитие соответствующей бизнес-модели, и подрывают конкурентоспособность национальной экономики в этой сфере. Это вызывает стремление к либерализации в этой сфере и расширению возможностей по сбору и использованию БД. В связи с этим предполагается замена концепции защиты индивидуальных данных, использовавшейся до сих пор в нормативных документах, на концепцию информационного суверенитета, позволяющую более свободный сбор и использование данных. Утверждается также, что сложившиеся в ЕС принципы конфиденциальности и защиты информации препятствуют развитию «экономики данных» и поэтому не должны применяться при формировании национального законодательства Германии, регулирующего эту сферу. В начале 2017 г. правительством Германии был представлен новый проект закона о защите информации. В нем существенно смягчаются многие ограничения и ответственность субъектов, осуществляющих сбор данных в Сети. Этот проект встретил сильное сопротивление со стороны целого ряда общественных организаций, занимающихся проблемой защиты данных, и в результате его наиболее радикальные положения не были реализованы в окончательной редакции документа. Тем не менее произошел явный сдвиг политики в данной сфере в сторону предоставления все большей свободы организациям, осуществляющим сбор и коммерциализацию БД.

Развитие информационных технологий и рост количества данных ведут к глубокой трансформации современного общества. Это порождает целый комплекс проблем. По мнению автора статьи, существующий конфликт между интересами использования данных об индивидуальном поведении, создаваемых при помощи современных технологий, и защитой интересов и конфиденциальности рядовых пользователей сегодня разрешается на базе принципов обусловленности и индивидуальной ответственности, которые приняты в современном развитом обществе. Они предполагают, что ответственность за защиту интересов индивида распределяется определенным образом между индивидом и государством и что сфера ответственности индивида достаточно велика и имеет тенденцию к возрастанию.

*М.О. Лихачев*

**С.М. Пястолов**

**ЭКОНОМИЧЕСКИЕ И СОЦИАЛЬНЫЕ ЦЕННОСТИ,  
РИСКИ И ОРГАНИЗАЦИОННЫЕ ПАРАДОКСЫ  
БОЛЬШИХ ДАННЫХ  
(Обзор)**

В современной науке о технологиях и в деловых кругах большие данные (БД) рассматриваются как средство для технологического развития и осуществления прорыва. Принимая технологии БД, частные и общественные организации ожидают получения больших выгод во многих областях деятельности – в развитии электронной коммерции, электронного правительства, науки, а также в здравоохранении и в сфере безопасности.

Таким образом, теории и практики БД имеют как социальную, так и экономическую ценность. Социальная ценность определяется приростом общественного блага в таких областях, как образование, здравоохранение, государственная безопасность, безопасность производства и граждан и т.д. Ценность обеспечивается как для отдельных пользователей, так и для всего социума, и выражается в росте занятости, производительности и потребительского излишка. Экономическая ценность может быть измерена величинами роста прибыли, экономического роста и конкурентных преимуществ. Например, есть эмпирические данные, подтверждающие факт того, что организации, которые реализуют стратегии и ежедневные операции с использованием БД, имеют лучшие финансовые показатели, чем организации, которые этого не делают. В ходе дальнейшего развития рассматриваемых технологий экономическая и социальная ценности могут быть получены за счет расширения возможностей принятия тактических решений и роста информационного обеспечения процессов разработки стратегий.

Понятие больших данных обычно определяется на основе представления о больших объемах экстенсивно различных данных, которые произведены, получены и обработаны на высокой скорости [1; 3]. Кроме того, их сложно обрабатывать при помощи существующих технологий.

Эксперты отмечают, что дискуссии на этапе возникновения феномена БД характеризуются «беспочвенным оптимизмом», появлению которого способствовало раздувание ажиотажа, «моды на ИТ». Соответственно, актуализировалась потребность в критическом размышлении о том, как организации понимают практическую ценность БД, что требует дополнительного анализа потенциалов отдельных организаций и экономической системы в целом.

Группа исследователей из Амстердамского свободного университета (Vrije Universiteit Amsterdam) идентифицировала в общей сложности шесть проблем, обсуждающихся в настоящее время в организациях разного уровня в связи с пониманием ценности БД. Эти проблемы были разделены на три группы: 1) практика работы с БД; 2) организационные модели; 3) согласование интересов различных сторон [1]. Далее в этих группах были выделены основные вопросы: а) индуктивные и дедуктивные методы анализа БД; алгоритмический и человеческий интеллект; б) централизованные и децентрализованные структуры в сфере БД; возможности развития инноваций и бизнес-моделей; в) контролируемые и открытые БД; минимизация и пренебрежение социальными рисками в этой сфере [1, с. 201]. Экспертная группа отобрала научные публикации («корзина восьми» журналов в области информационных систем (IS) и материалы конференций ICIS, ECIS и AMCIS), изданные в период с 2000 по 2016 г., так как именно в этот период начались активные исследования работ с большими объемами неструктурированных данных.

Основные приложения результатов других исследований такого рода были обозначены на научном форуме «На переднем крае машинного обучения», который был проведен в январе 2017 г. в Вашингтоне (округ Колумбия) в штаб-квартире Национальных академий естественных, инженерных и медицинских наук. Форум был совместно организован Национальными академиями США и Королевским научным обществом Великобритании. К очному участию в форуме организационный комитет пригласил 60 известных ученых, и более чем 500 заинтересованных лиц наблюдали за интернет-трансляцией.

*Машинное обучение*, как наиболее явное проявление достижений теории и практики в сфере БД, имеет множество приложений, число которых постоянно растет. Одним из самых ранних применений данной технологии был «информационный поиск», источником в котором обычно является документ неструктурированной природы (текст). Системы распознавания речи стали использоваться еще в 1950-х годах, но тогда они могли переложить только небольшое количество речевых сигналов на слова и фразы. Более современное программное обеспечение продвинулось до уровня понимания и использования естественных речевых образцов (учитываются, например, скорость, тон, выбор слова), которые характерны для нормальной беседы. Использование многослойных нейронных сетей – метод, известный как «глубокое обучение», – позволяет преобразовать звуки речи в текст, однако все еще сохраняются трудности в различении голосов.

*Целевая реклама* – прием, ставший популярным в 1990-х годах, – увеличивает вероятность покупки или определенных действий потребителя. Чтобы добиться этого, компании собирают данные о поведении своих потребителей или покупают кейсы и применяют методы машинного обучения, чтобы прогнозировать модели будущего потребительского поведения.

*Глубокое обучение* – класс мощных методов машинного обучения, существенная особенность которых состоит в том, что они не только выполняют определенную функцию, но и являются композицией ряда подфункций для построения моделей. Глубокое обучение имеет много общего с работой человеческого мозга: искусственные нейронные сети действуют как реальные нейроны, соединяя различные понятия на входе с потенциальными результатами расчетов.

Почтовые фильтры электронного спама, интернет-поисковые системы, устройства для распознавания речи, переводчики, программы для распознавания фотоизображений, автоматические почтовые серверы, обнаружение объектов в автоматизированных транспортных системах, навигационные приложения и т.п. – все эти разработки призваны посредством глубокого (само)обучения ЭВМ упростить решение различных задач для пользователей.

Так, автоматизированные транспортные средства могли бы увеличить мобильность пожилых людей и людей с ограниченными возможностями, уменьшить загрязнение окружающей среды и поддержать новые модели транспортной логистики. Однако остаются вопросы о полном понимании предполагаемых выгод и

рисков, например в случае беспилотных автомобилей это риски для пешеходов и велосипедистов.

В медицинских исследованиях, например, широко используют генетические данные, записи показателей за длительное время, экологические данные и др., чтобы лучше понять все аспекты здоровья человека. Развитие статистических методов, связанных с машинным обучением (в том числе функциональная магнитно-резонансная томография), позволит исследователям проектировать медицинские процедуры, оперативные вмешательства, которые улучшат здоровье людей.

Одно из ответвлений машинного обучения – *искусственный интеллект* – следует интерпретировать как «*искусственный общий интеллект*». «Цель исследований в этой области состоит в том, чтобы создать системы, которые были бы способны адаптироваться к широкому диапазону сложных задач и могли бы развить способности, соответствующие интеллектуальным способностям людей или превышающие их» [3, с. 6].

В то же время в области машинного обучения специалисты сталкиваются с рядом «канонических» проблем. Прежде всего это проблемы классификации, в которые включают распознавание лиц, идентификацию изображения, постановку медицинского диагноза. Эти задачи решаются с помощью таких методов, как построение логистических регрессий, нейронных сетей, векторные вычисления, обработка случайных чисел.

В следующий класс проблем входит регрессионный анализ, который применяется в том числе в прогнозировании – финансовом, воздействия интернет-рекламы. Методы, применяемые в решении данных проблем, – линейная регрессия, нейронные сети, гауссовские процессы.

Типичной проблемой считается также кластеризация множеств данных. Этот прием используется в биоинформатике, исследованиях экспрессии гена, астрономии, моделировании документов и сетей. Общие методы включают применение  $k$ -средних, гауссовой смеси распределений и процессов распределения Дирихле.

Сокращение размерности необходимо в случаях, где исходные данные имеют большое число измерений. Анализ базовых компонентов, факторный анализ, многомерное масштабирование, Iso Map и гауссовские модели с латентными переменными используются в целях снижения размерностей данных в научном анализе, идентификации изображений и в других случаях отбора данных [3, с. 7].

Алгоритмы искусственного интеллекта, используемые в обучающихся компьютерах, все в большей степени становятся способны предсказывать поведение человека в реальном времени. Наряду с большими данными алгоритмы позволяют повысить уровень автоматизации принятия оперативных и стратегических решений, которое традиционно считалось сложным и поэтому требующим человеческого вмешательства. Развертывание таких алгоритмов все более «формирует мир», где автоматика будет способна работать без человеческого вмешательства: например, случаи, когда алгоритмы учатся по ходу работы, принимают в реальном времени решения, основанные на анализе дискретного поведения, таким образом одновременно влияя на решения человека.

Между тем некоторые ученые утверждают, что слишком большая уверенность в алгоритмах, присущая лицам, принимающим решения, может привести к потере или замене профессионального знания, особенно когда не ясно, каким образом алгоритмы достигают определенных результатов, паттернов и решений. Так, наблюдения за поведением оператора мобильной сети показывают, что определение аудитории по данным о сетевой активности (затем эта информация может быть продана рекламодателям) в значительной степени зависит от результатов исследования данных оператором-человеком. Вывод по итогам таких исследований состоит в том, что процессы аналитики больших данных «будут разрабатываться так, чтобы в них был задействован человек» [1, с. 196].

Аналогично в более ранних исследованиях роли технологий в автоматизации и «информатизации» рабочих мест внимание было обращено на эффекты, обусловленные различиями между «умными людьми» и «умными машинами». Это могло бы означать, что еще рано судить о фактических изменениях в организации работ по внедрению методов больших данных в сравнении с теми, которые уже вызвали такие технологии, как системы управления знаниями. Пока еще остается неясным, какие именно особенности человеческого и алгоритмического интеллекта будут более востребованы в различных ситуациях (например, устойчивые и обычные против временных ситуаций на стадии становления).

Развитие организационных моделей вследствие ввода технологий БД обсуждается в следующих аспектах: централизация и децентрализация оргструктур; инновации; совершенствование бизнес-моделей.

Эффективный метод централизации ресурсов с целью компенсации недостатка аналитических навыков, по мнению многих

экспертов, – создание центров аналитической компетентности. Например, Банк Англии, центральный банк Соединенного Королевства, сформировал специальные подразделения, занимающиеся сбором и анализом БД, – центры компетенции: аналитическое подразделение (Advanced Analytics Division) с целью анализа БД и Лабораторию данных (Data Lab) с целью сбора, хранения, обработки, визуализации и анализа сегментированных, неструктурированных данных, доступных служащим [1, с. 197].

Однако результаты опроса руководителей разных организаций показывают, что централизация увеличивает вероятность принятия моделей управления на основе БД, тогда как децентрализация затрудняет управление данными в организационных единицах и подразделениях. Более того, в исследовании распространения инноваций (2015) были получены подтверждения того, что БД сами по себе практически не децентрализуются вследствие многих причин: из соображений безопасности; из-за распространенного мнения, что БД – это ключевой ресурс, за которым должен следить один «главный управляющий данными»; из-за роста коммуникационных затрат; из-за растущих потребностей в сотрудничестве и централизованных закупках, а также в дорогих сопутствующих технологиях, навыках и знаниях [1, с. 197].

Централизованные структуры в проектах БД потенциально ограничивают коммуникацию и причастность различных заинтересованных акторов бизнеса. Однако такое взаимодействие значимо для реализации соответствующих проектов. Организации также обновляют свои бизнес-модели, когда само наличие технологий БД, помимо прочего, принуждает их развивать новые ценности, формировать новые суждения или по-разному строить взаимоотношения с разными клиентами. Так, ИТ-компания Netflix отказалась от своей традиционной модели проката дисков и перешла к потоковой модели. В частности, компания теперь предлагает потоковую передачу мультимедиа для клиентов, разрабатывает динамические рекомендации, основанные на поведенческих паттернах, и использует текущие данные, чтобы сформировать адекватное содержание. Еще один пример изменения бизнес-модели – компания Nike, которая от исключительного производителя обуви продвинулась до владельца цифровой платформы по оказанию фитнес-услуг, управляемых данными. «Нью-Йорк Таймс», традиционная газета, вложила капитал в эксперименты с данными и теперь использует их результаты, чтобы привлечь читателей в цифровую среду. Исследования показывают, что превращение БД

в капитал посредством усовершенствования бизнес-моделей, инновационных подходов может быть стимулировано внедрением новых вычислительных средств. Одним из таких средств мог бы стать квантовый компьютер. Однако есть ряд еще не решенных проблем, которые были освещены в специальном отчете академий наук США [2].

Классический компьютер использует биты, чтобы представить величины, которыми он оперирует; квантовый компьютер использует квантовые биты, или кубиты (qubit). Если в первом случае бит может быть записан как 0 или 1, то кубит может быть представлен величинами 0 или 1 или их комбинацией (суперпозицией). Если информация классического компьютера представлена сочетаниями битов в единственном состоянии, то состояние квантового компьютера с тем же самым количеством квантовых битов может охватить все возможные состояния соответствующего классического компьютера, и таким образом новое устройство работает в экспоненциально большем пространстве задач. Однако условия использования этого пространства требуют того, чтобы все кубиты были неразрывно связаны (запутанны), хорошо изолированы от внешней среды и очень четко управляемы.

Многие инновации, разработанные за прошедшие 25 лет, позволили построить физические системы, которые уже начинают обеспечивать изоляцию и контроль, необходимые для квантовых вычислений. В 2018 г. в большинстве квантовых компьютеров использовались две технологии («пойманные в ловушку» ионы и искусственные «атомы»), произведенные за счет свойств сверхпроводимости). Кроме того, в настоящее время тестируются другие технологии с целью физической реализации кубитов (физические квантовые биты). Но несмотря на быстрое развитие данной области, еще слишком рано «делать ставки» хотя бы на одну из технологий квантовых вычислений.

Однако даже если будут созданы кубиты высокого качества, использование квантовых компьютеров влечет за собой новые проблемы. Технические риски, например, включают проблему «системного шума». Поскольку кубит может быть любой комбинацией единицы и ноля, квантовые фильтры могут пропустить небольшие ошибки – шум, который создают потоки в малых физических частицах. В итоге самые мелкие ошибки в квантовых операциях или любые случайные сигналы могут в конечном счете привести к неправильным результатам вычислений. Для устранения подобных и других ошибок необходимы алгоритмы квантовой

коррекции ошибок (quantum error correction – QEC). Без QEC маловероятно, чтобы сложная квантовая программа, осуществляющая, например, алгоритм Шора, будет работать правильно на квантовом компьютере.

Проблематично также осуществление загрузки в квантовый компьютер больших объемов данных. В настоящее время нет метода быстрого преобразования большого объема классических данных в квантовое состояние (этого не требуется, если данные могут быть произведены алгоритмически). Реализация существующих методов создания квантового состояния на входе, как правило, сильно уменьшает преимущество квантового компьютера. Кроме того, достижение «квантового ускорения» требует совершенно новых видов дизайна и принципов разработки алгоритма. Эта задача – критический аспект данной сферы [2, гл. 3]. Необходимы новые цепочки инструментов, аппаратного и программного обеспечения. Этому должны быть посвящены соответствующие научные исследования [2, раздел 6.1].

На более высоком организационном уровне большое значение приобретают контакты с лицами, принимающими решения. Вне организационных границ соответствующие заинтересованные лица представляют университеты и научно-исследовательские институты, правительства, источники данных, пользователей и клиентов [1, с. 198].

Предполагается, что для получения выгоды от БД организации должны обмениваться данными со своими партнерами. И здесь можно выделить ряд проблемных полей. Организации могут неохотно делиться или обмениваться данными с сетевыми партнерами, например из-за проблем конфиденциальности и безопасности, защиты частной жизни или коммерческой тайны, когда раскрытие информации подвергает опасности уникальное стратегическое положение организации. По этим причинам организации, как правило, контролируют доступ к данным с помощью многих способов, например посредством механизмов рынка, где собственность и права могут быть явно определены путем заключения формального контракта. Кроме того, доступ к данным и алгоритмам может регулироваться иерархическим способом, задаваемым доминирующей организацией.

Коммерческие организации и правительства могут также открывать доступ к своим данным, например для того чтобы стимулировать инновации и обеспечивать прозрачность. Так, Банк Англии в части его One Bank Research Agenda взял на себя обязательство

открыть для общественности ранее проприетарные наборы данных для краудсорсинга и решения сложных политических вопросов. Местные органы власти в США также внедрили открытые порталы данных, чтобы делиться БД с гражданами и предприятиями. В случаях открытия доступа эти данные структурированы инфраструктурой ИТ, которая «устанавливает правила производства данных» [1, с. 199]. Так, социальные организации, СМИ, возможно, не всегда сообщают о том, каким образом получены данные, которые они делают доступными через свои интерфейсы. Организации могут также предоставлять данные, полученные на трастовых (доверительных), а не на формальных основаниях. Активная коммуникация и прозрачные соглашения между сетевыми партнерами считаются крайне важными для сотрудничества в области обмена данными. Кроме того, потребители, пользователи и общественность (агенты, которые хранят данные и получают информационные продукты) становятся все более и более активными. Они также являются законными участниками сети. Как следствие, организации должны будут учитывать юридические и этические последствия использования данных, например, им, вероятно, придется просить согласия пользователей на использование данных. На сверхорганизационном уровне этот факт поднимает такие вопросы, как социальные риски использования БД, например, практика объединения источников анкетных данных может открыть доступ к очень личной и деликатной информации. Точно так же существующая сегодня тенденция предлагать клиентам персонифицированные продукты и услуги, основанные на анализе БД, ставит много вопросов, связанных с частной жизнью, «кражей личности», незаконной дискриминацией, несправедливой классификацией и даже «эксплуатацией уязвимых» (exploitation of the vulnerable) [1, с. 199]. Более того, даже если надзор за отдельными лицами со стороны правительств и других организаций может привести к улучшению общественного контроля и безопасности, меры наблюдения также затрагивают частную жизнь, чувства свободы и автономии людей. Это говорит о том, что, помимо регулирования, организациям часто приходится иметь дело с общественными ожиданиями и этическими соображениями в процессе реализации ценности БД.

Таким образом, БД ставят перед учеными много дополнительных вопросов. До настоящего времени дискуссии, посвященные этой проблеме, демонстрировали оптимистический тон. Однако значительная доля обсуждений, связанных с юридическими и этическими вопросами, вращается вокруг социальных рисков, снижая этот опти-

мизм. Эти риски возникают тогда, когда определенные акторы в экосистеме (как правило, организации) проявляют свою власть над другими (как правило, людьми) посредством аналитики БД. К настоящему времени эмпирические исследования только затронули некоторые аспекты того, как организации пытаются решить соответствующие юридические и этические проблемы. В связи с этим отмечен интересный парадокс: с одной стороны, механизмы (например, стандарты, принципы, политика) должны быть достаточно устойчивыми, чтобы учитывать мобильность и взаимосвязанность в пределах и между организациями, а с другой – «правила игры» необходимо постоянно пересматривать.

Когда создание социальной ценности является важной целью для организации, это влияет на этические соображения акторов относительно сбора и анализа данных: в исследованиях и при принятии решений они ограничены тем, что считают этичным и юридически обоснованным. Коммуникационные структуры, нормы и ценности отличаются в зависимости от контекста (например, в производственном или национальном секторах). И такие аспекты должны учитываться в дальнейших исследованиях.

### Список литературы

1. Debating big data: A literature review on realizing value from big data / Günther W.A., Rezazade Mehrizi M.H., Huysman M., Feldberg F. // *Journal of strategic information systems*. – 2017. – Vol. 26. – P. 191–209.
2. Quantum computing: Progress and prospects // *National academies of sciences, engineering, and medicine*. – Wash., DC: The National academies press, 2018. – XVII, 202 p. – DOI: 10.17226/25196.
3. The frontiers of machine learning: 2017 Raymond and Beverly Sackler U.S.-U.K. scientific forum // *National academies of sciences, engineering, and medicine*. – Wash., DC: The National academies press, 2017. – 32 p. – DOI: 10.17226/25021.

**Ки К.Ф., Шрок А.Р.**  
**ЛУЧШИЕ СОЦИАЛЬНЫЕ  
И ОРГАНИЗАЦИОННЫЕ ПРАКТИКИ  
УСПЕШНЫХ НАУЧНЫХ ИНТЕРНЕТ-ПОРТАЛОВ  
И КИБЕРИНФРАСТРУКТУРНЫХ ПРОЕКТОВ  
(Реферат)**

Ref. ad op.: Kee K.F., Schrock A.R. Best social and organizational practices of successful science gateways and cyberinfrastructure projects // Future generation computer systems. – 2018. – In press. – Mode of access: <https://doi.org/10.1016/j.future.2018.04.063>

Американские эксперты рассмотрели 12 лучших методик, применяемых в организации успешных проектов киберинфраструктуры науки (КИН).

Для эффективной работы ученых сегодня нужна хорошая КИН, частью которой являются научные интернет-порталы – так называемые научные врата (science gateways), призванные упростить проведение исследований, обмен данными, укрепить сотрудничество между учеными и снизить барьеры на пути к электронной науке (e-science). В связи с рассматриваемыми примерами научных интернет-порталов авторы выделяют несколько вопросов, которые не совсем обычны для практики научного исследования. Например, федеральные агентства, как источник финансирования, являются объектом поиска в академической сфере, но не в организациях частного сектора. Основная причина такого положения заключается в том, что цикл финансирования таких проектов конечен. Кроме того, проблемой является постоянная текучесть кадров, которая частично возникает вследствие привлечения к исследованиям студентов и аспирантов, имеющих недостаточно времени для работы над долгосрочными задачами. Это в меньшей степени относится к таким тех-

нологическим компаниям, как Microsoft, Apple и Google, которые и являются основными разработчиками научных интернет-порталов и других проектов КИН.

Организуя сбор информации для своего исследования посредством интервью с технологически искушенными участниками киберпроектов, авторы были несколько удивлены тем, насколько эта тематика оказалась важной в анализе данных. Но они быстро поняли, что определенные каналы связи являются адекватными лишь для определенных коммуникационных целей. Например, коммуникация лицом к лицу – более подходящее средство для установления отдельных целей и нахождения общего языка. Респонденты подчеркнули, что встречи с глазу на глаз важны в начале сотрудничества и должны периодически происходить на всех этапах жизненного цикла проекта. Было бы полезно для научных команд КИН учитывать опыт этой практики, чтобы преодолеть то, что Дж.Н. Каммингс и С. Кислер (J.N. Cummings, S. Kiesler) называют «затратами координации мультиинституционального сотрудничества» [с. 7].

Было проведено 98 глубоких интервью с представителями различных заинтересованных сторон, включая доменных ученых как пользователей, вычислительных технологов как разработчиков, администраторов центров по исследованиям суперкомпьютеров в США (n = 93) и ЕС (n = 5). От США в общей сложности были представлены 26 штатов с наибольшими выборками: Иллинойс (n = 19), Индиана (n = 16) и Калифорния (n = 13) [с. 2].

Социальные методики успешных проектов, выявленные по результатам исследования, объединяет одна характерная черта – наличие команды мультидисциплинарных экспертов в составе сотрудников и в организациях-партнерах. Необходимость мультидисциплинарной экспертизы обусловлена тем, что научные интернет-порталы часто оказываются местом сборки различных решений определенных проблем.

Успешные порталы и / или КИН обычно устанавливают общие цели для разнообразных групп сотрудников из различных научных областей. Администратор такого центра в Калифорнии заявил в интервью: «Я имел обыкновение думать, что деньги играют большую роль. Деньги играют роль, но фактически цели важны в большей степени» [цит. по: с. 3].

Одной из ошибок руководителя проекта, считают авторы, является его стремление сначала определить цель, а затем поставить ее перед сотрудниками. Но у многих талантливых сотрудни-

ков есть свои собственные устремления и цели, что может привести к «некоаксиальностям цели» в дальнейшем. В таком случае руководителю следует выявить отдельные цели, установить обратные связи и быть открытым для обсуждения и объединения новых целей.

Успешные команды КИН формируют общий язык. Он имеет особое значение в тех случаях, где нет явных точек соприкосновения – «общих знаний, общих верований и общих предположений» [с. 3]. Тогда общий язык базируется на понимании всеми участниками терминов, которые могут иметь различные значения в разных дисциплинах.

