

DATA JOURNALISM: НОВЫЕ ВОЗМОЖНОСТИ ДЛЯ ЖУРНАЛИСТОВ-РАССЛЕДОВАТЕЛЕЙ

Д. В. Неренц

Российский государственный гуманитарный университет

Поступила в редакцию 17 февраля 2018 г.

Аннотация: журналистика данных становится все более востребованной в среде журналистов-расследователей. В данной статье анализируются причины появления этой тенденции, а также рассматривается вопрос об использовании баз данных в качестве основного источника информации в расследовательской журналистике.

Ключевые слова: журналистские расследования, журналистика данных, базы данных, инструменты для анализа.

Abstract: data journalism is becoming more and more popular among investigative journalists. This article analyzes the reasons for this trend and it also considers the use of databases as the main source of information in investigative journalism.

Keywords: journalistic investigations, data journalism, databases, tools for analysis.

Новые технологии и возможности Интернета послужили толчком к стремительному видоизменению основных способов работы журналистов в целом и журналистов-расследователей в частности. Теперь, помимо активного вовлечения пользователей в работу над анализом документов и интервью, они все чаще обращаются к компьютерным базам данных, в которых содержится большинство документации и многочисленные отчеты компаний и правительственных секторов. Этот феномен получил название «журналистика данных» (data journalism).

С. И. Симакова в статье «Дата-журналистика как медиатренд» указывает, что «дата-журналистика — это формат грамотного анализа данных, еще один способ критического исследования мира и, по сути, инструмент, помогающий получить и представить общественности важные сведения, которые могли остаться незамеченными в огромных массивах информации о современной реальности. Для журналистов дата-журналистика во многом означает усложнение процесса их работы. При этом конечный журналистский продукт должен быть с легкостью воспринят аудиторией» [1, 484].

М. Г. Шилина в материале «Data Journalism — дата-журналистика, журналистика метаданных — в структуре медиакommunikации: к вопросу формирования теоретических исследовательских подходов» дает схожее определение: «Специфика указанного формата актуальной журналистики состоит в том, что авторы используют цифровые данные не просто как инструмент при создании текста или иллюстративного ряда к нему, но как основной ресурс формирования темы, концепции, ключевое условие выбора

формата/жанра сообщения с точки зрения соотношения вербального и визуального компонента, интерактивности потребления информации» [2].

Американские исследователи и авторы «Настольной книги 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 журналистского процесса: 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 ChicagoCrime, а затем с EveryBlock» [3].

Общеизвестно, что журналисты-расследователи при разработке темы в первую очередь обращаются к открытым источникам информации. Сейчас этими открытыми источниками стали базы данных, с изучения которых начинается работа и перепроверкой которых работа заканчивается. Открытые регистры и реестры — первый инструмент в поиске связей между бизнесменами, чиновниками, политиками и их активами.

«Журналистика данных стала главным источником информации в журналистской среде. База данных — это ключ к расследованию. После составления полного списка вопросов по выбранной теме необходимо обратиться к открытым записям, которые помогут ответить на большинство из них. Благодаря этому ресурсу можно узнать статистику (например, если расследование о наркокартелях, в базе данных можно найти: сколько преступлений в этой

области, сколько смертей от передозировки, сколько дилеров было зафиксировано, сколько поймано, сколько изъято наркотических средств, каких именно и т.д.) Таким образом, благодаря базе данных можно составить детальную картину нынешней обстановки в том или ином городе, штате», — рассказывает Джозеф Бергантино, директор «Нового английского центра расследовательской журналистики» Бостонского университета.

Следовательно, первый шаг в разработке темы — сформулировать все возможные запросы по базам данных (найти все открытые записи); иногда на это приходится тратить деньги, зачастую это требует также много времени, поскольку ответа приходится ждать до нескольких месяцев, а иногда и около года.

Правительства и предприятия во всем мире публикуют в открытом доступе свои отчеты, пытаются стать максимально прозрачными, следовательно, все больше и больше данных становятся открытыми, и задача расследователя — уметь эти данные собирать и использовать в работе. «Международный консорциум журналистов-расследователей» (ICIJ) на своем официальном сайте предлагает расследователям видеуроки по основным функциям Excel, инструкции по работе со справочной службой LexisNexis и по поиску документов федеральных судебных заседаний, корпоративной документации различных компаний.

Журналистика данных становится все более важной для расследовательской журналистики. Прежде всего, работа с данными помогает раскрыть базовую информацию о проблеме или явлении, подтвердить или усомниться в официальных заявлениях чиновников, разобраться с корпоративной отчетностью, позволяет использовать новые способы подачи материала (визуальные элементы, интерактивные карты и т.д.).

Однако при работе с базами данных нельзя забывать, что это не идеальный источник информации (неточности и ошибки встречаются везде), в базах данных нет всей информации, это начальный этап работы, требующий значительных доработок и проверки, и они не могут заменить профессиональных репортерских навыков.

