

**Московский государственный университет  
имени М. В. Ломоносова**

**ФАКУЛЬТЕТ ЖУРНАЛИСТИКИ**

Кафедра новых медиа и теории коммуникации

**Журналистика расследований на основе данных: опыт  
отечественных медиа**

Выпускная квалификационная работа  
студентки IV курса бакалавриата  
дневного отделения  
Ермаковой Виктории Валерьевны

Научный руководитель:  
к.ф.н., старший преподаватель  
Крашенинникова Мария Алексеевна

К ЗАЩИТЕ

\_\_\_\_\_/\_\_\_\_\_/\_\_\_\_\_  
«\_\_» \_\_\_\_\_ 2020 г.

Зав. кафедрой

К ЗАЩИТЕ

\_\_\_\_\_/ Засурский И.И /  
«\_\_» \_\_\_\_\_ 2020 г.

## АННОТАЦИЯ

Выпускная квалификационная работа направлена на изучение того, как данные изменили журналистские расследования в России. Феномен больших и открытых данных подтолкнул медиа к созданию новой коммуникационной стратегии, включающей в себя использование инфографики, взаимодействие с аудиторией особым образом, количественный подход и другие особенности, изученные в ВКР. Новизна работы заключается в том, что журналистика на основе данных находится на начальном этапе своего практического и теоретического развития, а также академического осмысления в России.

ВКР основывается на 54 дата-журналистских расследовательских материалах из 15 различных медиа. Статьи были проанализированы с точки зрения источников данных, визуализаций, тематики, интерактивности, инструментов для анализа данных и реплицируемости.

## ABSTRACT

The graduation paper “Investigative data journalism: the experience of Russian media” is aimed to understand the way data changed journalism investigations in Russia. The phenomenon of big data and open data creates a new communication scheme for media, which includes advanced infographics, interactivity, quantitative turn and other features covered in the paper. The novelty of the current work lies in that data journalism today is making first steps in the development of theory, practice, and academic research in Russia

This research is based on 54 journalistic investigations from 15 mass media. The articles are analyzed in terms of types of data sources, visualizations, topics, interactivity, tools for data analysis and replication.

*Работа написана мною самостоятельно и не содержит неправомерных заимствований.*

«02» мая 2020г. Ермакова В.В.

## Оглавление

Введение.....	4
<b>Глава 1. Журналистика данных в мировой и отечественной медиапрактиках</b>	
1.1. Исследования журналистики данных .....	8
1.2. Использование данных в мировой медиапрактике .....	19
1.2.1 Первые примеры использования данных в журналистике.....	19
1.2.2 Мировое законодательство в области раскрытия данных.....	21
1.2.3 Журналистика данных и ведущие мировые СМИ .....	23
1.2.4 Международные премии для дата-журналистов .....	26
1.3. Использование данных в отечественной медиапрактике .....	29
1.3.1 Российский опыт: законодательный и практический аспекты.....	29
1.3.2 Общественные проекты с использованием открытых данных.....	32
Выводы по главе 1.....	39
<b>Глава 2. Журналистские расследования на основе данных в России</b>	
2.1. Подход к анализу .....	40
2.2. Методика анализа.....	45
2.3. Основные характеристики журналистских расследований на основе данных .....	51
2.4. Журналистские расследования на основе данных в «Новой газете».....	61
2.5. Журналистские расследования на основе данных в «Трансперенси Интернешнл Россия» .....	69
2.6. Журналистские расследования на основе данных в «Проекте» .....	73
2.7. Журналистские расследования на основе данных в «ФБК» .....	76
2.8. Журналистские расследования на основе данных в «РБК» .....	79
2.9. Самые яркие примеры российских дата-расследований .....	82
Выводы по главе 2.....	89
<b>Заключение.....</b>	<b>90</b>
<b>Библиографический список .....</b>	<b>94</b>
<b>Приложения.....</b>	<b>100</b>

## ВВЕДЕНИЕ

Благодаря развитию современных компьютерных технологий и Интернета журналисты могут находить и анализировать огромные массивы данных с помощью алгоритмов, а посредством графических редакторов – наглядно визуализировать эти данные для читателей.

Мировое законодательство также способствовало формированию журналистики данных как самостоятельного направления, потребовав от правительств раскрывать государственные данные. Как следствие журналисты получили огромный материал для проведения журналистских расследований.

Как пишут авторы «Настольной книги по журналистике данных», теперь данные могут служить источником для дата-журналистики (равноценно журналистике данных; от англ. “data” – данные), инструментом, дополняющим журналистскую историю, или же и тем, и другим.<sup>1</sup>

Наша гипотеза состоит в том, что данные существенно изменили исследовательскую журналистику. Теперь журналист может проводить расследования, опираясь на наборы данных в Интернете или же с помощью специальных алгоритмов самостоятельно находить информацию для расследования.

Государство, общественные организации, компании располагают большими объемами данных, которыми они готовы или вынуждены (по законодательству) делиться. Для журналистов-расследователей это возможность узнать то, что ранее было недоступно, и рассказать о результатах своей аудитории.

---

<sup>1</sup> Bradshaw P. What Is Data Journalism? // Data Journalism Handbook 1. URL: <https://datajournalism.com/read/handbook/one/introduction/what-is-data-journalism> (дата обращения: 20.01.2020)

В мире журналистика на основе данных развивается в течение последних десяти лет. В России дата-журналистика как направление только зарождается. Этим обосновывается выбор темы, а также ее *актуальность*.

*Научная новизна данной работы* заключается в том, что в российском академическом сообществе на данный момент уделяется недостаточно внимания использованию открытых данных, больших данных, визуализаций на основе данных, а также программных алгоритмов в журналистских материалах. А тот аспект дата-журналистики, который мы намереваемся рассмотреть – то, как данные повлияли на журналистские расследования, – мало изучен в принципе.

*Объектом исследования* выступают журналистские расследования на основе данных.

*Предметом исследования* являются особенности расследовательской дата-журналистики в России.

В данной ВКР мы ставим себе *цель* рассмотреть особенности журналистских расследований на основе данных в России и выявить типичные характеристики, свойственные подобным материалам.

Для достижения цели нам необходимо решить следующие *задачи*:

- найти расследования на основе данных в российских медиа;
- выявить тематику дата-материалов;
- охарактеризовать источники данных для расследований;
- проанализировать наборы данных, которыми пользовался журналист;
- охарактеризовать визуализации, используемые в материалах;
- выявить алгоритм работы журналиста с данными;

- отследить общие паттерны в расследованиях на основе данных.

Теоретически журналистика на основе данных слабо изучена в мире и в России. Автор данной ВКР в основном будет опираться на англоязычные источники – научные и публицистические статьи профессоров американских и европейских университетов (Рич Гордон, Брюс Гаррисон, Марк Коддингтон и др.), а также экспертов в области (Эдриан Головатый, Эрик Юлкен, Саймон Роджерс и др.). К работам русскоязычных исследователей (Шилина М.Г., Симакова С.И., Неренц Д.В. и др.) и экспертов (Бегтин И., Богачев А., Дорожный А. и др.) мы также обращались во время написания ВКР.

При работе над исследованием мы решили использовать комплекс методов, таких как сопоставление, систематизация и классификация, контент-анализ.

*Эмпирической базой* для ВКР послужили 54 расследования (или материалы с признаками расследования) на основе данных в 15-ти различных отечественных медиа, которые нам удалось найти. Среди данных медиа оказались: «Новая газета», «Трансперенси Интернешнл Россия», «Проект», «ФБК», «РБК», «Деловой Петербург», «Инфометр», “The Bell”, OCCRP, Ivan Begtin Blog, фонд «Нужна помощь», «Открытые медиа», Институт проблем правоприменения, «Другой город», “Meduza”.

В качестве положений на защиту автор выделяет следующие *тезисы*:

- «Новая газета» – главное российское медиа, занимающееся расследовательской журналистикой на основе данных; большинство расследований на основе данных публикуются данным СМИ в настоящий момент;
- Для проведения расследования на основе данных в среднем требуется 2 источника данных; журналисты стремятся находить

вспомогательные наборы данных, чтобы дополнить свои истории;

- В большинстве случаев расследования строятся на государственных источниках данных; к общественным, коммерческим и альтернативным источникам данных журналисты обращаются существенно реже;
- Редакции стремятся сделать свои расследования более прозрачными для читателей, чтобы не потерять доверие той части аудитории, которая может усомниться в дата-материале;
- Расследований на основе больших данных в российском медиапространстве почти нет, материалы, которые строятся на малых данных, существенно доминируют

# ГЛАВА 1. ЖУРНАЛИСТИКА ДАННЫХ В МИРОВОЙ И ОТЕЧЕСТВЕННОЙ МЕДИАПРАКТИКАХ

## 1.1 Исследования журналистики данных

Несмотря на то, что первые примеры использования данных в журналистике появились еще в XIX веке<sup>2</sup>, научно-теоретическая часть значительно отстает.

Только 2010 году дата-журналистика была выделена как самостоятельное направление<sup>3</sup> на первой конференции по дата-журналистике Data-driven Journalism, проведенной Европейским центром журналистики (EJC). На конференции дискутировали на те же темы, что интересуют нас сейчас: возможности визуализации данных, их использования, сторителлинг на основе данных, дальнейшие перспективы.<sup>4</sup>

В 2001 году в своем исследовании “Diffusion of Online Information Technologies in Newspaper Newsrooms”<sup>5</sup> Брюс Гаррисон, профессор Университета Майами, констатировал, что интернет-технологии, такие как поисковые системы и службы баз данных, проникли в американские новостные редакции. Так, начиная с поздних 1990-х, Интернет стал главным источником сбора новостей.

В 2006 году Эдриан Головатый – один из разработчиков фреймворка (рабочая среда программиста) Django, создатель дата-проекта

---

<sup>2</sup> The first Guardian data journalism: May 5, 1821 // The Guardian. 2011.  
URL:<https://www.theguardian.com/news/datablog/2011/sep/26/data-journalism-guardian> (дата обращения: 20.01.2020)

<sup>3</sup> Лисицин М. Е. Понятие "журналистика данных" в современных научных публикациях // Коммуникативные исследования. 2018. №3 (17). С. 145. [Электронный ресурс].  
URL:<https://cyberleninka.ru/article/n/ponyatie-zhurnalistika-dannyh-v-sovremennyh-nauchnyh-publikatsiyah> (дата обращения: 21.01.2020)

<sup>4</sup> Data-driven journalism: What is there to learn. 2010.  
URL:[http://mediapusher.eu/datadrivenjournalism/pdf/ddj\\_paper\\_final.pdf](http://mediapusher.eu/datadrivenjournalism/pdf/ddj_paper_final.pdf) (дата обращения: 20.01.2020)

<sup>5</sup> Garrison, B. (2001). Diffusion of Online Information Technologies in Newspaper Newsrooms // Journalism, 2(2). P. 221–239. (дата обращения: 20.01.2020)



chicagocrime.org, веб-разработчик и журналист – опубликовал манифест (так его назвали в книге по цифровой журналистике расследований)<sup>6</sup> дата-журналистики “A fundamental way newspaper sites need to change”.<sup>7</sup> В своем манифесте Головатый призывает остановить «сюжетно-ориентированное мировоззрение» (от англ. “to stop the story-centric worldview”) журналистов и включить в истории данные. Теперь данные могут дополнять текст журналистских публикаций. Журналисты должны уметь не только рассказывать истории, а находить и обрабатывать данные. О журналистах-программистах писал следующий медиаисследователь – Рич Гордон.

В 2007 году Рич Гордон, профессор Северо-Западного университета в штате Иллинойс, в своей публицистической статье “What Will Journalist-Programmers Do?”<sup>8</sup> констатирует, что самый популярный вопрос, который он получает от СМИ, – «Когда мы сможем их нанимать?». Гордон предвосхищает тот факт, что спрос на специалистов, владеющих одновременно навыками программирования и журналистики, будет возрастать. 13 лет назад было неочевидно, зачем нужны «журналисты-программисты» и чем они *будут* заниматься, а теперь крупные мировые редакции создают свои дата-отделы.

Рич Гордон разделяет журналистику данных на несколько уровней сложности (по нарастанию), устанавливая различные виды деятельности дата-журналистов.

1. Доставка данных (от англ. data delivery): СМИ получают данные и публикуют их без дополнительных комментариев; функционал для онлайн-пользователей ограничен;

---

<sup>6</sup> Hahn O., Stalph F. Digital Investigative Journalism: Data, Visual Analytics and Innovative Methodologies in International Reporting. Palgrave Macmillan, 2018. (дата обращения: 20.01.2020)

<sup>7</sup> Holovaty A. A Fundamental Way Newspaper Sites Need to Change. 2006.

URL:<http://www.holovaty.com/writing/fundamental-change/> (дата обращения: 20.01.2020)

<sup>8</sup> Gordon R. What Will Journalist-Programmers Do? // Mediashift. 2007. URL:<http://mediashift.org/2007/11/what-will-journalist-programmers-do005/> (дата обращения: 20.01.2020)

2. Поиск данных (от англ. data search): данные, которые могут искать сами пользователи в базе данных, вводя нужное слово в строку поиска;
3. Разведка данных (от англ. data exploration): поиск данных с расширенным функционалом для пользователей;
4. Визуализация данных (от англ. data visualization): визуализация того, что показывают данные;
5. Истории на основе данных (от англ. data experiences and storytelling): соединение традиционного репортажа и данных.

Рич Гордон далее приводит примеры того, как уже в 2007 году мировые СМИ сочетали данные и журналистику. Многие из этих проектов уже устарели и закрылись, поэтому оценить их можно по словам самого исследователя.

Из всех перечисляемых уровней сложности, нас как журналистов интересуют только последние два (визуализация данных и истории на основе данных). Первыми тремя пунктами теперь занимаются совсем другие компании. К примеру, СМИ сейчас не заинтересованы в публикации интерактивных дата-баз с ценами на недвижимость, как это делали The Cincinnati Enquirer несколько лет назад (пример из статьи). Для этого в России есть такие интернет-сервисы, как ЦИАН или Авито.

Подобное также произошло с объявлениями о работе. Если раньше желающие трудоустроиться изучали публикации в газетах, то сейчас достаточно зайти на сайт российской компании интернет-рекрутмента HeadHunter. Электронные дата-базы с вакансиями больше не интересуют журналистов – этим в настоящее время занимаются узкоспециализированные электронные сервисы.

Однако некоторые интересные примеры создания базы данных журналистами сохранились и действительны до сих пор. Например, карта

убийств в Лос-Анджелесе “The Homicide Report”, созданная The Los Angeles Times.<sup>9</sup> Этот проект рассматривается в статье 2008 года “Building the data desk: lessons from the L.A. Times” Эрика Юллена, опытного диджитал-редактора, в настоящее время директора по продукту (от англ. product director) сети USA Today.<sup>10</sup> Автор рассказывает, что карта убийств запустилась в начале 2007 года (подобный проект уже запускался Эдрианом Головатым ранее – ChicagoCrime.org; сейчас уже закрыт и не действует). Изначально проект представлял собой только блог (данные об убитых людях, их истории), а не карту. Со временем стало очевидно, что исключительно блог не сумеет охватить в полном объеме ту информацию, которую хотели донести журналисты The Los Angeles Times. Нужно было создать проект, подобный карте происшествий Чикаго (проект Головатого [chicagocrime.org](http://chicagocrime.org)), чтобы пользователи сами смогли работать с разными фильтрами, находить тенденции. Таким образом, через несколько месяцев с помощью талантливых разработчиков проект превратился в карту, по которой можно отслеживать неблагополучные районы Лос-Анджелеса. На рис. 1 мы видим часть этой карты, созданной редакцией The Los Angeles Times.

.

---

<sup>9</sup> The Homicide Report // The Los Angeles Times. [URL:https://homicide.latimes.com](https://homicide.latimes.com) (дата обращения: 23.01.2020)

<sup>10</sup> Ulken E. Building the data desk: lessons from the L.A. Times // Online Journalism Review. 2008. [URL:http://www.ojr.org/p1581/](http://www.ojr.org/p1581/) (дата обращения: 23.01.2020)

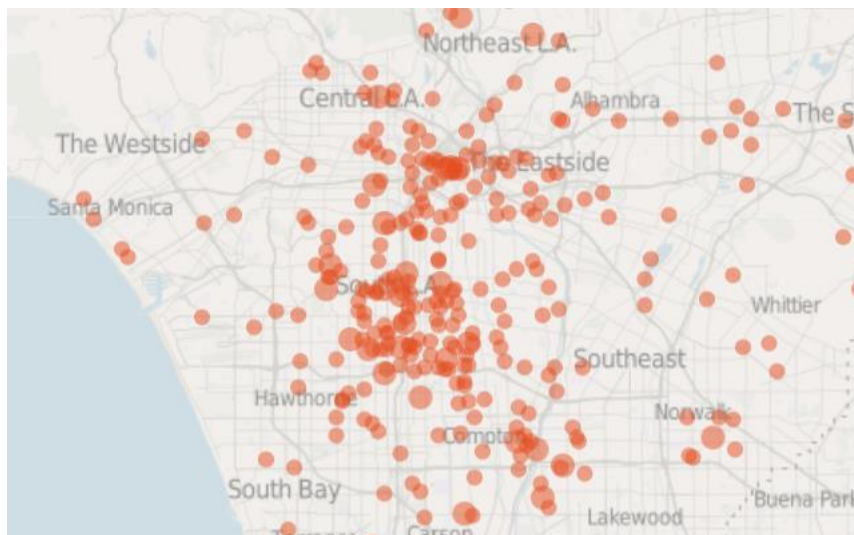


Рис. 1. Часть карты проекта “The Homicide Report”; Лос-Анджелес, штат Калифорния, США<sup>11</sup>

В российском медиапространстве есть свой частичный аналог «The Homicide Report» – это интерактивная карта заказных убийств бизнесменов в Петербурге с 1992 по 2005 годы<sup>12</sup> авторства Андрея Дорожного, ведущего дата-журналиста России, бывшего бильд-редактора «Делового Петербурга». Материал основывается на данных об убийствах, которые собрало издание «Петербургский предприниматель». По словам Андрея Дорожного, убийства в основном совершались у подъездов и офисов.<sup>13</sup>

Одним из самых цитируемых источников (по данным сервиса Google Scholar, на работу ссылались 475 раз на 25.04.2020) на данный момент является «Настольная книга по журналистике данных», выпущенная коллективом авторов в 2012 году. В данном издании речь идет о том, что такое данные, почему они значимы, как работают различные мировые редакции с данными, как добывать, анализировать и «доставлять» аудитории

<sup>11</sup> Там же.

<sup>12</sup> Заказные убийства бизнесменов в Петербурге 1992-2005 гг. [Электронный ресурс]. URL:[https://www.google.com/maps/d/embed?mid=1fFcGyENx-yOc3gng-hvFWeDJHv\\_5sr4q&ll=59.90944375183603%2C30.41769619921888&z=10](https://www.google.com/maps/d/embed?mid=1fFcGyENx-yOc3gng-hvFWeDJHv_5sr4q&ll=59.90944375183603%2C30.41769619921888&z=10) (дата обращения 24.01.2020)

<sup>13</sup> В Петербурге составили карту заказных убийств бизнесменов в 90-е. Деловой Петербург. 2018. [Электронный ресурс]. URL:[https://www.dp.ru/a/2018/07/29/V\\_Peterburge\\_sostavili\\_k](https://www.dp.ru/a/2018/07/29/V_Peterburge_sostavili_k) (дата обращения: 24.01.2020)

данные. Дата-журналистика, согласно книге, поможет журналистам создавать более сложные истории с помощью инфографики.

Самые первые статьи о журналистике данных несли скорее прикладной, а не научный характер. В первую очередь медиаисследователи были заинтересованы в редких удавшихся примерах работы СМИ с данными и в разборе конкретных кейсов. Впоследствии дата-журналистикой как феноменом заинтересовались с научной точки зрения, стали изучать американский и европейский опыт.

Следующая важная работа опубликована в 2014 году Марком Коддингтоном, профессором Университета Техаса в Остине, под названием “Clarifying Journalism’s Quantitative Turn”.<sup>14</sup> Согласно сервису Google Scholar, на работу Коддингтона ссылались 348 раз (на 25.04.2020).

Марк Коддингтон утверждает, что новые формы журналистики вовлекают все больше программистов в работу новостных редакций, а профессиональные журналисты все чаще обращаются к техническим возможностям программирования. На становление журналистики данных, по мнению медиаисследователя, повлияло движение в поддержку открытого исходного кода, зародившееся в 1990-х годах. Это движение касалось программного обеспечения, а не журналистики. Однако впоследствии, как разъясняют другие медиаисследователи, профессор Университета Миннесоты Сет Льюис и профессор Университета Джорджа Вашингтона Никки Ашер, принципы, лежащие в его основе, — прозрачность, итерация, исследование и участие (от англ. transparency, iteration, tinkering, and participation) — оказались применимы также и к журналистике. И, таким образом, свели журналистов и программистов.

---

<sup>14</sup> Coddington M. Clarifying Journalism’s Quantitative Turn // Digital Journalism. Vo. 3. 2015. №3. P. 331-348. URL:<https://www.tandfonline.com/doi/full/10.1080/21670811.2014.976400> (дата обращения: 26.01.2020)

Марк Коддингтон выделяет следующие виды применения данных в журналистике: CAR (Computer-assisted reporting), дата-журналистика (от англ. data journalism) и вычислительная журналистика (от англ. computational journalism). Каждый из них не является взаимоисключающим. CAR — это технология, с помощью которой создается журналистский материал на основе компьютерных вычислений (пример мы рассмотрим на стр. 21). Дата-журналистике обычно дается довольно широкое определение, подразумевая под ней практически любую деятельность, сочетающую данные и текст. А под вычислительной журналистикой преимущественно подразумевается сочетание CAR и дата-журналистики.

Круг русскоязычных исследователей журналистики данных на данный момент довольно ограничен. Среди представителей академического сообщества, уделяющих внимание этому феномену, Неренц Д.В., Симакова С.И., Шилина М.Г., Лисицин М.Е., Бережная В.С. и некоторые другие авторы. Кроме того, в России развивается сильное экспертное сообщество, среди представителей которого отмечают Бегтина И., Пархимович О., Дорожного А., Богачева А., Устинова С., Линделл Д. и некоторых других.

Медиаисследователь и доцент ЧелГУ С.И. Симакова в своей работе «Дата-журналистика как медиатренд» дает такое определение журналистике данных: «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»<sup>15</sup>.

---

<sup>15</sup> Симакова С. И. Дата-журналистика как медиатренд // Вестн. Нижегородского ун-та им. Н. И. Лобачевского. 2014. № 2 (2). С. 481–485. [Электронный ресурс]. URL: <https://cyberleninka.ru/article/n/data-zhurnalistika-kak-mediatrend/viewer> (дата обращения: 27.01.2020)

Нам кажется, что данное определение достаточно полно описывает направление дата-журналистики. Действительно, сейчас журналисты не только черпают новости из информационных агентств, выезжают на места происшествий, берут комментарии по телефону, встречаются с героями для интервью, но также анализируют большие объемы открытых данных и находят инфоповоды в базах данных. Если бы не дата-журналисты, некоторые факты могли бы затеряться в потоках информации и остаться незамеченными для общественности.

Благодаря тому, что в России формируется направление журналистики данных, органам государственной власти становится сложнее скрыть факты коррупции. О том, как журналисты проводят расследования на основе данных, мы поговорим в следующей главе.

Иван Бегтин, ведущий специалист по открытым данным в России, эксперт Счетной палаты РФ, генеральный директор Ассоциации участников рынка данных (АУРД) и учредитель АНО «Информационная культура», определяет журналистику данных как «совмещение журналистом роли переводчика с экономического/государственного/цифрового языка» на «язык понятный всем» и в то же время «роли аналитика, представляющего материалы в предельно наглядной форме».<sup>16</sup>

Роль «переводчика» для журналиста не нова. Журналисты и ранее «переводили» с языка государственных чиновников, официальных документов, языка терминологии на язык, понятный читателю. Однако в современную эпоху аудитория требует от журналиста стать переводчиком еще и с «цифрового языка». Раскрытие больших государственных данных не

---

<sup>16</sup> Бегтин И. Памятка по журналистике данных // Сайт Ивана Бегтина. 2011.[Электронный ресурс]. URL:<http://ivan.begtin.name/2011/05/22/datajournalism/> (дата обращения: 29.01.2020)

имело бы смысла, если бы не появились журналисты, способные переработать их и проанализировать.

Как пишет медиаисследователь РГГУ Лисицин в одной из своих работ, «главный принцип журналистики данных – выстраивание материала не вокруг новости, а вокруг статистической справочной информации, цифр, отчетной документации, сводок». Вот почему говоря о журналистике данных, мы также неизбежно затрагиваем статистический анализ, программирование, дизайн<sup>17</sup>.

