

УДК 316.77

DOI: 10.31857/S032120680004360-0

«Журналистика данных» в США

Д.В. Неренц

Институт массмедиа РГГУ. Российская Федерация, Москва
ORCID: 0000-0002-6370-0157 e-mail: ya.newlevel@yandex.ru

Статья поступила в редакцию 22.01.2019.

Резюме: С развитием интернета и новых технологий каждый человек ежедневно сталкивается с большим количеством информации, которую он не в силах систематизировать и обобщить. Вследствие этого особой популярностью сейчас пользуются новостные ленты, представляющие собой максимально короткое сообщение о событии без лишних деталей. Однако эти новости не всегда публикуют профессиональные журналисты. Безусловно, сейчас очевидец или участник события может обнародовать любые данные гораздо быстрее репортёра, который потратит время на проверку, сбор и обработку информации. Очевидно, что основная функция журналистики начинает меняться: от информирования к анализу и обработке данных. Крупные американские СМИ делают акцент не на скорости публикации материалов, а на их качестве. Так появилось новое направление, журналистика данных (дата-журналистика), в которой представлены материалы, основанные на анализе больших массивов данных. Задача современного дата-журналиста – не просто опубликовать свой материал, а предоставить аудитории все исходные данные в их первоначальном виде и оформить их в визуально привлекательной и упрощённой форме. Статья посвящена истории появления и развития дата-журналистики как современного направления, которое демонстрирует смещение целей и задач журналиста в сторону аналитической работы и работы с компьютерным программным обеспечением.

Ключевые слова: журналистика данных, дата-журналистика, большие данные, исследовательская журналистика, визуализация, инфографика

Для цитирования: Неренц Д.В. «Журналистика данных» в США. США & Канада: экономика, политика, культура, 2019;49(4):75-91. DOI: 10.31857/S032120680004360-0

Data Journalism as a New Trend of Journalism in the USA

Darya V. Nerentz

Institute of Massmedia of Russian State University for the Humanities
Moscow, Russian Federation

ORCID: 0000-0002-6370-0157 e-mail: ya.newlevel@yandex.ru

Abstract: The development of the Internet and new technologies has contributed a lot of information, and now many people cannot systematize and generalize this information at all. As a result, news feeds became very popular because they are representing the shortest message

without unnecessary details. However, professional journalists do not always publish this news. Of course, now an eyewitness or an event participant can public any data much faster than a reporter who spends time for checking, collecting and processing information. Therefore, it is obvious that the main function of journalism is beginning to change: from informing to analysis and data processing. Major US media do not focus on the speed of publication of materials, but on their quality. Therefore, it has appeared a new trend, data journalism, and a journalistic material became based on the analysis of large data sets. The task of a modern data journalist is not to publish his material based on data, but give the audience opportunity to see the initial data in its original form and then they should arrange this data in a visually attractive and simplified form. This article is devoted to the history of the emergence and development of data journalism as a modern trend in American journalism, which demonstrates a shift in the goals and objectives of a journalist towards analytical work and work with computer software.

Keywords: data journalism, data-driven journalism, big data, investigative journalism, visualization, infographics

For citation: Nerentz D.V. Data-journalism as a New Trend of Journalism in the USA. *USA & Canada: Economics, Politics, Culture*. 2019; 49(4):75-91.

DOI: 10.31857/S032120680004360-0

ВВЕДЕНИЕ

Обилие информации, с которой ежедневно сталкивается аудитория, всё чаще приводит к так называемой информационной перегрузке, вследствие чего читатель или зритель практически не может выбрать и обработать необходимые ему сведения, проигнорировав лишнее. Эта проблема стала первым шагом к новым требованиям, предъявляемым журналистам со стороны аудитории. Одно из основных – выделить, обработать и передать в доступной форме самое важное из ежеминутного потока сведений. Сложившаяся ситуация привела к всплеску интереса к аналитическим материалам, лонгридам, журналистским расследованиям, работа над которыми подразумевает обработку большого количества информации или «больших данных».

Использование в материале «больших данных» предоставляет аудитории, прежде всего, возможность увидеть картину целиком, не обрывочные сведения или только главные факты, а именно всю информацию с мельчайшими деталями. Так аудитория имеет возможность изучить проблему не на основе того, что выбрал журналист для демонстрации своим читателям или зрителям, а на основе всех фактов, которые удалось получить.

Данными журналисты пользуются всегда, примеры инфографики, созданной на основе данных, можно обнаружить в любых СМИ, в том числе в изданиях XIX века. Однако эти данные, как правило, представляли собой графики с числовыми показателями неизвестного происхождения. Другими словами, читатель не смог бы выяснить, откуда эти цифры, поскольку отсутствовал исходный материал, на основании которого были получены результаты. На современном

этапе новые технологии и большое количество информации в открытом доступе предполагает максимальную объективность и открытость в работе с «большими данными» со стороны журналистов, поскольку теперь эту информацию стало возможно проверить, обратившись к альтернативным источникам.

Эти тенденции породили новое направление журналистики, которое можно обозначить как «дата-журналистика», или «журналистика данных».