Многие практикующие расследователи и в России, и за рубежом говорят о том, что в будущем будет создана глобальная база данных, которая включит в себя все существующие базы данных в мире. Это, безусловно, будет эффективный инструмент, который позволит без дополнительных поисковых запросов найти необходимую информацию в открытом доступе о любой стране. Но пока эта идея остается только на уровне обсуждения. Между тем многие некоммерческие центры расследовательской журналистики и ведущие мировые СМИ (New York Times, Guardian, Financial Times) размещают на своих страницах в Сети созданные журналистами базы данных, которые может посмотреть любой пользователь.

Например, в ProPublica на официальном сайте организации есть отдельная рубрика, посвященная базам данных. Она называется «Расскажите вашу собственную историю с помощью базы данных от ProPublica» [4]. Существующие у них базы данных поделены на рубрики: Здоровье, Уголовное правосудие, Образование, Политика, Бизнес, Транспорт, Военные силы и Окружающая среда.

Соответственно, например, если мы выбираем раздел «Образование», то видим несколько баз данных, первая из которых платная (база данных о сегрегации в школе). На сайте указано, что платные, или премиум-базы, предназначены для предоставления информации с данными, анализом и практической документацией, которые вы больше нигде не найдете, то есть уникальные. Остальные базы бесплатные (альтернативные школы, данные по школьной десегрегации, данные об ограничении и изоляции). Однако для скачивания базы данных, надо оставить заявку онлайн, но в целом это не вызывает особых затруднений.

Большая база данных есть у «Центра исследования коррупции и организованной преступности» (OCCRP). Атанас Чобанов, редактор болгарской расследовательской сети bivol.bg, в своей статье «Работа с «большими данными» описывает принципы работы с этой базой данных: «ID располагает также хранилищем ссылок на более чем 750 электронных баз данных в 120 странах, доступ к которому предоставляется по принципу «самообслуживания». Эта коллекция полезных открытых ресурсов для журналистских расследований составляется исследователями OCCRP и охватывает более чем 450 коммерческих реестров, более чем 150 реестров интеллектуальной собственности, судебные базы данных, а также базы по земельным участкам и государственным закупкам. Базы удобно сгруппированы по странам и регионам. Имеется также специальная закладка для офшоров» [5].

Что касается российской практики, то здесь стоит отметить недавно появившееся объединение журналистов-расследователей «Четвертый сектор» в Перми. На их сайте собраны ссылки на все имеющиеся в открытом доступе базы данных. В частности, указаны российские базы: картотеки, базы дел и судебных решений, закупки, карточные базы, реестры организаций, данные о частных лицах, нормативные акты. Есть иностранные и международные базы: Investigative Dashboard — публичные регистры со всего мира, WikiLeaks, Offshore Leaks Database (созданное Международным консорциумом журналистов-расследователей на основе «Панамских архивов»), Enigma — «гугл» среди баз данных, OpenCorporates — самая большая открытая база данных по компаниям со всего мира. Кроме того, указаны и базы недвижимости в Европе, картотеки компаний в Европе, судебные базы, открытые базы США.

Еще одна рубрика на этом сайте посвящена открытым реестрам. Есть общие: «Портал открытых данных России», «Портал административной реформы», «Портал открытых данных Государственной думы РФ», «Единый портал бюджетной системы РФ», «Федеральная информационная адресная система», «Открытая полиция». Присутствуют реестры открытых данных государственных служб РФ, реестры открытых данных министерств РФ, реестры открытых данных федеральных агентств, открытые данные городов и регионов России, международные реестры, открытые данные стран Европы.

Безусловно, вид расследования требует работы с конкретными ресурсами. Например, если говорить о коррупционных расследованиях, то практикующие журналисты едины во мнении, что начинать следует с ЕГРЮЛ, где есть все сведения о юридических лицах и индивидуальных предпринимателях. Также используется платная база данных «Контур фокус» (focus.kontur.ru), которая содержит истории создания компании, список акционеров, владельцев, ИНН, финансовую информацию, даты регистрации, арбитражные дела, номера телефонов, закупки.

И самая подробная платная база на сегодняшний день на русском языке — «Спарк» (spark-interfax.ru), в которой, помимо общедоступной информации, есть кредитные истории компаний, отчеты об арбитражных делах, финансовая отчетность, поиск аффилированности, указана предварительная оценка факторов риска.

В последнее время в российской практике наблюдается тенденция ухода от работы с закрытой информацией. Все чаще расследователи предпочитают работу с открытыми источниками, например с данными Росреестра. Роман Шлейнов и Елена Виноградова из газеты «Ведомости» провели расследование «На берегах Злодейки» о поселке Акулинино и его обитателях, то есть всех, кто владел там недвижимостью близ дачи В. И. Якунина, бывшего президента ОАО «РЖД». Вся эта информация была взята только в Росреестре, больше не потребовалось никаких источников.