Создание новости вокруг статистики и цифр требует от журналиста определенных технических навыков, которыми на данный момент владеют немногие. Мы предполагаем, что через несколько лет журналистика может перестать считаться исключительно гуманитарной профессией.

Доктор филологических наук Шилина М.Г. выделяет следующие этапы работы дата-журналиста как основные: «нахождение темы, информации, анализ с помощью программ MySQL, Python и т.п., визуализация и оригинальная подача статистических данных в форме журналистского сюжета (или в качестве равноправной составляющей сюжета)».<sup>18</sup>

И.В. Бегтин говорит, что это и есть самое главное ограничение в дата-журналистике: далеко не все журналисты владеют необходимыми навыками программирования и программами визуализации данных.<sup>19</sup> И, несмотря на то, что использование данных усложняет работу журналиста, медийный

---

<sup>17</sup> Лисицин М. Е. Понятие "журналистика данных" в современных научных публикациях // Коммуникативные исследования. 2018. №3 (17). [Электронный ресурс]. URL:<https://cyberleninka.ru/article/n/ponyatie-zhurnalistika-dannyh-v-sovremennyh-nauchnyh-publikatsiyah> (дата обращения: 29.01.2020)

<sup>18</sup> Шилина М.Г. Data Journalism – дата-журналистика, журналистика метаданных – в структуре медиакommunikации: к вопросу формирования теоретических исследовательских подходов // Электронный научный журнал Медиаскоп. 2013. № 1 [Электронный ре-сурс]. URL:<http://www.mediascope.ru/node/1263> (дата обращения: 29.01.2020)

<sup>19</sup> Бегтин И. Памятка по журналистике данных // Сайт Ивана Бегтина. 2011.[Электронный ресурс]. URL:<http://ivan.begtin.name/2011/05/22/datajournalism/> (дата обращения: 30.01.2020)



продукт «должен быть с легкостью воспринят аудиторией» — пишет С.И. Симакова.<sup>20</sup>

Трансформация журналистики как профессии в техническую сторону началась даже в России. В 2016 году Высшая школа экономики (НИУ ВШЭ) запустила первую магистерскую программу образования по журналистике данных, где студентам преподают машинное обучение, нейронные сети, программирование.

Кандидат филологических наук Д.В. Неренц видит в дата-журналистике скорее дополнительный инструмент, с помощью которого можно «раскрыть базовую информацию о проблеме или явлении», а также визуализировать журналистский материал.<sup>21</sup> По мнению исследователя, данные не являются идеальным источником информации, так как в базах данных могут встречаться неточности, а также отсутствует полная информация. Д.В. Неренц говорит об использовании данных в журналистике как о «начальном этапе работы, требующем значительных доработок».

Однако мы не можем согласиться с исследователем. Неточности, о которых пишет Д.В. Неренц, компенсируются большим объемом данных. Чем больше объем, тем выше точность. Эту позицию также разделяют авторы книги «Большие данные: революция, которая изменит то, как мы живем, работаем и мыслим».<sup>22</sup> Кроме того, качественные открытые государственные базы данных уже становятся реальностью в некоторых государствах. Россия также стремится к этому, и через несколько лет

---

<sup>20</sup> Симакова С. И. Дата-журналистика как медиатренд // Вестн. Нижегородского ун-та им. Н. И. Лобачевского. 2014. № 2 (2). С. 481–485. [Электронный ресурс]. URL: <https://cyberleninka.ru/article/n/data-zhurnalistika-kak-mediatrend/viewer> (дата обращения: 30.01.2020)

<sup>21</sup> Неренц Д.В.. Data journalism: новые возможности для журналистов-расследователей? // Вестник ВГУ. Серия: Филология. Журналистика. 2018. №3. С 147-150. (дата обращения: 31.01.2020)

<sup>22</sup> Майер-Шенбергер В., Кукьер К. Большие данные: революция, которая изменит то, как мы живем, работаем и мыслим. М., 2014. (дата обращения: 30.01.2020)

качество, полнота и актуальность российских открытых государственных данных существенно возрастет.

С.И. Симакова в своей статье указывает на изменение роли журналистики в современном мире. Если раньше журналисты стремились «быстрее всех найти», то теперь важнее становится «грамотно проанализировать большие массивы информации». Таким образом, журналистика данных сможет противостоять журналистике гражданской: у блогера или очевидца нет тех навыков, которыми обладает дата-журналист.

Гражданские журналисты или телеграм-каналы (примеры – Mash, Vaza, Топор и др.), активно развивающиеся в последнее время в российском медиапространстве, способны добывать эксклюзивную информацию, но они не способны анализировать и грамотно перерабатывать большие объемы даты (данных). Современная журналистика так отвечает на вызовы, которые перед ней ставят гражданские журналисты, – будущее в данных. Цифры, а не слова, теперь будут рассказывать истории.

М.Г. Шилина в одной из своих работ также подтверждает эту гипотезу, подчеркивая важность научных изысканий: «теоретические исследования в обозначенном направлении дадут возможность преодолеть творческий кризис корпоративной идентичности современной журналистики на практике».<sup>23</sup>

---

<sup>23</sup> Шилина М.Г. Data Journalism – дата-журналистика, журналистика метаданных – в структуре медиакоммуникации: к вопросу формирования теоретических исследовательских подходов // Электронный научный журнал Медиаскоп. 2013. № 1 [Электронный ресурс]. URL:<http://www.mediascope.ru/node/1263> (дата обращения: 30.01.2020)

## 1.2. Использование данных в мировой медиапрактике

### 1.2.1. Первые примеры использования данных в журналистике

Самые ранние примеры использования данных для общественного блага можно встретить уже в XIX веке.

Одним из первых примеров считается работа журналистов The Guardian. Они опубликовали материал, основанный на данных, 5 мая 1821 года.<sup>24</sup> Материал нашел хороший отклик у читателей, поскольку раскрыл данные, ранее скрытые от аудитории. Статья включала в себя таблицу с информацией о среднем количестве учеников в школах Манчестера и Салфорда, а также о стоимости обучения. Впервые журналистская работа раскрывала количество учеников из бедных семей, а также тех, кто учится бесплатно. Данные показали, что оценки властей о восьми тысячах бесплатно обучающихся детях не верны. Всего таких учеников оказалось около 25 тысяч.

Увидев схожий материал в прессе сейчас, мы бы не удивились, однако в то время подобная информация не раскрывалась для публики. База данных, которая легла в основу материала, оказалась в распоряжении у The Guardian благодаря источнику, именовавшему себя «NH».

Следующим примером использования данных для пользы общества является работа, написанная в 1858 году Флоренс Найтингейл<sup>25</sup>, известной сестрой милосердия и общественной деятельницей Великобритании. Найтингейл, можно сказать, стала дата-журналистом своей эпохи. После Крымской войны (1853-1856гг.), в которой Британская империя потеряла

---

<sup>24</sup>The first Guardian data journalism: May 5, 1821 // The Guardian. 2011.

URL:<https://www.theguardian.com/news/datablog/2011/sep/26/data-journalism-guardian> (дата обращения: 02.02.2020)

<sup>25</sup>Rogers S. Florence Nightingale, datajournalist: information has always been beautiful // The Guardian. 2010.

URL:<https://www.theguardian.com/news/datablog/2010/aug/13/florence-nightingale-graphics> (дата обращения: 03.02.2020)

больше 20 тысяч человек, Флоренс написала труд, где доказала с помощью диаграмм, графиков и таблиц, что смертей многих солдат можно было избежать, улучшив санитарные условия содержания. Данные из трех различных источников, которыми располагала Найтингейл, были неточны и разнились (что также является проблемой для современных дата-журналистов, в особенности, если эти данные государственные), поэтому «леди с лампой» установила новый четкий порядок ведения статистических данных для военных госпиталей и собрала достаточно даты для своего исследования.<sup>26 27 28</sup>

Кроме того, Флоренс Найтингейл сумела наглядно визуализировать ее. На рис. 2 ниже мы видим круговую диаграмму «петушиный гребень», изобретенную Найтингейл и опубликованную в «Заметках о факторах, влияющих на здоровье, эффективность и управление госпиталями британской армии». Правая диаграмма показывает первые 12 месяцев войны, левая – следующие 12, то есть до и после того, как из Лондона прислали комиссию по улучшению гигиены. Синий цвет обозначает смертность от болезней, красный – от ранений, черный – от других причин. Видно, что смертность существенно сократилась после улучшения санитарных условий.

---

<sup>26</sup> Kopf E.W. Florence Nightingale as Statistician // Publications of the American Statistical Association. Vol. 15, No. 116. 1916. P. 388-404. URL:[https://www.jstor.org/stable/2965763?seq=3#metadata\\_info\\_tab\\_contents](https://www.jstor.org/stable/2965763?seq=3#metadata_info_tab_contents) (дата обращения: 03.02.2020)

<sup>27</sup> Симакова С.И. Журналистика данных как объективное требование времени и ее влияние на формирование визуальной журналистики // Знак: проблемное поле медиаобразования. 2016. №1 (18). С. 18-25. [Электронный ресурс] URL:<https://cyberleninka.ru/article/n/zhurnalistika-dannyh-kak-obektivnoe-trebovanie-vremeni-i-eyo-vliyanie-na-formirovanie-vizualnoy-zhurnalistiki/viewer> (дата обращения: 04.02.2020)

<sup>28</sup> Florence Nightingale: The Lady with the Data // This is Statistics. 2016. URL:<https://thisisstatistics.org/florence-nightingale-the-lady-with-the-data/> (дата обращения: 04.02.2020)

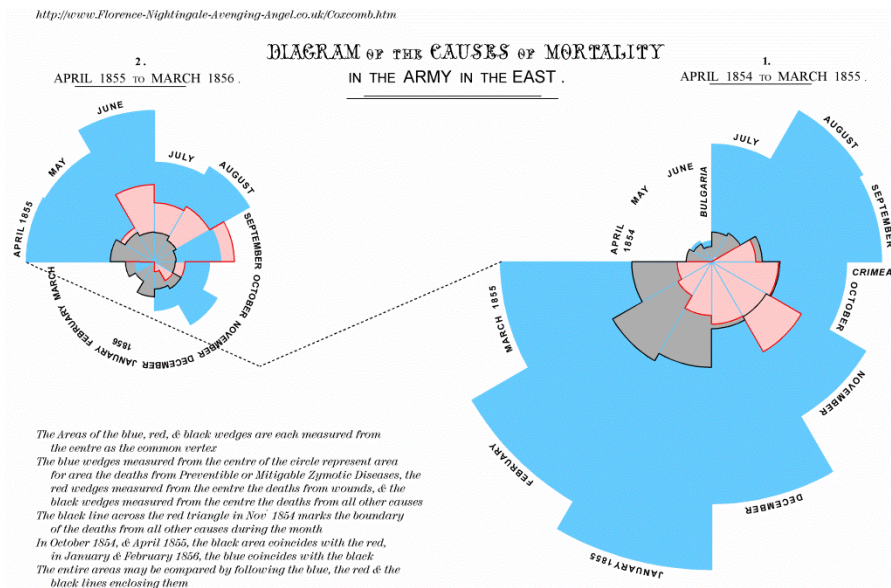


Рис. 2. Диаграмма Найтингейл «Петушиный гребень»

Ещё один пример использования данных в журналистике имел место в 1952 году во время выборов 34-го президента США. Американский канал CBS решил арендовать компьютер UNIVAC, чтобы предсказать, кто одержит победу в президентской гонке. Работники канала опросили всего 1% населения США, и на основе этих данных компьютер смог вычислить, что Дуайт Эйзенхауэр станет новым американским президентом с точностью до 98%. Этому виду журналистики впоследствии дадут название Computer-assisted-reporting (CAR).

### 1.2.2 Мировое законодательство в области раскрытия данных

Современное мировое законодательство предусматривает раскрытие государственных данных. Таким образом, у журналистов появились открытые базы данных, на основе которых теперь можно проводить расследования.

«Хартия открытых данных» была принята в июне 2013 года в Северной Ирландии на саммите G8, куда входили Канада, Франция, Германия, Италия,

Япония, Россия, Великобритания, США и Европейский союз.<sup>29</sup> Согласно документу, страны-участники должны следовать следующим принципам: «открытые данные по умолчанию», «качество и количество», «использование всеми», «опубликование данных для оптимизации управления» и «опубликование данных для инноваций». Это значит, что актуальные и точные государственные данные должны публиковаться в открытом доступе. В качестве примеров таких данных можно привести криминальную статистику, метеоданные, реестр организаций, государственный бюджет, потребление энергии, общегосударственные карты, перепись населения, результаты научных экспериментов и др.<sup>30</sup>

Однако доктрина, которая лежит в основе «Хартии открытых данных», появилась годами раньше. Впервые ее сформулировал герцог Ланкастер в докладе Парламенту Великобритании в 1993 году и дал ей название «открытое государство» (от англ. Open Government). Иван Бегтин приводит в своей статье определение этой доктрины, взятое из документа: «Открытое государство – это часть эффективной демократии. Граждане должны иметь адекватный доступ к информации и анализу, на основе которого государство работает. Министры и публичные служащие обязаны разъяснять их законы, решения и действия общественности. Государство, тем не менее, нуждается в том, чтобы хранить некоторые секреты, и обязано защищать личную жизнь тех, кем оно управляет».<sup>31</sup>

Далее определение открытого государства формулировалось неоднократно, и, в конце концов, в сентябре 2011 года была создана международная организация Open Government Partnership (OGP), чья

---

<sup>29</sup> Хартия открытых данных "Группы восьми" // Открытые данные России. [Электронный ресурс] URL:<https://data.gov.ru/hartiya-otkrytyh-dannyh-gruppy-vosmi> (дата обращения: 07.02.2020)

<sup>30</sup> Бюллетень по открытым данным. 2015. №1. [Электронный ресурс] URL:<https://ac.gov.ru/archive/files/publication/a/5572.pdf> (дата обращения: 10.02.2020)

<sup>31</sup> Бегтин И. Открытое государство. История и новый смысл // Полит.ру. 2012. [Электронный ресурс] URL:[https://polit.ru/article/2012/06/28/open\\_gov/](https://polit.ru/article/2012/06/28/open_gov/) (дата обращения: 10.02.2020)

деятельность направлена на раскрытие и доступность данных по всему миру. Цель OGP состоит в том, чтобы больше государств становились более прозрачными и подотчетными гражданам, чтобы в итоге повысить качество управления и доверие между государством и людьми. Странами-основателями являются Бразилия, Индонезия, Мексика, Норвегия, Филиппины, ЮАР, Великобритания и США. Стран-участниц на данный момент 78.<sup>32</sup> Россия планировала войти в OGP, однако в 2012 году это решение было отменено.<sup>33</sup>

### *1.2.3 Журналистика данных и ведущие мировые СМИ*

Благодаря открытым данным, многие мировые СМИ смогли создавать свои дата-проекты и собирать дата-отделы. The Guardian, The Washington Post, The New York Times, The Daily Telegraph, The Los Angeles Times, ABC, BBC, Zeit Online, Reuters, La Nacion, The Financial Times, Der Spiegel, Bloomberg News и многие другие СМИ включили в свои материалы данные. Рассмотрим самые яркие примеры.

Новостной редактор The Guardian Саймон Роджерс после беседы с Эдрианом Головатым в 2009 году осознал, что можно использовать данные для того, чтобы читатели понимали этот мир лучше, и запустил в газете дата-проект под названием «Datablog». Саймон Роджерс предпринял первую системную попытку включить общедоступные данные в новости.<sup>34</sup> Сначала в The Guardian решили опубликовать 200 различных наборов данных в открытом формате: уровень преступности, экономические показатели, данные с недели моды и другие, а через время уже взялись за полноценные дата-истории.

---

<sup>32</sup>About Open Government Partnership // Open Government Partnership.  
URL:<https://www.opengovpartnership.org/about/> (дата обращения: 12.02.2020)

<sup>33</sup>Бегтин И. Открытое правительство – миф или история? // Ivan Begtin Blog. 2018. [Электронный ресурс].  
URL:<https://begtin.tech/failed-gov/> (дата обращения: 12.02.2020)

<sup>34</sup>The Guardian Datablog // Multiple Journalism. URL:<http://multiplejournalism.org/case/the-guardian-datablog>  
(дата обращения: 14.02.2020)

В том же 2009 году случилась революция в сфере открытых данных – тогда правительство США запустило сайт с государственными открытыми данными, включающий в себя на тот момент всего 47 наборов данных. Затем примеру США последовали другие государства по всему миру, и журналисты получили доступ к открытым данным.

Год спустя, благодаря международной некоммерческой организации «Wikileaks», в руки команды журналистов-расследователей The Guardian попали военные данные США из Ирака и Афганистана. Так дата-журналистика попала в отдел новостей. К концу 2011 года появился совместный проект The Guardian и LSE (The London School of Economics and Political Science) «Reading Riots», созданный с помощью технологии CAR, о беспорядках августа 2011 года в Англии.<sup>35</sup>

The Washington Post запустил дата-колонку «Wonkblog» о политике и экономике в 2009 году. Над колонкой единолично работал журналист Эзра Клейн в течение пяти лет, став настоящей звездой The Washington Post. Затем команда расширилась до пяти человек, а в 2014 году ему на смену пришел Джим Танкерслей. Как пишет Лисицин в одном из своих исследований, «Wonkblog» до сих пор остается одной из «самых прогрессивных специализированных на журналистике данных колонок».<sup>36</sup>

С 2010 по 2013 годы в The New York Times действовал успешный дата-проект «FiveThirtyEight» Нейта Сильвера. Причина массового успеха проекта Нейта Сильвера заключалась в статистическом подходе при рассказе о

---

<sup>35</sup> Gambini L. A decade of working in data journalism: what has changed? Interview with Simon Rogers, data editor at the Google News Lab // Medium. 2019. URL:<https://medium.com/we-are-the-european-journalism-centre/a-decade-of-working-in-data-journalism-what-has-changed-8d950d99935e> (дата обращения: 15.02.2020)

<sup>36</sup> Лисицин М.Е. Понятие журналистика данных в освещении Российских и зарубежных СМИ. 2018. [Электронный ресурс]. URL:[http://www2.rsuh.ru/binary/object\\_28.1551881843.02622.pdf](http://www2.rsuh.ru/binary/object_28.1551881843.02622.pdf) (дата обращения: 15.02.2020)



политике.<sup>37</sup> «FiveThirtyEight» прославился тем, что верно предсказал победителя президентских выборов США 2008 года.

В 2012 году The New York Times публикует интерактивную карту<sup>38</sup>, где показано, как двигался ураган «Сэнди», нанесший наибольший урон северу восточного побережья США. Посмотреть на карту можно на рис. 3 ниже.



Рис. 3. Путь урагана Сэнди, проект The New York Times<sup>39</sup>

Через два года, в 2014 году, газета запускает собственную дата-колонку «The Upshot» на смену ушедшему Нейту Сильверу со своим проектом «FiveThirtyEight» в ESPN. Цель «The Upshot» – освещать политику и повседневную жизнь по-новому – с помощью визуализаций данных и

---

<sup>37</sup> McDuling J. “The Upshot” is the New York Times’ replacement for Nate Silver’s FiveThirtyEight // Quartz. 2014. URL: <https://qz.com/185922/the-upshot-is-the-new-york-times-replacement-for-nate-silvers-fivethirtyeight/> (дата обращения: 15.02.2020)

<sup>38</sup> Hurricane Sandy // The New York Times. 2012. URL: <http://archive.nytimes.com/www.nytimes.com/packages/html/newsgraphics/2012/0820-hurricane-map/sandy-phone.html> (дата обращения: 15.02.2020)

<sup>39</sup> Там же.

интерактивных функций. Созданием проекта занимался редактор Дэвид Леонхардт и 14 человек его команды.

Среди лучших работ своей дата-команды, которые отмечает сам The New York Times, можно найти следующие материалы: «Medicare спасет миллиарды или будет стоить миллиарды?», «Как шеф полиции, губернатор и социолог потратили бы \$100 миллиардов на решение опиумного кризиса?», «Максимально детализированная карта президентских выборов 2016», «Смерти от наркотиков в Америке растут быстрее, чем когда-либо», «Лучшие и худшие места для взросления», «Типичный американец живет всего лишь в 18 милях от мамы», «Лучше снимать или покупать?» и многие другие.<sup>40</sup>

#### ***1.2.4 Международные премии для дата-журналистов***

Журналистика данных стала настолько общественно значимой, что очень важно поддерживать журналистов, включающих данные в свои истории. Кроме того, необходимо повышать престиж дата-журналистики как профессии, чтобы привлекать новых талантливых дата-журналистов. Для этого существуют премии в области журналистики данных.

Самой главной международной премией в области журналистики данных до недавнего времени была Data Journalism Awards. Премия появилась в 2011 году и просуществовала вплоть до закрытия ее организатора – Глобальной сети редакторов (от англ. Global Editors Network, GEN) – в 2019 году.<sup>41</sup>

Победители определялись в 12 номинациях: за новаторство в дата-журналистике, среди студентов и молодых дата-журналистов года, за лучшее

---

<sup>40</sup> The Upshot, Five Years In // The New York Times. 2019.  
URL:<https://www.nytimes.com/interactive/2019/04/22/upshot/upshot-at-five-years.html> (дата обращения: 16.02.2020)

<sup>41</sup> Лисицин М.Е. Феномен журналистики данных в зарубежных СМИ // Litera. 2019. № 1. С. 102-115 [Электронный ресурс]. URL:[https://e-notabene.ru/fil/article\\_29056.html](https://e-notabene.ru/fil/article_29056.html) (дата обращения: 16.02.2020)

портфолио дата-команды (среди крупных редакций), за лучшее портфолио дата-команды (среди небольших редакций), за лучшее индивидуальное портфолио, за лучшее использование данных в срочных новостях (в течение первых 36 часов после события), за лучший веб-сайт в области журналистики данных, за визуализацию данных года, за расследование года, за новостное приложение для работы с данными, за использование открытых данных, выбор публики.<sup>42</sup>

В 2019 году российский журналист «РБК» Дада Линделл выиграла Data Journalism Awards в номинации «Лучший молодой дата-журналист» по работам из собственного портфолио.<sup>43</sup> Кроме Дады Линделл, была номинирована другая журналистка «РБК» – Полина Звезда – в номинации «Приз зрительских симпатий» за исследование «В России сменилась тройка регионов-лидеров по смертности от ВИЧ».<sup>44</sup>

Вознаграждения за победу в Data Journalism Awards были совсем небольшие – \$1801 в дополнении к статуэтке. Сумма была выбрана неслучайно: в 1801 году Уильям Плейфэр, в честь которого утверждена премия, придумал круговую диаграмму. По сравнению с Пулитцеровской премией, лауреаты Data Journalism Awards получали более чем в 8 раз меньшее вознаграждение.<sup>45</sup>

На смену Data Journalism Awards в 2020 году пришла новая награда – Sigma Awards. Награждаются девять победителей в шести категориях: за

---

<sup>42</sup> 2019 Shortlist // Data Journalism Awards. URL:<https://datajournalismawards.org/2019-shortlist/> (дата обращения: 17.02.2020)

<sup>43</sup> Журналистка РБК получила международную премию Data Journalism Awards // РБК. 2019. [Электронный ресурс] URL:<https://www.rbc.ru/rbcfreenews/5d024eeb9a79473f90310a49> (дата обращения: 17.02.2020)

<sup>44</sup> Линделл Д., Звезда П., Тарасенко Е. В России сменилась тройка регионов — лидеров по смертности от ВИЧ // РБК. 2019. [Электронный ресурс]. URL:<https://www.rbc.ru/society/11/03/2019/5c6c0ae49a79477703d915b7> (дата обращения: 17.02.2020)

<sup>45</sup> Подходит к концу конкурс Data Journalism Awards 2018, в последний день мая уже будут вручены премии // Телеграм-канал Журналистика данных. 2018. [Электронный ресурс]. URL:<https://t.me/digitalfan/268> (дата обращения: 18.02.2020)

лучший репортаж на основе данных (среди крупных и небольших редакций), за лучшую визуализацию данных (среди крупных и небольших редакций), за новаторство (среди крупных и небольших редакций), среди лучших молодых журналистов, за использование открытых данных, за лучшее приложение для работы с данными.<sup>46</sup>

Теперь в награду лучшие дата-журналисты получают статуэтку, возможность поучаствовать в дискуссиях и воркшопах Международного журналистского фестиваля (от англ. International Journalism Festival) и полную компенсацию затрат на поездку для 2 членов журналистской команды.<sup>47</sup>

Жюри нового конкурса Sigma Awards уже отметили достижения в том числе российских журналисток Алеси Мароховской и Ирины Долининой за дата-расследование «Тройка Ландромат» вместе с OCCRP (Центр по исследованию коррупции и организованной преступности).<sup>48</sup> Журналисты, участвовавшие в проекте, победили в номинации «Лучшие репортажи на основе данных», отследив миллиарды долларов, отмываемые через сеть офшоров. Над расследованием работали журналисты из 23 медиа по всему миру.<sup>49</sup>

Еще один престижный конкурс в области визуализации данных называется Kantar Information is Beautiful Awards. Награда появилась в 2014 году и дается за лучшие визуализации данных.<sup>50</sup> Всего награждается 35

---

<sup>46</sup> Sigma Awards: new data journalism competition launched // Online Journalism Blog. URL:<https://onlinejournalismblog.com/2020/01/09/sigma-awards-new-data-journalism-competition-launched/> (дата обращения: 18.02.2020)

<sup>47</sup> The 2020 Sigma Awards // Data Journalism. URL:<https://datajournalism.com/awards> (дата обращения: 18.02.2020)

<sup>48</sup> Ландромат Тройка. OCCRP. 2019. [Электронный ресурс]. URL:<https://www.occrp.org/ru/troikalaundromat/> (дата обращения: 18.02.2020)

<sup>49</sup> Announcing: The Sigma Awards 2020 winners. Data Journalism. 2020. URL:<https://datajournalism.com/awards>

<sup>50</sup> 6 Years of Outstanding Outfits // Kantar Information is beautiful awards. URL:<https://www.informationisbeautifulawards.com/news/539-6-years-of-outstanding-outfits> (дата обращения: 18.02.2020)

победителей в 10 категориях: досуг, игры и спорт; политика и глобальный мир; наука и технологии; искусство, развлечения и культура; восходящая звезда; визуализация на иностранном языке; самая красивая визуализация и другие.