Известный журналист и блоггер Пол Брэдшоу, один из авторов «Пособия по журналистике данных» в предисловии к изданию очертил понятие следующим образом: «Что же отличает журналистику данных от остальной? Вероятно, это новые возможности, которые открываются, когда вы объединяете традиционное «чутье на новости», умение выведать всё, что случилось, и способность рассказать захватывающую и наглядную историю с огромным количеством разнообразной цифровой информации, которая теперь стала доступной» [1].

Американские исследователи и авторы, представленные в «Пособии по журналистике данных», выделяют следующие отличительные особенности этого направления: «...это новые возможности, которые открываются, когда вы совмещаете традиционный “новостной текст” и способность рассказывать интересную историю с широким масштабом и диапазоном цифровой информации, доступной сейчас. И эти возможности могут использоваться на любом этапе: с помощью программирования для автоматизации процесса сбора и объединения информации из местных органов власти, полиции и других гражданских источников, как это сделал Адриан Холовати с “ЧикагоКрайм”, а затем “ЭвриБлок”» [1].

Одними из первых к визуализации «больших данных» прибегли американские журналисты, в частности журналисты-расследователи. Благодаря интернет-источникам процесс сбора информации расследователи смогли упростить до максимального уровня. Помимо активного вовлечения пользователей в работу над анализом документов и интервью, они все чаще обращались к компьютерным базам данных, в которых (особенно в США с установкой на полную прозрачность всей деятельности) содержалось большинство документации и многочисленные отчёты компаний и правительственных секторов. Эти «большие данные» стали главным источником информации в журналистской среде. К ним можно отнести открытые данные в Интернете, базы данных, оцифрованные архивные материалы, результаты в поисковых системах, даже данные в социальных сетях. Этот источник информации позволяет составить детальную картину нынешней обстановки в городе, регионе, стране и даже в мире. Статистические сведения, числовые показатели, оформленные в виде яркого и простого графика или схемы, позволяют, с одной стороны, отразить серьёзность проблемы, с другой – привлечь внимание аудитории за счёт доступной формы подачи информации.

Так можно сделать вывод, что «журналистика данных» («дата-журналистика») – это новое направление, основу которого составляет сбор, об-

работка и систематизация «больших данных» и последующая привлекательная визуализация полученной информации.

ЭТАПЫ РАЗВИТИЯ

Понятие «журналистика данных» сочетает в себе два слова, которые сами по себе являются расплывчатыми терминами, но имеют важное значение для демократического общества. Несмотря на расхождения в значении терминов «компьютерная журналистика» (*computer-assisted reporting*, или CAR) и «журналистика данных», суть их и цель похожи. Первый пример «компьютерной журналистики» восходит к 1952 г., когда цель заключалась в том, чтобы предсказать результаты выборов в США. В то время это было политически спорным новым явлением, включающим оцифровку информации.

Однако история появления термина «журналистика данных» несколько сложнее. В зависимости от методики работы и целевых установок дата-журналистов в разных условиях функционирования СМИ в развитии дата-журналистики можно выделить три основных этапа.

Первый этап: с момента появления (1950-е годы) до начала 1990-х годов. В этот период репортёры создавали электронные таблицы для обобщения и систематизации большого количества данных, чтобы выявить взаимосвязи и сделать выводы из полученной информации. «Несмотря на то что определить конкретные сроки появления этого понятия достаточно сложно, эксперты и сами журналисты чаще всего отталкиваются от двух значимых дат в смежных областях, без которых появление «журналистики данных» было бы невозможно. Первая – 1952 год, когда в работе журналиста впервые отмечено использование компьютера... Вторая дата – 1967 год. Именно тогда стал очевиден анализ данных, потому что Филип Майер из «Детройт фри пресс» использовал компьютер для систематизации полученных в ходе опроса жителей Детройта ответов для того, чтобы выяснить и объяснить причины серьёзных беспорядков, которые разразились в городе тем летом» [2]. Проведя опрос бунтовщиков и составив сводную таблицу в электронном виде, Ф. Мейер доказал, что среди вышедших на протест людей, окончивших колледж, было не меньше, чем жителей, не окончивших среднюю школу.

Ф. Майер продолжал использовать этот метод и в 1970-х годах, в частности сотрудничая с репортёрами «Филадельфия инквайер» Дональдом Барлеттом и Джеймсом Стиллом, а также с Ричем Морином из «Майями геральд» для анализа записей оценки имущества. Позже он написал книгу под названием «Прецизионная журналистика», в которой призывал коллег к активному использованию анализа баз данных и методов социальных исследований в публикациях.

Стоит отметить, что именно Дональд Барлетт и Джеймс Стил, занимающиеся журналистскими расследованиями, стали активно использовать электронные таблицы для обобщения большого количества данных из отчётов и судебных

решений. Благодаря такому методу работы они смогли сделать то, с чем другие никогда бы не справились из-за обилия документов и невозможности обработки всей информации вручную. Сами журналисты, вспоминая своё первое задание с «большими данными» указывали, что редактор газеты попросил их проанализировать и оценить работу Федеральной администрации жилищного строительства на предмет незаконных операций при реализации программы капитального ремонта старых домов. В результате длительной работы, журналисты обнаружили, что у Министерства жилищного строительства и городского развития возникли проблемы с выкупом закладных. Малообеспеченные семьи отказывались покупать дома из-за некачественного ремонта, который никак не отразился на «аварийности» домов. В течение двух месяцев Д. Барлетт и Дж. Стил изучали всевозможные жилищные акты и закладные, проводили интервью с жильцами, выясняли, как работают спекулянты, и кто на самом деле покупает эти дома. В итоге была опубликована серия статей, из-за которой агентства по недвижимости отозвали свою рекламу и прекратили сотрудничество с газетой.