Эрик Толер, исследователь международной исследовательской сети Bellingcat, рассказывает о плюсах работы только по открытым источникам: «...если мы используем в исследовании открытые источники, мы получаем гораздо больше прозрачности между нами и нашим зрителем или слушателем. Если мы опубликовали материал о каком-то чиновнике, и кто-то считает, что наша информация неверна или читатели нам не доверяют, у них есть точно такой же доступ к информации, которую мы использовали, и они могут самостоятельно все проверить. Если кто-то утверждает, что предоставленная нами информация недостоверна, мы можем ответить, что каждый имеет абсолютно равный доступ ко всем материалам, которые мы использовали» [5]. В качестве основных

открытых ресурсов он указывает социальные сети, видео на сайтах типа YouTube и спутниковую фотосъемку. «Иногда мы покупаем фотографии со спутников, но обычно находятся бесплатные варианты, такие как Google Earth или «Яндекс Карты». Есть также множество материалов, опубликованных в СМИ, будь то телевидение, радио, цифровые СМИ — анализу подвергается все, что открыто и доступно для всех» [5], — добавляет он.

Однако работа с базами данных требует от журналиста определенных навыков и способностей. И, прежде всего, это относится к умению пользоваться определенным набором компьютерных программ. Иван Бегтин, директор некоммерческого партнерства «Информационная культура», выделяет следующие необходимые инструменты: «Необходимые навыки работы с рядом аналитических и технических инструментов... включает такие инструменты, как Excel, Google Docs, IBM ManyEyes, Wordle и многие другие. Далеко не все журналисты обладают навыками работы с этими инструментами, и большинство из тех, кто такими навыками обладает, понимает, как необходимо подавать материалы в понятном медиформате» [6].

На данный момент среди расследователей США главными инструментами работы являются Microsoft Excel и Microsoft Access. Таблицы в Excel — 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 объему таблице. Таблицы в Access стоит использовать при работе с большим количеством данных, группировать эти данные по конкретным свойствам, объединять в единую таблицу.

Один из наиболее востребованных сервисов для работы с большим массивом информации — Google Docs, преимущество которого — возможность получения доступа 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 Google Docs есть возможность создания мультимедийных презентаций. Еще один инструмент от Google — Chart Tools, с помощью которого стало возможно создание многоуровневых диаграмм, а также интерактивных карт с возможностью выделения объектов на выбранной местности без особых затруднений.

Для визуализации полученных результатов и создания инфографики современные расследователи чаще всего используют ManyEyes. Этот ресурс позволяет создать привлекательные и яркие схемы, таблицы, диаграммы или графики, оформить их

в соответствии с выбранным журналистом стилем и цветовым решением. Однако у российских журналистов-расследователей пока нет возможности использовать многочисленные инструменты, созданные для англоязычной аудитории. Например, на русском языке из перечисленных выше ресурсов есть только инструменты Google и Microsoft Excel. Другие полезные ресурсы, такие как ManyEyes, Timetric, IBM, Wordle, не имеют русскоязычной версии.

Таким образом, журналистика данных становится все более важной для исследовательской журналистики. Прежде всего, работа с данными помогает раскрыть базовую информацию о проблеме или явлении, подтвердить или усомниться в официальных заявлениях чиновников, разобраться с корпоративной отчетностью, позволяет использовать новые способы подачи материала (визуальные элементы, интерактивные карты и т.д.).

Однако при работе с базами данных нельзя забывать, что это не идеальный источник информации (неточности и ошибки встречаются везде), в базах данных нет всей информации, это начальный этап работы, требующий значительных доработок и проверки, и они не могут заменить профессиональных репортерских навыков.

Российский государственный гуманитарный университет

Неренц Д. В., кандидат филологических наук, старший преподаватель кафедры журналистики

E-mail: ya.newlevel@yandex.ru

ЛИТЕРАТУРА

1. Симакова С. И. Дата-журналистика как медиатренд / С. И. Симакова // Вестник Нижегородского университета им. Н. И. Лобачевского. — 2014. — Вып. 2 (2). — С. 481—484.
2. Шилина М. Г. Data Journalism — дата-журналистика, журналистика метаданных — в структуре медиакommunikation: к вопросу формирования теоретических исследовательских подходов // Медиаскоп. — 2013. — Вып. 1. — URL: <http://www.mediascope.ru/node/1263> (дата обращения: 14.01.2018).
3. Gray J. The Data Journalism Handbook / J. Gray, L. Chambers, L. Bounegru. — Available at: http://datajournalismhandbook.org/1.0/en/introduction_0.html (accessed 10.02.2018).
4. Propublica Data Store // Propublica. — Available at: <https://www.propublica.org/datastore/> (accessed 10.02.2018).
5. Журналистские расследования случаев трансграничной коррупции. Методы и примеры / Ред. кол.: А. Калих, А. Кириленко, Г. Гуммель. — URL: http://eu-russia-csf.org/fileadmin/Docs/Trans-Border_Corruption/ManualTransborderCorruption02.02.2018_ru.pdf (дата обращения: 14.02.2018).
6. Бергин И. Памятка по журналистике данных / И. Бергин // Сайт Ивана Бергина. — URL: <http://ivan.begtin.name/2011/05/22/datajournalism/> (дата обращения: 15.02.2018).

Russia State University for the Humanities
Nerents D. V., Candidate of Philology, Senior Lecturer of the Journalism Department
E-mail: ya.newlevel@yandex.ru