В 2017 году работа студии инфографики ТАСС «Когда пришел Наполеон» взяла золото в номинации «Окружающая среда и карты». ТАСС стал первым из российских СМИ, кому вручили награду.<sup>51</sup>

Призовой фонд премии составляет \$27,000: каждый победитель в зависимости от места (бронза, серебро, золото) и категории получает \$250, \$1000 или \$5,000. Самая крупная сумма в \$5,000 достается за самую красивую визуализацию.<sup>52</sup>

Также награды в области дата-журналистики вручают на Online Journalism Awards в категории «Награда Университета Флориды за лучшую журналистику расследований на основе данных». Награда выдается с 2014 года крупным и небольшим редакциям. Победители получают \$15,000.<sup>53</sup>

### **1.3. Использование данных в отечественной медиапрактике**

#### ***1.3.1 Российский опыт: законодательный и практический аспекты***

Согласно рейтингу Open Data Barometer, по уровню открытости правительственных данных Россия на 2017 год (более новые данные отсутствуют) занимает 13 место. Канада, Великобритания, Австралия,

---

<sup>51</sup> ТАСС получил "золото" престижной премии Kantar Information is Beautiful Awards // ТАСС. 2017. [Электронный ресурс]. URL: <https://tass.ru/obschestvo/4766823> (дата обращения: 18.02.2020)

<sup>52</sup> Information is Beautiful Awards 2019 // Kantar Information is beautiful. 2019. URL:<https://informationisbeautiful.net/2019/information-is-beautiful-awards-2019-now-open/> (дата обращения: 18.02.2020)

<sup>53</sup> The University of Florida Award for Investigative Data Journalism // Online Journalism Awards. URL:<https://awards.journalists.org/awards/investigative-data-journalism/> (дата обращения: 18.02.2020)

Франция и Южная Корея входят в первую пятерку.<sup>54</sup> Некоторые государственные данные до сих пор недоступны или слабо доступны в России для массовой аудитории.

В России первое определение понятия открытых данных появилось в 2006 году в федеральном законе «Об информации, информационных технологиях и о защите информации», однако тогда под этим термином имелась в виду в принципе любая информация, размещаемая ее обладателями в Интернете.<sup>55</sup>

Следующее определение публикуется в «Методических рекомендациях по публикации открытых данных» в версии 3.0, утвержденных протоколом заседания Правительственной комиссии по координации деятельности открытого правительства (то же, что и открытое государство) в мае 2014 года. Теперь уже под него попадала информация «в виде систематизированных данных» для «свободного и бесплатного использования».<sup>56</sup>

Сами открытые данные регулируются в России с 2012 года<sup>57</sup> после подписания Указа Президента «Об основных направлениях совершенствования системы государственного управления».<sup>58</sup> Согласно Указу, до июля 2013 года должен быть предоставлен доступ к открытым данным информационных систем органов государственной власти РФ.

---

<sup>54</sup> The Open Data Barometer // Open Data Barometer.

URL:[https://opendatabarometer.org/?\\_year=2017&indicator=ODB](https://opendatabarometer.org/?_year=2017&indicator=ODB) (дата обращения: 19.03.2020)

<sup>55</sup> N 149-ФЗ Об информации, информационных технологиях и о защите информации // Консультант Плюс. 2006. [Электронный ресурс]. URL:[http://www.consultant.ru/document/cons\\_doc\\_LAW\\_61798/](http://www.consultant.ru/document/cons_doc_LAW_61798/) (дата обращения: 20.02.2020)

<sup>56</sup> Методические рекомендации по публикации открытых данных Версия 3.0 // Открытые данные России. 2014. С. 6 [Электронный ресурс]. URL:<https://data.gov.ru/metodicheskie-rekomendacii-po-publikacii-otkrytyh-dannyh-versiya-30> (дата обращения: 20.02.2020)

<sup>57</sup> Бюллетень по открытым данным. 2015. №1. [Электронный ресурс]

URL:<https://ac.gov.ru/archive/files/publication/a/5572.pdf> (дата обращения: 20.02.2020)

<sup>58</sup> Указ Президента Российской Федерации №601 Об основных направлениях совершенствования системы государственного управления // Официальный интернет-портал правовой информации. 2012. [Электронный ресурс]. URL:<http://pravo.gov.ru/proxy/ips/?docbody=&nd=102156335> (дата обращения: 20.02.2020)

В июне 2013 года была принята редакция федерального закона «Об обеспечении доступа к информации о деятельности государственных органов и органов местного самоуправления», по которому государственные данные должны публиковаться в открытой форме.<sup>59</sup> Благодаря этому закону, количество порталов с открытыми государственными данными в РФ возросло.<sup>60</sup>

Источники открытых государственных данных в России следующие: портал открытых данных РФ ([data.gov.ru](http://data.gov.ru)), открытые данные Министерства финансов ([minfin.ru/opendata](http://minfin.ru/opendata)), открытые данные Министерства культуры ([opendata.mkrf.ru/opendata](http://opendata.mkrf.ru/opendata)), статистика Центрального Банка ([cbr.ru/statistics](http://cbr.ru/statistics)), открытые данные ФНС (<https://www.nalog.ru/opendata/>), федеральная информационная адресная система ([fias.nalog.ru](http://fias.nalog.ru)), портал открытых данных города Москвы ([data.mos.ru](http://data.mos.ru)), Росстат ([gks.ru/opendata/](http://gks.ru/opendata/)), открытые данные Министерства юстиции ([minjust.ru/ru/opendata](http://minjust.ru/ru/opendata)), портал закупок ([zakupki.gov.ru](http://zakupki.gov.ru)) и другие.<sup>61</sup>

Все эти государственные проекты открытых данных реализуются в рамках «открытого государства», о котором мы писали на стр. 22. Однако до сих пор некоторые государственные данные в России полностью или частично закрыты от публики, также часто встречаются ошибки. Ольга Пархимович, руководитель разработки проекта Счетной палаты РФ «Госрасходы», и ведущий эксперт в области открытых данных Иван Бегтин в своей статье для «РБК» сообщают, что часто в процессе создания некоммерческого проекта «Госрасходы» сталкивались с публикацией

---

<sup>59</sup> N 112-ФЗ Об обеспечении доступа к информации о деятельности государственных органов и органов местного самоуправления // Официальный интернет-портал правовой информации. [Электронный ресурс]. URL:<http://pravo.gov.ru/proxy/ips/?docbody=&nd=102127629> (дата обращения: 20.02.2020)

<sup>60</sup> Лисицин М.Е. Проекты в области журналистики данных в российских СМИ (на примере газеты «Деловой Петербург») // Знак: проблемное поле медиаобразования. 2018. №4 (30). [Электронный ресурс]. URL:<https://cyberleninka.ru/article/n/proekty-v-oblasti-zhurnalistiki-dannyh-v-rossiyskih-smi-na-materiale-gazety-delovoy-peterburg/viewer> (дата обращения: 20.02.2020)

<sup>61</sup> Там же.

неверной информации госорганами.<sup>62</sup> Одна из главных причин низкого качества данных – «отсутствие механизмов контроля ввода информации». Кроме того, например, Минобрнауки до сих пор не публикует открытые данные, а в Минкомсвязи отсутствуют данные почти по всем государственным информационным системам, таким как Госуслуги, ФГИС КИ (Федеральная государственная информационная система координации информатизации) и другие<sup>63</sup>.

### *1.3.2 Общественные проекты с использованием открытых данных*

Благодаря раскрытию важных государственных данных, стало возможным создание некоммерческих общественных проектов, собирающих данные из разных государственных источников в одну базу. Таким образом, Россия все ближе становится к тому, что называется «открытое государство», стандарты которого намного более успешно выполняются на западе.

Проекты на основе открытых данных помогают журналистике исполнять «сторожевую функцию» (от англ. watchdog journalism), а также делают более удобной работу журналистов с данными. Теперь журналисты могут обращаться к общественным проектам, чтобы собрать данные для написания материала. В особенности, это касается журналистики расследований. Чаще всего на порталах проектов объединены данные из различных открытых источников, поэтому журналистам зачастую удобнее пользоваться базами данных проектов, чем искать и собирать данные вручную с разных государственных сайтов.

---

<sup>62</sup> Пархимович О., Бегтин И. Данные неверны: почему российские госсайты вводят граждан в заблуждение // РБК. 2019. [Электронный ресурс].

URL:<https://www.rbc.ru/opinions/society/20/12/2019/5dfb4e169a79470b46b4f98c> (дата обращения: 20.02.2020)

<sup>63</sup> Бегтин И. Немного о том, как сейчас публикуют открытые данные ФГИС // Телеграм-канал Ivan Begtin. 2020. [Электронный ресурс]. URL: <https://t.me/begtin/1720> (дата обращения: 20.02.2020)



Одними из самых видных игроков являются АНО «Информационная культура»<sup>64</sup>, «Трансперенси Интернешнл Россия»<sup>65</sup> и Комитет гражданских инициатив (КГИ)<sup>66</sup>. Именно они создают большинство общественных проектов на основе открытых государственных данных в России на данный момент.

Один из самых значимых общественных проектов – портал «Декларатор»<sup>67</sup>, публикующий декларации о доходах и имуществе российских чиновников. Проект создан в 2011 году «Трансперенси Интернешнл Россия». «Декларатор» собирает все исходные файлы с нужными сведениями, которые публичные должностные лица обязаны публиковать в соответствии с российским законодательством, и переводит их в машиночитаемый формат. Главная проблема в том, что эти данные публикуются не в едином формате и на сайте разных ведомств. «Декларатор» видит свою цель в том, чтобы собирать эти данные в одну большую базу, полезную для журналистов-расследователей.<sup>68</sup>

Например, «Новая газета» использовала данные портала «Декларатор» в своем дата-расследовании «ВИП-кочевники».<sup>69</sup> В материале журналисты пишут о том, как команды российских чиновников перемещались с одной должности на другую при распределении госконтрактов вместе своими

---

<sup>64</sup> АНО «Информационная культура» появилась в марте 2012 года силами основателей конкурса Apps4Russia («конкурс посвящен разработке управленческих и бизнес проектов на основе открытых данных»), среди которых также есть Иван Бегтин.

<sup>65</sup> «Трансперенси Интернешнл Россия» – это российское отделение неправительственной международной организации по противодействию коррупции. «Трансперенси Интернешнл Россия» выступает «за принципы прозрачности, подотчетности, порядочности и честности». В России организация появилась в 1999 году.

<sup>66</sup> Комитет гражданских инициатив создавался Алексеем Кудриным, председателем Счетной палаты РФ, Министром финансов РФ с 2000 по 2011 гг., и рядом других общественных деятелей в апреле 2012 года как сообщество экспертов из различных сфер деятельности в ответ на протесты на Болотной площади. Стало понятно, что граждане хотят принимать участие в политической жизни страны и влиять на внутренние процессы.

<sup>67</sup> Декларатор. [Электронный ресурс]. URL:<https://declarator.org> (дата обращения: 21.02.2020)

<sup>68</sup> О проекте // Декларатор. [Электронный ресурс]. URL:<https://declarator.org/about/> (дата обращения: 21.02.2020)

<sup>69</sup> Долинина И., Мароховская А. ВИП-кочевники // Новая газета. 2018. [Электронный ресурс]. URL: <https://novayagazeta.ru/articles/2018/09/17/77842-svoih-ne-brosaem> (дата обращения: 21.02.2020)

земляками и бывшими коллегами. Для этого Ирина Долинина и Алеся Мароховская, журналисты data-отдела «Новой», проанализировали «миграции» 59 тысяч чиновников.

Еще один пример – расследование «Трансперенси Интернешнл Россия» «Квартирный вопрос генерала Беляева»<sup>70</sup>, где журналисты расследовали происхождение множества квартир, принадлежащих генералу и его семье. Для этого журналисты активно изучили декларации Беляева на портале «Декларатор». Кроме того, важное преимущество проекта в том, что «Декларатор» сохраняет все декларации, когда-либо опубликованные чиновником. Таким образом, журналисты смогли заметить, что одна из квартир, находившихся в пользовании генерала Беляева, впоследствии перешла в его собственность.

Кроме того, проект использовали в следующих дата-журналистских материалах, которые нам удалось найти: «Штабы на лямках»<sup>71</sup> («Новая газета»), «Квартира для тюремщика»<sup>72</sup> («Трансперенси Интернешнл Россия»), «Царь-квартира Игоря Шувалова стоит как 600 обычных квартир, над которыми он смеялся»<sup>73</sup> («ФБК»).

Еще один не менее важный общественный проект называется «ГосЗатраты».<sup>74</sup> Проект совместно создали АНО «Информационная культура» и Комитет гражданских инициатив в 2013 году на базе уже существовавшего с 2010 года портала «РосГосЗатраты»<sup>75</sup>. Цель проекта –

---

<sup>70</sup> Квартирный вопрос генерала Беляева // Трансперенси Интернешнл Россия. 2017. [Электронный ресурс]. URL:<https://transparency.org.ru/special/belyaev/> (дата обращения: 21.02.2020)

<sup>71</sup> Шпилькин С., Заякин А. Штабы на лямках //Новая газета. 2019. [Электронный ресурс]. URL:<https://novayagazeta.ru/articles/2019/08/01/81457-shtaby-na-lyamkah> (дата обращения: 22.02.2020)

<sup>72</sup> Квартира для тюремщика // Трансперенси Интернешнл Россия. 2016. [Электронный ресурс]. URL:<https://transparency.org.ru/special/fsin/> (дата обращения: 22.02.2020)

<sup>73</sup> Царь-квартира Игоря Шувалова стоит как 600 обычных квартир, над кото-рыми он смеялся // ФБК.2016. [Электронный ресурс]. URL:<https://fbk.info/investigations/post/230/> (дата обращения: 22.02.2020)

<sup>74</sup> ГосЗатраты. [Электронный ресурс]. URL:<https://clearspending.ru> (дата обращения: 22.02.2020)

<sup>75</sup> FAQ // ГосЗатраты. [Электронный ресурс]. URL:<https://clearspending.ru/page/about/faq/#Istoria> (дата обращения: 22.02.2020)

«повышение общественной осведомленности в области расходования государственных средств». Проект охватывает открытые данные о государственных контрактах и закупках с 2011 года. Используется журналистами-расследователями в качестве базы для своих материалов.

На основе данных проекта «ГосЗатраты» журналисты «РБК» провели исследование «Как ИП в России получают миллиарды от государства».<sup>76</sup> Для этого «РБК» нашли и проанализировали все контракты с января 2013 года по ноябрь 2017 года с помощью «ГосЗатрат». Журналисты обнаружили, что в России зарегистрировано десять ИП, получивших миллиардные контракты за последние 5 лет от государства.

Еще один пример использования портала «ГосЗатраты» – рейтинг Forbes «Короли госзаказа» за 2014 год.<sup>77</sup> Используя данные о государственных контрактах с «ГосЗатрат», журналисты смогли обнаружить крупнейшие господряды. В дополнении к «ГосЗатратам», журналисты также использовали базы данных СПАРК и ЕИС (Единая информационная система в сфере закупок).

«Медуза» использовала данные проекта «ГосЗатраты» для своего теста «Госзакупка или выдумка?». Журналисты вместе с Комитетом гражданских инициатив изучили самые интересные государственные сделки 2014-2015 гг. и составили по ним тест. Среди них, к примеру, раскраски и видеоуроки по рисованию за 320 миллионов рублей Министерству здравоохранения, пять тонн конины за 1,2 миллиона рублей управлению ФСИН по Иркутской

---

<sup>76</sup> Житкова В., Линделл Д., Гришин Н. Исследование РБК: как ИП в России получают миллиарды от государства // РБК. 2017. [Электронный ресурс]. URL:[https://www.rbc.ru/own\\_business/13/12/2017/5a1d68bb9a7947745d083338](https://www.rbc.ru/own_business/13/12/2017/5a1d68bb9a7947745d083338) (дата обращения: 22.02.2020)

<sup>77</sup> Рейтинг // Forbes. 2015. [Электронный ресурс]. URL:<https://www.forbes.ru/forbes/issue/2015-03/280925-reiting> (дата обращения: 22.02.2020)

области, портреты Владимира Путина за полмиллиона рублей администрации главы Чечни.<sup>78</sup>

Счетная палата РФ запустила в 2019 году похожий проект «Госрасходы»<sup>79</sup>, цель которого сделать доходы и расходы государства более прозрачными. Проект в первую очередь акцентирует внимание на национальных и федеральных проектах, госпрограммах, нефтегазовых доходах, госконтрактах, субсидиях и грантах.<sup>80</sup> Портал агрегирует данные из разных открытых источников: ЕИС, электронного бюджета, портала госпрограмм, паспорта нацпроектов, сайтов органов власти. На данный момент портал содержит данные о более чем 38 миллионах контрактов (на 25.04.2020).

На основе данных «Госрасходов» «Новая газета» выпустила материал «Розы, пианино и жилье: как «Роскосмос», Минпромторг и Минобороны стали крупнейшими госзаказчиками за последние годы».<sup>81</sup> Благодаря порталу, российские читатели узнали о том, кто из госкорпораций и министерств тратит больше всего, сколько именно и на что.

Некоммерческий проект «Росправосудие» был создан в апреле 2012 года и просуществовал до июля 2018 года. Сайт представлял собой справочно-правовую систему с опубликованными решениями российских судов. Роскомнадзор заблокировал проект в 2018 году из-за нарушения закона о персональных данных. Более 87 миллионов судебных актов по

---

<sup>78</sup> Госзакупка или выдумка? Совместный тест проекта КГИ «Госзатраты» и «Медузы» // Медуза. 2015. [Электронный ресурс]. URL:<https://meduza.io/quiz/goszakupka-ili-vydumka> (дата обращения: 22.02.2020)

<sup>79</sup> Госрасходы. [Электронный ресурс]. URL:<https://spending.gov.ru> (дата обращения: 22.02.2020)

<sup>80</sup> О проекте // Госрасходы. [Электронный ресурс]. URL:<https://spending.gov.ru/pages/about/> (дата обращения: 22.02.2020)

<sup>81</sup> Джорджевич А. Розы, пианино и жилье: как «Роскосмос», Минпромторг и Минобороны стали крупнейшими госзаказчиками за последние годы // Новая газета. 2019. [Электронный ресурс]. URL:<https://novayagazeta.ru/articles/2019/11/18/82775-rozy-pianino-i-zhilie> (дата обращения: 22.02.2020)

данным на ноябрь 2015 года было опубликовано на сайте «Росправосудия».<sup>82</sup> Похожие проекты (например, «Самосуд», «Судебные решения») также блокировались Роскомнадзором.

Пример использования сервиса «Росправосудие» – расследование «Новой газеты» «Мелкие преступники: как административный штраф в России может превратиться в уголовное дело».<sup>83</sup> Материал подготовили журналисты data-отдела «Новой» Андрей Заякин (сооснователь «Диссернета», редактор data-отдела) и Алексей Смагин. Они утверждают, что штраф в 1000 рублей, наложенный на должностное лицо и оплаченный со счета предприятия, может превратиться в уголовное дело по статье 160 УК РФ «Присвоение или растрата». Журналисты выгрузили решения судов с «Росправосудия» по ключевым фразам и, таким образом, нашли 88 решений, которые они затем проанализировали вручную, и пришли к выводу, что за «мелочи» люди получали реальные сроки.

Следующий пример – «Судьи держат нас за болванку»<sup>84</sup>, расследование «Новой газеты». Для написания материала журналисты взяли из системы «Росправосудие» 653 решения судов первой инстанции по 228 статье УК РФ – это те дела, в текстах которых использовалось словосочетание «кондитерский мак». В фабулах некоторых решений нашлись совпадения, что говорит об отсутствии индивидуального подхода к подозреваемым.

---

<sup>82</sup> Маляренко Е. Сайт «Росправосудие» решили заблокировать из-за публикации решения суда // РБК. 2018. URL:<https://www.rbc.ru/society/18/07/2018/5b4f6b8e9a79476a4e8951f2> (дата обращения: 23.02.2020)

<sup>83</sup> Заякин А., Смагин А. Мелкие преступники // Новая газета. 2019. Электронный ресурс. URL:<https://novayagazeta.ru/articles/2019/07/03/81119-melkie-prestupniki> (дата обращения: 23.02.2020)

<sup>84</sup> Заякин А., Смагин А. Судьи держат нас за болванку // Новая газета. 2019. [Электронный ресурс]. URL: <https://novayagazeta.ru/articles/2019/03/20/79929-sudi-derzhat-nas-za-bolvanku> (дата обращения: 23.02.2020)

Кроме того, система также использовалась в еще одном расследовании «Новой газеты» – «Стыдно сказать».<sup>85</sup> Журналисты выяснили, что решения судов по делам об экстремизме почти недоступны, их публикуемость всего 2%. Авторы материала предполагают, что это связано с абсурдностью обвинений. Для того чтобы прийти к данным выводам, журналистам пришлось обратиться к решениям, опубликованным в нескольких онлайн-системах, в том числе «Росправосудии».

Проект «Открытые НКО»<sup>86</sup> создан в 2017 году АНО «Информационная культура». Для создания портала использовались базы данных реестра НКО, президентских грантов, реестра субсидий, ЕИС, ЕГРЮЛ (Единый государственный реестр юридических лиц). На портале доступна картотека с данными обо всех НКО, которые можно найти в открытом доступе.<sup>87</sup> Всего на данный момент проект включает в себя данные о 727,808 (на 25.04.2020) НКО. В эти данные входит информация о финансировании, учредителях и руководителях, роде деятельности.<sup>88</sup>

Проект Комитета гражданских инициатив «Открытая полиция»<sup>89</sup> создан силами АНО «Информационная культура» от лица КГИ в 2012 году для мониторинга деятельности полиции, сбора идей от граждан и выявлению критичных проблем.<sup>90</sup> Проект полностью основан на открытых данных, публикуемых на интернет-ресурсах МВД. На данный момент сайт не обновляется, есть доступ только к архиву, актуальному на май 2019 года.

---

<sup>85</sup> Смагин А., Заякин А. Стыдно сказать // Новая газета. 2018. [Электронный ресурс]. URL:<https://novayagazeta.ru/articles/2018/06/29/76970-stydno-skazat> (дата обращения: 23.02.2020)

<sup>86</sup> Открытые НКО. [Электронный ресурс]. URL:<https://openngo.ru> (дата обращения: 23.02.2020)

<sup>87</sup> Открытые НКО // Комитет гражданских инициатив. [Электронный ресурс]. URL:<https://komitetgi.ru/projects/3467/> (дата обращения: 23.02.2020)

<sup>88</sup> О проекте // Открытые НКО. URL:<https://openngo.ru/about/> (дата обращения: 23.02.2020)

<sup>89</sup> Открытая полиция. [Электронный ресурс]. URL:<https://openpolice.ru> (дата обращения: 23.02.2020)

<sup>90</sup> О проекте // Открытая полиция. [Электронный ресурс]. URL: <https://www.openpolice.ru/pages/about/> (дата обращения: 23.02.2020)

## ВЫВОДЫ ПО ГЛАВЕ 1

В данной главе мы рассмотрели журналистику данных в мировой и отечественной медиапрактиках. Это было необходимо нам для того, чтобы перед проведением собственного исследования понять, что представляет собой дата-журналистика, какие факторы повлияли на ее развитие и как данные используются в мировых редакциях.