В 1974 году Д. Барлетт и Дж. Стил обратили внимание на «налоговую» проблему в Америке. Проанализировав годовые отчёты Налоговой службы более чем за десятилетие, они обнаружили, что претензии по уплате налогов к крупным конгломератам значительно меньше их реальной задолженности и смогли доказать неравномерность уплаты налогов в американском обществе. Согласно их расследованию, средний и низшие классы американцев исправно отдавали положенные им налоговые выплаты, в то время как представители богатых слоёв общества платили гораздо меньше того, что на самом деле были должны. При анализе «больших данных» они применили хронологический принцип: при сопоставлении информации, полученной из различных источников, в хронологическом порядке они получили возможность увидеть полную картину. В итоге за серию расследований «Аудиторская налоговая служба» в 1975 г. исследователи получили первую Пулитцеровскую премию. Именно тогда технологию работы с «большими данными» Дональда Барлетта и Джеймса Стила стали постепенно применять другие репортёры.

Пять лет спустя они получили вторую Пулитцеровскую премию за разоблачение деятельности Конгресса по предоставлению мультимиллионерам льгот и освобождений от уплаты налогов. Журналисты получили список из 650 юридических и физических лиц, которые попали под «временные правила» закона о совершенствовании налогообложения, сформулированные комитетом по финансам. Они сравнили этот список с положениями 900-страничного закона для подтверждения того, что список корпораций, имеющих право на льготы, исчерпан. Барлетт и Стил обнаружили тысячи организаций и физических лиц, которым неофициально предоставлялись налоговые льготы [3].

В итоге, каждая статья была посвящена одному или нескольким крупным и влиятельным лицам, которым были неправомерно предоставлены налоговые льготы. В отличие от многих журналистских расследований эта серия привела к

некоторым практическим переменам. Публикации вынудили Конгресс отклонить пункты реформы, которые включают специальные налоговые льготы для политиков и бизнесменов.

За время работы журналисты провели более десяти серьёзных крупных расследований, касающихся бизнес- и политструктур, налоговых реформ, фармацевтического бизнеса, злоупотреблений властью со стороны чиновников. Д. Барлетт и Дж. Стил смогли показать расследователям, как значительно сэкономить временные ресурсы, используя компьютерные базы данных при обработке документов, и сами смогли систематизировать и изучить огромное количество документов различной направленности, включая судебные протоколы, отчёты о заседаниях компаний, финансовые отчёты. Они обозначали собой новый этап в методологии работы журналистов.

Несмотря на явные преимущества такого подхода, журналисты других СМИ редко использовали базы данных для выявления причинно-следственных связей. Переломным моментом стал материал Эллиота Джаспина для «Провиденс джорнел бюллетень» в середине 1980-х годов. Эта публикация вызвала большой общественных резонанс, поскольку Э. Джаспин смог детально описать проблемы большого количества неблагонадёжных водителей школьных автобусов и историю политического скандала с привлечением займов на жильё исключительно благодаря анализу баз данных. В 1989 г. американские журналисты признали ценность систематизации и анализа данных с использованием компьютеров, поскольку этот метод помог «Атланта джорнел конститиьюшн» получить Пулитцеровскую премию за публикации о расовой дискриминации при одобрении/отказе ипотечного кредитования.

Активное использование такого метода работы стало очевидным, когда в 1990 г. профессор Университета штата Индиана Джеймс Браун в Индианаполисе провёл первую конференцию по использованию баз данных в работе журналиста.

Первый этап развития характеризуется новым методом анализа и обработки большого количества информации. Такой метод работы позволил журналистам не только существенно снижать временные затраты на анализ информации, но и выявлять скрытые ранее факты, подтверждающие совершение преступления. Стоит отметить, что «большие данные» на этом этапе применялись только при проведении журналистских расследований, для выявления правонарушений, а социально значимые проекты и исследования в тот период в редакциях СМИ не практиковались.

Временные рамки *второго этапа* развития можно обозначить как: начало 1990-х – 2010-е годы. Он характеризуется использованием в качестве основного источника информации электронные базы данных. Журналисты по-прежнему анализировали большое количество документов в печатном виде, однако с активным развитием Интернета и переводом информации в цифровую форму данных стало в десятки или сотни раз больше и появились специальные базы

данных по различным областям, которые стали главным источником информации. Так, практика появления «журналистики данных» потребовала от журналистов определённого уровня подготовки, поскольку освоить схему работы самостоятельно было крайне трудно. В связи с этим в США стали появляться центры по обучению журналистов работе с большими массивами данных в электронном виде.

Одним из первых таких центров был Национальный институт компьютерной помощи (НИКП), созданный в 1994 г. директором по тренингам Дженнифер Лафлер и Брантом Хьюстоном. Целью создания отдела была обучающая программа продолжительностью до 50 семинаров в год. К 1996 г. успехи американских журналистов заставили иностранных репортёров посещать интенсивные недельные семинары в НИКП.