В команде необходим коммуникатор, обеспечивающий взаимодействия между пользователями и разработчиками в проекте. Пользователи часто не имеют глубоких знаний о вычислительной науке, а разработчики киберинфраструктуры часто недостаточно разбираются в научной дисциплине, чтобы предложить пользователям полезный инструмент для научных исследований.

Стабильность продолжающихся операций и непрерывность процесса обеспечивает установление производственных рутин. Это тем более важно для рассеянных команд и организаций, работающих над текущими задачами. Возможно, практика регулярных коротких встреч вместо нечастых и недлительных, по мнению авторов, была бы более эффективной в плане использования общего времени. Регулярные встречи с глазу на глаз также важны для успеха проекта, даже если у сотрудников есть предшествующая история успешного сотрудничества онлайн.

К разряду организационных методов авторы относят: демонстрацию альтруистического лидерства; четкое определение ролей; наличие обратной связи с пользователями; финансирование, обеспечивающее чувство уверенности и стабильности; низкий уровень текучести кадров.

Общий вывод исследователей состоит в том, что динамически развивающиеся проекты научных интернет-порталов требуют соблюдения баланса динамизма и предсказуемости.

*С.М. Пястолов*

**Масланов Е.**

**КРАУДСОРСИНГ В НАУКЕ:  
НОВЫЙ ЭЛЕМЕНТ НАУЧНОЙ ИНФРАСТРУКТУРЫ //**  
**Философия науки и техники. – 2018. – Т. 23, № 1. –**  
**С. 141–155. – DOI: 10.21146/2413-9084-2018-23-1-141-155**  
**(Реферат)**

Автор, сотрудник Института философии РАН, обсуждает технологию краудсорсинга, которая все чаще применяется учеными и научно-исследовательскими подразделениями корпораций для привлечения непрофессионалов к решению научно-исследовательских задач. Технология краудсорсинга – это использование времени, интеллектуальных или других ресурсов «толпы», большого количества людей, для решения какой-либо задачи. Задача автора статьи – показать, что краудсорсинговые проекты в области науки и техники могут стать одним из элементов научно-исследовательской инфраструктуры.

Традиционно научная инфраструктура включает в себя научное оборудование различного назначения и масштаба, ресурсы, электронные инфраструктурные элементы и услуги, используемые научным сообществом для проведения исследований<sup>1</sup>. Автор предлагает рассматривать систему краудсорсинговых площадок как часть нового инфраструктурного элемента науки, способствующего вовлечению все большего количества людей в деятельность, связанную с решением научно-технических задач, и росту инновационной культуры.

---

<sup>1</sup> Инфраструктура исследований и разработок, большая наука и международное научно-технологическое сотрудничество / под ред. Г.В. Трубниковой. – М.: Министерство образования и науки РФ: Нац. исслед. ун-т «Высшая школа экономики», 2016. – 43 с. – Режим доступа: [https://issek.hse.ru/data/2016/06/29/1115938709/05\\_Trubnikov.pdf](https://issek.hse.ru/data/2016/06/29/1115938709/05_Trubnikov.pdf)

Чтобы решить поставленную задачу, автор анализирует систему распределенного знания. Активное участие в краудсорсинговых проектах приводит к тому, что ситуативные знания становятся частью системных знаний участника проекта, но часть знаний может использоваться лишь однократно. Формируется сообщество профессионалов, которые, не обладая формальным институциональным признанием, все же владеют достаточными компетенциями для решения научно-технических задач на высоком профессиональном уровне. И в результате применения технологии краудсорсинга создается массив «распределенного знания», который может использоваться любым участником проекта.

Автор перечисляет уже сложившиеся элементы современной инфраструктуры, которые сделали возможным быстрое развитие научного краудсорсинга. Так, новые информационно-коммуникационные технологии (ИКТ) позволили создать пространства, в которых люди могут активно взаимодействовать друг с другом, не выходя из дома, и получать доступ практически к любой информации. В результате были сформированы онлайн-сообщества, объединившие пользователей по всему миру. С помощью современных ИКТ краудсорсинговые платформы могут привлечь к решению научно-технических задач большое количество людей, которые на данном этапе не вовлечены в производство научных и технических знаний.

Участники краудсорсинговых проектов используют в том числе и знания, почерпнутые из интернет-источников, но часто они оказываются отрывочными и недостаточными. В качестве примера автор приводит опыт известного проекта Fold.it, в котором в игровой форме решалась научная задача по упорядочиванию структуры белка<sup>1</sup>. В проекте приняли участие большое количество добровольцев, которые могли не только «собирать» головоломки, но и получать научные знания из этой игры.

При этом краудсорсинговые проекты в области науки и технологий, которые ориентируются не только на использование времени и интереса участников, предполагают активное использование данных из различных источников. К таким проектам относится, например, один из проектов NASA, в котором участникам нужно решать техническую задачу. В этом случае использу-

---

<sup>1</sup> Predicting protein structures with a multiplayer online game / Cooper S., Khatib F., Treuille A., Barbero J., Lee J., Beenen M., Leaver-Fay A., Baker D., Popović Z.; Foldit players // Nature. – 2010. – Vol. 466, N 7307. – P. 756–760.

мые «данные» могут постоянно обновляться. После решения определенной задачи часть из них может больше никогда не использоваться и быть заменена на новые, необходимые для участия в следующем проекте.

Автор выделяет две группы знаний: первая группа носит системный характер и формируется в процессе образования и самообразования, а вторая используется исключительно для решения конкретной задачи. Дело в том, что сама технология поиска формирует два типа используемых в исследованиях знаний: обладающих системным характером и ситуативных. Участник краудсорсингового проекта может не выстраивать системные связи между этими группами, так как вторая группа является лишь ситуативной. Однако успешное решение сложных фундаментальных научных и технических задач предполагает не только использование определенных быстродоступных «данных», но и умение находить нетривиальные связи между феноменами, выстраивать сложную систему аргументации, предполагает встраивание нового знания в систему уже имеющегося знания. Достижение этих результатов, опираясь только на ситуативные «данные», представляется достаточно сложным [с. 147].

Важный элемент любого краудсорсингового проекта – нацеленность на решение определенной задачи. В проектах в области науки и техники задачу ставит группа специалистов. Они обладают контролем над проектом, поэтому экспертное мнение выступает фильтром, отсеивающим спорные предложения. В таких проектах участвуют люди, которым либо интересна поставленная задача, либо они имеют опыт по решению схожих задач. Они обладают экспертными навыками, основанными на опыте, но не обладают необходимыми навыками взаимодействия с учеными. Краудсорсинговый проект может использоваться как инструмент встраивания общего языка.

К числу краудсорсинговых проектов большой значимости автор относит и всемирное хакерское движение. Показано, что хакерская культура оказала влияние на развитие сети Интернет. Во-первых, культура хакеров – это питательная среда для выдающихся технических инноваций, в связи с тем что она основана на принципах сотрудничества и свободной коммуникации. Во-вторых, она выступает в качестве передаточного звена между знаниями, порожденными так называемой техномеритократической культурой и предпринимательской деятельностью, которая способствует распространению Интернета в мировом сообществе [с. 148].

Проекты в области краудсорсинга могут рассматриваться как элемент формирования массива «распределенного знания» в целях решения научной или технической проблемы. При этом ни один из участников проекта не является носителем всех знаний, используемых и созданных в проекте. С одной стороны, построение краудсорсинговой площадки предполагает, что иногда часть работы по оценке проектов может быть осуществлена самими участниками проекта, а поэтому откровенно слабые предложения не анализируются экспертами. С другой стороны, эксперты могут вынести на совместное обсуждение только те проекты, которые уже прошли предварительную оценку со стороны части экспертного сообщества. Поэтому хотя потенциально эксперты и обладают доступом ко всем знаниям, сосредоточенным на платформе, актуально они взаимодействуют лишь с небольшим массивом знаний.

В результате сама краудсорсинговая площадка становится материальным носителем распределенного знания, оно аккумулируется на ней и используется участниками для решения общих задач. Построение краудсорсинговой площадки в области науки и техники на основе принципов сотрудничества, а не соревнования приводит к интересному эффекту – субъектом познания становится не отдельный индивид, а все участники краудсорсингового проекта. В этом случае сложно сказать, кто внес основной вклад в решение научной проблемы и можно ли вообще выделить такого участника.

Развитие краудсорсинговых проектов в области науки и техники ставит вопрос и о статусе экспертов и участников таких проектов. Распространение краудсорсинговых площадок может привести к снижению статуса научного эксперта. Не только участники проектов, но и все общество может поверить в то, что для решения научно-технической задачи совершенно не обязательно обладать специальными знаниями – достаточно лишь принять участие в краудсорсинговом проекте.

Дилетантизм участников краудсорсинговых проектов в области науки и техники носит скорее институциональный характер. Они не являются сотрудниками компаний или научных организаций, запустивших краудсорсинговый проект, но, имея университетское образование, обладают компетенциями, позволяющими решать поставленные задачи. Работа на краудсорсинговых площадках напоминает работу в проектных командах, которые создаются для решения конкретной задачи, а после того как задача решена, команда может быть распущена или направлена на решение следующей задачи. Формирование обычной проектной команды требует подбора

персонала, налаживания процесса коммуникации между ее участниками, размещения на территории исследовательского центра, предоставления оборудования. В случае использования технологии краудсорсинга всего этого можно избежать. Профессиональным исследователям необходимо четко сформулировать задачу, предложить решить ее участникам проекта, а затем заниматься отбором наиболее перспективных предложений.

В заключение автор отмечает, что использование этой технологии связано с некоторыми рисками. В настоящее время поиск информации в Сети все чаще становится персонализированным, т.е. поисковые алгоритмы ориентируются на поведение конкретного пользователя. В итоге история поиска влияет на полученные результаты. Доверие непрофессионала к полученным при помощи поиска в Сети данным может повлиять на результаты краудсорсинговых проектов. В свою очередь, успех или неуспех краудсорсингового проекта влияет на степень охвата сферы науки, технологий и образования новыми элементами инфраструктуры – краудсорсинговыми площадками.

*С.В. Егоров*

**Тхайил Н.**

**ПОСТРОЕНИЕ ГЛОБАЛЬНЫХ ДАННЫХ:  
АВТОМАТИЗИРОВАННЫЕ МЕТОДЫ ЭКОЛОГИЧЕСКОГО  
МОНИТОРИНГА, ОСТОРОЖНОГО ОТНОШЕНИЯ  
К РИСКАМ И ИХ НАГЛЯДНОГО ПРЕДСТАВЛЕНИЯ  
(Реферат)**

Ref. ad op.: Thayyil N. Constructing global data: Automated techniques in ecological monitoring, precaution and reification of risk // Big data & society. – 2018. – P. 1–14. – DOI: 10.1177/2053951718779407

Автор реферируемой статьи, сотрудник Индийского института технологий (Дели), рассматривает проблемы, связанные с автоматической агрегацией больших данных (БД), и ее последствия для экологических наук. Анализируются алгоритмы автоматизации обработки данных и алгоритмы анализа массивов данных при долгосрочном мониторинге биоразнообразия.

Автора интересуют сценарии возможного признания общественностью автоматически полученных массивов данных в качестве надежного источника для использования в процессе поддержания биоразнообразия, а также влияние тех или иных форм сбора данных на способы управления рисками. Особое внимание уделяется автоматизированному получению и анализу данных в ходе долгосрочного экологического мониторинга, изучения и сохранения глобального биоразнообразия.

При изучении процессов сбора БД всегда важно понять, каким образом, казалось бы, несопоставимые данные после автоматизированной обработки становятся общепризнанными последовательными и «глобальными» массивами. Автор предполагает, что именно эвристика осторожного дискурса помогает принять новые методы обработки данных, и знакомит с основными способами все

более широкого использования агрегированных данных в науках об окружающей среде. Ранее в метеорологических исследованиях уже вырабатывались меры превращения разрозненных данных в полный и согласованный глобальный массив.

Автор исследует перспективу очевидного соответствия между эвристикой осторожных подходов и автоматизированным мониторингом на основе данных. Это соответствие, по его мнению, и есть фактор укрепления доверия и общественного признания определенных сообществ, включая «непрофессиональные» группы, которые уже участвуют в сохранении биоразнообразия. Отмечены также парадоксы, возникающие в работе по повышению доверия к методам обработки данных при мониторинге биоразнообразия, и показано, каким образом конкретные формы данных соотносятся с конкретными формами управления рисками. Современные дискурсы риска в статье рассматриваются в увязке с институтами, которые составляют своеобразную регулирующую экосистему. Реферируемая статья состоит из трех больших блоков, раскрывающих: а) современные методы автоматизации сбора данных; б) современные подходы к проблемам БД; в) методы управления рисками.

Вопрос о масштабах массивов данных анализируется с двух сторон: 1) с точки зрения большого объема данных в режиме реального времени; 2) с точки зрения охвата больших территорий в региональном и глобальном масштабе. Этот вопрос всегда поднимается в дискуссиях о доверии общественности и приемлемости таких методов в принятии решений.

**Автоматизация.** Методы автоматизации используются для сбора, систематизации, проверки и распространения данных, необходимых для получения знаний об окружающей среде. Такие знания используются для экологического регулирования, в том числе посредством долгосрочного экологического мониторинга. Методы автоматизации и агрегирования данных нашли применение в исследованиях биоразнообразия, в том числе в долгосрочном экологическом мониторинге. Интенсивные статистические исследования и интеллектуальный анализ данных, в которых применяется широкий спектр алгоритмов, все чаще используются для обнаружения и оценки рисков в исследованиях окружающей среды. При использовании глобальных данных в ходе формирования экологической политики возникают большие правовые проблемы. Однако технология частично противоречит факторам, обеспечивающим осторожный подход к рискам, например страдают прозрачность и наглядное представление рисков.

Методы автоматизированного агрегирования данных при выявлении угроз биоразнообразию требуют более полного понимания в рамках существующей практики регулирования рисков. Эпистемологические и политические аспекты анализа рисков способствуют появлению «осторожного» подхода к принятию решений в экологии, а также в области охраны здоровья.

Использование автоматизированных объективных моделей анализа для вытеснения существующего экспертно-субъективного анализа в метеорологических исследованиях имеет важное сходство с предполагаемым поворотом в подходе к обработке данных в науке об окружающей среде в целом и долгосрочном мониторинге биоразнообразия в частности. Автоматизированные исследования, применяющие БД, имеют большие перспективы в качестве основы для разработки политики и ее реализации в структурах управления. Это может стать важным координирующим этапом в глобальном управлении экологическими рисками в целом и в сохранении биоразнообразия в частности.

**Массивы данных.** Запрос на использование автоматизированных методов агрегирования данных в науках об окружающей среде растет. Для развития этих методов сегодня важно учитывать предыдущий успешный опыт метеорологических прогнозов. Как известно, массив БД должен удовлетворять критерию «четырех V»: volume (объем), velocity (скорость), variety (разнообразие) и veracity (достоверность). Большое значение имеет не только тот факт, что данные имеют огромные объемы, но и то, что эти данные поступают из различных источников с возрастающей скоростью. Эта множественность способствует достоверности и целостности массива и в конечном итоге обеспечивает необходимое доверие. Известен лозунг «экологам нужны большие данные и большая экология» (цит. по: с. 4). Соответственно, знания сегодня создаются в различных точках – там, где большие массивы данных имеют перспективы стать движущей силой как для научных исследований, так и для управления в области окружающей среды. Густые сети беспроводных датчиков позволяют собирать данные из различных источников для архивирования, обработки, анализа и интеграции в удаленных лабораториях.

Например, Институт горных исследований (Mountain Research Institute, MRI) в течение последнего десятилетия стремится объединить существующие разрозненные наборы данных в глобальный набор наблюдений и унифицировать их. Подобные усилия всегда вызывают вопросы по поводу надежности и доступности данных. Инициатива MRI направлена на увязку и использование дан-

ных экологических и социально-экономических наблюдений по всем регионам в целях проведения сопоставлений на основе фактических данных, принятия обоснованных решений и разработки управленческих решений. Акцент на использовании данных в глобальной экологической политике и управлении заметен и в других инициативах MRI. Так, MRI инициировал диалог между различными дисциплинами социальной науки по вопросу взаимодействия человеческого воздействия и общественного экологического управления в горных районах. Усилия MRI направлены на оценку устойчивого развития горных районов с использованием массивов БД.

Задолго до того, как Интернет вещей стал популярным и модным, структуры для использования больших агрегированных данных (глобальных данных) уже внедрялись в другие области знаний, например в метеорологические исследования. Применение воздушных шаров, радиометров, самолетов – разведчиков погоды, метеорологических спутников, режим наблюдения в реальном времени при централизованном управлении обеспечивают создание массива глобальных данных. Большие массивы глобальных данных, полученных из подобных источников по всему миру, включают также данные, перенесенные из «богатых данными регионов» в регионы, где их недостает [с. 5].

Важной составной частью технологии наполнения массива глобальных данных в ходе метеорологических исследований является выход за пределы охвата реальности приборами и переход к автоматическому созданию объективных моделей. Автоматическое агрегирование больших разрозненных данных все чаще рассматривается в качестве главного элемента в производстве экологических знаний, а также в качестве основы для разработки глобальной экологической политики. Фактор глобальности масштаба является определяющим по мере приближения к предполагаемому пересмотру данных в экологических науках. Этот и другие факторы являются ключевыми для того, чтобы в будущем массивы данных обеспечили бы подлинную научную революцию в экологии, которая характеризуется большой неоднородностью фактов, поступающих в ходе исследования.

Эксперты приходят к выводу, что в целях прогресса в понимании экологических проблем «следует перейти на следующий уровень управления информацией, предоставляя революционно новые приложения управления данными»<sup>1</sup>.

---

<sup>1</sup> Reichman O.J., Jones M.B., Schildhauer M.P. Challenges and opportunities of open data in ecology // Science. – 2011. – Vol. 331. – P. 703.

Массивы больших данных определенным образом представлены в современном регулировании экологических рисков. Вопросы доверия и общественного признания в отношении методов автоматизации и базирующихся на данных исследований сохранения биоразнообразия в глобальном масштабе неразрывно связаны с противоречиями между существующими парадигмами традиционного риска и осторожного риска.

**Риски и управление.** Дискурс осторожного обращения с рисками способствует, по мнению автора, признанию таких методов общественностью. Автоматизированный мониторинг биоразнообразия делает осторожное отношение к рискам предпочтительным, так как дает более точную картину за счет более быстрого, экономически эффективного и менее запутанного сбора данных в больших масштабах в течение длительных периодов времени. Традиционный риск понимается как объективное явление, допускающее наглядное представление, и может определяться посредством экспертных оценок. В свою очередь, для такой оценки выполняется расчет вероятности того, что могут произойти конкретные неблагоприятные события, идентифицируемые как опасности. Этот вид количественной оценки и расчета вероятностей опирается на инженерию, экономику, психологию, эпидемиологию. Сегодня это основа современного регулирования рисков.

В регулировании традиционных рисков можно выделить техническую оценку и политическое управление. Парадигмы, связанные с осторожными рисками, предусматривают более публичный способ создания данных, поддерживающих регулирование рисков. При этом признается спорный характер сопоставления, анализа и использования поступающей информации, с учетом пробелов в измерениях, двусмысленности данных и просто невежества поставщиков информации.

Превращение исследований биоразнообразия в комплексную дисциплину будет, возможно, связано с принятием на вооружение мер осторожного риска правительствами. Дело в том, что попытки построения полного, последовательного и непротиворечивого глобального массива данных с использованием неполной и непоследовательной информации о состоянии окружающей среды сегодня представляются ключом к формированию новой правительственной практики осторожного риска.

В важном разделе «осторожный риск и обработка данных в мониторинге биоразнообразия» основное внимание уделяется управлению традиционными рисками, а также роли экспертных методов

генерирования данных. Обсуждается также развитие осторожных подходов к управлению рисками, которое приводит к необходимости наглядного представления рисков.

Далее автор подробно рассматривает применение БД в целях экологического мониторинга. Он выделяет два аспекта: долгосрочный экологический мониторинг и обеспечение широкого общественного участия в оценке рисков. Стремясь дополнить традиционную архитектуру рисков, сторонники принципа осторожности заявляют о ее актуальности при разработке и проведении научных исследований. Вопрос изменения возможного поворота в сторону автоматизированной науки, основанной на БД, также связан с сохраняющейся конкуренцией между двумя парадигмами управления рисками. Эта конкуренция вызывает озабоченность общественности по поводу недостаточно наглядного представления экспертных знаний в области регулирования традиционных рисков. Напротив, принцип осторожности призван смягчить озабоченность и демократизировать представление рисков. По-видимому, методы автоматизации обработки вскоре станут нормой, соответственно, следует понять, каким образом использование этих методов в научных исследованиях и в ходе принятия решений станет общепринятым. Осторожный подход также рассматривается и как способ восстановления доверия, утраченного традиционными парадигмами управления рисками.

Доверие – необходимая составляющая в использовании массивов данных, полученных с помощью автоматизированных методов агрегирования. Оно не может быть установлено лишь на основе горячих научных дискуссий, проходящих в технократических сообществах. Поэтому, считают эксперты, важно рассмотреть вопрос о том, какие новые варианты восстановления доверия к технике сбора и обработки данных можно найти в рамках осторожной парадигмы управления в экологии.

Внутренняя согласованность массивов данных дает им возможность стать основой для последующего регулирования экологических рисков и нормотворчества на глобальном уровне. Сторонники крупномасштабного объединения огромных массивов данных указывают на качественные сдвиги в производстве знаний. Сегодня используются алгоритмы для обработки больших объемов неструктурированных данных из различных источников в целях прогнозирования тенденций, закономерностей и определения причинно-следственных связей. Агломерация данных способствует вариативности результатов при решении различных вопросов. Так,

при мониторинге биоразнообразия факторы уязвимости отдельных видов должны быть установлены на региональном, если не на глобальном уровне, включая численность видов, мигрирующих на большие расстояния или подверженных воздействиям, которые не могут быть ограничены на субрегиональном масштабе. Например, в ряде международных договоров определены глобальные принципы защиты биоразнообразия, а также предусмотрены рамки и методы сбора соответствующих данных.

Более поздние общественные дискуссии по поводу эффективных и законных мер регулирования в интересах окружающей среды порождают напряженность между пересекающимися (но концептуально различными) парадигмами регулирования традиционных рисков и мер осторожного подхода. Все чаще осторожные подходы, а именно стремление к более наглядному представлению рисков и обеспечению участия общественности в анализе и менеджменте рисков, пропагандируются в качестве меры борьбы с различными ограничениями в регулировании традиционных рисков.

Научные дискуссии в кругу современных экологов напоминают аналогичные дискуссии в кругу метеорологов, происходившие в 1980-х годах. В результате этих дискуссий можно ожидать продвижения новых методов формирования «глобальных данных» в целях осуществления разумной экологической политики. Цель современных дискуссий состоит в решении проблем автоматизированных исследований, основанных на БД, в парадигме осторожного риска. Можно предвидеть использование этих методов в различных эпистемологических сообществах, например в кругах ученых-экологов, экспертов – оценщиков риска и других сообществах. Однако большое значение имеет признание новых методов и со стороны волонтерских групп, и со стороны широкой общественности.

**Роль «науки граждан».** Большие автоматизированные агрегированные данные стимулируют сдвиги в подходах к стандартизации в экологических науках. Благодаря этому появились перспективы применения осторожных подходов в двух важных областях – в привлечении общественности и долгосрочном экологическом мониторинге. Важными предварительными шагами в этом направлении в области экологических наук являются получение больших объемов разнородных по содержанию и разнородных по источникам данных, обеспечение их достоверности и доступа к соответствующим научным и нормативным материалам.

Проекты «науки граждан» хорошо справляются с получением таких разрозненных данных по целому ряду дисциплинарных об-

ластей в области экологии и обеспечивают оценку рисков силами общественности. Помимо генерирования и сбора экологических данных вклад граждан используется для проверки экологической информации в других добровольческих проектах – особенно в таких областях, как орнитология, палеонтология, астрономия и атмосферные науки. Например, Корнелльская лаборатория орнитологии осуществляет несколько проектов «науки граждан» различного масштаба, привлекая тысячи людей к сбору и представлению данных наблюдений за птицами. Эти проекты направлены на изучение влияния изменений окружающей среды на темпы размножения, распространение инфекционных заболеваний среди диких животных, влияния кислотных дождей на популяцию птиц и изменение распределения популяций птиц во времени и пространстве.

Привлечение граждан к сбору научной информации о биоразнообразии полезно не только при решении вопросов, имеющих большой пространственный или временной охват, но и при долгосрочном мониторинге. Тем не менее есть опасность, что массив данных, собранных «непрофессионалами» при долгосрочном мониторинге, может оказаться поддельным. Данные, которые предоставляют граждане, даже сравнивают с рассказами туземного обитателя, на которые опираются колониальные этнографы. Тем не менее сохраняется значительный потенциал укрепления доверия и общественной приемлемости методов, использующих участие «непрофессиональных» групп, проявляющих активный интерес к защите биоразнообразия.

Многие экологи считают, что их дисциплина готова к трансформации на основе массивов БД. Этому способствует появление стандартов метаданных как в сообществах, так и на уровне объединенных хранилищ данных. Однако пока неясно, будут ли наборы данных, полученные с помощью автоматизированных методов, приняты в качестве глобальных экологических данных даже в пределах конкретных научных сообществ.

Эксперты предвидят трудности в продвижении автоматизированных подходов и напоминают, что выбор гипотез, основанных на данных, не может рассматриваться как альтернатива выбору гипотез экспертами или прямому получению знаний. В лучшем случае – это комплементарный подход. Оба подхода могут быть полезны, однако соотношение затрат и выгод для каждого из них различно. Наглядное представление возможно и для автоматизированных моделей даже с учетом того, что основанный на наблю-

дениях интенсивный экологический мониторинг обещает более подходящую платформу для участия граждан.