К журналистике данных можно отнести материалы, в основе которых лежат данные или же данные дополняют журналистскую историю с помощью инфографики.

Стоит отметить, что одни из самых первых примеров применения данных для общественного блага появились еще в XIX веке, а уже на сегодняшний день данные используются в главных редакциях мира: The Guardian, The Washington Post, The New York Times, ABC, BBC, Zeit Online, Reuters, Spiegel и других.

На данный момент теоретическая часть в значительной степени отстает от практической. Российских исследователей, заинтересованных в данной теме, не очень много, однако экспертное сообщество активно развивается в стране, продвигая инициативы и создавая различные проекты, основанные на данных.

Изменения в российском законодательстве и создание общественных проектов на основе данных являются, на наш взгляд, основными драйверами становления дата-журналистики в России.

Правительства вынуждены раскрывать данные о государственных закупках, декларациях чиновников, уголовных делах и так далее, благодаря чему у журналистов появились дополнительные ресурсы, на основе которых можно проводить расследования.

## ГЛАВА 2. ЖУРНАЛИСТСКИЕ РАССЛЕДОВАНИЯ НА ОСНОВЕ ДАННЫХ В РОССИИ

### 2.1. Подход к анализу

Во второй главе мы ставим себе цель провести анализ журналистских расследований на основе данных в России. Всего мы нашли 54 публикации, которые подходят, по нашему мнению, под данное определение, в 15 различных медиа.

Данный анализ не претендует на абсолютно полный охват публикаций, которых можно отнести к журналистским расследованиям на основе данных в России. Вероятно, есть материалы, которые мы упустили или посчитали, что к дата-расследованиям их отнести нельзя, так как журналист пользовался и другими источниками, помимо данных.

Данное исследование стоит воспринимать как попытку понять методику работы дата-журналистов на определенной выборке публикаций (хотя мы включили в данное исследование абсолютно все материалы, которые смогли обнаружить). Сначала мы рассмотрим методику анализа, затем определим основные черты расследований на основе данных, а после проанализируем самые заметные публикации.

Для начала мы считаем целесообразным определить, что такое «журналистика расследований», и напомнить, что мы подразумеваем под «данными».

Во время анализа публикаций перед нами встал вопрос, какие журналистские материалы относятся к расследованиям, а какие – к исследованиям. Например, в «РБК» журналисты Дада Линделл, Полина Звездина и Евгений Тарасенко создают множество дата-материалов, как, например, «В России сменилась тройка регионов – лидеров по смертности от



ВИЧ». <sup>91</sup> Нам показалось, что небольшие признаки журналистского расследования в этом материале, безусловно, есть, но все-таки отнести его к расследовательской журналистике не представляется возможным.

Однако некоторые материалы с признаками расследования мы решили включить в наш анализ. Чаще всего данные публикации представляли собой новость, раскрывающую информацию, крайне значимую для общества и ранее не известную, написанную благодаря изучению данных. Например, новости о выявлении нарушений на выборах. Сами по себе такие новости нельзя назвать полноценными расследованиями, они даже по своему объему довольно незначительны, однако раскрывают важную информацию для общества.

Для того чтобы понять, что можно считать журналистскими расследованиями, а что – нет, мы обратились к работе доктора филологических наук, профессора А.А. Тертычного «Расследовательская журналистика». <sup>92</sup> «В качестве предмета журналистского расследования выступают прежде всего разного рода преступления, происшествия, конфликты, существование которых кто-то пытается скрыть от общества, исторические и иные (кроме государственной и военной) тайны <...>. Если же они берутся открыть нечто скрытое, то в этом случае будут ставить и решать задачу журналиста-расследователя,» – пишет А.А. Тертычный. Таким образом, журналистским расследованием можно назвать тот материал, в котором раскрывается информация, ранее скрытая от общественности, и в основе которого лежит факт нарушения закона или столкновения нескольких сторон (например, государства и человека).

---

<sup>91</sup> Линделл Д., Звездина П., Тарасенко Е. В России сменилась тройка регионов — лидеров по смертности от ВИЧ // РБК. 2019. [Электронный ресурс]. URL: <https://www.rbc.ru/society/11/03/2019/5c6c0ae49a79477703d915b7> (дата обращения: 29.02.2020)

<sup>92</sup> Тертычный А.А. Расследовательская журналистика: учебное пособие для вузов. М.: Аспект Пресс. 2002.

Под данными мы подразумеваем в первую очередь открытые большие объемы информации, которые должны соответствовать нескольким критериям: свобода доступа, свобода использования и машиночитаемость.<sup>93</sup> То есть это как раз те данные, которые, согласно принципам открытого государства и Международной хартии открытых данных (их мы упоминали в первой главе), должны раскрывать правительства стран. Свобода доступа означает, что любой человек, имея выход в интернет, может получить бесплатный доступ к этим данным. Свобода использования предполагает отсутствие авторских прав, то есть эта свобода ничем не ограничена ни для кого из пользователей. Машиночитаемость – это возможность для компьютера прочитать эти данные (а затем и проанализировать), то есть они должны быть предоставлены в структурированном виде.

Журналистские материалы, основанные на открытых данных, отвечающих всем трем критериям, в большинстве своем могут быть реплицируемы, то есть потенциально воспроизведены кем-либо из читателей. На основе методологии и указываемых источников потенциально любой человек должен иметь возможность воспроизвести те же самые результаты, которые указывает журналист. В действительности, скорее всего, это сможет сделать человек с опытом в анализе данных, однако никаких иных препятствий быть не должно. В этом и заключается крайне значимая характеристика дата-материалов – реплицируемость. Далее мы рассмотрим, сколько публикаций из 54 мы можем назвать реплицируемыми.

О ценности реплицируемости для дата-материалов говорила Дада Линделл, дата-журналист «РБК» и победитель Data Journalism Awards, на Дне открытых данных в Москве 2020 во время дискуссии «Считать больно.

---

<sup>93</sup> Открытые данные // сайт Экспертного совета при Правительстве РФ. [Электронный ресурс]. URL:<https://opendata.open.gov.ru/event/5598184/> (дата обращения: 29.02.2020)

Проблемы создания дата-материалов»<sup>94</sup>. По словам Линделл, именно благодаря возможности повторного воспроизведения материала, аудитория доверяет дата-журналисту. Для достижения реплицируемости, а, следовательно, и доверия читателей, должна быть четко прописана методология, то есть последовательность действий дата-журналиста. На данные, которыми пользуется дата-журналист, должны быть оформлены ссылки, а также те данные, которые журналист собирает сам из различных источников, должны быть выложены в публичный доступ (например, на GitHub).<sup>95</sup>

Однако в реальности российские дата-журналисты так делают редко. Причина этого кроется в том, что на данный момент журналисты не понимают, где и как нужно публиковать эти данные (на сайте газеты, GitHub, Google drive<sup>96</sup> или др.). Кроме того, не хватает ресурсов на публикацию данных, так как для этого нужно привести таблицы в достаточно читаемый вид, а на это также требуется время журналиста.

Еще одна причина, по которой редакции могут не давать ссылки, не публиковать данные, а тем более скрипты (компьютерный код), – недобросовестное последующее использование этих материалов другими коллегами-журналистами. К примеру, после дата-публикаций «РБК», по словам Дады Линделл<sup>97</sup>, другие СМИ могли позаимствовать результаты исследования и затем опубликовать их на сайте как свои собственные без ссылки на «РБК».

Ирина Долинина, корреспондентка дата-отдела «Новой газеты» и победительница конкурса Sigma Awards, во время дискуссии ответила на это тем, что российским журналистам пора перестать конкурировать между

---

<sup>94</sup> Линделл Д. Дискуссия Считать больно // День открытых данных. 2020.

<sup>95</sup> GitHub. [Электронный ресурс]. URL: <https://github.com/> (дата обращения: 01.03.2020)

<sup>96</sup> Google Drive. [Электронный ресурс]. URL: <https://www.google.com/drive/> (дата обращения: 01.03.2020)

<sup>97</sup> Там же.

собой и начать объединяться и кооперироваться, отмечая, что заграничные издания уже приходят к этому.<sup>98</sup> Делиться данными, по словам журналистки, может быть полезно для тех коллег, которые захотят продолжить историю и смогут увидеть в данных что-то новое.

Именно благодаря кооперации журналистов, существует организация OCCRP (Центр по исследованию коррупции и организованной преступности)<sup>99</sup>, расследования которой мы будем анализировать далее. Для создания своих масштабных расследований OCCRP привлекает журналистов со всего мира, в том числе Ирину Долинину и Алесю Мароховскую, дата-журналисток «Новой газеты». А с недавнего времени к OCCRP также присоединилась Дада Линделл.

Мы также будем рассматривать журналистские материалы, основанные на данных, которые трем критериям открытости данных (свобода доступа, свобода использования и машиночитаемость) могут соответствовать не полностью. Хотя основу исследования будут составлять все-таки материалы на тех данных, о которых мы описали выше.

Мы это делаем по причине того, что журналистика данных – молодое явление в российском медиапространстве, и для нас важно охватить всевозможные ее аспекты. Например, открытые данные могут дополняться собственными данными редакции, или же весь материал может строиться только на своих данных. Так поступил Институт проблем правоприменения при работе над проектом «Преступность в тени»<sup>100</sup>: для создания материала исследователи опросили 16 тысяч человек.

---

<sup>98</sup> Долинина И. Дискуссия Считать больно // День открытых данных. 2020.

<sup>99</sup> OCCRP. [Электронный ресурс]. URL:<https://www.occrp.org/ru/home> (дата обращения: 08.03.2020)

<sup>100</sup> Бублик М., Тоганова Н. Преступность в тени // Институт проблем правоприменения. 2019. [Электронный ресурс]. URL:<https://readymag.com/u85821431/russiancrimesurvey/> (дата обращения: 08.03.2020)

Также мы рассмотрели несколько расследований Фонда борьбы с коррупцией («ФБК»)<sup>101</sup>. Чаще всего такие расследования строятся не на данных, а на открытой информации (интервью героев, фотографии и геолокации в социальных сетях). Однако данным в их расследованиях тоже есть место (например, выписки из Росреестра, декларации чиновников).

Между информацией и данными есть существенные различия. Данные – это необработанные факты и цифры без нарратива. Данные могут считываться компьютером, а информация – нет. Кроме того, данные сами по себе смысла не несут, в отличие от информации. Смысл в данных находит журналист и доносит его до аудитории.

## **2.2. Методика анализа**

Для того чтобы проанализировать 54 публикации, мы создали таблицу (см. таблицу 1 в приложении) и для удобства разделили ее на следующие столбцы: №, медиа, заголовок, авторы, дата, объем, тематика, просмотры, вид ИД (источника данных), ИД, данные, срок взятых данных, количество ИД, доступность данных для публики, количество визуализаций, типы визуализаций, иллюстрации и фотографии, комментарии экспертов, интерактивность, реплицируемость, публикация данных, техника работы с данными, ссылка. Каждый из найденных материалов мы заносили в таблицу и анализировали по указанным критериям.

Тематика материалов нам представляется важной для изучения, так как мы пытаемся понять, про что больше всего пишут дата-журналисты-расследователи. Варианты тематик у нас оказались следующие: бизнес, городская среда, ЖКХ, медицина, наркотики, недвижимость, образование,

---

<sup>101</sup> Фонд борьбы с коррупцией. [Электронный ресурс]. URL:<https://fbk.info/> (дата обращения: 10.03.2020)

общество, пиратство, политика, правосудие, экономика. Найти материалы на другие тематики нам не удалось.

Нас интересуют также медиа, где был опубликован материал, дата публикации, объем материала, количество просмотров (если доступно) и заголовок.

Источник данных – один из ключевых критериев для анализа. Типы источников данных мы разделили в соответствии с классификацией Андрея Дорожного (одного из самых видных дата-журналистов), которую он давал в ходе своего семичасового онлайн-курса «Data-вебинары с Андреем Дорожным: данные для начинающих»<sup>102</sup> в марте 2020 года: государственные, коммерческие, общественные, альтернативные (например, краудсорсинг, скрейпинг – автоматизированный сбор данных с сайтов).

Государство – самый крупный владелец данных. Данные государства обычно открыты (или их хотя бы можно запросить и попытаться раскрыть). Запрашивать данные могут как журналисты (по статье 39 Федерального закона о СМИ<sup>103</sup>; ответ на запрос занимает до 7 дней), так и обычные граждане (по статье 18, пункт 6 Федерального закона №8 «Об обеспечении доступа к информации о деятельности государственных органов и органов местного самоуправления»<sup>104</sup>; ответ на запрос занимает до 30 дней). Правила оформления запроса стоит смотреть на сайте органа, куда вы обращаетесь. Качество государственных данных обычно невысокое: встречаются ошибки, данные могут не обновляться или отсутствовать вовсе.

---

<sup>102</sup> Дорожный А. Data-вебинары с Андреем Дорожным: данные для начинающих // Теплица социальных технологий. 2020.

<sup>103</sup> Закон РФ от 27.12.1991 N 2124-1 (ред. от 01.03.2020) "О средствах массовой информации" // КонсультантПлюс [Электронный ресурс].  
URL:[http://www.consultant.ru/document/cons\\_doc\\_LAW\\_1511/1599bf6f7ccc24d4966727daa0243ffe117f5ac3/](http://www.consultant.ru/document/cons_doc_LAW_1511/1599bf6f7ccc24d4966727daa0243ffe117f5ac3/)  
(дата обращения: 15.03.2020)

<sup>104</sup> Федеральный закон от 09.02.2009 N 8-ФЗ (ред. от 28.12.2017) "Об обеспечении доступа к информации о деятельности государственных органов и органов местного самоуправления"// КонсультантПлюс [Электронный ресурс].  
URL:[http://www.consultant.ru/document/cons\\_doc\\_LAW\\_84602/bc8f145b9b0cf81ba35eefedf48e421cf362e0a5/](http://www.consultant.ru/document/cons_doc_LAW_84602/bc8f145b9b0cf81ba35eefedf48e421cf362e0a5/)  
(дата обращения: 12.03.2020)

Общественные организации (НКО и другие) также владеют данными. Такие данные открыты, и их качество также обычно невысоко.

Коммерческие данные, или данные компаний, обычно очень высокого качества, так как зачастую такие данные используются для продажи, и конкуренция побуждает компании отвечать за качество своего продукта. Однако, как уже понятно, компания либо продает данные, либо делает их доступными по специальным соглашениям.

Альтернативные источники – это любые другие источники, которые не подходят под определение государственных, коммерческих и общественных. Под альтернативными источниками могут подразумеваться скрейпинг и парсинг данных, то есть их автоматизированный сбор с помощью программного кода или специальных расширений браузера (например, Instant Data Scraper, Table Capture, Web Scraper). Кроме того, краудсорсинг также является альтернативным источником данных. Под краудсорсингом мы подразумеваем самостоятельный сбор данных с помощью волонтеров или, к примеру, Google форм.

В рамках одной публикации могут быть задействованы самые разные источники данных. Зачастую они совмещаются и дополняют друг друга. По данной причине нас интересует количество источников данных, чтобы понять, сколько в среднем источников данных необходимо для написания расследования.

Также мы смотрим на сами данные, которые исследует медиа, чтобы понять, что именно заинтересовало журналистов в дата-сети (наборе данных), а также масштаб самих данных. Некоторые материалы, проанализированные в данной работе, основаны на настоящих больших данных, однако это скорее исключения.

Срок взятых данных также важен для данной работы, так как этот критерий может помочь молодым специалистам при прочтении этого исследования определиться с тем, какой срок данных может быть

репрезентативен для их дата-расследования. Журналист должен понимать, за какой период данные ему необходимы. Вероятно, для того, чтобы проследить определенные тенденции, одного года будет мало, так как в этот промежуток могли произойти нетипичные события, оказавшие влияние на данные.

В ходе нашего анализа мы также учитываем доступность данных для широкой публики, так как именно этот критерий напрямую влияет на реплицируемость, о которой мы говорили ранее. Если данные доступны и читатель сам может убедиться в правдивости журналистского материала, доверие аудитории к данной публикации повышается. Если данные доступны только для журналиста, то у пытливого читателя могут появиться вопросы в реальности данных.

Если данные не доступны, то редакция может их опубликовать. Если данные доступны, но на основе данных велись специализированные расчеты или каким-то образом данные обрабатывались, то такие данные журналисты тоже могут опубликовать. Таким образом, повышается доверие аудитории к данным и самому материалу. Отсюда возник еще один критерий – публикация данных.

Визуализации – значимый аспект дата-журналистики, поэтому в данной работе мы изучаем количество визуализаций и их типы. Кроме того, нам интересно соотношение обычных иллюстраций и дата-визуализаций, чтобы понять, насколько остаются важны обычные иллюстрации.

В данной работе мы не будем рассматривать целесообразность и качество визуализаций, так как это отдельная тема для исследования. Однако стоит отметить, что журналист всегда предварительно должен задаваться вопросом о необходимости диаграммы или графика. В некоторых случаях может оказаться, что объяснить словами легче, чем показать на визуализации. Кроме того, стоит помнить о том, для какой аудитории (насколько она подготовлена) пишет журналист. Некоторые визуализации могут оказаться слишком сложны для восприятия читателя.



Типы визуализаций мы определяли в соответствии с классификацией, представленной в электронной книге, вышедшей в апреле 2020 года, «Графики, которые убеждают всех»<sup>105</sup> одного из главных специалистов по визуализации данных в России Александра Богачева. Если не удавалось определить тип графика или диаграммы по книге, то мы использовали классификацию, представленную на сайте Каталога визуализации данных, который рекомендовал один из самых видных дата-журналистов Андрей Дорожный.<sup>106</sup> В случае, если тип визуализации было невозможно найти ни в одном из классификаторов, мы писали «инфографика».

Мы также намерены проанализировать материалы на наличие или отсутствие комментариев экспертов. У нас появилась гипотеза, согласно которой дата-материалам не нужны комментарии экспертов, потому что их роль выполняют сами данные и журналист. Согласно одному из пунктов «Догмы газеты “Ведомости”»<sup>107</sup>, «любые утверждения, выводы или оценки, которые дает журналист, должны быть подтверждены *либо* фактами и цифрами, *либо* мнениями экспертов». Мы предполагаем, что при наличии фактов и цифр мнение эксперта не обязательно. В течение нашего исследования мы проверим эту гипотезу.

Интерактивность – еще один критерий из нашей таблицы. Под интерактивностью мы понимаем любую нестатичную визуализацию, с которой может взаимодействовать наш читатель, то есть это и кликабельные материалы, и те, где читатели могут сами, например, вводить и проверять информацию.

---

<sup>105</sup> Богачев А. Графики, которые убеждают всех. 2020. [Электронный ресурс].

<sup>106</sup> Каталог визуализации данных. [Электронный ресурс]. URL:<https://datavizcatalogue.com/RU/> (дата обращения: 03.04.2020)

<sup>107</sup> Догма газеты Ведомости. [Электронный ресурс]. URL:<http://mediakar.org/wp-content/uploads/2012/10/d0b4d0bed0b3d0bcd0b0-d0b2d0b5d0b4d0bed0bcd0bed181d182d0b8.pdf> (дата обращения: 29.04.2020)

Следующий критерий – реплицируемость, то есть возможность читателя повторить ход расследования самому – найти данные или информацию, произвести те же вычисления, что и автор. Необходимо это, как мы уже писали ранее, для того, чтобы аудитория верила журналисту. В ходе нашего исследования мы выявляли, может материал быть реплицируемым или нет. Для того чтобы признать материал реплицируемым, он должен соответствовать следующим критериям: открытые или опубликованные данные, есть ссылка на данные или их легко найти, четко прописанная методология.

Кроме того, мы намереваемся понять, какие техники работы с данными используют дата-журналисты. Не все авторы указывают используемую технику. Возможно, это происходит, потому что они использовали обычный Excel. Возможно, потому что не считают это важным и интересным для аудитории. Мы рассмотрим те случаи, когда техника указывалась.

В течение данного исследования мы будем приводить короткие выдержки из нашей полной таблицы, которую можно найти в приложениях.

Сначала мы проанализируем журналистские расследования на основе данных в целом и, таким образом, узнаем, какие черты свойственны дата-расследованиям. Затем перейдем к анализу публикаций в «Новой газете»<sup>108</sup>, «Трансперенси Интернешнл Россия»<sup>109</sup>, «Проекте»<sup>110</sup>, «ФБК»<sup>111</sup>, «РБК»<sup>112</sup>, так как в данных медиа мы насчитали больше всего таких материалов. И после рассмотрим несколько наиболее интересных, на наш взгляд, примеров.

---

<sup>108</sup> Новая газета. [Электронный ресурс]. URL: <https://novayagazeta.ru/> (дата обращения: 10.04.2020)

<sup>109</sup> Трансперенси Интернешнл Россия. [Электронный ресурс]. URL: <https://transparency.org.ru/> (дата обращения: 10.04.2020)

<sup>110</sup> Проект. [Электронный ресурс]. URL: <https://www.proekt.media/> (дата обращения: 10.04.2020)

<sup>111</sup> Фонд борьбы с коррупцией. [Электронный ресурс]. URL: <https://fbk.info/> (дата обращения: 10.04.2020)

<sup>112</sup> РБК. [Электронный ресурс]. URL: <https://www.rbc.ru/> (дата обращения: 10.04.2020)

### 2.3. Основные характеристики журналистских расследований на основе данных

В этом подпункте мы проанализируем агрегированные данные обо всех 54 публикациях, чтобы узнать, какие тенденции доминируют в журналистике расследований на основе данных. Мы составим сводные таблицы из данных нашего анализа и, таким образом, сможем проанализировать все 54 публикации.

Самый первый материал из нашего анализа был опубликован в июне 2016 года, самый поздний – в феврале 2020 года.

В таблице 2 ниже приведены медиа и количество дата-расследований, найденных на их онлайн-ресурсах.

Медиа	Количество публикаций
Новая газета	16
Трансперенси Интернешнл Россия	7
Проект	7
ФБК	6
РБК	5
Деловой Петербург	2
Инфометр	2
The Bell	1
ОCCRP, Meduza	1
ОCCRP, Новая газета	1
Meduza, Трансперенси Интернешнл Россия	1
Ivan Begtin Blog	1
Фонд «Нужна помощь»	1
Открытые медиа	1
Институт проблем правоприменения	1
Другой город	1
<b>Итого</b>	<b>54</b>

Таблица 2. Количество дата-расследований в медиа

Сначала рассмотрим, на какие темы больше всего проводилось расследований на основе данных. В таблице 3 ниже можно увидеть распределение материалов по тематикам.

Тематика	Количество публикаций
Политика	21
Правосудие	12
Экономика	5
Городская среда	5
ЖКХ	3
Бизнес	2
Пиратство	1
Общество	1

Образование	1
Недвижимость	1
Наркотики	1
Медицина	1
<b>Итого</b>	<b>54</b>

Таблица 3. Тематика расследований на основе данных в России

Больше всего (21 из 54) расследований пришлось на политические темы. Журналисты расследовали перемещения российских чиновников целыми командами, влияние надомного голосования на выборы, лоббирование интересов, недошедшие до бюджета города деньги, аномалии в голосовании на выборах губернатора, покупку элитной недвижимости родственниками чиновников и многое другое.

Мы предполагаем, что политических расследований оказалось больше всего по причине того, что в основном журналисты использовали государственные источники информации, раскрывающие, в большинстве своем политическую информацию: госзакупки, декларации чиновников, итоги голосований на выборах и т.д.

На втором месте по количеству публикаций – правосудие. На правосудие пришлось 12 публикаций. Журналисты анализировали материалы уголовных дел, судебные вердикты, данные о помиловании, а в одном случае исследователи даже сами опросили 16 тысяч человек.<sup>113</sup> Расследования на данную тематику в основном раскрывали скрытые тенденции или доказывали их существование в российском правовом регулировании. Например, журналисты на основе данных смогли доказать, что «чувство несправедливости» по отношению к фигурантам «Московского дела» оказалось небезосновательным.<sup>114</sup> Кроме того, журналисты, использовав данные, опровергли высказывание замминистра юстиции, что

<sup>113</sup> Бублик М., Тоганова Н. Преступность в тени // Институт проблем правоприменения. 2019. [Электронный ресурс]. URL:<https://readymag.com/u85821431/russiancrimesurvey/> (дата обращения: 18.04.2020)

<sup>114</sup> Мароховская А., Щенников А., Мартынова Е. Если подойдешь, я воткну тебе вилы в горло // Новая газета. 2020. [Электронный ресурс]. URL:<https://novayagazeta.ru/articles/2020/02/13/83892-ya-tebya-zarublyu-musor> (дата обращения: 18.04.2020)

проблема домашнего насилия преувеличена.<sup>115</sup> Еще один пример – журналисты доказали, что институт помилования превратился в инструмент улучшения репутации режима при Владимире Путине.<sup>116</sup>

Основные источники в дата-расследованиях о правосудии в России – база судебных решений ГАС РФ «Правосудие»<sup>117</sup>, общественный проект «Росправосудие» (уже не действует, был заблокирован Роскомнадзором)<sup>118</sup>, общественный проект «Судебные и нормативные акты РФ»<sup>119</sup>.