Благодаря созданным тренингам журналисты из Финляндии, Швеции, Новой Зеландии, Венесуэлы, Аргентины, Нидерландов, Норвегии, Бразилии, Мексики, России, Боснии и Канады смогли увидеть, как использовать данные международных или американских ресурсов, а также как получить данные и проанализировать их касательно их страны.

Между тем, в Лондоне в 1997 г. профессор журналистики Мильвертон Уоллас начал проводить ежегодную конференцию под названием «НетМедиа», на которой участникам предлагались занятия по работе с компьютерными отчётами под руководством специалистов из НИКП и датских журналистов. Тренинг охватывал основные виды использования Интернета, электронных таблиц и баз данных. Такие занятия часто посещали журналисты из США, европейских стран и Африки.

«С начала XXI века использование информации, основанной на работе с электронными массивами данных расцвело, в основном благодаря семинарам, проводимым в Миссури и во всём мире Содружеством репортёров и редакторов (СРР) [4] и НИКП. В первое десятилетие XXI века Глобальная сеть журналистов-расследователей стала играть решающую роль в этом движении, начиная с первой конференции в 2001 г. в Копенгагене, в ходе которой был подготовлен серьёзный доклад о подготовке и практическом обучении журналистов с помощью компьютера» [2].

В 2001–2002 гг. формируется понятие «Биг Дата» (большие объёмы данных), завязанное на увеличении объёмов, типов и скорости поступления новых данных, о чём (ещё не употребляя термин «Биг Дата», но уже формулируя все его основы) впервые говорит исследовательская компания «Гартнер» в 2001 г. Специалист по сбору данных компании Гугл С. Стивенс-Давидовиц указывает, что «большие данные» очень расплывчатое понятие, поскольку никто не может ответить, с какого конкретно количества данных они считаются «большими». Однако он отмечает, что «мы переживаем взрывной рост количества и качества различных видов доступной информации. Новые потоки влились через Гугл и социальные сети. Некоторые из них – продукт оцифровки информации, кото-

рая раньше была спрятана в шкафах и папках, другие получены в результате увеличения ресурсов, выделяемых на маркетинговые исследования» [Стивенс-Давидовиц С., 2018: 27]. Подобное мнение высказывает и профессор Калифорнийского университета А. Вайгенд: «Социальные данные – самое ценное сырьё XXI века, новая нефть... К счастью, информация как ресурс радикальным образом отличается от нефти. Запасы нефти на планете конечны, и по мере истощения этого ресурса его эксплуатация обходится всё дороже. Количество же информации, напротив, растёт в геометрической прогрессии, а стоимость технологий её передачи и обработки неуклонно снижается... использование информации практически не ограничено, учитывая её изобилие и цифровую форму» [Вайгенд А., 2018: 32–33].

Именно на этом этапе «большие данные» стали использовать в редакциях СМИ повсеместно, не только в расследовательских материалах, но и в социальных проектах, так называемых новостных приложениях. Американские журналисты анализировали базы данных о состоянии государственных школ, больниц, приютов для создания целостной картины, которая помогла бы аудитории сделать выводы о том, как на самом деле обстоит дело в той или иной общественной сфере в их городе или стране в целом. Крупные СМИ активно применяют журналистику данных, создавая таблицы, сверяя данные и публикуя полученную информацию через истории конкретных людей.

Третий этап начался в 2010 г. и продолжается до сих пор. Именно в 2010 г. состоялась конференция, где и было обозначено появление этого отдельного направления в журналистике. Первая международная конференция, посвящённая журналистике данных, прошла в Амстердаме, после чего это направление журналистики считается официально сформированным. «Начиная со второй половины 2010 г. понятие журналистики данных уже активно используется как на конференциях и в академических журналах» [2].

Интересно также, что в 2010–2011 гг. в редакциях СМИ стали проводить так называемые «хакатоны»*, представляющие собой многочасовые «мозговые штурмы» команды из журналистов, программистов и хакеров. Их взаимное общение и передача опыта позволили СМИ понять принципы работы в дата-журналистике и начать организовывать собственные базы данных для удобства журналистов и читателей. «Хакатоны» проводили журналисты из Нидерландов, Аргентины, Финляндии. Главной идеей мероприятия стал тезис о том, что разработчики, которые знают, где искать информацию, не могут придать ей социальной значимости. Журналисты, в свою очередь, зная о возможностях журна-

* Англ. *hackathon*, от *hack* – хакер и *marathon* – марафон) – это возможность в течение нескольких часов создать медийный проект или рабочий прототип программного обеспечения для решения какой-либо проблемы. Журналисты, дизайнеры, разработчики, объединяясь в одну команду, используют свои знания и практические навыки для создания уникального продукта.

листки данных, не могут представить их в привлекательной форме и не знают, как эти данные обнаружить.

Пик проведения подобных мероприятий в США пришёлся на 2015 год, когда сразу несколько крупнейших университетов и редакций СМИ стали объявлять набор на участие в «хакатонах». Например, в ноябре 2015 г. на базе Колумбийского университета в Нью-Йорке прошёл «хакатон» «Данные + журналистика». Студенты, заинтересованные в работе с данными, объединялись в команды по четыре человека и демонстрировали свою лучшую визуализацию информации.