### **Пересмотр роли БД и парадоксы признания массивов.**

Биологи уже доказали явные преимущества более быстрого и более точного способа сбора БД на больших отрезках времени посредством автоматизированного мониторинга. Учитывая быструю утрату местообитаний видов, сохранение биоразнообразия требует срочного расширения масштабов мониторинга с помощью этих методов для создания более однородных, надежных и постоянных массивов данных. По-прежнему ощущается запрос со стороны государственных регулирующих органов и международных природоохранных учреждений на привлечение этих методов для мониторинга биоразнообразия и его сохранения.

В свое время произошел пересмотр роли данных в метеорологическом моделировании. В этой области знаний переход к изображению данных сделал «глобальные» данные прозрачными и доступными, в то время как сами базовые модели оставались непрозрачными. В экологии сегодня складывается похожая картина: непрозрачность в управлении рисками, основанном на данных, может лишить осторожного дискурса всю институциональную практику. Это воспроизводство непрозрачной технократической парадигмы при наличии открытых и наглядных массивов данных и есть важный очевидный парадокс. При этом данная парадигма претендует на эмпирическую точность и реалистичность и стремится обеспечить базу для осторожного подхода. Таким образом, возникает препятствие для общественного признания таких массивов данных, как глобальные экологические данные.

Преодоление выявленных в данной работе парадоксов имеет важное значение для признания методов, которые будут использоваться для сохранения биоразнообразия. Тем не менее приемлемость новых методов для общественности будет достигнута или не достигнута в зависимости от эффективности мер активного стимулирования доверия.

*С.В. Егоров*

**Бёрнер К.**  
**НАУЧНАЯ ПОЛИТИКА,**  
**ОСНОВАННАЯ НА УПРАВЛЕНИИ ДАННЫМИ**  
**(Реферат)**

Ref. ad op.: Börner K. Data-driven science policy // Issues in science and technology. – 2016. – Vol. 32, N 3, Spring. – Mode of access: <http://issues.org/32-3/perspective-data-driven-science-policy/>

Автор, сотрудница трех ведущих университетов США, Нидерландов и Германии, рассматривает проблемы, связанные с выбором новых подходов, основанных на управлении большими данными (БД). Такие подходы применяются администраторами при решении вопросов, касающихся разработки научной политики и финансирования научных исследований.

Самая важная проблема чиновников, занимающихся вопросами научной политики, – как использовать ограниченные ресурсы наиболее эффективным образом. Чтобы решить ее, они должны обладать огромным объемом знаний: прежде всего понимать внутреннюю работу научной, технологической и инновационной (НТИ) системы (science, technology and innovation system), процесс производства знаний, а также знать результаты работы конкретных ученых и организаций. В этом им могут помочь специальные подходы, основанные на использовании компьютерных мощностей и больших объемов разнообразных данных, касающихся динамики развития науки и технологий: извлечение и анализ данных (data mining), моделирование процесса (modeling process), визуализация алгоритмов (visualization algorithms) и т.д.

Все эти подходы давно используются в промышленности, например фирмы Amazon и Netflix применяют компьютерное моделирование, чтобы определить направления поведения своих

клиентов, а финансовые компании – чтобы выявить факты мошенничества. Ученые, работающие в разных областях, устанавливают дорогое оборудование для совместного использования с партнерами из промышленности и правительства: в метеорологии – для предсказания погоды и предупреждения о надвигающихся штормах; в здравоохранении – для определения и предупреждения эпидемий и т.д. Однако в исследованиях динамики развития НТИ-системы до сих пор, по мнению автора статьи, эти методы используются не в полной мере, несмотря на то что некоторые эксперты все же пытаются применять их при принятии важных решений. Компьютерная симуляция и компьютерное моделирование помогают определить пути развития науки и техники и предсказать возможный результат вмешательства государства. С их помощью администраторы могут легко найти ответы на важные вопросы: какие научные организации будут наиболее продуктивными в будущем; какая научная политика финансирования окажется наиболее эффективной; какие социальные факторы влияют на работу ученых – демографические изменения, альтернативные экономические траектории и т.д.

Современная наука – это совместное производство знаний. Она требует затраты огромных усилий целого ряда различных акторов, создания сетей, состоящих из отдельных исследователей, организаций, идей и технологий. Иногда такая сеть включает несколько сотен участников из разных стран. Некоторые проекты имеют такой размах, что необходимы усилия и знания множества ученых и исследователей. Многочисленные исследования показывают, что наиболее продуктивными и эффективными оказываются большие команды. Однако применение новых подходов показало противоположные результаты. По словам С. Милошевич из Университета штата Индиана, проведенное ею компьютерное моделирование продемонстрировало, что успех больших команд основан на эффективной работе малых команд и даже отдельных исследователей. Это открытие может стать решающим при принятии решений, связанных с финансированием научных проектов.

Й. Боллен и коллеги из Университета штата Индиана попытались выяснить, сколько усилий и средств тратится учеными на то, чтобы получить субсидии на свои исследования. Проведя компьютерное моделирование, они обнаружили, что ученые вкладывают огромное количество времени и энергии в написание и рецензирование исследовательских заявок, а финансирующие агентства используют многочисленные ресурсы, которые могли бы быть потра-

чены на сами исследования. Й. Боллен и коллеги подсчитали, что подготовка исследовательской заявки на грант требует работы четырех профессоров в течение четырех недель, и их работа оценивается в 35 тыс. долл. В среднем грант ННФ составляет около 165 тыс. долл. в год, часть суммы идет на возмещение косвенных расходов университета, и остается примерно 110 тыс. долл. на исследование. Таким образом, средний успех приводит к чистому убытку для исследователей. При этом расчеты не учитывают время, затрачиваемое на написание и рецензирование заявок.

Й. Боллен и коллеги предлагают другой способ распределения финансовых средств, направленных на исследовательские проекты, – так называемую модель Fund Rank. Она заключается в том, что каждый авторитетный ученый один раз год получает определенную сумму на исследования, например 100 тыс. долл. 50% суммы он самостоятельно распределяет среди своих коллег, принимающих участие в данном проекте, в соответствии со степенью их участия и вкладом в работу. Вместо того чтобы тратить массу времени на написание и проверку заявок, ученые могли бы использовать его более плодотворно – на более детальное обсуждение, планирование и реализацию научно-исследовательского проекта. Распределение финансов могло бы послужить мерой оценки вклада, сделанного каждым исследователем, а использование цифровых систем могло бы помочь в решении проблем, связанных с возникающими в процессе работы спорами, распределением средств на основе личных симпатий, выявлением конфликтов интересов или случаев взаимных благодеяний.

Политики и администраторы должны не только понимать, как использовать БД и основанные на них методы, но и доверять им и активно применять их на практике. Свою эффективность доказал метод визуализации процесса моделирования и его результатов. Команда под руководством Уильяма Роуза (William Rouse) из Института технологии Стивенса совместно с Национальными академиями инжиниринга и медицины разработали «симулятор полетов для политиков» (policy flight simulator), позволяющий «слетать в будущее, прежде чем выписать чек». Эта методика помогает экспертам прогнозировать направления развития науки и техники, для того чтобы принять правильные и ответственные решения. Работа симулятора заключается в следующем: группа экспертов (от пяти до 15 человек) собирается в зале, оснащенном техническими средствами. Они получают нужные им данные, обсуждают все возможные варианты, касающиеся данных вопросов, и затем пропускают их через симулятор. Графические

презентации результатов появляются на экранах. Например, одна из таких встреч была посвящена проблемам выбора мощностей для автомобиля: сравнивались различные системы – внутреннего сгорания, электрические, гибридные или водородные. Результаты моделирования показали, что спрос на электрические автомобили возрастет, если возрастут государственные вложения в инфраструктуру зарядных устройств.

Разработка и применение простых и действенных моделей, основанных на управлении БД, для принятия решений в области НТИ не только несут большие возможности, но и создают большие проблемы. Для того чтобы выявить «критические вызовы» и ответить на них, необходимо использование междисциплинарного подхода, а также тесное сотрудничество государства, науки и промышленности. Вычислительные модели должны быть тщательно проверены экспертами и заслужить доверие научного сообщества. Ключ к созданию такого доверия – прозрачность и вовлечение всех заинтересованных сторон в разработку и применение новых подходов, основанных на управлении данными.

*М.П. Булавинова*

# ПРАВОВОЕ РЕГУЛИРОВАНИЕ БОЛЬШИХ ДАННЫХ<sup>1</sup>

**Т.В. Захаров**

## ПРАВОВЫЕ ПРОБЛЕМЫ ОБОРОТА БОЛЬШИХ ДАННЫХ В УСЛОВИЯХ ЦИФРОВОЙ ЭКОНОМИКИ (Статья)

*Аннотация:* В статье рассматривается правовое регулирование сбора, хранения и использования больших данных в экономических отношениях, освещается проблема «информационной филантропии», анализируется влияние больших данных на конкуренцию, вопросы защиты персональных данных.

*Ключевые слова:* цифровая экономика; большие данные; информационно-коммуникационные технологии; конфиденциальная информация; персональные данные.

**T.V. Zakharov**

### Legal problems of big data circulation in digital economy

*Abstract:* The article deals with the issues of legal regulation in the sphere of accumulation, storage and exploitation of big data in economy, the question of «data philanthropy». The influence of big data on competition and the problem of personal information safety are analyzed.

*Keywords:* digital economy; big data; information and communication technologies; confidential information; personal data.

Широкое применение информационно-коммуникационных технологий в экономических отношениях позволяет говорить о становлении неизвестного ранее явления – «цифровой экономики». Одной из ее базовых основ стал анализ транзакционной информации, аккумулируемой средствами электронного протоколирования, накопленные массивы которой получили наименование «большие

---

<sup>1</sup>Раздел подготовлен сотрудниками Отдела правопедения Центра социальных научно-информационных исследований ИНИОН РАН.

данные» (*big data*). В современной экономике выделился самостоятельный сегмент хозяйственной деятельности (бизнес-модель), всецело построенный на аккумулировании и эксплуатации информации. Так работают зарубежные компании Google, Amazon, E-bay и российские – «ВКонтакте», «Яндекс», «Авито», «Юлмарт».

Правовое регулирование общественных отношений, связанных с аккумулированием информации и использованием результатов ее анализа, является сложным, комплексным. В публичном праве задействованы институты правовой защиты информации, контроля за ее оборотом, защиты неприкосновенности частной жизни, конфиденциальной информации, государственной и коммерческой тайны, обеспечения информационной безопасности. В частном праве на первый план выходят право интеллектуальной собственности, конкуренции, нормы защиты прав потребителей. Данные отношения развиваются крайне динамично, затрагивая иные институты публично-правового и частноправового регулирования. Однако в значительной части рассматриваемые отношения не получили правового определения.

Сложность регулирования связана с разграничением режимов благоприятствования и ограничения использования больших данных. Первый привлекателен и для экономической сферы, и для сферы государственного управления. К тому же правовое регулирование не способно решить всех проблем. Жесткие правовые рамки существенно ограничат возможные выгоды, а отсутствие правового поля создает риск появления «серой» зоны оборота и использования информации. Данные зачастую вторично используются вне их изначального функционального контекста. Встают сложные вопросы – как и для каких целей осуществляется оборот информации, на каких условиях она собиралась и т.п.

Самостоятельную проблему представляет определение понятия «большие данные», что препятствует их правовому регулированию. Что касается экономических отношений, то сложно установить механизмы формирования стоимостного выражения больших данных, потенциальные способы их использования при взаимодействии конкурирующих хозяйствующих субъектов и потребителей.

**Оборот больших данных как неотъемлемая часть цифровой экономики. Ее основные характеристики.** Термин «большие данные» вошел в научный оборот, хотя определить его исчерпывающее смысловое значение непросто. Ключевым признаком новой категории информации стали особенности электронной – «цифровой» – фиксации отдельных, «базовых» элементов структуры информации, предоставляющие возможность сложной автоматизированной обра-

ботки. Определяя понятие, следует особо выделить такую очевидную, но тем не менее имеющую принципиальное значение черту, как доступность для анализа данных сложной персонифицированной выборки. Результаты анализа составляющей большие данные информации могут иметь правовое значение<sup>1</sup>, актуализируя проблемы сбора, хранения (предоставления) и использования информации.

Ученые определяют понятие больших данных по-разному. Лаконичное определение предлагает профессор права и профессор философии Школы права Пенсильванского университета (University of Pennsylvania) (США) А. Аллен. Термин «большие данные» обозначает комплексный феномен – совокупность многообразной информации, поддающейся суммированию в обширные массивы, хранению и анализу. В использовании данной информации исследователи могут применять комплексные алгоритмы и технологии искусственного интеллекта для выявления производных данных, их комбинаций, определенных родовыми признаками, связей, стереотипов, тенденций, идентификационных признаков и практических знаний<sup>2</sup>.

Информация, составляющая большие данные, поступает из разных источников – от государственных органов и коммерческих организаций, реестров транзакций потребителей, цифровых программных приложений оборудования – так называемого Интернета вещей (*Internet of things*). Индивиды незримо дополняют большие данные. Многие проживают собственную виртуальную жизнь в сети Интернет, осуществляя поиск информации в электронных информационных системах, размещая информацию в социальных сетях и т.п. Большинство обращаются к электронным сервисам государственного и муниципального управления, вовлечены в механизмы цифровой экономики, расплачиваясь кредитными картами, получают медицинскую помощь в учреждениях здравоохранения, ведущих электронный учет данных о пациентах, подают заявления о приеме на работу посредством сети Интернет и т.п.<sup>3</sup>

---

<sup>1</sup>См.: *Spindler G.* Internet intermediary liability reloaded: The new German Act on responsibility of social networks and its (in-) compatibility with European law // Journal of intellectual property, information technology and e-commerce law. – Berlin; Karlsruhe, 2017. – Vol. 8, N 2. – P. 166–167; *Belli L., Sappa C.* The intermediary conundrum cyber-regulators, cyber-police or both? // Journal of intellectual property, information technology and e-commerce law. – Berlin; Karlsruhe, 2017. – Vol. 8, N 3. – P. 184, 185–186.

<sup>2</sup>См.: *Allen A.* Protecting one's own privacy in a big data economy // Harvard law review. – Cambridge, 2016. – Vol. 130, N 2. – P. 71.

<sup>3</sup>Ibid.

Став неотъемлемой частью цифровой экономики, оборот больших данных в полном объеме воспринял ее существенные характеристики. Цифровая экономика, с одной стороны, характеризуется устойчивым ростом, динамичностью, непрерывным появлением новых услуг, концентрацией рынков, особым положением участников с доминирующей позицией, с другой – растущим неравенством между большими компаниями и малыми и средними компаниями и потребителями<sup>1</sup>.

Традиционно большие данные характеризуются своим объемом (*volume*), скоростью оборачиваемости (*velocity*) и многообразием (*variety*)<sup>2</sup>. Профессор права Пресвитерианского университета Маккензи (Mackenzie Presbyterian university) (Бразилия) В. Баньоли считает, что для идентификации проблем больших данных с позиции права следует выделить также такие их существенные характеристики, как ценность (*value*), достоверность (*veracity*) и подтвержденность (*validation*)<sup>3</sup>.

В. Баньоли выступает за введение в научный оборот понятия «релевантный рынок больших данных», для правового анализа которого необходима сегментация. Автор предлагает выделить следующие сегменты: приобретение, хранение и анализ больших данных<sup>4</sup>. Потребность в сегментации обусловлена различиями в складывающихся правовых отношениях, составе их участников, природе сопутствующих проблем.

Следует учитывать, что информация, генерируемая в больших данных, не доступна всем участникам рынка в том же объеме и качестве, в каком она доступна аккумулирующим ее компаниям. Лишь 15% из пяти сотен крупнейших мировых компаний (*Fortunate 500*) имели возможность использовать преимущества больших данных к концу 2015 г. Лишь 8% из пяти сотен крупнейших мировых компаний в 2015 г. были способны применять анализ больших данных<sup>5</sup>.

---

<sup>1</sup> См. Big data: The next frontier for innovation, competition and productivity / McKinsey global institute: – N.Y., 2011. – P. 123–125.

<sup>2</sup> См.: *Ridgeway G.* Policing in the Era of Big Data // Annual review of criminology. – Philadelphia, 2018. – Vol. 1. – P. 403–405; *Richards N., King J.* Big data ethics // Wake Forest law review. – Winston-Salem, 2014. – Vol. 49, N 2. – P. 394.

<sup>3</sup> См.: *Bagnoli V.* The big data relevant market // Concorrenza e mercato. – Rome, 2016. – Vol. 23. – P. 79.

<sup>4</sup> См.: *Bagnoli V.* Op. cit. – P. 80–82, 88–93.

<sup>5</sup> См.: *Kemp R.* Legal aspects of managing Big Data // Computer law & security review. – Amsterdam, 2014. – Vol. 30, N 4. – P. 482–483.

Большие данные и большая аналитика поднимают множество правовых, моральных и этических вопросов. Наиболее важные – кибербезопасность и ответственность государственных органов и негосударственных, коммерческих организаций за применяемые аналитические алгоритмы. С одной стороны, большие данные становятся основой для принятия стратегических решений все большим числом участников управленческих и хозяйственных отношений. С другой стороны, государственные антимонопольные органы уделяют недостаточное внимание способности больших данных влиять на функционирование рынков<sup>1</sup>.

Отдельную проблему представляют современные агрессивные стратегии участников рынка при сборе персональной информации потребителей. Так, условием доступа к продуктам и сервисам на электронных рынках является согласие с предлагаемыми правилами пользования, которые зачастую становятся попросту средством извлечения персональных данных потребителя<sup>2</sup>. Распространенная практика предоставления услуг бесплатно фактически означает – в обмен на доступ к персональным данным. Денежная компенсация взамен предоставления данных в настоящее время практически недоступна<sup>3</sup>.

Актуальна проблема достоверности и подтвержденности больших данных. Их анализ должен проводиться с учетом локальных, культурологических и этнографических особенностей<sup>4</sup>. Географическое измерение рынка должно приниматься в расчет при анализе, равно как и хронологический угол зрения<sup>5</sup>. Сложность алгоритмов автоматизированного анализа данных крайне затрудняет возможность выявлять и обжаловать случаи «непреднамеренной» дискриминации<sup>6</sup>.

---

<sup>1</sup> См.: *Stucke M., Grunes A.* Big data and competition policy. – Oxford, 2016. – P. 1.

<sup>2</sup> См.: *Metzger A.* Data as counter-performance: What rights and duties do parties have? // *Journal of intellectual property, information technology and e-commerce law.* – Berlin; Karlsruhe, 2017. – Vol. 8, N 1. – P. 2–8; *Yergin R.* Consent in the age of Facebook: Applying the telephone consumer protection act to text messages from social media platforms // *Columbia law review online.* – N.Y., 2016. – Vol. 116. – P. 81, 84–85.

<sup>3</sup> См.: *Bagnoli V.* Op. cit. – P. 84–85.

<sup>4</sup> См.: *Sarfaty G.* Can big data revolutionize international human rights law? // *University of Pennsylvania journal of international law.* – Philadelphia, 2017. – Vol. 39, N 1. – P. 97.

<sup>5</sup> См.: *Bagnoli V.* Op. cit. – P. 81.

<sup>6</sup> См.: *Barocas S., Selbst A.* Big Data's disparate impact // *California law review.* – Berkeley, 2016. – Vol. 104, N 3. – P. 674.

Контроль за сохранностью больших данных составляет особую группу проблем. Главная из них – ненадлежащее использование государственными органами и организациями имеющихся у них персональных данных<sup>1</sup>.

Несмотря на широкий спектр проблем, связанных со сбором и электронным анализом обширных информационных массивов, в основу исследования феномена больших данных должен быть положен тезис о совокупности их позитивных эффектов. Доступность электронной аналитики делает ее неизбежной.

Профессор права Университета Вашингтона в Сент-Луисе (Washington University in St. Louis) (США) М. Ричардс и вице-президент по облачным стратегиям и развитию бизнеса компании Century Link technology solutions (США) Д. Кинг настаивают на необходимости предоставить правительствам возможность использовать большие данные в целях защиты населения. Кроме того, следует шире подходить к пониманию информации, разрабатывая этику конфиденциальности, нормы и правовые механизмы защиты таких важных социальных ценностей, как тайна частной жизни, доверие, прозрачность и защита от угрозы подчинения новым возможностям анализа информации<sup>2</sup>.

Говоря о потенциале больших данных в предупреждении нарушений прав человека, профессор Юридической школы П. Алларда Университета Британской Колумбии (Allard school of law, University of British Columbia) (Канада) Г. Сарфати указывает на необходимость побуждать компании к «информационной филантропии» (*data philanthropy*) – добровольной передаче данных в публичных общественных интересах (например, для предотвращения нарушений прав человека)<sup>3</sup>. Директор инновационной инициативы ООН в сфере данных для целей развития «Глобальный пульс» Р. Киркпатрик занимает более радикальную позицию – информационная филантропия должна стать приоритетом частного сектора<sup>4</sup>.

---

<sup>1</sup> См.: Towards a modern approach to privacy-aware government data releases / Altman M., Wood A., O'Brien D., Vadhan S., Gasser U. // Berkeley technology law journal. – Berkeley, 2015. – Vol. 30, N 3. – P. 1967–2072.

<sup>2</sup> См.: Richards N., King J. Op. cit. – P. 432. См.: также, например: Scassa T. Law enforcement in the age of big data and surveillance intermediaries: Transparency challenges // SCRIPT ed. – Edinburgh, 2016. – Vol. 14, N 2. – P. 239–284.

<sup>3</sup> См.: Sarfaty G. Op. cit. – P. 97–98.

<sup>4</sup> См.: Kirkpatrick R. A new type of philanthropy: Donating Data // Harvard business review. – Cambridge, 2013. – Digital article. – Mode of access: <https://hbr.org/2013/03/a-new-type-of-philanthropy-don>; Sarfaty G. Op. cit. – P. 97–98.

Информационная филантропия может достигаться различными (экономическими, регулятивными) средствами, например, формируются доступные информационные базы данных, которые становятся публичными после надлежащей деперсонализации и обобщения. Могут быть установлены «сигнальные» системы (*alerting network*), когда наиболее уязвимые данные, не становясь публичными, тем не менее анализируются компаниями (в пределах аппаратных средств сетевой защиты) в определенных чрезвычайных ситуациях<sup>1</sup>.

Информационная филантропия способна служить долгосрочным интересам как компаний, так и государственного управления и общества в целом<sup>2</sup>. Ответственная социальная позиция привлечет и объединит работников, для которых характерна высокая моральная ответственность, этически ориентированных потребителей. Вовлеченность в партнерские отношения с широким кругом единомышленников (государственными и негосударственными организациями, академической средой, общественными группами и т.п.) принесет долгосрочные выгоды в виде дополнительного опыта, общественной поддержки, повышения доверия, вовлеченности в коммерческие и некоммерческие формы взаимодействия и др.<sup>3</sup>

**Влияние больших данных на конкуренцию в условиях цифровой экономики.** Большие данные расширяют поле и условия конкурентной борьбы коммерческих организаций и требуют внимательного изучения для понимания новой роли антимонопольного права. С одной стороны, большие данные могут наделять исключительной властью, предоставляя конкурентные преимущества – оперативность и эффективность принятия решений. С другой стороны, они могут быть использованы для развития хитроумных механизмов ограничения конкуренции, консолидации средств доминирования на рынке, одностороннего вмешательства в функционирование рынка<sup>4</sup>.

В. Баньоли выделяет следующие группы отношений, требующих особого внимания антимонопольного права в связи с использованием больших данных: слияния и поглощения; ограничительные действия, такие как отказ в доступе, дискриминационный доступ к данным, эксклюзивные контракты, принуждение к при-

---

<sup>1</sup> См.: *Sarfaty G.* Op. cit. – P. 97–98.

<sup>2</sup> См.: *Porter M., Kramer M.* The competitive advantage of corporate philanthropy // *Harvard business review.* – Cambridge, 2013. – Vol. 80, N 12. – P. 56, 57.

<sup>3</sup> См.: *Sarfaty G.* Op. cit. – P. 99–100.

<sup>4</sup> См.: *Bagnoli V.* Op. cit. – P. 74.

обретению сопутствующих товаров и услуг; ценовая дискриминация; нарушение конфиденциальности<sup>1</sup>.

Условия конкуренции на цифровых рынках имеют существенные характеристики, включая такие как: 1) победитель получает все; 2) эффект сетевой выгоды; 3) двусторонность (многосторонность) рынков; 4) стремительность инноваций; 5) большие объемы инвестиций<sup>2</sup>. Тем не менее такие ценности конкуренции, как инновации, благополучие потребителей, свободный доступ к рынкам, остаются неизменными.

В. Баньоли, а также профессор права Университета Теннесси (University of Tennessee) (США) М. Штюке и экс-поверенный антимонопольного отдела Министерства юстиции США, соучредитель компании The Konkurrenz Group А. Грюнс считают, что с конкурентными преимуществами, связанными с аккумулярованием и использованием в своих интересах персональных данных, должны коррелировать специальные обязательства участников рынка<sup>3</sup>. Они ссылаются на позицию руководителя Федерального ведомства по надзору за деятельностью картелей Германии (Bundeskartellamt) А. Мундта (А. Mundt) при возбуждении производства в отношении компании Facebook. Компании, занимающие доминирующее положение на рынке, должны использовать корректные (соответствующие условиям рынка) правила пользования цифровыми продуктами и услугами, надлежаче осведомлять потребителей о характере и объеме собираемой информации.