Для таких расследований журналисты в некоторых случаях анализировали несколько тысяч строк данных. Анализу подвергались 60 тысяч судебных вердиктов, 12 тысяч судебных решений, 4,5 тысячи судебных приговоров. Без помощи компьютера журналисты изучали бы такие данные годами.

Экономических расследований из 54 материалов, анализируемых нами, мы насчитали пять. Журналисты посчитали, во сколько миллионов рублей обойдется празднование Нового года, и выясняли, почему некоторые из подрядчиков оказались скрыты.<sup>120</sup> В другом материале – высчитывали долги населения за ЖКХ, искали самых главных должников среди регионов и выясняли, сколько денег недополучили компании.<sup>121</sup> В еще одном

---

<sup>115</sup> Я тебя сейчас, сука, убивать буду // Новая газета. 2019. [Электронный ресурс]. URL: <https://novayagazeta.ru/articles/2019/11/25/82847-ya-tebya-seychas-suka-ubivat-budu> (дата обращения: 18.04.2020)

<sup>116</sup> Аренина Е. Немилостивый государь // Проект. 2019. [Электронный ресурс]. URL: <https://www.proekt.media/research/statistika-pomilovaniya/> (дата обращения: 18.04.2020)

<sup>117</sup> Государственная автоматизированная система Российской Федерации «Правосудие» интернет-портал. [Электронный ресурс]. URL: <https://sudrf.ru/> (дата обращения: 18.04.2020)

<sup>118</sup> Маляренко Е. Сайт «Росправосудие» решили заблокировать из-за публикации решения суда // РБК. 2018. URL: <https://www.rbc.ru/society/18/07/2018/5b4f6b8e9a79476a4e8951f2> (дата обращения: 18.04.2020)

<sup>119</sup> Судебные и нормативные акты РФ. [Электронный ресурс]. URL: <https://sudact.ru/> (дата обращения: 18.04.2020)

<sup>120</sup> Линделл Д. Новый год обойдется Москве почти в миллиард рублей // РБК. 2019. [Электронный ресурс]. URL: <https://www.rbc.ru/society/27/12/2019/5dfcd8769a79476a5a3ef489> (дата обращения: 18.04.2020)

<sup>121</sup> Савина С. Суверенный долг // Проект. 2019. [Электронный ресурс]. URL: <https://www.proekt.media/research/kommunalnye-dolgi/> (дата обращения: 18.04.2020)

расследовании журналисты раскрыли схему отмывания денег, по которой из России вывели 700 миллиардов рублей.<sup>122</sup>

Заметно, что отличительной чертой экономических расследований является наличие в материале конкретных сумм денег, которые были украдены, выведены из страны или кем-то не получены.

Два из пяти экономических материала основаны на «утекших» данных, и это единственные публикации из всего нашего анализа, которые строились на таком типе данных. Это два расследования-проекта OCCRP: «Российская финансовая "мегапрачечная"» (совместно с «Новой газетой» и другими СМИ)<sup>123</sup> и «Тройка Ландромат» (совместно с «Meduza» и другими СМИ).<sup>124</sup>

Еще пять расследований мы отнесли к городской среде, хотя некоторые из них подошли бы и под экономическую тематику. Журналисты расследовали, кто зарабатывает на ремонте дворов в Москве<sup>125</sup>, почему в Москве ремонтируют одни и те же улицы<sup>126</sup>, почему в Самаре в 2018 году сбили больше пешеходов, чем годом ранее,<sup>127</sup> и т.д. В основном для написания этих материалов использовались следующие источники: статистика ГИБДД<sup>128</sup> и портал госзакупок.<sup>129</sup>

---

<sup>122</sup> Российская финансовая «мегапрачечная» // OCCRP. 2014. [Электронный ресурс]. URL:<https://www.occrp.org/ru/laundromat/> (дата обращения: 18.04.2020)

<sup>123</sup> Там же.

<sup>124</sup> Ландромат Тройка // OCCRP. 2019. [Электронный ресурс]. URL: <https://www.occrp.org/ru/troikalaundromat/> (дата обращения: 18.04.2020)

<sup>125</sup> Линделл Д. Кто зарабатывает на ремонте дворов в Москве // РБК. 2019. [Электронный ресурс]. URL: <https://www.rbc.ru/society/25/09/2019/5d7f5ad09a7947436751c99b> (дата обращения: 18.04.2020)

<sup>126</sup> Линделл Д, Тарасенко Е. Разрытие покажет: почему в Москве ремонтируют одни и те же улицы // РБК. 2019. [Электронный ресурс]. URL:<https://www.rbc.ru/society/21/08/2019/5d5a951b9a7947b5b3dc6022> (дата обращения: 18.04.2020)

<sup>127</sup> Место смерти // Другой город. 2019. [Электронный ресурс]. URL: <https://drugoigorod.ru/place-of-death/> (дата обращения: 18.04.2020)

<sup>128</sup> Показатели состояния безопасности дорожного движения. [Электронный ресурс]. URL:<http://stat.gibdd.ru/> (дата обращения: 18.04.2020)

<sup>129</sup> Единая информационная система в сфере закупок. [Электронный ресурс]. URL:<https://zakupki.gov.ru/epz/main/public/home.html> (дата обращения: 18.04.2020)

На тематику ЖКХ мы обнаружили три материала, о бизнесе – два, о пиратстве, обществе, образовании, недвижимости, наркотиках и медицине – по одному.

Следующий важный аспект дата-журналистики, который мы намереваемся проанализировать, – источники данных. Согласно нашему анализу, журналисты использовали от 1 до 6 источников в одной публикации, среднее значение – 2 источника на одно расследование. В 12 случаях нам не удалось сосчитать, сколько источников использовалось, так как не в каждом материале методология была четко прописана. Статистика говорит нам о том, что для написания расследования одного источника данных недостаточно. Журналистам приходится в среднем искать два набора данных для публикации.

Теперь рассмотрим, к каким типам источников данных журналисты обращались чаще всего (см. таблицу 4). Согласно нашему исследованию, в 41 случаях использовались государственные источники, в 17 – общественные, в девяти – коммерческие, лишь в двух – альтернативные. Мы затруднились определить, к какому типу источников относятся утечки данных, поэтому выделили их в отдельную категорию. Публикаций, которые строились на таком типе данных, оказалось две. Не удалось определить, какие источники использовались в трех случаях. Стоит не забывать о том, что в публикации могли использоваться несколько видов источников данных одновременно.

<b>Вид источника данных</b>	<b>Частота использования в материалах</b>
Государственный	41
Общественный	17
Коммерческий	9
Альтернативный	2
Утечка данных	2
Не удалось определить	3

Таблица 4. Частота использования разных видов источников данных в материалах

Как мы уже писали ранее, реплицируемость – одно из важнейших качеств дата-материалов. От того, может материал быть реплицируем или

нет, зависит доверие читателей. Рассмотрим наши 54 расследования на основе данных и узнаем, сколько материалов соответствуют данному критерию.

Оказалось, что большинство расследований соответствуют критерию реплицируемости: 34 материала (63%) читатель, обладающий некоторыми навыками, может самостоятельно воспроизвести. Остальные 20 расследований (37%) не отвечают данному критерию.

Доступность данных для широкой публики – один из факторов, влияющих на реплицируемость материала. Согласно нашему анализу, 41 расследование (76%) строится на доступных данных. Материалов, где данные доступны частично (означает, что один из источников доступен, а другой – нет) всего пять. Тех, где источник недоступен для широкой аудитории, мы насчитали еще пять. Точно не известно, доступны источники или нет, в трех случаях.

Еще один фактор – публикация данных. Медиа может публиковать данные по-разному и на разных платформах, а может не публиковать в принципе.

Как нам удалось обнаружить, данные были опубликованы всего в 14 случаях, в остальных 40 – нет. Вероятно, публикация данных не была необходима, так как данные взяты из открытого источника и расчеты с ними не производились. Возможны также и другие причины, о которых мы писали на стр. 43.

Один из вариантов публикации данных, который часто использовали анализируемые нами медиа, – разместить файл Excel на собственном сайте. Так в большинстве случаев делает «Трансперенси Интернешнл Россия». Следующий способ опубликовать данные – выложить необходимые файлы на Google Drive. Его применяет «ФБК». Кроме этого, редакции могут публиковать данные на GitHub, веб-сервисе для хостинга IT-проектов. Таким способом пользовалась «Новая газета» (статья «Продам свой голос за



свитшот»<sup>130</sup>), OCCRP (проект «Тройка Ландромат»)<sup>131</sup> и Иван Бегтин в своем блоге (статья «Государство как пират»)<sup>132</sup>.

Визуализации принято считать одной из основ журналистики данных, однако, как нам удалось обнаружить, не каждое дата-расследование сопровождается графиками и диаграммами. «ФБК», «РБК», “The Bell”<sup>133</sup>, «Открытые медиа»<sup>134</sup> в некоторых случаях вообще не использовали визуализации. Таких материалов мы насчитали восемь (15%). Максимальное количество визуализаций в одном материале – 32, но это скорее редкость. В среднем на одно расследование приходилось 4,5 визуализации. Чаще всего использовались следующие визуализации: столбиковая диаграмма (в 24 случаях), таблица (в 18 случаях), фоновая картограмма (хороплет) (в 8 случаях). В 11 случаях использовалась инфографика – визуализация, которую не удалось типизировать.

Мы изначально предполагали, что иллюстраций и фотографий будет не так много по сравнению с визуализациями, однако изображения также активно используются в дата-журналистике. В среднем – по 6,4 иллюстрации на одну публикацию, что почти в полтора раза больше по сравнению с визуализациями.

Нам также было любопытно узнать, как часто российские медиа стремятся взаимодействовать с аудиторией в своих дата-расследованиях. Результаты получились следующие: только в 18 материалах (33%) присутствовали интерактивные элементы. В основном это были

---

<sup>130</sup> Заякин А., Смагин А. Продам свой голос за свитшот // Новая газета. 2018. [Электронный ресурс]. URL:<https://novayagazeta.ru/articles/2018/09/21/77914-prodam-svoy-golos-za-svitshot> (дата обращения: 19.04.2020)

<sup>131</sup> Ланромат Тройка // OCCRP. 2019. [Электронный ресурс]. URL: <https://www.occrp.org/ru/troikalaundromat/> (дата обращения: 19.04.2020)

<sup>132</sup> Бегтин И. Государство как пират // Ivan Begtin Blog. 2019. [Электронный ресурс]. URL: <https://begtin.tech/government-piracy/> (дата обращения: 19.04.2020)

<sup>133</sup> The Bell. [Электронный ресурс]. URL:<https://thebell.io/> (дата обращения: 19.04.2020)

<sup>134</sup> Открытые медиа. [Электронный ресурс]. URL:<https://openmedia.io/> (дата обращения: 19.04.2020)

кликабельные визуализации, которые позволяли получать дополнительную информацию после нажатия мышью. Однако были и очень любопытные примеры – материал «Делового Петербурга» «"ДП" создал сервис для подсчета, сколько денег "потеряла" управляющая компания на вашем доме в Петербурге».<sup>135</sup> Читатель может сам ввести адрес своего дома и узнать сумму денег, которую УК собрала с жильцов этого дома в 2016 году и не отразила в своих отчетностях.

Нам было также любопытно посмотреть на наличие или отсутствие комментариев экспертов в расследованиях на основе данных. Наша первоначальная гипотеза была такова: так как расследования строятся на данных, комментарии экспертов не нужны, данные сами все «расскажут». Предположение не подтвердилось: в большинстве материалов комментарии экспертов были. В 30 публикациях (55%) журналисты обращались за комментарием, в 24 (45%) – нет. Однако, возможно, если бы мы рассматривали обычные журналистские расследования, то соотношение было бы больше.

Теперь рассмотрим, какие техники работы с данными использовали журналисты при работе над дата-расследованиями. В 41 случае (76%) из 54 авторы материалов не указывали, какие технологии они использовали, поэтому наш анализ будет строиться на тех 13 публикациях, чьи способы работы нам известны. Однако стоит отметить, что даже в этих 13 случаях не всегда было понятно, что именно журналисты имеют в виду.

В шести расследованиях журналисты использовали программный код для анализа или сбора данных (точно не известно). В трех материалах – специальные сервисы: «Вотум» (в данный момент зайти на сайт сервиса не

---

<sup>135</sup> Панкратова И. "ДП" создал сервис для подсчета, сколько денег "потеряла" управляющая компания на вашем доме в Петербурге // Деловой Петербург. 2017. [Электронный ресурс]. URL:[https://www.dp.ru/a/2017/06/14/DP\\_sdelal\\_servis\\_pozvo](https://www.dp.ru/a/2017/06/14/DP_sdelal_servis_pozvo) (дата обращения: 19.04.2020)

получается), «Антиплагиат»<sup>136</sup>, систему обработки и поиска Aleph, разработанную OCCRP.<sup>137</sup> В одном случае – API (интерфейс для взаимодействия двух компьютерных программ) для парсинга (автоматизированного сбора) данных. Таким способом журналисты собрали данные с портала «Реформа ЖКХ».<sup>138</sup> При работе над еще одним расследованием – сами писали парсер.

Получить данные можно следующими способами: скачать уже готовые данные (или запросить их), собрать самому с сайтов с помощью парсеров, собрать данные с помощью API.

Кроме того, нам удалось найти целую группу материалов, объединенных схожей техникой работы с данными, – это выявление заимствований в тексте. Мы обнаружили четыре расследования, подходящих под данную категорию: «Они написали протесты»<sup>139</sup>, «Штабы на лямках»<sup>140</sup>, «Судьи держат нас за болванку»<sup>141</sup>, «Атака клоунов».<sup>142</sup>

В расследовании «Они написали протесты»<sup>143</sup> журналистам удалось выявить 8 шаблонов фабул дел, которые использовались в 64 делах фигурантов «Московского дела» в судах разных районов, что говорит нам о том, что наказания не могли быть написаны судьями независимо – кто-то составил их централизованно.

---

<sup>136</sup> Антиплагиат. [Электронный ресурс]. URL:<https://www.antiplagiat.ru/> (дата обращения: 19.04.2020)

<sup>137</sup> OCCRP Aleph. [Электронный ресурс]. URL:<https://aleph.occrp.org/> (дата обращения: 19.04.2020)

<sup>138</sup> Реформа ЖКХ. [Электронный ресурс]. URL:<https://www.reformagkh.ru/> (дата обращения: 19.04.2020)

<sup>139</sup> Заякин А., Щенников А. Они написали протесты // Новая газета. 2019. [Электронный ресурс]. URL:<https://novayagazeta.ru/articles/2019/08/13/81587-oni-napisali-protesty> (дата обращения: 19.04.2020)

<sup>140</sup> Шпилькин С., Заякин А. Штабы на лямках // Новая газета. 2019. [Электронный ресурс]. URL:<https://novayagazeta.ru/articles/2019/08/01/81457-shtaby-na-lyamkah> (дата обращения: 19.04.2020)

<sup>141</sup> Заякин А., Смагин А. Судьи держат нас за болванку // Новая газета. 2019. [Электронный ресурс]. URL:<https://novayagazeta.ru/articles/2019/03/20/79929-sudi-derzhat-nas-za-bolvanku> (дата обращения: 19.04.2020)

<sup>142</sup> Карягин М., Романов В., Pitkänen M. Атака клоунов // Инфометр. [Электронный ресурс].

URL:<https://read.infometer.org/clone> (дата обращения: 19.04.2020)

<sup>143</sup> Там же.

Материал «Штабы на лямках»<sup>144</sup> выявляет общие паттерны в финансовых отчетностях кандидатов, баллотирующихся от разных партий в Мосгордуму, что говорит о централизованном изготовлении этих документов.

Авторы материала «Судьи держат нас за болванку»<sup>145</sup> нашли 50 тысяч судебных решений, совпадающих почти дословно, что говорит о неиндивидуальности подхода к каждому решению.

При работе над публикацией «Атака клоунов»<sup>146</sup> журналисты обнаружили почти полное (в некоторых случаях частичное) сходство политических программ партий, что говорит нам о том, что реальной политической конкуренции в России нет, партии просто крадут части программ друг у друга.

Над первыми тремя материалами работал сооснователь «Диссернета» и редактор data-отдела «Новой газеты» Андрей Заякин. Вероятно, именно благодаря его опыту работы по выявлению плагиата в научных работах чиновников в рамках «Диссернета», мы смогли увидеть такой нестандартный вид расследований.

Кроме того, стоит заметить, что в основном материалы строились на малых, а не больших данных. Только два материала содержали данные более 1 миллиона строк: «Вся эта дурь»<sup>147</sup> и «Тройка Ландромат».<sup>148</sup> Данные для первого материала было собраны самостоятельно с помощью парсеров (алгоритмов для автоматизированного сбора данных), источник для второй публикации был получен вследствие утечки банковских транзакций.

---

<sup>144</sup> Там же.

<sup>145</sup> Там же.

<sup>146</sup> Там же.

<sup>147</sup> Дорожный А., Хачатурянц А. Вся эта дурь // Проект. 2019. [Электронный ресурс].

URL:<https://www.proekt.media/research/narkotiki-v-darknete> (дата обращения: 20.04.2020)

<sup>148</sup> Ландромат Тройка // OCCRP. 2019. [Электронный ресурс].

URL:<https://www.occrp.org/ru/troikalaundromat/> (дата обращения: 20.04.2020)

Вероятно, небольшое количество таких материалов связано с тем, что большие дата-сетов сложнее найти, а также труднее обработать.

## 2.4. Журналистские расследования на основе данных в «Новой газете»

Самым ярким примером нашего исследования является «Новая газета», опубликовавшая, как нам удалось обнаружить, 17 расследований (одно из них совместно с OCCRP<sup>149</sup> – Центром по исследованию коррупции и организованной преступности) на основе данных. Девять материалов из 17 посвящены правосудию, пять – политике, еще по одному – на образование, бизнес и экономику (см. таблицу 5 ниже).

Тематика	Количество материалов в «Новой газете»
Правосудие	9
Политика	5
Экономика	1
Бизнес	1
Образование	1
<b>Итого</b>	<b>17</b>

Таблица 5. Тематика расследований на основе данных «Новой газеты»

Перечислим самые просматриваемые материалы в порядке убывания: «Я тебя и убью»<sup>150</sup>, «Судьи держат нас за болванку»<sup>151</sup>, «ВИП-кочевники»<sup>152</sup>, «Штабы на лямках»<sup>153</sup>, «Спец и специи»<sup>154</sup>, «Я тебя сейчас, сука, убивать буду»<sup>155</sup>, «Пиши, Емеля»<sup>156</sup>. Каждая из публикаций набрала от 73 тысяч до

<sup>149</sup> OCCRP [Электронный ресурс]. URL:<https://www.occrp.org/ru/> (дата обращения: 20.04.2020)

<sup>150</sup> Долинина И., Мароховская А. Я тебя и убью // Новая газета. 2019. [Электронный ресурс]. URL: <https://novayagazeta.ru/articles/2019/09/23/82057-ya-tebya-i-ubyu> (дата обращения: 20.04.2020)

<sup>151</sup> Заякин А., Смагин А. Судьи держат нас за болванку // Новая газета. 2019. [Электронный ресурс]. URL: <https://novayagazeta.ru/articles/2019/03/20/79929-sudi-derzhat-nas-za-bolvanku> (дата обращения: 20.04.2020)

<sup>152</sup> Долинина И., Мароховская А. ВИП-кочевники // Новая газета. 2018. [Электронный ресурс]. URL: <https://novayagazeta.ru/articles/2018/09/17/77842-svoih-ne-brosaem> (дата обращения: 20.04.2020)

<sup>153</sup> Шпилькин С., Заякин А. Штабы на лямках // Новая газета. 2019. [Электронный ресурс]. URL: <https://novayagazeta.ru/articles/2019/08/01/81457-shtaby-na-lyamkah> (дата обращения: 20.04.2020)

<sup>154</sup> Долинина И., Мароховская А. Спец и специи // Новая газета. 2019. [Электронный ресурс]. URL: <https://novayagazeta.ru/articles/2019/02/04/79417-spetsy-i-spetsii> (дата обращения: 20.04.2020)

<sup>155</sup> Я тебя сейчас, сука, убивать буду // Новая газета. 2019. [Электронный ресурс]. URL: <https://novayagazeta.ru/articles/2019/11/25/82847-ya-tebya-seychas-suka-ubivat-budu> (дата обращения: 20.04.2020)

100 тысяч просмотров (на 25.04.2020). Остальные материалы можно найти в таблице 1 в приложении.

Самая интересная черта дата-расследований «Новой газеты» в том, что их материалы расследуют скорее сложившиеся негативные тенденции в обществе, а не конкретные преступления, как это делают, например, в «Трансперенси Интернешнл Россия». «Антагонист» почти половины расследований (8 из 17) «Новой газеты» – это целая система, а не какой-то конкретный персонаж (к примеру, чиновник). В данном случае дата-расследования – это исследования важных для общества проблем. Журналисты «Новой» с помощью данных находят скрытые тенденции, что делает такие дата-расследования уникальными. На наш взгляд, без помощи данных невозможно было бы создать такие материалы. Данные совершили революцию.

Перечислим такие материалы: «Я тебя и убью»<sup>157</sup>, «Я тебя сейчас, сука, убивать буду»<sup>158</sup>, «Если подойдешь, я воткну тебе вилы в горло»<sup>159</sup>, «Мелкие преступники»<sup>160</sup>, «Судьи держат нас за болванку»<sup>161</sup>, «Продам свой голос за свитшот»<sup>162</sup>, «Стыдно сказать»<sup>163</sup>, «Кто поднимает Россию на дыбу»<sup>164</sup>.

Например, таким образом, журналисты смогли обнаружить, что:

---

<sup>156</sup> Заякин А., Смагин А. Пиши, Емеля // Новая газета. 2019. [Электронный ресурс]. URL:<https://novayagazeta.ru/articles/2019/04/20/80297-pishi-emelya> (дата обращения: 20.04.2020)

<sup>157</sup> Там же.

<sup>158</sup> Там же.

<sup>159</sup> Мароховская А., Щенников А., Мартынова Е. «Если подойдешь, я воткну тебе вилы в горло» // Новая газета. 2020. [Электронный ресурс]. URL: <https://novayagazeta.ru/articles/2020/02/13/83892-ya-tebya-zarublyu-musor?ref=tjournal.ru> (дата обращения: 20.04.2020)

<sup>160</sup> Заякин А., Смагин А. Мелкие преступники // Новая газета. 2019. [Электронный ресурс]. URL:<https://novayagazeta.ru/articles/2019/07/03/81119-melkie-prestupniki> (дата обращения: 20.04.2020)

<sup>161</sup> Там же.

<sup>162</sup> Заякин А., Смагин А. Продам свой голос за свитшот // Новая газета. 2018. [Электронный ресурс]. URL:<https://novayagazeta.ru/articles/2018/09/21/77914-prodam-svoy-golos-za-svitshot> (дата обращения: 20.04.2020)

<sup>163</sup> Смагин А., Заякин А. Стыдно сказать // Новая газета. 2018. [Электронный ресурс]. URL:<https://novayagazeta.ru/articles/2018/06/29/76970-stydno-skazat> (дата обращения: 20.04.2020)

<sup>164</sup> Мароховская А., Долинина И. Кто поднимает Россию на дыбу // Новая газета. 2018. [Электронный ресурс]. URL:<https://novayagazeta.ru/articles/2018/10/08/78095-kto-podnimaet-rossiyu-na-dybu> (дата обращения: 20.04.2020)

1. 80% насилия над детьми совершается в семье; для 40% людей, избивших или убивших ребенка, суд признавал смягчающим обстоятельством беременность или наличие других детей у подсудимых; чтобы избежать ответственности за преступление, подсудимым, не причинившим серьезный вред здоровью, было достаточно помириться с ребенком.
2. Российские суды за применение насилия к представителям власти (брошенные стаканчики) выносят участникам митингов намного более строгие приговоры, чем лицам, которые действительно нападали на полицейских.
3. 91% женщин, осужденных за превышение самообороны, защищались от своих партнеров или других родственников-мужчин; а суды и правоохранительные органы склонны вставать на сторону агрессора.

Увидеть такие тенденции журналисты могли, только проанализировав несколько тысяч или сотен (для «Московского дела») судебных решений, опубликованных в открытом доступе.

Ирина Долинина и Алеся Мароховская, авторы расследования «Я тебя и убью»<sup>165</sup>, смогли обнаружить настоящие противоречия в законодательстве, позволяющие убийцам и насильникам детей получать более мягкие сроки, если они имеют других детей, или же избегать наказания в принципе, если родитель (опекун) примирился с жертвой (ребенком), находящейся в абсолютной зависимости от него.