В этом же году Остинская школа журналистики на базе Техасского университета организовала восьмичасовой «хакатон», где могли поучаствовать практикующие журналисты и студенты для обмена опытом и получения информации от дизайнеров и разработчиков по визуальным инструментам и оформлению своих данных.

В июле 2015 г. в Вашингтоне при участии газеты «Вашингтон пост» был организован масштабный «хакатон», в котором приняли участие журналисты, редакторы, владельцы средств массовой информации, разработчики, исследователи данных и креативные технологи для создания новых продуктов и инструментов при использовании данных в ньюсрумах.

Такое взаимовыгодное сотрудничество привело к активному распространению журналистики данных по всему миру, и теперь в любых крупных СМИ есть хотя бы один материал, сделанный с помощью данных и визуальных инструментов.

За последние 20 лет такие журналистские организации, как Национальный институт компьютерной помощи и Датский международный центр аналитической отчётности, были созданы исключительно для того, чтобы содействовать использованию компьютерных технологий при создании новостного контента. Многие другие организации – Общество профессиональных журналистов, Канадская ассоциация журналистов и Университет Королевского колледжа в Галифаксе – предлагают обучение или семинары по работе с программным обеспечением.

Таким образом, три этапа, выделенные на основе принципов работы с данными указывают, что дата-журналистика появилась сравнительно недавно, но за небольшой промежуток времени стала активно развиваться и превратилась в отдельное направление, которое сосредоточено на анализе и обработке больших объёмов данных.

ХАРАКТЕРИСТИКА И ПРИНЦИПЫ РАБОТЫ

Журналистика данных представляет собой весьма специфическое направление, работа в котором предполагает наличие большого количества навыков и умений и в том, что касается определённого подхода к изучению информации,

и в том, что необходимо обладать знаниями специализированных компьютерных программ.

Главная задача дата-журналиста – найти и трансформировать «большие данные» в визуально привлекательную форму, помогая аудитории сформировать собственное мнение и понять важность проблемы. «Вместо того, чтобы просто перерабатывать пресс-релизы и пересказывать истории и сюжеты, которые они где-то слышали, журналисты, работающие в секторе журналистики данных, могут давать читателям ясную, всеобъемлющую и настраиваемую по предпочтениям перспективу или точку зрения, снабженную интерактивной графикой, а также прямой доступ к первичным источникам» [5].

Таким образом, для успешного создания дата-материала необходимо умение работать с большим набором данных, понимать схемы составления и группировки полученных данных, умение эти данные объединить и истолковать, а также анализировать статистическую информацию.

Анализ и понимание данных – это процесс, отнимающий много времени, даже при наличии всех необходимых инструментов. «Подогнать его под новостной цикл, который, как известно, очень короткий, сложно, поэтому журналистика данных часто используется в долгосрочных, долгоиграющих исследовательских или расследовательских проектах» [6].

Проведение исследования или расследования, как правило, способствует накоплению большого количества информации, относящейся к выбранной теме. И в последние несколько лет журналисты пришли к выводу, что все эти данные, полученные ими в ходе работы, должны быть доступны аудитории не только в виде готовой публикации, но и в виде исходного материала. Теперь крупные традиционные СМИ и новые медиа, занимающиеся аналитикой и исследованиями, создают собственные базы данных на основе собранной за месяцы и даже годы информации. Например, своя большая база данных есть у Центра исследования коррупции и организованной преступности, которая называется «Панель данных». Это универсальный инструмент для международных расследований коррупции и организованной преступности, который содержит в себе ссылки на 750 открытых существующих баз данных из 120 стран. Причём сам ресурс представляет собой поисковую машину, отличительным достоинством которой является возможность поиска информации по заголовкам, что довольно удобно, учитывая, что база насчитывает более 97 миллионов заголовков статей из различных источников.

Кроме глобальных ресурсов по самой разнообразной информации, есть некоммерческие расследовательские организации, которые создают свои базы данных по результатам проведённых расследований. Так у «Международного консорциума журналистов-расследователей» (МКЖР) есть база данных, посвящённая архивным данным по документам, составляющим «Панамское досье» и «Архивы Райских островов». База данных поделена на рубрики: «Офшор ликс», «Панама пейперс», «Багамас ликс» и «Парадайс пейперс».

Ещё один некоммерческий центр, имеющий свою базу данных – «Пропаб-лика». Их данные поделены на несколько рубрик (от здравоохранения до образования) и классифицируются как премиум-базы (платные) и обычные (бесплатные). Для доступа к базе необходимо заполнить небольшую регистрационную форму.

База данных американского издания «Лос-Анжелес таймс» представляет собой выборку всей существующей информации по определённой теме. Среди тематических рубрик есть, например, «Калифорнийские пожары» или «Заболевания учителей в Лос-Анжелесе».

Помимо использования в качестве главного источника информации баз данных, журналистика данных как отдельное направление, имеет также ряд отличительных особенностей. «В традиционной журналистике из-за линейного характера печатных и вещательных СМИ нам приходится думать о начале, концовке, сюжетной линии, длительности сюжета и о том, под каким углом зрения смотреть на проблему. В журналистике данных всё по-другому. Тут есть начало, да. Люди заходят на веб-сайт и получают первое впечатление от интерфейса. Но потом они начинают действовать сами по себе. Они могут задержаться на сайте на минуту, а могут на полчаса. Наша задача как журналистов, работающих с данными, заключается в том, чтобы предоставить текущую конструкцию, основу, каркас, окружающую среду, условия для этих действий наших читателей/пользователей» [7]. Журналист, строящий свой материал на основе «больших данных», всегда обращает внимание аудитории на все исходные найденные данные, в связи с этим сгруппированные и обработанные факты изложены самым простым и логичным способом, как правило, с помощью графического интерфейса.

Таким образом, медиаисследователями были сформулированы некоторые теоретические основы работы в журналистике данных. Помимо принципов работы, в 2011 г. Пол Брэдшоу, основатель проекта *HelpMeInvestigate.com*, создал модель, которую назвал «перевернутой пирамидой журналистики данных» [8]. Согласно этой модели, для получения эффективного результата при работе с большими данными необходимо поэтапно выполнить ряд действий.

Прежде всего, следует обратиться ко всей уже имеющейся и общедоступной информации по выбранной теме. Данные можно получить непосредственно из правительственных баз данных, таких как *data.gov*, а также путём подачи запросов в государственные учреждения. Несмотря на то, что прослеживается всеобщая тенденция к открытым данным, существуют национальные различия в отношении того, насколько эта информация доступна для граждан. В США персональными данными является только номер социального страхования, а данные о месте жительства, недвижимости, заработной плате, дате рождения и т.д. являются открытыми, что делает условия поиска гораздо проще, нежели в ряде европейских стран.

Следующий этап – фильтрация и систематизация данных, подготовка к визуализации. Отбор и систематизация данных относится к использованию определённого программного обеспечения. Именно создание таблиц в электронном виде – основной инструмент журналиста, работающего с большим количеством информации. В целом набор программ для создания таблиц в мировой практике весьма однообразен: «Майкрософт эксель», «Гугл рефайн», «Гугл спидшитс» или «Дата вронглер». Эти ресурсы позволяют загружать, извлекать или форматировать данные.

Третий этап – визуализация (отображение данных в виде графики или анимации). Для визуализации данных в виде графиков и диаграмм доступны такие приложения, как «Мэни айз» или «Тэблзу паблик», «Яхуу пейперс» или «Опен эдьюкейшн датабейс» [9], являющиеся примерами инструментов, которые позволяют создавать карты на основе электронных таблиц данных. Расширяется количество опций и платформ. Некоторые ресурсы, такие как «Тайметрик» [10] предоставляют возможность поиска, отображения и копирования данных. Для создания значимой и релевантной визуализации журналисты используют всё большее число инструментов, многие из которых предлагают использование шаблонов, что актуально для журналистов, только начинающих работать в этом направлении.

Четвёртый этап – публикация (включение данных в журналистский материал). Существуют различные варианты публикации данных и визуализации. Основной подход заключается в том, чтобы прикрепить данные к журналистскому тексту. Более серьёзные и профессиональные концепции позволяют создавать отдельные досье, например, для отображения ряда визуализаций, статей и ссылок на данные на одной странице.

Следующий шаг – распространение (создание мультимедийного материала, доступного на различных устройствах, таких как планшеты или мобильные телефоны). Предоставление доступа к существующим данным является ещё одним этапом, который приобретает всё большее значение. В условиях постоянного совершенствования технологий представляется крайне актуальным размещать данные на различных платформах и структурировать их для корректной работы на всех цифровых устройствах. Эти данные может просмотреть максимальное количество пользователей, а коллеги из других СМИ могут использовать их, например, для дальнейшего исследования темы. Предоставление доступа к данным и создание групп для обсуждения того, какая информация может быть извлечена из этих данных, является основной идеей ресурса «Базздата», по схожему принципу работают сервисы «Хелп ми инвестигейт» [11] или «Тайметрик».

Последний этап – отслеживание использования опубликованных данных с течением времени и во всех сферах деятельности. В контексте дата-журналистики такое отслеживание может быть использовано для маркетинговых целей. Некоммерческий центр «Пропаблика» использует для измерения программу «ПиксельПинг», созданную им совместно с ресурсом «Документ-

Клауд». Кроме того, существуют соответствующие организации, которые занимаются сбором таких сведений на коммерческой основе.

Использование определённых алгоритмов работы, применение специальных техник и навыков при создании материала, определённые требования, предъявляемые к журналистам, позволяют говорить о дата-журналистике как отдельном направлении, имеющем свои характерные особенности.

ИНФОГРАФИКА – НЕОТЪЕМЛЕМАЯ СОСТАВЛЯЮЩАЯ ДАТА-ЖУРНАЛИСТИКИ

Работа с данными требует от журналиста определённых навыков и способностей и, как следствие, расширяет сферу его работы. Области, которые включает в себя «журналистика данных», весьма многочисленны: большие базы данных, компьютерные отчёты и инфографика, визуализация данных, интерактивная визуализация, а также структурированная журналистская система управления информацией, где зафиксированы основные фактические сведения.

На данный момент среди исследователей США главными инструментами работы являются «Майкрософт эксель» и «Майкрософт аксес». Таблицы в «Экселе» – идеальный способ сортировки данных, использование математических формул позволяет работать с числовой информацией, вычленять интересные цифры, используя фильтры, а также легко возвращаться к информации, с которой работал, например, неделю назад. Кроме того, специальные функции в этой программе позволяют искать все запрашиваемые строки в любой по объёму таблице. Таблицы в «Аксес» стоит использовать при работе с большим количеством данных, группировать эти данные по конкретным свойствам, объединять в единую таблицу. Эти программы используют журналисты центра «Пропаблика», издания «Файненшл таймс», в Школе журналистики им. Уолтера Кронкайта.

Один из наиболее востребованных сервисов для работы с большим массивом информации – «Гугл докс», преимущество которого – возможность получать доступ к документу сразу несколькими пользователями без подключения к Интернету. В числе возможностей этого сервиса – создание текстовых документов, таблиц различной степени сложности. Помимо перечисленного, в программе «Гугл докс» есть возможность создавать мультимедийные презентации. Ещё один инструмент от «Гугл» – «Чарт тулс», с помощью которого стало возможно сделать многоуровневые диаграммы, а также интерактивные карты с возможностью выделять объекты на выбранной местности без особых затруднений. Помимо перечисленных выше инструментов, «Гугл» предлагает также специальную программу для работы с таблицами – «Гугл таблицы» и программу для очистки и преобразования данных – «Гугл рефайн». Эти инструменты используют журналисты центра «Пропаблика», изданий «Нью-Йорк таймс» и «Вашингтон пост», «Международный консорциум журналистов-расследователей» (МКЖР).

Однако некоторые проекты требуют более профессионального подхода к работе с данными, порой доступные и известные всем программы могут не справиться с чтением таблицы в несколько десятков тысяч строк, и дата-журналисты используют специальный язык программирования «Питон», который создан для производства программ разного назначения и активно применяется для специальных программ, которые могут обрабатывать неограниченные массивы данных. Этим языком пользуются большинство американских дата-отделов в СМИ, в том числе «Чикаго трибьюн», «Нью-Йорк таймс» и другие.

Самый важный элемент визуализации дата-материала – инфографика, которая стала незаменимой частью текста, содержащего большие данные. Сейчас журналисты отдают предпочтение графикам и диаграммам, поскольку в таблицах не всегда можно чётко увидеть взаимосвязи и закономерности.

Графические элементы решают сразу две задачи: привлекают аудиторию за счёт красивой и яркой картинки и помогают этой привлечённой аудитории понять сложную информацию большого объёма. Так появилось отдельное направление – визуальная журналистика, которая является неотъемлемым элементом журналистики данных. К визуальной журналистике относятся фотоизображения, рисунки, карикатуры, видеоматериалы, а также графические вставки: таблицы, диаграммы, графики, интерактивные карты.

Сейчас различные приложения и программы по созданию графики позволяют создавать всё что угодно. «Например, для организации информации в хронологическом порядке можно использовать временную шкалу» [12]. Для сравнения каких-то явлений или предметов можно выбрать диаграмму или таблицу. Гистограммы часто используются для сравнения категорий числовых значений. Однако диаграммы всё же остаются лучшей графикой, потому что они могут показывать вещи, которые трудно объяснить в тексте. Для демонстрации соотношения между двумя частями какого-то явления или проблемы часто используется специальная категория диаграммы. Их можно назвать «реляционными диаграммами». Они включают в себя диаграмму Венна, которая показывает, насколько велико совпадение между элементами, а также организационные диаграммы, блок-схемы, *playoff bracket* (вид схемы) и другие способы визуализации того, как отдельные лица или группы связаны друг с другом. Графики показывают взаимосвязи между данными в целом, а также между отдельными элементами данных. «Графики позволяют изображать величины, придавая им визуальные характеристики геометрических фигур» [13].

В настоящее время существует множество различных инструментов для создания инфографики, сейчас большое количество интернет-ресурсов посвящено обучению созданию графических схем и всем необходимым инструментам для создания нужного графического элемента в онлайн-режиме. Из самых распространённых – уже представленная выше программа «Гугл докс», работающая с презентациями, графиками, диаграммами и текстами. Не менее часто используются программы «Мэни айз» и «Тэблору паблик». Эти ресурсы позволяют

создать привлекательные и яркие схемы, таблицы, диаграммы или графики, оформить их в соответствии с выбранным журналистом стилем и цветовым решением.

С развитием журналистики данных визуализация текста заняла важное место в привлечении внимания аудитории. Практика показала, что самым оптимальным является использование диаграмм, схем и таблиц, которые благодаря нескольким словам или словосочетаниям, поставленным в логической связи, позволяют донести информацию, которая в тексте занимала бы два – три крупных абзаца.

На основе проведённого исследования можно утверждать, что журналистика данных открывает для репортёра новые возможности: сбор и объединение информации из множества источников, поиск взаимосвязей по документам из разных учреждений или компаний (благодаря специальному компьютерному обеспечению), выявление общих показателей среди сотен тысяч документов, создание общественно важных проектов, которые могут помочь аудитории в социально-бытовых аспектах.