По мнению Н. Ричардса и Д. Кинга, желание потребителя делиться информацией с коммерческими структурами должно обуславливаться (исключительно) целью повышения уровня его обслуживания<sup>4</sup>. Хорошим примером В. Баньоли считает принцип «ограничения целей» (*purpose limitation*), закрепленный в п. b ст. 6.1 Директивы Европейского парламента и Совета ЕС от 24 октября 1995 г. «О защите прав частных лиц применительно к обработке персональных данных о свободном движении таких данных»<sup>5</sup>. Согласно данному принципу персональные данные должны собираться для объявленных, явных и законных целей и в дальнейшем не могут обрабатываться каким-либо образом, несовместимым с этими целями.

---

<sup>1</sup> См.: Bagnoli V. Op. cit. – P. 74.

<sup>2</sup> См.: Ibid. – P. 75–81.

<sup>3</sup> См.: Ibid. – P. 85–86; Stucke M., Grunes A. Op. cit. – P. 3.

<sup>4</sup> См.: Richards N., King J. Op. cit. – P. 432.

<sup>5</sup> См.: Bagnoli V. Op. cit. – P. 82.

Отмечая такие вызываемые большими данными проблемы, как внеконкурентные (в традиционном понимании) слияния и поглощения, дискриминационные бизнес-стратегии, М. Штюке и А. Грюнс указывают на 10 «мифов», требующих развенчания.

Миф № 1. Право защиты частной информации и антимонопольное право имеют разные цели.

Миф № 2. Современные публичные антимонопольные органы в полной мере способны решать проблемы, создаваемые большими данными.

Миф № 3. Внутренние силы рынков способны решить проблемы, создаваемые большими данными.

Миф № 4. Компании онлайн-индустрии, основанные на использовании больших данных, не зависят от эффектов сетевой выгоды (*network effects*).

Миф № 5. Основанные на использовании больших данных рынки имеют низкий порог доступа.

Миф № 6. Информация имеет небольшое, если имеет, конкурентное значение, так как она повсеместно распространена, обладает низкой стоимостью и широкодоступна.

Миф № 7. Информация имеет небольшое, если имеет, конкурентное значение, так как доминирующие компании не способны отстранить малые компании от доступа к ключевой информации или использовать информацию для приобретения конкурентных преимуществ.

Миф № 8. Публичные антимонопольные органы не должны беспокоиться об индустрии больших данных, так как конкуренция всегда непредсказуема.

Миф № 9. Публичные антимонопольные органы не должны беспокоиться об индустрии больших данных, так как потребители только выигрывают от наличия бесплатных (цифровых) продуктов и сервисов.

Миф № 10. Потребители, приобретающие бесплатные продукты и сервисы, не могут иметь обоснованных ожиданий конфиденциальности.

**Проблема правовой защиты конфиденциальности в анализе больших данных.** По мнению А. Аллен, наблюдаемый в настоящее время агрессивный сбор коммерческими организациями информации, поддающейся анализу, представляет главную угрозу конфиденциальности частной жизни, включающей право на ограниченный доступ к персональной информации и контроль за ее оборотом. Современный сбор больших данных сравнивается с

цифровым «шпионажем», приводящим к неправоначальному раскрытию, хищениям персональных и идентификационных данных, дискриминации при трудоустройстве, найме жилища, предоставлении финансовых услуг и т.п. Борцы за гражданские права стараются привлечь внимание к тому, что большинство потребителей сервисов цифровой экономики и пользователей сети Интернет не осознают масштабов генерируемых их действиями данных, которые собираются, анализируются и используются в правительственных и коммерческих целях<sup>1</sup>.

Н. Ричардс, Д. Кинг, а также профессор права, директор лаборатории управления в сфере технологий Университета Вашингтона (University of Washington Techpolicy lab.) (США), сотрудник Центра исследования Интернета и общества Стэнфордской школы права (Stanford law school Center for Internet and society) (США) Р. Кало, равно как и многие другие ученые, указывают на угрозу конфиденциальности информации, тайне частной жизни и идентичности со стороны больших данных<sup>2</sup>. В то же время они рассчитывают на право. Нормы, регулирующие оборот в обществе частной информации, являются неотъемлемыми и неизбежными в той либо иной форме.

Задача установления баланса между необходимостью защиты частной информации, с одной стороны, и потребностью в сокрытых в ней общественных ценностях – с другой, видится крайне сложной<sup>3</sup>. Как и в других областях правового регулирования, для нормализации общественных отношений иногда бывает достаточно эффективных процедурных норм. Однако в (потенциально) меняющейся среде появляется необходимость в материально-правовых ограничениях. В попытках обуздать преимущества новых технологий, не став жертвой их потенциального вреда, осмысление этики больших данных становится первоочередной необходимостью<sup>4</sup>.

---

<sup>1</sup> См.: *Allen A.* Op. cit. – P. 71–72.

<sup>2</sup> См.: *Richards N., King J.* Op. cit. – P. 432; *Calo R.* Digital market manipulation // *The George Washington law review.* – Washington, 2014. – Vol. 82, N 4. – P. 1050–051.

<sup>3</sup> См.: *Mayer-Schönberger V., Padova Y.* Regime change? Enabling big data through Europe's new data protection regulation // *The Columbia science & technology law review.* – N.Y., 2016. – Vol. 17 (2015/2016). – P. 317.

<sup>4</sup> См.: *Richards N., King J.* Op. cit. – P. 432; *Rubenstein I.* Regulating privacy by design // *Berkeley technology law journal.* – Berkeley, 2011. – Vol. 26, N 3. – P. 1454–1456.

А. Аллен, исходя из философской точки зрения, подчеркивает важность моральной ответственности индивидов по защите информации о своей частной жизни. Автор ставит ее в один ряд с этической ответственностью надлежащего государственного управления и ведения экономической деятельности<sup>1</sup>. В то же время она приходит к выводу, что особенность угроз, создаваемых технологиями анализа больших данных, указывает на необходимость коллективных, политических подходов к самозащите. К счастью, хотя и бизнес, и правительства вовлечены в использование больших данных, проблемы неприкосновенности частной информации привлекают заслуженное внимание политиков и исследователей<sup>2</sup>.

Г. Сарфати останавливается на особом значении международного права в защите конфиденциальности в условиях глобальной цифровой экономики<sup>3</sup>.

В международном праве прав человека нет обобщенного регулирования или руководящих принципов, имеющих целью защиту конфиденциальности при использовании больших данных. ООН делает первые шаги в этом направлении, назначив Специального докладчика по вопросу о праве на неприкосновенность личной жизни со специальным мандатом в отношении защиты данного права в виртуальной среде.

Ряд негосударственных организаций (Всемирный экономический форум (WEF), Ассоциация Глобальной системы мобильной коммуникации (GSMA), Форум «Данные с полной ответственностью» (Responsible data forum)) разработали методические рекомендации, касающиеся условий отказа потребителей от получения рекламной рассылки по электронной почте, стандартов конфиденциальности, которым компании должны следовать при передаче данных в публичных целях. Согласно этим рекомендациям, госу-

---

<sup>1</sup> См.: *Allen A. Op. cit.* – P. 72–74.

<sup>2</sup> См.: *Golla S. Is data protection law growing teeth? The current lack of sanctions in data protection law and administrative fines under the GDPR // Journal of intellectual property, information technology and e-commerce law.* – Berlin; Karlsruhe, 2017. – Vol. 8, N 1. – P. 70–78; *Allen A. Op. cit.* – P. 72, 78; *César J., Debussche J. Novel EU legal requirements in Big Data security: Big Data – big security headaches? // Journal of intellectual property, information technology and e-commerce law.* – Berlin; Karlsruhe, 2016. – Vol. 8, N 1. – P. 79–88; *Van Alsenoy B. Liability under EU data protection law: From directive 95/46 to the General data protection regulation // Journal of intellectual property, information technology and e-commerce law.* – Berlin; Karlsruhe, 2016. – Vol. 7, N 3. – P. 271–288.

<sup>3</sup> См.: *Sarfaty G. Op. cit.* – P. 94.

дарствам следует принимать нормы, обеспечивающие средствами защиты лиц, чье право на охрану личной информации было нарушено, устанавливать ограничения на пользование информацией (запрет произвольного изменения цели использования данных), предупреждать неправомерное обращение к данным. Гармонизация норм обмена государствами данными с общими идентификаторами способна предупреждать трансграничные гуманитарные и эпидемиологические кризисы<sup>1</sup>.

Важное значение в нормировании использования больших данных в экономических отношениях играют «Руководящие принципы предпринимательской деятельности в аспекте прав человека: Осуществление рамок ООН, касающихся “защиты, соблюдения и средств правовой защиты”», особенно в отношении компаний сектора информационных и коммуникационных технологий. Компаниям следует оценивать вероятность и степень возможного воздействия их услуг или политики сбора и передачи информации о пользователях на права человека<sup>2</sup>.

**Заключение.** В течение последних 50 лет революция в области больших данных изменила общество. Потенциал больших данных проявился не только в открытии новых «уровней» осведомленности, но и в возникновении «источника власти», способного видеть вариативность сложных, неочевидных решений. Такие решения могут как быть эффективными с позиции стратегий хозяйственного и публичного управления, так и негативно сказываться на жизни каждого<sup>3</sup>. Угрозы ненадлежащего использования информации требуют от государств и международных институтов выработки критериев оценки выгод и рисков аккумулирования и анализа больших данных<sup>4</sup>. Коммерческие организации должны быть осмотрительными не только в отношении того, как они используют большие данные, но и в отношении влияния на рынок того, кому они их предоставляют<sup>5</sup>.

---

<sup>1</sup> См.: *Sarfaty G. Op. cit.* – P. 94–95.

<sup>2</sup> См.: *Ibid.* – P. 95–96. См. также: Право на неприкосновенность личной жизни в цифровой век: Доклад Управления Верховного комиссара ООН по правам человека. – 2014. – Режим доступа: <https://www.ohchr.org/RU/Issues/DigitalAge/Pages/DigitalAgeIndex.aspx>

<sup>3</sup> См.: *Richards N., King J. Op. cit.* – P. 432.

<sup>4</sup> См.: *Van der Sloot B., Van Schendel S. Ten questions for future regulation of Big Data: A comparative and empirical legal study // Journal of intellectual property, information technology and e-commerce law.* – Berlin; Karlsruhe, 2016. – Vol. 7, N 2. – P. 110–145; *Sarfaty G. Op. cit.* – P. 99–100.

<sup>5</sup> См.: *Bagnoli V. Op. cit.* – P. 75.

**Д.В. Красиков**

**СБОР И ИСПОЛЬЗОВАНИЕ БОЛЬШИХ ДАННЫХ  
В ПРАВООЩИТНОЙ ДЕЯТЕЛЬНОСТИ:  
ПОТЕНЦИАЛ И РИСКИ С ПОЗИЦИИ  
МЕЖДУНАРОДНОГО ПРАВА ПРАВ ЧЕЛОВЕКА  
(Обзор)**

Различные методы, инструменты и практики, связанные с использованием больших массивов данных, генерируемых пользователями персональных мобильных и иных устройств, очень эффективны при оценке рынков, прогнозировании поведения их субъектов, выявлении различных тенденций и т.д. Это заставляет задуматься о возможности использования больших данных в иных сферах общественной жизни. В контексте прав человека их сбор и использование традиционно вызывают негативные ассоциации: на ум могут прийти известные факты о «массовой слежке» государств за своими гражданами или иностранцами, о злоупотреблениях корпораций с использованием персональных данных своих клиентов и иных лиц. Вместе с тем в настоящее время все больше исследователей обращают внимание на позитивный потенциал больших данных для обеспечения прав человека.

Г.А. Сэрфэти (G.A. Sarfaty) – адъюнкт-профессор Школы права им. Питера А. Элларда из Университета Британской Колумбии (Peter A. Allard School of Law at the University of British Columbia) (Ванкувер, Канада) – утверждает, что в самой возможности использования больших данных заложен значительный потенциал для предотвращения нарушений прав человека. Вместе с тем, по утверждению автора, подход к правам человека, ориентированный на обращение к большим данным, несет в себе определенные риски.

Усилия, предпринимаемые в сфере международного обеспечения прав человека, в основном представляют собой expost-

реакцию на нарушения прав человека и сфокусированы на соблюдении международных договоров о правах человека, а предотвращение нарушений прав человека находится на периферии внимания [3, с. 75–76]. В то время как режим международного обеспечения прав человека, созданный после Второй мировой войны, формировался главным образом для предотвращения грубых нарушений прав человека, большинство существующих сегодня механизмов и процедур не преследуют эту цель: они в основном направлены на поощрение и защиту прав человека путем установления стандартов и контроля за соблюдением договоров [там же].

Новые знания и методы, основанные на работе с большими данными (а именно развитие систем сбора информации и ее анализа, а также возможности прогнозирования потенциальных вызывающих нарушений), способны помочь сместить центр внимания в сфере международного права прав человека на предотвращение нарушений [3, с. 78–82]. Возможности использования больших данных для этой цели изучены мало [3, с. 75]. И хотя основным препятствием для предотвращения нарушений прав человека остается отсутствие политической воли со стороны государств, большие данные могут восполнить информационные пробелы в тех случаях, когда необходимая воля существует [3, с. 76].

Проекты, использующие большие данные в сфере международного права прав человека, уже реализуются, однако они ограничены пространственно (имеют лишь региональный охват), предметно (посвящены отдельным видам нарушений прав человека) и во времени (как правило, носят краткосрочный характер) и существуют в основном в рамках *ad hoc* партнерств частных компаний, неправительственных организаций и научных коллективов [3, с. 83]. Подобные инициативы осуществляются в сферах международного развития, здравоохранения, предотвращения социальных конфликтов, противодействия торговли людьми и рабству и др.

Подход к правам человека, ориентированный на обращение к большим данным, сопряжен с определенными рисками. Во-первых, такой подход пригоден лишь для предотвращения отдельных нарушений прав человека; во-вторых, он сам может породить дополнительные нарушения прав человека (например, нарушение права на неприкосновенность частной жизни); а в-третьих, чрезмерное полагание на анализ больших данных может привести к «приватизации» режима международного права прав человека за счет возрастания той степени, в какой глобальное со-

общество в деле защиты прав человека вынуждено будет опираться на корпорации, а не на государства [3, с. 88–93].

Для снижения этих рисков необходимо налаживать диалог о более ответственном, этическом и легальном подходе к использованию больших данных в сфере международного обеспечения прав человека. В целях усиления ответственности в правозащитной сфере, недопущения манипуляций с данными и минимизации вреда, причиняемого уязвимым категориям населения, Г.А. Сэрфэти предлагает принять следующие меры:

- больше внимания уделять обеспечению неприкосновенности частной жизни. ООН необходимо рассмотреть вопрос о разработке руководящих правил соблюдения компаниями соответствующих стандартов [3, с. 94]. Государства должны обеспечить средства защиты в связи со сбором и использованием больших данных, ввести ограничения на использование информации, собранной для одной легитимной цели, в иных целях, а также гармонизировать законодательство о трансграничном обмене большими данными [3, с. 95]. Компании должны соблюдать соответствующие требования, в частности установленные Руководящими принципами предпринимательской деятельности в аспекте прав человека, принятыми в рамках ООН; следует устранить проблемы «нахождения общего языка» между специалистами по сбору и обработке данных и экспертами в сфере прав человека [3, с. 95–96];

- работа с большими данными должна быть обеспечена гарантиями их репрезентативности, отсутствия злоупотреблений при выборке и надлежащего толкования [3, с. 97]. Следует исходить из того, что анализ больших данных дает представление лишь о корреляционных показателях, а не о причинах или объективности тех или иных явлений, и надлежащее толкование данных требует использования также качественных методов;

- необходимо побуждать компании заниматься информационной благотворительностью (*data philanthropy*) и по своей инициативе передавать имеющиеся у них данные для использования во благо общества [3, с. 97–100].

А. Бедаши (A. Beduschi) – доцент Эксетерского университета (University of Exeter) (Эксетер, Соединенное Королевство) – исследует потенциал и риски использования больших данных в контексте продолжающегося глобального миграционного кризиса. Основной тезис, отстаиваемый автором, состоит в том, что вопреки общепринятому восприятию анализа больших данных исключительно в качестве инструмента слежения и контроля большие дан-

ные могут и должны использоваться как средство для защиты прав мигрантов. Обращение к ним может усовершенствовать процесс принятия решений, а также помочь в осуществлении мер для предотвращения смертей на море, жестокого обращения с мигрантами и торговли ими.

Мигранты и беженцы активно пользуются мобильной связью и социальными сетями для общения с членами своих семей, для поиска маршрутов передвижения, а также для получения информации о положении дел в транзитных государствах или предполагаемом принимающем государстве [1, с. 982]. При этом в настоящее время уровень развития компьютерных технологий позволяет со значительной степенью точности анализировать информацию, генерируемую мобильными устройствами и поисковыми системами, определяя конкретные модели поведения и взаимодействия людей [там же]. Такие возможности влияют на содержание международно-правовых обязательств государств в области прав человека.

А. Бедаши исходит из того, что современные технологии предоставляют государствам уникальную возможность идентифицировать лиц, нуждающихся в помощи и защите. Обращение к большим данным может стать вспомогательным средством для управления процессами в сфере международной миграции. Прогнозирование миграционных трендов, основанное на традиционных статистических инструментах, использующих административные источники, данные переписи населения и выборочные опросы, имеет серьезные недостатки: они измеряют лишь поведение, имевшее место в прошлом, и являются затратными с точки зрения ресурсов и времени. Более эффективно комбинированное использование традиционных методов и методов, основанных на анализе больших данных [1, с. 985–986].

В той степени, в какой это совместимо с международно-правовыми нормами о правах человека, государства могут иметь позитивные обязательства, требующие использования доступных им технологий для идентификации находящихся в уязвимом положении мигрантов и оказания им помощи [1, с. 991–994]. В то время как негативные обязательства по правам человека требуют от государств воздерживаться от вмешательства в реализацию прав человека, позитивные обязательства предполагают обязанность принять все необходимые меры для уважения прав человека.

Для предотвращения смертности мигрантов на море (эта ситуация достигла угрожающих масштабов), по мнению автора, государства обязаны соблюдать и обеспечивать их право на жизнь в

пределах своей юрисдикции, и в этих целях использовать новейшие технологические средства при условии, что:

1) они имеют возможность делать это, о чем может свидетельствовать практика использования этих средств для иных целей (например, практика массового слежения в развитых государствах);

2) применение этих методов не налагает непосильного или непропорционального бремени на них (обязанность возникает в том случае, если государства обладают информацией о соответствующих рисках);

3) есть основания считать, что обращение к таким средствам является наиболее эффективным способом соблюдения соответствующего обязательства [1, с. 999–1002].

Аналогичные позитивные обязательства государства имеют в отношении запретов жестокого обращения и торговли людьми, к нарушениям которых мигранты и их дети особенно уязвимы [1, с. 1004–1009]. Анализ больших данных, в частности аналитические методы прогнозирования, может использоваться для выявления и картирования зон риска, для отслеживания используемых торговцами людьми мобильных телефонов и Интернета для поиска жертв, в целях рекламы своих услуг и проведения переговоров.

Вместе с тем при отсутствии необходимых ограничений использование этих методов может порождать серьезные риски для безопасности мигрантов, а также представлять собой необоснованное вмешательство в их право на частную жизнь и на защиту персональных данных. Необходимо выработать гарантии того, что использование этих методов не будет повышать риски выдворения мигрантов и беженцев и способствовать установлению чрезмерно жестких режимов пересечения границ государств. Кроме того, криминальные группы могут использовать большие данные для поиска мест сосредоточения мигрантов, выбираемых ими маршрутов или для установления местонахождения конкретных лиц. Наличие этих рисков свидетельствует о необходимости усиления контроля за сбором, хранением и обработкой больших данных в сфере миграции [1, с. 1015].

М. Лэтонино (M. Latonero) – ведущий научный сотрудник программы «Информация и права человека» (Data & Human Rights) научно-исследовательского института «Информация и общество» (Data & Society) (Нью-Йорк, США) – выявляет и оценивает риски, которые порождает использование аналитики больших данных в правозащитной деятельности. При этом автор акцентирует внимание на соответствующих угрозах для частной жизни людей. По его

убеждению, обработка и использование больших данных в контексте мониторинга соблюдения прав человека могут порождать этические и иные проблемы, аналогичные возникающим в связи с осуществлением неправомерной «слежки» со стороны государств или корпораций [2, с. 149]. Это связано с тем, что методы, основанные на сборе и анализе больших данных, могут применяться без ведома субъектов соответствующих данных, в отсутствие их согласия либо без должного понимания ими характера действий, производимых с их данными.

В исследованиях, посвященных использованию больших данных для целей правозащитной деятельности (например, данных о жертвах среди гражданского населения в ходе вооруженных конфликтов или гуманитарных катастроф), нередко поднимаются проблемы их репрезентативности и достоверности, однако вопросы неприкосновенности частной жизни остаются за пределами внимания [2, с. 153].

Основной сложностью при оценке того, насколько тот или иной метод работы с большими данными создает угрозу для частной жизни, является то, что ни один из них не применяется изолированно: аналитика больших данных представляет собой комбинацию социотехнических процессов, которые можно разделить на две категории – сбор и использование [2, с. 151–153]. Сбор включает в себя поиск, выявление и мониторинг информации, ее вычленение, классификацию, индексирование, сохранение и хранение. Использование означает анализ больших данных, осуществление доступа к ним и их распространение. При этом каждый из таких процессов может создавать риски для частной жизни человека.

По мнению автора, при определении допустимости использования больших данных для целей обеспечения прав человека ключевым фактором является степень безотлагательности вмешательства, необходимого для достижения правомерной цели. Учитывая, что право на частную жизнь не является абсолютным, оно может быть ограничено при определенных условиях, однако степень вмешательства в данное право должна быть пропорциональной соответствующей цели. При повышении степени угрозы для безопасности человека (например, неминуемой угрозы смерти) снижается степень обеспокоенности по поводу неприкосновенности его частной жизни [2, с. 154–157].

Иные подходы должны применяться в отношении сбора и использования информации для продолжительного мониторинга в сфере прав человека. Открытым является вопрос о том, насколько

неопределенные преимущества долгосрочного правозащитного контроля «перевешивают» вполне конкретные риски для частной жизни, которые несут в себе сбор и использование больших данных. Например, каковы пределы допустимости сбора информации неправительственными организациями в сфере прав человека? Сам факт наличия у них «лучших побуждений» не является достаточной гарантией неприкосновенности частной жизни или оправданием вмешательства в нее. Данная деятельность должна регулироваться путем как распространения на нее действующих принципов и норм, относящихся к защите персональных данных, так и выработки специальных правил, учитывающих особенности современных технологических методов сбора и обработки информации [2, с. 159–161].

### **Список литературы**

1. Beduschi A. The big data of international migration: Opportunities and challenges for states under international human rights law // *Georgetown journal of international law*. – Wash., 2018. – Vol. 49. – P. 981–1017.
2. Latonero M. Big data analytics and human rights: Privacy considerations in context // *New technologies for human rights law and practice* / Ed. by M.K. Land. – Cambridge: Cambridge univ. press, 2018. – P. 149–161.
3. Sarfaty G.A. Can big data revolutionize international human rights law // *University of Pennsylvania journal of international law*. – Philadelphia, 2017. – Vol. 39, N 1. – P. 73–101.

**Галиндо Ф., Марко Х.Г.**  
**СВОБОДА И ИНТЕРНЕТ:**  
**НАДЕЛЕНИЕ ГРАЖДАН ПОЛНОМОЧИЯМИ**  
**И РЕШЕНИЕ ПРОБЛЕМЫ ДЕФИЦИТА ПРОЗРАЧНОСТИ**  
**ПОИСКОВЫХ СИСТЕМ**  
**(Реферат)**

Ref. ad op.: Galindo F., Marco J.G. Freedom and the Internet:  
Empowering citizens and addressing the transparency gap in search  
engines // European journal of law and technology. – Belfast, 2017. –  
Vol. 8, N 2. – P. 1–18. – Mode of access: <http://ejlt.org/article/view/476>

Авторы статьи – профессор философии права Университета Сарагосы (University of Zaragoza) Фернандо Галиндо и профессор документалистики Университета Сарагосы Хавьер Гарсия Марко – исследуют два ключевых вопроса: как соотносятся свобода воли и Интернет и может ли Интернет способствовать упрочению свободы выражения мнения и обмена информацией без ограничения свободы воли его пользователей? Их вывод разочаровывает: наблюдается явное противоречие между правом на свободу и практикой использования сети Интернет.

Философское понимание категории «свобода» содержится в трудах мыслителей древних времен: Р. Декарт рассматривал свободу как свободу «мышления», зависящую от личности, а не от внешней власти; И. Кант понимал свободу человека как свойство разумной воли и способность «не подчиняться никакому закону, кроме закона, на который гражданин дал свое согласие»; Ж.-Ж. Руссо утверждал, что «человек рождается свободным, но везде он в оковах» [цит. по: с. 3].

Свобода определяется авторами как право на автономию воли в повседневной жизни граждан или членов общественных и поли-

тических организаций, уважающих использование этого права. При этом они исходят из положения, закрепленного в преамбуле Хартии ЕС об основных правах 2000 г.: «Осознающий свое духовное и нравственное историческое наследие Союз базируется на всеобщих и нераздельных ценностях – человеческом достоинстве, свободе, равенстве и солидарности; он опирается на принцип демократии и принцип правового государства. Он помещает человеческую личность в центр своей деятельности посредством введения гражданства Союза и создания пространства свободы, безопасности и справедливости».