Еще одна отличительная черта дата-расследований «Новой газеты» – сильное эмоциональное воздействие на читателя. Несмотря на то, что расследования «Новой» строятся на наборах данных, эмоциональная

---

<sup>165</sup> Там же.

составляющая материалов невероятно сильна. По нашему мнению, «Новая газета» является в данном случае прекрасным примером того, как создавать истории на основе данных, не превращая их в сухую статистику.

Журналисты газеты добиваются такого эффекта, добавляя личные истории людей в расследования. К примеру, в материале «Я тебя сейчас, сука, убивать буду»<sup>166</sup> приведены типичные выдержки из приговоров осужденным женщинам за убийство своих партнеров и мужей. В конце каждой выдержки указан срок, на который осудили женщину. Журналисты показывают, какие реальные истории стоят за страшными цифрами.

Для написания одного расследования «Новая газета» обращалась к 1-4 источникам. В среднем в одной публикации журналисты использовали 1,8 источника. Государственные и общественные источники данных доминировали среди остальных: в 10 случаях источниками выступало государство, в 8 – общественные организации (в одной публикации могли использоваться источники данных разных типов).

Так как о правосудии журналисты «Новой» писали чаще всего, доминируют источники, раскрывающие или агрегирующие данные о судебных решениях и материалах уголовных дел, такие как ГАС РФ «Правосудие»<sup>167</sup>, общественный проект «Росправосудие» (заблокирован Роскомнадзором, не действует), общественный проект «Судебные и нормативные акты РФ».<sup>168</sup> Кроме того, журналисты обращались к следующим источникам данных: общественный проект «Декларатор»<sup>169</sup>, сайт

---

<sup>166</sup> Там же.

<sup>167</sup> Государственная автоматизированная система Российской Федерации «Правосудие» интернет-портал. [Электронный ресурс]. URL:<https://sudrf.ru/> (дата обращения: 21.04.2020)

<sup>168</sup> Судебные и нормативные акты РФ. [Электронный ресурс]. URL:<https://sudact.ru/> (дата обращения: 21.04.2020)

<sup>169</sup> Декларатор. [Электронный ресурс]. URL:<https://declarator.org> (дата обращения: 21.04.2020)



МГИК (Мосгорисбирком)<sup>170</sup>, портал госзакупок<sup>171</sup>, сервисы отслеживания полетов и некоторым другим.

«Новая газета» использует визуализации в каждом своем дата-расследовании. На материал приходится от 1 до 32 визуализаций, в среднем – 5,8 визуализаций на одну публикацию. Журналисты «Новой» используют самые разные визуализации: пузырьковую карту, столбиковые диаграммы, тримэп, линейные графики, карту взаимосвязей, таблицы, графики рассеяния, столбиковые с накоплением, круговые диаграммы, вафельную диаграмму, фоновую картограмму (хороплет), сетевые графы, укладку круга, древовидные диаграммы, точечную карту. Чаще всего «Новая газета» использовала столбиковые диаграммы и графики рассеяния, которые, согласно книге «Графики, которые убеждают всех»<sup>172</sup>, используются для сравнения и отслеживания корреляции соответственно.

Другие иллюстрации и фотографии «Новая» также активно использует, однако в среднем на одну публикацию приходится по 4,1 картинке, что меньше по сравнению с визуализациями. Особенно запоминаются эмоциональные иллюстрации художницы, которые появились в публикациях.

В 13 из 17 материалах журналисты «Новой газеты» добавляли комментарии экспертов в свои расследования. Несмотря на то, что материалы строились на данных, в большинстве случаев понадобились комментарии экспертов для дополнительного разъяснения. Журналисты брали комментарии у юристов правозащитных центров, руководителя проекта «Декларатор», председателя совета по этике Ассоциации научных редакторов и издателей и других.

---

<sup>170</sup> Вестник Московской городской избирательной комиссии [Электронный ресурс]. URL:<http://mosgorizbirkom.ru/> (дата обращения: 21.04.2020)

<sup>171</sup> Единая информационная система в сфере закупок. [Электронный ресурс]. URL:<https://zakupki.gov.ru/epz/main/public/home.html> (дата обращения: 21.04.2020)

<sup>172</sup> Богачев А. Графики, которые убеждают всех. 2020. [Электронный ресурс]. (дата обращения: 21.04.2020)

В некоторых случаях «Новая газета» указывала технику работы с данными, но чаще всего – нет. Если посмотреть на те случаи, где журналисты указывали, то в основном «Новая газета» использовала алгоритмы машинного обучения и программные коды. Перечислим используемые технологии: алгоритм машинного обучения (Random Forest Classifier), алгоритм для анализа формулировок для определения пола обвиняемого, регулярные выражения (вид поиска нужной информации в текстах с применением шаблонов).

На наш взгляд, очень важно подробно объяснять читателю, какими технологиями воспользовался журналист, чтобы повысить доверие к материалу. Приведем в анти-пример спецпроект газеты «Госзаказ и тайные холдинги»<sup>173</sup>, где журналисты использовали собственную программу, позволяющую «одним нажатием кнопки находить скрытые связи бизнес-элит». Нам как читателям такое описание технологии кажется неубедительным. Маловероятно, что аудитория поверит в чудесную программу, которая в один клик раскроет связи бизнес-элит. Как нам кажется, аудитория до сих пор скептически относится к технологиям, поэтому очень важно объяснять, как можно детальнее, что представляет собой программа.

«Новая газета» стала первой и единственной на данный момент из российских СМИ, кто запустил собственный дата-отдел, говорит Алеся Мароховская, корреспондентка дата-отдела и победительница конкурса Sigma Awards, в интервью изданию «Журналист».<sup>174</sup>

Журналистка рассказывает, что идея посвятить данным полноценный отдел возникла в 2018 году. Инициаторами создания дата-отдела стали

---

<sup>173</sup> Госзаказ и тайные холдинги // Новая газета. 2018. [Электронный ресурс]. URL:<https://goszakaz.novayagazeta.ru/> (дата обращения: 21.04.2020)

<sup>174</sup> Филиппова А. Что такое дата-журналистика и чем она может помочь любой редакции. Журналист. 2019. [Электронный ресурс]. URL:<https://jrnlst.ru/data-journalism-novaya> (дата обращения: 21.04.2020)

главный редактор «Новой» Дмитрий Муратов и несколько других журналистов. Для начала «Новая газета» решила провести хакатон по дата-журналистике, в ходе которого можно бы было выбрать кадры для нового отдела. Возглавил дата-отдел Андрей Заякин, один из основателей «Диссернета».

На данный момент, исходя из наших наблюдений, следующие журналисты принимают больше всего участия в создании дата-расследований в «Новой газете»: Андрей Заякин, Алеся Мароховская, Ирина Долинина и Алексей Смагин.

Мы предполагаем, что необходимо знать бэкграунд российских дата-журналистов для того, чтобы понять, как функционируют дата-отделы СМИ. Андрей Заякин<sup>175</sup>, редактор дата-отдела, – ученый-физик, кандидат физико-математических наук, сооснователь «Диссернета». Алеся Мароховская<sup>176</sup> и Ирина Долинина<sup>177</sup> изучали медиакоммуникации в НИУ ВШЭ, а после закончили в данном вузе магистерскую программу «Журналистика данных». Алексей Смагин – профессиональный программист, также окончивший магистратуру по направлению «Журналистика данных».<sup>178</sup> Мы видим, что в дата-журналистику приходят специалисты как гуманитарного образования, так и технического. Алеся Мароховская и Ирина Долинина, по словам журналисток на Дне открытых данных в Москве 2020, несмотря на свой журналистский бэкграунд, умеют программировать на питоне (язык программирования).<sup>179</sup>

---

<sup>175</sup> Андрей Заякин // Диссернет. [Электронный ресурс]. URL:<https://www.dissnet.org/team/andrew-zayakin.htm> (дата обращения: 22.04.2020)

<sup>176</sup> Алеся Мароховская // Facebook. [Электронный ресурс]. URL:<https://www.facebook.com/AlesyaMarohovskaya> (дата обращения: 22.04.2020)

<sup>177</sup> Ирина Долинина // Facebook. [Электронный ресурс]. URL:[https://www.facebook.com/iradolina?comment\\_id=Y29tbWVudDoxNDk4Njc1OTgzNjE0MzMzXzE0OTkwODI3NTY5MDY5ODk%3D](https://www.facebook.com/iradolina?comment_id=Y29tbWVudDoxNDk4Njc1OTgzNjE0MzMzXzE0OTkwODI3NTY5MDY5ODk%3D) (дата обращения: 22.04.2020)

<sup>178</sup> Алексей Смагин // Facebook. [Электронный ресурс]. URL:<https://www.facebook.com/blackpn> (дата обращения: 22.04.2020)

<sup>179</sup> Долинина И., Мароховская А. Дискуссия Считать больно // День открытых данных. 2020.

Однако в принципе умение программировать не является обязательным навыком дата-журналиста. Например, Андрей Дорожный, один из наиболее видных дата-журналистов, по его собственному признанию<sup>180</sup>, программировать не умеет, ему как дата-журналисту хватает Google Spreadsheets, Excel и некоторых других программ. Дада Линделл программировать на питоне умеет и использует его для парсинга (автоматизированного сбора) данных с сайтов, однако и она считает, что уметь программировать не обязательно – для многих операций хватает Excel.<sup>181</sup>

Журналистика данных оказывается на стыке гуманитарных наук и технических. Дата-журналистика формирует запрос на то, чтобы журналисты отныне владели не только навыками сторителлинга и поиска информации, но и умели программировать. Спрос именно на таких «журналистов-программистов» будет только расти, как писал профессор Северо-Западного университета в штате Иллинойс Рич Гордон, на статью которого мы ссылались в первой главе.<sup>182</sup>

Алеся Мароховская, комментируя работу дата-отдела «Новой» на Дне открытых данных 2019 во время дискуссии «Данные, истории, инструменты: как работает дата-отдел «Новой газеты»» заявила, что отдел специализируется на расследованиях, так как их дата-отдел находится при отделе расследований.<sup>183</sup> По словам журналистки, дата-журналистика и журналистика расследований довольно близки, потому что «именно в расследованиях мы чаще всего имеем дело с данными». «Умение

---

<sup>180</sup> Дорожный А.. Курс Как стать дата-журналистом? // Институт Стрелка. 2020.

<sup>181</sup> Линделл Д.. Курс Как стать дата-журналистом? // Институт Стрелка. 2020.

<sup>182</sup> Gordon R. What Will Journalist- Programmers Do? // Mediashift. 2007. URL:<http://mediashift.org/2007/11/what-will-journalist-programmers-do005/> (дата обращения: 22.04.2020)

<sup>183</sup> Данные, истории, инструменты: как работает дата-отдел «Новой газеты» // Infoculture. [Электронный ресурс]. URL:<https://youtu.be/YUgFd16b5w> (дата обращения: 22.04.2020)

пользоваться данными может сделать расследование качественнее» – говорит Мароховская.<sup>184</sup>

Журналисты «Новой» также активно принимают участие в обсуждениях использования данных в журналистике. Например, 6 и 7 марта 2020 года Алеся Мароховская и Ирина Долинина стали одними из ведущих спикеров на Дне открытых данных в Москве. Журналистки вместе с другими своими коллегами дискутировали о проблемах процесса создания дата-материалов, организации работы дата-отдела в издании и представили свои лучшие кейсы за последний год («Я тебя и убью»<sup>185</sup> и «Тройка Ландромат»<sup>186</sup>).

## **2.5. Журналистские расследования на основе данных в «Трансперенси Интернешнл Россия»**

Еще одно медиа, создающее расследования на основе данных, – российская некоммерческая организация «Трансперенси Интернешнл Россия», о которой мы уже писали на стр. 33. Напомним, что один из самых важных российских проектов на основе данных, портал «Декларатор»<sup>187</sup>, создали именно в «Трансперенси Интернешнл Россия». Дата-отдела у Трансперенси нет, однако их журналисты часто обращаются к данным.

Всего нам удалось обнаружить восемь (одно из них вместе с “Meduza”) дата-расследований «Трансперенси Интернешнл Россия», семь из которых посвящены политике, еще одно – городской среде. Нам понятно, почему политические расследования доминируют: некоммерческая организация ставит своей целью противодействовать коррупции в России. По данной

---

<sup>184</sup> Что такое дата-журналистика и чем она может помочь любой редакции // Журналист. 2019. [Электронный ресурс]. URL:<https://jrnlst.ru/data-journalism-novaya> (дата обращения: 23.04.2020)

<sup>185</sup> Долинина И. Мароховская А. Я тебя и убью // Новая газета. 2019. [Электронный ресурс]. URL: <https://novayagazeta.ru/articles/2019/09/23/82057-ya-tebya-i-ubyu> (дата обращения: 23.04.2020)

<sup>186</sup> Ландромат Тройка // OCCRP. 2019. [Электронный ресурс]. URL: <https://www.occrp.org/ru/troikalaundromat/>

<sup>187</sup> Декларатор. [Электронный ресурс]. URL:<https://declarator.org> (дата обращения: 23.04.2020)

причине даже то единственное расследование о городской среде раскрывает схемы, по которым картели и аффилированные фирмы берут больше 50% всех подрядов по ремонту дорог в шести сибирских городах.

Перечислим дата-расследования, которые нам удалось найти: «Небо. Самолет. Министр»<sup>188</sup>, «Номер с мраморной ванной за 1,4 млн Р»<sup>189</sup>, «Сибирские дороги»<sup>190</sup>, «Квартирный вопрос генерала Беляева»<sup>191</sup>, «Дума Бинго»<sup>192</sup>, «Квартира для тюремщика»<sup>193</sup>, «Распорядители земли Владимирской»<sup>194</sup>, «Соседи. Как устроена власть в российских регионах»<sup>195</sup> (совместно с «Meduza»).

Особенность деятельности «Трансперенси» состоит в том, что некоммерческая организация не только расследует факты коррупции, но также и сама предпринимает действия. Например, после каждого расследования создатели материала включают блок «Наши действия», где перечисляют заявления, собственно поданные в Генпрокуратуру, Счетную палату и др.

Еще одна значимая черта расследований «Трансперенси» – четко прописанная методология. Журналисты по пунктам расписывают, что и как они делали, чтобы прийти к выводам. Кроме того, авторы материалов дают

---

<sup>188</sup> Автонагов А., Вавилов А., Романовский Р., Сергеев И. Небо. Самолет. Министр // Трансперенси Интернешнл Россия. 2020. [Электронный ресурс]. URL:<https://travel.transparency.org.ru/2/> (дата обращения: 23.04.2020)

<sup>189</sup> Автонагов А., Вавилов А., Романовский Р., Сергеев И. Номер с мраморной ванной за 1,4 млн Р // Трансперенси Интернешнл Россия. 2019. [Электронный ресурс]. URL:<https://travel.transparency.org.ru/1/> (дата обращения: 23.04.2020)

<sup>190</sup> Сибирские дороги // Трансперенси Интернешнл Россия. 2017. [Электронный ресурс]. URL:<https://transparency.org.ru/special/dorogisibiri/> (дата обращения: 23.04.2020)

<sup>191</sup> Квартирный вопрос генерала Беляева // Трансперенси Интернешнл Россия. 2017. [Электронный ресурс]. URL:<https://transparency.org.ru/special/belyaev/> (дата обращения: 23.04.2020)

<sup>192</sup> Дума Бинго // Трансперенси Интернешнл Россия. [Электронный ресурс]. URL:<https://dumabingo.ru/> (дата обращения: 23.04.2020)

<sup>193</sup> Квартира для тюремщика // Трансперенси Интернешнл Россия. 2016. [Электронный ресурс]. URL:<https://transparency.org.ru/special/fsin/> (дата обращения: 23.04.2020)

<sup>194</sup> Распорядители земли Владимирской // Трансперенси Интернешнл Россия. 2016. [Электронный ресурс]. URL: <https://transparency.org.ru/special/shokhin/> (дата обращения: 23.04.2020)

<sup>195</sup> Соседи. Как устроена власть в российских регионах // Meduza. [Электронный ресурс].

URL:<https://meduza.io/feature/2016/09/08/sosedi-kak-ustroena-vlast-v-rossiyskih-regionah>  
<https://transparency.org.ru/special/sosedi/> (дата обращения: 23.04.2020)

ссылки на любую информацию или данные, с которыми они работают. Иногда прилагают Excel-файл со своими расчетами. Все перечисленные пункты, по нашему мнению, вызывают доверие у читателей к материалу. Четыре из восьми материалов, как нам кажется, являются реплицируемыми, то есть придирчивый читатель при желании сможет сам повторить расследование.

В шести случаях из восьми хотя бы один из источников был государственным, в одном случае – общественным, в еще одном – источники указаны не были. В двух случаях использовалось несколько видов источников данных: государственный и коммерческий, государственный и общественный. В пяти материалах одним из источников являлся портал госзакупок.<sup>196</sup> Также журналисты обращались к сервисам отслеживания полетов (Flightradar<sup>197</sup>, PrivateFly<sup>198</sup>, JetVIP<sup>199</sup>), Госзатратам<sup>200</sup>, «Декларатору»<sup>201</sup>, официальным государственным сайтам, картотеке арбитражных дел<sup>202</sup>, статистике ГИБДД<sup>203</sup>, федеральному проекту «Безопасные и качественные дороги».<sup>204</sup> Количество источников для одной статьи варьировалось от 1 до 4, в среднем – по 2,2 источника на материал.

Визуализации «Трансперенси Интернешнл Россия» используют в каждой публикации. На одно расследование приходится от 1 до 6 визуализаций, в среднем – по 2,9 визуализаций на материал. Типы визуализаций самые разные: сетевые диаграммы, фоновые картограммы

---

<sup>196</sup> Единая информационная система в сфере закупок. [Электронный ресурс]. URL:<https://zakupki.gov.ru/epz/main/public/home.html> (дата обращения: 23.04.2020)

<sup>197</sup> FlightRadar24. [Электронный ресурс]. URL:<https://www.flightradar24.com/> (дата обращения: 23.04.2020)

<sup>198</sup> PrivateFly. [Электронный ресурс]. URL:<https://www.privatefly.com/> (дата обращения: 23.04.2020)

<sup>199</sup> JetVip. [Электронный ресурс]. URL:<https://jetvip.ru/> (дата обращения: 23.04.2020)

<sup>200</sup> ГосЗатраты. [Электронный ресурс]. URL:<https://clearspending.ru> (дата обращения: 23.04.2020)

<sup>201</sup> Декларатор. [Электронный ресурс]. URL:<https://declarator.org> (дата обращения: 23.04.2020)

<sup>202</sup> Электронное правосудие. [Электронный ресурс]. URL:<https://kad.arbitr.ru/> (дата обращения: 23.04.2020)

<sup>203</sup> Показатели состояния безопасности дорожного движения. [Электронный ресурс]. URL:<http://stat.gibdd.ru/> (дата обращения: 23.04.2020)

<sup>204</sup> Безопасные и качественные автомобильные дороги. [Электронный ресурс]. URL:<https://bkdrf.ru/> (дата обращения: 23.04.2020)

(хороплет), инфографика, тримэп (дерево), точечная карта, столбиковые диаграммы, таблицы, карты потоков. Чаще всего использовались сетевые и столбиковые диаграммы.

Другие иллюстрации и картинки «Трансперенси» также активно использует, за исключением материала «Дума Бинго»<sup>205</sup>. В среднем на одну публикацию приходится по 5,8 картинок (не визуализаций). Несмотря на то, что от дата-журналистики ожидают визуализации данных, иллюстрации, привычные для обычной журналистики, также используются, притом более активно (5,75 картинок против 2,9 визуализаций в среднем на один материал).

Вероятно, из соображений безопасности организация редко указывает авторов материалов. Авторство указано только в трех публикациях из восьми. Перечислим имена тех, кто нам известен: Александр Автонагов, Александр Вавилов, Роман Романовский, Игорь Сергеев, Алеся Мароховская, Ирина Долинина. Данные журналисты, кроме последних двух, работали над двумя частями расследования о министре промышленности и торговли Денисе Мантурове («Номер с мраморной ванной за 1,4 млн Р»<sup>206</sup> и «Небо. Самолет. Министр»)<sup>207</sup>.

Бэкграунд Алеси Мароховской и Ирины Долининой нам хорошо известен. Попробуем найти остальных авторов на сайте «Трансперенси», чтобы понять, кто работает над дата-материалами в организации. Александр Автонагов<sup>208</sup> – эксперт по юридическим вопросам антикоррупционного

---

<sup>205</sup> Там же.

<sup>206</sup> Автонагов А., Вавилов А., Романовский Р., Сергеев И. Номер с мраморной ванной за 1,4 млн Р // Трансперенси Интернешнл Россия. 2019. [Электронный ресурс]. URL:<https://travel.transparency.org.ru/1/> (дата обращения: 24.04.2020)

<sup>207</sup> Автонагов А., Вавилов А., Романовский Р., Сергеев И. Небо. Самолет. Министр // Трансперенси Интернешнл Россия. 2020. [Электронный ресурс]. URL:<https://travel.transparency.org.ru/2/> (дата обращения: 24.04.2020)

<sup>208</sup> Наша команда // Трансперенси Интернешнл Россия. [Электронный ресурс]. URL:<https://transparency.org.ru/nasha-komanda/aleksandr-vladimirovich-avtonagov.html> (дата обращения: 24.04.2020)



центра в Калининграде. Александр Вавилов<sup>209</sup> – аналитик регионального антикоррупционного центра в Калининграде. Роман Романовский<sup>210</sup> – аналитик регионального антикоррупционного центра в Калининграде. Игорь Сергеев<sup>211</sup> – руководитель регионального антикоррупционного центра в Калининграде. Можем прийти к выводу, что, вероятно, над дата-материалами в «Трансперенси Интернешнл Россия» могут работать не только журналисты.

## 2.6. Журналистские расследования на основе данных в «Проекте»

«Проект»<sup>212</sup> – молодое онлайн-СМИ, запущенное в 2018 году. Редакция так пишет о себе на своем сайте: «Мы независимое медиа, объединившее журналистов-расследователей и репортеров. Мы делаем то, что умеем лучше всего, — находим скрытое и важное и подробно рассказываем об этом вам».<sup>213</sup> Помимо того, что «Проект» специализируется на расследовательской журналистике, медиа также часто использует данные в своих историях. В «Проекте» неоднократно публиковались известные дата-журналисты: Андрей Дорожный, Юлия Апухтина, Софья Савина и дата-аналитик Сергей Устинов.

В данном СМИ нам удалось найти семь расследований на основе данных. Самые яркие из них, на наш взгляд: «Шоссе в никуда»<sup>214</sup> и «Вся эта дурь»<sup>215</sup>. О них мы расскажем в подпункте 2.9.

---

<sup>209</sup> Наша команда // Трансперенси Интернешнл Россия. [Электронный ресурс]. URL:<https://transparency.org.ru/nasha-komanda/aleksandr-evgenevich-vavilov.html> (дата обращения: 24.04.2020)

<sup>210</sup> Наша команда // Трансперенси Интернешнл Россия. [Электронный ресурс]. URL:<https://transparency.org.ru/nasha-komanda/roman-igorevich-romanovskiy.html> (дата обращения: 24.04.2020)

<sup>211</sup> Наша команда // Трансперенси Интернешнл Россия. [Электронный ресурс]. URL:<https://transparency.org.ru/nasha-komanda/igor-nikolaevich-sergeev.html> (дата обращения: 24.04.2020)

<sup>212</sup> Проект. [Электронный ресурс]. URL:<https://www.proekt.media/> (дата обращения: 24.04.2020)

<sup>213</sup> О «Проекте» // Проект . [Электронный ресурс]. URL:<https://www.proekt.media/about/> (дата обращения: 24.04.2020)

<sup>214</sup> Жолобова М., Сотников Д., Баданин Р. Шоссе в никуда // Проект . 2020. [Электронный ресурс]. URL:<https://rublevka.proekt.media/> (дата обращения: 24.04.2020)

Если посмотреть на тематику материалов, то «Проект» – самое разнообразное СМИ среди других медиа. Нам удалось найти два материала о правосудии, и по одному – о медицине, наркотиках, недвижимости, экономике и городской среде (см. таблицу 6 ниже).

Тематика	Количество материалов
Правосудие	2
Медицина	1
Наркотики	1
Недвижимость	1
Экономика	1
Городская среда	1
<b>Итого</b>	<b>7</b>

Таблица 6. Тематика расследований на основе данных в «Проекте»

Особенность расследований на основе данных «Проекта» в том, что материалы содержат очень много визуализаций, графиков и диаграмм, в то время как обычных иллюстраций довольно мало (исключение – материал «Шоссе в никуда»<sup>216</sup>). В одном расследовании мы находили от 3 (исключение) до 13 визуализаций, в среднем – 9,4. Картинок оказалось от 1 до 27 (исключение), в среднем – 5,4 иллюстрации на одну публикацию. В этом «Проект» сильно отличается от других медиа, рассматриваемых в данной работе.