Данное направление характеризуется двумя особенностями: это возможность работать с базами данных для поиска необходимых сведений и «украшение» текста инфографикой (схемами, таблицами, яркими графиками), которые позволяют сделать материал более достоверным и наглядным, а также способствуют более лёгкому усвоению аудиторией сложных и многочисленных сведений.

Таким образом, «журналистика данных» становится всё более востребованной в силу открывающихся возможностей. Работа с данными помогает раскрыть мельчайшие детали о какой-то проблеме, подтвердить или опровергнуть официальные заявления чиновников, политиков, бизнесменов, а также разобраться с корпоративной отчётностью, финансовыми данными, увидеть и проанализировать социальные показатели (качество медицинского обслуживания, образования, жилищных условий).

ИСТОЧНИКИ

[1] Bradshaw P., Gray J., Chambers L., Bounegru L. The Data Journalism Handbook / P. Bradshaw, J. Gray, L. Chambers, L. Bounegru. Available at: http://datajournalismhandbook.org/1.0/en/introduction_0.html (accessed 08.01.2019).

[2] Houston B. Fifty Years of Journalism and Data: a Brief History. Global Investigative Journalism Network. Available at: <https://gijn.org/2015/11/12/fifty-years-of-journalism-and-data-a-brief-history/> (accessed: 15.01.2019).

[3] Barlett D., Steel J. Wasted Billions: How Your Tax Dollars Were Lost on Synthetic Fuel. Official Site of Donald Barlett and James Steel. Available at: http://www.barlettandsteel.com/journalism/inq_billions_1 (accessed: 15.01.2019).

[4] Investigative Reporters & Editors // Official Site of IRE. Available at: <https://www.ire.org> (accessed: 06.01.2019)

[5] Лоренц М. Бизнес-модели для журналистики данных. Пособие по журналистике данных; под ред. Д. Грея. Available at: <https://unotices.com/book.php?id=125825> (accessed 10.01.2019).

[6] Чемберс Л., Грей Дж. Освещение темы госбюджета при помощи Open-Spending.org. Пособие по журналистике данных; под ред. Д. Грея. Available at: <https://unotices.com/book.php?id=125825> (accessed 10.01.2019).

[7] Матцат Л. Наши сюжеты делаются в виде кода. Пособие по журналистике данных; под ред. Д. Грея. Available: <https://unotices.com/book.php?id=125825> (accessed 10.01.2019).

[8] Bradshaw P. The inverted pyramid of data journalism. Online Journalism. Available at: onlinejournalismblog.com/2011/07/07/the-inverted-pyramid-of-data-journalism/ (accessed: 06.01.2019).

[9] Do-It-Yourself GIS: 20 Free Tools & Data Sources for Creating Data Maps. Open Education Database. Available at: <https://oedb.org/ilibrarian/do-it-yourself-gis-20-free-tools-data-sources-for-creating-data-maps/> (accessed: 16.01.2019).

[10] Timetric. Official Site of Timetric. Available at: <http://www.timetric.com/> (accessed: 17.01.2019).

[11] About Project. Help Me Investigate. Available at: <http://helpmeininvestigate.com/> (accessed: 17.01.2019).

[12] Cherubini F. The Future of Investigative Journalism in the U.K. According to the House of Lords. World association of newspapers and news publishers. Available at: <http://www.editorsweblog.org/2012/03/05/the-future-of-investigative-journalism-in-the-uk-according-to-the-house-of-lords> (accessed 23.01.2019).

[13] Эйш Г. Как визуализировать данные. Пособие по журналистике данных; под ред. Д. Грея. Available: <https://unotices.com/book.php?id=125825> (accessed 10.01.2019).

СПИСОК ЛИТЕРАТУРЫ

Вайгенд А. 2018. Big Data. Вся технология в одной книге; [пер. с англ. С. Богданова]. М.: Эксмо. 384 с.

Стивенс-Давидовиц С. 2018. Все лгут. Поисковики, Big Data и Интернет знают о вас все ; [пер. с англ. Л.И. Степановой]. М.: Эксмо. 382 с.

REFERENCES

Weigend A. Data for the People: How to Make Our Post-Privacy Economy Work for You. New York: Basic Books, 2017. 272 pp.;

Vaygend A. Big Data. Vsyа technologiya v odnoy knige ; [per. s angl. S. Bogdanova]. M.: Eksmo, 2018. 384 s.

Stephens-Davidowitz S. Everybody Lies: Big Data, New Data, and What the Internet can tell us about who we really are. New York: Dey Street Books, 2017. 338 p.

Stivens-Davidovitz S. Vse lgut. Poiskoviki, Big Data i Internet znayut o vas vse ; [per. s angl. L.I. Stepanovoy]. M.: Eksmo, 2018. 382 s.

ИНФОРМАЦИЯ ОБ АВТОРЕ / INFORMATION ABOUT THE AUTHOR

Неренц Дарья Валерьевна, кандидат филологических наук, доцент кафедры журналистики факультета журналистики Института массмедиа РГГУ. Российская Федерация, Москва.

Nerentz Darya Valeryevna, Cand. Sci. (Philology), Assistant Professor. Institute of Massmedia of Russian State University for the Humanities. Moscow, Russian Federation.