Опираясь на этот принцип, авторы рассматривают взаимосвязь свободы и Интернета. Вместе с тем, подчеркивают они, Интернет – это не просто сеть компьютеров и коммуникаций, расширяющая знания. Он становится кибернетическим инструментом контроля посредством принятия и исполнения приказов, издаваемых людьми, или инструкций и директив, запрограммированных на это. Интернет также является инструментом политического контроля: он позволяет осуществлять дистанционный мониторинг с помощью камер и датчиков деятельности и процессов, происходящих в любой социальной среде. То есть в самом Интернете заложены большие ресурсы ограничения свободы. Авторы выделяют и рассматривают три вида таких ограничений: 1) свободы выражения мнения; 2) экономических; 3) политических свобод.

*Феномен ограничения свободы выражения мнений*, порождаемый Интернетом, настолько очевиден, что признается даже теми, кто считает, что Интернет способствует распространению мнений гораздо большего числа людей, чем традиционные средства массовой информации. Однако связь между свободой выражения мнений и Интернетом относительна: «Защита некоторых прав человека или свобод часто оказывает прямое и непосредственное воздействие на другие права и свободы» [с. 4]. То есть сохранение одной свободы может ограничить другую. Сбалансированность этих противоречивых ценностей и интересов, как полагают авторы, может быть достигнута только путем нормативно-правового регулирования и общественного договора.

*Экономические ограничения* – одно из основных ограничений свободы действий через Интернет: если человек не имеет ресурсов для доступа в Интернет, то, очевидно, он не в состоянии предпринять какие-либо действия. Поэтому Интернет – это не то же самое, что свобода действий и общения для всех, это право может быть ограничено теми, кто обладает достаточными ресурсами [с. 4].

Еще одним базовым ограничением является наличие необходимых технических знаний и навыков для пользования Интернетом. Однако это ограничение все менее и менее актуально, поскольку образование и профессиональная подготовка вовлекают все больше людей в жизнь общества, основанного на знаниях.

*Ограничение политических свобод* легко достигается с помощью контроля со стороны государственных и частных органов за использованием Интернета. К примеру, Китай и Турция в значительной мере регулируют доступ к определенным интернет-ресурсам, включая и ту информацию, которая выкладывается в Сеть, и ту, к которой могут получить доступ пользователи [с. 4].

И еще один серьезный аспект: компании, предлагающие интернет-услуги и контент, могут скрытно управлять поведением пользователей, тем самым они уменьшают их реальную свободу. Это, например, происходит, когда предоставляется только часть запрашиваемой информации или используются методы возбуждения интереса пользователя к определенному информационному контенту. Данная проблема, считают авторы, наиболее актуальна для поисковых систем. Решить ее можно посредством введения в интернет-компаниях кодексов этики.

Таким образом, Интернет позволяет осуществлять контроль и тем самым ограничивать свободу воли, даже если этот принцип соблюдается пользователями. Бесспорно, через социальные сети, такие как Facebook, Twitter, Instagram, LinkedIn, YouTube и др., люди могут «свободно» выражать свое мнение, отправлять видео и информацию. Однако это не отменяет того факта, что те, кто отвечает за Интернет, могут изменять передачу этой информации, у них есть ресурсы для вмешательства в ее содержание. Это приводит к парадоксу, который трудно объяснить: Интернет может как помогать, так и препятствовать упрочению свободы. В связи с этим он подпадает под действие ряда сложных правил, касающихся хранения информации и доступа к ней, а также норм и принципов, поощряющих свободы и запрещающих злоупотребления.

*Юридические рамки хранения, получения и поиска информации.* До Интернета компьютеры предназначались для хранения и обработки информации в произвольном порядке или с использованием закрытых коммуникационных сетей. В связи с развитием информационно-коммуникационных технологий в бизнесе, торговле и государственном управлении было замечено, что право граждан на свободу нарушается владельцами компьютерных систем, особенно теми, кто отвечает за управление базами данных, содержащими

персональную информацию для конкретных целей: приобретение продуктов, предоставление медицинских услуг, страховые полисы и т.д. Вскоре было признано, что личная свобода владельцев персональных данных, которые дали согласие на использование их информации в конкретных целях, является уязвимой. Персональные данные могут использоваться получателями и лицами, ответственными за базы данных, для других целей без получения дополнительного согласия, т.е. вне реализации свободы воли соответствующих лиц.

Практика судебной защиты персональных данных в США отличается от подобной практики в Европе. Первое подобное дело рассматривалось в США. Оно было разрешено на основании прецедента, касавшегося вмешательства в частную жизнь, относящегося к началу XX в. В Европе такой подход показался довольно странным, поскольку судья вынес приговор, который не был основан на каком-либо действующем законе. Однако эти осложнения не были слишком проблематичными. Для судебного вмешательства достаточно было предыдущего приговора по аналогичному делу. В данном случае прецедент квалифицировал неприкосновенность частной жизни как индивидуальное право, которое было нарушено.

Учитывая те же потенциальные нарушения личных свобод и следуя стандартам, установленным демократическими конституциями, европейские суды пошли другим путем. В Евросоюзе были приняты специальные законы, направленные на защиту персональных данных. В каждой стране ЕС созданы независимые административные органы – Управления по защите данных (Data Protection Office), которым органы и структуры, отвечающие за программирование, информационные ресурсы и управление ими, должны сообщать об общих характеристиках своих баз данных и способах их использования. В то же время были введены процедуры, позволяющие заинтересованным сторонам консультироваться по поводу их личных данных и проверять, используется ли персональная информация в соответствии с явно выраженным согласием их собственника.

Уровень защиты информации в Европе растет. Как отмечают авторы, в США подобная законодательно установленная система защиты отсутствует, все находится в руках судьи, принимающего решение о наличии / отсутствии нарушения прав человека на основе прецедента. Столь разный подход к регулированию свободы в Интернете неизбежно сказывается и на развитии технологий, приводит к конфликту между разработанными в США программами (использующими такие механизмы сбора информации о пользователе, как

«куки» (cookies)) и регулированием рассматриваемых прав в Европе. Как отмечают авторы, этот факт наносит ущерб компаниям, которые адаптируют программы к европейским потребностям, не зная, как избежать санкций со стороны пользователей или других компаний за несоблюдение законодательства о защите персональных данных. Это само по себе предполагает ограничение свободы, поскольку несоблюдение регламента или принципа информационного самоопределения пользователей и обращение с персональными данными без явно выраженного согласия являются нарушениями действующих европейских регламентов [с. 7].

С этой точки зрения культурная традиция США в отношении Интернета может вызвать проблемы в других странах, имеющих иной подход к вопросу ограничений свободы и прав. Именно в Европе впервые было введено законодательство о защите свободы и использовании персональных данных компаниями и организациями, работающими с информационно-коммуникативными технологиями. Некоторые практические следствия из этих правил уже упоминались, но есть и другие, например «право на забвение», или удаление информации, выполнять которое Европейским судом в 2014 г. были призваны Google и аналогичные компании [с. 8]. Такого рода правило о защите данных было предусмотрено также Европейским парламентом и Европейским советом в 2016 г.<sup>1</sup>

Еще одним примером является право на защиту данных, которые подлежат регулированию, в Испании и в других европейских странах. Компаниям сложно применять это право на практике при разработке приложений, так как необходимо учитывать мнение учреждений, занимающихся защитой персональных данных.

Трудности очевидны: те, кто получает доступ к информации в Интернете, нередко используют ежедневный ресурс, который близок к незаконному. Применение cookies-программ, собирающих личную информацию о пользователях веб-сайтов, характерно для поисковых систем, радиостанций, телевидения и даже госу-

---

<sup>1</sup> См.: Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) = Регламент (ЕС) 2016/679 Европейского парламента и Совета от 27 апреля 2016 г. О защите физических лиц в отношении обработки персональных данных и о свободном перемещении таких данных, а также об отмене директивы 95/46/EC (Общий регламент о защите данных). – Mode of access: <https://publications.europa.eu/en/publication-detail/-/publication/3e485e15-11bd-11e6-ba9a-01aa75ed71a1>

дарственных органов. Они, как правило, помогают освоить новые методы ведения бизнеса (с рекламой и базовым размещением продукта одновременно с отправкой сообщений пользователям через социальные сети – Facebook, Twitter и др.). С помощью cookies все виды информации собираются и используются в коммерческих целях. В Европе существуют правила, регулирующие эти злоупотребления свободой.

*Открытые данные.* В данном разделе статьи авторы пытаются оспорить аргументы тех специалистов, которые считают, что информация и данные, доступные в Сети, являются «открытыми данными», должны использоваться свободно и вне прав собственности (интеллектуальной / промышленной / личной), что понятия «Интернет» и «свобода» тождественны. По их мнению, эта идея противоречит концепции свободы, которая прописана в Хартии ЕС об основных правах и в конституциях всех демократических стран. Согласно Хартии, свобода – это право на свободу и безопасность: «Каждый человек имеет право на свободу и личную неприкосновенность» (ст. 6); «Каждый человек имеет право на уважение своей частной и семейной жизни, своего жилья и своих коммуникаций» (ст. 7); «1. Каждый человек имеет право на защиту относящихся к нему данных личного характера. 2. Обработка подобных данных должна производиться без манипуляций, в четко определенных целях, с согласия заинтересованного лица либо при наличии других правомерных оснований, предусмотренных законом. Каждый человек имеет право на получение доступа к собранным в отношении него данным и право на устранение в них ошибок. 3. Соблюдение этих правил подлежит контролю со стороны независимого органа» (ст. 8). Иными словами, при осуществлении свободы необходимо уважать свободу и права других лиц.

Концепция открытых данных ограничивает корреляцию между свободой и Интернетом. В данном случае Интернет дает ошибочное понимание свободы мнений. Это ограничение самой свободы, потому что те, кто выступает за нее, нарушают права тех, кто публикует информацию в Интернете, позволяя использовать ее в целях, которые не были прямо согласованы.

Проблема увеличивается, когда открытые данные доступны через поисковые системы, такие как Google, которые не являются прозрачными и состоят из математических формул, неизвестных пользователям. Европейское законодательство может быть нарушено: 1) техническими специалистами, когда те думают, что данные открыты (хотя это не так), и делают их доступными через

Сеть; 2) учреждениями, ответственными за обеспечение законности открытых данных; 3) компаниями, владеющими поисковыми системами и алгоритмом доступа к информации, при несоблюдении требований закона; 4) лицами, получившими доступ к информации и использующими ее без учета ее законности [с. 9].

*Открытое правительство.* Специалисты, которые поддерживают движение «открытые данные», также выступают за то, чтобы государственные органы предоставляли данные и информацию, связанную с их функционированием и принятием решений. Эта доктрина названа «открытое правительство». Они утверждают, что открытые данные и открытое правительство повышают эффективность Интернета, поскольку правительства должны публиковать информацию о своих услугах: пособия по безработице; медицинское обслуживание; выдача лицензий на строительство; статистическая информация; услуги государственной регистрации; управление бюджетом; государственные расходы; заработная плата гражданских служащих и т.д.

На основе этих данных государственные органы принимают решения. Технические специалисты считают, что эти данные должны быть в свободном доступе для граждан и компаний. С помощью этих данных специалисты могут разработать новые программы и приложения, однако часто их цели сильно отличаются от целей, для которых данные собирались.

В определенной степени этот процесс уже идет, в соответствии с демократическими обязательствами правительств и администраций эффективно управлять государственными средствами и обеспечивать большую транспарентность. Кроме того, по крайней мере в Европе, правительства принимают меры предосторожности при публикации данных, защищая права граждан, которые дали согласие на использование их данных в конкретных целях. В противном случае правительства, администрации и гражданские служащие могут быть признаны виновными в нарушении свободы и прав личности.

Поэтому логично, что использование открытых данных программистами (специалистами) компаний не предполагает каких-либо затрат для административных органов, или если существуют какие-либо расходы, связанные с созданием программ и баз данных и защитой персональных данных граждан, эти расходы будут покрываться за счет организаций или физических лиц, которые хотят использовать информацию.

*Поисковые системы и тезаурусы.* Авторы утверждают, что ограничения, налагаемые на свободу использования Интернета, не смягчаются простым соблюдением действующих правил в отношении информации, которая предоставляется пользователям. Существуют ограничения, связанные с функционированием методов доступа; программ и приложений общего пользования, поисковых систем, таких как Google, нарушающих закон из-за отсутствия прозрачности в методах их работы. Основное внимание авторы уделяют последствиям применения этих методов. Вместе с тем существуют другие пути доступа к информации, например тезаурусы, которые могут дополнять Google и аналогичные поисковые системы способами, более совместимыми с правами граждан. В отличие от непрозрачного функционирования поисковых систем, стандарты тезауруса (ISO 25964–1:2011, ISO 25964–2:2013) могут, по мнению авторов, обеспечить прозрачность процедуры поиска. Такой подход, считают они, может способствовать преодолению разрыва между Интернетом и действующим законодательством, преодолению препятствий на пути к свободе, присущей Интернету.

Таким образом, исследования показали, что взаимоотношения между Интернетом и свободой воли сопряжены с проблемами, которые сложно преодолеть. Трудности, возникающие в результате доступа в поисковые системы, прежде всего связаны с тем, что используемые алгоритмы не являются прозрачными. В связи с этим авторы предлагают в конкретном случае поиска информации использовать тезаурус как метод представления информации способом, совместимым (с точки зрения пользователя) с требованиями законодательства. По их мнению, модель тезауруса подходит для того, чтобы в простой схеме увидеть различные синонимы и многозначные термины. Таким образом, сами пользователи могут решить, является ли поиск оптимальным, а термины релевантными.

В идеале поисковые системы могли бы предоставить пользователям инструменты для включения этих терминов в поиск, чтобы улучшить его. Эти новые функции не должны нарушать простоту интерфейсов поисковых систем. Использование модели тезауруса для повышения прозрачности поисковых систем может стать решением многих проблем: для большого числа пользователей Интернет является эффективным поставщиком информации. Безусловно, пользователи имеют определенную степень свободы, ведь в принципе они могут выбирать между несколькими поисковыми системами и интернет-порталами. Но это не настоящая сво-

бода: появляется все больше и больше данных, использование которых ограничено из-за проблем доступа, требований закона, в том числе получения информированного согласия. Кроме того, индустрия поисковых систем очень развита, и большая прозрачность поиска обеспечивается только в экспериментальных поисковых системах с очень ограниченным охватом в Интернете.

В этом контексте реализация решения, построенного на модели тезауруса, считают авторы, может состояться, поскольку поисковые системы все больше полагаются на семантические графики, например предоставляемые викиданными. В результате поисковые системы станут более прозрачными, расширятся возможности пользователей, а общественное и личное мнение избавятся от манипуляций, осуществляемых с помощью непрозрачных автоматических процедур, которые представляют большую опасность для свободы и демократии [с. 14].

*О.Л. Алферов*

**Кокколи Я.**

**ВЫЗОВЫ НОВЫХ ТЕХНОЛОГИЙ  
В ОБЛАСТИ РЕАЛИЗАЦИИ ПРАВ ЧЕЛОВЕКА:  
АНАЛИЗ НЕКОТОРЫХ КРИТИЧЕСКИХ ВОПРОСОВ  
В ЦИФРОВУЮ ЭПОХУ  
(Реферат)**

Ref. ad op.: Coccoli J. The challenges of new technologies in the implementation of human rights: An analysis of some critical issues in the digital era // Peace human rights governance (PHRG). – Venice, 2017. – Vol. 1, N 2. – P. 223–250.

В статье Якопо Кокколи (Папский лютеранский университет, Италия) исследуются проблемы влияния цифровых технологий на права человека. С одной стороны, эти технологии усиливают права, например благодаря возможности реализации права на свободу выражения своего мнения, с другой – подвергают их беспрецедентным рискам в случае фильтрации информации и / или блокирования веб-доступа [с. 223].

Любой анализ влияния новых технологий на права человека является чрезвычайно сложным и требует предварительного учета двух аспектов: первый обусловлен эволюционным промежутком времени, который разделяет достижения технического прогресса и их юридическое оформление, второй – отражает тенденцию цифрового развития на международном уровне.

В статье рассматриваются два ключевых вопроса: 1) реинтерпретация традиционных прав человека в свете технологических инноваций; 2) новые виды прав человека, которые могут быть идентифицированы как *suigeneris* поколения цифровых прав [с. 226].

В первой части статьи автор исследует *развитие традиционных прав человека в цифровом пространстве*. Пересмотр преж-

него мнения о традиционных правах человека в свете научно-технических достижений рассматривается автором как решающий шаг для защиты достоинства человека в цифровой век, толкования свободы мнения в свете цифровых инноваций как главного средства, позволяющего людям участвовать в развитии культуры, которую они хотят видеть более демократичной. Право на свободу выражения мнения – в первую очередь интерактивное право; оно возможно среди общающихся людей, как действующих и говорящих и как слушающих одновременно. Это право основано на способности человека владеть и пользоваться элементами культуры, критикуя ее или предлагая что-то новое [с. 227].

Идея о том, что отношения между свободой мнения и информационно-коммуникативными технологиями (ИКТ) приносят бесспорную пользу для общества, широко признана. Первый положительный опыт накоплен благодаря действию в Сети принципа прозрачности, который стал эффективным правоохранительным средством борьбы с коррупцией, и главное – ИКТ делают явными любые нарушения прав человека и усиливают все формы взаимодействия между людьми, облегчая обмен идеями и мнениями. Тем самым, подчеркивает автор, они способствуют успеху общества в целом [с. 228].

Тем не менее цифровые технологии подвергают права человека многим рискам. В связи с этим возникает ряд проблем, требующих внимания и законодателей, и исследователей. Первая: в достаточной ли степени сегодняшняя нормативная база о правах человека гарантирует свободу выражения мнения в юридических ситуациях, созданных новыми технологиями. Вторая касается того, что право искать, получать и передавать информацию, кажется, становится нестабильным в цифровую эру из-за угрозы вмешательства правительств посредством введения различных ограничений. Ряд государств, несмотря на наличие в их конституциях глав, посвященных основным правам человека, по-прежнему проводят пагубную политику, направленную на предотвращение широкого использования ИКТ гражданами. В качестве примера автор приводит Китай, который владеет одной из самых сложных и широко распространенных систем управления информацией в Сети и механизмом фильтрации контента, известного под названием «Great Fire wall». Китайское правительство прибегает к карательным мерам, основанным на многочисленных средствах цензуры: сотрудники полиции блокируют веб-сайты и следят за доступом физических лиц в Интернет, платные обозреватели постоянно

отслеживают любую критику в адрес правительства, распространение новостей из других стран должно быть одобрено информационным агентством Госсовета. Таким образом, с одной стороны, в Китае действует глобальная сеть Интернет и, соответственно, извлекается выгода из международной онлайн-торговли, с другой – ограничивается свободный доступ пользователей к слишком большому количеству информации [с. 229].

Автор приводит и другие примеры: принятый в Российской Федерации закон налагает на интернет-провайдеров обязанность расшифровывать коммуникации. Это, по его мнению, повлечет появление «черных» ходов шифрования, которые, вероятно, непропорционально подорвут безопасность всех пользователей. Соединенное Королевство и Франция также предложили наделить сотрудников правоохранительных органов и разведки полномочиями потребовать, чтобы компании предоставляли им доступ к зашифрованным сообщениям их пользователей. В Бразилии конституционное право запрещает анонимность в онлайн и офлайн полностью [с. 234].

Согласно международному и национальному законодательству ряда стран мира, право на свободу выражения мнения может быть ограничено только в том случае, если это ограничение является необходимым и соразмерным для защиты конкретных законных целей. Это – непреложное условие. Свобода выражения мнений, как заявлено в резолюции Совета по правам человека ООН 20/8 «Поощрение и защита всех прав человека, гражданских, политических, экономических, социальных и культурных прав, включая право на развитие», является «одной из важнейших основ демократического общества и одним из основных условий его развития». Свободный доступ к технологиям рассматривается ООН как средство построения инклюзивных обществ знаний и демократий и содействия межкультурному диалогу, миру и эффективному управлению [цит. по: с. 231].

Цифровые технологии не только ввели новые способы реализации традиционных прав человека, но и способствовали появлению таких прав и свобод, которые еще требуют конституционного закрепления и дальнейшего регулирования со стороны правительств. В связи с этим во *второй части статьи «Защита человеческого достоинства в современных обществах: Новое поколение цифровых прав человека»* Я. Кокколи рассматривает гипотезу о признании существования новых прав человека по мере развития инноваций, а также тех специфических прав, которые считаются важнейшими для человеческого достоинства в цифровую эпоху. Речь идет о праве на

защиту персональных данных, праве на использование достижений науки и техники, праве на доступ в Интернет.

Рассмотрение проблем реализации *права на защиту персональных данных* автор начинает с анализа опыта Европейского союза в этой сфере. В апреле 2016 г. Европейским парламентом и Советом был принят Регламент (ЕС) 2016/679 «О защите физических лиц в отношении обработки персональных данных и о свободном перемещении таких данных и отмене Директивы 95/46/ЕС (Общие положения о защите данных)» (вступил в силу 25 мая 2018 г.). Принятие такого нормативного акта в рамках Евросоюза, по мнению Я. Кокколи, будет гарантировать единообразное и последовательное регулирование защиты персональных данных. Кроме того, важное нововведение связано с кодификацией права на исправление персональных данных и права на их удаление по запросу владельца.

Вторым ключевым актом, свидетельствующим об актуальности защиты данных в Евросоюзе, считается Соглашение ЕС – США о защите персональных данных, утвержденное Европейским парламентом. Соглашение предусматривает обязательные стандарты защиты персональных данных, которыми обмениваются между собой полиция ЕС и США и судебные органы для предупреждения, раскрытия, расследования и судебного преследования преступлений, включая терроризм [с. 238].

Далее автор рассматривает два аспекта *права использовать достижения науки и техники*: во-первых, право отдельных лиц на пользование достижениями науки и техники и право доступа к научно-техническому прогрессу, во-вторых, право ученых на свободное ведение научной деятельности, которое понимается как право или свобода оценивать и выбирать предпочтительный путь научного и технологического развития. Кроме того, среди элементов права на пользование результатами научно-технического прогресса выделяются защита от возможного вредного воздействия науки, а также международное сотрудничество в этой области.

В связи с этим автор описывает вклад ООН в разработку «трехсторонней теории обязательств государств». Данная теория классифицирует позитивные и негативные обязательства по трем различным категориям, а именно: обязанности *уважать, защищать и выполнять* [с. 240].

*Обязательство уважать* означает, что государства должны воздерживаться от вмешательства и, следовательно, не принимать меры, которые приводят к нарушению этого права. Такие обяза-

тельства должны также включать автономию высшего образования, его институтов и свободу преподавателей и студентов, в частности, выражать мнения об учреждении или системе, в которых они работают, без дискриминации или страха наказания.

*Обязательство защищать* требует от государства мер по обеспечению законных интересов физических лиц и защите их прав. В обязанности правительств должны входить: установление и охрана баланса соответствующих интересов в целях научного прогресса и в соответствии с правами человека; принятие мер по предотвращению и недопущению использования третьими сторонами достижений науки и техники в ущерб правам человека; защита прав лиц, на которых распространяется научно-исследовательская деятельность.

*Выполнение* включает такие обязанности, как: облегчать доступ к научной информации; обеспечивать качественное научное образование и возможность участия общественности в принятии решений, связанных с наукой; содействовать международному сотрудничеству; создавать учреждения, способствующие распространению научных знаний; развивать мониторинг и информировать общественность о потенциально вредном воздействии науки и техники.

Анализ реализации права на использование достижений науки и техники показывает, что это право не является достаточно известным правом человека и не может считаться наиболее убедительным в обеспечении человеческого достоинства. Однако, как замечает автор, его значение, несомненно, растет, и государства обязаны соблюдать это право в рамках международных норм в области прав человека [с. 241].

Следующий аспект осуществления права на достоинство, рассматриваемый в статье, – *проблемы реализации права на доступ в Интернет*. Автор замечает, что такая универсальная услуга, как доступ всем и ко всему в Интернете, с одной стороны, отсутствует ввиду того, что пока нет возможности разрешить такой общий доступ, с другой – предоставляется, но крайне ограничено.

В настоящее время дебаты в основном ведутся о цифровом неравенстве. Цифровое неравенство имеет два измерения: географическое (городские и сельские районы, развитые и слаборазвитые регионы) и социальное (цифровая грамотность, доступ для уязвимых групп или язык). Последнее может быть обусловлено физическими препятствиями, инфраструктурной неэффективностью, несовершенством информационных технологий и др.

В специальном докладе Совета по правам человека ООН по вопросу о поощрении и защите права на свободу мнений и их сво-

бодное выражение (16 мая 2011) отмечено, что доступ к Интернету имеет два параллельных компонента. Первый – доступ к онлайн-контенту должен быть обеспечен без каких-либо ограничений, кроме нескольких случаев, предусмотренных международным правом о правах человека. Второй – наличие в первую очередь необходимой инфраструктуры и информационных коммуникационных технологий, таких как кабели, модемы, компьютеры и программное обеспечение [цит. по: с. 242].

Исследование показало, что только у трех юрисдикций есть уже конституционно закрепленное обязательство государства гарантировать возможность доступа к информации в электронном виде. Так, Конституция Греции в ст. 5а предусматривает: «Все люди имеют право на участие в информационном обществе. Облегчение доступа к переданной в электронном виде информации, так же как и обмен информацией и ее распространение, являются обязательствами государства всегда при соблюдении гарантий конфиденциальности, сохранности персональных данных и корреспонденции». В решении N 580 (2009) французский Конституционный совет указал, что доступ к Интернету нужно рассматривать как право человека на свободу выражения мнения. Аналогичное решение N 12 790 (2010) принято Конституционным судом Коста-Рики, заявившим, что обязанностью органов государственной власти является поощрение и обеспечение во благо граждан доступа к новым технологиям.

Другие страны пытаются обеспечить доступ в Интернет, закрепляя это право в обычном законодательстве. Так, в Эстонии принят закон, объявляющий доступ в Интернет правом человека. Испания и Финляндия рассматривают такой доступ как универсальную услугу. В Нидерландах принят закон о сетевом нейтралитете, который впервые среди законов, действующих в других странах, закрепил революционный, по мнению автора, принцип, требующий от интернет-провайдеров управления всем законным контентом нейтральным образом [с. 244].

И даже когда доступ в Интернет уже гарантирован, могут возникнуть новые и спорные вопросы. Основная проблема, по мнению автора, связана с риском ограничений, который характеризует свободное функционирование прав человека: есть противоречие между идеей о том, что Интернет – открытое пространство без инструкций, и озабоченностью вмешательством государственной власти. В этом контексте самая большая проблема закона состоит в том, чтобы найти баланс между безопасностью пользователей и уважением их свободы, не прибегая к несоразмерным

мерам, как то: закрытие социальных сетей, блокировка веб-сайтов, приостановка услуг мобильной связи и др. Например, в 2014 г. перед выборами Турция закрыла доступ к Twitter и YouTube. Такие ограничения, с точки зрения Управления Верховного комиссара ООН по правам человека, могут подорвать законность избирательного процесса и подвергнуть сомнению гарантии бесплатного и справедливого осуществления гражданских и политических прав людей [с. 242].

Любые проблемы, связанные с цифровым доступом (политика цензуры, цифровое неравенство, универсальная возможность общения), могут быть решены путем формализации этого права человека, его закрепления в национальном законодательстве. Международному сообществу предлагается сделать первый шаг: содействовать процессам регулирования, которые могли бы развиваться так же быстро, как и внедрение цифровых технологий, в целях предотвращения конфликтов и обеспечения соответствующего баланса между правами человека [с. 245–246].

*Е.В. Алферова*

**Хинг А.**

**НЕКОТОРЫЕ РАЗМЫШЛЕНИЯ О ДОСТОИНСТВЕ  
КАК АЛЬТЕРНАТИВНОЙ ПРАВОВОЙ КОНЦЕПЦИИ  
В РЕГУЛИРОВАНИИ ЗАЩИТЫ ДАННЫХ  
(Реферат)**

Ref. ad op.: Hing A. Some reflections on dignity as an alternative legal concept in data protection regulation // German law journal. – Frankfurt a. M., 2018. – Vol. 19, N 5. – P. 1269–1290.

По мере роста использования Интернета и онлайн-платформ увеличиваются масштабы сбора и обработки персональных данных. Все чаще высказываются опасения по поводу превращения Сети в настоящую машину для профилирования и размещения рекламы, а также использования ее в качестве основного инструмента для наблюдения и контроля. Откровения Сноудена, злоупотребления Cambridge Analytica в целях политического микротаргетинга лишь усилили обеспокоенность со стороны общественности. Автор статьи Анна Хинг – доцент кафедры информационного права юридического факультета Амстердамского свободного университета – считает, что современное правовое регулирование не справляется с решением проблем информационной безопасности. Необходимо иное, сочетающее этические и правовые элементы обоснование недопустимости современных способов сбора, обработки и хранения данных. Таким обоснованием автору видится концепция человеческого достоинства.

Современные технологии позволяют обрабатывать, анализировать, покупать и продавать, использовать, передавать и распространять большое количество информации об отдельных людях. Она стала особым источником дохода и власти как для частных, так и для публичных субъектов. Кто-то объясняет это наступлением эры «но-

вой экономики». Кто-то же видит в этом серьезную угрозу для конфиденциальности в Сети, опасность возникновения новой религии «dataism», наступление информационного капитализма [с. 1270].

Основные опасения сводятся к тому, что в настоящее время рынок, полностью завязанный на торговле персональными данными, действительно существует и развивается в экстремальных масштабах. Личная информация, которая собирается в основном в целях извлечения экономической выгоды, впоследствии используется и для других целей, таких как политический таргетинг или наблюдение. Это, по мнению А. Хинг, противоречит принципу уважения человеческого достоинства. Человека нельзя рассматривать как ресурс для получения информации, которую впоследствии можно купить и продать. Кроме того, следует разграничить перемещение данных между различными сферами и субъектами. Отсутствие рамок между частным и публичным, противоправным и правомерным стало результатом процесса бесконечной переработки информации. Подобная коммодитизация нашла отражение в недавно принятом регламенте Европола, который уделяет особое внимание процессам обработки, а не отдельным базам данных [с. 1272].

Современное использование персональных данных ставит под сомнение эффективность существующего правового регулирования конфиденциальности личной информации. И дело не только в экономическом неравенстве участников информационных отношений: даже если пользователи будут надлежащим образом осведомлены о последствиях дачи согласия на обработку данных, беспокойство от того, что эти данные будут доступны третьим лицам, не станет меньше. Различные социальные опросы показывают, что доверие пользователей к крупным социальным платформам (Facebook, Twitter, Google) неизбежно падает. Исследования КМРГ показали, что из 7 тыс. интернет-потребителей в 24 странах менее 10% считают, что имеют адекватный контроль над сбором и использованием своих персональных данных [с. 1273].

Можно утверждать, что проблема коммерческого использования информации о людях выходит за границы защиты персональных данных. Она поднимает более фундаментальный вопрос: что значит быть человеком вообще? В то же время сложно сформулировать и юридически обосновать, почему существующую практику сбора, обработки и хранения данных нельзя назвать приемлемой.

Наиболее распространенным объяснением является то, что подобные операции могут повлечь нарушение конфиденциальности или противоречить принципам защиты данных. Многие видят

решение возникших проблем в расширении прав на автономию и самоопределение личности. Однако автор считает это недостаточным для понимания и устранения этических проблем, вызванных современными цифровыми технологическими разработками.

Дискуссии в данной области ведутся достаточно долго. В сентябре 2015 г. было опубликовано мнение Европейского надзора по защите данных (EDPS) 4/2015, согласно которому современные тенденции информационного рынка (масштабный и тотальный сбор данных, отсутствие транспарентности в процессе их обработки) поднимают важные этические и практические вопросы в сфере применения принципов защиты данных. В данном документе предлагается новаторский подход, согласно которому «достоинство человека должно быть в центре новой цифровой этики» [с. 1275].

Выбор достоинства в качестве отправной точки информационной безопасности особенно эффективен ввиду наличия обширного законодательства Евросоюза, в котором понятие незыблемости человеческого достоинства играет основную роль. В Хартии ЕС подчеркивается не только наличие самого права на достоинство как фундаментального права человека, но и то, что оно служит основой для других основных прав и свобод, в том числе и права на неприкосновенность частной жизни. Человеческое достоинство является основой для интерпретации Общего регламента по защите данных (GDPR) и содержащегося в нем понятия информационной безопасности [с. 1276].

Конечно, можно сказать, что концепция человеческого достоинства не нова. Однако стоит учитывать, что она многогранна и изменчива. И мнение EDPS выявило новые ее аспекты с целью введения новых принципов в регулирование информационной безопасности, например принципа некоммерциализации. К сожалению, не все они нашли отражение в GDPR. Автор считает, что более эффективным было бы использование человеческого достоинства как средства ограничения субъектов информационной индустрии, а не как основы для расширения возможностей субъектов данных.

Современная коммодитизация информации привела к формированию глобальной сети, выходящей за пределы отдельных хранилищ персональных данных. Персональные данные, вне всяких сомнений, сегодня имеют коммерческую ценность. Все аспекты человеческого бытия повседневной жизни превращаются в товары, а индивиды начинают рассматриваться одновременно как субъекты персональных данных, объекты купли-продажи и источники прибыли [с. 1278]. Большая часть доходов компаний, будь то электрон-

ная коммерция, интернет-платформы и др., в значительной степени зависит от сбора, анализа и использования данных. Вероятно, в ближайшем будущем информация станет ключевым активом любой бизнес-модели. Большинство потребителей дают согласие на обработку персональных данных, потому что считают это необходимым для существования в виртуальном мире XXI в. Впоследствии их личная информация бесконечно циркулирует в интернет-пространстве, служа объектом купли-продажи между различными субъектами.

С точки зрения господствующей теории подобные процессы легитимны, поскольку субъект дал «конкретное, информированное и недвусмысленное» автономное согласие. Таким образом, все сводится к презумпции хорошо информированных, обладающих высоким уровнем цифровых навыков потребителей.

Эффективность такой концепции является предметом дискуссий на протяжении последних нескольких лет из-за целого ряда проблем. Во-первых, когда лицо дает свое согласие на обработку данных, оно не знает третьих лиц, которым впоследствии его данные будут переданы. Во-вторых, фактически у потребителей нет другого выбора. Компании провозглашают свободу и автономию пользователей, утверждая, что те могут просто отказаться от использования их услуг. Однако в большинстве случаев индивиды просто не могут не пользоваться этими услугами. По словам М. Сандела: «Мы перешли от рыночной экономики к рыночному обществу, в котором решение всевозможных социальных и гражданских проблем – это не проблемы морали, а законы рынка; денежные стимулы теперь являются главным механизмом, с помощью которого делается выбор» [цит. по: с. 1279]. Концепция человеческого достоинства, предусматривающая расширение прав и свобод личности, на самом деле просто перекладывает на нее все большую ответственность, а не придает фактическую значимость в информационных отношениях.

Т. Бернерс-Ли в связи с этим предлагает ввести механизм дифференцированного согласия на обработку персональных данных. Это решило бы проблему, вызванную тем, что данная процедура ограничивает пользователя бинарным решением: либо принять его, либо нет; не имея полноценных альтернатив. Дифференцированное согласие предоставит возможность, например, тем, кто предпочитает не просматривать рекламу, просто заплатить за это дополнительную денежную сумму. Однако это решение лишь подтверждает, что персональные данные имеют значительную экономическую ценность. Ситуация, когда индивидуальному субъекту

будет предлагаться автономный выбор между оплатой собственными данными или оплатой деньгами, только усилит тенденцию к коммерциализации личной информации.

Таким образом, какой бы ни была форма такого согласия, она не в состоянии решить фундаментальные проблемы коммодитизации персональных данных. Как подчеркивал М. Сандел, это связано с тем, что в результате рыночной оценки некоторые товары деградируют, их нравственная ценность снижается. Коммодитизация является важнейшей «моральной дилеммой, которую либерализация рынка решить не может» [цит. по: с. 1281].

В то же время не только бизнес, но и правительства активно используют возможности сбора, обработки и анализа данных, особенно в сфере борьбы с преступностью и терроризмом. Для этого правоохранные органы применяют самый широкий спектр способов: от мониторинга открытых источников до покупки данных у третьих лиц, в том числе полученных незаконным путем. Это делала, например, голландская секретная служба AIVD в 2016–2017 гг. Как подметил Б. Шнайер, «правительства просто забирают себе копию того, что уже собрал корпоративный мир» [цит. по: с. 1283]. В связи с этим некоторые авторы предполагают, что нынешний недостаток государственного регулирования вызван тем, что он просто выгоден правительственным структурам.

И следующий вопрос, который неизбежно возникает в сфере обмена персональными данными, связан со степенью осознания субъектом данных пределов этого обмена. Границы перемещения личной информации, полученной через онлайн-транзакции, весьма размыты. Должен ли пользователь учитывать, что в итоге его данные могут оказаться в базе данных голландских служб безопасности?

Положения GDPR не могут применяться в отношении службы безопасности Нидерландов, однако стоит вспомнить общие принципы защиты данных, изложенные в Конвенции Совета Европы 1981 г. В ней закреплено, что данные хранятся для определенных и законных целей и не используются иным образом, несовместимым с этими целями. Исключения из этого правила возможны лишь в случае их законности, необходимости и соразмерности.

Бесконтрольность потока данных, размывание границ между частным и публичным, законным и незаконным, экономическим и политическим выходят за рамки простой защиты конфиденциальности. Наиболее важной проблемой является то, что уникальные персональные данные превращаются в универсальный массовый продукт, т.е. происходит процесс коммодитизации. И концепция

человеческого достоинства с точки зрения расширения прав и возможностей субъектов персональных данных не способна справиться со всеми этими недостатками цифровой экономики.

Случаи с Cambridge Analytica и голландской службой безопасности лишней раз подтверждают откровения Сноудена. Фундаментальную проблему современности, когда личность фактически становится совокупностью данных, которые впоследствии перерабатываются, передаются, продаются и распространяются, нынешнее правовое регулирование, в частности GDPR, решить не в состоянии.

С точки зрения биоэтики существуют два аспекта человеческого достоинства: субъективный и объективный. Субъективное измерение рассматривает человеческое достоинство с точки зрения самоопределения и автономии личности, тогда как объективное измерение включает в себя такие ценности, как уважение, ограничение и коллективная ответственность. В регулировании биотехнологий ограничительные элементы играют доминирующую роль, воплощаясь в юридические запреты на превращение частей человеческого тела в объекты права и использование их в качестве инструментов, ресурсов, товаров. Как правило, законодатель лишь в исключительных случаях позволяет использование частей человеческого тела в качестве источников финансовой выгоды [с. 1286]. Такой подход был выдвинут еще в категорическом императиве И. Канта, согласно которому человеческие существа должны рассматриваться как цель, но не как средство.

Комплексная интерпретация человеческого достоинства могла бы внести неопределимый вклад в дискуссии о коммодитизации персональных данных. Торговля персональными данными, по существу, аналогична торговле частями тела человека. По справедливому замечанию Л. Флориди: «‘Мое’ в отношении моих данных не совпадает с ‘моим’ в отношении моей машины, оно, скорее, совпадает с ‘моим’ в отношении моей руки» [цит. по: с. 1287].

Спустя пару лет после публикации мнения 4/2015 Европейского надзора по защите данных (EDPS) был опубликован доклад «На пути к цифровой экономике», согласно которому необходимо переосмысление фундаментальных ценностей, лежащих в основе европейского регулирования защиты персональных данных. В частности, в нем подчеркивается, что «расхождение между основными концепциями, принципами защиты данных и эпистемической парадигмой больших данных» существенно ограничивает применение GDPR [с. 1288]. Необходима разработка новой концепции ввиду

возникновения постоянных угроз в связи со стремительным развитием экосистемы больших данных.

По мнению автора, в конце доклада также следовало сделать вывод, что определенные бизнес-модели, основанные главным образом на коммодитизации персональных данных и стимулирующие ее, стоит ограничить и даже в некоторых случаях запретить. Как заявил Т. Бернерс-Ли, в наше время «два мифа ограничивают процесс поиска оптимальных решений: миф о том, что реклама – это единственно возможная бизнес-модель для онлайн-компаний, и миф о том, что слишком поздно менять способ функционирования онлайн-платформ... нужно быть более креативными» [цит. по: с. 1288–1].

Таким образом, практика обработки личных данных, превратившая их в особый товар, всерьез подорвала представление о том, что человек находится вне сферы коммерческого оборота. И дело не только в угрозе неприкосновенности частной жизни: риску подвергаются и такие ценности, как свобода, чувство персонального контроля, правовая определенность. Для прерывания губительного процесса коммодитизации необходима более глубокая правовая концепция, которая не просто предоставила бы широкие полномочия автономным пользователям, но и устанавливала бы границы деятельности участников информационного рынка. Например, это можно осуществить через наложение запретов на неограниченный сбор, распространение и передачу данных.

*А.П. Иванова*

**Роен М., Фэн Ц.И.**

**ПОЧЕМУ «КОМПЬЮТЕР ГОВОРИТ НЕТ»:  
ИЛЛЮСТРАЦИЯ РИСКА ДИСКРИМИНАЦИИ  
ПРИ ИСПОЛЬЗОВАНИИ БОЛЬШИХ ДАННЫХ  
С ПОМОЩЬЮ НАУКИ О СЛОЖНЫХ СИСТЕМАХ  
(Реферат)**

Ref. ad op.: Rhoen M., Feng Q.Y. Why the «Computer says no»: Illustrating big data's discrimination risk through complex systems science // International data privacy law. – Oxford, 2018. – Vol. 8, N 2. – P. 140–159. – Mode of access: <https://doi.org/10.1093/idpl/ipy005>

В статье предложен подход к изучению законодательства Европейского союза о персональных данных через призму технологий обработки больших данных. В частности, речь идет об Общем регламенте Европейского парламента и Совета № 2016/679 от 27.04.16 «О защите физических лиц при обработке персональных данных и о свободном обращении таких данных», который вступил в силу в мае 2018 г. (далее – Регламент ЕС).

Актуальность исследования обусловлена тем, что современные технологии обработки больших данных способны разрушить все антидискриминационные механизмы. Например, научно доказано, что анализ больших данных о покупках в магазинах позволяет выявить будущих матерей, принадлежность к определенной расе или сексуальные предпочтения из анализа лайков в соцсетях. Таким образом, «датафикация» предоставляет всем, кто может получить доступ к соответствующей информации, возможность тотального контроля за каждым человеком.

Прежде всего это касается «особых категорий персональных данных», или «конфиденциальных данных», перечисленных в ст. 9 Регламента ЕС, как то: расовое или этническое происхождение, поли-

тические взгляды, религиозные или философские убеждения, членство в профсоюзах, генетические данные, биометрические данные с целью однозначной идентификации физического лица, данные о состоянии здоровья, о половой жизни или сексуальной ориентации.

Все это может привести к дискриминации, усилить маргинализацию отдельных социальных групп. Многие ученые отмечают, что алгоритмы могут быть предвзятыми и действовать дискриминационно в отношении защищаемых групп.

Регламент ЕС призван создать механизмы защиты персональных данных, особенно при их передаче за границы ЕС, однако, как отмечают авторы статьи, он был разработан еще до того, как стали очевидны риски технологии обработки больших данных. Поэтому они предлагают использовать науку о сложных системах для лучшего понимания рисков и средств защиты, связанных с передачей данных, для повышения эффективности антидискриминационных положений Регламента ЕС.

Общество представляет собой типичную комплексную систему, состоящую из множества компонентов, взаимодействующих на разных уровнях. Люди, образующие его, также представляют собой сложные системы, объединяясь в различные сообщества, коммуницируя между собой и с этими сообществами. В настоящее время все эти коммуникации регистрируются с помощью датчиков и оцифровываются различными способами (примерами могут служить смартфоны, журналы загрузок компьютеров, финансовые транзакции и т.д.), превращая многие аспекты жизни людей и обществ в большие данные.

Многие сложные системы обладают так называемыми эмерджентными свойствами (*emergent property*). Эмерджентное свойство – это свойство, отображаемое сложной системой, которое не прогнозируется непосредственно из свойств элементов этой системы. Примером может служить реакция человеческого организма на болезнь, которая напоминает реакцию других людей или схожие поведенческие стереотипы. Большие данные позволяют выявить эти эмерджентные свойства сложных систем.

Для анализа больших данных (например, для анализа большой клиентской базы для поиска групп «похожих» клиентов при сегментации рынка без предварительной идентификации сегментов) все чаще используется искусственный интеллект (самообучаемые алгоритмы). Однако если алгоритм обучается с использованием предвзято составленных данных, это находит отражение в результатах его работы. Связано это с тем, что поведение человека определяется его индивидуальными чертами и взглядами, а также пове-

денческими шаблонами, которые формируются под влиянием групп и обществ, окружающих его. Анализ больших данных позволяет выявлять эти шаблоны. И если оператор алгоритма, загружая в него большие данные, задает параметры, при которых персональные данные субъектов в зависимости от различий их поведенческих шаблонов обрабатываются по-разному, это может иметь дискриминационные последствия. В этих случаях алгоритмы могут быть названы «расистскими», «сексистскими» или иным образом предвзятыми по отношению к группам субъектов данных, имеющих общие признаки, в том смысле, что эти субъекты данных рассматриваются иначе, чем другие субъекты данных с аналогичными признаками. Например, предложение скидок лицам, покупающим алкоголь, может рассматриваться как дискриминация по отношению к тем, кто алкоголь не покупает по религиозным причинам.

Причем чем больше данных можно почерпнуть из открытых источников, тем больше риск получения на основе их анализа «персональных данных». И в этом случае, как отмечают авторы, эффективность законодательных гарантий защиты от их обработки существенно снижается.

В законодательстве ЕС обработка конфиденциальных данных, как правило, запрещена, если это специально не разрешено; в то же время положение об «общественных интересах» предоставляет государствам-членам значительную свободу действий.

Рекомендация ОЭСР 1980 г. «О руководящих принципах защиты частной жизни и трансграничных потоков персональных данных» не содержит аналогичного положения. Вместо этого в ней говорится: является спорным то, в какой степени люди, принадлежащие к определенной группе (например, психически недееспособные лица, иммигранты, этнические меньшинства), нуждаются в дополнительной защите от распространения информации, касающейся этой группы. В США, так же как и в странах ОЭСР, не запрещена обработка данных, касающихся этнической принадлежности или религиозных убеждений.

Как полагают авторы, законодательное регулирование имеет своей целью контроль рисков. Однако возможность установления неограниченного количества исключений (как это сделано в Регламенте ЕС) не позволяет эти риски эффективно контролировать. Кроме того, запрет на обработку персональных данных прежде всего обусловлен риском дискриминации. Однако вопросам дискриминации посвящены специальные нормативные акты с собственной доктриной и сформировавшейся на их базе судебной прак-

тикой. Тем не менее раньше, когда возможности распространения информации были относительно невелики, эти инструменты действовали эффективно. Сейчас же, в эпоху датафикации, субъекты персональных данных сталкиваются с большим количеством лиц и алгоритмов, обрабатывающих их данные, зачастую даже не подозревая, что такая обработка позволяет выявлять их специфические черты.

В этом смысле Регламент ЕС неоднозначен. С одной стороны, он направлен на защиту от дискриминации, но с другой – рассматривает «получение согласия на обработку персональных данных» как достаточное средство правовой защиты, хотя после получения такого согласия данные могут не только использоваться, но и передаваться в другие страны. С позиции европейского законодателя это считается оправданным компромиссом, тем более что обработка персональных данных может быть полезна и государству.

Авторы делают вывод, что, идя на такой компромисс, европейский законодатель не имеет четкого представления о потенциальном ущербе. Действительно, Регламент ЕС содержит более жесткие требования к работе с «конфиденциальными данными», но в нем не учитывается тот факт, что «конфиденциальные данные» с помощью современных технологий можно получить и из простых «персональных данных», требования к обработке которых гораздо мягче.

Авторы приводят пример, когда на основе анализа данных, добровольно предоставляемых беременными женщинами, можно обнаружить закономерность (переход к лосьонам без запаха на определенном сроке беременности), а затем применять этот «шаблон» ко всем женщинам, выявляя среди них беременных (затем это можно использовать в целях направления им определенной рекламы). Однако они видят неопределенность в норме ст. 9 Регламента ЕС об обработке персональных данных: касается ли она только «раскрытия» конкретных данных конкретного лица или выявления указанных «шаблонов»? С одной стороны, общий смысл Регламента ЕС свидетельствует о верности первого варианта. Но с другой – «обработка» определяется как «любая операция или набор операций, которые выполняются с персональными данными или наборами персональных данных». Следует отметить, что положение ст. 9 Регламента ЕС звучит по-разному в английской, французской и немецкой версиях в отношении того, распространяется ли запрет только на данные, помогающие идентифицировать

их носителя, или на обработку, позволяющую создать «шаблоны», с помощью которых лицо может быть идентифицировано. Тем не менее судебная практика однозначно толкует данную статью только применительно к «данным», а не к их «обработке».

Еще один пример слабой защиты, обеспечиваемой Регламентом ЕС, – это ст. 22, согласно которой предоставляются гарантии при составлении профилей человека с помощью программных продуктов. Согласно этой статье, любое лицо может обжаловать решение, принятое на основании такого профиля, а также отказаться от его составления, за исключением случаев, когда это требуется на основании договора, закона или личного согласия. Однако это касается только профилей, включающих конфиденциальные данные. Авторы описывают ряд ситуаций, демонстрирующих, что использование профилей при принятии решений алгоритмами может приводить к непреднамеренной дискриминации, причем выявить это иногда просто невозможно, поскольку требуется обработка персональных данных, которая запрещена ст. 9 Регламента. Кроме того, субъекты персональных данных, которые теоретически могут отказаться от составления и использования их профиля, зачастую поставлены в такие условия, когда это невозможно, либо предложенный им профиль может не включать именно «конфиденциальные» данные в значении Регламента ЕС. Конечно, они вправе обжаловать дискриминационное решение, но для этого они должны знать и понимать, что решение нарушает их права и принято на основании анализа компьютерным алгоритмом их профиля, что в большинстве случаев маловероятно.

Регламент направлен на предотвращение нарушения неприкосновенности персональных данных путем закрепления обязательных для исполнения технологически нейтральных принципов их обработки: законности, справедливости, прозрачности, ограничения цели, минимизации данных, точности, ограничения хранения, целостности, конфиденциальности и подотчетности. Именно за нарушение этих основополагающих принципов предусмотрено максимально возможное наказание – до 20 млн евро, или 4% мирового годового оборота (ст. 83 Регламента ЕС).

Однако, по мнению авторов, приведенные выше примеры показывают, что большинство из этих принципов сами по себе не являются достаточными для предотвращения дискриминации в результате обработки персональных данных, тем более что даже для контролирующих органов зачастую оказывается крайне затруднительно разобраться в тонкостях пользовательских соглаше-

ний, в которые включены пункты о праве операторов на обработку персональных данных пользователей. Так, положения и условия операционной системы Windows 10 использовались в течение двух лет, регулируя более 4 млн соглашений только в Нидерландах, прежде чем голландский надзорный орган рассмотрел их в 2017 г. и признал не соответствующими законодательству.

Авторы статьи рассматривают различные технологии защиты персональных данных, однако приходят к выводу, что ни одна из них не может гарантировать достаточно высокую степень защиты, тем более что развитие общества и технологий будет постоянно создавать новые возможности для персонализации на основании больших данных при обработке их искусственным интеллектом.

*С.И. Коданева*

# **ВЗАИМООТНОШЕНИЯ НАУКИ И ОБЩЕСТВА В ЭПОХУ БОЛЬШИХ ДАННЫХ: ЭТИЧЕСКИЕ, ЭПИСТЕМОЛОГИЧЕСКИЕ И ОРГАНИЗАЦИОННЫЕ ИЗМЕРЕНИЯ**

**Михаэль М., Луптон Д.**

## **К СОЗДАНИЮ МАНИФЕСТА ДЛЯ ИЗУЧЕНИЯ ПРОБЛЕМЫ «ПОНИМАНИЯ ПУБЛИКОЙ БОЛЬШИХ ДАННЫХ» (Реферат)**

Ref. ad op.: Michael M., Lupton D. Toward a manifesto for the «public understanding of big data» // Public understanding of science. – 2016. – Vol. 25, N 1. – P. 104–116. – DOI: 10.1177/0963662515609005

В своей статье австралийские авторы намечают контуры «манифеста», который может помочь в изучении «понимания публичкой больших данных» (ППБД). В частности, они изучают отношение людей к сложным вопросам, поднятым новым феноменом больших данных (БД).

Термин «большие данные» используется для описания массивных и постоянно разрастающихся цифровых баз данных, которые генерируются в результате взаимодействия с онлайн-технологиями. Эти интеракции могут включать как людей, так и неодушевленных акторов, особенно с развитием Интернета вещей<sup>1</sup> и «сенсорного общества» (sensor society)<sup>2</sup>.

Новым явлением последних лет стала экономика цифровых знаний (digital data knowledge economy), возникшая в результате

---

<sup>1</sup> Интернет вещей – концепция вычислительной сети физических объектов («вещей»), оснащенных встроенными технологиями для взаимодействия друг с другом или с внешней средой. Организация таких сетей рассматривается как явление, способное перестроить экономические и общественные процессы, из части действий и операций необходимость участия человека исключается. – *Прим. реф.*

<sup>2</sup> Понятие «сенсорное общество», или «общество датчиков», обеспечивает концептуальные основы для понимания характеристик возникающих форм мониторинга и контроля. – *Прим. реф.*

того, что цифровые данные приобрели коммерческую, исследовательскую и управленческую ценность для многих акторов и видов деятельности. Представляя собой агрегацию персональных (или «малых») данных в массивные базы, большие данные репрезентируются как предлагающие беспрецедентную возможность уловить общественные и экономические тренды и модели поведения. В своих множественных проявлениях БД стали «большим вопросом» для социальных наук: они одновременно являются и темой (каков социальный вклад больших данных), и ресурсом (какой вызов использование БД бросает существующим описаниям общественных процессов и как социальные науки ответят на этот вызов).

В данном случае авторы предпринимают попытку связать БД со спорами, которые ведутся в широкой междисциплинарной области, занимающейся проблемами «понимания обществом науки» и «включенности публики в науку и технологию». Будучи одновременно описанием, ресурсом и продуктом, которые зависят от общественных практик, БД приобретают разные формы, и их отношения с публикой высоковариативны. Такие практики ранжируются от «активных» (использование социальных сетей) до «пассивных» (осуществление сбора данных с помощью компьютерных технологий).

В ранних исследованиях, направленных на изучение отношений между наукой и обществом, применялись преимущественно количественные методы (прежде всего опросы), например для измерения научной грамотности населения или отсутствия таковой. Затем так называемая дефицитная модель отошла на периферию, а основное внимание стало уделяться качественному анализу знаний непрофессионалов (народному знанию), который включал и кейс-стади, посвященные изучению доверия населения к отдельным научным институтам. Преобладающими стали качественные методы (такие как этнографические наблюдения, полуструктурированные интервью и фокус-группы), которые оценивали, каким образом знания обычных людей связаны с вопросами идентичности, властными отношениями и неявными навыками.

Следующие изменения произошли, когда было пересмотрено разделение между экспертной и общественной составляющими – наукой и обществом. Так, согласно Модели 2 науки, по мере того как обычные люди привлекаются в качестве непрофессиональных экспертов или ученых граждан, границы между экспертным сообществом и непрофессионалами постепенно размываются. Привлечение общественности к решению вопросов, связанных с развитием науки и технологий, стало предметом осознанной и целенаправлен-

ной политики. Авторы называют эти изменения «поворотом к участию», в ходе которого проблема «понимания обществом науки» (ПОН) трансформировалась в проблему «включенности публики в науку и технологию» (ВПНТ).

Возник целый ряд интерактивных методов (participatory methods), таких как жюри граждан, согласительные конференции, групповые дискуссии с использованием карточек и пр. Основное внимание исследователи ВПНТ стали уделять тому, как обеспечить надлежащее функционирование этих методов, а также задумались о допущениях, встроенных в эти процедуры, и о более общих проблемах управления, связанных с ними.

По словам Н. Маррес<sup>1</sup>, исследователи ВПНТ в основном сосредоточились на «процедурных» элементах взаимоотношений между обществом и наукой. Она считает, что необходимо понять, как вопросы (особенно те, которые избежали институционального оформления) совозникают вместе с создаваемой ими публикой. Таким образом, Н. Маррес предполагает, что публика способна артикулировать вопросы, которые институционально не смогли быть уловлены. Именно внутри этого меняющегося ландшафта ПОН и ВПНТ авторы намерены расположить свой манифест, касающийся изучения проблемы, которую они называют «понимание публикой больших данных» [с. 106].

Появился уже достаточно большой объем литературы, которую можно описать как критические исследования БД, в которой последние позиционируются как социотехнический ассамбляж, включающий комплексное взаимодействие людей и не-людей.

Предполагаемая объективность и нейтральность БД были поставлены под сомнение. Критически настроенные ученые усомнились в понятии «сырых» данных, которое лежит за преобладающими репрезентациями БД, указывая на то, что данные всегда носят частичный, выборочный характер, «состряпаны» социальными акторами и не бывают «чистыми». Было также доказано, что соавторами БД выступают гетерогенные акторы: представители публики; программное обеспечение; платформы; датчики; эксперты, занимающиеся составлением программ; эксперты по анализу данных; алгоритмы, использующиеся при отборе и интерпретации данных.

Ассамбляж цифровых данных создается каждый раз, когда пользователь взаимодействует с технологиями, встроенными в

---

<sup>1</sup> Marres N. Material participation: Technology, the environment and everyday publics. – Basingstoke: Macmillan, 2012. – 206 p.

Интернет, и таким образом поставляются данные, которые затем будут агрегированы с другими данными. Пользователь одновременно выступает в двух ролях – и создателя, и потребителя контента. Это совмещение двух ролей обозначается термином «просьюмер» (prosumer), который используется в исследованиях медиа и Интернета, чтобы описать интеракции пользователя с онлайн-технологиями. Важно отметить, что эти ассамбляжи данных реактивны, динамичны и постоянно реконфигурируются, по мере того как генерируются новые данные, а также происходит перенастройка баз данных [с. 107].

Ассамбляжи данных одновременно опираются на персональные данные индивидов и вносят вклад в большие цифровые базы данных, поэтому они могут быть использованы в качестве обратной связи для индивидов. В качестве примера авторы приводят приложение и платформу Glow, позволяющие женщинам точно определять момент овуляции [с. 108].

В то же время БД могут с успехом использоваться для осуществления надзора за населением. В последние годы в прессе появилось много сообщений о том, как персональные данные людей и метаданные используются секретными службами в США, Великобритании и Австралии. Самый яркий пример – разоблачения, сделанные Э. Сноуденом. Эти сообщения вместе с описанием хакерских атак, а также случаями продажи персональных данных третьим лицам продемонстрировали публике, в какой опасности находятся их данные. Однако у людей практически нет выбора. Кроме того что пользователи цифровых девайсов находятся под надзором друг у друга (на социальных медиасайтах), публичное и частное пространство все больше насыщается цифровыми сенсорами, которые отслеживают самые мелкие аспекты поведения и действия посетителей этих пространств [с. 108].

«Понимание», или (как писали раньше) «знание», как экспертов, так и непрофессионалов, обычно репрезентировалось как некое статичное владение. Затем ситуация изменилась, и эти концепты стали привязывать к вопросам, касающимся доверия, идентичности, диалога и пр. Однако реактивный, динамичный и скоростной характер знаний, создаваемых онлайн-технологиями, изменил ситуацию радикально. Что конкретно следует считать знанием, или пониманием, если данные сами представляют собой постоянно меняющийся поток и для обычных людей, и для экспертов?

Более того, знания, инвестируемые в экономику цифровых знаний, обладают разными формами ценности. В отличие от

большинства видов непрофессионального знания, которые исследуют ПОН и ВПНТ, знание, которое генерируется и потребляется публикой как часть баз данных, имеет прямую коммерческую стоимость или обладает иной ценностью. Коль скоро данные входят в экономику цифровых знаний, те, кто их генерирует, часто теряют контроль над ними. В связи с этим возникает целый ряд вопросов, касающихся отношений знания и «собственности», а также фигуры «просьюмера» и его прав [с. 109].

Понятие «экспертиза» также должно быть пересмотрено в свете новой экономики цифровых знаний. Новые разновидности «науки граждан» представляют собой формы краудсорсинга, направленные на обеспечение людей цифровыми девайсами для сбора научных данных, которые затем объединяются и анализируются. Еще один пример, который приводят авторы, – концепция «умного города». Представляя собой систему коммуникативных и информационных технологий, взаимосвязанную с Интернетом вещей, «умный город» выполняет две главные задачи: сбор и передачу данных административным структурам, которые намного упрощают управление его внутренними процессами; налаживание обратной связи между администрацией и горожанами в целях защиты экологии и создания благоприятной окружающей среды.

Подобные инициативы, в свою очередь, меняют представление о том, кого следует считать «гражданином». Быть активным и ответственным гражданином значит согласиться активно использовать различные цифровые девайсы и датчики для мониторинга оставляемых следов и состояния окружающей среды, а также обмена этими знаниями с другими. «Так же как мы можем говорить об “умном городе”, мы можем говорить об “умном гражданине” – точке в Интернете вещей, испускающей цифровые данные. Все это меняет концепт экспертизы. Она распределяется между одушевленными и неодушевленными актерами, что включает “умный ландшафт”, начиненный датчиками и девайсами, в котором совместно генерируются данные» [с. 109].

Важно отметить, считают авторы, что движение БД осуществляется не беспрепятственно: блокировка и сопротивление представляют собой часть экономики цифровых знаний. Возможности непрофессионалов и массовых организаций использовать БД ограничены, если они не обладают соответствующими экспертными знаниями. Но не только отсутствие технических знаний мешает людям использовать БД для собственных целей. Вопреки рассуждениям об «открытом доступе» к «открытым данным», такой доступ уменьшается, по

мере того как растет коммерческая стоимость данных. Интернет-империи (Google, Facebook, Amazon, Twitter и Apple) вместе с более мелкими корпорациями и правительственными учреждениями осуществляют контроль над архивами цифровых данных; доступ к ним становится вопросом, связанным с корпоративной властью, и щедрым даром со стороны правительства [с. 109].

Непрофессионалы в той или иной мере осознают значимость БД для благополучия общества, например для обеспечения национальной безопасности, контроля за преступностью, развития здравоохранения и пр. В то же время многие также осознают, что их данные обладают коммерческой ценностью, и протестуют против того, чтобы правительственные учреждения продавали их, вместо того чтобы использовать на благо общества.

Таким образом, можно видеть, что публика имеет разные «эпистемные лица» в отношении БД: «непрофессиональные пользователи» (пользователи, или потребители, БД); «непрофессиональные вкладчики» (те, кто генерирует и поставляет БД); «непрофессиональные эксперты» (те, кто владеет определенными компьютерными знаниями и занимается процессами, качеством и политикой в области БД); «эксперты-непрофессионалы» (те, кто пишет программы и изобретает технологии).

Эти категории не являются стабильными или взаимоисключающими, но они дают представление о том, насколько сложны проблемы вовлеченности публики в динамику БД. Поэтому одна из задач исследователей, которых интересует эта тема, состоит в идентификации онтологии этих категорий и их пересечений. «Другими словами, понимание большими данными публики и понимание публикой больших данных – это взаимные и соразвивающиеся конфигурации» [с. 110].

Из всего вышесказанного следует, что изучение того, каким образом публика вовлечена в БД, реагирует на них и использует их, предполагает изменение привычного стиля мышления, что уже нашло отражение в таких областях, как цифровая социология, цифровые культуры, исследования программного обеспечения и исследования Интернета. Более того, учитывая то, как цифровое знание о привычках людей и их предпочтениях генерируется и входит в базы БД, исследование ППБД потребует обращения к современным концепциям собственности, частной жизни и информационной безопасности, которые мало звучали в прежних областях ПОН и ВПНТ.

Исследование ППБД означает неизбежное столкновение с новыми моделями производства знания и его распространения, новой академической литературой и новым осмыслением концепта БД, практики их получения и распространения. Авторы формулируют основные принципы своего манифеста. Они разделили его на три части: эмпирическую, концептуальную и методологическую.

### **Эмпирическая часть**

– Публика – это изменчивая комбинация пользователей и производителей.

– Эмпирические данные ППБД имеют многообразный и подвижный характер, они переплетены и запутанны и могут включать ППБД.

– Экспертиза носит мобильный и распределенный характер, соединяя непрофессионалов и профессионалов (человеческих и неодушевленных акторов).

– БД включают данные, которые сложным образом ассоциированы с вопросами собственности, контроля и частной жизни.

– Понимание и вовлеченность в создание и использование БД необходимо рассматривать в связи с понятием «осознание» и различными его уровнями [с. 111–112].

### **Концептуальная часть**

– ППБД нуждается в разработке концептов, способных анализировать вклад БД, поскольку они включают комбинацию относительно самостоятельных баз данных, обеспечивающих более или менее стабильные и новые идентичности публики.

– ППБД нуждается в расширении концептуального аппарата, чтобы изучать, как идентичности, порождаемые БД, пересекаются с уже сложившимися идентичностями.

– Учитывая сложность общественных практик, касающихся БД, ППБД нуждается в пересмотре ключевых идей ПОН / ВПНТ, а именно: как такие понятия, как идентичность, доверие, вовлеченность, народное знание и участие, ведут себя в контексте БД [с. 113]?

### **Методологическая часть**

– ППБД нуждается в принятии и адаптации методологий, которые способны уловить многоликую вариабельность публики, БД и отношений между ними.

– В частности, такие методологии должны будут включать не только понимание и вовлеченность в БД, но и изменчивые формы осознания, а также имажинативные и аффективные отношения к БД.

– Такие методологии должны учитывать особую процессуальность и повторяемость событий в исследованиях ППБД и рассматривать вопросы их собственной особой формы перформативности в генерировании социальных данных.

В заключение авторы еще раз подчеркивают, что они лишь попытались наметить контуры подхода к изучению ППБД, который будет учитывать комплексный характер БД. Очевидно, что это только общие соображения и начало изучения этой проблемы [с. 114].

*Т.В. Виногорова*

**Сивков Д.Ю.**

**БОЛЬШИЕ ДАННЫЕ В ЭТНОГРАФИИ:  
ВЫЗОВЫ И ВОЗМОЖНОСТИ // Социология науки  
и технологий. – 2017. – Т. 8, № 1. – С. 56–67.  
(Реферат)**

Автор, российский философ, намерен «выявить и обсудить коллаборативный потенциал этнографии (и, шире, – качественных исследований вообще) и больших данных (БД). Он пытается понять, какие вызовы исходят от больших данных и какие возможности они могут предоставить этнографии» [с. 57].

Появление феномена БД в маркетинге и сфере управления стало серьезным вызовом для социальных наук, который заставил пересмотреть их возможности. 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 вступают в противоречие с методами разработки БД. Несмотря на эпистемологический оптимизм БД, в бизнес-этнографии, в первую очередь в маркетинговых исследованиях, заговорили об их недостатках 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 БД.

Однако все не так однозначно. В начале антропологии у У. Тайлора и особенно у Б. Малиновского стоял вопрос об использовании статистических данных в этнографии. Этот вопрос тесно связан с требованием научной объективности антропологических изысканий. В определенном смысле этнография – это всегда напря-

жение между поэзией насыщенного описания и квантификацией строгой науки. Антропология балансирует между нерепрезентативными техниками и «доверием к числам» (trust in numbers). 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 2000-х годах наметилась тенденция к квантификации качественных данных и использованию дополнительных количественных данных для анализа результатов полевых исследований.

При каталогизации отношений БД и этнографии важно отметить противопоставление их методологий и эпистемологий. БД делают то, что не могут «схватить» этнографы, и наоборот: этнография способна видеть вещи, недоступные алгоритмам машинного обучения. Появление аналитики БД было отмечено оптимизмом технических специалистов и корпоративного сектора, так как отпала необходимость в социальных и иных исследователях для понимания тех или иных процессов и явлений.

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 состоит в коллаборации между производителями информации, исследователями, хранителями, распространителями и конечными пользователями. При этом предлагается общее цифровое решение: Интернет мог бы быть для различных групп общим интерфейсом для обмена, анализа и хранения данных.

Коллаборация этнографов со специалистами – аналитиками БД показала, что БД предлагают исследователям охват, а этнография – необходимую для интерпретации данных глубину. Кроме того, коллаборация предлагает этнографу доступ к БД. Сам по себе этот доступ является серьезной исследовательской проблемой, и он есть только у корпоративных этнографов. Совместная работа в рамках коллаборации позволяет ученым расширять свои навыки и учиться друг у друга.

Кроме противопоставления БД и этнографии автор отмечает такую тенденцию, как дополнение или взаимное использование их возможностей в междисциплинарной перспективе. Анализ БД и его результаты 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.

Этнография может предложить рамки для интерпретации БД и результатов машинного анализа. Кроме того, объемы данных требуют от исследователя определенной фокусировки на конкретных проблемах или сегментах. БД требуются «маленькие паттерны» (small patterns), которые говорят, где искать данные, что именно следует исследовать и как это использовать. Вместе с рамками для анализа данных этнографы могли бы составить коллаборации со специалистами по БД.

Один из ключевых тезисов звучит так: «Без контекста большие данные теряют смысл» [с. 63]. 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.

Автор задается вопросом: «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?» [с. 63]. БД 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. БД 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.

Итак, этнография и БД, несмотря на различие подходов и методов и несмотря на противопоставление, могут совместно решать исследовательские задачи и быть полезными друг для друга. Есть даже точка зрения, согласно которой и БД, и этнография решают одну задачу, а именно «как большие данные создают культурные профили того, чем мы являемся» [с. 63]. Коллаборативный потенциал этнографии и БД раскрывается во взаимной проблематизации неких узловых моментов, концептов, методов и даже фоновых оснований. Кроме того, малый фокус или малый паттерн БД существенно облегчают доступ к меньшинствам (например к пожилым женщинам, много времени проводящим за онлайн-играми), которые не только не заметны на фоне анализа БД, но доступ к

которым зачастую закрыт. Принцип БД «чем больше, тем правдивее», как оказывается, работает далеко не всегда.

В то же время для качественных исследований существует другая серьезная проблема, 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. Таким образом, БД с малым фокусом – это еще один способ решения важной этнографической проблемы, а именно проблемы выявления границ некоторого не представленного или плохо представленного сообщества, не имеющего пространственной локализации. Контекстуальность анализа БД – одно из главных требований со стороны социальных и гуманитарных наук.

В заключение автор отмечает, что несмотря на шум вокруг БД и на то, что они становятся зонтичным понятием, обозначающим все и ничего, появление этого феномена и новых методов ставит под вопрос основания социальных наук и проблематизирует концепты. Поэтому необходимо очертить контуры перспективных исследований в области взаимодействия этнографии и БД [с. 67].

*Т.В. Виноградова*

**Флориди Л.**  
**БОЛЬШИЕ ДАННЫЕ**  
**И ИХ ЭПИСТЕМОЛОГИЧЕСКИЙ ВЫЗОВ**  
**(Реферат)**

Ref. ad op.: Floridi L. Big data and their epistemological challenge // Philosophy and technology. – 2012. – Vol. 25, N 4. – P. 435–437. – Mode of access: <https://doi.org/10.1007/s13347-012-0093-4>

Посчитано, что с момента изобретения письма и до 2006 г. человечество аккумулировало 180 EB (EB – 1 эксабайт = 1000 петабайтов) данных, а в период с 2006 по 2011 г. эта величина выросла приблизительно в 10 раз, достигнув 1600 EB. Эта цифра возрастает примерно в четыре раза каждые три года. Каждый год, пишет автор, создается достаточно данных, чтобы восемь раз заполнить все библиотеки США. В результате возникает много разговоров о так называемых больших данных (БД), без которых были бы невозможны многие научные исследования, например современная генетика, которая является сегодня одной из самых ресурсоемких наук. Национальные институты здравоохранения США (National Institutes of Health – NIH) и Национальный научный фонд (National Science Foundation – NSF) определили БД как свой программный фокус, имея в виду общие инициативы, направленные на разработку ключевых технических средств и технологий, необходимых для развития науки о БД.

Несмотря на важность самого явления, все еще остается неясным, отмечает автор, что именно обозначает термин «большие данные» и к чему он относится. Вышеупомянутый документ определяет его как «фразу, относящуюся к широким, разнообразным, сложным, удлиненным и / или распределенным кластерам данных, генерируемых различными инструментами, сенсорами, интернет-

транзакциями, имейлами, потоками видео, кликами и / или другими цифровыми источниками, доступными сейчас и в будущем». Не надо быть философом аналитической школы, замечает Л. Флориди, чтобы увидеть нечеткость и расплывчатость подобных формулировок. Википедия здесь тоже не помощница – она дает обобщенную неудовлетворительную дефиницию: «Наборы данных столь обширные и сложные, что с ними неудобно работать, используя имеющиеся для баз данных подручные инструменты».

Автор указывает на круговую проблему в определении, когда «big» интерпретируется через «large», а также на связь значений обоих терминов лишь по отношению к текущим вычислительным мощностям, что, по его мнению, ведет к заблуждению. «Большой», как и многие другие, является относительным предикатом: пара обуви слишком велика для одного человека, но прекрасно подходит другому. Тривиально также признание, что в данном случае есть стремление определять вещи безотносительно чего-либо, как абсолютно большие, – каждый раз, когда референтный фрейм достаточно велик, чтобы остаться имплицитным. Конь – достаточно большое животное, что бы там киты ни думали [с. 435–436].

Эти две простые точки могут создать впечатление, что в действительности нет реальных проблем с БД, неряшливо определяемыми как термин, относящийся к тому факту, что современные компьютеры не в состоянии достаточно эффективно обрабатывать столь много «газообразных» данных.

Здесь, продолжает Л. Флориди, появляются два недоразумения. Эпистемологическая проблема с БД заключается, во-первых, в том, что их слишком много (этическая проблема – в том, как ими пользуются), и, во-вторых, в том, что технологическое решение эпистемологической проблемы – более совершенные технологии, которые «сожмут» БД до размеров, с которыми можно работать. Эпистемологическая проблема иная, она требует эпистемологического же решения. Термин «большие данные» стал формулироваться после других навязчивых выражений, таких как «информационное перенасыщение» (infoglut) или «перегруженность информацией», которые уже начали исчезать, но идея оставалась все той же и относилась к «оглушающему» смыслу [с. 436].

Автор приводит яркий образ людей, которые откусили кусок больше, чем могут прожевать, сравнивая людей с насильно перекормленными гусями, так что их интеллектуальные печенки уже взрываются, что, полагает Л. Флориди, ошибочно. Несомненно, имеет место экспоненциальный рост объема данных во все боль-

шем числе областей знания. Но кивать на подобное сверхозобилие – это все равно что жаловаться на какой-нибудь банкет, на котором предлагается больше, чем может съесть человек.

Данные остаются активом, ресурсом для использования. Никто не принуждает «поедать и переваривать» каждый доступный байт. С каждым днем человечество все богаче и богаче данными, но это не может быть фундаментальной проблемой. Она заключается не в увеличении объема данных, становящихся доступными, дело не в количестве данных, которые можно обработать. Большие и лучшие технические устройства и технологии приведут только к генерированию еще большего количества данных. Если бы проблема была в слишком большом количестве данных, компьютеры лишь обострили бы ее. Выращивание все больших систем «переваривания» информации, которое и имеет место, – это не путь вперед.

Подлинная эпистемологическая проблема с БД – в маленьких паттернах. Именно потому, что такое огромное количество данных сейчас может производиться и обрабатываться так быстро и так дешево, давление как на данные вновь возникших богатеев вроде Facebook, Walmart, Amazon и Google, так и на данные уже давно конституированных гигантов потребления активов вроде генетики, медицины, неврологии, способно помочь отследить, считает Л. Флориди, не только то, где лежат новые паттерны с действительно ценной новизной в их безмерно обширных хранилищах данных, но и то, как они могут быть употреблены для создания и богатства, и прогресса в знаниях.

Эти небольшие паттерны значимы, поскольку репрезентируют новые соревновательные и конкурентные границы, начиная от науки и заканчивая бизнесом, от правительственных структур и до социальных программ включительно. А если кто-нибудь сможет использовать их раньше и успешнее на «бэконовском» открытом рынке идей, то другие вскоре могут оказаться не у дел, как это произошло с компанией «Кодак», или же упустить какое-нибудь фундаментальное открытие.

Небольшие паттерны могут таить в себе и риски, поскольку расширяют пределы предсказуемого и могущего быть предвиденным – и не только относительно природы, но и людей, их поведения. И это – этическая проблема. Для ее иллюстрации автор приводит пример с фирмой, которая, основываясь на анализе паттернов покупательности 25 продуктов, отмечала для каждой покупательницы ее черту «предсказанной беременности», оценивала время и сроки, посылая купоны к определенным стадиям ее предполагаемой беремен-

ности. Однажды такой подход вызвал серьезные последствия, когда эта компания послала купоны в семью, где дочь-тинейджер не информировала родителей о своем новом положении.

Однако маленькие паттерны, замечает ученый, значимы только в том случае, если их соответствующим образом объединять, т.е. – в приведенном примере – если надежны и предположения о покупках, и сами расчеты. Аналогично в банковском деле, когда какой-нибудь банк умеет использовать БД, чтобы бороться с мошенниками и / или учитывать различные временные колебания на финансовых рынках.

Поскольку информация индикативна даже тогда, когда она отсутствует, мелкие паттерны значимы и тогда, когда их нет. Шерлок Холмс распутал одно из своих дел благодаря молчанию собаки, которая должна была бы лаять. Так и с БД: если они «не лают», пишет автор, когда должны были бы, значит, нечто происходит, как, должно быть, известно «финансовым сторожевым псам» [с. 436–437].

Все более ценные подводные течения во все более расширяющихся океанах данных невидимы для не вооруженного компьютером человека, поэтому количественное увеличение и качественное улучшение техники и технологий оказывают значительную помощь. Однако сами по себе они недостаточны. Простое накопление данных в ожидании появления более мощных компьютеров и софта не работает: с 2007 г. мир стал производить больше данных, чем имеется возможностей для их хранения. Человечество сместилось от проблемы «что сохранять» к проблеме «что стирать».

Автор статьи предлагает своему читателю представить, что его смартфон переполнен слишком большим количеством фотографий, и вообразить это как глобальную проблему. Уже в течение многих лет в инфосфере не осталось памяти, даже чтобы просто сваливать данные в кучу. Но это не так плохо, как выглядит, ведь половина их – просто мусор; только люди не знают, какая именно половина. Таким образом, люди должны понять, какие данные заслуживают хранения. А это уже дело понимания того, какие вопросы интересны сейчас или же будут таковыми в будущем. Другими словами, поскольку проблема с БД заключается в небольших паттернах, то игру в конце концов выиграют те, цитирует автор Платона, кто «знает, как задавать вопросы и отвечать на них». И они знают, какие данные могут быть полезными и релевантными и в силу этого стоят аккумуляции и заботы в целях использования их ценных паттернов [цит. по: с. 437].

Люди нуждаются как во все больших количествах усовершенствованных техники и технологий, чтобы видеть паттерны данных, так и во все больших и лучших эпистемологиях, чтобы отсеивать самые ценные из них. БД уже здесь, и их объем будет только увеличиваться. И единственный путь совладать с ними – это знать, что искать. Сейчас такие эпистемологические умения, иронизирует Л. Флориди, применяются аналитикой – «черной магией». Тем не менее столь многое зависит от нее, что, очевидно, пришло время начать философское исследование ее методов. Философы, возможно, должны чему-то научиться и сами преподавать пару уроков. Платон, заканчивает статью автор, наверное, согласился бы.

*Г.В. Хлебников*

**Прайс У.Н., Коэн И.Г.**  
**ЧАСТНАЯ ЖИЗНЬ В ЭПОХУ**  
**БОЛЬШИХ МЕДИЦИНСКИХ ДАННЫХ**  
**(Реферат)**

Ref. ad op.: Price W.N., Cohen I.G. Privacy in the age  
of medical big data // Nature Medicine. – 2019. – Vol. 25. – P. 37–43. –  
DOI: 10.1038/s41591-018-0272-7

Авторы – ведущие специалисты в области права и наук о жизни (США) – рассматривают этические аспекты проблемы конфиденциальности больших данных (БД) в медицине. Они отмечают, что с БД сегодня многие эксперты связывают перспективы существенного улучшения здравоохранения. БД часто определяют как «три V»: объем (большие объемы данных), скорость (высокая скорость доступа и анализа) и разнообразие (существенная неоднородность данных по отдельным лицам и типам данных).

Исследовательские приложения БД в медицине можно представить в виде двух групп: давно применяемые методы анализа и новые методы, использующие технологии машинного обучения и искусственного интеллекта (ИИ). БД позволяют значительно улучшить оценку качества и эффективности медицинской помощи, что весьма актуально на фоне постоянных доказательств неэффективности лечения, медицинских ошибок и значительных финансовых затрат. Они также могут использоваться в разработке научных гипотез, при сравнении эффективности различных вмешательств, для контроля безопасности лекарств и устройств, как в системе Sentinel Управления по контролю за продуктами и лекарствами США (FDA).

Кроме того, применение технологий ИИ для поиска закономерностей о состоянии здоровья на основе БД позволит прогнозировать течение болезни и предоставлять рекомендации по уходу за

пациентами. Самый известный из этих инструментов – анализ изображений – уже вводится в клиническую практику. Алгоритмы могут идентифицировать раковые поражения кожи по изображениям так же точно, как и профессиональные дерматологи, а система диагностики диабетической ретинопатии (IDx-DR) на основе ИИ не так давно получила одобрение Управления по контролю за продуктами и лекарствами США. Алгоритмы ИИ могут использоваться во многих областях: для предсказания необходимости оперативного вмешательства у пациентов с травмой, прогнозирования смерти пациентов, для своевременного перехода к паллиативной помощи или для принятия решений о распределении ресурсов.

Обучить ИИ или оценить успешность обучения невозможно без больших массивов медицинской информации, которая должна включать не только персональные данные о пациенте, его образе жизни, поведении в отношении здоровья, но и сведения о семейной истории болезней. Развитие обучающихся систем здравоохранения – это еще один этап эволюции БД в здравоохранении, который постепенно набирает обороты. В этих системах граница между клиническими исследованиями и медицинской помощью размывается, хотя даже в более традиционных разработках существует некоторая неопределенность с точки зрения нормативных аспектов относительно разграничения исследований, улучшения качества и инноваций. Развитие обучающихся систем здравоохранения, несомненно, будет способствовать улучшению его качества, включению новых идей в процессы оказания медицинской помощи, будь то явным образом (в практических руководствах или публикациях) или неявным образом (в контексте рекомендаций или процедур, автоматически включаемых в системы электронных медицинских карт (EHR)).

Понятие конфиденциальности довольно сложно определить. Одно из определений связано с контекстом: кто хранит информацию; кто к ней имеет доступ; почему и как она может передаваться. Нарушение контекстных правил означает нарушение конфиденциальности и, как правило, влечет за собой две категории проблем: *консеквенциалистские* и *деонтологические*. Некоторые нарушения конфиденциальности могут затрагивать оба типа. Кроме того, некоторые проблемы также актуальны для «небольших данных», но в случае с БД различные негативные эффекты значительно увеличиваются.

*Консеквенциалистские* проблемы касаются нежелательных последствий вмешательства в личную жизнь человека. Это может быть рост стоимости страховки в результате доступа к информа-

ции, дискриминация при приеме на работу или распространение информации о ВИЧ-статусе в кругу общения пациента. Речь также может идти об эмоциональных переживаниях, связанных с осознанием того, что личная медицинская информация «существует» и потенциально может использоваться другими.

*Деонтологические* проблемы не зависят от переживания негативных последствий. Беспокойство относительно нарушения конфиденциальности проявляется даже в тех случаях, если никто не использует информацию против человека или если он никогда не узнает, что нарушение имело место. Например, организация недобросовестно или по неосторожности получает доступ к данным, хранящимся в смартфоне, являющимся частью более крупного хранилища данных. Признав их бесполезными для своих целей, организация уничтожает записи. Человек может даже не узнать, что это случилось. Люди, имевшие доступ к этим данным, живут очень далеко и никогда не встретят владельца смартфона или его знакомых. Трудно сказать, что человеку был нанесен ущерб в строгом смысле, но многие считают, что вторжение и потеря контроля над данными сами по себе этически проблематичны, даже если вреда нет.

Сбор медицинских данных поднимает много правовых и этических вопросов конфиденциальности. Авторы рассматривают проблемы обработки данных о здоровье и методы их решения в США и сравнивают их с подходами, применяемыми в ЕС. Данные о состоянии здоровья поступают из разных источников: электронные истории болезни, страховые претензии, устройства Интернета вещей, сообщения в социальных сетях и т.д. Закон США о конфиденциальности трактует данные о здоровье в зависимости от того, как они создаются и кто их обрабатывает, т.е. кто хранит. Регламент ЕС, напротив, устанавливает единый широко определенный режим для данных о здоровье (а также других данных), независимо от того, в каком формате и как они собираются или кто является хранителем. В ЕС категория «данные о здоровье» определяется в широком смысле – как «персональные данные, относящиеся к физическому или психическому здоровью физического лица, включая предоставление медицинских услуг, которые раскрывают информацию о его или ее состоянии здоровья»<sup>1</sup>.

---

<sup>1</sup> Commission Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with regard to the Processing of Personal Data and on the Free Movement of such Data, and Repealing Directive, 95/46/EC, 2016 O.J. (L 119) 1, 34 (General Data Protection Regulation). – Mode of access: <http://www.privacy-regulation.eu/en/article-4-definitions-GDPR.htm> (2016).

Хранителями, на которых сфокусировано законодательство США, являются врачи, система здравоохранения и их деловые партнеры. Основной федеральный закон США, регулирующий конфиденциальность медицинских данных, – Правило конфиденциальности, созданное в соответствии с Законом о переводе медицинского страхования [в целях его непрерывности] и контроле (Health Insurance Portability and Accountability Act – HIPAA). Существуют также законы о конфиденциальности для отдельных штатов и федеральное законодательство, которое защищает субъектов исследований. В соответствии с Правилom конфиденциальности HIPAA запрещено использовать или раскрывать «защищенную медицинскую информацию», за исключением определенного списка обстоятельств. Эта информация может использоваться во всех видах лечения, для оплаты услуг, в целях общественного здравоохранения и правоприменения, но ее невозможно использовать в научных исследованиях без разрешения институционального контрольного совета (Institutional Review Board – IRB) или пациента. Один из пробелов HIPAA – защита пациентов от нарушений конфиденциальности при одновременном предоставлении доступа к данным с помощью удаления набора из 18 идентификаторов, таких как имена и адреса электронной почты. Однако, как показывают публикации, деидентифицированные данные могут стать реидентифицируемыми. Кроме того, HIPAA фокусирует регулирование на конкретных субъектах и их деятельности, а не на самих данных.

Но более сложная проблема заключается в том, что большинство данных о здоровье вообще не охвачено HIPAA. Принимая закон в 1996 г., конгресс предполагал режим, при котором большинство данных будет храниться в медицинских записях, и поэтому он сосредоточился на поставщиках медицинских услуг и других организациях, охваченных страховкой. В настоящий момент источники, включенные в HIPAA, являются лишь небольшой частью более обширной экосистемы данных о здоровье. Закон не распространяется на данные о медицинском обслуживании, полученные за пределами организаций, на которые распространяется действие страховки, он не охватывает данные о состоянии здоровья, которые не связаны с медицинской помощью, но генерируются множеством людей или продуктов. Он не распространяется на пользовательскую информацию из мобильных приложений для смартфона или запросы Google по определенным симптомам. Кроме того, есть огромный объем данных, которые вообще не

касаются здоровья, но позволяют сделать выводы о нем, например информация о покупках. Человечество вступает в будущее, в котором традиционные области здравоохранения, охватываемые субъектами HIPAA, вытесняются в пространстве данных о здоровье такими гигантами, как компании Google, Apple и IBM, работающими вне режима закона. Соответственно, акцент американского законодательства на данных, вытекающих непосредственно из лечебных учреждений, контрастирует с европейским регулированием данных, касающихся здоровья, в целом. Хотя некоторые законы могут защищать конкретные виды использования данных, в целом существует значительная проблема, которую необходимо решать в ближайшем будущем.

Другая проблема касается справедливости сбора данных. С этической точки зрения сбор данных лучше всего рассматривать как своего рода сделку между тем, кто их предоставляет, и тем, кто их использует. Любое нарушение баланса между этими сторонами может сорвать сделку или негативно повлиять на анализ и прогноз. Важно учитывать также слабую представленность некоторых групп граждан в БД и необходимость статистической корректировки. Например, программа «Все мы»<sup>1</sup>, разработанная Национальными институтами здравоохранения, направлена на разработку репрезентативной для всей страны геномной выборки.

Множество сложных вопросов возникает относительно роли пациента в процессе сбора данных и доступа к ним. В частности: в какой степени данные об индивиде должны быть доступны для использования в прогнозной аналитике без его согласия? Должны ли некоторые медицинские данные рассматриваться как своего рода общественное благо и быть собраны для некоторых потенциально публичных целей?

Этический анализ в значительной степени будет зависеть от типа данных, их идентификации, доступа к ним и целей доступа. Можно по-разному относиться к Центру контроля заболеваний (Center for disease control – CDC), имеющему доступ к данным для отслеживания гриппа, или к больничной системе, использующей данные для переоценки своего персонала и рабочего процесса, или к фармацевтической компании, применяющей данные для разработки продукта. Даже если нарушается неприкосновенность частной жизни, возможно, учитывая все обстоятельства, нарушение в

---

<sup>1</sup> National Institutes of Health: All of us research program. – Mode of access: <https://allofus.nih.gov/about/about-all-us-research-program>

некоторых случаях перевешивается справедливым распределением выгод (принцип взаимности), а там, где риски для пациента низки, а интерес пациентов невелик по сравнению с выгодой, – принципом пропорциональности.

Даже если пациенты согласны с тем, чтобы их данные использовались, какую роль они должны играть при принятии решения о том, какие виды применения их данных допустимы? Авторы рассматривают два варианта ответа: 1) максимизация автономии пациента и получение его согласия на каждый доступ; 2) отчуждение деидентифицированных данных от пациента, лишение его права контроля над ними. Однако, опираясь на контекстуальное понимание конфиденциальности, они утверждают, что во многих случаях оптимальный режим управления может находиться где-то посередине. Например, решение может принимать руководящий совет, включающий представителей пациентов; другой вариант – передача данных в благотворительный фонд, при этом попечители (некоторые из них будут представителями пациентов) принимают решения об условиях доступа и разрешенных видах использования. Еще один подход, который Б. Эванс, профессор права в Университете Хьюстона, определяет как «общедоступные данные, ориентированные на потребителя», предполагает сбор БД группами людей по обоюдному согласию для использования в научных медицинских исследованиях на условиях, которые установят сами члены группы [с. 40]. По мнению авторов, проблемы с использованием данных будут возникать и в будущем, когда разные наборы БД будут применяться для разных целей, что затруднит установление одного режима управления для всех данных человека.

Использование БД, полученных от пациентов, может привести к дискриминации: если работодатели или страховые компании узнают важную информацию о человеке из медицинских данных, они могут принять решение не нанимать или не страховать его. Существует точка зрения, согласно которой этот тип дискриминации оправдан в соответствии с принципом «актуарной справедливости», когда каждый должен платить или получать оплату в соответствии со своими рисками. Возникает очень сложный вопрос о том, следует ли отдавать предпочтение понятию «каждому в соответствии с его риском», а не более солидарному взгляду на страхование, в котором происходит некоторое перераспределение рисков через объединение страховых взносов. Американское законодательство в области медицинского страхования запрещает некоторые, но не все виды дискриминации. Но сущест-

вует сопутствующий вред, который трудно устранить с помощью закона, в частности стигма, которая может возникнуть из-за того, что другие знают об инфекции, передаваемой половым путем, у данного человека, или узнают, что родитель данного ребенка не является его биологическим родителем [с. 41].

Специфика генетической информации заключается в том, что она затрагивает интересы генетических родственников, которые могут не давать согласия на обмен данными. В то же время генетические данные базы GEDmatch позволили полиции штата Калифорния поймать преступника, действовавшего в период с 1976 по 1986 г.<sup>1</sup>

Одна из реакций на нарушение конфиденциальности в отношении здоровья – резкое ограничение доступа к данным пациентов. Если деонтологические и косвенные проблемы трудно уменьшить постфактум, превентивное ограничение БД кажется привлекательным решением. Но чрезмерная защита конфиденциальности может стать тормозом, а фрагментация медицинских данных может создавать как технологические, так и экономические препятствия для инноваций. Во многих случаях компромисс между приватностью и инновациями будет существовать и дальше.

Однако существует множество инноваций, секретность которых из-за коммерческих интересов противоречит интересам всего общества. Так, компания Myriad Genetics поддерживает собственную базу данных генетических последовательностей и историй болезней женщин, которые тестировались на предрасположенность к раку молочной железы и яичников (BRCA1 и BRCA2). Но если бы общество имело доступ к этим данным, другие компании могли бы их перепроверять и улучшать тесты.

Еще одна проблема, на которую обращают внимание авторы статьи, – подрыв доверия к нововведениям в области БД вследствие секретности алгоритмов их обработки и непрозрачности алгоритмов машинного обучения. Конечно, есть много медицинских процессов, внутренняя работа которых скрыта коммерческой тайной и недоступна для пациентов, но внимание средств массовой информации и новизна БД и ИИ могут вызывать особенное напряжение из-за их интеграции в медицинскую помощь.

---

<sup>1</sup> May T. Sociogenetic risks – ancestry DNA testing, third-party identity, and protection of privacy // The new England journal of medicine. – 2018. – Vol. 379, N 5. – P. 410–412. – DOI: 10.1056/NEJMp1805870.

В заключение авторы отмечают, что как недостаточная, так и чрезмерная защита конфиденциальности может создавать ощутимый вред для пациентов. Достижение правильного баланса, обеспечивающего доверие к потенциальным преимуществам БД, является важной задачей, решение которой зависит от типа данных и того, кто ими владеет и кто хранит, а также от возможности сопоставления данных из разных источников.

*Е.Г. Гребенщикова*

**Эдвардс С.**  
**ПОСТОЯННОЕ СОГЛАСИЕ**  
**ИЛИ ОСТОРОЖНЫЙ КОНТРОЛЬ**  
**ЗА ИСПОЛЬЗОВАНИЕМ ЦИФРОВЫХ ДАННЫХ?**  
**(Реферат)**

Ref. ad op.: Edwards S. Continuous consent to, or discreet control over digital data? // Research ethics. – 2014. – December, Vol. 10, N 4. – P. 184–186. – DOI: 10.1177/1747016115575892

Доктор Сара Дж.Л. Эдвардс – специалист в области исследовательской этики, сотрудник Университетского колледжа Лондона. В редакционной статье, предваряющей специальную подборку, она отмечает, что компьютерные технологии открывают новые возможности для науки и здоровья, а мощные средства анализа больших данных (БД) – новые возможности и для персонализированного здравоохранения. Таким образом, они несут пользу всем и каждому. Однако постоянно возникает важная проблема частной жизни, которая не поддается простому решению без ущерба для тех или иных групп интересов.

Политики из Министерства здравоохранения Великобритании подошли к этой проблеме с двух сторон. Для того чтобы более успешно способствовать развитию науки и инновациям, они предложили совместное использование разными группами исследователей идентифицируемых данных о здоровье. Однако выполнение этого предложения было приостановлено, после того как в 2014 г. правительство Великобритании пришло к выводу, что общественность не получила адекватной информации о нем. Это противоречило первому традиционному принципу защиты данных – принципу «справедливой обработки» или прозрачности

использования. Неудача в информировании означает, что возможность отказа оказывается бессмысленной.

Вместо того чтобы отказываться от этого предложения, можно сделать более привлекательными для отдельного пациента возможности принятия цифровой технологии, пообещав ему онлайн-доступ к касающимся его записям и использование новых персонализированных средств. Однако принудительные блага от онлайн-доступа ничего не дают для защиты интересов пациентов против несанкционированного использования их персональных данных.

У защитников частной жизни сегодня есть два повода для усиливающегося беспокойства: один давно знакомый, а другой, очевидно, новый, порождаемый недавним расцветом телемедицины и удаленного мониторинга.

Во-первых, используя схему, исключающую отказ, каждый должен будет принять все возможные способы использования его персональных данных (ограничиваемые только назначенным посредником – Информационным центром национального здоровья и социального ухода). Считается, что европейские директивы<sup>1</sup> угрожают этой позиции с двух сторон. Европейская комиссия не одобряет исследователей, которые извлекают пользу «по умолчанию», ценой активного и выраженного участия пациентов. В любом случае комиссии не нравится, что любое активно выражаемое согласие может быть настолько общим или «широким», чтобы покрыть все устремления исследователей, не нарушая при этом имплицитных социальных норм и персональных ожиданий. Однако настаивать на том, чтобы каждый пациент давал разрешение на каждое специфическое использование персональных данных, – значит перейти через прагматический предел в том, о чем исследователи могут просить пациентов.

Во-вторых, нынешнее всеобщее использование мобильных приборов, а также все более широкое применение телемедицины ставят перед пуристами – защитниками частной жизни (к числу которых можно отнести и Европейскую комиссию) – новые трудности. Персональные данные, такие, например, как показатели кровяного давления, могут теперь обрабатываться непрерывно и дистанционно в реальном времени, что делает выраженное и спе-

---

<sup>1</sup> European Commission. Commission proposes a comprehensive reform of the data protection rules. – 2012. – Mode of access: [http://ec.europa.eu/justice/newsroom/data-protection/news/120125\\_en.htm](http://ec.europa.eu/justice/newsroom/data-protection/news/120125_en.htm) (Accessed: 12 February 2015.)

цифическое (выбор за применение) согласие затруднительным как с концептуальной, так и с прагматической точки зрения. Можно попытаться на концептуальном уровне «втиснуть» постоянное разрешение, данное пациентами, что, по-видимому, требуется для защиты частной жизни из-за рисков, вызванных этими технологиями, в традиционное дискретно выраженное согласие (выбор за) с постоянной возможностью отказа (или выбора против). Возможно, будет только один период, в течение которого пациент захочет держать в тайне свое состояние. Это могут быть данные, связанные со стрессовым состоянием при какой-то форме деятельности, например возрастание кровяного давления. Но поскольку сбор данных может происходить в реальном времени, бремя планирования наперед для защиты частной жизни может лежать на пациенте, заранее делающем выбор против; в зависимости от технологии отказ может оказаться более сложным, чем простое нажатие кнопки выключения. Кроме того, многие мобильные приборы и онлайн-сервисы вообще не дают пользователю какого-либо выбора при обработке их цифровых данных, что затрудняет использование таких технологических достижений без отказа от всех претензий на защищенную частную жизнь.

Поэтому необходимо сохранять возможности не только использования новых цифровых технологий для науки и пациента, но и защиты частной жизни, например с помощью создания отдельного законодательства для каждого вновь возникающего повода для дискриминации против прав человека или его значимых интересов. Исходя из этого правительство Великобритании стремится стимулировать доверие общественности к более широким исследовательским акциям. Принятые защитные меры призваны обеспечивать такое положение дел, чтобы доступ к идентифицируемым данным пациентов имели только те исследователи, которые работают в более или менее технологически безопасных и социально аккредитованных средах, в «надежных местах»<sup>1</sup>. Таких мер, однако, может быть недостаточно для защиты общественного интереса. Ответом могли бы быть новые технологии. Например, система, разработанная в Массачусетском технологическом институте, позволяет индивидам не только проследить, кто использует их данные и для каких целей, но и

---

<sup>1</sup> Data in safe havens // Academy of medical sciences. – 2014. – March. – Mode of access: <http://www.acmedsci.ac.uk/policy/policy-projects/data-in-safe-havens>

накладывать ограничения на доступ к ним. Пациентам предлагается много возможностей для контроля над тем, кто имеет доступ к их данным и образцам тканей. Для этого им нужно зарегистрироваться на веб-сайте исследования и отслеживать все запросы на доступ. В то же время исследователи понимают, что система работает только тогда, когда поддерживается доверие пациентов, поэтому именно на них ложится обязанность мудро пользоваться данными.

*Б.Г. Юдин*

# **БОЛЬШИЕ ДАННЫЕ В СОЦИАЛЬНЫХ И ГУМАНИТАРНЫХ НАУКАХ**

## **Сборник обзоров и рефератов**

Оформление обложки И.А. Михеев  
Компьютерная верстка Н.В. Афанасьева  
Корректор Я.А. Кузьменко

Гигиеническое заключение  
№ 77.99.6.953. П. 5008.8.99 от 23.08.1999 г.  
Подписано к печати 20/VI – 2019 г. Формат 60x84/16  
Бум. офсетная № 1. Печать офсетная  
Усл. печ. л. 12,0 Уч.-изд. л. 10,0  
Тираж 300 экз. (1–100 экз. – 1-й завод)  
Заказ № 16

**Институт научной информации  
по общественным наукам РАН,  
Нахимовский проспект, д. 51/21,  
Москва, В-418, ГСП-7, 117997**

**Отдел маркетинга и распространения  
информационных изданий  
Тел. / Факс: (925) 517-36-91  
E-mail: [inion@bk.ru](mailto:inion@bk.ru)**

Отпечатано по гранкам ИНИОН РАН  
ООО «Амирит»  
410004, Саратовская обл., г. Саратов  
ул. Чернышевского, д. 88, литера У