В отличие от большинства других медиа «Проект» в качестве визуализаций активно использует (в пяти из семи публикаций) различные карты: точечную и хороплет. Также лидируют столбиковые диаграммы, таблицы, линейные графики.

Причем четыре из семи публикаций можно назвать интерактивными. Читатель может кликнуть на визуализации и получить дополнительную информацию. Таким образом, можно не перегружать график

---

<sup>215</sup> Дорожный А., Хачатурянц А. Вся эта дурь // Проект. 2019. [Электронный ресурс]. URL: <https://www.proekt.media/research/narkotiki-v-darknete/> (дата обращения: 25.04.2020)

<sup>216</sup> Там же.

необязательными данными, однако есть риск, что читатель не поймет, что визуализация кликабельная.

Еще одна важная особенность – четко прописанная методология в публикациях. Каждое расследование на основе данных журналисты начинают с блока «как мы считали», где прописывают свои действия и их порядок. Может показаться, что, возможно, читателям больше интересен результат, чем процесс, однако в этом случае также есть решение. Оно заключается в том, чтобы сделать этот блок кликабельным, то есть открывающимся при нажатии мышью, чтобы его текст отображался только для тех, кто заинтересован. Журналисты «Проекта» так делали в некоторых своих публикациях.

Несмотря на прописанную методологию, только четыре материала из семи мы смогли назвать реплицируемыми. Причина кроется в том, что данные не всегда доступны для аудитории и что «Проект» не публикует свои данные в принципе.

Источники данных в большинстве случаев были государственными (в шести из семи публикаций), в одном – коммерческими, общественными и альтернативными. Чаще всего в качестве источника встречалась статистика ГИБДД.<sup>217</sup> Также были и другие источники данных: ЕМИСС<sup>218</sup>, Росстат<sup>219</sup>, сайты президента РФ<sup>220</sup>, региональных УФСИН, статистика судебного департамента при Верховном суде<sup>221</sup>, ГАС РФ «Правосудие»<sup>222</sup>, данные запрещенной российской площадки по торговле наркотиками и другие.

---

<sup>217</sup> Показатели состояния безопасности дорожного движения. [Электронный ресурс]. URL:<http://stat.gibdd.ru/>

<sup>218</sup> ЕМИСС Государственная статистика [Электронный ресурс]. URL:<https://fedstat.ru/> (дата обращения: 25.04.2020)

<sup>219</sup> Федеральная служба государственной статистики. [Электронный ресурс]. URL:<https://www.gks.ru/> (дата обращения: 25.04.2020)

<sup>220</sup> Президент России. [Электронный ресурс]. URL:<http://kremlin.ru/> (дата обращения: 25.04.2020)

<sup>221</sup> Судебный департамент при Верховном суде Российской Федерации. [Электронный ресурс]. URL:<http://www.cdep.ru/index.php?id=79> (дата обращения: 25.04.2020)

<sup>222</sup> Государственная автоматизированная система Российской Федерации «Правосудие» интернет-портал. [Электронный ресурс]. URL:<https://sudrf.ru/> (дата обращения: 25.04.2020)

В пяти из семи случаев мы смогли обнаружить комментарии экспертов. Несмотря на обилие графиков и диаграмм, которые, предположительно, должны все разъяснить читателю, журналисты в большинстве случаев все-таки обращались к аналитикам.

## 2.7. Журналистские расследования на основе данных в «ФБК»

«Фонд борьбы с коррупцией»<sup>223</sup> – некоммерческая организация, созданная Алексеем Навальным в 2011 году. Пожертвования – единственный источник финансирования фонда, утверждается на сайте ФБК. Всего в штате фонда работает 30 человек, занимающихся антикоррупционными расследованиями.<sup>224</sup>

Документальные фильмы «ФБК» о генеральном прокуроре РФ (до 22 января 2020 года) Юрии Чайке<sup>225</sup>, председателе правительства РФ (до 16 января 2020 года) Дмитрие Медведеве<sup>226</sup> и многие другие набирают миллионы просмотров на YouTube. Мы решили не включать самые громкие расследования «ФБК» в наш анализ, вместо этого мы взяли одни из рядовых материалов с сайта фонда (с 2018 года расследования «ФБК» публикуются на личном сайте Алексея Навального).

Материалов для анализа у нас получилось шесть, больше мы решили не брать по причине того, что они довольно схожи по источникам и техникам работы с данными между собой. Расследования «ФБК» в классическом представлении не являются дата-материалами. Расследователи, несомненно, пользовались открытыми источниками информации и данных, однако во

---

<sup>223</sup> Фонд борьбы с коррупцией. [Электронный ресурс]. URL:<https://fbk.info/> (дата обращения: 25.04.2020)

<sup>224</sup> О фонде // Фонд борьбы с коррупцией. [Электронный ресурс]. URL: <https://fbk.info/about/> (дата обращения: 25.04.2020)

<sup>225</sup> «Чайка». Фильм Фонда борьбы с коррупцией // Алексей Навальный. [Электронный ресурс]. URL:<https://youtu.be/eXYQbgvzxdM> (дата обращения: 25.04.2020)

<sup>226</sup> Он вам не Димон // Алексей Навальный. [Электронный ресурс]. URL:[https://youtu.be/qrwlk7\\_GF9g](https://youtu.be/qrwlk7_GF9g) (дата обращения: 25.04.2020)

многих случаях они не отвечали критерию машиночитаемости. Чаще всего сотрудники фонда обращались к выпискам из различных реестров, интервью чиновников, а также социальным сетям их самих и их родственников. Обойти «Фонд борьбы с коррупцией» для нас также не представлялось возможным, так как фонд является одной из крупнейших организаций, занимающихся расследовательской деятельностью в России. По данной причине мы решили взять хотя бы несколько примеров из «ФБК» для нашего анализа.

Все шесть материалов, как и ожидалось, посвящены политике. «ФБК» расследовал, откуда деньги у бывшей жены Дмитрия Пескова на покупку дорогой недвижимости во Франции<sup>227</sup>; каким образом Владимир Соловьев приобрел квартиры, дачу и итальянскую виллу<sup>228</sup>; какой незадекларированной элитной недвижимостью владеют родственники мэра Нижнего Новгорода в Майами<sup>229</sup> и другие факты нарушения законодательства.

Особенность материалов «ФБК» состоит в том, что они почти всегда (в пяти из шести случаев) публикуют данные, с которыми работали, на своем Google Drive в виде таблиц и PDF-документов. Читатели могут скачать файлы и убедиться в правдивости материала. Кроме того, в некоторых случаях авторы расследований дают ссылки на реестры, приглашая читателя проверить информацию самому. Например, в материале «Квартира в Майами для мэра Нижнего Новгорода», таким образом можно узнать в реестре города

---

<sup>227</sup> Доказываем, что Песков — коррупционер // Фонд борьбы с коррупцией. 2017. [Электронный ресурс]. URL:<https://fbk.info/investigations/post/392/> (дата обращения: 25.04.2020)

<sup>228</sup> Квартиры, дача и итальянская вилла Владимира Соловьёва // Фонд борьбы с коррупцией. 2017. [Электронный ресурс]. URL:<https://fbk.info/investigations/post/364/> (дата обращения: 25.04.2020)

<sup>229</sup> Квартира в Майами для мэра Нижнего Новгорода // Фонд борьбы с коррупцией. 2016. [Электронный ресурс]. URL:<https://fbk.info/investigations/post/285/#> (дата обращения: 25.04.2020)

Майами, действительно родственники главы города владеют там недвижимостью или нет.<sup>230</sup>

Чаще всего (в 5 случаях из 6) «Фонд борьбы с коррупцией» использовал государственные источники информации в своих публикациях. К общественным и коммерческим источникам данных расследователи обращались одинаково часто: в трех материалах из шести. Главный источник расследований «ФБК» – различные реестры недвижимости, российские и зарубежные. Кроме того, фонд использовал портал госзакупок<sup>231</sup>, «Декларатор»<sup>232</sup>, ЦИАН<sup>233</sup>, сервисы по отслеживанию полетов (PlaneFinder<sup>234</sup>, PrivateFly<sup>235</sup>), телефонный справочник Парижа<sup>236</sup>, материалы из интервью и социальных сетей.

При работе над одним расследованием «ФБК» использовали от одного до шести источников, в среднем на одну публикацию – 3,4. Как мы видим, одного источника данных оказалось недостаточно.

Почти во всех случаях данные оказывались доступны для широкой публики, за исключением одного материала. Благодаря доступности данных и их публикации на Google Drive, пять из шести расследований оказались, по нашему мнению, реплицируемыми.

Однако, по нашим предположениям, этого не вполне достаточно для того, чтобы читатель доверял расследованиям. Манера подачи материала у «ФБК» достаточно тенденциозная: используются стилистически не нейтральные слова, оскорбления героев, заметно манипулирование эмоциями. Мы предполагаем, что подобная риторика может вызывать

---

<sup>230</sup> Там же.

<sup>231</sup> Единая информационная система в сфере закупок. [Электронный ресурс]. URL:<https://zakupki.gov.ru/epz/main/public/home.html> (дата обращения: 26.04.2020)

<sup>232</sup> Декларатор. [Электронный ресурс]. URL:<https://declarator.org> (дата обращения: 26.04.2020)

<sup>233</sup> ЦИАН. [Электронный ресурс]. URL:<https://www.cian.ru/> (дата обращения: 26.04.2020)

<sup>234</sup> PlaneFinder. [Электронный ресурс]. URL:<https://planefinder.net/> (дата обращения: 26.04.2020)

<sup>235</sup> PrivateFly. [Электронный ресурс]. URL:<https://www.privatefly.com/> (дата обращения: 26.04.2020)

<sup>236</sup> Tel.Fr. [Электронный ресурс]. URL:<https://www.tel.fr/> (дата обращения: 26.04.2020)

сомнения в правдивости материалов у некоторой части аудитории. К тому же, ни в одной из публикаций не использовались комментарии экспертов.

Визуализации авторы материалов не использовали совсем, однако обычные иллюстрации использовали очень активно. В среднем на одну публикацию приходилось 15,6 картинок, среди которых можно было найти выписки из реестров, скриншоты из социальных сетей, фотографии недвижимости.

Интерактивных материалов также не было. Вероятно, особые техники по работе данных расследователи не использовали, поэтому нам и не удалось найти упоминания о них.

## **2.8. Журналистские расследования на основе данных в «РБК»**

«РБК»<sup>237</sup> – российский мультимедийный холдинг деловых СМИ. Компания ведет свою деятельность в Интернете, телевидении и прессе.<sup>238</sup>

У «РБК» своего дата-отдела, как у «Новой газеты», нет. Однако журналистика данных активно развивается в редакции. Среди самых видных дата-журналистов «РБК» – Дада Линделл, победитель Data Journalism Awards, активный участник дней открытых данных, преподаватель курса от Strelka Institute «Как стать дата-журналистом?».<sup>239</sup> По словам Линделл, она, возможно, является первым человеком в России, у кого в трудовой книжке записано: «Дата-журналист».<sup>240</sup>

---

<sup>237</sup> РБК. [Электронный ресурс]. URL:<https://www.rbc.ru/> (дата обращения: 26.04.2020)

<sup>238</sup> О компании // РБК. [Электронный ресурс]. URL:<http://www.rbcholding.ru/about.shtml> (дата обращения: 26.04.2020)

<sup>239</sup> Как стать дата-журналистом? [Электронный ресурс]. URL:<https://data-journalism.strelka.com/>

<sup>240</sup> Дата-журналистика в деловом СМИ // Infoculture. 2019. [Электронный ресурс]. URL:<https://youtu.be/HM0Q7RRDXys> (дата обращения: 26.04.2020)

В основном «РБК» пишет исследования и новостные заметки на основе данных, однако расследования и материалы с признаками расследования нам также удалось найти. Их оказалось пять.

Двое из них посвящены экономике, еще двое – городской среде, одно расследование приходится на бизнес (см. таблицу 7 ниже).

Тематика	Количество публикаций
Экономика	2
Городская среда	2
Бизнес	1
<b>Итого</b>	<b>5</b>

Таблица 7. Тематика расследований на основе данных в «РБК»

Перечислим самые популярные расследования «РБК» на основе данных по просмотрам на сайте: «Водка и бриллианты: какие компании больше других задолжали налоговикам»<sup>241</sup>, «В России нашлось 26 тыс. не платящих налоги и страховые взносы компаний»<sup>242</sup>, «Кто зарабатывает на ремонте дворов в Москве». Публикации набрали от 81 от 91 тысячи просмотров (на 25.04.2020).<sup>243</sup>

Во всех расследованиях использовались государственные источники данных, в одном из случаев – совместно с коммерческим. В этом «РБК» существенно выделяется среди остальных медиа.

«РБК» обращался к следующим источникам: данные ФНС<sup>244</sup>, портал госзакупок<sup>245</sup>, база решений ФАС<sup>246</sup>, Росреестр<sup>247</sup>, СПАРК<sup>248</sup>, портал

<sup>241</sup> Линделл Д., Агеева О. Водка и бриллианты: какие компании больше других задолжали налоговикам // РБК. 2018. [Электронный ресурс].

URL:<https://www.rbc.ru/economics/04/12/2018/5c0540019a7947d135d961ae>(дата обращения: 26.04.2020)

<sup>242</sup> Линделл Д., Агеева О., Фейнберг А. России нашлось 26 тыс. не платящих налоги и страховые взносы компаний // РБК. 2018. [Электронный ресурс].

URL:<https://www.rbc.ru/economics/05/10/2018/5bb617bd9a7947876c5bd190> (дата обращения: 26.04.2020)

<sup>243</sup> Линделл Д. Кто зарабатывает на ремонте дворов в Москве // РБК. 2019. [Электронный ресурс].

URL:<https://www.rbc.ru/society/25/09/2019/5d7f5ad09a7947436751c99b> (дата обращения: 26.04.2020)

<sup>244</sup> Федеральная налоговая служба. [Электронный ресурс]. URL:<https://www.nalog.ru/rn77/>

<sup>245</sup> Единая информационная система в сфере закупок. [Электронный ресурс].

URL:<https://zakupki.gov.ru/epz/main/public/home.html> (дата обращения: 26.04.2020)

<sup>246</sup> База решений и правовых актов // Федеральная Антимонопольная служба. [Электронный ресурс].

URL:<https://br.fas.gov.ru/> (дата обращения: 26.04.2020)

<sup>247</sup> Росреестр. [Электронный ресурс]. URL:<https://rosreestr.ru/site/> (дата обращения: 26.04.2020)



открытых данных Москвы<sup>249</sup>. Обычно для написания расследования хватало одного источника данных. Только при написании одного материала – «Кто зарабатывает на ремонте дворов в Москве»<sup>250</sup> – журналисты обращались сразу к четырем ресурсам. Этот же материал, на наш взгляд, можно назвать самым серьезным у «РБК» среди остальных пяти.

Методологию работы авторы материалов прописывают не в каждом расследовании, а только в некоторых из них. Сами данные и расчеты с ними не публикуют. Зато ссылки на данные, которыми пользовались журналисты, есть в каждой из публикаций.

По причине того, что «РБК» дает ссылки на необходимые дата-сети и что почти все из них доступны для широкой публики, реплицируемыми мы можем назвать четыре материала из пяти. Пятый материал мы не смогли признать реплицируемым, так как один из источников данных доступен только по подписке – это система СПАРК от Интерфакса.<sup>251</sup>

Важный аспект дата-журналистики – визуализации – редакция «РБК» применяет не слишком активно. В «РБК» не встретишь обилие визуализаций, как мы это видим в материалах «Проекта». В среднем на одну публикацию приходится 2,2 диаграммы. Журналисты могли использовать от 0 до 4 визуализаций в одной статье. Иллюстрациями авторы «РБК» пользуются еще менее активно: в среднем по 1,6 картинок на один материал. Стоит сказать, что ни одна публикация не была интерактивной.

Зато к комментариям экспертов журналисты обращались в каждом из материалов. Вероятно, это может быть связано с внутренним кодексом редакции, где, возможно, комментарии экспертов являются обязательной

---

<sup>248</sup> СПАРК. [Электронный ресурс]. URL:<http://www.spark-interfax.ru/> (дата обращения: 27.04.2020)

<sup>249</sup> Портал открытых данных правительства Москвы. [Электронный ресурс]. URL:<https://data.mos.ru/> (дата обращения: 27.04.2020)

<sup>250</sup> Там же.

<sup>251</sup> Там же.

частью материала. Такое правило, например, прописано в кодексе другого главного делового СМИ России – Догме «Ведомостей» (если нет подтверждения цифрами или фактами).

## **2.9. Самые яркие примеры российских дата-расследований**

Рассмотрим теперь самые интересные, на наш взгляд, расследования и материалы с признаками расследования, которые заслуживают отдельного внимания. Некоторые из них были отмечены российскими или международными дата-премиями, однако какие-то публикации просто показались нам любопытными с читательской точки зрения.

«Дума Бинго»<sup>252</sup> – проект «Трансперенси Интернешнл Россия», раскрывающий лоббизм в Госдуме седьмого созыва. Авторы проекта проанализировали медийную активность членов Госдумы, их биографии, связи с бизнесом и другими организациями и властями, а потом на основе данных о принятых депутатами законопроектах и продолжительности их прохождения составили список самых влиятельных чиновников Думы.<sup>253</sup> Инфографика на сайте проекта в очень удобной форме рассказывает о каждом из депутатов седьмого созыва (см. на рис. 4 ниже).

---

<sup>252</sup> Лоббизм в Госдуме. [Электронный ресурс]. URL:<https://dumabingo.ru/> (дата обращения: 27.04.2020)

<sup>253</sup> О проекте // Лоббизм в Госдуме. [Электронный ресурс]. URL:<https://dumabingo.ru/about/> (дата обращения: 27.04.2020)

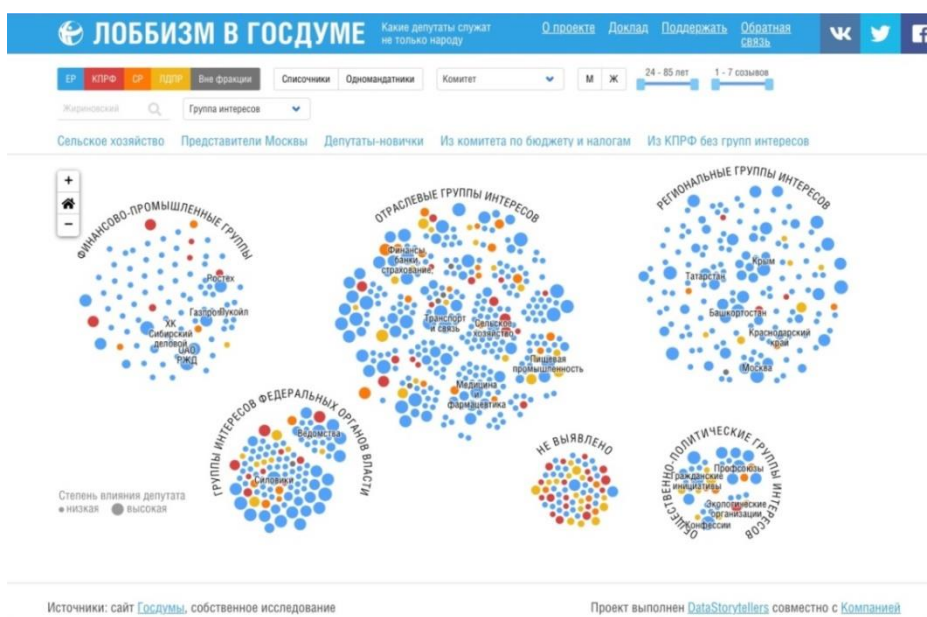


Рис. 4. Проект «ДумаБинго». Скриншот сайта проекта

Проект стал победителем в номинации «Приз зрительских симпатий» на первой Всероссийской датавиз-премии – 2020, проведенной 7 марта на Дне открытых данных в Москве.<sup>254</sup>

При работе над проектом журналисты использовали 48 тысяч страниц самых различных источников. Среди них: официальные сайты Государственной Думы, политических партий и парламентских фракций, аккаунты в социальных сетях депутатов, сайты СМИ, блоги, официальные сайты органов исполнительной власти Российской Федерации, субъектов РФ, муниципальных образований, Единый государственный реестр юридических лиц, портал госзакупок, официальные сайты коммерческих и некоммерческих организаций, отраслевых объединений и ассоциаций, финансовые отчеты кандидатов в депутаты и др.

Основываясь на законопроектах, которые продвигали депутаты, их успешности и бэкграундах самих чиновников, исследователи смогли

<sup>254</sup> Всероссийская датавиз-премия — 2020. День открытых данных. [Электронный ресурс]. URL:<https://opendataday.ru/dataviz-award> (дата обращения: 27.04.2020)

определить, какие депутаты служат не только гражданам, но и интересам определенных групп.

«Дума Бинго» почти не содержит текста, весь материал основывается на визуализации – сетевой диаграмме, которую читатели могут сами фильтровать по различным признакам. Например, можно посмотреть, кто лоббирует интересы табачной отрасли, РПЦ, «Новатэка» и т.д. При нажатии на «кружочки» появляется дополнительная информация о члене Госдумы седьмого созыва. Читатели могут изучить группу интересов депутата, количество внесенных им и принятых законов. Авторы материала даже подсчитали, за сколько дней в среднем принимаются законы от определенного депутата и по итогам в виде отдельного PDF-файла составили рейтинг самых влиятельных чиновников в Госдуме.<sup>255</sup> Кроме того, к инфографике прикладывается доклад на 60 страницах, описывающий проект.<sup>256</sup>

Еще один интересный материал создан журналистами «Проекта» – «Шоссе в никуда».<sup>257</sup> Сами авторы называют свой проект так: «Путеводитель по Рублевке, главной дороге России». Изучив 4 тысячи выписок о собственности в 76 поселках вдоль Рублево-Успенского шоссе, журналисты пришли к выводу, что огромная часть самых богатых людей России, владеющих недвижимостью в самом элитном месте страны, связана с государством, работает на него, получает господряды или дружит с

---

<sup>255</sup> Рейтинг влияния депутатов Государственной Думы // Лоббизм в Госдуме. [Электронный ресурс]. URL:<https://dumabingo.ru/%D0%A0%D0%B5%D0%B9%D1%82%D0%B8%D0%BD%D0%B3%20%D0%B2%D0%BB%D0%B8%D1%8F%D0%BD%D0%B8%D1%8F%20%D0%B4%D0%B5%D0%BF%D1%83%D1%82%D0%B0%D1%82%D0%BE%D0%B2%20%D0%93%D0%BE%D1%81%D0%B4%D1%83%D0%BC%D1%8B.pdf> (дата обращения: 27.04.2020)

<sup>256</sup> Лоббизм в Государственной Думе Федерального собрания (седьмой созыв) // Лоббизм в Госдуме. [Электронный ресурс]. URL:<https://dumabingo.ru/%D0%9B%D0%BE%D0%B1%D0%B1%D0%B8%D0%B7%D0%BC%20%D0%B2%20%D0%93%D0%BE%D1%81%D0%B4%D1%83%D0%BC%D0%B5.pdf> (дата обращения: 27.04.2020)

<sup>257</sup> Жолобова М., Сотников Д., Баданин Р. Шоссе в никуда // Проект. 2020. [Электронный ресурс]. URL:<https://rublevka.proekt.media/> (дата обращения: 27.04.2020)

чиновниками. Многие из них являются друг другу соседями, как видно по карте, составленной журналистами.

Кроме того, как выяснил «Проект», Кутузовский проспект, выводящий трафик на Рублевку, – самая опасная улица Москвы, именно там гибнут люди в каждой седьмой аварии. Причину данной ситуации журналисты видят в том, что по трассе ездит главная российская элита, а потому на проспекте отсутствуют камеры и отбойники. Эта же элита регулярно становится участниками резонансных аварий.

«Шоссе в никуда» – это целый лонгрид, рассказывающий об истории Рублевки и современном ее состоянии. Каждая точка на карте обозначает конкретного бизнесмена или чиновника, владеющего недвижимостью. Если журналисты предполагали, что данный человек аффилирован с государством, то ставили соответствующую пометку. Там, где это возможно, отмечали годовой доход и стоимость земли, в некоторых случаях добавляли фотографии.

Кликнув на «точки», можно получить дополнительную информацию о каждом владельце недвижимости. Кроме того, на карте также есть фильтры. Читатели сами могут отфильтровать для показа на карте персон из госкомпаний, окружения Путина, криминала, господрядчиков и т.д.

Другой интересный, на наш взгляд, материал «Проекта» – «Вся эта дурь» – рассказывает о том, на чем «сидит» Россия, и как российская торговая площадка в даркнете стала мировым лидером по продаже наркотиков.<sup>258</sup> Это первый материал, в ходе которого журналисты проанализировали работу крупнейшей нелегальной электронной площадки по продаже наркотиков, оружия, услуг хакеров и т.д. Расследование

---

<sup>258</sup>Дорожный А., Хачатурянц А. Вся эта дурь // Проект. 2019. [Электронный ресурс]. URL:<https://www.proekt.media/research/narkotiki-v-darknete> (дата обращения: 27.04.2020)

журналистов показало, что спрос в России на наркотики растет вместе с масштабом площадки, а правоохранительные органы уже годами не предпринимают никаких действий.

С помощью специально разработанных алгоритмов (парсеров) журналисты смогли собрать 3 миллиона отзывов на наркотические вещества, цены на каждое из моментальных наркотических предложений, а также более 393 тысяч аккаунтов людей, зарегистрированных на площадке. Таким образом, на основе этих данных журналисты смогли сделать вывод о работе площадки и о масштабе употребления наркотиков в России. Это действительно серьезная дата-журналистская работа, для которой требуются не только навыки сторителлинга, но также и программирования.

Данная публикация – одна из редких примеров использования альтернативных источников данных. Журналисты не брали и не покупали эти данные – они смогли сами собрать их с запрещенного сайта.

В данном случае возникает вопрос этичности написания такого рода материала. С одной стороны, подобное расследование очень значимо для общества, с другой – несмотря на то, что название площадки не используется, такая публикация может быть расценена как побуждающая к употреблению наркотиков или же рекламная. К тому же, в дополнение к своему анализу журналисты включали комментарии людей, работающих на эту площадку, что может наводить на мысль об их возможной связи.

Следующий расследовательский дата-журналистский проект, который нас заинтересовал, – «Место смерти», созданный самарским интернет-журналом «Другой город».<sup>259</sup> Это единственный региональный проект,

---

<sup>259</sup> Место смерти // Другой город. 2019. [Электронный ресурс]. URL:<https://drugoigorod.ru/place-of-death/> (дата обращения: 27.04.2020)

основанный на данных. Помимо этого, «Место смерти» – уникальный проект в России, детально изучивший дорожную безопасность в городе.

Цель спецпроекта – изменить восприятие ДТП в обществе, которое обычно возлагает вину на водителя или пешехода. Хотя, помимо человеческого фактора, плохое устройство дорожной сети – не менее значимая причина ДТП.

«Другой город» собрал все ДТП, произошедшие в Самаре и области за три года (2016-2018гг.) из различных источников: сайт ГИБДД<sup>260</sup>, некоммерческий проект «Карты ДТП 100 городов России»<sup>261</sup> красноярского журналиста Владимира Жаринова («главный по ДТП в России проект», как пишут журналисты «Другого города»), некоммерческий проект Сергея Устинова и Алексея Радченко «Карта ДТП».<sup>262</sup> И на основе этих данных смог определить самые опасные места в Самаре и причины, почему так получилось, раскрыв страшную статистику о том, как на конкретных самарских улицах умирают взрослые и дети. Среди причин аварий СМИ приводит в основном бездействие городских служб: неровные дороги, отсутствие ограждений, снежные завалы, темные переходы, опасные перекрёстки и т.д.

Журналисты создали карту ДТП Самары и области по годам (2016, 2017, 2018), по исходу аварии (ранен, умер) и по возрасту пострадавшего (ребенок, взрослый). Читатели также могут сами фильтровать информацию, которая появляется на сайте.

Подобные проекты имеют колоссальное значение для развития города в сторону более безопасного дорожного движения, в особенности в регионах.

---

<sup>260</sup> Показатели состояния безопасности дорожного движения. [Электронный ресурс]. URL:<http://stat.gibdd.ru/> (дата обращения: 27.04.2020)

<sup>261</sup> Карты ДТП 100 городов России. [Электронный ресурс]. URL:<https://bigdatarussian.ru/dtp-map-regions.php> (дата обращения: 27.04.2020)

<sup>262</sup> Карта ДТП. [Электронный ресурс]. URL:<https://dtp-stat.ru> (дата обращения: 27.04.2020)

Без данных создать что-либо похожее было бы невозможно. Благодаря данным, журналистам удалось найти проблемные участки Самары, выяснить причины аварийности и, вероятно, подтолкнуть местных властей к тому, чтобы изменить ситуацию.

Последний журналистский дата-проект, который мы рассмотрим, – это «Ландромат Тройка». Это целая серия расследований, публиковавшихся на разных языках во многих международных СМИ, сотрудничающих с OCCRP. В российском медиапространстве публикации вышли на самом сайте OCCRP<sup>263</sup> и «Медузе».<sup>264</sup> Над материалами работал целый журналистский коллектив, в том числе Ирина Долинина и Алеся Мароховская. Проект победил на The 2020 Sigma Awards в категории «лучший репортаж на основе данных» (среди крупных редакций).<sup>265</sup>

Расследователи раскрыли схему, позволявшую российским олигархам и политикам отмывать деньги, уклоняться от налогов и вкладывать средства, полученные незаконным путем. В центре расследования оказался частный инвестбанк России «Тройка Диалог», принадлежащий теперь ПАО «Сбербанк России». Банк создал целую сеть офшорных компаний, с помощью которых из России вывели миллиарды рублей.

Расследование строится на глобальной утечке 1,3 миллиона транзакций из 238 тысяч компаний литовского банка Ukiu. Банк входил в пятерку крупнейших в Литве, однако в 2013 году у него отозвали лицензию, а владельца обвинили в мошенничестве и растрате, а также заподозрили в отмывании денег из России. Таким образом, утекшие данные попали в распоряжение OCCRP и литовского издания “15min.lt”, а затем OCCRP

---

<sup>263</sup> Ландромат Тройка // OCCRP. 2019. [Электронный ресурс]. URL:<https://www.occrp.org/ru/troikalaundromat/>

<sup>264</sup> Шмагун О., Долинина И., Кустикова А., Мароховская А. Кошелек российской элиты // Meduza. 2019. [Электронный ресурс]. URL:<https://meduza.io/feature/2019/03/04/koshelek-rossiyskoy-elity> (дата обращения: 27.04.2020)

<sup>265</sup> The 2020 Sigma Awards // Data Journalism. [Электронный ресурс]. URL:<https://datajournalism.com/awards> (дата обращения: 27.04.2020)



поделился данными со своими центрами, которые находятся по всему миру. В итоге, над данным проектом работало более 20 изданий.

Данное расследование уникально тем, что оно строится на настоящих больших данных, что довольно большая редкость, как мы поняли в процессе нашего исследования. Раскрыть такую крупную схему отмывания средств было бы невозможно без данных и журналистов, умеющих с этими данными работать.

## **ВЫВОДЫ ПО ГЛАВЕ 2**

В данной главе мы провели собственный анализ 54 дата-публикаций из 15 различных медиа. На основе этого анализа мы смогли определить некоторые характерные черты для расследований на основе данных в целом, а также для пяти медиа, которые больше других публиковали подобные материалы. Среди таких медиа нам удалось обнаружить «Новую газету», «Трансперенси Интернешнл Россия», «Проект», «ФБК» и «РБК».

«Новую газету» можно, на наш взгляд, назвать главной российской газетой, занимающейся дата-журналистикой на данный момент. СМИ более чем в два раза опередило следующего своего конкурента по количеству дата-публикаций, в 2018 году сформировало свой дата-отдел, а еще журналисты газеты регулярно принимают участие в публичных дискуссиях.

## ЗАКЛЮЧЕНИЕ

Первые примеры использования данных в журналистике можно найти еще в XIX веке. Однако в качестве самостоятельного направления дата-журналистику выделили только в 2010 году. Журналисты осознали, что с помощью данных и программных алгоритмов можно создавать уникальные истории. Находить их было бы невозможно без современного мирового законодательства, требующего от правительств раскрытия данных, компьютерных мощностей и Интернета.

Использование данных в журналистике может поставить точку в противостоянии между профессиональными и гражданскими журналистами. Блогеры могут оказываться на местах происшествий быстрее уважаемых СМИ, однако они не обладают специальными компетенциями, необходимыми для написания дата-материала. В настоящий момент данные совершают революцию в журналистике.

Главные мировые редакции уже включают данные в свои истории, некоторые из них создают специализированные дата-отделы. Спрос на журналистов, обладающих навыками программирования и анализа данных, возрастает во всем мире.

В России дата-журналистика также активно развивается, российские журналисты выигрывают престижные международные премии в области журналистики данных. В 2016 году появилась первая магистерская программа (в НИУ ВШЭ) по подготовке дата-журналистов в России. С 2017 года стали проходить Дни открытых данных, в ходе которых можно прослушать лекции от экспертов в области дата-журналистики, а также посетить мастер-классы. В 2018 году был создан первый российский дата-отдел (в «Новой газете»), в котором начали работать выпускники магистратуры НИУ ВШЭ Ирина Долинина, Алеся Мароховская и Алексей Смагин. В реальности данные пришли в отечественные медиа еще раньше,

однако тогда в России никто еще не выделял дата-журналистику в отдельное направление.

Крайне существенно данные повлияли на расследовательскую журналистику. Наша гипотеза подтвердилась. В ходе исследования мы смогли обнаружить 54 материала, соответствующих заданным критериям, в 15-ти различных медиа. Большинство расследований было бы невозможно провести без помощи данных.

Больше всего расследований на основе данных (17 материалов) опубликовала «Новая газета», став основным российским медиа, занимающимся дата-расследовательской журналистикой. Расследования «Новой» по большей части рассказывают об опасных тенденциях в обществе, а не о конкретных преступлениях, что является одной из особенностей расследований на основе данных в целом.

Кроме того, пример «Новой газеты» показателен с точки зрения того, как СМИ удается не превращать свои материалы в сухую статистику. Журналисты добавляют эмоции и личные истории героев в свои материалы: к примеру, приводят выдержки из уголовных дел, если анализируют определенную судебную практику.

**Тематика.** Если посмотреть в целом на дата-расследования, то среди всех тематик существенно доминирует политика (21 материал), на втором месте – правосудие (12 материалов), после идут экономика и городская среда (по 5 материалов). Правосудие оказалось на втором месте, потому что, благодаря публикации фабул дел в открытом доступе, у журналистов наконец-таки появилась возможность изучить реальную судебную практику и понять, как судят в России по разным уголовным статьям.

**Источники данных.** Для проведения расследования на основе данных российским медиа требовалось в среднем два источника данных. Значит, зачастую одного источника бывает недостаточно, журналисту приходится

задумываться о том, какие еще наборы данных можно найти, чтобы дополнить историю. Государственные источники данных для журналистов-расследователей на данный момент приоритетны. К ним обращались в 41 случае. Общественные источники данных также важны для журналистской деятельности, ими пользовались в 17 материалах. Коммерческие оказались менее востребованы: их мы смогли обнаружить в 9 расследованиях. К альтернативным источникам данных российские медиа прибегали, но только в двух случаях.

**Реплицируемость.** 63% расследований нам удалось признать реплицируемыми, из чего следует сделать вывод, что в большинстве своем редакции стремятся делать свои расследования как можно более прозрачными для читателей.

**Доступность данных.** 76% расследований строятся на свободно доступных данных, для которых не требуется особый статус или платная подписка. Это означает, что читатели могут сами проверить достоверность заявляемых журналистом фактов. К тому же, в большинстве случаев журналисту не приходилось искать недоступное: все необходимые наборы данных опубликованы в Интернете в свободном доступе.

**Публикация данных.** В тех случаях, когда данные были недоступны для читателей или с ними производились вычисления, данные можно опубликовать. Это делается для того, чтобы читатели могли сами удостовериться в верности материала, а также для того, чтобы другие журналисты могли раскрыть ту же историю с другой стороны, найдя в наборе данных новые факты. Мы пришли к выводу, что авторы материалов публиковали данные только в 26% случаев, в остальных 74% – нет. Мы рассматриваем это скорее как негативную тенденцию в журналистике данных.

**Визуализации данных.** Оказалось, что не для каждого дата-расследования необходимы визуализации. Почти в 15% материалов визуализаций не было. В среднем на одно расследование пришлось 4,5 визуализации. Самыми часто встречаемыми оказались столбиковая диаграмма, таблица, фоновая картограмма (хороплет). На наш взгляд, данные визуализации довольно легко воспринимаются самыми разными аудиториями, поэтому их большинство. Сложные визуализации, такие как диаграмма Санкея и укладка круга, встречались лишь однократно.

**Интерактивность.** Оказалось, что в основном журналисты предпочитали не делать свои материалы интерактивными. Только 33% публикаций мы смогли признать интерактивными. Под интерактивными мы понимаем любые нестатичные визуализации, с которыми может взаимодействовать читатель.

**Комментарии экспертов.** В 55% публикаций журналисты обращались за комментариями к экспертам. Несмотря на то, что расследования строятся на данных, которые, казалось бы, сами могут все «рассказать», роль эксперта все еще значительна.

**Большие данные против малых.** Среди анализируемых нами 54 материалов нашлись также те, которые были основаны на больших данных (от 1 миллиона строк), однако таких расследований оказалось всего два, или менее 4% всех публикаций. В основном журналисты оперировали малыми данными (десятки, сотни, тысячи строк).

## Библиографический список

1. Бегтин И. Открытое государство. История и новый смысл // Полит.ру. 2012. [Электронный ресурс] URL:[https://polit.ru/article/2012/06/28/open\\_gov/](https://polit.ru/article/2012/06/28/open_gov/)
2. Бегтин И. Памятка по журналистике данных // Сайт Ивана Бегтина. 2011.[Электронный ресурс]. URL:<http://ivan.begtin.name/2011/05/22/datajournalism/>
3. Бегтин И. Ivan Begtin Blog. [Электронный ресурс]. URL:<https://begtin.tech>
4. Бегтин И. Ivan Begtin. [Электронный ресурс]. URL: <https://t.me/begtin>
5. Богачев А. Графики, которые убеждают всех. 2020. [Электронный ресурс].
6. Бублик М., Тоганова Н. Преступность в тени // Институт проблем правоприменения. 2019. [Электронный ресурс] URL: <https://readymag.com/u85821431/russiancrimesurvey/>
7. Бюллетень по открытым данным. 2015. №1. [Электронный ресурс] URL:<https://ac.gov.ru/archive/files/publication/a/5572.pdf>
8. Деловой Петербург. [Электронный ресурс]. URL:<https://www.dp.ru>
9. День открытых данных. [Электронный ресурс]. URL:<https://opendataday.ru>
10. Другой город. [Электронный ресурс]. URL: <https://drugoigorod.ru>
11. ГосЗатраты. [Электронный ресурс]. URL:<https://clearspending.ru>
12. Госрасходы. [Электронный ресурс]. URL:<https://spending.gov.ru>
13. Декларатор. [Электронный ресурс]. URL:<https://declarator.org>
14. День открытых данных. 2020.
15. Дорожный А. Data-вебинары с Андреем Дорожным: данные для начинающих // Теплица социальных технологий. 2020.
16. Если быть точным // Фонд нужна помощь. [Электронный ресурс]. URL: <https://tochno.st/>

17. Журналистика данных. [Электронный ресурс].  
URL:<https://t.me/digitalfan>
18. Информационная культура. [Электронный ресурс].  
URL:<https://www.infoculture.ru>
19. Каталог визуализации данных. [Электронный ресурс].  
URL:<https://datavizcatalogue.com/RU/>
20. Комитет гражданских инициатив. [Электронный ресурс]. URL:<https://komitetgi.ru/>
21. Консультант Плюс. 2006. [Электронный ресурс].  
URL:<http://www.consultant.ru>
22. Курс Как стать дата-журналистом? // Институт Стрелка. 2020.
23. URL:<https://www.rbc.ru/society/11/03/2019/5c6c0ae49a79477703d915b7>
24. Лисицин М.Е. Понятие журналистика данных в освещении российских и зарубежных СМИ. 2018. [Электронный ресурс].  
URL:[http://www2.rsuh.ru/binary/object\\_28.1551881843.02622.pdf](http://www2.rsuh.ru/binary/object_28.1551881843.02622.pdf)
25. Лисицин М. Е. Понятие "журналистика данных" в современных научных публикациях // Коммуникативные исследования. 2018. №3 (17). С. 145. [Электронный ресурс]. URL:<https://cyberleninka.ru/article/n/ponyatie-zhurnalistika-dannyh-v-sovremennyh-nauchnyh-publikatsiyah>
26. Лисицин М.Е. Проекты в области журналистики данных в российских СМИ (на примере газеты «Деловой Петербург») // Знак: проблемное поле медиаобразования. 2018. №4 (30). [Электронный ресурс].  
URL:<https://cyberleninka.ru/article/n/proekty-v-oblasti-zhurnalistiki-dannyh-v-rossiyskih-smi-na-materiale-gazety-delovoy-peterburg/viewer>
27. Лисицин М.Е. Феномен журналистики данных в зарубежных СМИ // Litera. 2019. № 1. С. 102-115 [Электронный ресурс]. URL:[https://e-notabene.ru/fil/article\\_29056.html](https://e-notabene.ru/fil/article_29056.html)
28. Лоббизм в Госдуме. [Электронный ресурс]. URL:<https://dumabingo.ru/>
29. Майер-Шенбергер В., Кукьер К. Большие данные: революция, которая изменит то, как мы живем, работаем и мыслим. М., 2014.

30. Макеенко М., Шилина А. Журналистика данных в качественной прессе: жанровые и тематические особенности // МедиаАльманах. 2019. №6. [Электронный ресурс]. URL: <http://mediaalmanah.ru/upload/iblock/2e8/22-35-6-95.pdf>
31. Медуза. [Электронный ресурс]. URL:<https://meduza.io/>
32. Методические рекомендации по публикации открытых данных Версия 3.0 // Открытые данные России. 2014. [Электронный ресурс]. URL:<https://data.gov.ru/metodicheskie-rekomendacii-po-publikacii-otkrytyh-dannyh-versiya-30>
33. Неренц Д.В.. Data journalism: новые возможности для журналистов-расследователей? // Вестник ВГУ. Серия: Филология. Журналистика. 2018. №3. С 147-150.
34. Новая газета. Электронный ресурс. URL:<https://novayagazeta.ru>
35. Открытая полиция. [Электронный ресурс]. URL:<https://openpolice.ru>
36. Открытые данные России. [Электронный ресурс] URL:<https://data.gov.ru>
37. Открытые НКО. [Электронный ресурс]. URL:<https://openngo.ru>
38. Официальный интернет-портал правовой информации [Электронный ресурс]. URL:<http://pravo.gov.ru/>
39. Пархимович О., Бегтин И. Данные неверны: почему российские госсайты вводят граждан в заблуждение // РБК. 2019. [Электронный ресурс]. URL:<https://www.rbc.ru/opinions/society/20/12/2019/5dfb4e169a79470b46b4f98c>
40. Проект. [Электронный ресурс]. URL:<https://www.proekt.media>
41. Проектный центр Инфометр. [Электронный ресурс]. URL: <https://read.infometer.org/>
42. РБК. URL:<https://www.rbc.ru>
43. Симакова С. И. Дата-журналистика как медиатренд // Вестн. Нижегородского ун-та им. Н. И. Лобачевского. 2014. № 2 (2). С. 481–485. [Электронный ресурс]. URL: <https://cyberleninka.ru/article/n/data-zhurnalistika-kak-mediatrend/viewer>



44. Симакова С.И. Журналистика данных как объективное требование времени и ее влияние на формирование визуальной журналистики // Знак: проблемное поле медиаобразования. 2016. №1 (18). С. 18-25. [Электронный ресурс] URL:<https://cyberleninka.ru/article/n/zhurnalistika-dannyh-kak-obektivnoe-trebovanie-vremeni-i-eyo-vliyanie-na-formirovanie-vizualnoy-zhurnalistiki/viewer>
45. ТАСС. [Электронный ресурс]. URL: <https://tass.ru>
46. Тертычный А.А. Расследовательская журналистика: учебное пособие для вузов. М.: Аспект Пресс. 2002.
47. Трансперенси Интернешнл Россия. [Электронный ресурс]. URL:<https://transparency.org.ru/>
48. Филиппова А. Что такое дата-журналистика и чем она может помочь любой редакции. Журналист. 2019. [Электронный ресурс]. URL:<https://jrnlst.ru/data-journalism-novaya>
49. Фонд борьбы с коррупцией. [Электронный ресурс]. URL:<https://fbk.info>
50. Шилина М.Г. Data Journalism – дата-журналистика, журналистика метаданных – в структуре медиакommunikации: к вопросу формирования теоретических исследовательских подходов. Электронный научный журнал Медиаскоп. 2013. № 1 [Электронный ресурс]. URL:<http://www.mediascope.ru/node/1263>
51. Экспертный совет при Правительстве РФ. [Электронный ресурс]. URL:<https://opendata.open.gov.ru/>
52. American Statistical Association. URL:<https://thisisstatistics.org/>
53. Bradshaw P. What Is Data Journalism? // Data Journalism Handbook 1. URL: <https://datajournalism.com/read/handbook/one/introduction/what-is-data-journalism>
54. Coddington M. Clarifying Journalism’s Quantitative Turn // Digital Journalism. Vo.3. 2015. №3. P. 331-348. URL:<https://www.tandfonline.com/doi/full/10.1080/21670811.2014.976400>

55. Data-driven journalism: What is there to learn. 2010. URL:[http://mediapusher.eu/datadrivenjournalism/pdf/ddj\\_paper\\_final.pdf](http://mediapusher.eu/datadrivenjournalism/pdf/ddj_paper_final.pdf)
56. Data Journalism Awards. URL:<https://datajournalismawards.org>
57. Data Journalism. URL:<https://datajournalism.com>
58. Gambini L. A decade of working in data journalism: what has changed? Interview with Simon Rogers, data editor at the Google News Lab // Medium. 2019. URL:<https://medium.com/we-are-the-european-journalism-centre/a-decade-of-working-in-data-journalism-what-has-changed-8d950d99935e>
59. Garrison, B. (2001). Diffusion of Online Information Technologies in Newspaper Newsrooms. *Journalism*, 2(2). P. 221–239.
60. Gordon R. What Will Journalist- Programmers Do? // Mediashift. 2007. URL:<http://mediashift.org/2007/11/what-will-journalist-programmers-do005/>
61. Hahn O., Stalph F. Digital Investigative Journalism: Data, Visual Analytics and Innovative Methodologies in International Reporting. Palgrave Macmillan, 2018.
62. Holovaty A. A Fundamental Way Newspaper Sites Need to Change. 2006. URL:<http://www.holovaty.com/writing/fundamental-change/>
63. Infoculture. URL:<https://www.youtube.com/channel/UCve6269kQEJ1LW52HXKtTng>
64. Kopf E.W. Florence Nightingale as Statistician // Publications of the American Statistical Association. Vol. 15, No. 116. 1916. P. 388-404. URL:[https://www.jstor.org/stable/2965763?seq=3#metadata\\_info\\_tab\\_contents](https://www.jstor.org/stable/2965763?seq=3#metadata_info_tab_contents)
65. Kantar Information is beautiful awards. URL:<https://informationisbeautiful.net>
66. Multiple Journalism. URL:<http://multiplejournalism.org/case/the-guardian-datablog>
67. OCCRP. [Электронный ресурс]. URL:<https://www.occrp.org>
68. Online Journalism Awards. URL:<https://awards.journalists.org>
69. Online Journalism Blog. URL:<https://onlinejournalismblog.com>
70. Open Data Barometer. URL:<https://opendatabarometer.org>

71. Open Government Partnership. URL:<https://www.opengovpartnership.org>
72. Quartz. URL:<https://qz.com>
73. The Bell. [Электронный ресурс]. URL:<https://thebell.io/>
74. The Guardian. URL:<https://www.theguardian.com/>
75. The New York Times. URL:<https://www.nytimes.com>
76. Ulken E. Building the data desk: lessons from the L.A. Times // Online Journalism Review. 2008. URL:<http://www.ojr.org/p1581/>

## Приложение

### Макет таблицы для анализа российских журналистских расследований на основе данных

№	Ме диа	Заго лово к	Авт ор ы	Д ат а	Об ье м	Тем атик а	Прос мотр ы	В и д И Д	И Д	Да нн ые	Срок взят ых данн ых	Ко л- во ИД	Досту пност ь данн ых	Кол- во визу ализац ий	Типы визуа лизац ий	Иллю страц ии	Комм ентар ии экспе ртов	Интерак тивност ь	Реплици руемост ь данных	Публика ция данных	Техника работы с данным и	Ссыл ка	
1																							
2																							
...																							

Ссылка общего доступа на заполненную таблицу:

<https://docs.google.com/spreadsheets/d/1QSWCVvS8FMYw-i6LPtiaxO7yySMHdNmKFnDiyQOJKnI/edit?usp=sharing>

QR-код:

