

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М. В. Ломоносова
ФАКУЛЬТЕТ ЖУРНАЛИСТИКИ

На правах рукописи

ШИЛИНА АЛЕКСАНДРА ГЕННАДЬЕВНА

**ОСНОВНЫЕ ХАРАКТЕРИСТИКИ МАТЕРИАЛОВ
ЖУРНАЛИСТИКИ ДАННЫХ
В ЗАРУБЕЖНОЙ И РОССИЙСКОЙ КАЧЕСТВЕННОЙ ПРЕССЕ**

Специальность 10.01.10 – Журналистика

Диссертация на соискание ученой степени
кандидата филологических наук

Научный руководитель:
кандидат филологических наук
М. И. Макеев

Москва - 2019

СОДЕРЖАНИЕ

Введение.....	3
Глава 1. Журналистика в условиях датафикации: теоретико-методологические подходы к изучению.....	14
1.1. Журналистика данных: генезис, сущность, известные характеристики.....	14
1.2. Основные направления исследований журналистики данных....	42
Глава 2. Зарубежный опыт функционирования журналистики данных: кейсы ведущих изданий.....	58
2.1. Методика исследования	58
2.2. Характеристики журналистики данных в <i>The Guardian Datablog</i> британской газеты <i>The Guardian</i>	69
2.3. Характеристики журналистики данных раздела <i>The Upshot</i> американской газеты <i>The New York Times</i>	81
2.4. Выявленные характеристики зарубежной «журналистики больших данных».....	93
Глава 3. Журналистика данных: современный российский опыт.....	104
3.1. Формирование и развитие журналистики данных в России.....	104
3.2. Характеристики российских публикаций журналистики данных в качественной прессе.....	117
Заключение.....	139
Библиографический список.....	145

Введение

Актуальность темы исследования. В начале XXI века темы, связанные с изменениями, которые принесли в политическую, экономическую и социальную реальность процессы и явления, связанные с активно развивающейся цифровизацией, стали важнейшей частью общественного, профессионального и, конечно, академического дискурса. Однако направления исследовательского интереса к разным сферам, связанным с цифровизацией, получают иногда заметно несопоставимое внимание со стороны специалистов. Например, в то время как российские исследователи журналистики, массовой коммуникации и медиа широко интегрированы в изучение эффектов цифровизации и, в частности, медиаконвергенции, как одного из ключевых её последствий, внимание к исследованиям еще одного важного результата всестороннего внедрения цифровых технологий – так называемой «датафикации» – на журналистику было пока минимальным.

При этом воздействие *датафикации* как процесса представления явлений в количественном цифровом формате, преобразования данных в цифровую информацию, доступную для аналитики¹ представляется сегодня крайне актуальной темой для изучения теми, кто пытается зафиксировать происходящие в медиа изменения. Нельзя не заметить, в частности, что параллельно с ростом интереса к датафикации растет и заинтересованность в понимании и изучении во многом связанного с ней явления – журналистики данных. На наш взгляд, многостороннее изучение этого явления, которое пока в отечественном академическом дискурсе практически отсутствует, позволяет представителям как научных, так и индустриальных кругов получить дополнительные ответы на крайне **актуальные** для них в настоящий момент вопросы. Например, понять то, насколько масштабны в реальности те изменения в деятельности и продукте работы журналистов,

¹ Mayer-Schönberger V., Cukier K. (2013) Big Data: A Revolution That Will Transform How We Live, Work, and Think. UK: John Murray.

которые широко обсуждаются теоретиками, или какие компетенции журналисту необходимо приобретать или совершенствовать наиболее активно, а также оценить уровень специфичности новых направлений журналистики в сравнении с более традиционными её формами.

В рамках данной работы в качестве операционного мы будем использовать следующее ***определение ключевого термина***:

журналистика данных – направление в журналистике, основой материалов которого являются результаты обработки количественных данных, которые могут служить инфоповодами или быть частью материала, обогащая его. При этом особое значение при презентации материала аудитории приобретает визуализация результатов анализа выбранных данных.

Степень разработанности темы исследования. Практики журналистики данных пока в достаточной степени не систематизированы и не отражены комплексно в научной литературе. Основная часть работ по исследуемой тематике носит научно-практический характер. Существующие исследования в обозначенной области находятся на стыке различных научных направлений и концепций и носят междисциплинарный характер.

Базовыми для изучения характеристик больших данных, особенностей журналистики данных, а также для изучения работы специалистов, создающих публикации на основе цифровых форматов данных, являются работы зарубежных авторов по эпистемологии журналистики данных: научные статьи С. Паризи, С. К. Льюиса и О. Уэстлунда посвящённые изучению влияния феномена больших данных на современную журналистику; исследование влияния инноваций на работу журналистов А. Гюннильда и анализ последствий появления вычислительной журналистики и журналистики данных на работу в редакциях М. Коддингтона².

² Parasić S. (2015) Data-driven revelation? Epistemological tensions in investigative journalism in the age of 'big data'. *Digital Journalism* 3(3): 364-380; Gynnild A. (2014) Journalism Innovation Leads to Innovation Journalism: The Impact of Computational Exploration on Changing Mindsets. *Journalism*, 15(6): 713–730; Coddington M. (2015)

Значительное место занимают работы посвящённые изучению специфики журналистики данных разных стран, в частности исследования специфики американской журналистики данных Л. Финк и К. Андерсона, исследование работы журналиста с цифровыми данными в редакциях Чикаго С. Параси и Е. Дагираль и Нью-Йорка В. Уэбера и Х. Ролла. Первая попытка осмысления становления британской журналистики данных описана в статьях С. Роджерса и в работах Э. Борхес-Рэя, Л. Ханнафорд и М. Найт³, изучающих её качественными методами.

Для теоретико-методологической базы нашего исследования наиболее важными представляются несколько работ, в которых сформированы подходы к построению типологии журналистики данных по разным основаниям. К ним необходимо отнести, прежде всего, статьи Э. Борхеса-Рэя⁴ и исследовательской группы, состоявшей из Д. Де Мейер, Д. Доминго и других⁵. В первой из них классификация выстраивается на основе выделения форматов журналистских материалов (разделение на краткую / ежедневную

Clarifying Journalism's Quantitative Turn: A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digital Journalism* 3(3): Journalism in an Era of Big Data: Cases, Concepts, and Critiques: 331-348; Lewis S. C., Westlund O. (2015) Big Data and Journalism: Epistemology, expertise, economics, and ethics. *Digital Journalism* 3(3): Journalism in an Era of Big Data: Cases, Concepts, and Critiques.

³ Fink K. and Anderson C. (2015) Data Journalism in the United States: Beyond the "usual suspects". *Journalism Studies*, 16(4): 467-481; Parasie S. and Dagiral E. (2012) Data-driven journalism and the public good: "Computer-assisted-reporters" and "programmer-journalists" in Chicago. *New Media & Society* 15(6): 853-871; Weber W. and Rall H. (2013) We are journalists. Production Practices, Attitudes and a Case Study of the New York Times Newsroom. *Interaktive Infografiken*, 161-172; Borges-Rey E. (2016) Unravelling data journalism: A study of data journalism practice in British newsrooms. *Journalism Practice* 10(7): 833-843; Borges-Rey, E. (2017) Towards an epistemology of data journalism in the devolved nations of the United Kingdom: Changes and continuities in materiality, performativity and reflexivity. *Journalism*. DOI: 10.1177/1464884917693864; Hannaford L. (2015) Computational Journalism in the UK newsroom: hybrids or specialists? *Journalism Education*, 4(1): 6-21; Knight M. (2015) Data journalism in the UK: a preliminary analysis of form and content. *Journal of Media Practice*, 16(1): 55-72; Rogers S. (2011) Data journalism at the Guardian: what is it and how do we do it? // Guardian.co.uk. Available at: <https://www.theguardian.com/news/Datablog/2011/jul/28/data-journalism>

⁴ Borges-Rey E. (2016) Unravelling data journalism: A study of data journalism practice in British newsrooms. *Journalism Practice* 10(7): 833-843.

⁵ De Maeyer, J., Libert M., Domingo, D., Heinderyckx F. & Le Cam F. (2015) Waiting for Data Journalism. A qualitative assessment of the anecdotal take-up of data journalism in French-speaking Belgium. *Digital Journalism* 3(3): 432-446.

форму и развернутую расследовательскую форму журналистики данных), во второй – исходя из рабочих практик подготовки материалов (выделение «обычного» и «основательного» типа журналистики данных). Также необходимо отметить труд Ф. Штальфа 2017⁶, разработавшего наиболее полную, на наш взгляд, к настоящему времени схему анализа журналистских публикаций интересующего нас направления современной журналистики.

Теоретическому осмыслению использования новых форматов цифровых данных в российской журналистике посвящены работы М. Шилиной, С. Вартанова, И. Радченко и А. Сакоян и др.⁷. При этом пока в отечественном исследовательской практике об эмпирических исследованиях журналистики данных можно было говорить только на примере работы А. Валеевой, которая рассмотрела использование открытых данных отечественными журналистами на примере кейсов нескольких изданий и на материале полуструктурированных интервью с их авторами и редакторами⁸.

В целом, ознакомление с имеющимся корпусом научной литературы о функционировании журналистики данных позволяет обратить внимание на то, что к настоящему моменту в нем не были представлены работы, подготовленные по результатам эмпирических исследований текстов и рабочих практик журналистики данных в России; сравнительного анализа отечественных и зарубежных процессов и явлений журналистики данных, проведенного на основе эмпирической базы, полученной по единой методике и методологии; а также исследований, в центре которых находились бы

⁶ Stalpf F. (2017) Classifying Data Journalism. A content analysis of daily data-driven stories. *Journalism Practice*, 12(10): 1332-1350.

⁷ Radchenko I., Sakoyan A. (2014) The view on open data and data journalism: Cases, educational resources and current trends. In: Ignatov, D.I., Khachay, M.Y., Panchenko, A., et al. (eds) *Analysis of Images, Social Networks and Texts*. Cham: Springer, pp. 47–54; Вартанов С. А. Большие данные в онлайн-СМИ: подходы и стратегии использования // Медиаскоп. 2017. Вып. 4; Шилина М. Г. Data Journalism – дата-журналистика, журналистика метаданных – в структуре медиакommunikации: к вопросу формирования теоретических исследовательских подходов. Медиаскоп. 2013. Вып. 1.

⁸ Valeeva A. (2017) Open Data in a Closed Political System: Open Data Investigative Journalism in Russia. Reuters Institute for the Study of Journalism. University of Oxford. Available at: http://reutersinstitute.politics.ox.ac.uk/sites/default/files/2017-10/Valeeva_Open%20data%20and%20investigative%20journalism%20in%20Russia.pdf

феномены, связанные с интеграцией в продукты журналистской деятельности «больших данных», являющихся одним из важнейших последствий цифровизации социальных и экономических реалий.

В этой ситуации **научная новизна** проведенного исследования обусловлена тем, что впервые в отечественной научной практике сформирована и опробована оригинальная методика и методология изучения российских текстов журналистики данных, а также тем, что впервые закономерности развития и функционирования журналистики данных выявлены на эмпирическом материале, полученном в ходе исследования журналистских произведений, в основе которых лежат инфоповоды, ставшие результатом обработки «больших данных».

Соответственно, **объектом** исследования выступили журналистские материалы, которые можно отнести к различным формам журналистики данных в зарубежной и отечественной качественной прессе.

Предметом исследования стали характеристики текстов журналистики данных, связанные с их тематикой, жанрами, используемыми типами данных и источниками информации, иллюстративным материалом и авторством публикаций и позволяющие выявить наличие или отсутствие у них специфических черт, влияющих на их позиции в современной журналистике в целом.

Под журналистикой данных в данном случае понимается направление в журналистике, основой материалов которого являются результаты обработки количественных данных, которые могут служить инфоповодами или быть частью материала, обогащая его; под датафикацией – преобразование привычных процессов и объектов действительности в цифровую форму, приемлемую для анализа и интерпретации; а под большими данными – массивы цифровых данных, которые из-за их объема и сложности, не могут быть собраны, сохранены и проанализированы без помощи специализированных горизонтально масштабируемых программных инструментов обработки

Цель исследования – выявление характеристик журналистики данных и особенностей текстов, подготовленных с использованием больших данных как важнейшего результата современных процессов датафикации социальной и экономической деятельности.

Для достижения поставленной цели было необходимо решение следующих исследовательских **задач**:

- определить возможность и обоснованность диверсификации определений журналистики данных, в первую очередь, за счет использования критериев, связанных с современной типологией имеющихся данных и непосредственно с понятием «большие данные»,

- сформировать представление о тенденциях развития и характеристиках, присущих зарубежным образцам журналистики данных, позволяющих создавать более широкий контекст для анализа и интерпретации результатов, полученных на отечественном материале, а также выявить универсальные закономерности в развитии исследуемого направления журналистики,

- определить проявляющиеся в контенте журналистики данных возможные новые компетенции сотрудников редакций новостных медиа, связанные с технологическим усложнением современной журналистской работы,

- выявить возможные особенности развития журналистики данных как потенциально передового направления на фоне развития под воздействием цифровизации современной новостной журналистики в целом.

Рабочая гипотеза исследования. Появление новых типов цифровой информации, в частности больших данных, предполагает и новые способы её поиска, обработки, представления и визуализации в журналистике (создание и эксплуатация специального программного обеспечения, увеличение интерактивных элементов в контенте и т.д.). Эти изменения могут находить отражение в текстах журналистских материалов, подготовленных в форматах журналистики данных. Однако велика вероятность, что к середине 2010-х гг.

влияние использования данных на источники информации, структуру, жанровые характеристики материалов ведущих новостных медиа в России и за рубежом оставалось минимальным.

Методология и методика исследования. Как будет подробнее раскрыто далее (см. параграф 2.1), *методологические рамки* исследования заданы обоснованной тенденцией развития датафикации необходимостью в разделении используемых в журналистике данных на разные типы, в частности, на «количественные» и «большие». В этом контексте *количественные данные* могут быть определены как наборы данных небольших объемов и разнообразия, которые могут быть обработаны человеком с использованием любого из существующих традиционных инструментов обработки.

Кроме того, как уже оговаривалось выше (раздел «Степень разработанности темы исследования») мы воспользовались имеющимися теоретическими подходами к типологии журналистики данных, выделяющими в ней «краткую ежедневную» / «крупную / развернутую исследовательскую» и «обычную» / «основательную» формы.

Непосредственными *эмпирическими объектами* стали 139 публикаций, основанных на результатах обработки больших данных и опубликованных в специализированных онлайн-ресурсах *Guardian Datablog* (Великобритания) и *The Upshot* (США), и 171 журналистский материал, основанный на больших данных, из авторитетных качественных российских изданий «Коммерсант daily», «Новая Газета», «Деньги», «Власть», «РБК», *Forbes*, *The New Times*.

Хронологические рамки исследования охватывают 2014-2016 гг. – три года в середине 2010-х гг., период, достаточный для сбора статистических данных на уровне пилотажного исследования, на этапе, когда журналистика данных уже достаточно широко распространена в медиа и основные её практики можно считать устоявшимися и представляющими интерес для изучения.

Основной использованный *метод* исследования – контент-анализ попавших в выборку публикаций. Метод кейс-стади применен при изучении онлайн-ресурсов *Guardian Datablog* и *The Upshot*, ставших наиболее крупными специализированными медиа, публиковавшими материалы, маркировавшиеся как журналистика данных, на первом этапе её широкого распространения. Сравнительный анализ использовался при сопоставлении данных, полученных при изучении публикаций журналистики данных и более традиционных направлений журналистики на отечественных и зарубежных примерах, а также при сравнении результатов обработки российского материала и зарубежных кейсов.

При проведении контент-анализа были использованы *кодификаторы* по следующим группам параметров:

- тематика публикации,
- жанр публикации,
- типа источников информации,
- иллюстративный материал как результат визуализации данных,
- авторство,
- число профессий, компетенции которых потенциально необходимы при её создании (например, навыки графического дизайнера и разработчика программного обеспечения).

За основу при этом взята дополненная схема анализа публикаций журналистики данных Ф. Штальфа, дополненная типологией жанров А. В. Колесниченко и классификацией источников информации М. А. Разумовой⁹.

⁹ Stalpf F. (2017) Classifying Data Journalism. A content analysis of daily data-driven stories. *Journalism Practice*, 12(10): 1332-1350; Колесниченко А. В. Практическая журналистика. Учебное пособие. М.: Изд-во Моск. ун-та, 2008; Колесниченко А. В. Востребованность жанров журналистских текстов в онлайн-СМИ // Вестник Московского университета. Серия 10. Журналистика. 2018. № 1. С. 26-42; Разумова М. А. Тематические приоритеты и база источников информации деловых СМИ (на примере газет "Коммерсантъ" и "Ведомости"): диссертация ... кандидата филологических наук: 10.01.10 / Место защиты: ФГБОУ ВО «Московский государственный университет имени М. В. Ломоносова», 2016.

Положения, выносимые на защиту:

- В ситуации отсутствия в индустриальном и в академическом сообществах устоявшихся определений журналистики данных представляется обоснованным выделять различные типы журналистики данных, применяя как основной критерий отличия использование при подготовке материалов результатов обработки так называемых «больших данных». В таком случае изучаемое направление может быть разделено на «журналистику количественных данных» в её широком понимании, охватывающем целый спектр материалов, содержащих любые обработанные с применением традиционных статистических инструментов данные, и «журналистику больших данных» как специализированный тип, материалы которого создаются с использованием информации, полученной из соответствующих по масштабам и сложности обработки цифровых баз данных.

- Отобранные для изучения образцы зарубежной и отечественной журналистики данных при сравнительном анализе демонстрируют минимальные отличия и наличие целого ряда общих, универсальных характеристик. В кратком описании публикацию журналистики больших данных можно представить как комментарий к сведениям экономической и политической тематики, сопровождающиеся одной или двумя формами преимущественно статической визуализации данных.

- К специфическим чертам российской журналистики данных можно отнести то, что информационными поводами для создания публикаций часто чаще становятся данные социальных медиа и поисковых систем, а также собственные данные изданий; обработанные данные часто комментируются представителями властей; большее внимание в публикациях уделяется культурной проблематике и теме образования.

- Авторство изученных материалов показывает, что и журналистике данных свойственна моносубъектность – у большинства публикаций один автор. Источники данных крайне редко обрабатываются самими

журналистами, что свидетельствует об отсутствии используемых навыков программирования. В отечественной практике в четверти публикаций отсутствует визуализация данных, а в тех случаях, где она есть, велика вероятность работы специализированных дизайнерских команд. В целом, при создании материалов требуется не более двух компетенций. Все эти характеристики журналистики данных дают основания считать, что сохраняющаяся популярность в отечественном индустриальном и образовательном дискурсах концепций «универсальных журналистов» как редакционных специалистов нового типа с широким кругом компетенций и навыков пока основана на завышенных ожиданиях.

- Изменения в характеристиках рабочих практик и производимого контента у современных отечественных журналистов происходят медленнее, чем это представляется при изучении работ о потенциальных эффектах цифровизации. На этом фоне можно говорить о том, что и трансформации, которые происходят в новостной журналистике под влиянием датафикации, также пока не резкие и точечные и идут во многом теми же темпами и в тех же областях, что и в более традиционных направлениях журналистской деятельности.

Теоретическая значимость исследования состоит в концептуализации понятия журналистики больших данных, в выработке оригинальных методических и методологических решений для изучения этого явления, а также в решении научной проблемы обобщения характеристик журналистики, основанной на новых форматах цифровых данных, в качественной прессе России в коннотации с анализом зарубежных форм журналистики данных.

Практическая значимость исследования. С прикладной точки зрения результаты работы могут быть применены в образовательном процессе, в научной и учебной работе преподавателей и студентов, в практике СМИ и всех иных сферах, в которых используются новые форматы цифровых данных.

Структура работы обусловлена целью и задачами исследования, работа состоит из введения, трех глав, заключения и библиографического списка.

ГЛАВА 1. ЖУРНАЛИСТИКА В УСЛОВИЯХ ДАТАФИКАЦИИ: ТЕОРЕТИКО-МЕТОДОЛОГИЧЕСКИЕ ПОДХОДЫ К ИЗУЧЕНИЮ

1.1. Журналистика данных: генезис, сущность, известные характеристики

Ускоряющиеся перемены в современном обществе, обусловленные развитием технологий, подвергают журналистику трансформациям во всех направлениях – производстве, распространении и использовании продуктов журналистской деятельности.

С появлением новых технологий, работа СМИ становится все более технологически опосредована. *Цифровизация* – «перевод информации в цифровую форму, цифровая трансмиссия информационных данных, закодированных в дискретные сигнальные импульсы»¹⁰, перевод содержания СМИ во всех его формах — текстовой, графической, звуковой — в цифровой формат, понятный современным компьютерам трансформирует работу современных массмедиа. Устраняя различия между отдельными СМИ и создавая единую основу для их содержания, цифровизация прокладывает дорогу к их конвергенции¹¹. Именно цифровизация и конвергенция создают новую модель коммуникации для работников СМИ и аудитории, где базы данных становятся одними из основных характеристик цифровой коммуникации и коммуникации в Сети¹².

Наиболее значимой характеристикой интернет-коммуникации в цифровую эпоху является гипертекстуальность¹³, определяющая большинство свойств технологического и коммуникативного уровней взаимодействия в интернет-пространстве. Благодаря Интернету, характеристиками новой модели коммуникации становятся открытость,

¹⁰ Князев А. А. Энциклопедический словарь СМИ. Бишкек: Издательство КРСУ. 2002.

¹¹ Медиаэкономика зарубежных стран: Учеб. пособие / под ред. Е. Вартановой. М.: Аспект Пресс, 2003.

¹² Интернет-СМИ: Теория и практика. Учебное пособие / под ред. М.М. Лукиной. М.: Аспект Пресс, 2010.

¹³ Термин «гипертекст» ввел Тед Нельсон в 1965 г. для обозначения текста «ветвящегося или выполняющего действия по запросу».

неограниченный доступ к информации и скорость представления и распространения информации. Мультимедийность, как комплексное представление информации в цифровом формате, позволяет создавать уникальный тип контента, применимый в любом канале медиакommunikации. Значимым при подготовке журналистских материалов становится использование такого базового свойства гипертекста, как возможность благодаря гиперссылкам привлекать множество источников (первоисточников) информации – весь Интернет как общую базу данных¹⁴.

Под влиянием этих факторов претерпевают изменения и профессиональные приемы и методы использования факта (и написания текста). В частности, это влияет на обнаружение журналистских сюжетов и их представление аудитории. Компьютеризация способствует продвижению инноваций в журналистике (в обнаружении, анализе, персонализации, агрегации, визуализации и выявлении смыслов информации)¹⁵.

Актуальные направления развития журналистики строятся на использовании *вычислительных методов и инструментов*¹⁶. В контексте современных исследований журналистики, это означает сосредоточение на технологических и технических аспектах, которые определяют журналистику. Исследователи и практики массмедиа заняты поиском оптимальных экономических и творческих моделей медиа. Зарубежные специалисты отмечают, что СМИ способны оставить старую бизнес-модель, основанную на продаже внимания своей аудитории рекламодателям, и сосредоточиться на поиске альтернативы. Они должны приспособливаться к

¹⁴ Шилина М. Г. *Data Journalism* – дата-журналистика, журналистика метаданных – в структуре медиакommunikации: к вопросу формирования теоретических исследовательских подходов // Медиаскоп. 2013. Вып. 1.

¹⁵ Cohen S. (2014) Research Chat: Sarah Cohen of the New York Times on the state of data journalism and what reporters need to know. Harvard Kennedy School. Shorenstein Center on Media, Politics and Public Policy. Available at: <http://www.journalistsresource.org/skills/reporting/research-chat-new-york-times-Sarah-Cohen-state-data-journalism-what-reporters-need-know#>

¹⁶ Cohen S., Hamilton J. T., Turner F. (2011) Computational Journalism. Communications of the ACM. №10.

новым технологиям и возможностям, становятся центрами, платформами надежных данных¹⁷.

Влияние датафикации и больших данных на массмедиа

С накоплением больших объемов цифровой информации и появлением цифровых форматов данных связан новейший период цифровизации. Сегодня цифровые данные становятся источником новостей и анализа, инструментом, с помощью которого аудитории подается сюжет. Журналисты, осуществляющие деятельность в цифровую эпоху, находят сюжеты среди больших массивов информации и представляют их аудитории. Располагая массивами данных, они дополняют их анализом, объяснением для читателей. Таким образом, одной из задач журналиста XXI в. становится разъяснение смысла новых типов данных, основанной на них информации, их потенциала для аудитории.

Ключевой тенденцией в мире цифровой информации становится датафикация, связанная с понятием данных – «представлением фактов, понятий или инструкций в форме, приемлемой для общения, интерпретации, или обработки человеком или с помощью автоматических средств»¹⁸. Она представляет собой процесс представления явлений в количественном формате, преобразование данных в информацию, доступную для обработки¹⁹.

В рамках нашей работы мы определяем *датафикацию* как преобразование привычных форм восприятия действительности в цифровую форму, приемлемую для анализа и интерпретации.

Датафикация обладает различными проявлениями. Известны примеры датафикации литературных произведений: инициативы по датафикации

¹⁷ Lorenz M., Kayser-Bril N., McGhee G. (2011) Media companies must become trusted data hubs // Ownieu. Available at: <http://www.owni.eu/2011/28/02/media-companies-must-become-trusted-data-hubs-catering-to-the-trust-market/>

¹⁸ Systems and software engineering. ISO/IEC/IEEE 24765-2010. Vocabulary. Available at: <https://www.iso.org/obp/ui/#iso:std:iso-iec-ieee:24765:ed-1:v1:en>

¹⁹ Mayer-Schönberger V., Cukier K. Big Data: A Revolution That Will Transform How We Live, Work, and Think. – UK: John Murray, 2013.

Google и *Amazon*, проект «Гутенберг» – общественная инициатива по размещению различных произведений в Интернете для общего пользования. Датафикация местоположений – возможность сбора геолокационных данных о пользователях – становится чрезвычайно ценной (на уровне отдельных лиц она позволяет, к примеру, нацеливать рекламу, исходя из местоположения человека или его предполагаемого пункта назначения), геолокационные данные используются для выявления важных тенденций, которые в свою очередь, могут быть использованы журналистами²⁰. Датафикация принципов работы человеческого тела (например, анализ походки человека с целью использования этих данных в системе безопасности²¹) открывает широкое поле для изучения журналистами, представления в СМИ. Инфоповодами для написания журналистских материалов все чаще становятся результаты датафикации информации о взаимодействии людей, которую сегодня генерируют социальные медиа.

Во втором десятилетии XXI в. датафикация тесно связывается с распространением цифровизации и инновационных информационных технологий, и, как следствие, с концепцией больших данных. Согласно этой концепции *большие данные* представляют собой цифровые базы данных значительного многообразия и огромных объёмов, исчисляемых в петабайтах²², а также набор инструментов и методов их обработки для получения воспринимаемых человеком результатов и дальнейшего анализа. Термин «большие данные» характеризует «совокупность данных с возможным экспоненциальным ростом, которые слишком велики, слишком неформатированы или слишком неструктурированы для анализа

²⁰ Mayer-Schönberger V., Cukier K. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. – UK: John Murray, 2013.

²¹ Derawi Biometrics. Your walk is your PIN-code. Press release. Feb. 21, 2011. Available at: <http://biometrics.derawi.com/?p=175>.

²² Петабайт (ПБ, ПБайт) — единица измерения количества информации, равная 10^{15} (квадриллион) байт.

традиционными методами»²³. Они относятся к наборам данных, «размер которых выходит за пределы возможностей типичных программных средств сбора, хранения, управления и анализа баз данных»²⁴.

Первые попытки количественной оценки темпов роста объема данных предпринимались семьдесят лет назад в контексте процессов, которые в Оксфордском словаре английского языка в 1941 г. были обозначены термином «информационный взрыв» (англ. *information explosion*). Гил Пресс фиксирует первое осмысленное использование термина «большие данные» в октябре 1999 г. в цифровой библиотеке Ассоциации компьютерных вычислений (*ACM Digital Library*) – в статье исследователей NASA, посвященной проблемам визуализации информации²⁵. Популяризируется термин «большие данные» К. Линчем, редактором журнала *Nature*, подготовившим в 2008 г. специальный номер на тему «Как могут повлиять на будущее науки технологии, открывающие возможности работы с большими объёмами данных?». В нем размещаются материалы о феномене взрывного роста объёмов и многообразия обрабатываемых данных и технологических перспективах в парадигме вероятного скачка «от количества к качеству». Термин предлагается по аналогии с употребляемыми в деловой англоязычной среде метафорами «большая нефть», «большая руда».

В качестве определяющих характеристик больших данных исследователи отмечают «три V»: объём (англ. *volume*, в смысле величины физического объёма), скорость (англ. *velocity* в смыслах, как скорости прироста, так и необходимости высокоскоростной обработки и получения результатов), многообразие (англ. *variety*, в смысле возможности

²³ Моррисон А. Большие Данные: как извлечь из них информацию // «Технологический прогноз». PricewaterhouseCoopers, 2010. Вып. 3. С. 42.

²⁴ Manyika J. et al. (2011) Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute, June, 2011. McKinsey (9 August 2011).

²⁵ Cox M., Ellsworth D. Application-controlled demand paging for out-of-core visualization. Proceedings of the 8th conference on Visualization '97. Phoenix, Arizona, USA, 1997. Available at: https://www.evl.uic.edu/cavern/rg/20040525_renambot/Viz/parallel_volviz/paging_outofcore_viz97.pdf

одновременной обработки различных типов структурированных и полуструктурированных данных)²⁶.

Исследователи отмечают, что размер больших данных неустойчив, он постоянно меняется, в конкретных случаях составляя от нескольких десятков терабайт²⁷ до многих эксабайт²⁸ данных. Считается, что размер больших данных также зависит и от возможностей инструментов их обработки. «Для некоторых организаций, впервые сталкивающихся лишь с сотнями гигабайт данных, может возникать необходимость пересмотра вариантов управления ими. Для других могут потребоваться десятки или сотни терабайт, прежде чем размер данных станет для них существенным фактором»²⁹.

Базовыми источниками возникновения больших данных являются «непрерывно поступающие данные с измерительных устройств, от радиочастотных идентификаторов, потоки сообщений из социальных сетей, метеорологические данные, данные дистанционного зондирования Земли, потоки данных о местонахождении абонентов сетей сотовой связи, устройств аудио- и видеорегистрации»³⁰, а также внутренние информационные системы предприятий и организаций.

В рамках данной работы мы определяем *большие данные*, как массивы цифровых данных, которые из-за их объема и сложности, не могут быть собраны, сохранены и проанализированы без помощи специальных горизонтально масштабируемых программных инструментов обработки, появившихся в начале XXI века (*NoSQL, MapReduce, Hadoop, R*). Их источниками чаще всего являются транзакционные данные, данные, генерируемые социальными медиа, веб-сервисами, мобильными

²⁶ Канаракус К. Машина Больших Данных // Сети. 2011. № 4.

²⁷ Терабайт (Тбайт, ТБ; TB, TByte) — единица измерения количества информации; может означать 10^{12} (триллион) или 2^{40} байт.

²⁸ Эксабайт (Эбайт, ЭБ, Eb) — единица измерения количества информации, равная 10^{18} (квинтиллион) байт.

²⁹ Magoulas R., Lorica B. Introduction to Big Data. Release 2.0. – Sebastopol CA: O'Reilly Media, 2009.

³⁰ Черняк Л. Большие Данные — новая теория и практика // Открытые системы. 2011. № 10.

приложениями, интернетом вещей и т.д., которые оформляются в базы данных огромных объемов размером от нескольких петабайт.

Так, например, экономические или демографические данные Росстата можно отнести к обычным количественным данным, в то время как сведения о перемещениях смартфонов, собираемые мобильными операторами, формируют большие данные.

Развитие и распространение датафикации способствует проникновению технологий больших данных в сферу государственного управления, коммерческий сектор и массмедиа. Однако доступ журналистов к данным и их использование, как правило, контролируется государственными и частными организациями, что сегодня может замедлять развитие журналистики данных, особенно в России. Данный контроль производится через патенты, лицензии, требования оплаты доступа к данным или их повторного использования. В противовес этому в середине 2010-х гг. оформляется концепция открытых данных. Сторонники данной концепции полагают, что данные должны быть доступны любому гражданину, любые ограничения не соответствуют идее гражданского общества.

Концепция открытых данных представляет собой идею свободно доступных данных для машиночитаемого использования и публикаций без ограничений авторского права, как правило, из государственных источников. Обращение исследователей к концепции открытых данных во второй половине 2000-х гг. связано с запуском правительственных инициатив. Государственные данные востребованы обществом и массмедиа, журналисты и активисты добиваются открытости государственной информации в машиночитаемой форме.

Основным источником цифровых данных для журналистов остается Интернет. В нем функционируют лидирующие поисковые системы Яндекс, *Google*, *Bing*, а также присутствуют общественные проекты, такие как *Wikipedia*, *OpenStreetMap* и др. Появляются инициативы по интеграции информации в базы данных. Например, инициатива поисковой системы

Schema.org, публикующей данные в специальных форматах, инициативы *Semantic Web* и *Linked Data*, заключающиеся в приведении всех данных к общим форматам и специальном способе их публикации подходящем для извлечения информации сразу из нескольких источников³¹. В последние годы происходит интеграция интернет-сервисов (например, коллекция веб-сервисов на ресурсе *ProgrammableWeb*), открываются порталы и каталоги открытых данных³².

С распространением Интернета, цифровизации, датафикации, становится востребованной журналистика, опирающаяся на анализ информации из обширных баз цифровых данных. Со второй половины 2010-х гг. трансформации в использовании больших данных в медиа продолжают, «накопление данных и неограниченный доступ к цифровой информации становятся важными составляющими выживания журналистики как элемента общественной и экономической сферы»³³.

Таким образом, цифровизация и датафикация в современном мире приводят к следующим изменениям параметров журналистики: происходит увеличение объемов цифровой информации, изменяются её форматы и появляются новые, контент становится интерактивным, в результате подобных трансформаций и дальнейшего распространения новых технологий источники больших объемов данных и сервисов становятся доступными для изучения и работы журналистов.

³¹ Бегтин И. Готовы ли мы к журналистике данных? // Полит.ру. 29 апр. 2013. Режим доступа: http://polit.ru/article/2013/04/29/data_journalism/

³² Data.gov (US), Available at: <http://www.data.gov>; Data.gov.uk (UK) Available at: <http://data.gov.uk>; Worldbank Data, Available at: <http://data.worldbank.org>; Scaperwiki, Available at: <http://scaperwiki.com>; OpenGovData.ru (Ru), Available at: <http://www.opengovdata.ru>; Data.Gov.ru (Ru), Available at: <http://www.datagov.ru>; Open Knowledge Foundation, Available at: <http://okfn.org/>

³³ Van der Haak B., Parks M., Castells M. The Future of Journalism: Networked Journalism. *International Journal of Communication*. 2014. Available at: <http://ijoc.org/index.php/ijoc/article/view/1750/832>

Развитие журналистики данных

В зарубежных массмедиа различные варианты использования, анализа, интерпретации, представления, в том числе визуализация существенных объемов данных, присутствуют с XIX в. – еще до появления современных цифровых технологий и методов анализа³⁴.

Примеры первых попыток сравнения различных типов данных для создания журналистского материала исследователи относят к первой трети XIX в., упоминая статью *The Guardian* 1821 года. Статья содержала список школ Манчестера с указанием числа школьников различных категорий³⁵, а автор на основе сравнительного анализа выявлял реальное число учеников, получающих бесплатное образование, которое оказывалось гораздо выше показателей, содержащихся в официальных государственных данных³⁶. Также можно упомянуть исследование Флоренс Найтингейл 1852 года – «Смертность в британской армии» (*The Mortality in the British Army*), в

³⁴ Отметим, что в 1812-1813 гг. французский инженер Чарльз Джозеф Минард впервые использовал графики в инженерии и статистике. Линейным графиком он проиллюстрировал события Отечественной войны 1812 года после Бородинского сражения: возвращение русской армии в Москву и отход французских войск. На графике также была выведена зависимость смертности во французской армии от падения температуры. Другой пример визуализации данных относится к 1854 году, когда доктор Джон Сноу, анализируя данные о числе заболевших и умерших от холеры, составил карту Лондона, на которой чёрными столбиками отметил количество смертей в каждом доме. Он обратил внимание, что большинство смертей сосредоточено у колонки на Брод-стрит. Визуализация стала красноречивым доказательством, которое убедило мэра Лондона закрыть колонку, после чего эпидемия пошла на спад. См. Friendly M. (2002) Visions and Re-Visions of Charles Joseph Minard. *Journal of Educational and Behavioral Statistics*, 27(1): 31-51.

³⁵ The first Guardian data journalism // *The Guardian*. May 5, 1821. Available at: <https://www.theguardian.com/news/Datablog/2011/sep/26/data-journalism-guardian>

³⁶ Однако существует мнение, что журналистика данных появляется ещё раньше. В частности, С. Роджерс, основатель и редактор *The Guardian Datablog*, пишет, что журналистика данных берет свое начало не в 2009 г. (год создания раздела), и даже не в 1960-х и 1970-х гг. с появлением компьютерной и прецизионной журналистики, работ Ф. Мейера, не с появлением публикацией исследования Ф. Найтингейл или первого выпуска *The Guardian*. Она так же стара, отмечает С. Роджерс, как прогноз погоды или биржевой тикер. Цифровизация, социальные сети, размещение материалов качественной прессы в Интернете лишь трансформируют существование данного феномена. (Rogers S. What does data journalism look like today? // *Medium*. 2017. Available at: <https://medium.com/data-journalism-awards/what-does-data-journalism-look-like-today-a-10-step-guide-6dd90c1f0c25>).

котором автор визуализирует информацию о смертях круговыми секторными диаграммами, показывающими, что подавляющее большинство смертельных случаев в британской армии связано не с вражескими пулями, а с предотвратимыми заболеваниями³⁷. Оно, по мнению исследователей, является одним из значимых прообразов современной журналистики, основанной на статистических методах работы с данными и их визуализацией.

Журналистика, основанная на использовании цифровых данных и информационных технологий, зарождается в 1950-е гг. в США. Так, в истории становления журналистики, основанной на работе автора с цифровыми данными, исследователи выделяют первоначальный, предварительный период³⁸. Его начало связывается с появлением «компьютерной», или «вычислительной», журналистики (*computer-assisted reporting – CAR*). Систематический подход к использованию компьютеров для сбора и анализа данных способствовал улучшению качества новостей: так, в 1952 г. телеканал *CBS* использует компьютер *UNIVAC I*, единственный компьютер общего назначения того времени, для предсказания результатов президентских выборов³⁹.

В 1960-е гг. сформировались по-настоящему значимые журналистские практики с использованием ЭВМ и вычислительного инструментария различного рода для анализа и представления информации. Например, в 1967

³⁷ Edwin W. (2016) *Kopf's, Florence Nightingale as Statistician*. Publications of the American Statistical Association. 15(116): 388-404; Rogers S. *Florence Nightingale, datajournalist: information has always been beautiful // The Guardian Datablog*. 13 Aug. 2010. Available at: <https://www.theguardian.com/news/Datablog/2010/aug/13/florence-nightingale-graphics>

³⁸ Borges-Rey E., Stalph F. (2018) *A random walk through data + journalism: assessing the data journalism ecosystem*. Paper presented on NODA-2018 (The Nordic Data Journalism Conference), Sweden: Södertörn University. March 15-17 2018.

³⁹ Соколов В.С. *Американский «новый журнализм» – концепция социальной мифологии // Демократический журналист*. 1977. № 4–6; Уллмен Дж. *Журналистские расследования: современные методы и техника*. М., 1998; Dione E. (1991) *Why Americans Hate Politics*. Washington; Cox, M. (2000) *The development of computer-assisted reporting* Archived 2011-09-28 at the Wayback Machine. Paper presented to the Newspaper Division, Association for Education in Journalism and Mass Communication, Southeast Colloquium, March 17–18, 2000, University of North Carolina, Chapel Hill.

г. Ф. Мейер, журналист газеты *Detroit Free Press*, освещая беспорядки в Детройте и исследуя их причины, использовал компьютер для работы с социологическими данными⁴⁰. В 1980-е годы Б. Дедман – журналист *The Atlanta Journal-Constitution* – получает Пулитцеровскую премию за серию материалов «Цвет денег» (*The Color of Money*)⁴¹ о том, что малообеспеченные белые семьи, проживающие в Атланте, чаще получают кредиты банков, чем афроамериканцы среднего достатка, наглядно демонстрируя информацию о систематических расовых предрассудках в кредитной политике ведущих финансовых институтов США. Следуя образовавшейся традиции в начале 1990-х годов С. Доиг в своей работе «Что пошло не так» (*What Went Wrong*)⁴² применяет статистические методы анализа информации и визуализацию обработанных им данных при исследовании разрушений во время урагана «Эндрю», чтобы проанализировать ущерб и понять, в какой степени на масштаб этого ущерба оказали влияние недостатки в области политики городского развития. Подобные репортажи на основе компьютерных данных постепенно получают широкое общественное признание и приносят журналистам профессиональные награды.

В 1970-х гг. для описания данного нового типа сбора новостей вводится термин «прецизионная журналистика» (англ. *precision journalism*), в основании которой лежали представления об устаревании традиционных методов сбора и обработки информации, которые уже не позволяли своевременно выявлять потенциально важные социально-политические явления, акцентируют внимание лишь на сиюминутных событиях. Для устранения имеющихся недостатков, по мнению Ф. Мейера⁴³, в журналистике становятся актуальными социологические и статистические

⁴⁰ Meyer F. (1967) A Survey of Attitudes of Detroit Negroes After the Riot of 1967 // Detroit Urban League.

⁴¹ Dedman, B. (1989) The Color of Money // *The Atlanta Journal-Constitution*. Available at: <http://powerreporting.com/color/>

⁴² Doig, S. What Went Wrong // *The Miami Herald*. December 20, 1992. Available at: <http://www.derivativeworks.com/2008/09/complete-scans.html>

⁴³ Meyer P. Precision Journalism: A Reporter's Introduction to Social Science Methods. – Rowman & Littlefield Publishers Inc, 2002.

методы сбора и обработки информации. В 1973 г. в своей книге «Прецизионная журналистика», первой попытке систематизировать процессы, с помощью которых компьютеры трансформируют работу журналиста, Ф. Мейер утверждал, что журналист должен использовать базы данных и собственные исследования совместно, объединяя их с помощью компьютера. В редакции книги 2002 г. он идет еще дальше, заявляя, что сегодня «журналист должен стать управляющим (англ. *manager*) баз данных»⁴⁴.

В 1970-е гг. технология, инструменты и методы изучения опросных данных и данных социологических исследований дополняются для журналистов новыми возможностями: использованием компьютера для анализа огромных объемов данных, содержащихся в государственных документах, находящихся в открытом доступе. Первым примером такого типа материала служит публикация К. Джонса – журналиста *The Miami Herald* – который в 1969 г., применяя компьютерные технологии, выявил закономерности в системе уголовного правосудия США⁴⁵. Среди других известных ранних примеров – публикации журналиста *The New York Times* Д. Бернхэма, который в 1972 г. использовал компьютер для выявления расхождений в показателях преступности, о которых сообщала полиция⁴⁶. А Э. Джаспин – журналист издания *The Providence Journal* – в 1986 г. создает ряд материалов, основанных на сопоставлении баз данных, для выявления нарушений водителей школьных автобусов, работников с плохой репутацией и судимостью⁴⁷.

К 2001 г. компьютеризация американских редакций достигает критической отметки с точки зрения использования ЭВМ журналистами при

⁴⁴ Там же.

⁴⁵ DeFleur M. H. *Computer-assisted Investigative Reporting: Development and Methodology*. – UK: Routledge, 2013.

⁴⁶ Burnham D. A Wide Disparity Is Found in Crime Throughout City // *The New York Times*, February 14, 1972, p. 1.

⁴⁷ Jaspin E. Sons, Daughters of State Leaders Got 8 1/2 Percent RIHMFC Loans // *The Providence Journal*. June 2, 1985, p. 2.

написании материалов, онлайн-исследований, поиска материалов и ежедневной частоты использования Интернета⁴⁸, что свидетельствует о том, что компьютерная, вычислительная журналистика укореняется в практике СМИ. Для продвижения технологий вычислительной журналистики в СМИ создаются профессиональные журналистские организации, такие как Национальный институт вычислительной журналистики (*The National Institute for Computer-Assisted Reporting, NICAR*) – программа американской некоммерческой организации репортёров и редакторов-расследователей (*Investigative Reporters and Editors, IRE*), Датский международный центр аналитической журналистики (*The Danish International Center for Analytical Reporting, DICAR*) и др. «Глобальная сеть расследовательской журналистики» (*Global Investigative Journalism, GIJN*) размещает список образовательных ресурсов по журналистике данных на сайте⁴⁹. Также создаются профессиональные списки рассылок специалистов *CAR* (*NICAR-L, CARR-L* и *JAGIS-L* и др.), способные помочь журналистам в написании материалов⁵⁰.

Именно появление *CAR* в 1952 г. как основы анализа данных журналистами с помощью компьютерных технологий; зарождение прецизионной журналистики в 1973 г. как социально-ориентированного подхода к созданию материалов на основе данных; открытие института *NICAR* в 1989 г. в качестве академического и профессионального консолидатора, по мнению современных исследователей⁵¹, приводит к тому, что с конца 2000-х гг. оформляется новое направление в журналистике – журналистика данных (англ. *data journalism*), опирающееся на

⁴⁸ Garrison B. (2001) Diffusion of online information technologies in newspaper newsrooms. *Journalism*, (2): 221-239.

⁴⁹ GIJN – Global Investigative Journalism Network. Available at: <https://gijn.org/ресурсный-центр-gijn-глобальной-сети-жур/>

⁵⁰ Для журналистов, занимающихся изучением журналистики данных, ресурсы *Data Driven Journalism List*, *Project Wombat*, списки рассылок фонда *Open Knowledge Foundation* и *theInfo* и сегодня остаются отправными точками.

⁵¹ Borges-Rey E. and Stalph F. (2018) A random walk through data + journalism: assessing the data journalism ecosystem. Paper presented on NODA-2018 (The Nordic Data Journalism Conference). Södertörn University, Sweden, 15-17 Mar. 2018.

предшествующие традиции. Его особенность заключается в том, что для поиска информационного повода и создания истории ведущими медиа все чаще используются цифровые форматы информации, открытые базы данных. Сегодня цифровизация, компьютеризация, распространение программного обеспечения изменяют работу современных журналистов, способных находить инфоповоды в базах данных, анализировать данные с помощью электронных таблиц и статистических программ, изучать политические и демографические изменения с помощью геоинформационных систем и т.п.

Термин «журналистика данных» предлагается в 2006 г. американским журналистом, сторонником внедрения сложных методов компьютерного программирования в рабочий процесс производства медиаконтента Э. Головатым для обозначения структурированных, машиночитаемых данных, используемых в журналистике вместе с традиционным текстом⁵². Э. Головатый разбивает производство публикаций журналистики данных на три этапа: сбор, «очищение» (англ. *distilling*) и представление информации, отмечая значение автоматизации как ценного фактора, способствующего динамичности и эффективности данных этапов⁵³. С именем Э. Головатого связан один из первых примеров журналистского проекта, основанного на больших данных – использование журналистом алгоритмов сбора и анализа информации с сайтов муниципальных служб Чикаго⁵⁴. Сначала проект существует в рамках раздела *ChicagoCrime*⁵⁵ газеты *Chicago Tribune* со временем становится самостоятельным генератором новостей в качестве проекта *EveryBlock*⁵⁶.

Постепенно термин «журналистика данных» приобретает более широкое значение. Современные исследователи, в частности, Э. Аппельгрэн и Г.

⁵² Holovaty A. (2006) A fundamental way newspaper sites need to change. Available at: <http://www.holovaty.com/writing/fundamental-change/>

⁵³ Niles R. (1995) Finding Data on the Internet. How to Make Money Publishing Community News Online. Available at: <http://www.robertniles.com/data/>

⁵⁴ Сайт Э. Головатого. Режим доступа: [Holovaty.com](http://www.holovaty.com)

⁵⁵ Crime in Chicagoland (2014) // *Chicago Tribune*. Available at: <http://crime.chicagotribune.com>

⁵⁶ EveryBlock // *Chicago Tribune*. Available at: Chicago.everyblock.com

Нюгрен определяют: журналистика данных – это форма повествования на основе журналистских расследований, реализованная путем сбора, накопления, обработки, анализа и представления больших объемов данных. Эта форма журналистики часто представляется в виде интерактивных графиков и простых интерфейсов для пользователя на веб-страницах или на других мобильных каналах публикации. Она публикуется вместе с новостными статьями, аудио- или видеоконтентом⁵⁷.

В конце 2000-х гг. (2009 - 2010 гг.) публикации *WikiLeaks* и скандалы с парламентскими расходами придают термину «журналистика данных» массовую известность. Самые технически и технологически развитые ведущие медиаорганизации – *The Guardian* и *The New York Times* – используют огромные объемы цифровых данных для написания значимых материалов и возглавляют внедрение практик журналистики данных в работу СМИ. Другие медиа следуют их примеру: в США – *Huffington Post*, *Chicago Tribune*, *ProPublica*, в Великобритании – *Financial Times*, *BBC* и Бюро расследовательской журналистики (*The Bureau of Investigative Journalism*), в Аргентине – *La Nación*, во Франции – *Le Monde*, в Германии – *Der Spiegel*, *Deutsche Welle*, *ZeitOnline*, в Норвегии – *VerdensGang*, в Швеции – *SVT*, в Финляндии – *HelsinginSanomat*.

В начале 2010-х гг., несмотря на отсутствие четкого определения, журналистика данных оформляется в одно из популярных направлений журналистики, продолжает развиваться и распространяться, качественные издания размещают на своих страницах публикации, основанные на анализе и интерпретации больших массивов данных, создают отдельные проекты в Интернете и открывают разделы на своих сайтах. Среди примеров первых

⁵⁷ Appelgren E. and Nygren G. 'Data Journalism in Sweden - Opportunities and Challenges'. In: Clare Mills, Michael Pidd and Esther Ward. *Proceedings of the Digital Humanities Congress 2012*. Studies in the Digital Humanities. Sheffield: The Digital Humanities Institute, 2014. Available at: <https://www.dhi.ac.uk/openbook/chapter/dhc2012-appelgren>

публикаций в США серия статей в *Las Vegas Sun* о больничной помощи⁵⁸ и интерактивные проекты *Chicago Tribune* в специальном разделе, посвящённом журналистике данных, иллюстрирующие, к примеру, статистику преступлений в Чикаго с 2001 г. по настоящее время с картами с мест преступлений и именами жертв⁵⁹. На сайте издания *Texas Tribune* также появился раздел, в котором публикуются статьи и проекты на основе анализа данных (например, интерактивная карта Техаса с 7000 мест утилизации нефтегазовых операций⁶⁰).

Первые попытки использования больших данных журналистами основываются на спонтанно выбранной теме, актуальной и интересной читателям, являются примерами еще не отрегулированных технологий работы. Так, британская ежедневная газета *The Daily Telegraph* использует простые программы визуализации и методы выявления связей внутри массива отчетов членов Парламента о расходах, которые они осуществляют в рамках своей депутатской деятельности. В ходе расследования журналисты обнаружили траты представительских средств на приобретение недвижимости, на личные расходы, а также махинации с налогами⁶¹. Парламент в ответ на обвинения газеты рассекречивает еще около 460 тысяч страниц отчетов, которые *The Guardian* размещает на своем сайте как базу данных и более 33 тысяч читателей принимают участие в изучении статей расходов⁶².

Постепенно журналистика данных становится набирающим силу трендом в европейских СМИ. В 2010-е гг. качественные издания вырабатывают свои практики создания публикаций на основе использования

⁵⁸ Do No Harm (2010) // *The Las Vegas Sun*. Available at: <http://www.lasvegassun.com/hospital-care/>

⁵⁹ Crime in Chicagoland (2014) // *Chicago Tribune*. Available at: <http://crime.chicagotribune.com>; Chicago homicides (2014) // *Chicago Tribune*. Available at: <http://crime.chicagotribune.com/chicago/homicides>

⁶⁰ Map: Texas Disposal Wells (2013) // *Texas Tribune*. Available at: <http://www.texastribune.org/series/water-for-fracking/map/>

⁶¹ Parliament // *The Daily Telegraph*. Available at: Parliament.telegraph.co.uk

⁶² Mps-expenses // *The Guardian*. Available at: Mps-expenses.guardian.co.uk

новых форматов данных, расширяются возможности подачи материала и взаимодействия с аудиторией. Например, проекты *The Guardian Datablog* и *Datastore* предоставляют пользователям доступ к программам, способным генерировать собственные визуализации данных⁶³; американский проект «Район на миллион долларов» (*One Million Dollar Block*)⁶⁴ представляет интерактивные визуализации городских бюджетов в разрезе кварталов крупных городов и др. На сайте *BBC News* размещается ряд материалов, основанных на больших данных (ещё в 1999 г. публикуются рейтинги школ, в которых используются данные, ежегодно обнародываемые правительством⁶⁵), *Financial Times* публикует статьи в отдельном разделе на сайте (к примеру, о взаимодействии делегатов на Всемирном экономическом форуме в Давосе⁶⁶). Продолжают появляться новые американские проекты, размещающие на своих сайтах в Сети публикации журналистики данных. Так, в 2014 г. создаются *The Upshot*⁶⁷, *Vox*⁶⁸ и *FiveThirtyEight*⁶⁹.

Появляется практика создания баз данных различных организаций, в том числе международных и государственных, создаются государственные порталы открытых данных, инструменты анализа и визуализации цифровых данных, упрощающие работу журналистов и способствующие дальнейшему распространению журналистики данных⁷⁰. Сегодня профессионалы,

⁶³ Проекты, посвященные большим и открытым данным, на сайте британской газеты *The Guardian*. Режим доступа: www.theguardian.com/data

⁶⁴ *One Million Dollar Block* // Available at: <http://www.spatialinformationdesignlab.org/projects.php>

⁶⁵ Проект *BBC News*. Режим доступа: <http://www.bbc.com/news/education-11950098>

⁶⁶ *The tribes of Davos* // *Financial Times*. – 2015, 22 Jan. Available at: <http://blogs.ft.com/ftdata/page/2/>

⁶⁷ *The Upshot*. Available at: http://www.nytimes.com/upshot/?_r=2

⁶⁸ *Vox*. Available at: <http://www.vox.com>

⁶⁹ *FiveThirtyEight*. Available at: <http://fivethirtyeight.com>

⁷⁰ Каталоги открытых данных: *OpenGovData.ru* (Россия). Режим доступа: <http://www.opengovdata.ru>; *DataGov.ru* (Россия). Режим доступа: <http://www.datagov.ru>; *Data.gov* (US). Режим доступа: <http://www.data.gov>; *Data.gov.uk* (UK). Режим доступа: <http://data.gov.uk>; *Worldbank, Data*. Режим доступа: <http://data.worldbank.org>; *Open Knowledge Foundation*. Режим доступа: <http://okfn.org>. Ссылки по визуализации: *Visual Complexity*. Режим доступа: <http://www.visualcomplexity.com/vc>; *Flowing Data*. Режим доступа: <http://flowingdata.com>; *Well-formed Data*. Режим доступа: <http://well-formed-data.net/>; *Infografistas*. Режим доступа: <http://infografistas.blogspot.com>; *Visual Editors*. Режим

специализирующиеся на данном направлении журналистики, взаимодействуют с различными видами данных – базами данных, данными, представленными в Интернете, данными, генерируемыми социальными медиа и др.

В последние годы исследователи начинают говорить о де-вестернизации журналистики данных⁷¹. Однако новейшие исследования подтверждают зависимость журналистики данных развивающихся стран от влияния развитых медиасистем. Так, опровергает тезис о де-вестернизации журналистики данных исследование Д. Черуйот и Р. Феррер Конилл⁷², в котором авторы сопоставляют и наглядно визуализируют использование данных африканскими медиа, отмечая заметное участие западных сторонних организаций, низкое качество данных и сложность поиска их источников.

Аналогичные тенденции можно наблюдать в странах АСЕАН, где обучение журналистике данных обеспечивается Агентством по сотрудничеству Министерства иностранных дел Франции в Европе (*Cooperation Agency of the French Ministry of Europe and Foreign Affairs, CFI*)⁷³, которому подконтрольны 12 организаций в Камбодже, Индонезии, Малайзии, Мьянме, Вьетнаме, Таиланде и на Филиппинах, а также Программой посольства США по обучению журналистике данных во Вьетнаме во главе с профессорами Филадельфийского университета.

доступа: <http://www.coolinfographics.com>; Cool Infographics. Режим доступа: <http://www.coolinfographics.com>; Datenjournalist. German blog on ddj. Режим доступа: <http://www.datenjournalist.de>. Инструменты: Document Cloud. Режим доступа: <http://www.documentcloud.org/home>; Google Code Playground: Helps to explore Google data and tools. Режим доступа: <http://code.google.com/apis/ajax/playground>; Open Calais. Режим доступа: <http://www.opencalais.com>; API Playground – helping journalists to understand API Data. Режим доступа: <http://apiplayground.org>; Gapminder Desktop. Режим доступа: <http://www.gapminder.org/desktop>; Yahoo Pipes. Режим доступа: <http://pipes.yahoo.com/pipes>

⁷¹ Borges-Rey E. and Stalpf F. (2018) A random walk through data + journalism: assessing the data journalism ecosystem. Paper presented on NODA-2018 (The Nordic Data Journalism Conference). Södertörn University, Sweden, 15-17 March.

⁷² Cheruiyot D. and Ferrer Conill R. (2018) Expanding journalistic practice through civic technologies and data journalism. Paper presented on NODA-2018 (The Nordic Data Journalism Conference). Södertörn University, Sweden, 15-17 March.

⁷³ Cooperation Agency of the French Ministry of Europe and Foreign Affairs. Available at: <https://www.diplomatie.gouv.fr/en/the-ministry-and-its-network/ministry-for-europe-and-foreign-affairs-implementing-agencies/>

Пакистанский журналист, сотрудник *The News International*, Х. Хаттак в 2015 г. запускает веб-сайт *Data Stories*⁷⁴ о работе с большими данными и их использовании в журналистских материалах после изучения курса журналистики данных в «Центре передового опыта в области журналистики» в Карачи, который организуется совместно с «Международным центром для журналистов» (*International Center for Journalists, ICFJ*)⁷⁵.

Значимые материалы журналистики данных продолжают публиковаться в СМИ по всему миру (например, размещаются в постоянно обновляемом разделе на сайте аргентинской газеты *La Nación*, на страницах других латино-американских качественных изданий), исследователи из разных стран изучают публикации журналистики данных, конференции в заданной области проводятся регулярно и повсеместно (например, Конференция в области журналистики данных и цифровой методики в Бразилии⁷⁶). В 2017 г. активистами платформы «Открытые данные в Европе и Центральной Азии» (*Open Data in Europe and Central Asia, ODECA*) публикуется руководство по созданию публикаций журналистики данных, переведенное на многие национальные языки⁷⁷.

Сегодня исследователи говорят о появлении «гиперлокальной» журналистики данных, объединяющей специалистов из разных профессиональных областей, разных стран и регионов, но при этом способных работать вместе благодаря распространению новых информационных технологий. Появление термина связано с созданием Бюро локальной расследовательской журналистики (*Bureau Local – The Bureau of Investigative Journalism*)⁷⁸ – инновационного проекта, объединяющего

⁷⁴ Data Stories. Available at: <http://www.datastories.pk/>

⁷⁵ International Center for Journalists, ICFJ. Available at: <https://www.icfj.org/>

⁷⁶ Coda BR: Brazilian Conference on Data Journalism and Digital Methods. Available at: <https://www.journalism.co.uk/events-calendar/coda-br-brazilian-conference-on-data-journalism-and-digital-methods/s29/a712474/> <https://coda.escoladedados.org/>

⁷⁷ Data Journalism Manual (2017) Open Data in Europe and Central Asia, ODECA. Available at: <http://www.odcanet.org/data-journalism-manual/>

⁷⁸ The Bureau Local-Bureau of Investigative Journalism. Available at: <https://www.thebureauinvestigates.com/local>

местных журналистов для разбора баз данных и создания публикаций журналистики данных, которые имеют значение на уровне сообщества; а также с созданием других небольших дата-сообществ и объединений. В рамках Партнерства местных новостей «Би-би-си» (*BBC's Local News Partnerships*) создается Общий отдел по работе с новыми форматами цифровых данных (*BBC's Shared Data Unit*)⁷⁹ – подразделение, объединяющее опытных журналистов в области анализа и обработки данных с журналистами из более широких областей медиаиндустрии.

С начала 2010-х гг. направление журналистики данных можно считать уже признанным научным сообществом. Исследователи отмечают, что она состоит из комбинации компетенций и навыков нескольких профессий⁸⁰, представляет собой «пересекающийся набор компетенций, взятых из разнородных полей»⁸¹. Журналистика данных широко используется для объединения с журналистикой ряда технологических концепций, рассматриваемых как стадии, ведущие от простого к более сложному уровню использования новых технологий в журналистском процессе⁸².

Определения журналистики данных

При этом необходимо отметить, что общепринятого определения журналистики данных пока нет ни в академической среде, ни среди практиков.

Существует несколько подходов к её дефиниции. Первоначально термин «журналистика данных»⁸³ ещё в середине 2000-х при описании

⁷⁹ BBC's Shared Data Unit. Available at: <http://www.bbc.co.uk/lnp/sdu>

⁸⁰ Appelgren E. and Nygren G. 'Data Journalism in Sweden - Opportunities and Challenges'. In: Clare Mills, Michael Pidd and Esther Ward. *Proceedings of the Digital Humanities Congress 2012*. Studies in the Digital Humanities. Sheffield: The Digital Humanities Institute, 2014. Available at: <https://www.dhi.ac.uk/openbook/chapter/dhc2012-appelgren>

⁸¹ Thibodeaux T. (2011) 5 tips for getting started in data journalism // *Poynter*. Available at: <http://www.poynter.org/2011/5-tips-for-getting-started-in-data-journalism/147734/>

⁸² Minkoff M. (2010) Bringing data journalism into curricula. Available at: <http://michelleminkoff.com/2010/03/24/bringing-data-journalism-into-curricula/>

⁸³ Holovaty A. (2006) A fundamental way newspaper sites need to change. Available at: <http://www.holovaty.com/writing/fundamental-change/>

возможностей использования в журналистике больших данных. Сегодня ученые сходятся на том, что источником для подобных публикаций являются новые форматы цифровых данных, получившие распространение в конце 2000-х гг.

Некоторые авторы определений журналистики данных делают акцент на изменении рабочего процесса в редакции с появлением компьютерных технологий⁸⁴. В этих определениях подчеркивается преемственность журналистики данных по отношению к компьютерной журналистике, и роль больших данных как дополнительного источника в процессе создания публикации, работа с которым требует специальных навыков.

Другие специалисты отмечают, что журналистика данных – это особый процесс, прежде всего, создание публикации на основе анализа данных, с

⁸⁴ Appelgren E. and Nygren G. 'Data Journalism in Sweden - Opportunities and Challenges'. In: Clare Mills, Michael Pidd and Esther Ward. *Proceedings of the Digital Humanities Congress 2012*. Studies in the Digital Humanities. Sheffield: The Digital Humanities Institute, 2014. Available at: <https://www.dhi.ac.uk/openbook/chapter/dhc2012-appelgren>; Bounegru L. (2012) Data journalism in perspective. In: Gray, J., Bounegru, L. and Chambers, L. (eds) *The Data Journalism Handbook: How Journalists Can Use Data to Improve the News*. – Sebastopol, CA: O'Reilly, pp. 17–22; Davenport L., Fico F. and Detwiler M. (2000) Computer-assisted reporting in Michigan daily newspapers: More than a decade of adoption. In: *Proceedings of the national convention of the association for education in journalism and mass communication*, Phoenix, AZ, August 9–12; Diakopoulos N. (2011) A functional roadmap for innovation in computational journalism. Available at: <http://www.nickdiakopoulos.com/2011/04/22/a-functional-roadmap-for-innovation-in-computational-journalism/>; Felle T. (2016) Digital watchdogs? Data reporting and the news media's traditional 'fourth estate' function. *Journalism* 17(1): 85–96.; Flew T., Daniel A. and Spurgeon C. L. (2010) The promise of computational journalism. In: *Media, democracy and change: Refereed proceedings of the Australian and New Zealand Communication Association annual conference* (ed K McCallum), Canberra, ACT, Australia, July 7–9; Karlsen J. and Stavelin E. (2014) Computational journalism in Norwegian newsrooms. *Journalism Practice* 8(1): 34–48; Parasić S. and Dagiral E. (2013) Data-driven journalism and the public good: 'Computer-assisted reporters' and 'programmer-journalists' in Chicago. *New Media & Society* 15(6): 853–871; Tandoc E. C. Jr. and Oh S.-K. (2015) Small departures, big continuities? Norms, values, and routines in The Guardian's big data journalism. *Journalism Studies*. DOI: 10.1080/1461670X.2015.1104260; Uskali T. and Kuutti H. (2015) Models and streams of data journalism. *The Journal of Media Innovations* 2(1): 77–88; Weber W. and Rall H. (2012) Data visualization in online journalism and its implications for the production process. In: *Proceedings of the 16th international conference on information visualization (IV)*, July 11–13, pp. 349–356; Weinacht S. and Spiller R. (2014) Datenjournalismus in Deutschland: Eine explorative Untersuchung zu Rollenbildern von Datenjournalisten. *Publizistik* 59(4): 411–433.

необходимостью включения в материал различных форм интерактивной визуализации, например, карт или диаграмм⁸⁵.

Большинство авторов утверждают, что журналистика данных является одновременно и процессом и продуктом (публикацией контента, основанного на цифровых данных)⁸⁶.

Британский исследователь М. Найт предлагает широкую дефиницию журналистики данных, как истории, основанной на количественном первоисточнике и включающей в себя элементы анализа цифровых данных и

⁸⁵ Baack S. (2013) A new style of news reporting: Wikileaks and data-driven journalism. In: Rambatan B. and Johanssen J. (eds) *Cyborg Subjects: Discourses on Digital Culture*. CreateSpace Independent Publishing, pp. 113–122; Hullman J., Diakopoulos N., Momeni E., et al. (2015) Content, context, and critique: Commenting on a data visualization blog. In: *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, Vancouver, BC, Canada, March 14–18, pp. 1170–1175.

⁸⁶ Aitamurto T., Sirkkunen E. and Lehtonen P. (2011) Trends in data journalism. Available at: http://virtual.vtt.fi/virtual/nextmedia/Deliverables-2011/D3.2.1.2.B_Hyperlocal_Trends_In%20Data_Journalism.pdf; Appelgren E., Nygren G. (2014) Data journalism in Sweden – Opportunities and challenges: A case study of Brottspejl at Sveriges Television (SVT). In: Mills, C., Pidd, M. and Ward, E. (eds) *Proceedings of the Digital Humanities Congress 2012*. Available at: <https://www.hrionline.ac.uk/openbook/chapter/dhc2012-appelgren>; Ausserhofer J. (2015) ‘Die Methode liegt im Code’: Routinen und digitale Methoden im Datenjournalismus. In: Maireder A., Ausserhofer J., Schumann, C. and et al. (eds) *Digitale Methoden in der Kommunikationswissenschaft*. Berlin: Digital Communication Research, pp. 87–111; Coddington M. (2015) Clarifying journalism’s quantitative turn: A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digital Journalism* 3(3): 331–348; Cohen S., Hamilton J.T. and Turner F. (2011) Computational journalism: How computer scientists can empower journalists, democracy’s watchdogs, in the production of news in the public interest. *Communications of the ACM* 54(10): 66–71; Gynnild A. (2014) Journalism innovation leads to innovation journalism: The impact of computational exploration on changing mindsets. *Journalism* 15(6): 713–730; Hamilton J.T. and Turner F. (2009) Accountability through algorithm: Developing the field of computational journalism. In: *Proceedings of the center for advanced study in the behavioral sciences summer workshop*, Palo Alto, C.A. Available at: <http://web.stanford.edu/~fturner/Hamilton%20Turner%20Acc%20by%20Alg%20Final.pdf>; Hannaford L. (2015) Computational journalism in the UK newsroom: Hybrids or specialists? *Journalism Education* 4(1): 6–21; Howard A. B. (2014) *The Art and Science of Data-Driven Journalism*. New York: Tow Center for Digital Journalism. Available at: <http://towcenter.org/wp-content/uploads/2014/05/TowCenter-Data-Driven-Journalism.pdf>; Knight M. (2015) Data journalism in the UK: A preliminary analysis of form and content. *Journal of Media Practice* 16(1): 55–72; Loosen W., Reimer J. and Schmidt F. (2015) When data become news: A content analysis of data journalism pieces. In: *Proceedings of the future of journalism 2015 conference*. Cardiff, September 10–11; Stavelin E. (2013) *Computational journalism: When journalism meets programming*. PhD Dissertation, University of Bergen, Bergen; Tabary C., Provost, A.-M. and Trottier A. (2016) Data journalism’s actors, practices and skills: A case study from Quebec. *Journalism* 17(1): 66–84.

визуализацию⁸⁷. Сотрудник *The New York Times* А. Пилхофер считает, что журналистика данных – это некий зонтичный термин, который «охватывает многое, включая в себя все новый и новый набор инструментов, техник и подходов к рассказыванию историй, описанию событий, созданию сюжетов. (...) Объединяющая цель – журналистская: предоставление информации и анализа»⁸⁸.

Популяризатор журналистики данных и создатель первого тематического раздела, посвященного ей, в крупной газете (*The Guardian Datablog*) Саймон Роджерс считает, что журналистика данных является форматом журналистики, комбинирующим анализ данных, основанный на них новостной сюжет и их визуализацию⁸⁹. При этом, публикация может представлять собой по формату либо новостную историю в виде текста, либо визуализацию данных в виде графиков, таблиц или карт, иногда же достаточно просто публикации данных⁹⁰. Он также подчеркивает важность интерактивного формата подачи материала, ссылаясь на исследование «Института Рейтер» (*Reuters Institute for the Study of Journalism*), согласно которому половина аудитории сегодня потребляет медиаконтент на мобильных устройствах. Автор пишет о значении использования визуальных инструментов, которые способны привлекать аудиторию и рассказывать сложные истории, основанные на анализе данных⁹¹.

⁸⁷ Knight M. (2015) Data journalism in the UK: a preliminary analysis of form and content. *Journal of Media Practice* 16(1): 55-72.

⁸⁸ Грей Д. Пособие по журналистике данных. М.: РИА «Новости», 2013. С. 18.

⁸⁹ Rogers S. *Facts are Sacred: The Power of Data*. – London: Guardian books, 2011.

⁹⁰ Rogers S. (2011) Data journalism at the Guardian: what is it and how do we do it? // *The Guardian*. 28 Jul. 2011. Available at: <https://www.theguardian.com/news/Datablog/2011/jul/28/data-journalism>

⁹¹ Rogers S. (2017) What does data journalism look like today? // *Medium*. Available at: A 10-step guide <https://medium.com/data-journalism-awards/what-does-data-journalism-look-like-today-a-10-step-guide-6dd90c1f0c25>

Британский журналист П. Брэдшоу, автор основных принципов журналистики данных⁹², за основу которых взята градация принципов журналистики по Б. Ковачу и Т. Розенштилю⁹³, а также автор перевернутой пирамиды журналистики данных (*the inverted pyramid of data journalism*)⁹⁴, отмечает приоритетность визуальной составляющей в журналистике, определяя её как производство новостной графики, как журналистские практики с элементами дизайна и приоритетом интерактивности⁹⁵. Согласно П. Брэдшоу, визуализация и интерактивность – необходимые элементы презентации журналистом анализа данных аудитории⁹⁶. Значение визуализации данных подчеркивается и в последующих работах С. Баака⁹⁷, Д. Хулльмана, Н. Диакопоулоса, Е. Моменти и др⁹⁸.

В отличие от С. Роджерса, П. Брэдшоу, С. Баака, Д. Хулльмана, Н. Диакопоулоса и Е. Моменти, голландский исследователь Х. ван Эсс полагает, что визуализация не является обязательным элементом публикации журналистики данных⁹⁹. Ван Эсс также считает, что журналистика данных может быть основана на любом типе данных: главное, что они должны быть обработаны с помощью специальных инструментов.

⁹² Bradshaw P. (2017) 10 principles for data journalism in its second decade. *Online Journalism*. Available at <https://onlinejournalismblog.com/2017/08/07/10-principles-for-data-journalism-in-its-second-decade/#more-25052>

⁹³ Kovach B., Rosenstiel T. (2007) *The Elements of Journalism: What Newspeople Should Know and the Public Should Expect Completely Updated and Revised*. Three Rivers Press.

⁹⁴ Bradshaw P. (2011) The inverted pyramid of data journalism. Available at: <https://onlinejournalismblog.com/2011/07/07/the-inverted-pyramid-of-data-journalism/>

⁹⁵ Bradshaw P. (2010) How to be a data journalist // *The Guardian*. 10 Oct. 2010. Available at: <https://www.theguardian.com/news/Datablog/2010/oct/01/data-journalism-how-to-guide>

⁹⁶ Bradshaw P. (2017) The next wave of data journalism? // *Medium*. Available at: <https://medium.com/@paulbradshaw/the-next-wave-of-data-journalism-7e2e10087bb3>

⁹⁷ Baack S. (2013) A new style of news reporting: Wikileaks and data-driven journalism. In: Rambatan, B. and Johanssen, J. (eds) *Cyborg Subjects: Discourses on Digital Culture*. CreateSpace Independent Publishing, pp. 113–122.

⁹⁸ Hullman J., Diakopoulos N., Momeni E. et al. (2015) Content, context, and critique: Commenting on a data visualization blog. In: *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, Vancouver, BC, Canada, March 14–18, pp. 1170–1175.

⁹⁹ Van Ess H., van der Kaa H. (2012) *Handboek Datajournalistiek*. Available at: <https://www.boomhogeronderwijs.nl/kernproduct/141/Handboek-datajournalistiek>

Как полезный для общественности процесс анализа, фильтрации и визуализации данных, определяет журналистику данных немецкий специалист М. Лоренц, одним из первых описавший¹⁰⁰ процесс создания публикаций журналистики данных в 2010 году.

Попытки определения журналистики данных, предпринятые в середине 2010-х гг. отечественными исследователями, мало отличаются от зарубежных и не выявляют российской специфики. Так, например, журналистику данных определяют как направление в журналистике, в основе которого лежит обработка данных и последующее представление полученной из них информации в виде текстовой истории или инфографики¹⁰¹, или как «набор специфических навыков для поиска, анализа, визуализации информации из цифровых источников данных для формирования интерактивных форматов уникальной подачи авторского аналитического журналистского контента и эффективного взаимодействия СМИ и журналиста с аудиторией»¹⁰².

В большинстве определений журналистики данных присутствуют объединяющие их элементы, которые раскрывают базовые компоненты данного явления: это нахождение информационного повода путем анализа больших наборов цифровых данных в основном количественными, вычислительными методами, а также презентация результатов анализа в форме визуализации данных (часто интерактивной).

В рамках данной работы мы используем в качестве рабочего следующее обобщающее и уточняющее определение:

журналистика данных – направление в журналистике, основой материалов которого являются результаты обработки количественных данных, которые могут служить инфоповодами или быть частью материала,

¹⁰⁰ Lorenz M. (2010) Data Driven Journalism: What Is There to Learn? Paper presented at the IJ-7 Innovation Journalism Conference. Standford: CA, June 2010.

¹⁰¹ Бегтин И. (АНО «Информационная Культура»). Режим доступа: <https://www.infoculture.ru/glossary/data-journalistic/>

¹⁰² Шилина М. Г. Big Data, Open Data, Linked Data, метаданные в PR: актуальные модели трансформации теории и практики // Медиаскоп. 2014. Вып. 1.

обогащая его. При этом особое значение при презентации материала аудитории приобретает визуализация результатов анализа выбранных данных.

Основная особенность публикаций журналистики данных состоит в том, что сегодня они создаются в междисциплинарной области, на границе между гуманитарными, социальными и техническими науками, а журналистская деятельность приобретает все более интегрированный характер. Изменяется практика работы специалистов в области медиа, происходит усложнение рабочего процесса, журналистские проекты все чаще представляют собой многоуровневый конвергентный материал. Востребованным становится использование вычислительных технологий для извлечения, анализа и визуализации данных¹⁰³, применение навыков программирования для создания новостных приложений, интерактивных мультимедийных историй, инструментов для изучения данных¹⁰⁴. Журналисты все чаще объединяются или даже сами частично становятся специалистами из других областей – программистами, аналитиками, графическими дизайнерами.

Материалы СМИ, основанные на новых форматах цифровых данных, требуют особых навыков реализации, подготовки пишущих журналистов и присутствия специалистов технических профессий в редакции. Трансформируется рабочий процесс, усложняется коммуникация, происходит увеличение числа этапов создания публикации. Схема создания материала включает получение данных, их фильтрацию и анализ, визуализацию, построение истории¹⁰⁵. Структурно технология разработки текстового и визуального материала состоит из следующих этапов: во-

¹⁰³ Appelgren E. and Nygren G. (2014) Data journalism in Sweden – Opportunities and challenges: A case study of Brottspejl at Sveriges Television (SVT). In: Mills C, Pidd M and Ward E (eds) Proceedings of the Digital Humanities Congress 2012. Sheffield: HRI Online Publications. Available at: <https://www.hrionline.ac.uk/openbook/chapter/dhc2012-appelgren>

¹⁰⁴ Royal C. (2010) The journalist as programmer: A case study of the New York Times interactive news technology department. Paper presented at the Anais do International Symposium in Online Journalism, The University of Texas at Austin, TX, USA.

¹⁰⁵ Lorenz M. (2013) Information Architect. Available at: <http://www.mirkolorenz.com/?id=6>

первых, выбор цели и темы публикации, описание целевой аудитории, источника данных, что соответствует классической схеме. Этапы, характерные для журналистики данных: извлечение данных и их фильтрация с помощью математических и статистических методов; графический дизайн, визуализация данных; создание интерактивного облика публикации программными методами; тестирование и оценка эффективности публикации. Каждый из перечисленных этапов обладает собственной спецификой, что отличает схему создания материала журналистики данных от схемы создания любого другого материала современной журналистики.

Так, работа с данными при создании проекта и его визуализации, согласно Б. Фраю, состоит из нескольких этапов¹⁰⁶. На первом этапе происходит выбор источников данных, их получение и анализ. Основанием анализа данных на данном этапе является информатика и её инструментарий. Второй этап: извлечение нужных данных и их фильтрация. На этом этапе журналист использует математические и статистические методы работы с данными. На третьем этапе к работе над проектом подключается графический дизайнер, он представляет данные в виде визуализации, очищает их и представляет читателям. Завершающий этап – создание интерактивного дизайна программистом, результатом которого является готовый проект. Правильное взаимодействие с аудиторией, понятное и простое представление больших массивов данных для читателей – основные принципы создания проекта на основе использования новых форматов данных.

Таким образом, при создании публикации журналистики данных в работу журналистов интегрируются новые элементы. Цифровые данные становятся основой для создания материалов СМИ, или же их частью, сопровождающим материалом, обогащающим содержание.

¹⁰⁶ Fry B. Visualizing Data: Exploring and Explaining Data with the Processing Environment. – London: O'Reilly Media; 1 edition, 2008.

У каждой медиаорганизации своя специфика создания публикаций журналистики данных, крупные СМИ нанимают диверсифицированные команды, функционирующие на постоянной основе, в небольших редакциях проекты могут создаваться энтузиастами. Не существует общей схемы построения команды по созданию журналистики данных, понимания того, какие именно специалисты должны в нее входить и каково должно быть их число. В среде практиков и в академическом сообществе остается нерешенным вопрос о том, кого считать автором материалов журналистики данных, кто именно создает проекты в редакциях, чей вклад наиболее значим: только ли журналистов, или же и графических дизайнеров, аналитиков данных и программистов, появляется ли новое поколение программистов-журналистов, должны ли все журналисты обладать навыками кодирования. Идут споры о том, какую терминологию применять по отношению к подобным специалистам – программист-журналист, хакер-журналист, журналист-программист¹⁰⁷, и даже «единорог» в области медиа¹⁰⁸. Нет определенности, каким именно навыкам программирования должны обладать журналисты, каким учиться¹⁰⁹ из-за неспособности точно сформулировать, что означает «кодирование» в контексте журналистики (знание языков программирования или же компьютерная грамотность). До сих пор мало изучено, как фактически используются навыки программирования журналистами в редакциях и используются ли вообще.

Одно из крупных исследований журналистики данных¹¹⁰ фокусируется на комплексном описании характеристик, состояния и перспектив данного явления. Авторы провели 56 глубинных интервью с журналистами в США,

¹⁰⁷ Pilhofer A. (2010) Programmer-Journalist? Hacker-Journalist? Our Identity Crisis. Media Shift – Idea Lab: Available at: <http://www.pbs.org/idealab/2010/04/programmer-journalist-hacker-journalist-our-identity-crisis107>

¹⁰⁸ Hermida A. and Young M. L. (2017) Finding the Data Unicorn. *Digital Journalism*, 5(2): 159-176.

¹⁰⁹ Where to learn about data journalism. Available at: <http://pudo.org/blog/2013/11/13/ddj-resources.html> <https://noahveltman.com/>

¹¹⁰ Rogers S., Schwabish J., Bowers D. (2017) The state of data journalism in 2017 // *Google blog*. Available at: <https://www.blog.google/topics/journalism-news/data-journalism-2017/> <https://newslab.withgoogle.com/assets/docs/data-journalism-in-2017.pdf>

Великобритании, Германии и Франции, а также онлайн-опрос 900 журналистов, специализирующихся на создании публикаций журналистики данных по всему миру. Согласно результатам исследования, уже более 50% всех новостных организаций в США и Европе имели в штате специалиста по журналистике данных; свыше 40% журналистов регулярно используют данные для создания публикаций (более двух раз в неделю); почти 50% публикаций создаются за день или меньше; 20% журналистов используют собственные инструменты очистки и визуализации данных.

1.2. Основные направления исследований журналистики данных¹¹¹

Несмотря на большое количество академических работ, посвященных журналистике данных, можно констатировать, что к настоящему времени не сформирован универсальный концептуальный и методологический аппарат, а методически исследовательские подходы больше ориентируются на качественные, поисковые инструменты.

Среди наиболее известных – работа С. Коэн, Д. Т. Гамильтона и Ф. Тёрнера, в которой авторы утверждают, что для развития медиаиндустрии и преодоления ей финансового и технологического давления журналистам необходимо работать сообща с техническими специалистами, способными предлагать иные возможные алгоритмы создания контента, к примеру, путем извлечения данных¹¹². Исследование работы зарубежных редакций предприняли Т. Флю, К. Спурджион, А. Дэниэль, А. Свифт, утверждающие, что применение методов вычислительной журналистики приводит к появлению основ для оригинальных материалов расследовательской журналистики и расширяет возможности форм взаимодействия с аудиторией

¹¹¹ Параграф написан с использованием материалов и фрагментов из ранее опубликованной статьи: Шилина А. Г. Журналистика данных в зарубежной качественной прессе (кейс специализированных ресурсов газет *The Guardian* и *The New York Times*) // Вестник Московского университета. Серия 10. Журналистика. 2019. № 5. С. 135-162.

¹¹² Cohen S., Hamilton J. T., Turner F. (2011) Computational journalism. *Communications of the ACM*, 54(10), 66-71.

в Интернете¹¹³. В статье Н. Диакопулоса об инновациях в компьютерной журналистике автор, опираясь на теоретическое обобщение пятидесяти пяти концепций, методику системного анализа и ценностно-ориентированного проектирования, производит градацию аспектов влияния инноваций в журналистике и выявляет недооцененные области их применения¹¹⁴. Другие работы – исследование практик отдела информационных технологий *The New York Times* К. Рояла, работа по социологии вычислительной и алгоритмической журналистики К. В. Андерсона, работа С. Баака о датафикации в журналистике и о значении концепции открытых данных для возможностей создания медиаконтента¹¹⁵.

Отдельной группой стоят работы по эпистемологии журналистики данных: статьи С. Параси, С. К. Льюиса и О. Уэстлунда посвящённые изучению влияния феномена больших данных на современную журналистику; исследование влияния инноваций на работу журналистов А. Гюннильда и анализ последствий появления вычислительной журналистики и журналистики данных на работу в редакциях М. Коддингтона¹¹⁶.

¹¹³ Flew T., Spurgeon C., Daniel A., Swift A. (2012) The promise of computational journalism. *Journalism Practice*, 6(2), 157-171.

¹¹⁴ Diakopoulos N. (2012) Cultivating the Landscape of Innovation in Computational Journalism. Tow-Knight Center for Entrepreneurial Journalism; Diakopoulos N. (2015) Algorithmic Accountability. Journalistic investigation of computational power structures. *Digital Journalism* 3(3): 398-415.

¹¹⁵ Royal C. (2010) The journalist as programmer: A case study of the New York Times interactive news technology department. Paper presented at the Anais do International Symposium in Online Journalism, The University of Texas at Austin, TX, USA; Anderson C. W. (2013) Towards a sociology of computational and algorithmic journalism. *New Media & Society*, 15(7): 1005-1021; Baack S. (2015) Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism". *Big Data & Society* 2(2): 6.

¹¹⁶ Parasi S. (2015) Data-driven revelation? Epistemological tensions in investigative journalism in the age of 'big data'. *Digital Journalism*. 3(3): 364-380; Lewis S. C. and Westlund O. (2015) Big Data and Journalism: Epistemology, expertise, economics, and ethics. *Digital Journalism*. 3(3): Journalism in an Era of Big Data: Cases, Concepts, and Critiques; Gynnild A. (2014) Journalism Innovation Leads to Innovation Journalism: The Impact of Computational Exploration on Changing Mindsets. *Journalism* 15 (6): 713–730; Coddington M. (2015) Clarifying Journalism's Quantitative Turn: A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digital Journalism*. 3(3): 331-348.

Также необходимо выделить работы исследователей, посвящённые национальным различиям и особенностям журналистики данных разных стран, специфике как источников, так и работы журналистов в США¹¹⁷, Великобритании¹¹⁸, Канаде¹¹⁹, Германии¹²⁰, Италии¹²¹, Бельгии¹²², Швеции¹²³,

¹¹⁷ Fink K. and Anderson C. (2014) Data Journalism in the United States: Beyond the “usual suspects” *Journalism Studies*, 1-15; Parasio S. and Dagiral E. (2012). Data-driven journalism and the public good: “Computer-assisted-reporters” and “programmer-journalists” in Chicago. *New Media & Society*; Powers M. (2012) In Forms That Are Familiar and Yet-to-Be Invented: American Journalism and the Discourse of Technologically Specific Work. *Journal of Communication Inquiry* 36(1): 24–43; Weber W. and Rall H. (2013) We are journalists. Production Practices, Attitudes and a Case Study of the New York Times Newsroom. *Interaktive Infografiken*: 161-172; Uskali T. and Kuutti H. (2015) Models and Streams of Data Journalism. *The Journal of Media Innovations* 2(1): 77–88.

¹¹⁸ Borges-Rey E. (2016) Unravelling data journalism: A study of data journalism practice in British newsrooms. *Journalism Practice* 10(7): 833–843; Borges-Rey E. (2017) Towards an epistemology of data journalism in the devolved nations of the United Kingdom: Changes and continuities in materiality, performativity and reflexivity. *Journalism*. DOI: 10.1177/1464884917693864; Hannaford L. (2015) Computational Journalism in the UK newsroom: hybrids or specialists? *Journalism Education* 4(1): 6-21; Knight M. (2015) Data journalism in the UK: a preliminary analysis of form and content. *Journal of Media Practice* 16(1): 55-72; Rogers S. (2011) Data journalism at the Guardian: what is it and how do we do it? // *Guardian.co.uk* Available at: <https://www.theguardian.com/news/Datablog/2011/jul/28/data-journalism>; Mair J., Keeble R. L., Lucero M. Data Journalism: Past, Present and Future. – UK: Abramis, 2017.

¹¹⁹ Hermida A., Young M. L. and Fulda J. (2018) What Makes for Great Data Journalism? *Journalism Practice* 12(1): 115-135; Tabary C., Provost A. M. and Trottier A. (2016) Data journalism’s actors, practices and skills: A case study from Quebec. *Journalism: Theory, Practice, and Criticism* 17(1): 66–84.

¹²⁰ Weinacht S. and Spiller R. (2014): Datenjournalismus in Deutschland. Eine explorative Untersuchung zu Rollenbildern von Datenjournalisten [Data-journalism in Germany. An exploratory study on the role conceptions of data-journalists]. *Publizistik* 59(4): 411–433.

¹²¹ Porlezza C. (2016) Data Journalism in Italy. Between closed data, networked journalism and ethical conflicts. IAMCR Conference – Leicester, UK. Available at: https://prezi.com/p_gvprilp_uo/accountability-and-transparency-in-data-journalism-the-case-of-italy/; Porlezza C. and Splendore S. (2019) From Open Journalism to Closed Data: Data Journalism in Italy. *Digital Journalism*. DOI: 10.1080/21670811.2019.1657778

¹²² De Maeyer J., Libert M., Domingo D., Heinderyckx F., Le Cam F. (2015) Waiting for Data Journalism. A qualitative assessment of the anecdotal take-up of data journalism in French-speaking Belgium. *Digital Journalism* 3(3): 432-446.

¹²³ Appelgren E. and Nygren G. (2014) Data Journalism in Sweden: Introducing New Methods and Genres of Journalism into ‘Old’ Organizations. *Digital Journalism*, 2(3): 394–405; Appelgren E. (2017) An illusion of interactivity: The paternalistic side of data journalism. *Journalism Practice*. DOI: 10.1080/17512786.2017.1299032; Appelgren E. (2019) Remaining in Control with an Illusion of Interactivity: The Paternalistic Side of Data Journalism. *Journalism Practice*, 13(8): 956-960.

Финляндии¹²⁴, Норвегии¹²⁵, Китае¹²⁶, Австралии¹²⁷, арабских¹²⁸ и африканских¹²⁹ государствах использованию новых форматов данных в российской журналистике¹³⁰.

Присутствуют в академическом дискурсе журналистики данных и тексты, посвященные рассмотрению подходов к обучению журналистов программированию, статистике, базовым инструментам, необходимым при создании публикаций журналистики данных: исследование Д. Хьюетта о возможностях и ограничениях в обучении преподаванию журналистики данных¹³¹; изучение преподавания политической расследовательской журналистики на примерах публикаций журналистики данных К. Грэхама¹³²; исследование стратегий образования в журналистике данных на основе сравнительного исследования методик европейских стран¹³³, исследование готовности студентов Гонконга к созданию материалов журналистики

¹²⁴ Sirkkunen E., Aitamurto T. and Lehtonen P. (2011) Trends in Data Journalism. Finnish Strategic Centre for Science, Technology and Innovation in the field of ICT.

¹²⁵ Karlsen J. and Stavelin E. (2014) Computational Journalism in Norwegian Newsrooms. *Journalism Practice* 8 (1): 34–48.

¹²⁶ Zhang S. and Feng J. (2019) A Step Forward? Exploring the diffusion of data journalism as journalistic innovations in China. *Journalism Studies*, 20(9): 1281-1300.

¹²⁷ Wright S. and Doyle K. (2019) The Evolution of Data Journalism: A Case Study of Australia. *Journalism Studies*, 20(13): 1811-1827.

¹²⁸ Lewis N. P. and Al Nashmi E. (2019) Data Journalism in the Arab Region: Role Conflict Exposed. *Digital Journalism*, DOI: 10.1080/21670811.2019.1617041

¹²⁹ Cheruiyot D., Baack S. and Ferrer-Conill R. (2019) Data Journalism Beyond Legacy Media: The case of African and European Civic Technology Organizations. *Digital Journalism*, DOI: 10.1080/21670811.2019.1591166

¹³⁰ Valeeva A. (2017) Open Data in a Closed Political System: Open Data Investigative Journalism in Russia. Reuters Institute for the Study of Journalism University of Oxford. Available at: http://reutersinstitute.politics.ox.ac.uk/sites/default/files/2017-10/Valeeva_Open%20data%20and%20investigative%20journalism%20in%20Russia.pdf;

Вартанов С. А. Большие данные в онлайн-СМИ: подходы и стратегии использования // Медиаскоп. 2017. Вып. 4; Шилина М. Г. Big Data, Open Data, Linked Data, метаданные в PR: актуальные модели трансформации теории и практики // Медиаскоп 2014. Вып. 1.

¹³¹ Hewett J. (2015) Learning to teach data journalism: Innovation, influence and constraints. *Journalism*, 17(1): 119-137.

¹³² Graham C. (2015) By the Numbers. Data Journalism Projects as a Means of Teaching Political Investigative Reporting. *Asia Pacific Media Educator*, 25(2): 247-261.

¹³³ Splendore S., Di Salvo P., Eberwein T., Groenhardt H., Kus M., Porlezza C. (2015) Educational strategies in data journalism: A comparative study of six European countries. *Journalism* 17(1): 138-152.

данных¹³⁴; статьи раскрывающие значение обучения журналистов статистическим методам¹³⁵ и навыкам анализа данных в эпоху цифровых технологий¹³⁶.

Существуют работы, подчеркивающие роль журналистики данных в обществе, её значение как одного из важнейших инструментов открытости власти и демократии: научная статья К. В. Андерсона в которой автор формулирует исторические и теоретические основы для концептуализации феномена журналистики данных, рассматривая его в контексте «количественного поворота» в журналистской практике¹³⁷; статья Н. С. Коэн, посвящённая трансформации работы журналиста в эпоху цифровых технологий в историческом аспекте в рамках стратегии снижения затрат на рабочую силу¹³⁸; рассмотрение журналистики данных в рамках этики Й. Фэрфилдом и Х. Штайн¹³⁹ и «экологии публичной информации» М. Шудсоном¹⁴⁰; концептуализация предполагаемого будущего журналистики С. Бааком и открытости политики М. Бунзом на примере журналистики данных и публикаций *Wikileaks*¹⁴¹.

¹³⁴ Fan Yang, Ying Roselyn Du (2016) Storytelling in the Age of Big Data: Hong Kong Students' Readiness and Attitude towards Data Journalism. *Asia Pacific Media Educator*, 26(2): 148-162.

¹³⁵ Nguyen A. and Lugo-Ocando J. (2015) The state of data and statistics in journalism and journalism education: Issues and debates. *Journalism* 17(1): 3-17.

¹³⁶ Yarnall L., Johnson J. T., Rinne L. and Ranney M. A. (2008) How Post-secondary Journalism Educators Teach Advanced CAR Data Analysis Skills in the Digital Age. *Journalism & Mass Communication Educator*, 63(2): 146-164.

¹³⁷ Anderson C. W. (2015) Between the unique and the pattern: Historical tensions in our understanding of quantitative journalism. *Digital Journalism* 3(3): 349-363.

¹³⁸ Cohen N. S. (2015) From pink slips to pink slime: Transforming media labor in a digital age. *The Communication Review*, 18(2): 98-122.

¹³⁹ Fairfield J. and Shtein H. (2014) Big data, big problems: Emerging issues in the ethics of data science and journalism. *Journal of Mass Media Ethics*, 29(1): 38-51.

¹⁴⁰ Schudson M. (2010) Political observatories, databases & news in the emerging ecology of public information. *Daedalus* 139(2): 100-109.

¹⁴¹ Baack S. (2013) A new style of news reporting: Wikileaks and data-driven journalism. In: Rambatan, B. and Johanssen, J. (eds) *Cyborg Subjects: Discourses on Digital Culture*. CreateSpace Independent Publishing: 113-122; Bunz M. (2011) Das offene Geheimnis: Zur Politik der Wahrheit im Datenjournalismus. In: Geiselberger, H. (ed.), *Wikileaks und die Folgen: Netz - Medien - Politik*. Berlin: Suhrkamp: 134-151.

Большинство авторов существующих работ, рассматривающих зарубежную журналистику данных, опирается на качественные методы исследования данного явления: наблюдение за работой специалистов в редакции и интервьюирование.

В целом, в последние годы формируются два теоретических подхода к исследованиям компьютеризации и масштаба проникновения информационных технологий в редакции. Приверженцы первого – технологического детерминизма – подчеркивают, что журналистика всегда формировалась технологией¹⁴², её форма и содержание в решающей степени определяются доступным способом производства и распространения новостей¹⁴³. Другие исследователи этой группы более осторожны, отмечая, что при наличии сильного детерминирующего компонента сама технология не является единственным определяющим фактором изменений в журналистике¹⁴⁴.

Сторонники противоположного подхода предлагают отказаться от идеи технологического детерминизма как упрощенного объяснения изменений в редакционных практиках¹⁴⁵. В последующих научных исследованиях внимание ученых фокусируется на понимании редакций как сложных социальных конструкций, динамика в которых способна оказывать влияние на технологические разработки¹⁴⁶.

Одно из наиболее полных эмпирических исследований того, как новые технологии принимаются в редакциях – исследование П. Бочковски трех

¹⁴² Pavlik V. (2000) The impact of technology on journalism. *Journalism Studies* 1(2): 229–237.

¹⁴³ McNair B. The sociology of journalism. – UK: Arnold London, 1998.

¹⁴⁴ Bardoel J. and Deuze M. (2001) Network Journalism: converging competences of old and new media professionals. *Australian Journalism Review* 23(2): 91-103.

¹⁴⁵ Cottle S. and Ashton M. (1999) From BBC newsroom to BBC newscentre: On changing technology and journalist practices. *Convergence* 5(3): 22-43; Örnebring H. (2010). Technology and journalism-as-labour: Historical perspectives. *Journalism* 11(1): 57-74.

¹⁴⁶ Boczkowski P. J. (2004) The processes of adopting multimedia and interactivity in three online newsrooms. *Journal of Communication* 54(2): 197–213; Schudson M. (1989) The sociology of news production. *Media, culture and society* 11(3): 263-282; Anderson C. W. (2013) Towards a sociology of computational and algorithmic journalism. *New Media & Society*, 15(7): 1005-1021.

американских медиа. При решительном отказе от детерминистского подхода автор фокусируется на процессе принятия решений, утверждая, что применение новых мультимедийных и интерактивных технологий связано с различиями в организационных структурах и методах работы¹⁴⁷. Следуя традиции отказа от детерминизма, А. Вайс и Д. Доминго подчеркивают, что именно сетевые отношения и социальные иерархии в редакциях влияют на принятие инноваций¹⁴⁸. Подобные подходы, ориентированные на изучение социальных связей в коллективе, являются основой для понимания того, как журналистика данных включается в редакции и влияет на разграничение рабочих ролей, а также для сопоставления работы редакций разных масштабов и стран¹⁴⁹.

Одно из первых эмпирических исследований, посвящённых работе команд, создающих материалы журналистики данных в редакции – исследование американскими учеными С. Параси и Е. Дагралем влияния «программистов-журналистов» на работу команды над новостными проектами журналистики данных, основанное на интервью и контент-анализе. Оно выявляет, что включение элементов кодирования в работу редакций меняет эпистемологию журналистики благодаря применению методов, основанных на CAR, вниманию к детализации данных, а не статистическому анализу и сторителлингу, характерному для традиционной журналистской модели¹⁵⁰. Другой пример – этнографическое исследование отдела информационных технологий редакции *The New York Times*, объединяющее включенное наблюдение и интервью с ключевыми работниками. Его результаты показывают, что большинство респондентов

¹⁴⁷ Boczkowski P. J. (2004) The processes of adopting multimedia and interactivity in three online newsrooms. *Journal of Communication* 54(2): 197–213.

¹⁴⁸ Weiss A. and Domingo D. (2010) Innovations processes in online newsrooms as actor-networks and community of practice. *New Media & Society* 12(7): 1156–1171.

¹⁴⁹ Fink K. and Anderson C. (2014) Data Journalism in the United States: Beyond the “usual suspects”, *Journalism Studies*: 1-15.

¹⁵⁰ Parasie S. and Dagiral E. (2012) Data-driven journalism and the public good: “Computer-assisted-reporters” and “programmer-journalists” in Chicago. *New Media & Society* 15(6): 853-871.

являются энтузиастами-самоучками, обладают наборами навыков – кодированием и созданием журналистского контента, их комбинирование определяет гибридную рабочую модель¹⁵¹.

В. Уэбер и Х. Ролл подтверждают гипотезу, о том, что для американской журналистики данных характерен «целостный подход». Их масштабное исследование работы журналистов, дизайнеров и разработчиков, участвующих в производстве текстов журналистики данных в восьми медиа Германии, Швейцарии и США выявляет резкий контраст между европейскими и американскими практиками¹⁵². В отличие от США, в редакциях Германии и Швейцарии присутствует четкое разделение функций журналистов, отвечающих за идейное содержание проекта, и дизайнеров и программистов, разрабатывающих визуальный и интерактивный элементы.

Другие исследования работы над проектами журналистики данных в европейских редакциях – исследование норвежских редакций¹⁵³, материалы, представленные в пособии по журналистике данных¹⁵⁴, исследование практик двух новостных организаций Великобритании – *BBC* и *Financial Times*, в котором подчеркивается отсутствие гибридной модели работы, характерной для медиапрофессионалов США¹⁵⁵.

Об «иерархии гибридности» заявляют и А. Хермида и Л. Янг, употребляя данный термин по отношению к различиям команд, создающих

¹⁵¹ Royal C. (2010) The journalist as programmer: A case study of the New York Times interactive news technology department. Paper presented at the Anais do International Symposium in Online Journalism, The University of Texas at Austin, TX, USA.

¹⁵² Weber W. and Rall H. (2013) We are journalists. Production Practices, Attitudes and a Case Study of the New York Times Newsroom. *Interaktive Infografiken*: 161-172.

¹⁵³ Karlsen J. and Stavelin E. (2014) Computational Journalism in Norwegian Newsrooms. *Journalism Practice* 8 (1): 34–48.

¹⁵⁴ Грей Д. Пособие по журналистике данных. М.: РИА «Новости», 2013. Режим доступа: https://ria.ru/files/book/_site/%D0%B2%D0%B2%D0%B5%D0%B4%D0%B5%D0%BD%D0%B8%D0%B5_4.html

¹⁵⁵ Hannaford L. (2015) Computational Journalism in the UK newsroom: hybrids or specialists? *Journalism Education*. 4(1): 6-21.

материалы журналистики данных в новостных организациях Канады¹⁵⁶. Ученые отмечают, что канадские журналисты, так же как американские коллеги, считают себя многопрофильными работниками или «единорогами» – специалистами, способными писать статью, самостоятельно анализировать данные и кодировать. В медиасреде страны возникает четкое различие между редакциями, поскольку некоторые из них уже имеют разработанную культуру создания публикаций, которая объединяет технологическую и журналистскую компоненты, в то время как другие все еще пытаются набирать диверсифицированные команды, обладающие необходимыми навыками. «Интерактивные» или «научные» группы журналистов, создающих материалы журналистики данных, могут внедрять свои технологические знания в редакцию, способствуя культурному сдвигу. Так, на основании 17 интервью с журналистами и внештатными сотрудниками газет и вещательных компаний, ученые анализируют, как возникает профессиональная идентичность журналиста, участвующего в создании материалов журналистики данных, и как это влияет на рабочий процесс, сотрудничество внутри и вне медиа, трансформируя сложные профессиональные границы. При этом, согласно исследованию, подобные специалисты демонстрируют склонность к неформальным формам сотрудничества и привлечению ресурсов из разных источников, как внутри, так и за пределами редакции, их коммуникация характеризуется сильными над-медийными связями (в том числе с программистами, учеными в области науки о данных и др.). Подобное внешнее сотрудничество проявляется и в востребованности «внештатных ретрансляторов данных» (англ. *freelance data wrangler*), работающих с различными медиаорганизациями. Похожая ситуация наблюдается также в итальянских редакциях, где доминирующей

¹⁵⁶ Hermida A., Young M. L. and Fulda J. (2018) What Makes for Great Data Journalism? *Journalism Practice* 12(1): 115-135.

категорией является фрилансер, специализирующийся на создании публикаций журналистики данных¹⁵⁷.

Другое качественное исследование институционализации журналистики данных в британских редакциях, проведенное в последние годы Э. Борхес-Реем, основано на анализе 24 интервью со специалистами по созданию публикаций журналистики данных, редакторами, программистом и дизайнером¹⁵⁸. Полученные данные показывают, что опрошенные журналисты считают основополагающей задачей раскрытие динамики политических сил и корпораций и вносят вклад в работу редакции лишь в качестве специалистов по обработке информации из баз данных, остальные же задачи чаще решаются специалистами в области технологий или науки о данных.

В отличие от А. Гермиды и Л. Янг, Э. Борхес-Рэй рисует более полную картину британской журналистики данных, при появлении которой в редакции помимо нарушения традиционной логики рабочего процесса также происходит изменение в подаче контента – переход от линейного повествования к интерактивному, вовлекающему, ориентированному на аудиторию просьюмеров. В этом отношении практика журналистики данных укореняется в разных редакциях и диверсифицируется на три основные формы: а) краткая ежедневная форма (англ. *daily quick / brief form*) журналистики данных; б) крупная / развернутая расследовательская форма (англ. *extensive investigative form*) в) геймифицированная (англ. *light entertaining gamified form*) журналистика данных, ориентированная на развлечение аудитории. Однако подобная классификация ограничена, так как в опросе Э. Борхес-Рея принимали участие специалисты только из ведущих центров журналистики данных, таких как *The Guardian* или *BBC* в то время

¹⁵⁷ Porlezza C. (2016) Data Journalism in Italy. Between closed data, networked journalism and ethical conflicts. IAMCR Conference. Leicester, UK. Available at: https://prezi.com/p_gvprilp_uo/accountability-and-transparency-in-data-journalism-the-case-of-italy/

¹⁵⁸ Borges-Rey E. (2016) Unravelling data journalism: A study of data journalism practice in British newsrooms. *Journalism Practice* 10(7): 833–843.

как интеграция журналистики данных в региональных и местных медиа пока не изучена. Значимые исследования работы специалистов, создающих публикации журналистики данных в российских редакциях также до сих пор отсутствуют.

Исследования зарубежной журналистики данных, опирающиеся на количественные методы и рассматривающие сам контент и форму публикаций малочислены. Пример подобного исследования, в терминологии Э. Борхес-Рэя – материалов крупной / развернутой расследовательской формы журналистики данных¹⁵⁹ – работа В. Лузен, Д. Реймер и Ф. Де Сильва-Шмидт, в которой авторы анализируют 225 проектов журналистики данных, номинированных на *Data Journalism Awards* за 2013-2016 гг.¹⁶⁰. В работе делается вывод о том, что создание подобных крупных проектов требует наличия в редакции диверсифицированных команд (над 192 проектами из 225 работают более пяти человек), больших трудозатрат (более трети проектов создается совместно с внешними партнерами, которые дополняют анализ или помогают с визуализацией данных). Согласно исследованию, чаще всего на премию номинируются материалы качественной прессы, они же чаще всего получают награды: на газеты приходится 43,1% номинантов, среди лауреатов – 37,8% газетных публикаций¹⁶¹. При этом число публикаций-номинантов премии увеличивается с каждым годом (в 2016 г. 12 публикаций-победителей выбирается из 471 номинанта, в 2017 г. – из 573). Тематика публикаций

¹⁵⁹ Borges-Rey E. (2016) Unravelling data journalism: A study of data journalism practice in British newsrooms. *Journalism Practice* 10(7): 833–843.

¹⁶⁰ Loosen W., Reimer J. and De Silva-Schmidt F. (2017) Data-driven reporting: An on-going (r)evolution? An analysis of projects nominated for the Data Journalism Awards 2013–2016. *Journalism*. DOI: 10.1177/1464884917735691

¹⁶¹ Остальные номинанты – организации, связанные с расследовательской журналистикой (например, *ProPublica*) - 18,2%, журналы и онлайн-СМИ - по 8,4%, вещательные СМИ – 5,3%, новостные агентства – 4,4%, нежурналистские организации – 4%, университетские СМИ – 3,1%, другие типы авторов – 2,7%. (Loosen W., Reimer J., De Silva-Schmidt F. (2017) Data-driven reporting: An on-going (r)evolution? An analysis of projects nominated for the Data Journalism Awards 2013–2016. *Journalism*. DOI: 10.1177/1464884917735691 p.7).

стабильна: половина проектов касается политики, треть — преступности и социальной проблематики, менее 5% приходится на культуру, спорт и образование. Как отмечают исследователи – один из заметных трендов журналистики данных — рост критики в отношении общественных институтов и публичных лиц (в 2013 г. критика текущей ситуации содержалась в 46,4% проектов, в 2016 г. — в 63 %). Визуализация за три года также не становится сложнее, чаще всего используется статическая визуализация, интерактивные проекты остаются исключениями. Динамика в выборке за три года заметна лишь в отношении источников данных публикаций-лауреатов: сегодня журналисты чаще опираются на собственные данные и утечки, реже – на открытые базы данных.

Попытка анализа и классификации краткой, ежедневной формы журналистики данных (в терминологии Э. Борхес-Рэя¹⁶²) предпринята в исследовании Ф. Штальфа¹⁶³. Автор опирается на метод количественного контент-анализа публикаций журналистики данных европейских качественных изданий *The Guardian*, *Zeit Online*, *Spiegel Online*, и *Neue Zürcher Zeitung*, направленного на определение универсальных характеристик ежедневной формы журналистики данных. Результаты исследования показывают, что журналисты полагаются преимущественно на предварительно обработанные данные, полученные от национальных организаций и правительственных органов, визуализируют их простыми графиками, тематика публикаций – политические темы, параллельные традиционным новостям. Таким образом, исследование Ф. Штальфа подчеркивает различие между сложной и ежедневной формами журналистики данных.

¹⁶² Borges-Rey E. (2016) Unravelling data journalism: A study of data journalism practice in British newsrooms. *Journalism Practice*, 10(7): 833–843.

¹⁶³ Stalphy, F. (2017) Classifying Data Journalism. A content analysis of daily data-driven stories // *Journalism Practice*, 12(10): 1332-1350.

Публикации на основе больших объемов данных в современной российской журналистике изучаются с 2010-х гг.¹⁶⁴. Российские исследования журналистики данных представляют собой скорее концептуализацию и теоретическое описание данного явления. Крупные эмпирические исследования в заявленной области пока отсутствуют. Существуют работы российских исследователей о значении больших данных для медиа – работа С. А. Варганова, посвященная подходам и стратегиям использования больших данных в массмедиа, в частности в медиаизмерениях онлайн-СМИ и телевидения¹⁶⁵, статьи М. Г. Шилиной о значении больших данных для журналистики и в медиакоммуникации¹⁶⁶.

Среди других отечественных работ по данной проблематике статья С. А. Панюковой о роли концепции открытых данных в развитии журналистики данных и о значении баз открытых данных как первоисточников для публикаций журналистики данных¹⁶⁷. Исследование А. Валеевой доступности открытых цифровых данных как источника для журналистов в России, основанное на сочетании полуструктурированных интервью с журналистами и экспертами по открытым данным с качественным анализом контента СМИ и хакатонов, выявляет, что открытые данные редко используются в практике СМИ, в основном из-за недоступности или неструктурированности и невозможности обработки авторами, однако

¹⁶⁴ Первые проекты журналистики данных в России исследователи относят к 2013 г. (См. Левченко В. Data-журналистика. Учеб. пособие. М., 2013; Шилина М.Г. Data Journalism – дата-журналистика, журналистика метаданных как новый формат медиакоммуникации: к вопросу формирования исследовательских подходов // Медиаскоп. 2013. Вып. 1).

¹⁶⁵ Варганов С. А. Большие данные в онлайн-СМИ: подходы и стратегии использования // Медиаскоп. 2017. Вып. 4; Варганов С.А. Телевизионные измерения в эпоху Big Data: концепции и примеры // Вестн. Моск. ун-та. Сер.10. Журналистика. 2017. № 3. С. 37–57.

¹⁶⁶ Шилина М. Г. Big Data, Open Data, Linked Data, метаданные в PR: актуальные модели трансформации теории и практики // Медиаскоп. 2014. Вып. 1.; Шилина М. Г. Data Journalism – дата-журналистика, журналистика метаданных – в структуре медиакоммуникации: к вопросу формирования теоретических исследовательских подходов // Медиаскоп. 2013. Вып. 1.

¹⁶⁷ Панюкова С. А. Роль открытых данных в развитии журналистики данных // Знак: проблемное поле медиаобразования. 2015. №1 (15). Режим доступа: <https://cyberleninka.ru/article/n/rol-otkrytyh-dannyh-v-razvitii-zhurnalistiki-dannyh>

публикации журналистики данных могут создаваться на проводимых активистами хакатонах по открытым данным¹⁶⁸. Есть и обзорные статьи: С. Симаковой, в которой автор определяет журналистику данных как новый тренд в медиаиндустрии¹⁶⁹, а также исследует влияние журналистики данных на формирование визуальной журналистики¹⁷⁰; статья А. Л. Николова, посвященная проблеме использования новых компьютерных технологий в журналистике, в которой автор рассматривает явление журналистики данных в широком и узком смысле и предпринимает попытку классификации публикаций¹⁷¹; обзорная теоретическая статья М. Н. Шерстюковой¹⁷²; статья М. П. Бочарова о взаимодействиях сферы связей с общественностью и журналистики данных¹⁷³.

Также существуют работы практиков в области цифровых данных, больших данных, государственных данных, концепции открытости данных. Это работы И. Бегтина¹⁷⁴, члена совета по открытым данным при Правительственной комиссии по координации деятельности Открытого

¹⁶⁸ Valeeva, A. (2017) Open Data in a Closed Political System: Open Data Investigative Journalism in Russia. Reuters Institute for the Study of Journalism. University of Oxford. Available at: http://reutersinstitute.politics.ox.ac.uk/sites/default/files/2017-10/Valeeva_Open%20data%20and%20investigative%20journalism%20in%20Russia.pdf

¹⁶⁹ Симакова С. И. Дата-журналистика как медиатренд // Вестник ННГУ, 2014. №2-2. Режим доступа: <https://cyberleninka.ru/article/n/data-zhurnalistika-kak-mediatrend>

¹⁷⁰ Симакова С. И. Журналистика данных как объективное требование времени и её влияние на формирование визуальной журналистики // Знак: проблемное поле медиаобразования. 2016. №1 (18). Режим доступа: <https://cyberleninka.ru/article/n/zhurnalistika-dannyh-kak-obektivnoe-trebovanie-vremeni-i-eyo-vliyanie-na-formirovanie-vizualnoy-zhurnalistiki>

¹⁷¹ Николов А. Л. Новые направления в медиаиндустрии, или Что такое дата-журналистика // Медиа. Информация. Коммуникация. 2012. № 4. Режим доступа: <http://mic.org.ru/index.php/new/165-novye-napravleniya-v-mediaindustrii-ili-chto-takoe-data-zhurnalistika-data-journalism>

¹⁷² Шерстюкова М. Н. Дата-журналистика как новое направление в системе средств массовой коммуникации // Медиа. Информация. Коммуникация. 2012. № 1. Режим доступа: <http://mic.org.ru/1-nomer-2012/65-1-shestukova-2>

¹⁷³ Бочаров М. П. PR-технологии и журналистика данных. Новые формы работы со статистическими и фактологическими источниками информации о социальной сфере // Труд и социальные отношения. 2014. № 4.

¹⁷⁴ Бегтин И. Готовы ли мы к журналистике данных? // Полит.ру. 2013. Режим доступа: http://polit.ru/article/2013/04/29/data_journalism/

правительства, соучредителя АНО «Инфокультура»¹⁷⁵ создателя сайта *OpenGovData.Ru* и других инициатив в области открытости данных, и статья И. Радченко и А. Сакоян, активистов, в 2013 г. организовавших проект, посвященный журналистике данных в России¹⁷⁶, включающая обзор образовательных ресурсов и трендов в области журналистики данных¹⁷⁷.

Существующие концептуальные работы о сущности феномена журналистики данных, его месте в системе массмедиа и возможных социальных эффектах пока не изучены комплексно ни зарубежными, ни российскими авторами, что свидетельствует не только о его новизне, но и о сложности, обуславливает необходимость поиска релевантных междисциплинарных подходов к изучению.

Выводы

Известные к настоящему моменту результаты академического изучения и профессионального осмысления развития современной журналистики под влиянием «цифры» показывают, что датафикация, появление новых цифровых форм, таких как большие данные, обуславливает многоуровневые изменения в журналистике. Становясь основой журналистских материалов, данные расширяют возможности формирования фактографической базы, привносят в работу журналиста определенную специфику. Возникает необходимость поиска новых способов представления информации, в первую очередь, за счет визуализации, публикации приобретают интерактивные характеристики. При этом увеличивается количество этапов подготовки материалов и расширяется число авторов за счет привлечения технических специалистов.

¹⁷⁵ Сайт АНО «Инфокультура». Режим доступа: <https://www.infoculture.ru/>

¹⁷⁶ Журналистика данных. Режим доступа: <http://datadrivenjournalism.ru/>.

¹⁷⁷ Radchenko I., Sakoyan A. (2014) The view on open data and data journalism: Cases, educational resources and current trends. In: Ignatov D.I., Khachay M.Y., Panchenko A., et al. (eds) *Analysis of Images, Social Networks and Texts*. Cham: Springer: 47–54.

Традиции использования зарубежными журналистами больших объемов нецифровых данных, берущие начало в XIX в., обработка цифровых данных, получившая распространение во второй половине XX в. в США, соединение элементов компьютерной журналистики и прецизионного журнализма, появление профессиональных организаций, инициатив по созданию баз данных и их открытости, публикации пионеров журналистики данных начала XXI в. во многом обуславливают развитие и оформление журналистики данных – комплексного направления журналистики, состоящего из комбинации компетенций и навыков нескольких профессий.

С начала 2010-х гг. феномен журналистики данных признается академическим сообществом. На сегодняшний день существует несколько подходов к его исследованию – это качественные исследования отдельных публикаций журналистики данных первой волны; исследования, связанные с компьютерной, вычислительной журналистикой; работы по эпистемологии журналистики данных; работы, посвящённые национальным различиям и особенностям журналистики данных разных стран; работы, посвященные рассмотрению направлений обучения журналистике данных; работы, подчеркивающие роль журналистики данных в обществе, как одного из важнейших инструментов открытости власти и демократии.

Большинство существующих работ, посвященных изучению журналистики данных, основывается на качественных методиках исследования. Значимые исследования контента журналистики данных пока малочисленны. Российские работы о журналистике данных представляют собой концептуализацию и теоретическое описание данного явления, исследования, основанные на количественных методах, отсутствуют.

ГЛАВА 2. ЗАРУБЕЖНЫЙ ОПЫТ ФУНКЦИОНИРОВАНИЯ ЖУРНАЛИСТИКИ ДАННЫХ: КЕЙСЫ ВЕДУЩИХ ИЗДАНИЙ

В предыдущей главе нашей работы мы использовали преимущественно материал, отражающий факты истории и современное состояние журналистики данных за рубежом, в ведущих странах Северной Америки и Западной Европы. Именно в этих регионах она сформировалась, прошла несколько важных этапов в своем развитии, и именно в крупнейших западных медиасистемах оформились передовые и массовые практики изучаемого нами направления журналистской деятельности. В этом контексте нам представляется методологически корректным, прежде чем анализировать российский опыт, провести апробацию выбранной нами методики на кейсах показательных зарубежных изданий. Кроме того, мы ориентировались на то, что помимо результатов апробации, в завершении второго этапа исследования полученный эмпирический материал станет качественной оригинальной базой для сравнительного анализа российского и зарубежного опыта на основе результатов, полученных с использованием идентичной методологии и методики, что даст возможность выявить потенциальные универсальные характеристики изучаемых объектов.

2.1. Методика исследования

Методологические рамки исследования журналистики данных

Изучение существующего академического и профессионального дискурса журналистики данных показало, что в настоящее время в её понимании и трактовке есть ряд важных компонентов, которые должны быть приняты во внимание при формировании методологической базы нашего исследования.

В параграфе 1.1 мы уделили существенное внимание определению понятия «большие данные» и влиянию этого феномена на журналистские практики в начале XXI столетия. Можно обратить внимание на тот факт, что понятие «журналистики данных» стало по-настоящему активно

употребляться после формирования широкого интереса к большим данным во всех их проявлениях и к их воздействию на разные области экономической деятельности и социальной действительности. По сути, рост популярности и научного и индустриального освоения двух терминов шел практически параллельно.

При этом наш опыт предварительного знакомства с продуктами журналистики данных показывал, что сегодня журналисты всё ещё редко основывают свои материалы непосредственно на больших данных в их актуальном понимании. Большинство материалов современной журналистики данных в качестве источников информации содержат массивы данных, которые не достигают по скорости обработки, многообразию и объему характеристик больших данных.

Именно поэтому мы предлагаем называть их «количественными данными» и даем им следующее рабочее определение:

количественные данные – это наборы данных небольших объемов и разнообразия, которые могут быть обработаны человеком с использованием любого из существующих традиционных инструментов обработки. К подобным данным можно отнести статистическую информацию, данные социологических исследований, рейтинги и т.д., анализ которых формирует многие привычные информационные поводы в журналистике.

Данное положение вещей позволяет нам провести важное методологическое разграничение «журналистики данных» в её широком понимании, охватывающем целый спектр материалов, публикуемых в соответствующих рубриках многих изданий, и условной «журналистики больших данных». Последнее в этом случае будет специализированным направлением в журналистике, материалы которого создаются с использованием информации, полученной из соответствующих по масштабам и сложности обработки цифровых баз данных.

Соответственно, ориентируясь на данное нами выше (см. стр. 4) определение журналистики данных, можно считать, что если объем и

разнообразии данных находятся на уровне, который мы назвали выше «количественным», то журналистские тексты, на них основанные, можно относить к условной «журналистике количественных данных».

В то же время в работе отдельно в рамках нашего исследования будет обращено внимание на условную «журналистику больших данных» – направление, в основе которого лежит работа именно с большими данными в соответствии с их пониманием современными специалистами, как цифровой информации, которую невозможно обрабатывать традиционными способами.

В ходе работы мы, в частности, дополнительно эмпирически проверим обоснованность подобного разделения. При отборе непосредственных эмпирических объектов исследования мы намеренно сфокусировали внимание именно на журналистских текстах, в основе которых лежат инфоповоды, связанные с обработкой «больших данных», предполагая, что этот фактор может повлиять на то, что в них полнее, более «выпукло» проявятся возможные особенности, потенциально отличающие журналистику данных от более традиционных направлений и форматов современной журналистики.

Ещё один важный методологический компонент исследования связан с классификацией материалов журналистики данных по специфическим «формам» (подробнее см. Параграф 1.3). Среди выделенных Э. Борхесом-Реем¹⁷⁸ форм для нашей работы наиболее важны так называемые «краткая ежедневная» и «крупная / развернутая расследовательская». При этом журналистские материалы в обеих этих формах могут использоваться как количественные, так и большие данные.

Близким к предыдущей можно считать и ещё один вариант типологии – по практикам подготовки журналистских материалов. В рамках этого теоретического подхода журналистику данных можно разделить на «обычную» (англ. *ordinary data journalism*), в которой публикации могут

¹⁷⁸ Borges-Rey E. (2016) Unravelling data journalism: A study of data journalism practice in British newsrooms. *Journalism Practice* 10(7): 833–843.

готовиться одним журналистом в ежедневном режиме, и «основательную» (англ. *thorough data journalism*)¹⁷⁹, предполагающую работу группы специалистов в различными навыками, чаще всего в рамках отдельных проектов. Во многом этот вариант классификации пересекается с тем, что был нами описан в предыдущем абзаце.

В итоге, условную схему типологизации сегментов журналистики данных, актуальной для нашего исследования, будет выглядеть следующим образом (см. Схема 1):

По типу данных	По форматам материалов	По практикам подготовки материалов
Журналистика количественных данных (ЖКД)	Краткая / ежедневная форма журналистики данных (ЕФЖД)	Обычная журналистика данных (ОбЖД)
Журналистика больших данных (ЖБД)	Развернутая / крупная расследовательская форма журналистики данных (РФЖД)	Основательная журналистика данных (ОснЖД)

Схема 1. Методологически важные типы материалов журналистики данных

В индустриальном и, частично, научном дискурсе специфических особенностей журналистики данных чаще всего говорят о журналистских текстах / проектах, выстроенных на основе сочетания ЖБД (реже ЖКД) с РФЖД. Именно журналистские расследования и лонгриды различных жанров выдвигаются на соискание крупнейших профессиональных специализированных премий и попадают в поле зрения академического сообщества (особенно в России).

Однако нельзя забывать и о том, что в журналистике данных, как и в журналистике в целом, подавляющее большинство материалов составляют все же рутинно производящиеся тексты кратких форматов, «ежедневная форма» работы. При этом важно все-таки и то, насколько на нынешнем этапе

¹⁷⁹ De Maeyer J., Libert M., Domingo D., Heinderyckx F., Le Cam F. (2015) Waiting for Data Journalism. A qualitative assessment of the anecdotal take-up of data journalism in French-speaking Belgium. *Digital Journalism* 3(3): 432-446.

развития журналистика данных адаптирует именно большие данные как наиболее современный тип данных, потенциально, способный определять её специфические черты именно как особого направления в журналистике.

Именно поэтому мы разрабатывали методику исследования, ориентируясь на массив журналистских текстов, выстроенных вокруг сочетания журналистики больших данных (ЖБД) с ежедневной формой журналистики (ЕФЖД) по формату подготовки материалов.

Основные элементы методики¹⁸⁰

Для понимания значения больших данных в журналистике данных и для выявления характерных особенностей её контента на первом этапе исследования мы предприняем кейс-стади специализированных онлайн-ресурсов данного направления журналистики, принадлежащих технологически развитым и авторитетным зарубежным изданиям – *Guardian Datablog* газеты *The Guardian* и *The Upshot* газеты *The New York Times*.

Выбор практик и материалов этих изданий как наиболее показательных кейсов обусловлен не только их востребованностью среди читателей и авторитетностью у профессионального сообщества, но и более прикладными для нашего контекста характеристиками: они оказались среди первых медиа, начавших целенаправленно развивать журналистику данных, при этом с помощью специально созданных онлайн-разделов они «маркируют» подобные материалы. Последнее, с одной стороны, упрощает практическую задачу по отбору публикаций для исследования, а с другой, позволяет видеть в них концентрированный пример того, что сегодня признается журналистикой данных, как западным академическим сообществом, так и журналистами и редакторами, которые чаще всего при изучении или

¹⁸⁰ Раздел написан с использованием материалов и фрагментов из ранее опубликованной статьи: Шилина А. Г. Журналистика данных в зарубежной качественной прессе (кейс специализированных ресурсов газет *The Guardian* и *The New York Times*) // Вестник Московского университета. Серия 10. Журналистика. 2019. № 5. С. 135-162.

обсуждении данного предмета ссылаются именно на эти ресурсы как показательные.

На их примере мы рассматриваем, какое место в материалах журналистики данных отдельных газет занимают большие данные, которые в современных условиях потенциально должны определять её специфику, а также выделяем основные характеристики этого направления, представленные не в отдельных разовых проектах (как, например, в публикациях на основе «Панамского досье»), а в регулярных публикациях новостного типа.

Основной исследовательский метод, как уже указывалось – кейс-стади на базе результатов количественного контент-анализа. Опыт двух изданий может рассматриваться в большей степени именно как кейсы, но, при этом, за счет показательности объектов, в них отражается, на наш взгляд, и общее состояние журналистики данных в настоящий момент.

Хронологические рамки охватывают 2014-2016 гг. – три года ровно в середине 2010-х гг., период, достаточный для сбора статистических данных на уровне пилотажного исследования на этапе, когда основные практики можно считать устоявшимися и представляющими интерес для исследования, когда журналистика данных уже широко представлена в медиа и активно изучается академическим сообществом.

Для выявления тематической и жанровой специфики публикаций (подробнее методику см. далее в этом параграфе), относимых к журналистике данных, они сравниваются с материалами основных изданий *The Guardian* и *The New York Times*. Для этого специально изучаются материалы, размещенные на сайтах газет за один день – среду 28 марта 2018 года. Предварительная проверка показала, что в указанную дату и предшествовавшие дни не происходило событий, которые привлекали бы к себе повышенное внимание и потенциально могли исказить тематический и жанровый баланс, складывающийся в изучаемых изданиях. При отборе публикаций выборка сплошная: массив публикаций, вышедших за один день

(обычно не менее 120-130 для крупных газет в США и Великобритании), видится нам достаточным для того, чтобы дать представление об использовании в газетах определенных тематик и жанров.

Очевидно, что проводить анализ журналистских материалов, в зависимости от предмета исследования, можно по множеству разнообразных критериев. В рамках нашей работы мы сформировали набор критериев, исходя, с одной стороны, из необходимости выявить характеристики контента, позволяющие оценить специфические особенности изучаемого направления журналистики (типы используемых данных, подходы к созданию материалов, задействованные механизмы журналистской работы), а с другой, из возможности применить и более традиционные подходы к исследованию журналистики (жанры, тематика, источники), позволяющие поставить журналистику данных в более привычный для российских исследований медиа контекст.

Как мы уже упоминали, не всегда публикации, относимые сообществом к журналистике данных и публикуемые в соответствующих рубриках, содержат в качестве первоисточника именно большие данные в их признанном понимании – часто материалы основываются на большом наборе количественных данных (статистических, социологических и т.п.). Наличие в анализируемых публикациях больших данных, позволяющих разделять журналистику больших и журналистику количественных данных, определялось нами, исходя из информации, предоставленной в статье. Учитывались, в частности, прямые указания автора на то, что при создании материала используются большие данные; упоминания автором названия ресурса или размещение в публикации ссылок на базы данных, которые в дальнейшем просматривались нами на предмет содержания больших данных; и ряд других критериев.

Для понимания значения больших данных в материалах журналистики данных мы обращаем внимание на то, используются ли они как информационный повод или как вспомогательная часть публикации. Для

выявления особенностей работы журналистов над публикацией журналистики данных, уделялось внимание анализу источника обработки данных – сами ли журналисты обладают навыками работы с инструментами обработки больших данных, находят ли и анализируют данные с помощью специального программного обеспечения, или же опираются на готовый анализ данных (эксперта, аналитической компании, другого источника). Подобное разграничение представляется важным для понимания степени самостоятельности работы специалистов по созданию публикаций журналистики данных и степени оригинальности материалов.

Для релевантного описания специфики журналистики данных (анализа контента и субъекта) публикации сравниваются по набору параметров. За основу взята дополненная нами схема анализа публикаций журналистики данных Ф. Штальфа¹⁸¹.

Параметры анализа публикаций:

- тематика публикации,
- жанр публикации,
- источники информации и их типы.
- число иллюстраций как результата визуализации данных,
- официальное число авторов публикации (подписей под статьей),
- число профессий, компетенции которых необходимы при её создании (например, навыки графического дизайнера и разработчика).

Категория «тематика» разбивается на следующие достаточно крупные подкатегории: политика, экономика / бизнес, здравоохранение, спорт, преступность, культура, образование, технологии, социальные отношения, природные явления, другое (потенциально – наука, мода, путешествия, и пр.). Применяемая в данном исследовании классификация создавалась с опорой на классификатор Ф. Штальфа, который пользовался им при изучении журналистики данных в ведущих зарубежных изданиях. Мы так же

¹⁸¹ Stalphy F. (2017) Classifying Data Journalism. A content analysis of daily data-driven stories. *Journalism Practice*, 12(10): 1332-1350.

посчитали нецелесообразным выделять очень узкие тематические группы в рамках первичного анализа.

Жанр рассматривается по объединенной классификации А. В. Колесниченко из источников 2008 и 2018 годов. В работе 2008 г. автор выделял следующие жанры журналистики: комментарий, аналитическая статья, расширенная новость, короткая новость, «песочные часы», экспертное интервью, информационное интервью, личностное интервью, ньюс-фиче, фиче, репортаж, портрет¹⁸². В статье 2018 г. он использовал более широкий спектр жанров: расширенная новость, аналитическая статья, короткая новость, репортаж, портрет, колонка, случай / житейская история, подборка, рекомендация, реконструкция, тест, рецензия, тест-драйв, мультимедийный лонгрид (формат), опрос (знаменитостей, публики), кейс, ответ, карточки, онлайн-трансляция, расследование, анонсирование книг, обзор матча¹⁸³.

Соответственно, после интеграции двух классификаторов в нашей работе при изучении жанровых особенностей публикаций категория «жанр» разбивается на следующие подкатегории: «песочные часы», информационное интервью, ньюс-фиче, экспертное интервью, личностное интервью, комментарий (в том числе, как комментарий к данным), фиче, расширенная новость, аналитическая статья, короткая новость, репортаж, портрет, колонка, случай / житейская история, подборка, рекомендация, реконструкция, тест, рецензия, тест-драйв, мультимедийный лонгрид (формат), опрос (знаменитостей, публики), кейс, ответ, карточки, онлайн-трансляция, расследование, анонсирование книг, обзор матча.

Мы понимаем, что журналистика данных может не включать некоторые из перечисленных жанров, однако подробная классификация представляется нам необходимой при изучении публикаций традиционной

¹⁸² Колесниченко А.В. Практическая журналистика. Учебное пособие. М.: Изд-во Моск. ун-та, 2008.

¹⁸³ Колесниченко А.В. Востребованность жанров журналистских текстов в онлайн-СМИ // Вестник Московского университета. Серия 10. Журналистика. 2018. № 1. С. 26-42.

журналистики рассматриваемых изданий и для дальнейшего сравнения выявленных жанровых особенностей с характеристиками публикаций, основанных на больших данных.

Источники данных и их типы изучаются на основе классификации источников информации в журналистских материалах М. А. Разумовой, которая выделяет следующие их типы: PR-источники, под которыми подразумеваются официальные заявления компании, пресс-служба, руководство компании, пресс-конференции, сайт компании, корпоративные издания; IR-источники – IR-служба, финансовые отчёты, существенные факты, совет директоров, уставные документы, миноритарные акционеры, инвестиционные меморандумы, встречи с аналитиками и инвесторами, годовые отчёты, базы данных по компаниям (СПАРК, ЕГРЮЛ); государственные источники, под которыми подразумеваются законопроекты, заявления официальных лиц, министерства и ведомства, страницы официальных лиц в интернете, решения судов; экспертный блок источников информации: аналитика, участники рынка, профессиональные ассоциации, консультанты; источники – другие СМИ (российские и иностранные); неофициальные источники, а именно: анонимные источники и конфиденциальные документы; новые медиа – сайты и социальные сети¹⁸⁴.

В рамках нашего исследования, категорию анализа «источники данных» публикации мы разделяем на следующие подкатегории:

- Данные, которые могут являться большими данными или содержаться в публикации в качестве контекстного источника к ним. Это – данные организаций – международных (ЮНЕСКО, ООН, ОБСЕ, МВФ, ВОЗ и др.), государственных / общественных (министерств, национальных статистических служб, профессиональных ассоциаций и т.п.), аналитических компаний; данные социальных медиа / поисковиков / приложений (*Facebook*,

¹⁸⁴ Разумова М. А. Тематические приоритеты и база источников информации деловых СМИ (на примере газет "Коммерсантъ" и "Ведомости"): диссертация ... кандидата филологических наук: 10.01.10 / Место защиты: ФГБОУ ВО «Московский государственный университет имени М. В. Ломоносова», 2016.

Twitter, Google и др.); данные других СМИ / онлайн-ресурсов; собственные данные (данные журналистов издания); данные результатов исследований (академических исследований, опубликованных в научных журналах, например, Стэнфордское исследование «География расовых / этнических различий США»¹⁸⁵ и т.п.);

- Контекстные данные в публикациях: PR-источники (официальные заявления компании, пресс-служба, руководство компании, пресс-конференции, корпоративные издания); IR-источники компании (финансовые отчёты, уставные документы, инвестиционные меморандумы, годовые отчёты, базы данных по компаниям); другие государственные источники (законопроекты, заявления официальных лиц, министерств и ведомств, решения судов);

- Среди контекстных источников информации мы выделяем также экспертный блок источников информации: комментарии или данные научного / академического сообщества; общественных деятелей и представителей бизнеса; комментарии представителей государственных органов.

- Отдельно обращается внимание на то, содержат ли рассматриваемые публикации неофициальные источники (анонимные источники и конфиденциальные документы); или же источник неизвестен / не указан в публикации.

При рассмотрении публикаций также отмечается, к каким типам источников информации относится первоисточник: база данных, данные социальных медиа / поисковиков, документ.

¹⁸⁵ Rich M., Cox A., Bloch M. Money, Race and Success: How Your School District Compares // *The New York Times*. – 2016, 29 Apr. Available at: <https://www.nytimes.com/interactive/2016/04/29/upshot/money-race-and-success-how-your-school-district-compares.html?rref=collection%2Fsectioncollection%2Fupshot>; Source: “The Geography of Racial/Ethnic Test Score Gaps”, by Sean F. Reardon, Demetra Kalogrides and Kenneth Shores of Stanford. Available at: <https://cepa.stanford.edu/content/geography-racialethnic-test-score-gaps>

2.2. Характеристики журналистики данных в *The Guardian Datablog* британской газеты *The Guardian*

Британский *The Guardian Datablog* – показательный пример современной европейской журналистики данных. Он был создан в 2009 г. в качестве отдельного раздела издания *The Guardian* журналистом и впоследствии его редактором С. Роджерсом. Материалы, основанные на крупных цифровых данных публикуются регулярно, несколько раз в сутки, что иллюстрирует растущую востребованность подобной информации. Помимо текстового контента авторы предоставляют читателям возможность ознакомиться с проанализированными данными, размещая на сайте ссылки на базы данных, послужившие основой публикаций, визуализируя данные различными способами, в том числе, интерактивно.

При выборе эмпирических объектов использовалась простая вероятностная выборка. Этот инструмент оказался необходим в условиях, когда количество материалов, опубликованных в *The Guardian Datablog* в 2014-2016 гг. колебалось в районе 60-70 за год, а в *The Upshot* за тот же год размещалось около 600 публикаций. Для того чтобы работать с сопоставимыми данными для обоих ресурсов, было принято решение ориентироваться на средние гарантированные объемы публикационной активности *The Guardian Datablog* и отобрать на обоих изданиях по 60 материалов за год – по 180 за весь изучаемый период.

Результаты контент-анализа материалов *The Guardian Datablog* показывают, что, как мы и предполагали, журналистика данных может быть не связана с большими данными как таковыми – чуть более трети из рассматриваемых публикаций содержат большие данные (74 из 180). В изучаемом тематическом разделе размещаются совершенно разные по своей основе материалы, содержащие в основном «количественные данные».

Так, помимо больших данных часто встречаются следующие типы данных: статистические данные – например, данные национальной службы статистики Великобритании (*Office for National Statistics, ONS*), данные

генерируемые аналитической компанией *Ipsos Mori* и исследовательским центром *Pew Research*, открытые данные государственных порталов Великобритании (*Gov.Uk*¹⁸⁶), опросные данные (например, Института Гэллапа).

В трети публикаций (58 из 180) исходные данные обрабатываются самими журналистами. Как видно из контента подобных статей – автор находит базу данных и самостоятельно при помощи специальных программных инструментов анализирует информацию, содержащуюся в ней. Так, 12 публикаций *The Guardian Datablog* в 2014 г., 20 – в 2015 г., 26 – в 2016 г. включают оригинальную интерпретацию данных. Среди обработанных сотрудниками издания материалов выделяются 46 публикаций, в которых данные служат инфоповодом и 12 – в которых данные являются иллюстрацией или дополнением к основным источникам информации.

Таблица 1. Использование больших данных в публикациях *The Guardian Datablog* 2014-2016 гг.

Количество публикаций, n=74				
Год	Большие данные – инфоповод	Большие данные как инфоповод + обработаны автором	Большие данные как иллюстрация	Большие данные как иллюстрация + обработаны автором
2014	25	5	1	0
2015	20	4	4	2
2016	20	5	4	2
Итого	65	14	9	4

При этом анализ публикаций, основанных на больших данных, выявляет, что практически для всех из них (65 из 74) большие данные являются информационным поводом (для 25 статей в 2014 г., для 20 – в 2015 г., для 20 – в 2016 г.). Только в 9 из 74 рассматриваемых публикаций большие данные выполняют второстепенные функции иллюстрации или дополнения. В четверти публикаций (18 из 74) журналистами

¹⁸⁶ Gov.Uk. Available at: <https://gov.uk>

обрабатываются исходные большие данные. Среди них в 14 публикациях большие данные являются инфоповодом, в 4 – выполняют иллюстративные функции.

Например, для написания статьи «Турция и Россия стимулируют рост запросов на информацию в аккаунтах пользователей *Twitter*»¹⁸⁷ информационным поводом стали большие данные «Отчета о прозрачности¹⁸⁸» социальной сети. Примером обработанных автором больших данных, послуживших инфоповодом, стали данные портала *YouGov*, проанализированные автором в статье «Как партия Тори может оставаться самой крупной партией»¹⁸⁹.

Таким образом, выявляется, что большие данные в публикациях рассматриваемого специализированного раздела, посвящённого журналистике данных, не являются обязательной составляющей (менее половины изученных публикаций содержат большие данные – 74 из 180). В большинстве публикаций, которые включают в себя большие данные, именно они являются основным источником для написания материала (65 из 74), однако журналисты *The Guardian Datablog* редко самостоятельно обрабатывают большие данные (18 из 74), чаще берут за основу готовые исследования аналитических компаний (которые зачастую выполняются по заказу самого издания), или используют обработанные данные, находящиеся в открытом доступе.

¹⁸⁷ Sedghi A. Turkey and Russia drive rise in Twitter user account information requests // *The Guardian datablog*. – 2015, 11 Feb. Available at: <http://www.theguardian.com/news/datablog/2015/feb/09/turkey-and-russia-drive-rise-in-twitter-user-account-information-requests>

¹⁸⁸ Twitter transparency report. Available at: <https://transparency.twitter.com/>

¹⁸⁹ Nardelli A. How Tories could remain the largest party (part one): Ukip support drops // *The Guardian datablog*. – 2015, 12 Feb. Available at: <http://www.theguardian.com/politics/datablog/2015/feb/12/how-tories-could-remain-the-largest-party-part-one-ukip-support-drops>; Nardelli A. How Tories could remain the largest party (part two): vote green, get blue // *The Guardian datablog*. – 2015, 18 Feb. Available at: <https://www.theguardian.com/politics/datablog/2015/feb/18/how-tories-could-remain-the-largest-party-part-two-vote-green-get-blue>

Далее анализируется специфика содержания материалов условной журналистики больших данных – из заданных 180 публикаций изучаются материалы, основанные на больших данных как информационном поводе, то есть 65 публикаций отобранных в начале исследования.

Во многих статьях источники информации дополняют друг друга. Для понимания значения основных и контекстных источников в публикациях журналистики данных, производится их градация и анализ. Во-первых, рассматриваются категории информационных источников, которые могут являться большими данными или содержаться в публикации в качестве контекстного источника к ним: данные организаций, данные социальных медиа / поисковиков / приложений, данные других СМИ / онлайн-ресурсов, собственные данные журналистов издания и данные результатов академических исследований.

Так как далее анализируются статьи, основанные на больших данных, как основном информационном поводе – источник, обозначаемый нами как «инфоповод», соответствует источнику, представляющему собой большие данные.

Выявляется, что самым распространенными источниками информации в публикациях *The Guardian Datablog*, основанных на больших данных, являются данные государственных или общественных организаций (их содержат 34 публикации из 65, причем 29 из этих источников данных являются большими данными, то есть служат информационным поводом для написания материала). Данные аналитических компаний содержатся в 22 из 65 публикаций и в 15 из них служат инфоповодом (большие данные). Реже встречаются данные международных организаций (в 15 публикациях из 65, 12 источников являются инфоповодом = большими данными). Таким образом, материалы журналистики данных *The Guardian Datablog* чаще всего основываются на данных британских организаций.

Данные других СМИ и онлайн-ресурсов содержат 19 статей (в 4 из них являясь большими данными, инфоповодом = большими данными). Большие

данные социальных медиа, поисковых систем и приложений содержатся в 6 из 65 статей, в 4 они служат инфоповодом (большие данные). Материалы раздела, основанные на больших данных, также содержат в качестве источников академические исследования (в 5 из 65, результаты одного исследования, являющиеся большими данными, становятся инфоповодом для написания материала).

Собственные данные издания, собранные журналистами *The Guardian* содержатся лишь в одной статье, попавшей в выборку. Это статья, содержащая опросную информацию (присутствуют данные обработанных анкет, размещавшихся на сайте *The Guardian* за десять дней до публикации статьи) о восприятии людьми своей безопасности после наступления темноты и статистическими данными о вероятности возникновения преступлений (большие данные)¹⁹⁰.

Мы также обращаем внимание на то, какие дополнительные источники информации используются в материалах, в том числе, для создания контекста: PR-источники, IR-источники, государственные источники. Среди контекстных источников выделяется экспертный блок источников информации, выраженный в комментариях академического сообщества, государственных, общественных и бизнес-деятелей.

Так, в публикациях присутствуют PR-источники компаний (в 2 из 65 публикаций), IR-источники компаний (в 2 из 65), государственные источники (в 3 из 65). Например, в публикации о состоянии компании *Coca-Cola* на американском рынке помимо данных аналитической компании *Euromonitor* и данных о динамике поисковых запросов *Google* (большие данные)

¹⁹⁰ Duncan P., Fishwick C. Older women feel most unsafe but are least likely to be attacked // *The Guardian Datablog*. – 2016, 21 Apr. Available at: <https://www.theguardian.com/uk-news/2016/apr/21/women-feel-unsafe-least-likely-attacked-crime-survey-analysis>; Do you feel safe walking outside at night? Share your experiences // *The Guardian Datablog*. – 2016, 11 Apr. Available at: <https://www.theguardian.com/society/2016/apr/11/do-you-feel-safe-walking-outside-at-night-share-your-experiences>

присутствует информация о прибыли, полученная из документов самой компании¹⁹¹.

Таблица 2. Источники информации и типы данных в публикациях *The Guardian Datablog*, основанных на больших данных как информационном поводе, 2014-2016 гг.

Публикации 2014-2016 гг., n=65		
Данные организаций		
Типы источников	Источники больших данных, ставшие инфоповодом	Источники информации, создающие контекст
Государственные / общественные	29	34
Аналитические компании	15	22
Международные	12	15
Другие данные		
	Большие данные	Контекстный источник
Другие СМИ /онлайн-ресурсы	4	19
Социальные медиа / поисковые системы / приложения	4	6
Академические исследования	1	5
Собственные данные издания	0	1
Контекстные данные		
Государственные источники		3
PR-источники компании		2
IR-источники компании		2
Экспертный блок источников информации (комментарии)		
Общество / бизнес		16
Государственные органы		2
Научное/академическое сообщество		1
Другое		
Публикации, содержащие неофициальные источники		0
Публикации, в которых источник не известен / не указан		0
Тип источника данных		
Базы данных		59
Данные социальных медиа/поисковиков		5
Документы		1

¹⁹¹ Arnett G. How Coca-Cola is fighting against a US public losing the taste for it // *The Guardian Datablog*. – 2015, 13 Feb. Available at: <http://www.theguardian.com/news/Datablog/2015/feb/13/coca-cola-us-public-losing-taste>

Экспертный блок источников информации в публикациях *The Guardian Datablog* чаще всего выражен в комментариях общественных и бизнес-деятелей – в 16 из 65 материалов, редко – в комментариях представителей государственных органов и организаций (в 2 из 65), представителей академического сообщества (1 из 65).

При рассмотрении публикаций также отмечается, к каким типам источников информации относится первоисточник: база данных, данные социальных медиа / поисковиков, документ. Так, по типам источники информации в публикациях *The Guardian Datablog* чаще всего являются базами данных (59 из 65). Такие типы, как генерируемые социальными сетями и поисковыми системами большие данные, практически не встречаются (в 5 публикациях из 65). Основанием одной публикации являются большие данные, представленные в форме документа, что скорее представляет собой исключение.

Отдельно обращается внимание на то, содержат ли рассматриваемые публикации неофициальные источники или же, возможно, источник информации неизвестен или не указан. Публикаций *The Guardian Datablog*, содержащих подобную информацию не встретилось.

Таким образом, результаты исследования показывают, что публикации журналистики данных *The Guardian Datablog*, основанные на больших данных основном из баз данных, чаще всего содержат данные государственных организаций или других медиа, комментарии общественных и бизнес-деятелей.

Для понимания особенностей контента журналистики данных изучается тематика и жанр публикаций *The Guardian Datablog*, основанных на больших данных.

Таблица 3. Тематика публикаций *The Guardian Datablog*, основанных на больших данных, 2014-2016 гг.

Тема	Число публикаций, n=65
Экономика / бизнес	15
Политика	14
Здравоохранение	9
Преступность	7
Культура	1
Спорт	1
Природные явления	1
Социальные отношения	1
Технологии	1
Образование	0
Другое	15

Выявляется, что тематика публикаций журналистики данных не обладает особой спецификой. Большинство публикаций посвящено бизнесу и различным аспектам экономики (15 публикаций из 65), а также политическим темам (14 из 65). Теме здравоохранения посвящено 9 публикаций из 65, теме преступности – 7, культуре и спорту, природным явлениям и социальным отношениям – по одной. Значительная часть публикаций освещает круг тем, интересных самой широкой аудитории. В остальных 15 публикациях доминирует тематика, связанная, например, с модой, наукой и др.

При исследовании жанровых особенностей выявляется, что публикации журналистики данных чаще всего представляют собой комментарии к проанализированным данным (37 статей из 65), что представляется существенным отличием журналистики данных от других традиционных направлений. Также встречаются такие жанры, как расширенная новость (11 из 65), аналитическая статья (4 из 65), короткая новость (3 из 65). Три материала из выборки представлены в жанре карточек, две – в жанре кейса. Из классических жанров в выборку попал один репортаж и одно информационное интервью. Три публикации не были отнесены нами ни к одному жанру, предложенному в классификации А. В. Колесниченко, так как они представляют собой размещенные на сайте 2 карты и одну таблицу, содержащие данные, без какой-либо текстовой составляющей.

Таким образом, журналистика данных с одной стороны, не обладает заметной тематической спецификой. При этом, с другой стороны, преобладающий жанр в ней не очень характерен для традиционных форм журналистики – это чаще всего комментарий к данным, на которых основан публикация.

Таблица 4. Жанры публикаций *The Guardian Datablog*, основанных на больших данных, 2014-2016 гг.

Жанр	Число публикаций, n=65
Комментарий	37
Расширенная новость	11
Аналитическая статья	4
Короткая новость	3
Карточки	3
Кейс	2
Репортаж	1
Информационное интервью	1
Другое	3

Для понимания возможных отличий контента журналистики данных в *The Guardian*, её тематической и жанровой специфики, публикации, основанные на больших данных и размещенные в разделе *The Guardian Datablog*, сравниваются с другими материалами издания. Для этого, как упоминалось выше, дополнительно изучаются материалы основной редакции *The Guardian*, размещенные на сайте газеты за один день 28 марта 2018 года. При отборе материалов выборка была сплошная – всего 72 публикации.

Сравнение показывает, что материалы специализированного раздела по тематике отличаются от публикаций издания, а также обладают меньшим жанровым разнообразием.

Так, в публикациях на сайте *The Guardian* политика занимает основное место наравне со спортом (по 12 публикаций из 72). Встречается больше материалов посвященных культуре (9 из 72) и преступности (7 из 72). Экономическая тематика в публикациях встречается редко – всего 3 публикации из 72. Столько же публикаций посвящено здравоохранению.

Также в выборку включается одна публикация на тему образования и один некролог. Значительная часть публикаций раскрывает другие темы (24 из 72): наука (2), мода (3), технологии (6), путешествия (1) и др.

Таблица 5. Соотношение тематики публикаций *The Guardian Datablog*, основанных на больших данных, 2014-2016 гг. и публикаций *The Guardian* от 28 марта 2018 г.

Тема	Доля публикаций (%)	
	<i>The Guardian</i> от 28 марта 2018 г., n=72	<i>The Guardian Datablog</i> , 2014-2016 гг., n=65
Политика	17	22
Спорт	17	2
Культура	12	2
Преступность	10	11
Экономика / бизнес	4	23
Здравоохранение	4	14
Образование	1	0
Природные явления	1	2
Социальные отношения	1	2
Технологии	8	2
Другое	24	23

Данные публикации более разнообразны по жанру. Чаще всего они представлены в форме классических информационных жанров – расширенной новости (33 из 72 публикаций), короткой новости (7 из 72) и аналитической статьи (6 из 72). Также присутствуют традиционные жанры журналистики – экспертные интервью (4 статьи), личностное интервью (2 статьи), три рецензии, три комментария, портрет, репортаж, фиче, житейская история, а также некролог и анкета. Помимо этого четыре публикации представляют по жанру колонки, три – обзоры и рекомендации, одна – обзор матча.

Таким образом, основные публикации *The Guardian* и публикации журналистики данных *The Guardian Datablog* обладают определенными различиями: первые остаются ориентированными на оперативное информирование аудитории по актуальным темам, обладают большим жанровым разнообразием; вторые, в свою очередь, отличаются от

традиционной журналистики издания по жанру – представляют собой комментарии к проанализированным данным.

Таблица 6. Соотношение жанров публикаций *The Guardian Datablog*, основанных на больших данных 2014-2016 гг. и публикаций *The Guardian* от 28 марта 2018 г.

Жанр	Доля публикаций (%)	
	<i>The Guardian</i> от 28 марта 2018 г., n=72	<i>The Guardian Datablog</i> , 2014-2016 гг., n=65
Расширенная новость	46	17
Короткая новость	10	5
Аналитическая статья	8	6
Колонка	6	0
Экспертное интервью	6	0
Личностное интервью	3	0
Рецензия	4	0
Обзор / рекомендация	4	0
Комментарий	4	57
Портрет	1	0
Репортаж	1	2
Фиче	1	0
Житейская история	1	0
Некролог	1	0
Анкета	1	0
Обзор матча	1	0
Карточки	0	5
Кейс	0	3
Информационное интервью	0	2
Другое	0	5

Далее выявляется значение визуального компонента и особенностей подачи материала журналистики данных – отслеживается наличие иллюстративного материала (диаграмма, карта, график и т.п.) в публикациях *The Guardian Datablog*, основанных на больших данных. Исследование показывает, что практически во всех публикациях результаты обработки больших данных визуализированы. Из всей выборки (65 статей) только в одной публикации отсутствует какая-либо визуализация. В большинстве материалов данные визуализируются один (11 из 65), два (13 из 65), три (10 из 65) или четыре (13 из 65) раза. Более трех раз визуализация происходит в половине рассматриваемых публикаций (30 из 65). Исследование визуальных компонентов подтверждает, что в журналистике данных презентация

аудитории материала, в частности, визуализация проанализированных и обработанных данных занимает особо значимое место.

Таблица 7. Визуализация данных в публикациях *The Guardian Datablog*, основанных на больших данных, 2014-2016 гг.

Число иллюстраций к данным	Число публикаций, n=65
Одна	11
Две	13
Три	10
Четыре	13
Пять	8
Шесть	5
Семь	3
Двенадцать	1
Визуализация отсутствует	1

Далее предпринимается попытка описания субъекта журналистики данных *The Guardian Datablog*. На основе изучения числа авторов и компетенций, необходимых для создания публикации, выявляется специфика умений специалиста, работающего в этом направлении.

Выявляется, что авторство большинства публикаций – единоличное: 45 из 65 проанализированных публикаций создаются одним специалистом (подписываются одной фамилией). Реже над публикацией работают двое специалистов (17 из 65). Одна публикация, попавшая в выборку, создана тремя авторами, одна – пятью авторами, одна – двенадцатью. При этом необходимое число компетенций – или профессий, навыки которых нужны для создания материала – всегда больше одной.

Для определения числа компетенций анализируются следующие составляющие публикаций – текстовая, визуальная и интерактивная. Предполагается, что каждая составляющая соответствует наличию у автора определенной компетенции: для создания текста необходимы навыки пишущего журналиста; для создания инфографики, для визуализации данных – навыки графического дизайнера (например, работа с *HTML5*, *CSS* и *Java*

Script); для сложных программных решений и придания интерактивности элементам публикации – навыки программиста¹⁹².

Таблица 8. Авторство в публикациях *The Guardian Datablog*, основанных на больших данных, 2014-2016 гг.

Число авторов	Число публикаций, n=65
Один	45
Двое	17
Трое	1
Четыре	0
Пять	1
Двенадцать	1
Число компетенций	Число публикаций, n=65
Одна	0
Две	52
Три	13

Выявляется, что при создании большинства публикаций *The Guardian Datablog*, основанных на больших данных (52 из 65) требуется обладание минимум двумя компетенциями (навыками пишущего журналиста и графического дизайнера или программиста), при создании остальных 13 – тремя (пишущего журналиста, графического дизайнера и программиста).

Таким образом, журналистика данных *The Guardian Datablog* отличается моносубъектностью: над большинством материалов работает один специалист, при этом компетенции двух или трех профессий требуются при создании большинства публикаций раздела.

2.3. Характеристики журналистики данных раздела *The Upshot* американской газеты *The New York Times*

Американская журналистика данных находится на подъеме с 2014 г. с появлением профильных СМИ, разделов в изданиях и посвященных ей интернет-порталов. Сегодня платформы *FiveThirtyEight.com* и *VOX.com*

¹⁹² Работа со специализированными инструментами обработки больших данных и умение их анализировать рассматриваются нами как отдельная компетенция только на первом этапе исследования – для понимания уровня самостоятельности журналиста при поиске информационного повода и работе с ним, и степени оригинальности материала.

конкурируют с *QZ.com* – порталом, созданным *Atlantic Media*, материалы, основанные на больших данных, публикуются в специализированных разделах на сайтах газет, например, *The New York Times* и *Washington Post*. Растущее число ресурсов и практик постепенно изменяет представление читателей и авторов о журналистике данных и методике создания проектов.

Мы рассматриваем публикации *The New York Times* – одного из наиболее технологически развитых американских печатных изданий, которое одним из первых практикует создание публикаций, основанных на новых форматах цифровых данных: в 2012 году оно запускает проект *2012: The Year in Graphics*¹⁹³, представив главные события года (выборы в США, Олимпиаду, разрушения от урагана и др.) в формате интерактивных визуализаций, После этого 2013 год *The New York Times* объявляет годом интерактивного сторителлинга (*2013: The Year in Interactive Storytelling*)¹⁹⁴ и запускает проект *Data Visualization*. Ежегодно журналисты *The New York Times* создают интерактивные онлайн-проекты, основанные на больших данных, в рамках серий публикаций, озаглавленных «Год в интерактивном сторителлинге, графике и мультимедиа» (*The Year in Interactive Storytelling, Graphics and Multimedia*). С апреля 2014 г. материалы журналистики данных размещаются в специализированном новостном разделе – *The Upshot*, созданном лауреатом Пулитцеровской премии Д. Леонхардтом.

Нами изучаются публикации *The Upshot* 2014-2016 гг., с первого года существования раздела. На первом этапе исследования рассматриваются 180 публикаций (по 60 в год), отобранные методом простого случайного отбора (простая вероятностная выборка) из генеральной совокупности равной примерно 1800 (подробнее о выборке в параграфе 2.2).

Вначале нами выявляется специфика контента *The Upshot*, анализируются источники данных в отобранных 180 публикациях.

¹⁹³ 2012: The Year in Graphics // *The New York Times*. Available at: <http://www.nytimes.com/interactive/2012/12/30/multimedia/2012-the-year-in-graphics.html>

¹⁹⁴ 2013: The Year in Interactive Storytelling // *The New York Times*. Available at: <http://www.nytimes.com/newsgraphics/2013/12/30/year-in-interactive-storytelling/>

Публикации, основанные на «количественных данных», отделяются от материалов, содержащих большие данные.

Результаты показывают, что около половины изученных публикаций раздела содержат большие данные (78 из 180), и почти для всех из них (74 из 78) большие данные являются информационным поводом (для 25 публикаций – в 2014 г., для 20 – в 2015 г., для 29 – в 2016 г.). В четырех из 78 публикаций большие данные выполняют вторичные функции иллюстрации или дополнения.

В остальных 102 публикациях большие данные отсутствуют. Помимо больших данных источником публикаций *The Upshot* часто становятся статистические данные, данные Бюро переписи населения США (*National Fair Housing Alliance*), Бюро статистики труда (*U.S. Bureau of Labor Statistics*), данные аналитических компаний (например, исследовательского центра *Pew Research*), данные государственных порталов открытых данных США и др.

Исходные данные для большинства рассматриваемых публикаций (101 из 180) анализируются при помощи программ обработки больших данных самими сотрудниками издания (36 – в 2014 г., 28 – в 2015 г., 37 – в 2016 г.). При этом в большинстве публикаций, основанных на обработанной ими информации, эти данные служат инфоповодом (98 из 101), только в трех – иллюстрацией.

Большие данные, обработанные самими журналистами, присутствуют в половине публикаций (48 из 78), в большинстве из них являясь информационным поводом для написания материала (47 из 48). Представленные в виде иллюстрации к основным источникам информации обработанные авторами большие данные содержатся в одной статье, попавшей в выборку.

Таблица 9. Использование больших данных в публикациях *The Upshot* 2014-2016 гг.

Количество публикаций, n=78				
Год	Большие данные – инфоповод	Большие данные как инфоповод + обработаны автором	Большие данные как иллюстрация	БД как иллюстрация + обработаны автором
2014	25	21	0	0
2015	20	10	4	1
2016	29	17	0	0
Итого	74	48	4	1

Таким образом, значительная часть материалов журналистики данных рассматриваемого специализированного раздела *The New York Times* включает публикации, содержащие именно большие данные. Однако они не являются обязательной составляющей публикаций (менее половины изученных материалов содержат большие данные – 78 из 180). В большинстве публикаций, которые включают в себя большие данные, именно они являются основным источником для написания материала (74 из 78), и большинство из них находится, анализируется и обрабатывается сотрудниками *The Upshot* самостоятельно (48 из 78).

Например, информационным поводом для написания статьи «Как изменилась лояльность фанатов во время Чемпионата мира по футболу 2014»¹⁹⁵ стала публикация результатов анализа активности в социальной сети *Facebook*, показавшая, как менялась поддержка болельщиков во время турнира¹⁹⁶. Примером послуживших инфоповодом больших данных, обработанных самим автором, стали данные того же *Facebook*, статистику по

¹⁹⁵ Carter S., Quealy K. How Fan Loyalty Changed During the World Cup // *The New York Times*. – 2014, 14 July. Available at: <http://www.nytimes.com/interactive/2014/07/12/upshot/how-fan-loyalty-changed-during-the-world-cup.html?rref=upshot&abt=0002&abg=0>

¹⁹⁶ Diuk C. G., Filiz O. World Cup Fan Migration // *Facebook Data Science team*. – 2014, 12 July. Available at: <https://www.facebook.com/notes/facebook-data-science/world-cup-fan-migration/10152377138198859>

«лайкам» которого проанализировали и проиллюстрировали журналисты издания¹⁹⁷.

Далее из заданных 180 публикаций изучаются материалы, основанные на больших данных как информационном поводе (отобранные на первом этапе исследования 74 публикации). Так как далее изучаются статьи, основанные на больших данных, как основном информационном поводе – источник, обозначаемый нами как «инфоповод», – соответствует источнику, представляющему собой большие данные.

В публикациях *The Upshot* содержащих большие данные многие типы источников информации дополняют друг друга, поэтому для понимания значения основных и контекстных источников в материалах раздела производится их градация и последующий анализ.

При детальном изучении источников информации в публикациях *The Upshot*, основанных на больших данных, выявляется, что самыми распространенными являются данные, которые могут быть как инфоповодом, так и содержаться в публикации в качестве контекстного источника. Это данные государственных или общественных организаций (их содержат 33 публикации из 74, причем 21 из этих источников данных служит инфоповодом для написания материала, то есть являются большими данными), а также данные аналитических организаций, которые содержатся в 30 из 74 публикаций (в 22 из них служат инфоповодом = являются большими данными). Реже встречаются данные международных организаций (в 4 публикациях). Данные других СМИ и онлайн-ресурсов содержат 22 статьи (в 5 из них данные служат инфоповодом = являются большими данными). Большие данные социальных медиа, поисковых систем и приложений содержатся в 14 из 74 статей и служат инфоповодом для их написания.

¹⁹⁷ Giratikanon T., Katz J., Leonhardt D., Quealy K., Tracy M. N.C.A.A. Fan Map: How the Country Roots for College Football // *The New York Times*. – 2014, 3 Oct. Available at: <http://www.nytimes.com/interactive/2014/10/03/upshot/ncaa-football-fan-map.html?abt=0002&abg=0>

Собственные данные *The New York Times* содержат 6 статей, в 5 из них являясь инфоповодом (большими данными). Например, статья об отношении жителей девятнадцати стран к Чемпионату мира по футболу основана на больших данных совместно собранными журналистами *The Upshot* и аналитиками портала *YouGov*¹⁹⁸.

Материалы *The Upshot*, основанные на больших данных, довольно часто включают в качестве источников данные, представленные в результатах академических исследований (в 17 из 74, 3 результата исследования представляют собой большие данные и становятся поводом для написания публикации).

Помимо основных, рассматриваются источники информации, которые используются в материалах журналистики данных для создания контекста. Так, например, наряду с большими данными в публикациях *The Upshot* присутствуют различные государственные источники – законопроекты, заявления министерств и ведомств, решения судов и др. (в 5 из 74 публикациях), комментарии представителей академического сообщества (в 18 из 74) и высказывания общественных и бизнес-деятелей (в 15 из 74). В одной статье, попавшей в выборку, присутствует мнение представителя государственных органов.

В изученных материалах *The Upshot* не встретилось ни одной публикации, содержащей неофициальные источники или неизвестный источник данных.

По типам источники информации в публикациях раздела чаще всего представляют собой базы данных (55 из 74). Реже журналисты *The Upshot* берут за основу материала данные, генерируемые социальными сетями и поисковыми системами (лишь в 14 из 74 изученных публикаций). Основой

¹⁹⁸ Aisch G., Leonhardt D., Quealy K. World Cup Opinions in 19 Countries: Likes, Dislikes, Predictions // *The New York Times*. – 2014, 10 Jun. Available at: <https://www.nytimes.com/2014/06/11/upshot/world-cup-opinions-in-19-countries-likes-dislikes-predictions.html?rref=collection%2Fbyline%2Fkevin-quealy>

пяти публикаций являются данные, представленные в формате документов, основанных на больших данных.

Таблица 10. Источники информации и типы данных в публикациях *The Upshot*, основанных на больших данных как информационном поводе, 2014-2016 гг.

Публикации 2014-2016 гг., n=74		
Данные организаций		
	Источники больших данных, ставшие инфоповодом	Источники информации, создающие контекст
Государственные / общественные	21	33
Аналитические компании	22	30
Международные	4	4
Другие данные		
	Большие данные	Контекстный источник
Другие СМИ / онлайн-ресурсы	5	22
Социальные медиа / поисковые системы / приложения	14	14
Академические исследования	3	17
Собственные данные издания	5	6
Контекстные данные		
Государственные источники		5
PR-источники компании		0
IR-источники компании		1
Экспертный блок источников информации (комментарии)		
Общество / бизнес		15
Государственные органы		1
Научное/академическое сообщество		18
Другое		
Публикации, содержащие неофициальные источники		0
Публикации, в которых источник не известен / не указан		0
Тип источника данных		
Базы данных		55
Данные социальных медиа/поисковиков		14
Документы		5

Таким образом, результаты исследования показывают, что публикации журналистики данных *The Upshot*, созданные на основе больших данных, чаще всего содержат информацию государственных, общественных и

аналитических организаций, а также данные других СМИ, онлайн-ресурсов, социальных медиа, поисковых систем и приложений, которые по форме представляют собой базы данных. Многие публикации в качестве дополнительного источника информации содержат мнения представителей академического и бизнес-сообщества.

В рамках исследования особенностей контента журналистики данных изучается тематика и жанр публикаций *The Upshot*, основанных на больших данных.

Таблица 11. Тематика публикаций *The Upshot*, основанных на больших данных, 2014-2016 гг.

Тема	Число публикаций, n=74
Экономика / бизнес	22
Политика	15
Здравоохранение	11
Спорт	10
Преступность	3
Образование	2
Культура	2
Природные явления	1
Социальные отношения	1
Технологии	0
Другое	7

Выявляется, что большинство публикаций раздела посвящено бизнесу и различным аспектам экономики (22 публикаций из 74), а также политическим темам (15 из 74). Теме здравоохранения посвящено 11 публикаций, спорту – 10, преступности – три, образованию и культуре – по 2 публикации. В 7 публикациях доминирует другая тематика.

По жанру публикации чаще всего представляют собой комментарии автора к проанализированным данным (27 статей из 74). Также встречаются такие жанры, как аналитическая статья (21 из 74), расширенная новость (16 из 74). Две статьи из выборки представлены в жанре карточек, 2 – в жанре подборки, 2 – в жанре теста. Так, в выборку включается статья с проанализированными данными публикаций пользователей сети *Twitter*,

которая представляет собой интерактивную викторину из двенадцати вопросов по лингвистике¹⁹⁹.

Из классических жанров в выборку попал один портрет, ньюс-фиче и два экспертных интервью.

Таблица 12. Жанры публикаций *The Upshot*, основанных на больших данных, 2014-2016 гг.

Жанр	Число публикаций, n=74
Комментарий	27
Аналитическая статья	21
Расширенная новость	16
Экспертное интервью	2
Карточки	2
Подборка	2
Тест	2
Портрет	1
Ньюс-фиче	1
Другое	0

Для выявления тематической и жанровой специфики публикаций журналистики данных *The Upshot* они сравниваются с другими материалами *The New York Times*. Так, дополнительно изучаются материалы, размещенные на сайте *The New York Times* за один день 28 марта 2018 года. Выборка сплошная, всего 54 публикации.

Сравнение показывает, что публикации *The Upshot* обладают меньшим тематическим и жанровым разнообразием в сравнении с контентом основной редакции *The New York Times*. При анализе тем публикаций выявляется, что большая часть из них посвящена политике (12 из 54) и культурной проблематике (10 из 54). Экономическая тематика в публикациях встречается редко (4 публикации из 54). В выборку попали 3 публикации на тему здравоохранения, 2 некролога, одна статья о спорте, одна статья о преступности. Почти половина публикаций раскрывает другие темы (21 из 54): наука, мода, технологии, путешествия, недвижимость и др.

¹⁹⁹ Andrews W., Katz J. Language Quiz: Are You on Fleek? // *The New York Times*. – 2015, 22 Feb. Available at: <https://www.nytimes.com/interactive/2015/02/22/upshot/internet-language-quiz.html?ref=collection%2Fbyline%2Fjosh-katz>

Таблица 13. Соотношение тематики материалов *The Upshot*, основанных на больших данных, 2014-2016 гг. и публикаций *The New York Times* от 28 марта 2018 г.

Тема	Доля публикаций (%)	
	<i>The New York Times</i> от 28 марта 2018 г. n=54	<i>The Upshot</i> , основанных на больших данных 2014-2016 гг. n=74
Политика	22	20
Экономика / бизнес	7	30
Культура	19	3
Здравоохранение	6	15
Наука	6	0
Мода	6	0
Технологии	6	0
Природные явления	0	1
Социальные отношения	2	1
Образование	2	3
Недвижимость	6	0
Путешествия	4	0
Некролог	4	0
Преступность	2	4
Спорт	2	14
Другое	9	9

Публикации *The New York Times* разнообразны по жанру. Чаще всего они представлены в форме классических жанров – расширенной новости (13 из 54 публикаций), репортажа (7 из 54) и аналитической статьи (6 из 54). Реже встречаются: комментарий (4 из 54), портрет (4 из 54), короткая новость (2 из 54), фиче (2 из 54). Также присутствуют три рецензии, три обзора, два мультимедийных лонгрида, две колонки, два некролога, очерк, житейская история, личностное и экспертное интервью.

Таблица 14. Соотношение жанров публикаций *The Upshot*, основанных на больших данных, 2014-2016 гг. и публикаций *The New York Times* от 28 марта 2018 г.

Жанр	Доля публикаций (%)	
	<i>The New York Times</i> от 28 марта 2018 г. n=54	<i>The Upshot</i> , основанных на больших данных, 2014-2016 гг. n=74
Расширенная новость	24	22
Репортаж	13	0
Аналитическая статья	11	28
Комментарий	7	36
Портрет	7	1
Рецензия	6	0
Обзор	6	0
Короткая новость	4	0
Фиче	4	0
Мультимедийный лонгрид	4	0
Колонка	4	0
Некролог	4	0
Очерк	2	0
Экспертное интервью	2	3
Личностное интервью	2	0
Житейская история	2	0
Карточки	0	3
Подборка	0	3
Тест	0	3
Ньюс-фиче	0	1
Другое	0	0

Для выявления особенностей визуальной составляющей материалов журналистики данных *The Upshot* изучается иллюстративный материал в публикациях (графики, диаграммы, таблицы и др.). Из всей выборки (74 статьи) только в одной публикации отсутствует какая-либо визуализация. В большинстве материалов данные иллюстрируются одной (42 из 74), двумя (16 из 74) или тремя (9 из 74) визуализациями. Больше 10 визуализаций содержатся в 6 изучаемых публикациях, при этом их число может достигать пятидесяти. Так, увеличение количества визуального материала в публикации объясняется наличием, к примеру, необходимости

иллюстрировать различные тенденции отдельно, например, по каждому из штатов США²⁰⁰.

Таблица 15. Визуализация данных в публикациях *The Upshot*, основанных на больших данных, 2014-2016 гг.

Число иллюстраций к данным	Число публикаций, n=74
Одна	42
Две	16
Три	9
Другое	6
Визуализация отсутствует	1

На следующем этапе исследования нами предпринимается попытка анализа характеристик субъекта журналистики данных *The Upshot*. На основе изучения числа авторов и компетенций, необходимых для создания публикации, выявляется специфика работы задействованных специалистов.

В половине статей авторство – единоличное (34 из 74). Треть статей подписана двумя фамилиями (25 из 74), десятая часть – тремя (8 из 74).

Таблица 16. Авторство в публикациях *The Upshot*, основанных на больших данных, 2014-2016 гг.

Число авторов	Число публикаций, n=74
Один	34
Двое	25
Трое	8
Другое	7
Число компетенций	Число публикаций, n=74
Одна	0
Две	38
Три	36

В половине публикаций (38 из 74) присутствует два типа работы (текст, основанный на анализе данных, их статическая визуализация), требующие

²⁰⁰ Aisch G., Gebeloff R., Quealy K. Where We Came From and Where We Went, State by State// *The New York Times*. – 2014, 19 Aug. Available at: <https://www.nytimes.com/interactive/2014/08/13/upshot/where-people-in-each-state-were-born.html?rref=collection%2Fbyline%2Fkevin-quealy>; Katz J. ‘Duck Dynasty’ vs. ‘Modern Family’: 50 Maps of the U.S. Cultural Divide // *The New York Times*. – 2016, 27 Dec. Available at: <https://www.nytimes.com/interactive/2016/12/26/upshot/duck-dynasty-vs-modern-family-television-maps.html?rref=collection%2Fsectioncollection%2Fupshot>

навыков двух профессий – журналиста и графического дизайнера. Для создания половины (36 из 74) рассматриваемых публикаций необходимы компетенции трех профессий (журналиста, графического дизайнера и программиста). Таким образом, журналистике данных *The Upshot* присуща моносубъектность: над большинством материалов работает один специалист, при этом при создании публикаций требуются компетенции двух или трех профессий.

2.4. Выявленные характеристики зарубежной «журналистики больших данных»²⁰¹

Для понимания возможности концептуализации феномена журналистики данных и построения её условной нормативной модели, законов её развития и взаимодействия с другими направлениями журналистики и для последующего осуществления исследования журналистики данных в качественной российской прессе мы обобщаем характеристики данного явления, выявленные нами в рамках рассмотрения двух разобранных выше кейсов.

При исследовании контента показательных ресурсов зарубежной журналистики данных выявляется, что менее половины – 139 из 360 – публикаций содержат в качестве источника материала большие данные (65 в *Guardian Datablog* и 74 в *The Upshot*). Это позволяет нам сделать вывод о том, что журналистика данных сегодня не обязательно основывается на результатах обработки данных этого цифрового формата как информационном поводе для написания материала, непосредственно большие данные не являются необходимой составляющей публикаций.

²⁰¹ Параграф написан с использованием материалов и фрагментов из ранее опубликованных статей: Шилина А. Г. Журналистика данных в зарубежной качественной прессе (кейс специализированных ресурсов газет *The Guardian* и *The New York Times*) // Вестник Московского университета. Серия 10. Журналистика. 2019. № 5. С. 135-162; Шилина А. Г., Макеенко М. И. Журналистика данных в качественной прессе: тематические и жанровые особенности // МедиаАльманах. 2019. [Импакт-фактор РИНЦ 2018 – 0,314]

Также они не обязательно собираются, обрабатываются и анализируются самими журналистами – лишь в 62 публикациях из 360 большие данные анализируются при помощи специальных программ обработки больших данных журналистами самостоятельно.

Таблица 17. Использование больших данных в материалах *Guardian Datablog* и *The Upshot* 2014–2016 гг.

Год	Количество материалов, n=152			
	Большие данные – инфоповод	Большие данные как инфоповод + обработаны автором	Большие данные как иллюстрация	Большие данные как иллюстрация + обработаны автором
2014	50	26	1	0
2015	40	14	8	3
2016	49	22	4	2
Итого	139	62	13	5

Основные источники информации в публикациях зарубежной журналистики данных, основанных на больших данных: данные государственных и общественных (национальных) организаций – 68 (в 50 из них подобные данные являются инфоповодом и представляют собой большие данные), аналитических кампаний – 52 (в 37 – инфоповод = большие данные), международных организаций – 19 (в 16 – инфоповод = большие данные). Часто публикации создаются на основе данных других СМИ и онлайн-ресурсов (41 публикация из 139, в 9 информация из подобных источников является инфоповодом = большими данными). Данные социальных медиа, поисковых систем и приложений содержатся в 20 публикациях (в 18 – инфоповод = большие данные). Собственные данные журналистов *The Guardian* и *The New York Times* содержат 7 статей. Материалы *The Upshot*, основанные на больших данных, включают в качестве источников академические исследования 22 (в 4 – результаты исследования представляют собой большие данные). Также в публикациях присутствуют контекстные источники информации: другие государственные источники – законопроекты, заявления министерств и ведомств, решения судов и др. (в 8 из 139 публикаций), два PR-источника и три IR-источника.

Экспертный блок источников информации в публикациях, основанных на больших данных, формируется комментариями общественных и бизнес-деятелей (в 31 из 139), представителей академического сообщества (в 20 из 139), представителей государственных органов и организаций (в 3 из 139 публикаций). В выборку не попала ни одна статья, в которой источник данных не указывался бы, был неизвестен, или статья, содержащая неофициальные источники информации, что раскрывает важность концепции информационной открытости, а также подтверждает значение необходимости ознакомления читателей с первоисточником, послужившим автору поводом для написания материала.

Журналисты, готовящие материалы, чаще всего находят информацию в формате баз данных, содержащуюся в открытом доступе в Интернете, на порталах открытых больших данных и специализированных ресурсах. Так, самыми распространенными типами источников информации в зарубежной качественной прессе являются именно базы данных (114 из 139). Также в части публикаций журналистами используются данные, генерируемые социальными сетями и поисковыми системами (в 19 из 139 изученных публикаций). Основанием шести публикаций являются большие данные, представленные в форме документов.

Таким образом, публикации зарубежной журналистики данных, основанные на больших данных, чаще всего содержат в качестве информационного повода или в качестве дополнительного источника информации данные государственных, общественных и аналитических организаций, а также данные других СМИ, онлайн-ресурсов, социальных медиа и поисковых систем, которые по форме являются базами данных. Данные часто дополняются комментариями общественных и бизнес-деятелей, высказываниями представителей академического сообщества. Собственные данные, собранные сотрудниками изданий, практически не используются.

Таблица 18. Источники и типы данных в публикациях *The Guardian Datablog* и *The Upshot*, основанных на больших данных как информационном поводе, 2014-2016 гг.

Публикации 2014-2016 гг., n=139		
Данные организаций		
	Источники больших данных, ставшие инфоповодом	Источники информации, создающие контекст
Государственные / общественные	50	68
Аналитические компании	37	52
Международные	16	19
Другие данные		
	Большие данные	Контекстный источник
Другие СМИ /онлайн-ресурсы	9	49
Социальные медиа/поисковые системы/приложения	18	20
Академические исследования	4	22
Собственные данные издания	5	7
Контекстные данные		
Государственные источники		8
PR-источники компании		2
IR-источники компании		3
Экспертный блок источников информации (комментарии)		
Общество / бизнес		31
Государственные органы		3
Научное/академическое сообщество		20
Другое		
Публикации, содержащие неофициальные источники		0
Публикации, в которых источник не известен / не указан		0
Тип источника данных		
Базы данных		114
Данные социальных медиа/поисковиков		19
Документы		6

Далее обобщаются выявленные характеристики контента зарубежной журналистики данных – изученное тематическое и жанровое своеобразие публикаций.

Выявляется, что публикации *The Guardian Datablog* и *The Upshot*, основанные на больших данных, не представляют собой разбор особых и специализированных тем. Большинство публикаций посвящено бизнесу и различным аспектам экономики (37 публикаций из 139), а также

политическим темам (29 из 139). Теме здравоохранения посвящено 20 публикаций, спорту – 11, преступности – 10, культуре – 3, образованию, природным явлениям и социальным отношениям – по 2. В остальных 22 публикациях доминирует другая тематика, не охваченная классификатором.

Таблица 19. Тематика публикаций *The Guardian Datablog* и *The Upshot*, основанных на больших данных, 2014-2016 гг.

Тема	Число публикаций, n=139
Экономика / бизнес	37
Политика	29
Здравоохранение	20
Спорт	11
Преступность	10
Культура	3
Образование	2
Природные явления	2
Социальные отношения	2
Технологии	1
Другое	22

Доминирующий жанр в публикациях зарубежной журналистики данных – комментарий. Большинство материалов представляют собой авторский текст с пояснением к проанализированным данным (64 из 139). Другие востребованные жанры – аналитическая статья (27 из 139) и расширенная новость (25 из 139). Встречаются также материалы в форме карточек, подборок, тестов. Три публикации не были отнесены нами ни к одному жанру, предложенному в классификации А. В. Колесниченко, так как они представляют собой размещенные на сайте визуализации без какой-либо текстовой составляющей.

Для выявления тематической и жанровой специфики журналистики данных проводится сравнение с материалами основных изданий *The Guardian* и *The New York Times*. Дополнительно изучаются материалы, размещенные на сайтах этих газет за один день 28 марта 2018 года. При отборе публикаций выборка сплошная – всего 126 материалов.

Таблица 20. Жанры публикаций, *The Guardian Datablog* и *The Upshot*, основанных на больших данных, 2014-2016 гг.

Жанр	Число публикаций, n=139
Комментарий	64
Аналитическая статья	27
Расширенная новость	25
Карточки	5
Короткая новость	3
Кейс	2
Подборка	2
Тест	2
Экспертное интервью	2
Информационное интервью	1
Репортаж	1
Ньюс-фиче	1
Портрет	1
Другое	3

Выявляется, что, публикации *The Guardian* и *The New York Times* в основном посвящены политике (24 публикаций из 126), культуре (19 из 126), и спорту (13 из 126). Часть статей раскрывает темы преступности (8 из 126), технологий – 9, экономики (7 из 126), здравоохранения (6 из 126). Значительная часть материалов (25 из 126) посвящена моде, науке, путешествиям, вопросам, связанным с недвижимостью, проведением досуга, едой и др. В выборку попали также три некролога. Подобное разнообразие тем в публикациях классической журналистики наглядно подчеркивает её отличие от журналистики, материалы которой основываются на больших данных.

Данные публикации разнообразны по жанру. Чаще всего они представлены в форме расширенной новости (47 из 126 публикаций), аналитической статьи (17 из 126), репортажа (11 из 126), короткой новости (9 из 126), интервью (8 из 126). Реже встречаются: обзор (8 из 126), комментарий (7 из 126), рецензия (6 из 126), портрет (5 из 54).

Таблица 21. Соотношение жанров публикаций *The Guardian Datablog* и *The Upshot*, основанных на больших данных 2014-2016 гг., и публикаций *The Guardian* и *The New York Times* от 28 марта 2018 г.

Жанр	Доля в публикациях <i>The Guardian Datablog</i> и <i>The Upshot</i> , основанных на больших данных, n=139 (%)	Доля в публикациях <i>The Guardian</i> и <i>The New York Times</i> от 28 марта 2018 г., n=126 (%)
Комментарий	46	6
Аналитическая статья	19	10
Расширенная новость	18	37
Карточки	4	0
Короткая новость	2	7
Кейс	1	0
Подборка	1	0
Тест	1	0
Экспертное интервью	1	4
Информационное интервью	1	0
Репортаж	1	6
Ньюс-фиче	1	0
Портрет	1	4
Личностное интервью	0	2
Колонка	0	5
Рецензия	0	5
Обзор	0	5
Фиче	0	2
Житейская история	0	2
Некролог	0	2
Анкета	0	1
Обзор матча	0	1
Очерк	0	1
Мультимедийный лонгрид	0	2
Другое	2	0

Сравнение публикаций *The Guardian* и *The New York Times* и публикаций журналистики данных, размещающейся в специализированных разделах, показывает, что публикации, основанные на больших данных, обладают меньшим тематическим и жанровым разнообразием в сравнении с публикациями изданий в целом.

Таблица 22. Соотношение тематики публикаций *The Guardian Datablog* и *The Upshot*, основанных на больших данных, 2014-2016 гг. и публикаций *The Guardian* и *The New York Times* от 28 марта 2018 г.

Тематика	Доля в публикациях <i>The Guardian Datablog</i> и <i>The Upshot</i> , основанных на больших данных, n=139 (%)	Доля в публикациях <i>The Guardian</i> и <i>The New York Times</i> от 28 марта 2018 г., n=126 (%)
Экономика / бизнес	27	6
Политика	21	19
Здравоохранение	14	5
Спорт	8	10
Преступность	7	6
Культура	3	15
Образование	1	1
Природные явления	1	2
Социальные отношения	1	2
Технологии	1	7
Другое	16	26

Для понимания значения визуального компонента и особенностей подачи материала в журналистике данных изучается иллюстративный материал в рассматриваемых публикациях. Из всей выборки (139 статей) только в двух публикациях отсутствует какая-либо визуализация (диаграмма, карта, график и т.п.). В большинстве материалов данные иллюстрируются одной (43 из 139), двумя (24 из 139), тремя (17 из 74) или четырьмя примерами визуализации (17 из 74). Более десяти визуализаций содержатся в 7 изучаемых публикациях (их число может достигать 50).

Исследование визуальных компонентов подтверждает, что в журналистике данных презентация аудитории материала, в частности, визуализация обработанных проанализированных данных, имеет большое значение. Формируется необходимость релевантных способов подачи такой информации, в первую очередь за счет визуализации; большие данные обуславливают особый тип контента, представление информации в котором демонстрирует паритетное соотношение вербального текста и визуальной составляющей.

Таблица 23. Визуализация данных в публикациях *The Guardian Datablog* и *The Upshot*, основанных на больших данных, 2014-2016 гг.

Число иллюстраций к данным	Число публикаций, n=139
Одна	43
Две	24
Три	17
Четыре	17
Пять	12
Шесть	6
Семь	6
Восемь	2
Девять	1
Десять	2
Более десяти	7
Визуализация отсутствует	2

На следующем этапе обобщения характеристик зарубежных качественных СМИ описывается субъект журналистики данных. На основе изучения числа авторов и компетенций, необходимых для создания публикации, выявляется специфика работы специалистов по созданию материалов журналистики данных.

Таблица 24. Авторство в публикациях, *The Guardian Datablog* и *The Upshot*, основанных на больших данных, 2014-2016 гг.

Число авторов	Число публикаций, n=139
Один	79
Двое	42
Трое	9
Четверо	3
Пятеро	3
Шестеро	2
Двенадцать	1
Число компетенций	Число публикаций, n=139
Одна	0
Две	90
Три	49

Выявляется, что для зарубежной журналистики данных характерна моносубъектность. В большинстве публикаций авторство – единоличное (79 из 139), 42 статьи подписаны двумя фамилиями, 9 – тремя. В большинстве публикаций (90 из 139) присутствует два типа работы (текст, основанный на

анализе данных, их статическая визуализация или иллюстрация), требующие навыков двух профессий – журналиста, графического дизайнера или иллюстратора. Для создания 49 из 139 рассматриваемых публикаций необходимы компетенции трех профессий (журналиста, графического дизайнера и программиста).

Исходя из выявленного числа авторов материалов и используемых компетенций, можно было бы предположить, что сотрудникам присуща мультикомпетентность. Однако мы не можем с уверенностью утверждать, что число подписей под публикацией соответствует реальному числу её авторов, возможно технических специалистов не всегда вписывают. В данной работе мы опираемся лишь на ту информацию, которая представлена в публикации.

Выводы

Большинство существующих на сегодняшний день исследований зарубежной журналистики данных основываются на качественных методах описания, теоретизации и интервьюировании экспертов и журналистов. В нашей работе в рамках кейс-стади мы предпринимаем попытку исследования материалов ресурсов-эталонов журналистики данных *The Guardian Datablog* и *The Upshot* количественными методами.

Результаты исследования показывают, что контент материалов, размещенных в этих разделах, по подавляющему количеству критериев обладает схожими характеристиками.

Большинство материалов представляют собой комментарии к данным, посвящены экономическим и политическим темам. Как показывает исследование, журналистика данных сегодня не обязательно основывается на анализе больших данных как инфоповоде, они не являются необходимой составляющей публикаций. Также данные не обязательно анализируются при помощи специальных программ обработки данных самими журналистами – многие материалы являются результатом интерпретации уже готовых

данных. Самыми распространенными источниками информации в публикациях являются большие данные различных организаций (государственных, общественных) и аналитических компаний, других медиа, социальных сетей и поисковых систем.

Исследование визуальных компонентов подтверждает, что в публикациях журналистики больших данных ведущих зарубежных изданий презентация аудитории материала, в частности, визуализация обработанных данных, играет значительную роль. Большинство материалов сопровождается одной или двумя визуализациями. Отличие визуальной составляющей в материалах журналистики данных от иллюстраций в материалах других направлений журналистики заключается в том, что аудитории в понятном виде предъявляются сложные для восприятия массивы информации (в форме графиков, диаграмм, таблиц и др.).

Авторство большинства публикаций зарубежной журналистики данных – единоличное. Изучение характеристик субъекта подтверждает, что журналистика больших данных *The Guardian Datablog* и *The Upshot* подпадает под категорию «обычной» журналистики данных, которая может создаваться одним человеком ежедневно, в отличие от «основательной»²⁰², которая создается группами, обладающими диапазоном различных навыков.

Полученные на первом этапе исследования результаты, представленные во второй главе, также дали нам возможность сделать вывод о том, что выбранные нами методологические рамки и методика изучения журналистики данных работают и могут быть применены для изучения российской журналистики данных.

²⁰² De Maeyer J., Libert M., Domingo D., Heinderyckx F., Le Cam F. (2015) Waiting for Data Journalism. A qualitative assessment of the anecdotal take-up of data journalism in French-speaking Belgium. *Digital Journalism* 3(3): 432-446.

ГЛАВА 3. ЖУРНАЛИСТИКА ДАННЫХ: СОВРЕМЕННЫЙ РОССИЙСКИЙ ОПЫТ

Первый этап исследования, результаты которого были представлены в главе 2, позволил нам, во-первых, провести апробацию выбранной для работы методологии и методики, которые показали свою действенность, а во-вторых, дал возможность на основе обобщения результатов двух кейс-стади составить представление об условной нормативной модели современной зарубежной журналистики больших данных. Это позволило нам перейти к изучению российской журналистики данных, в котором будет применена та же методика и результаты которого будут подвергнуты сравнительному анализу с помощью сопоставления с уже имеющимися характеристиками журналистики больших данных, для того, чтобы встроить тенденции развития отечественной журналистики в более широкий международный контекст. Нам представляется, что в случае обнаружения сходств по ключевым показателям, мы сможем сделать вывод об универсальных форматах современного функционирования журналистики данных.

3.1. Формирование и развитие журналистики данных в России²⁰³

Российские практики использования больших данных в журналистике пока немногочисленны, находятся на стадии становления с начала 2010-х гг., однако предоставляют достаточный материал для анализа особенностей и идентификации журналистики данных.

В массмедиа одними из первых большие открытые данные в своих материалах используют журналисты РИА «Новостей» в начале 2010-х гг., иллюстрируя их интерактивной графикой. В тот же период проекты, основанные на новых форматах цифровых данных, публикуются на интернет-порталах «Лента.ру», «Полит.ру», «Слон.ру», *Look at Me* и на

²⁰³ Параграф написан с использованием материалов и фрагментов из ранее опубликованной статьи: Шилина А. Г. Журналистика данных в российской прессе: особенности источников информации // МедиаАльманах. № 5. 2019. С. 78-88.

сайтах других изданий. На данном направлении журналистики начинают специализироваться отраслевые эксперты-блогеры²⁰⁴.

Первые десять российских публикаций, основанных на новых форматах цифровых данных в сфере медиа, разрабатываются в 2013 г. в рамках программы «Большая Россия – большие данные»²⁰⁵. Реализованные проекты содержат интерактивную визуализацию с обязательным участием пользователей, и минимальный объем текстового контента. Публикации раскрывают социально значимые темы – например, интерактивная программа «Биографии России», демонстрирующая социальные характеристики населения страны, или проект «Порядочное государство. Детские сады», представляющий гражданам интерактивную тепловую карту детских садов в Татарстане. Над проектами работают авторы из таких изданий как «Российская газета», «Московские новости», «РБК», «Эскавайр» и др., и они становятся примерами первых опытов работы российских журналистов с большими данными, находящимися в открытом доступе. При этом все проекты реализуются в Интернете и остаются в формате бета-версий.

Вскоре появляются примеры журналистики данных в СМИ. Публикуются громкие спецпроекты «Новой Газеты», журналисты которой работают над Панамским архивом со специалистами Международного консорциума журналистов-расследователей (*International Consortium of Investigative Journalists, ICIJ*)²⁰⁶, и над «Ландроматом»²⁰⁷ в рамках Проекта по расследованию коррупции и организованной преступности (*Organized Crime*

²⁰⁴ Например, Егор Сусин регулярно пишет о состоянии макроэкономических рынков, подкрепляя каждый пост графиками на основе открытых финансовых данных (блог Егора Сусина. Режим доступа: <http://ugfx.livejournal.com>). Сергей Шпилькин, который во время выборов в Государственную думу сделал серию публикаций на основе открытых данных о выборах (блог Сергея Шпилькина. Режим доступа: <http://podmoskovnik.livejournal.com>).

²⁰⁵ Проекты были реализованы в рамках московского этапа хакатона, организованного Глобальной сетью редакторов (*Global Editors Network*) и группой «РИА Новости».

²⁰⁶ International Consortium of Investigative Journalists. Available at: <https://www.icij.org/>

²⁰⁷ Спецпроект. Золото партитуры // «Новая газета» от 3 Апр. 2016 г. Режим доступа: <http://krug.novayagazeta.ru/12-zoloto-partituri>

and Corruption Reporting Project, OCCRP). В 2015 г. журналисты «РБК» публикуют собственное расследование трат московского бюджета, исследуя базу данных госзакупок²⁰⁸.

Одним из важных факторов, повлиявших на активизацию использования в России количественных данных в качестве источника информации для журналистов, стало распространение концепции открытости информации и гражданских инициатив.

Источники информации для материалов журналистики данных – большие данные – становятся доступными для журналистов в контексте концепции открытых данных. Большие данные являются одним из основных видов открытых данных, создаваемых, к примеру, государственными организациями (информация о государстве, информация, собранная при выполнении функций государственными органами и др.).

Открытые данные в широком смысле — это та часть раскрываемой органами государственной власти и местного самоуправления информации, которая отвечает требованиям свободы доступа, свободы использования и автоматической обработки²⁰⁹. В узком смысле, открытые данные – это машиночитаемые данные, доступные для свободного повторного использования, основными принципами которых являются первичность, полнота, актуальность, пригодность к компьютерной обработке, отсутствие дискриминации по доступу, лицензионная чистота²¹⁰.

За рубежом идеология открытых данных поддерживается крупнейшими государствами мира и международными организациями. С конца 2000-х гг. создаются интернет-ресурсы, содержащие единые базы данных органов государственной власти. Среди международных организаций и ассоциаций,

²⁰⁸ Расследование РБК: кто зарабатывает на реконструкции Москвы // «РБК» 19 Окт. 2015 г. Режим доступа: <https://www.rbc.ru/investigation/society/19/10/2015/561b6c739a79474587968837>

²⁰⁹ Портал «Открытого правительства РФ». Режим доступа: <http://opendata.open.gov.ru/event/5598184/>

²¹⁰ Бегтин И. О том, как устроены открытые данные в России // «Хабрахабр». – 2015, 26 янв. Режим доступа: <http://habrahabr.ru/company/infoculture/blog/248801/>

наиболее вовлеченных в процесс открытия данных и их последующей публикации в виде цифровых баз данных выделяются Всемирный банк, ОЭСР, *Open Government Partnership*, *International Budget Partnership*, *Open Knowledge Foundation*.

Распространение больших данных, во многих развитых странах становится главным мотивом политических и административных преобразований, изменений в общественной сфере. В России тенденция зарождается в 2000-х гг., когда появляются новые модели взаимодействия между государством и обществом.

В 2013 г. страны-участницы «Большой восьмерки» принимают «Хартию открытых данных» (*G8 Open Data Charter*)²¹¹, которая подписывается Россией, состоявшей в международном клубе на тот момент, и претендовавшей на участие в «Партнерстве открытого правительства» (*Open Government Partnership*)²¹². После выхода страны из организации и расхождения по ряду вопросов с Западом, по мнению экспертов, происходит остановка на пути открытости цифровой информации. Тем не менее, в 2017 г. Россия занимает относительно высокие позиции в рейтингах открытости данных. В «Глобальном индексе открытости данных» (*Global Open Data Index*) – 38 место с 20% открытости²¹³, разделяя его с Пуэрто-Рико и островом Мэн. Согласно «Барометру открытых данных» (*Open Data Barometer*), уровень открытости данных в РФ составляет 49%, страна занимает 25 место в рейтинге²¹⁴.

²¹¹ G8 Open Data Charter – Canada's Action Plan. Available at: <http://open.canada.ca/en/g8-open-data-charter-canadas-action-plan>

²¹² Open Government Partnership. Available at: <http://www.opengovpartnership.org>

²¹³ Open Data Index. Available at: <http://index.okfn.org>; <https://index.okfn.org/place/ru/>

²¹⁴ Open Data Barometer. Available at: <http://barometer.opendataresearch.org/report/analysis/rankings.html>; https://opendatabarometer.org/4thedition/detail-country/?_year=2016&indicator=ODB&detail=RUS

Внедрение концепции открытых данных в России, по мнению экспертов²¹⁵, приводит к: повышению прозрачности органов государственной власти; формированию рынка приложений и сервисов; улучшению инвестиционного климата; повышению степени участия гражданского общества в формировании и создании общественно-политических институтов; улучшению качества жизни населения за счет повышения доступности информации. С помощью распространения концепции открытых данных обеспечивается понимание работы властей, создаются услуги для общества и бизнеса, формируется база для гражданского контроля.

Движение в сторону обеспечения доступа к открытым данным, содержащимся в информационных системах российских властей, начинается с появления системы «Открытого правительства», сформированной в 2011 г., принимает гражданско-ориентированный подход и охватывает все уровни государственной и муниципальной власти. Система обладает специальным порталом²¹⁶, а также рядом государственных и общественных порталов, направленных на создание новой модели связи между государством и гражданами («Открытое Министерство», «Открытый регион»²¹⁷ и др.). Одной из основных задач формирования системы «Открытого правительства» является раскрытие информации о деятельности российских государственных органов федеральной власти и органов местного самоуправления в форме открытых цифровых баз данных. Трансформация государственного управления под влиянием цифровизации и Интернета на

²¹⁵ Открытые данные. Совещание у Председателя Правительства Российской Федерации Д. А. Медведева по вопросу "Открытые данные". РАЭК. 6 Фев. 2013. Режим доступа: <http://static.government.ru/media/files/41d466695abe367e6dd3.pdf>

²¹⁶ Система «Открытое правительство». Режим доступа: <http://open.gov.ru/>

²¹⁷ Портал «Открытое Министерство» Режим доступа: <http://open.gov.ru/openministry/>;
Портал «Открытый регион». Режим доступа: <http://open.gov.ru/openregion/>

федеральном и местном уровне показывает положительную тенденцию в укреплении связи между государством и обществом²¹⁸.

Механизмом внедрения открытых цифровых данных в практику государственной работы в России является осуществляемый под эгидой «Открытого правительства» проект «Открытые данные». Целью проекта является реализация экономического и социального потенциала открытых данных, находящихся в распоряжении государства, путем вовлечения их в оборот негосударственного сектора – то есть содействие формированию информационного общества и экономики знаний. Проект является одним из механизмов реализации «Стандарта открытости федеральных органов исполнительной власти»²¹⁹, утвержденного Правительством РФ (N 93-р от 30 января 2014 г.).

Создаются нормативно-правовые акты, подкрепляющие концепцию открытости цифровых данных на новом уровне. В 2013 г. принимается Федеральный закон Российской Федерации от 7 июня 2013 г. N 112-ФЗ «О внесении изменений в Федеральный закон «Об информации, информационных технологиях и о защите информации» и Федеральный закон «Об обеспечении доступа к информации о деятельности государственных органов и органов местного самоуправления»». Закон содержит набор поправок в федеральный закон от 9 февраля 2009 года N 8-ФЗ «Об обеспечении доступа к информации о деятельности государственных органов и органов местного самоуправления» и в федеральный закон от 27 июля 2006 года N 149-ФЗ «Об информации, информационных технологиях и о защите информации» о том, что данные должны публиковаться в форматах открытых данных. Также в распоряжении Правительства Российской Федерации от 10 июля 2013 г. N 1187-р г. представлен перечень тех данных,

²¹⁸ Electronic Governance and Open Society: Challenges in Eurasia (EGOSE 2014). Available at: <http://egov-conf.ifmo.ru>

²¹⁹ «Стандарт открытости федеральных органов исполнительной власти» Стандарт включает методические рекомендации по внедрению принципов открытости и методику оценки уровня открытости. Режим доступа: openstandard.ru/

которые должны публиковаться в первую очередь. При Правительстве РФ функционирует Совет по открытым данным, работы по созданию федерального портала открытых данных контролирует Министерство экономического развития.

Сегодня «Концепция информационной открытости»²²⁰, создание открытых порталов правительства и системы оказания государственных услуг²²¹ обеспечивают обратную связь с обществом. Государство и граждане, как главные действующие лица современной модели коммуникации с использованием больших данных активно осваивают подобные форматы взаимодействия.

В России существует ограничение, связанное с большими данными – сравнительно невысокая доступность баз данных, небольшое количество общедоступной информации в структурированном виде, в отличие от, например, Великобритании или США, в которых присутствуют такие инициативы, как американский *Data.gov* или английский *Data.gov.uk* и другие каталоги больших данных, созданные для предоставления возможности гражданскому обществу и коммерческим организациям пользоваться предоставляемой властями информацией²²², а также содействующие распространению журналистики данных. Не существует общероссийского государственного портала по открытым данным, на сайтах многих органов власти отсутствуют специальные разделы с открытыми данными и др.²²³. Сегодня журналисту сложно создать материал, основываясь

²²⁰ Концепция открытости федеральных органов исполнительной власти. Распоряжение Правительства Российской Федерации от 30 января 2014 г. N 93-р // «Российская газета» от 3 фев. 2014 г. Режим доступа: <https://rg.ru/2014/02/03/koncepciya-site-dok.html>

²²¹ Портал государственных услуг Российской Федерации. Режим доступа: <https://www.gosuslugi.ru/>

²²² [Datacatalogs.org](https://www.datacatalogs.org) Guardian World Government Data. Available at: <https://www.theguardian.com/politics/government-data>

²²³ В России существуют базы данных о бюджетах, индикаторов Центробанка и статистика по сдаче ЕГЭ, климатические, транспортные и иные данные востребованные на практике. Некоторые издания публикуют инфографику и результат анализа некоторых больших и открытых данных, но, пока это происходит не системно и зачастую лишь для понятности печатного материала или для его иллюстрации.

на больших данных, нужную информацию необходимо искать в отдаленных разделах Сети и преобразовывать.

При этом существует множество негосударственных проектов открытых данных и ресурсов для поиска больших данных, призванных помочь журналисту: сообщество *GIsLab*, проекты *OpenGovData*, *Hubofdata* *Buzzdata*, *DataMarket*, *DataCouch*²²⁴ и др., а также многочисленные национальные или отраслевые агрегаторы исследовательских данных, такие как, например, «Британский архив данных» (*UK Data Archive*)²²⁵, онлайн-инструмент, облегчающий журналистам извлечение данных, как, например, *ScraperWiki*, форумы *Get The Data* или *Quora*²²⁶, посвященные вопросам, связанным с базами данных (где находить данные, как следует строить поисковые запросы или извлекать источники данных, какие инструменты использовать исследования баз данных, как очищать базу данных или преобразовывать в формат, пригодный для работы и др.).

Сегодня в мире потребителями больших открытых данных остаются коммерческие компании, некоммерческие организации и активисты, университеты и журналисты. Так, согласно отчету «Новостной лаборатории Гугл» (*Google News Lab*) 2017 г., большие открытые данные востребованы зарубежными журналистами, материалы на их основе создаются несколько раз в неделю²²⁷. В России у больших данных пока мало потребителей. Работа с базами больших открытых данных пока только осваивается российскими журналистами. Большие и открытые данные в России используются в общественных и коммерческих проектах, среди наиболее популярных у журналистов – базы госзакупок и сводные данные системы ГЛОНАСС.

²²⁴ *GIsLab*. Available at: <http://gis-lab.info>; *OpenGovData*. Available at: <http://opengovdata.ru>; *Hubofdata*. Available at: <http://hubofdata.ru>; *Buzzdata*. Available at: <https://www.crunchbase.com/organization/buzzdata>; *DataMarket*. Available at: <https://datamarket.com/>; *DataCouch*. Available at: datacouch.io/

²²⁵ *UK Data Archive*. Available at: www.data-archive.ac.uk/

²²⁶ *ScraperWiki*. Available at: <https://scraperwiki.com/>; *Get The Data - GetTheData.com*. Available at: <https://www.getthedata.com/>; *Quora*. Available at: <https://www.quora.com/>

²²⁷ Rogers S., Schwabish J., Bowers D. (2017) The state of data journalism in 2017. Available at: <https://www.blog.google/topics/journalism-news/data-journalism-2017/>
<https://newslab.withgoogle.com/assets/docs/data-journalism-in-2017.pdf>

Российские данные раскрываются не всегда эффективно (например, государственные геоданные, образовательная, криминальная и иная муниципальная статистика). Исследование доступности больших данных как источника информации для российских журналистов, подтверждает, что они редко используются в практике СМИ, в основном из-за их недоступности или невозможности обработки²²⁸.

В последние годы цифровизация и датафикация находят поддержку государства. В 2017 г. разрабатывается и утверждается распоряжением Правительства Российской Федерации от 28 июля № 1632-р программа по созданию условий для перехода страны к цифровой экономике «Цифровая экономика России 2024»²²⁹. Основными сквозными цифровыми технологиями, которые входят в рамки программы называются большие данные (наряду с нейротехнологиями, системами распределенного реестра, технологиями беспроводной связи, технологиями виртуальной и дополненной реальностей и др.), среди ключевых целей цифровой трансформации называется создание единой архитектуры государственной цифровой платформы, базирующейся на едином массиве данных. Так, цифровизация способствует появлению новой информационной инфраструктуры и упрощению доступа к новым форматам цифровых данных, их структурированию и публикации.

В июле 2017 г. консалтинговая компания *McKinsey* публикует исследование «Цифровая Россия: новая реальность»²³⁰ о роли цифровых технологий в стране. В нем представляется анализ текущего состояния развития цифровизации в России, ее влияния на основные сферы жизни населения и бизнес. Авторы исследования констатируют, что Россия уже

²²⁸ Valeeva A. (2017) Open Data in a Closed Political System: Open Data Investigative Journalism in Russia. Reuters Institute for the Study of Journalism. University of Oxford. Available at: http://reutersinstitute.politics.ox.ac.uk/sites/default/files/2017-10/Valeeva_Open%20data%20and%20investigative%20journalism%20in%20Russia.pdf

²²⁹ Программа «Цифровая экономика России 2024». Режим доступа: <https://data-economy.ru/>

²³⁰ Цифровая Россия: новая реальность. Отчет консалтинговой компании *McKinsey*. Режим доступа: <http://www.tadviser.ru/images/c/c2/Digital-Russia-report.pdf>

живет в цифровой эре: по количеству пользователей Интернета она занимает первое место в Европе и шестое – в мире, а количество пользователей порталов государственных и муниципальных услуг только за 2016 г. увеличилось в два раза и достигло 40 млн человек. Цифровизация постепенно влияет на работу журналистов, открывает новые возможности для поиска источников для материалов среди цифровой информации, среди больших данных, публикуемых на государственном уровне.

Большие данные в России как источник информации для журналистов являются доступными на законодательном уровне, однако в широкую практику использования качественными российскими СМИ пока не входят.

Цифровизация и преобразование медиапространства современной России стимулируют также появление образовательных программ, посвящённых изучению данного нового направления журналистики. Так, в середине 2010-х гг. создается магистерская программа НИУ ВШЭ «Журналистика данных»²³¹.

В 2013 г. «Пособие по журналистике данных» переводится на русский язык и размещается в Интернете во всеобщем доступе²³². В 2017 г. Региональным центром Программы развития Организации Объединенных Наций (ПРООН) для Европы и стран СНГ подготавливается и публикуется учебное пособие для студентов, интересующихся изучением журналистики данных, на русском языке²³³.

В Рунете появляются проекты активистов, например, в 2013 г. формируется портал о журналистике данных на русском языке «Журналистика данных», который помимо рассмотрения данного явления охватывает сферы знаний и умений, которые связаны с открытыми данными и их обработкой. Задачами портала являются аккумуляция имеющегося

²³¹ Магистерская программа НИУ ВШЭ «Журналистика данных». Режим доступа: <https://www.hse.ru/ma/datajournal/>

²³² Грей Д. Пособие по журналистике данных. РИА «Новости». — М., 2013. Режим доступа: https://ria.ru/files/book/_site/index.html

²³³ Data Journalism Manual (2017) Open Data in Europe and Central Asia, ODECA. Available at: <http://www.odecanet.org/data-journalism-manual/>

зарубежного опыта журналистики данных и централизация разбросанной по сети информации по данной теме при необходимости понимания конкретных российских реалий и учета местной специфики²³⁴.

Все чаще проводятся хакатоны, специализированные мероприятия в области журналистики данных: например, московский хакатон по журналистике данных, организованный магистерской программой «Журналистика данных», Школой дизайна при участии факультета компьютерных наук НИУ ВШЭ, *SociaDataHub*, АНО «Информационная культура» и Аналитическим центром при Правительстве РФ; хакатоны в Санкт-Петербурге (*JourHack*), в Оренбурге, «Дальневосточный фокус», хакатон «Новой газеты»²³⁵. Параллельно организуются экспедиции данных: в частности, экспедиция данных активистов проекта «Журналистика данных» в КарГТУ (Казахстан), онлайн-экспедиция данных, организованная проектами «Госзатраты» и «Журналистика данных» с исследованием баз данных «Госзакупок»²³⁶. Проводятся циклы вебинаров для журналистов по работе со статистическими данными²³⁷, ежегодно организуется Всероссийский Конкурс по открытым данным²³⁸.

Попытки концептуализации явления журналистики данных на первоначальном этапе её становления в российских медиа предпринимают практики. Так, И. Бегтин, директор некоммерческого партнерства «Информационная культура», отмечает, что начало развития журналистики данных в России становится следствием тенденции к открытости данных, в

²³⁴ Журналистика данных. Режим доступа: <http://datadrivenjournalism.ru/>; Концепция блога «Журналистика данных». Режим доступа: <http://www.datadrivenjournalism.ru/about-project/concept/>

²³⁵ Хакатон по журналистике данных «Новой газеты». 10-11 марта 2018 г. Режим доступа: <http://hackathon.novayagazeta.ru/>

²³⁶ Дата-Экспедиция №4: 2015. Режим доступа: http://www.datadrivenjournalism.ru/2015/01/12/de4_report/; Дата-Экспедиция №5: 2016. Режим доступа: <http://www.datadrivenjournalism.ru/2016/04/19/de5-intro/>

²³⁷ Цикл вебинаров для журналистов по работе со статистическими данными. Режим доступа: <https://github.com/iradche/webinars/blob/master/201709/webinar01.md>

²³⁸ Всероссийский конкурс «Открытые данные Российской Федерации». Режим доступа: opendatacontest.ru/

первую очередь – государственных, следующим же шагом со стороны государства становится представление данных не просто в открытом доступе, но и в удобном для обработки виде. В случае с журналистикой данных речь идет о «совершенно другом качестве работы с информацией (...), обладающей ценностью, как для национальной экономики знаний, так и для жесткого общественного контроля»²³⁹. Он считает, что журналистика данных – это «журналистика без журналистики в традиционном понимании этого слова /.../ Для формирования журналистики данных как жанра в России давно все готово, есть данные, есть разработчики, которые умеют их визуализировать, и единственное, чего не хватает, – это системного подхода изданий; возможно, кризис в отрасли, а также нарастающая конкуренция со стороны интернет-компаний, социальных сетей и развлекательных брендов, заставит издателей обратить внимание на этот новый перспективный жанр журналистики»²⁴⁰.

Главный редактор журнала «Русский репортер» В. Лейбин говорит о том, что в последние годы в медиапространстве России «растет интерес к фактуре, к реальности, а не к «большим идеям», вынутым из головы, как шар, проблема современного мира – это не недостаток данных, а их избыток». Он подчеркивает, что умение журналиста анализировать данные «является одним из самых востребованных /.../ аудитории нужен журналист (или аналитик, или ученый), чтобы увидеть хоть что-то значимое посреди нарастающего шума информации». Журналистика данных «открывает и новые возможности для изданий: иногда можно не бегать по полям, а просто внимательно анализировать, скажем, сайты госзакупок и прочие данные»²⁴¹.

С начала 2010-х гг. изучению феномена журналистики данных начинают уделять внимание представители российского академического сообщества.

²³⁹ Бегтин И. (АНО «Информационная Культура»). Режим доступа: <https://www.infoculture.ru/glossary/data-journalistic>

²⁴⁰ Бегтин И. Готовы ли мы к журналистике данных? // Полит.ру, 2013. Режим доступа: http://polit.ru/article/2013/04/29/data_journalism/

²⁴¹ Ларот Т. Журналистика данных: как заставить цифры говорить // Журналист. №3 от 13 2013.

Первое определение присутствует в работе М. Шилиной, в которой автор определяет журналистику данных «как набор специфических навыков для поиска, анализа, визуализации информации цифровых источников данных для формирования интерактивных форматов уникальной подачи авторского аналитического журналистского контента и эффективного взаимодействия СМИ, журналиста с аудиторией»²⁴². В 2014 г. вариант определения журналистики данных предлагается С. Симаковой: журналистика данных – это «инструмент, помогающий представить общественности важные сведения, которые могли остаться незамеченными в огромных массивах информации о современной реальности»²⁴³.

Дальнейшие попытки теоретизации журналистики данных представлены в работах А. Валеевой, М. П. Бочарова, А. Л. Николова, С. А. Панюковой, И. Радченко и А. Сакоян, С. И. Симаковой, М. Н. Шерстюковой²⁴⁴.

²⁴² Шилина М. Г. Big Data, Open Data, Linked Data, метаданные в PR: актуальные модели трансформации теории и практики // Медиаскоп. Вып. №1. – М., 2014.

²⁴³ Симакова С. И. Дата-журналистика как медиатренд // Вестник ННГУ. 2014. №2-2. Режим доступа: <https://cyberleninka.ru/article/n/data-zhurnalistika-kak-mediatrend>

²⁴⁴ Valeeva A. (2017) Open Data in a Closed Political System: Open Data Investigative Journalism in Russia. Reuters Institute for the Study of Journalism. University of Oxford. Available at: http://reutersinstitute.politics.ox.ac.uk/sites/default/files/2017-10/Valeeva_Open%20data%20and%20investigative%20journalism%20in%20Russia.pdf;
Бочаров М. П. PR-технологии и журналистика данных. Новые формы работы со статистическими и фактологическими источниками информации о социальной сфере» // Труд и социальные отношения. 2014, № 4; Николов А. Л. Новые направления в медиаиндустрии, или Что такое дата-журналистика // Медиа. Информация. Коммуникация. 4 номер, 2012. Режим доступа: <http://mic.org.ru/index.php/new/165-novyie-napravleniya-v-mediaindustrii-ili-cto-takoe-data-zhurnalistika-data-journalism>; Панюкова С. А. Роль открытых данных в развитии журналистики данных // Знак: проблемное поле медиаобразования. 2015, №1 (15). Режим доступа: <https://cyberleninka.ru/article/n/rol-otkrytyh-dannyh-v-razvitii-zhurnalistiki-dannyh>; Radchenko I., Sakoyan A. (2014) The view on open data and data journalism: Cases, educational resources and current trends. In: Ignatov DI, Khachay MY, Panchenko A, et al. (eds) Analysis of Images, Social Networks and Texts. Cham: Springer: 47–54.; Симакова С. И. Журналистика данных как объективное требование времени и её влияние на формирование визуальной журналистики // Знак: проблемное поле медиаобразования. 2016. №1 (18). Режим доступа: <https://cyberleninka.ru/article/n/zhurnalistika-dannyh-kak-obektivnoe-trebovanie-vremeni-i-eyo-vliyanie-na-formirovanie-vizualnoy-zhurnalistiki>; Шерстюкова М. Н. Дата-журналистика как новое направление в системе средств массовой коммуникации // Медиа. Информация. Коммуникация. №1 2012. Режим доступа: <http://mic.org.ru/1-nomer-2012/65-1-shestukova-2>

3.2. Характеристики российских публикаций журналистики данных в качественной прессе²⁴⁵

Существующие на сегодняшний день исследования российской журналистики данных опираются на теоретическое описание (М. Г. Шилина²⁴⁶) или интервьюирование экспертов и специалистов в данной области (А. Валеева²⁴⁷).

Как уже подробно разобрано выше (см. Параграф 2.1) основным методом сбора эмпирических данных для нас стал количественный контент-анализ. Все основные критерии анализа остаются теми же, что были использованы при изучении текстов зарубежных изданий (см. Параграф 2.1).

В рамках второго этапа исследования рассматриваются публикации российских качественных изданий, отобранных по критериям цитируемости, периодичности (выбираются ежедневные, еженедельные и ежемесячные издания), целевого назначения и характера информации (тематика), авторитетности (учитывается репутация издания). В выборку попали:

- «Коммерсант daily» – ежедневная общественно-политическая газета, выпускаемая издательским домом «Коммерсантъ» с периодичностью шесть раз в неделю.
- «Новая Газета» – общественно-политическое издание, известное журналистскими расследованиями, выходит с периодичностью три раза в неделю.
- *The New Times* (ранее — «Новое время») – общественно-политический журнал. Периодичность выхода на момент проведения исследования: еженедельно по понедельникам.

²⁴⁵ Параграф написан с использованием материалов и фрагментов из ранее опубликованной статьи: Шилина А. Г. Журналистика данных в российской прессе: особенности источников информации // МедиаАльманах. № 5. 2019. С. 78-88.

²⁴⁶ Шилина М. Г. Big Data, Open Data, Linked Data, метаданные в PR: актуальные модели трансформации теории и практики // Медиаскоп. Вып. 1.

²⁴⁷ Valeeva A. (2017) Open Data in a Closed Political System: Open Data Investigative Journalism in Russia. Reuters Institute for the Study of Journalism. University of Oxford. Available at: http://reutersinstitute.politics.ox.ac.uk/sites/default/files/2017-10/Valeeva_Open%20data%20and%20investigative%20journalism%20in%20Russia.pdf

- «Власть» – еженедельный общественно-политический и информационно-аналитический журнал, выпускаемый на момент проведения исследования издательским домом «Коммерсантъ».
- «Деньги» – еженедельный общественно-политический деловой журнал, выпускаемый на момент проведения исследования издательским домом «Коммерсантъ».
- Журнал «РБК» – ежемесячное издание медиагруппы «РосБизнесКонсалтинг». В журнале представлен анализ ситуации в различных отраслях экономики и информация о компаниях, а также международные новости.
- *Forbes* – ежемесячный финансово-экономический журнал, российская версия американского издания *Forbes*.

Данные издания были отобраны как наиболее цитируемые федеральные СМИ, согласно данным системы «Медиалогия» за 2014 год.

Хронологические рамки исследования охватывают 2014-2016 гг. – период становления журналистики данных в России и предполагаемого оформления её в самостоятельное направление журналистики.

Так как в отечественной качественной прессе специализированные ресурсы, самостоятельные рубрики, колонки или разделы, посвящённые исключительно журналистике данных пока отсутствуют, для формирования первичной выборки изучаются все публикации выбранных российских изданий на предмет наличия среди них материалов, которые по используемым нами критериям можно отнести к журналистике данных.

В результате мониторинга были выявлены публикации, содержащие большие данные, а также статьи, основанные на небольших по объёму и разнообразию «количественных» данных (статистических, открытых государственных данных, на данных исследовательских центров и социологических служб) – всего 1564 публикаций.

Таблица 25. Публикации российских качественных изданий, содержащие количественные и большие данные, 2014-2016 гг.

Число публикаций								
Год	«Новая Газета»	«Коммерсант daily»	«РБК»	«Деньги»	«Власть»	<i>Forbes</i>	<i>The New Times</i>	Итого
2014	30	163	37	83	111	36	34	494
2015	28	189	32	106	119	33	23	530
2016	47	202	38	79	126	33	16	541
Итого	105	554	107	268	356	101	73	1 564

Отметим, что чаще всего подобные материалы основываются на данных российских статистических и социологических служб – Росстата, ВЦИОМа, ФОМа, «Левада-центра» и др. Так, например, 156 из 554 публикаций «Коммерсант daily» содержат данные Росстата, 39 из 554 – данные ВЦИОМа и т.д. Часть публикаций является результатом обработки или пересказа тех же информационных поводов, что встречаются в публикациях зарубежной журналистики данных – это крупные цифровые данные международных общественных и аналитических организаций.

Зачастую публикации размещаются в специализированных разделах, или рубриках. Например, в журнале «Деньги» – это была рубрика «Аукцион», в журнале «Власть» – рубрики «Плюс-минус», «По большому счету», «События и цифры. Значимые события минувшей недели в цифрах», «Упоминаемость России в мировых СМИ. Какое внимание мировые СМИ и интернет уделяют России», в издании «Коммерсант daily» – публикации с пометкой «Мониторинг» и публикации Информационного центра.

Вначале в отобранных публикациях изучаются особенности контента на основе анализа источников информации. Материалы разграничиваются по типам данных, присутствующих в них: разделяются публикации, основанные на больших данных, и материалы, основой которых являются остальные типы количественных данных (опросные данные, статистические данные и др.). Для понимания значения основных и дополнительных источников информации они разделяются на информационные поводы для написания материала и на дополнения или иллюстрацию к нему.

Также исследуется источник обработки данных: при анализе публикаций обращается внимание на то, сам ли журналист, автор материала находит и обрабатывает данные, присутствующие в публикации, или же он опирается на готовый анализ данных (компании, эксперта, другого медиа и т.д.), что важно для определения степени оригинальности материалов и для понимания уровня самостоятельности работы специалистов по созданию публикаций журналистики данных в качественной прессе России.

Результаты выявляют, что среди публикаций очень мало материалов, содержащих большие данные – 195 публикаций из 1504, отобранных на предварительном этапе исследования. Причем в рассматриваемых изданиях в год было опубликовано примерно одинаковое количество подобных материалов: 69 – в 2014 г., 70 – в 2015 г., 64 – в 2016 году.

Таблица 26. Публикации российских качественных изданий, основанные на больших данных, 2014-2016 гг.

Число публикаций								
Год	«Новая Газета»	«Коммерсант daily»	«РБК»	«Деньги»	«Власть»	<i>Forbes</i>	<i>The New Times</i>	Итого
2014	1	19	11	13	11	10	4	69
2015	4	23	8	9	7	13	6	70
2016	3	23	5	8	10	11	4	64
Итого	8	55	26	30	28	34	14	195

В большинстве публикаций, которые включают в себя большие данные (171 из 195), именно они являются основным инфоповодом. В части публикаций (24 из 195) они выполняют дополнительные функции – иллюстрации или обогащения материала. Российские журналисты, создающие материалы журналистики данных, используют большие данные в основном в качестве главного инфоповода для написания статьи. Например, для написания статьи «Восток – дело съемное»²⁴⁸ автор журнала «Деньги» использовал большие данные сервиса аренды квартир у собственников

²⁴⁸ Восток - дело съемное // Журнал "Коммерсантъ Деньги" №2 от 18.01.2016, стр. 46. Режим доступа: <https://www.kommersant.ru/doc/2891307>

AirBnb, обозначив, что его клиентура предпочитает путешествовать в Азии. Информационным поводом для статьи «Персональная мобильность в информационном обществе»²⁴⁹ в журнале «Власть» послужили большие данные исследовательской компании *IDC* выявившие, что количество сотрудников, работающих удаленно по всему миру, продолжает расти. Основываясь на данных «Яндекса» журналист «Коммерсанта daily» выяснил, где в Москве чаще всего происходят ДТП²⁵⁰.

Таблица 27. Использование больших данных в публикациях российских качественных изданий 2014-2016 гг.

Публикации российской качественной прессы, n=195				
Год	Большие данные – инфоповод	Большие данные как инфоповод + обработаны автором	Большие данные как иллюстрация	БД как иллюстрация + обработаны автором
2014	54	15	4	1
2015	58	19	8	3
2016	59	17	12	2
	171	51	24	6

В половине отобранных публикаций (88 из 195) данные находятся и анализируются с помощью специальных программ обработки больших данных самими сотрудниками изданий (36 – в 2014 г., 28 – в 2015 г., 37 – в 2016 г.), что говорит об определенной степени самостоятельности российских журналистов. В основном это большие данные (в 57 из 88 публикаций), ставшие инфоповодом (51 из 57). Так, например, большие данные Всемирной туристской организации (*UNWTO*) и онлайн-агентств путешествий стали основным информационным поводом для статьи «Туристическая карта мира»²⁵¹ 2016 г. в журнале «Деньги» вместе с данными Всемирного совета по туризму и путешествиям (*WTTC*), *TripAdvisor*,

²⁴⁹ Персональная мобильность в информационном обществе // Журнал "Коммерсантъ Власть" №22 от 06.06.2016, стр. 38. Режим доступа: <https://www.kommersant.ru/doc/2996171>

²⁵⁰ Буранов И. Аварии расставили по трассам // Газета "Коммерсантъ" №112 от 27.06.2015, стр. 4. Режим доступа: <https://www.kommersant.ru/doc/2756791>

²⁵¹ Шкуренко О., Мишанина Т. Туристическая карта мира-3 // Журнал "Коммерсантъ Власть" от 05.09.2016. Режим доступа: <https://www.kommersant.ru/doc/3078236>

данными компании *Global Blue*, Ассоциации туроператоров России (АТОР) и др.

Далее отобранные публикации, основанные на больших данных (171 публикация), исследуются по тому же набору параметров, что и зарубежные публикации. За основу берется дополненная схема анализа публикаций журналистики данных Ф. Штальфа²⁵². Параметры анализа публикаций: тематика публикации, жанр (по объединенной классификации А.В. Колесниченко²⁵³), источники данных и их типы (на основе классификации источников информации М. А. Разумовой²⁵⁴), визуализация данных, официальное число авторов публикации (подписей под статьей), число профессий, компетенции которых необходимы при её создании (например, навыки графического дизайнера и разработчика).

Так как далее анализируются статьи, основанные на больших данных, ставших информационным поводом, источник, представляющий собой большие данные, обозначается нами как «инфоповод». Для понимания значения основных и контекстных источников информации в публикациях журналистики данных, производится их градация и анализ. Во многих статьях они дополняют друг друга.

Во-первых изучаются источники, которые могут являться результатами обработки больших данных или содержаться в публикации в качестве контекстного источника к ним: данные организаций, данные социальных медиа / поисковиков / приложений, данные других СМИ / онлайн-ресурсов, собственные данные журналистов издания и данные результатов академических исследований.

²⁵² Stalphy F. (2017) Classifying Data Journalism. A content analysis of daily data-driven stories. *Journalism Practice*, 12(10): 1332-1350.

²⁵³ Колесниченко А.В. Практическая журналистика. Учебное пособие. – М.: Изд-во Моск. ун-та, 2008; Колесниченко А.В. Востребованность жанров журналистских текстов в онлайн-СМИ // Вестник Московского университета. Серия 10: Журналистика, издательство Изд-во Моск. ун-та (М.), 2018, № 1, с. 26-42.

²⁵⁴ Разумова М. А. Тематические приоритеты и база источников информации деловых СМИ (на примере газет "Коммерсантъ" и "Ведомости"): диссертация ... кандидата филологических наук: 10.01.10 / Место защиты: ФГБОУ ВО «Московский государственный университет имени М. В. Ломоносова», 2016.

Выявлено, что самыми распространенными источниками информации в публикациях российских СМИ, основанных на больших данных, являются данные государственных или общественных организаций (их содержат 73 публикации из 171, причем 61 из этих источников служат инфоповодом для написания материала).

Таблица 28. Источники и типы данных в публикациях российских качественных изданий, основанных на результатах обработки больших данных как информационном поводе, 2014-2016 гг.

Публикации российской качественной прессы, основанные на результатах обработки больших данных как информационном поводе 2014-2016 гг., n=171		
Данные организаций		
	Источники больших данных, ставшие инфоповодом	Источники информации, создающие контекст
Государственные / общественные	61	73
Аналитические компании	60	67
Международные	12	19
Другие данные		
	Большие данные	Контекстный источник
Другие СМИ / онлайн-ресурсы	17	45
Социальные медиа / поисковые системы / приложения	10	41
Академические исследования	2	23
Собственные данные издания	11	14
Контекстные данные		
Государственные источники		12
PR-источники компании		9
IR-источники компании		6
Экспертный блок источников информации (комментарии)		
Общество / бизнес		57
Государственные органы		46
Научное/академическое сообщество		36
Другое		
Публикации, содержащие неофициальные источники		0
Публикации, в которых источник не известен / не указан		0
Тип источника данных		
Базы данных		153
Данные социальных медиа/поисковиков		18
Документы		0

Данные аналитических компаний содержатся в трети публикаций (67 из 171) и в большинстве из них служат инфоповодом (60 из 67). Реже встречаются данные международных организаций (в 19 публикациях, в 12 из которых являются инфоповодом = большими данными).

Четверть статей содержит данные других СМИ и онлайн-ресурсов, российских или зарубежных (45 из 171, в 17 из них большие данные являются инфоповодом). В четверти материалов присутствуют данные социальных медиа, поисковых систем и приложений (*Facebook, Twitter, Google* и др.) – в 41 из 171 статьи, в 10 из которых служат инфоповодом.

Также материалы российской качественной прессы, основанные на больших данных, включают в качестве источников академические исследования (в 23 из 171).

Собственные данные, генерируемые российскими изданиями самостоятельно, в том числе данные, собранные при помощи сторонних организаций, например по их заказу самого издания, содержат 14 статей (в 11 из них являясь инфоповодом = большими данными). Для периода становления журналистики данных данный показатель вполне высок.

Во-вторых, в материалах российской журналистики данных изучаются дополнительные источники информации, в том числе, создающие контекст: PR-источники компаний, IR-источники компаний, государственные источники. Среди контекстных источников выделяется экспертный блок источников информации, выражаемый в комментариях академического сообщества, мнениях государственных, общественных и бизнес-деятелей.

Так, в ряде публикаций российской журналистики данных присутствует контекстная информация из различных государственных источников – законопроекты, заявления министерств и ведомств, решения судов и др. (в 12 из 171 публикаций). Некоторые публикации содержат PR-источники (9 из 171) и IR-источники компаний (6 из 171).

Дополнительный блок источников информации в публикациях российской журналистики данных выражается также в комментариях

общественных и бизнес-деятелей (содержатся в 57 из 171 публикации), представителей государственных органов и организаций (в 46 из 171 публикации), представителей академического сообщества (в 36 из 171 публикации). Большие данные, присутствующие в материалах российской прессы, часто нуждаются в интерпретации специалистов, расширяющей авторскую трактовку, и во многом упрощающей восприятие информации.

При рассмотрении публикаций отмечается, к каким типам источников информации относится первоисточник: база данных, данные социальных медиа / поисковиков, документ. В российской качественной прессе источниками информации в основном являются базы данных (153 из 171). Данные, собранные из социальных сетей и поисковых систем, российскими журналистами практически не используются (присутствуют лишь в 18 из 171 изученной публикации).

Таблица 29. Тематика публикаций российских качественных изданий, основанных на больших данных, 2014-2016 гг.

Тема	Число публикаций, n=171
Экономика / бизнес	54
Политика	47
Культура	13
Здравоохранение	9
Преступность	4
Образование	4
Спорт	3
Природные явления	3
Социальные отношения	3
Технологии	0
Другое	31

На следующей стадии изучается тематика и жанр публикаций, основанных на больших данных. При изучении тематики было выявлено, что большинство рассматриваемых публикаций посвящено бизнесу и различным аспектам экономики (54 публикации из 171), а также политическим темам (47 из 171). Теме культуры (включающей кино, искусство и др.) посвящено 13 публикаций, здравоохранению и медицине – 9, преступности и

образованию – по четыре, спорту – три. В значительной части публикаций доминирует другая тематика, не охваченная классификатором (туризм, демография, мода, путешествия, наука и пр. – 31 из 171).

Жанровое своеобразие рассматриваемых публикаций заключается в том, что они чаще всего представляют собой комментарии к проанализированным данным (66 статей из 171). Также встречаются расширенная новость (35 из 171), аналитическая статья (29 из 171), расследование (8 из 171), мультимедийный лонгрид (7 из 171), репортаж (4 из 171). Часть статей, попавших в выборку, составлена в жанрах, характерных для онлайн-СМИ – в форме карточек, подборки, кейса.

Таблица 30. Жанры публикаций российских качественных изданий, основанных на больших данных, 2014-2016 гг.

Жанр	Число публикаций, n=171
Комментарий	66
Расширенная новость	35
Аналитическая статья	29
Расследование	8
Мультимедийный лонгрид	7
Карточки	6
Подборка	5
Короткая новость	4
Репортаж	4
Ньюс-фиче	2
Опрос (знаменитостей, публики)	2
Кейс	2
Колонка	1
Другое	0

Для выявления возможных отличий контента журналистики данных, её тематической и жанровой специфики, публикации, основанные на больших данных, сравниваются с другими материалами рассматриваемых изданий – дополнительно изучаются материалы за один день 25 июня 2016 г. – всего 168 публикаций.

Исследуются следующие номера изданий от 25 июня 2016 г.: «Власть» №29(1184), «Деньги» №29(1087), «Коммерсант» № 132/П (5882), «Новая газета» № 80. Для *The New Times* выбран выпуск №22 (412) от 27 июня 2016

г. (ближайший к условной дате номер, так как 25 июня 2016 г. журнал не выходил). Отдельно взяты июльские номера журналов «РБК» и *Forbes* за 2016 год, условно привязанные к выбранной дате. Сроки исследования в данном случае обусловлены теми же критериями, что и в ситуации зарубежных эмпирических объектов.

Сравнение показывает, что публикации, основанные на больших данных, отобранные в ходе мониторинга, не сильно отличаются от более традиционных публикаций, представленных в российской качественной прессе.

Таблица 31. Соотношение тематики публикаций российских качественных изданий за один условный день 25 июля 2016 г. и публикаций, основанных на больших данных, 2014-2016 гг.

Тема	Доля в публикациях российской качественной прессы от 25 июля 2016 г., n=168 (%)	Доля в публикациях российской качественной прессы, основанных на больших данных 2014-2016 гг., n=171 (%)
Экономика / бизнес	24	32
Политика	17	27
Культура	16	8
Преступность	11	2
Спорт	9	2
Здравоохранение	8	5
Образование	3	2
Природные явления	1	2
Социальные отношения	2	2
Технологии	2	0
Другое	7	18

Так, экономика, политика и культура являются наиболее часто встречающимися темами как в большинстве публикаций журналистики данных, так и в традиционных публикациях российской прессы. Однако при этом в материалах журналистики данных встречается меньше материалов посвященных теме преступности, характерных для общественно-политических изданий. Соотношение другой тематики примерно одинаково.

Публикации журналистики данных менее разнообразны по жанрам. Чаще всего они представлены в форме комментариев к данным. Остальные публикации изданий создаются в широком диапазоне жанров, от классических информационных – интервью, расширенной и короткой новости, песочных часов и др. – до тест-драйвов, некрологов, рейтингов и анонсов.

Таблица 32 Соотношение жанров публикаций российской качественной прессы за один условный день 25 июля 2016 г. и публикаций, основанных на больших данных 2014-2016 гг.

Жанр	Доля в публикациях российской качественной прессы от 25 июля 2016 г., n=168 (%)	Доля в публикациях российской качественной прессы, основанных на больших данных 2014-2016 гг., n=171 (%)
Короткая новость	21	2
Аналитическая статья	11	17
Колонка	9	1
Расширенная новость	8	20
Кейс	7	1
Экспертное интервью	5	0
Карточки	5	4
Репортаж	5	2
Реконструкция	4	0
Личностное интервью	4	0
Рецензия	3	0
Случай / житейская история	2	0
Обзор	2	0
Подборка	2	3
Информационное интервью	2	0
Расследование	1	5
Тест-драйв	1	0
Рейтинг	1	0
Ответ на вопрос	1	0
Песочные часы	1	0
Анонс	1	0
Портрет	1	0
Очерк	1	0
Ньюс-фиче	1	1
Некролог	1	0
Рекомендация	1	0
Комментарий	0	39
Мультимедийный лонгрид	0	4
Опрос (знаменитостей, публики)	0	1
Иное	0	0

Таким образом, сравнение показывает, что публикации журналистики данных по тематике не сильно отличаются от публикаций издания, но обладают меньшим жанровым разнообразием.

Для понимания значения визуального компонента и особенностей подачи материала российской журналистики данных изучается содержащийся в них иллюстративный материал.

Выявлено, что в российской журналистике данных текстовый компонент является доминирующим. В значительной части публикаций (42 из 171) отсутствует какая-либо визуализация (диаграмма, карта, график и т.п.). В некоторых публикациях присутствуют сопутствующие визуальные элементы – фото и иллюстрации – косвенно относящиеся к описанию информации доступной из обработанных автором больших данных. Также визуализации, размещаемые на сайтах изданий, не обладают никакими интерактивными функциями. В материалах данные иллюстрируются в основном одной визуализацией (69 из 171), реже – двумя (30 из 171) или тремя (11 из 171).

Таблица 33. Визуализация данных в публикациях российской качественной прессы, основанных на больших данных, 2014-2016 гг.%

Число иллюстраций к данным	Число публикаций, n=171
Одна	69
Две	30
Три	11
Более трех	19
Визуализация отсутствует	42

Исследование визуальных компонентов подтверждает, что в российской журналистике данных презентация аудитории материала, в частности, визуализация проанализированных автором данных, не является обязательной частью публикации, однако в большинстве изученных материалов, в которых она присутствует, ей отводится значительное место.

Далее описаны характеристики субъекта российской журналистики данных. На основе изучения числа авторов публикаций и компетенций, необходимых для их создания, выявляется специфика работы специалистов

по созданию материалов журналистики данных. Исследование показывает, что в большинстве рассматриваемых публикаций авторство единоличное (в 99 из 171). Часть статей подписана двумя фамилиями (35 из 171), часть – тремя (16 из 171).

В большинстве материалов российской журналистики больших данных (129 из 171) присутствует два типа работы – текст, основанный на анализе данных и их статическая визуализация, требующие навыков двух профессий – журналиста и графического дизайнера. Исходя из выявленного числа авторов и компетенций, необходимых для создания публикации, можно было бы предположить, что сотрудникам присуща мультикомпетентность. Однако в данной работе мы опираемся лишь на ту информацию, которая представлена в публикации, и не можем с уверенностью утверждать, что число подписей соответствует реальному числу авторов материала.

Таблица 34. Авторство в публикациях российской качественной прессы, основанных на больших данных 2014-2016 гг.

Число авторов	Число публикаций, n=171
Один	99
Двое	35
Трое	16
Другое	21
Число компетенций	Число публикаций, n=171
Одна	0
Две	129
Три	42

Таким образом, публикации российской журналистики данных отличает моносубъектность: над большинством материалов работает один специалист, при этом при создании большинства публикаций требуются компетенции двух профессий.

Сравнение выявленных характеристик публикаций зарубежной и российской журналистики данных²⁵⁵

Для выявления отличительных особенностей присущих журналистике данных в качественной прессе России, они сравниваются с обобщёнными характеристиками зарубежной журналистики данных, сформулированными в рамках кейс-стади *The Guardian Datablog* и *The Upshot*.

Контент публикаций обладает как очевидными сходствами, так и различиями. Сопоставление результатов исследования показывает, что публикации журналистики данных в показательных зарубежных изданиях и публикации российской журналистики данных основываются на схожих типах источников информации. В первую очередь, на данных организаций – государственных органов и аналитических компаний. В меньшей степени используются данные международных организаций и исследовательских центров. Источниками публикаций часто являются данные других СМИ, при этом основными информационными поводами для публикаций российской прессы часто становятся результаты обработки данные социальных медиа и поисковых систем, а также собственные данные, генерируемые самими изданиями. Часто публикации основываются на результатах академических исследований, представляющих собой большие данные.

Контекстные источники информации чаще встречаются в публикациях российской журналистики данных – это комментарии экспертов бизнес-сообщества и представителей органов государственной власти, высказывания членов академического сообщества. Стоит отметить, что для публикаций зарубежной журналистики данных комментирование данных представителями властей является нехарактерным в отличие от публикаций российской прессы (в 2% от числа публикаций и в 27% соответственно).

²⁵⁵ Раздел написан с использованием материалов и фрагментов из ранее опубликованной статьи: Шилина А. Г., Макеенко М. И. Журналистика данных в качественной прессе: тематические и жанровые особенности // МедиаАльманах. 2019. № 6.

Таблица 35. Источники информации и типы данных в публикациях зарубежной и российской качественной прессы, основанных на больших данных, 2014-2016 гг.

Публикации 2014-2016 гг. (%)				
Данные организаций	Источники больших данных, ставшие инфоповодом		Источники информации, создающие контекст	
	Зарубежные, n=139	Российские, n=171	Зарубежные, n=139	Российские, n=171
Государственные / общественные	36	36	49	43
Аналитические компании	24	35	37	39
Международные	12	7	14	11
Другие данные	Большие данные		Контекстный источник	
	Зарубежные, n=139	Российские, n=171	Зарубежные, n=139	Российские, n=171
Другие СМИ / онлайн-ресурсы	6	10	29	26
Социальные медиа / поисковые системы / приложения	13	6	14	24
Академические исследования	3	1	16	13
Собственные данные издания	4	6	5	8
Контекстные данные	Зарубежные, n=139		Российские, n=171	
Государственные источники	6		7	
PR-источники компаний	1		5	
IR-источники компаний	2		4	
Экспертный блок источников информации (комментарии)	Зарубежные, n=139		Российские, n=171	
Общество / бизнес	22		33	
Государственные органы	2		27	
Научное/академическое сообщество	14		21	
Другое	Зарубежные, n=139		Российские, n=171	
Публикации, содержащие неофициальные источники	0		0	
Публикации, в которых источник не известен / не указан	0		0	
Тип источника данных	Зарубежные, n=139		Российские, n=171	
Базы данных	82		89	
Данные социальных медиа/поисковиков	14		11	
Документы	4		0	

По типам подавляющее большинство источников информации для публикаций является базами данных. При этом, в отличие от публикаций зарубежной журналистики данных, материалы российской прессы не создаются на основе документов, содержащих большие данные (в выборку не попала ни одна подобная публикация).

Таблица 36. Соотношение тематики публикаций зарубежной и российской качественной прессы, основанных на больших данных, 2014-2016 гг.

Тема	Доля публикаций (%)	
	Зарубежные, n=139	Российские, n=171
Экономика / бизнес	27	32
Политика	21	27
Здравоохранение	14	5
Спорт	8	2
Преступность	7	2
Культура	3	8
Образование	1	2
Природные явления	1	2
Социальные отношения	1	2
Технологии	1	0
Другое	16	18

Присутствуют различия в тематике публикаций зарубежной и российской качественной прессы, основанных на больших данных. Несмотря на то, что наиболее часто встречаемыми являются публикации на экономические и политические темы, в российской прессе большее внимание уделяется культурной проблематике – 8% публикаций в отличие от 3% в зарубежных СМИ. В свою очередь, среди публикаций зарубежной журналистики данных чаще встречаются материалы на тему спорта (8% от публикаций в отличие от 2% в российской прессе), преступности (7% в отличие от 2%) и здравоохранения (14% в отличие от 5%). При этом рассматриваемые публикации не сильно отличаются по тематике от материалов традиционной журналистики, согласно проведенному исследованию публикаций отобранных изданий за один условный день.

Публикаций российской и зарубежной качественной прессы, основанные на больших данных, обладают определенной жанровой спецификой: большинство материалов представляют собой комментарии автора к обработанным данным. Особенностью российской журналистики данных является то, что часть её материалов создается в жанре расследований и в формате мультимедийных лонгридов, подобная практика полностью отсутствует в изученных публикациях зарубежной прессы. Возможно, на это соотношение повлиял тот факт, что в российскую выборку

попали журнальные издания, больше открытые материалам, ориентированным на крупные журналистские формы.

Таблица 37. Соотношение жанров публикаций зарубежной и российской качественной прессы, основанных на больших данных, 2014-2016 гг.

Жанр	Доля публикаций (%)	
	Зарубежные, n=139	Российские, n=171
Комментарий	46	39
Аналитическая статья	19	17
Расширенная новость	18	20
Карточки	4	4
Короткая новость	2	2
Кейс	1	1
Подборка	1	3
Тест	1	0
Экспертное интервью	1	0
Информационное интервью	1	0
Репортаж	1	2
Ньюс-фиче	1	1
Портрет	1	0
Расследование	0	5
Мультимедийный лонгрид	0	4
Опрос (знаменитостей, публики)	0	1
Колонка	0	1
Другое	2	0

При этом, в отличие от публикаций традиционной журналистики, материалы журналистики данных обладают меньшим жанровым разнообразием, согласно проведенному исследованию публикаций изучаемых изданий за один условный день.

Таблица 38. Визуализация данных в публикациях зарубежной и российской качественной прессы, основанных на больших данных 2014-2016 гг.

Число иллюстраций к данным	Доля публикаций (%)	
	Зарубежные, n=139	Российские, n=171
Одна	31	40
Две	17	18
Три	12	6
Более 10	5	11
Визуализация отсутствует	1	25

Визуальная составляющая в публикациях, основанных на больших данных, в зарубежной и российской прессе играет важную роль. Большинство материалов сопровождается одним или двумя результатами визуализации обработанных данных. Однако значительным отличием российской журналистики данных является то, что в четверти её публикаций, попавших в выборку, отсутствует какая-либо визуализация – 25% от числа публикаций в отличие от 1% в рассматриваемых зарубежных СМИ.

Исследование также показывает, что автором большинства рассматриваемых публикаций является один специалист (57% от числа зарубежных публикаций и 58% от числа публикаций российских изданий). Часть статей подписана двумя фамилиями, часть – тремя.

В большинстве материалов журналистики данных присутствует два типа работы (текст, основанный на анализе данных, их статическая визуализация), требующие навыков двух профессий – журналиста и графического дизайнера. В связи с тем, что материалы российской журналистики данных публикуются в Сети, дублируя печатные издания, и в них отсутствуют какие-либо интерактивные функции, анализ выявляет, что при их создании всегда требуется не более двух компетенций (компетенции программиста или разработчика не представляются необходимыми).

Таблица 39. Авторство публикаций зарубежной и российской качественной прессы, основанных на больших данных 2014-2016 гг.

Число авторов	Доля публикаций (%)	
	Зарубежные, n=139	Российские, n=171
Один	57	58
Двое	30	20
Трое	6	9
Другое	6	12
Число компетенций	Доля публикаций (%)	
	Зарубежные, n=139	Российские, n=171
Одна	0	25
Две	65	75
Три	35	0

Таким образом, публикации зарубежной и российской журналистики данных отличает моносубъектность: над большинством материалов работает один специалист, при этом при создании большинства публикаций требуются компетенции двух профессий.

Выводы

В третьей главе было проведено исследование журналистики данных в российских качественных новостных медиа по методике, опробованной в предыдущей главе. Был также выполнен сравнительный анализ отечественных и показательных зарубежных подходов к созданию текстов журналистики больших данных, который способствовал определению ряда специфических особенностей её российского варианта.

Российская журналистика (больших) данных характеризуется моносубъектностью – над большинством материалов работает один специалист. При этом при создании контента часто требуются компетенции двух профессий из-за необходимости визуализации данных. Значение визуального компонента в публикациях велико, однако наличие визуализации данных не является обязательным компонентом публикаций российской журналистики данных – в четверти изученных статей визуализация отсутствует.

В большинстве изученных публикаций, которые включают в себя большие данные, именно результаты их обработки стали информационным поводом для написания материала, однако менее половины из них анализируется с помощью программ обработки данных авторами публикаций самостоятельно. Чаще российские журналисты используют готовые данные государственных, общественных и аналитических организаций, а также данные других медиа, онлайн-ресурсов, социальных медиа и поисковых систем, которые по форме представляют базы данных.

Публикации российской журналистики данных не обладают особым тематическим и жанровым разнообразием: большинство из них посвящены

темам экономики, бизнеса и политики и представлено в жанрах комментария, аналитической статьи или расширенной новости.

Проведенное сравнение характеристик зарубежной и российской журналистики данных в качественной прессе помогло выявить основную специфику российской журналистики данных: от зарубежных практик её отличает то, что основными информационными поводами для создания публикаций часто становятся данные социальных медиа и поисковых систем, а также собственные данные, генерируемые самими изданиями. Дополнительные источники информации (комментарии экспертов, представителей органов государственной власти и членов академического сообщества) встречаются в российских публикациях намного чаще, чем в зарубежных. Характерной особенностью российских публикаций является комментирование данных именно представителями властей – в зарубежных публикациях подобная практика редка.

Несмотря на то, что наиболее часто встречаемыми являются публикации на экономические и политические темы, в российской прессе большое внимание уделяется культурной проблематике и теме образования.

Большинство материалов представляют собой комментарии автора к обработанным данным. Особенностью российской журналистики данных является то, что часть публикаций создается в жанре расследований и в формате мультимедийных лонгридов.

Публикации сопровождаются одной или двумя визуализациями данных. Однако в четверти публикаций российской журналистики данных отсутствует какая-либо визуализация.

Для российской журналистики данных, как и для зарубежной, характерна моносубъектность, однако при создании материала требуется не более двух компетенций – журналиста и графического дизайнера, в то время как компетенции программиста или разработчика не представляются необходимыми.

Несмотря на перечисленные выше специфические черты российской журналистики больших данных, главным результатом сравнительного анализа можно считать вывод о том, что по подавляющему числу показателей данные по отечественным и зарубежным текстам совпадают. И это, с одной стороны, позволяет нам говорить о включенности российской ежедневной формы журналистики данных в международный контекст, а с другой стороны, видеть в результатах сравнительного анализа базовые элементы универсальных характеристик журналистики больших данных.

ЗАКЛЮЧЕНИЕ

Для выявления ключевых характеристик журналистики данных в российских качественных медиа нами было проведено первое в отечественной научной практике эмпирическое исследование этого направления в современном развитии журналистики.

Одним из важных результатов проведенной исследовательской работы нам представляется то, что нам удалось сформировать действенную методологическую рамку и методику изучения журналистики данных, которая в ближайшее время может быть актуальной для начинающегося сейчас этапа разведывательных исследований в этой области. Методически корректным оказалось интегрирование в категории классификаторов, используемых для контент-анализа текстов журналистики данных, не только апробированных западными учеными критериев, но и ряда разработок отечественных специалистов, связанных, в частности, с типологией журналистских жанров и источников информации. Близость полученных на зарубежном и на российском эмпирическом материале результатов показывает, что методика адекватна современным реалиям журналистики данных и важна для создания базы для сравнительного анализа исследовательских материалов, полученных в разных национальных и профессиональных традициях и контекстах.

Методологически важным решением стало и выделение на этапе формирования теоретических рамок исследования разных типов журналистики данных по категориям исходного материала (количественные / большие данные), форматам подготовки (обычная / основательная работа) и представления текстов (краткая (ежедневная) / развернутая форма). Выбранное нами для выделения эмпирических объектов сочетание больших данных, кратких форм и обычных ежедневных практик позволило, с одной стороны, сделать акцент на типах данных, в наибольшей степени отражающих эффекты процессов датафикации в начале XXI века. А с другой стороны, провести исследование на материале тех журналистских практик,

которые сегодня связаны с ежедневным рутинным рабочим процессом, в рамках которого пока работает подавляющее большинство сотрудников новостных медиа и производится большая часть журналистских произведений.

Собранная в рамках данного исследования информация, на наш взгляд, дает возможность для развития общих представлений и академического дискурса журналистики данных. В настоящий момент в России она по-прежнему ассоциируется, в первую очередь, с крупными проектами, часто исследовательского характера, основанными на результатах обработки так называемых больших данных. Однако и анализ кейсов показательных зарубежных онлайн-ресурсов, принадлежащих ведущим англоязычным газетам, и результаты изучения работ, опубликованных в ведущих российских качественных медиа, дают основания для выделения нескольких важных для понимания современного состояния журналистики данных тенденций.

При этом анализ и интерпретацию полученных при изучении зарубежных кейсов и российского опыта результатов можно проводить в рамках минимум двух контекстов. Первый из них, более узкий, связан с выявлением характеристик непосредственно российской журналистики данных. Второй же может быть более широким и ориентированным на дискуссии о развитии современной отечественной (и не только) журналистики в целом.

Так, в условиях, когда пока ни в профессиональных, ни в академических кругах не сложилось устойчивого и четкого определения журналистики данных, полезным представляется ориентация на использование понятия «больших данных». Изучение попавших в выборку текстов показало, что при потоковом, конвейерном производстве материалов, определяемых редакциями как журналистика данных, непосредственно большие данные являются основой чуть более чем для трети публикаций. В этой ситуации актуальным становится использование при определении,

классификации, обсуждении и изучении журналистики данных дополнительных критериев, основанных на уточняющих характеристиках используемых данных. Это позволило нам, например, подтвердить сформулированные при разработке теоретической базы предположения и эмпирически разделять внутри журналистики данных условные «журналистику количественных данных» и «журналистику больших данных».

При этом контент-анализ материалов, которые в рамках рассматриваемого кейса можно отнести к «журналистике больших данных», показывает, что практики производства и характеристики текста не демонстрируют существенных отличий большинства изученных публикаций от вполне традиционных представлений о современной журналистике.

Тематические и жанровые характеристики (при сравнении, в нашем случае, с контентом материнских изданий), в целом, мало отличаются от массива материалов, производимых редакциями медиа в ежедневном режиме. Возможно, обращает на себя внимание только более узкий спектр используемых в журналистике данных жанров и явное доминирование среди них группы аналитических жанров. Анализ жанровых доминант выявил преобладание в журналистике данных комментария и аналитической статьи при меньшем разнообразии жанровой палитры по сравнению с контентом ведущих качественных изданий в целом. Доли расширенной новости в выбранных российских и зарубежных образцах журналистики данных тоже почти совпадают и составляют около одной пятой изученного контента.

Важным направлением для дальнейшего сбора информации и научной рефлексии может стать выявление того, насколько такое соотношение жанров «естественно» для журналистики данных и насколько использование больших данных как основы инфоповода влияет на выбор редакциями тех или иных жанров / групп жанров.

Собранные данные также показывают минимальные различия в тематических и жанровых доминантах отобранных образцов российской и

зарубежной журналистики данных, что может быть основанием для того, чтобы говорить о формировании универсальной модели «журналистики больших данных». В рамках нашего исследования замеры происходили по, так сказать, укрупненным тематическим группам и показали доминирование политической и экономической тематики. Немного более высокий интерес к криминалу и здравоохранению в англо-американских изданиях и к культуре в отечественных можно объяснить как национальными особенностями, так и форматом попавших в выборку изданий.

При этом потенциал дальнейшего развития «тематического подхода» видится нам расширении имеющихся тематических областей за счет детализации классификатора, например, с ориентацией на опыт российских контент-аналитических исследований последних лет²⁵⁶.

Что касается подготовки материалов, то можно заметить, что даже журналистика данных, потенциально предполагающая развитие технологических навыков работающих в этом сегменте авторов, пока не свидетельствует о формировании широкого круга специалистов нового типа. Лишь в небольшой доле текстов непосредственный инфоповод на базе результатов обработки больших данных генерируется редакцией (и, возможно, автором), а не сторонними организациями. Если основываться на информации об авторстве изученных материалов, наличие в большинстве публикаций одного автора теоретически может говорить о формировании у журналистов компетенций графического дизайнера или программиста (если исходить из того, что над материалом не работала команда и часть её членов в выходных данных просто не указана). Однако невысокая сложность решенных в большинстве материалов технологических задач может говорить и о справедливости предложенного зарубежными специалистами разделения

²⁵⁶ Свитич Л. Г., Смирнова О. В., Ширяева А. А., Шкондин М. В. и др. Газеты средних и малых городов России в 2010-х гг. Контент-аналитическое исследование. Коллективная монография. М.: Фак. журн. МГУ, 2016.

журналистики данных «обычную» и «основательную», в которой большая часть текстов укладывается в первую категорию.

Такие направления исследовательского интереса могут стать основой для развития эмпирических исследований и теоретического осмысления отечественными специалистами темпов и масштабов изменений, происходящих в журналистике под влиянием цифровизации и производной от неё медиаконвергенции. Изучение их влияния на журналистские культуры²⁵⁷, рабочие практики²⁵⁸ и конечный продукт – медиатексты²⁵⁹ показывает, что вызванные цифровизацией изменения в журналистике происходят медленнее и точнее, чем часто представляется. На наш взгляд, все более актуальной становится интеграция в отечественный научный дискурс и влияния датафикации, как ещё одного производного от цифровизации, на журналистскую работу и тексты.

Использование результатов обработки больших данных в журналистике потенциально должно оказывать влияние на сегмент журналистики данных и формировать у неё определенные специфические черты. Полученный нами опыт изучения журналистских практик и основных характеристик опубликованных текстов позволяет, с одной стороны, считать,

²⁵⁷ Anikina M. E. (2014) Journalism as a Profession in the First Decades of the 21st Century: the Russian Context. *World of Media. Journal of Russian Media and Journalism Studies*: 233-252; Anikina M., Hanitzsch T. et al. (2011) Mapping Journalism Cultures Across Nations. *Journalism Studies*, No. 12(3): 273-293.

²⁵⁸ Вырковский А. В., Галкина М. Ю., Колесниченко А. В., Образцова А. Ю. Конвергенция в работе российского журналиста // *Медиаскоп* (электронный журнал). 2016. Вып.1; Вырковский А. В., Галкина М. Ю., Колесниченко А. В., Образцова А. Ю., Варганов С. А. Мультимедийные элементы в современном медиатексте // *Медиаскоп* (электронный журнал). 2017. Вып. 3; Баранова Е. А. Новые журналистские компетенции в условиях медиаконвергенции: мифы и реальность // *Вестник РУДН. Серия: Литературоведение. Журналистика*. 2017. Т. 22. № 1. С. 177-188.

²⁵⁹ Колесниченко А.В. Востребованность жанров журналистских текстов в онлайн-СМИ // *Вестн. Моск. ун-та. Серия 10: Журналистика*. 2018. № 1. С. 26-42; Вырковский А. В., Галкина М. Ю., Колесниченко А. В., Образцова А. Ю. Жанровая структура работы журналиста // *Вестник Воронежского государственного университета. Серия: Филология. Журналистика*. 2016. № 2. С. 86-92; Фомичева И. Д., Образцова А. Ю. и др. Анализ качества медиатекстов на столичных информационно-разговорных радиостанциях // *Вестник Санкт-Петербургского университета. Язык и литература*. 2019. Т. 16. № 1. С. 48-65.

что по ключевым критериям даже условная «журналистика больших данных» в России практически не выбивается сегодня из актуальных тенденций развития журналистики в целом и вызывает такие же плавные и точечные трансформации. При этом, с другой стороны, все-таки отмеченные на настоящем этапе отдельные специфические черты могут говорить о потенциале исследований эффектов датафикации при выявлении воздействия цифровизации на современную журналистику.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Баранова Е. А. Конвергенция СМИ устами журналистов-практиков. – М.: Изд-во «Прометей», 2017.
2. Вартанова Е.Л. Медиаэкономика зарубежных стран: Учеб. пособие. — М.: Аспект Пресс, 2003. — 335 с.
3. Вартанова Е. Л. Теория СМИ: Актуальные вопросы. – М.: МедиаМир, 2009. – 488 с.
4. Вершинская О.Н. Информационно-коммуникационные технологии и общество. – М.: Наука, 2007. – 203 с.
5. Воройский Ф. С. Информатика. Энциклопедический словарь-справочник: введение в современные информационные и телекоммуникационные технологии в терминах и фактах. – М.: Физматлит, 2006. – 768 с.
6. Гиляревский Р.С., Михайлов А.И., Черный А.И. Основы научной информации. – М.: Наука, 1965. – 655 с.
7. Грей Д. Пособие по журналистике данных. – М.: РИА «Новости», 2013. – 240 с.
8. Дзялошинский И.М., Пильгун М.А. Медиатекст: особенности создания и функционирования. – М.: НИУ ВШЭ, 2011. – 377 с.
9. Засурский Я.Н. Система средств массовой информации России. – М.: Аспект Пресс, 2003. – 259 с.
10. Засурский Я.Н. Коммуникация в обществе знаний. Российская журналистика: 2008 – 2013. – М.: МедиаМир, 2013.
11. Иваницкий В.Л. Модернизация журналистики: методологический этюд. – М.: Изд-во Моск. ун-та; ф-т журн. МГУ, 2010. – 360 с.
12. Интернет-СМИ: Теория и практика / Под ред. М.М. Лукиной. – М.: Аспект Пресс, 2010. – 348 с.
13. Кастельс М. Информационная эпоха: экономика, общество и культура. – М.: ГУ ВШЭ, 2000. – 608 с.

14. Кастельс М. Галактика Интернет: Размышления об Интернете, бизнесе и обществе. – Екатеринбург: У-Фактория, 2004. – 328 с.
15. Колесниченко А. В. Практическая журналистика. Учеб. пособие. – М.: Изд-во Моск. ун-та, 2008.
16. Кульчицкая Д. Ю., Галустян А. А. Лонгриды в онлайн-СМИ. Особенности и технология создания. Учебное пособие. М: Аспект Пресс. 2016.
17. Луман Н. Реальность массмедиа. – М.: Праксис, 2005. – 256 с.
18. Олешко Е. В. Конвергентная журналистика: Профессиональная культура субъектов информационной деятельности: учебное пособие / Екб: Изд-во Уральского университета, 2015. – 128 с.
19. Разумова М. А. Тематические приоритеты и база источников информации деловых СМИ (на примере газет "Коммерсантъ" и "Ведомости"): диссертация ... кандидата филологических наук: 10.01.10 / Место защиты: ФГБОУ ВО «Московский государственный университет имени М. В. Ломоносова», 2016.- 182 с.
20. СМИ в меняющейся России / Под ред. Е.Л. Варгановой. – М.: Аспект Пресс, 2010. – 336 с.
21. Уэбстер Ф. Теории информационного общества. – М.: Аспект Пресс, 2004. – 400 с.
22. Bogost I., Ferrari S. and Schweizer B. Newsgames: Journalism at play. – Cambridge, Massachusetts: MIT Press, 2010.
23. Computer-Mediated Communication. – London: Sage Publishing, 2004. – 234 p.
24. Felle T., Mair J., and Radcliffe D. Data Journalism: Inside the global future. – UK: Abramis, 2015.
25. Fry B. Visualizing Data: Exploring and Explaining Data with the Processing Environment. – London: O'Reilly Media; 1 edition, 2008.

- 26.Gray J., Bounegru L. and Chambers L. (eds) The Data Journalism Handbook: How Journalists Can Use Data to Improve the News. – Sebastopol, CA: O'Reilly, 2012.
- 27.Harpold T. Conclusions. Hypertext theory. – Baltimore: John Hopkins University Press, 1994.
- 28.Kovach B. and Rosenstiel T. The Elements of Journalism: What Newspeople Should Know and the Public Should Expect Completely Updated and Revised. New York: Three Rivers Press; Revised, Updated edition (April 24, 2007).
- 29.McNair B. The sociology of journalism. – UK: Arnold London, 1998.
- 30.Mair J., Keeble R. L., Lucero M. Data Journalism: Past, Present and Future. – UK: Abramis, 2017.
- 31.Martin W. J. The Information Society. – London: Aslib, 1988. – 289 p.
- 32.Masuda Y. The Information Society as Postindustrial Society. – Washington: World Future Soc., 1983. – 212 p.
- 33.Mayer-Schönberger V. and Cukier K. Big Data: A Revolution That Will Transform How We Live, Work, and Think. – UK: John Murray, 2013.
- 34.Meyer P. Precision Journalism: A Reporter's Introduction to Social Science Methods. – Oxford: Rowman & Littlefield, 2002.
- 35.Rogers S. Facts are Sacred: the power of data. – London: Faber&Faber, 2013.
- 36.Rogers R. Digital Methods. – Cambridge, Massachusetts: MIT Press, 2013.

Научные статьи и тезисы научных конференций

- 37.Баранова Е. А. Новые журналистские компетенции в условиях медиаконвергенции: мифы и реальность // Вестник РУДН. Серия: Литературоведение. Журналистика. 2017. Т. 22. № 1. С. 177-188.

38.Бочаров М. П. PR-технологии и журналистика данных. Новые формы работы со статистическими и фактологическими источниками информации о социальной сфере» // Труд и социальные отношения. 2014. № 4.

39. Вартанов С. А. Большие данные в онлайн-СМИ: подходы и стратегии использования // Медиаскоп. 2017. Вып. 4. Режим доступа: <http://www.mediascope.ru/2375>

40. Вартанов С.А. Телевизионные измерения в эпоху Big Data: концепции и примеры // Вестн. Моск. ун-та. Сер.10, Журналистика. 2017. № 3. С. 37–57.

41.Вартанова Е. Л. СМИ и журналистика в пространстве постиндустриального общества // Медиаскоп. – 2009. – Вып. 2. – Режим доступа: <http://mediascope.ru/node/73-0420900082/0031>

42.Вартанова Е. Л. Журналистика в информационном обществе: новые возможности и новые вызовы / Материалы круглого стола. – М.: Институт развития информационного общества, 2005. – С. 36 – 48.

43.Вартанова Е. Л. К чему ведет конвергенция СМИ? // Информационное общество, 1999, вып. 5, с. 11 - 14. Режим доступа: <http://emag.iis.ru/arc/infosoc/emag.nsf/BPA/b59df6463a315de4c32568fd0038da32>

44.Вырковский А. В., Галкина М. Ю., Колесниченко А. В. и др. Трансформация журналистской работы под влиянием новых технологий: поиск информации, жанры медиатекстов, редакционная культура // Вестник Московского университета. Серия 10: Журналистика. — 2017. — № 5. — С. 51–71.

45.Вырковский А. В., Галкина М. Ю., Колесниченко А. В., Образцова А. Ю. Конвергенция в работе российского журналиста // Медиаскоп (электронный журнал). 2016. Вып.1.

46.Вырковский А. В., Галкина М. Ю., Колесниченко А. В., Образцова А. Ю. Жанровая структура работы журналиста // Вестник Воронежского государственного университета. Серия: Филология. Журналистика. 2016а. № 2. С. 86-92.

47.Вырковский А. В., Галкина М. Ю., Колесниченко А. В., Образцова А. Ю., Вартанов С. А. Мультимедийные элементы в современном медиатексте // Медиаскоп (электронный журнал). 2017. Вып. 3.

48.Дедова О. Графическая неоднородность как категория гипертекста // Вестник Московского университета. Серия 9: Филология, издательство Изд-во Моск. ун-та (М.), № 6, с. 91-103.

49.Канаракус К. Машина Больших Данных. Сети. – М.: Открытые системы, 2011. — № 4.

50.Колесниченко А.В. Востребованность жанров журналистских текстов в онлайн-СМИ // Вестник Московского университета. Серия 10: Журналистика, издательство Изд-во Моск. ун-та (М.), 2018, № 1, с. 26-42.

51.Макеенко М., Вырковский А. Конвергенция в российской ежедневной прессе: экономические особенности и перспективы // Вестник Московского университета. Серия 10. Журналистика. — 2012. — № 5.

52.Моррисон А. Большие Данные: как извлечь из них информацию // «Технологический прогноз» - 2010. – PricewaterhouseCoopers. Вып. №3.

53. Николов А. Л. Новые направления в медиаиндустрии, или Что такое дата-журналистика // Медиа. Информация. Коммуникация. № 4, 2012. Режим доступа: <http://mic.org.ru/index.php/new/165-novye-napravleniya-v-mediaindustrii-ili-chto-takoe-data-zhurnalistika-data-journalism>

54.Панюкова С. А. Роль открытых данных в развитии журналистики данных // Знак: проблемное поле медиаобразования. 2015. №1 (15). Режим доступа: <https://cyberleninka.ru/article/n/rol-otkrytyh-dannyh-v-razvitiizhurnalistiki-dannyh>

55.Свитич Л. Г., Смирнова О. В., Ширяева А. А., Шкондин М. В. и др. Газеты средних и малых городов России в 2010-х гг. Контент-аналитическое исследование. Коллективная монография. М.: Фак. журн. МГУ, 2016.

56.Симакова С. И. Дата-журналистика как медиатренд // Вестник ННГУ. 2014. №2-2. Режим доступа: <https://cyberleninka.ru/article/n/data-zhurnalistika-kak-mediatrend>

57. Симакова С. И. Журналистика данных как объективное требование времени и её влияние на формирование визуальной журналистики // Знак: проблемное поле медиаобразования. 2016, №1 (18). Режим доступа: <https://cyberleninka.ru/article/n/zhurnalistika-dannyh-kak-obektivnoe-trebovanie-vremeni-i-eyo-vliyanie-na-formirovanie-vizualnoy-zhurnalistiki>
58. Субботин М. Гипертекст: новая форма письменной коммуникации. Итоги науки и техники // Информатика. ВИНТИ. – М., 1994. – Т. 18.
59. Черняк Л. Большие Данные — новая теория и практика // Открытые системы. СУБД. — М.: Открытые системы, 2011. — № 10.
60. Шерстюкова М. Н. Дата-журналистика как новое направление в системе средств массовой коммуникации // Медиа. Информация. Коммуникация. 1 номер, 2012. Режим доступа: <http://mic.org.ru/1-nomer-2012/65-1-shestukova->
61. Шилина А. Г. Дата журнализм в газете *The Guardian*: особенности модели новейшего времени // Mass-media. Действительность. Литература, издательство Твер. гос. ун-т (Тверь), том 16, с. 77-83. 2016.
62. Шилина А. Г. Журналистика данных в качественных российских журналах: опыт идентификации // Вестник Тверского государственного университета. Серия Филология, № 3, с. 222-228. 2016.
63. Шилина М. Г. Data Journalism – дата-журналистика, журналистика метаданных как новый формат медиакommunikации: к вопросу формирования исследовательских подходов // Медиаскоп. Вып. 1. – М., 2013. Режим доступа: <http://www.mediascope.ru/node/1263>
64. Шилина М. Г. Big Data, Open Data, Linked Data, метаданные в PR: актуальные модели трансформации теории и практики // Медиаскоп. Вып. №1. – М., 2014. Режим доступа: <http://www.mediascope.ru/node/1486>
65. Anderson C. W. (2013) Towards a sociology of computational and algorithmic journalism. *New Media & Society* 15(7): 1005–1021.

66. Anderson C. W. (2015) Between the unique and the pattern: Historical tensions in our understanding of quantitative journalism. *Digital Journalism* 3(3): 349–363.

67. Anderson B., Borges-Rey E. (2019) Encoding the UX: User Interface as a Site of Encounter between Data Journalists and Their Constructed Audiences. *Digital Journalism*. DOI: 10.1080/21670811.2019.1607520

68. Aitamurto T., Sirkkunen E. and Lehtonen P. (2011) Trends in data journalism. Available at: http://virtual.vtt.fi/virtual/nextmedia/Deliverables_2011/D3.2.1.2.B_Hyperlocal_Trends_In%20Data_Journalism.pdf

69. Anderson C. W. (2016) Social survey reportage: Context, narrative, and information visualization in early 20th century American journalism. *Journalism*, 2016. 18:1, 81-100.

70. Alexe B., Chiticariu L., Miller R. J., Tan W. C. Muse: Mapping understanding and design by example. IEEE 24th International Conference on Data Engineering. – Cancun, Mexico, 2008.

71. Appelgren E. and Nygren G. (2014) Data journalism in Sweden – Opportunities and challenges: A case study of Brottspejl at Sveriges Television (SVT). In: Mills C, Pidd M and Ward E (eds) *Proceedings of the Digital Humanities Congress 2012*. Sheffield: HRI Online Publications. Available at: <https://www.hrionline.ac.uk/openbook/chapter/dhc2012-appelgren>

72. Appelgren E. and Nygren G. (2014) Data journalism in Sweden: Introducing new methods and genres of journalism into ‘old’ organizations. *Digital Journalism*, 2(3): 394–405.

73. Appelgren E. (2017) An Illusion of Interactivity. The paternalistic side of data journalism. *Journalism Practice*, 12:3, 308-325.

74. Appelgren E. and Salaverria R. (2018) The Promise of the Transparency Culture, *Journalism Practice*, 12:8, 986-996.

75. Appelgren E. (2019) Remaining in Control with an Illusion of Interactivity: The Paternalistic Side of Data Journalism. *Journalism Practice*, 13:8, 956-960.

76.Auer S. R., Bizer C., Kobilarov G., Lehmann J., Cyganiak R, Ives Z. (2007) DBpedia: A Nucleus for a Web of Open Data. *The Semantic Web*. Lecture Notes in Computer Science. DOI:10.1007/978-3-540-76298-0_52.

77.Ausserhofer J. (2015) ‘Die Methode liegt im Code’: Routinen und digitale Methoden im Datenjournalismu. In: Maireder A, Ausserhofer J, Schumann C and et al. (eds) *Digitale Methoden in der Kommunikationswissenschaft*. Berlin: Digital Communication Research, pp. 87–111.

78.Ausserhofer J., Gutounig R., Oppermann M., et al. (2017) The datafication of data journalism scholarship: Focal points, methods, and research propositions for the investigation of data-intensive newswork. *Journalism*. Epub ahead of print. DOI: 10.1177/1464884917700667.

79.Baack S. (2013) A new style of news reporting: Wikileaks and data-driven journalism. In: Rambatan B and Johanssen J (eds) *Cyborg Subjects: Discourses on Digital Culture*. CreateSpace Independent Publishing, pp. 113–122.

80.Baack S. (2015) Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data & Society*, 2015 2:2. Available at: <http://journals.sagepub.com/doi/full/10.1177/2053951715594634>

81.Bakker P. (2014) Mr. Gates returns: Curation, community management and other new roles for journalists. *Journalism Studies*, 15(5): 596–606.

82.Bardoel J. and Deuze M. (2001) Network Journalism: converging competences of old and new media professionals. *Australian Journalism Review*, 23(2), 91-103.

83.Bates J. (2012) ‘This is what modern deregulation looks like’: Co-optation and contestation in the shaping of the UK’s Open Government Data Initiative. *The Journal of Community Informatics*, 8(2).

84.Boczkowski P. J. (2004) The processes of adopting multimedia and interactivity in three online newsrooms. *Journal of Communication*, 54(2): 197–213.

85. Borges-Rey E. (2016) Unravelling Data Journalism. *Journalism Practice*, 10(7): 833-843.
86. Borges-Rey E. (2017) Towards an epistemology of data journalism in the devolved nations of the United Kingdom: Changes and continuities in materiality, performativity and reflexivity. *Journalism*. DOI: 10.1177/1464884917693864.
87. Borges-Rey E. and Stalph F. (2018) A random walk through data + journalism: assessing the data journalism ecosystem. Paper presented on NODA-2018 (The Nordic Data Journalism Conference). Södertörn University: Sweden, 15-17 Mar. 2018.
88. Boyd D., and Crawford K. (2012) Critical questions for Big Data. *Information, Communication & Society*, 15(5): 662–679.
89. Boyles J. L., Meyer E. (2016) Letting the data speak: Role perceptions of data journalists in fostering democratic conversation. *Digital Journalism* 4(7): 944–954.
90. Boyles J. L. (2017) Laboratories for news? Experimenting with journalism hackathons. *Journalism*, Published October 17, 2017 <https://doi.org/10.1177/1464884917737213>
91. Bradshaw P. How to be a data journalist // *The Guardian datablog*. – 2010, 1 Oct. Available at: <http://www.theguardian.com/news/Datablog/2010/oct/01/data-journalism-how-to-guide>
92. Bradshaw P. (2011) The inverted pyramid of data journalism. Available at: <https://onlinejournalismblog.com/2011/07/07/the-inverted-pyramid-of-data-journalism/>
93. Bradshaw P. (2014) Scraping for Journalists: Leanpub. 20 Journalism.
94. Brandão R. F. (2019) Challenging Data-Driven Journalism. *Journalism Practice*, 13(8): 927-930.
95. Broussard M. and Boss K. (2018) Saving Data Journalism. *Digital Journalism*, 6(9): 1206-1221.

96. Bunz M. (2011) Das offene Geheimnis: Zur Politik der Wahrheit im Datenjournalismus. In: Geiselberger H (ed.), Wikileaks und die Folgen: Netz - Medien - Politik. Berlin: Suhrkamp: 134–151.

97. Cheruiyot D., Baack S. and Ferrer-Conill R. (2019) Data Journalism Beyond Legacy Media: The case of African and European Civic Technology Organizations. *Digital Journalism*. DOI: 10.1080/21670811.2019.1591166

98. Chignard S. A brief history of Open Data. *ParisTechReview* – 2013, 29 March. Available at: <http://www.paristechreview.com/2013/03/29/brief-history-open-data/>

99. Coddington M. (2015) Clarifying journalism's quantitative turn: A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digital Journalism* 3(3): 331–348.

100. Cohen N. S. (2015) From pink slips to pink slime: Transforming media labor in a digital age. *The Communication Review* 18(2): 98–122.

101. Cohen S., Hamilton J. T. and Turner F. (2011) Computational journalism: How computer scientists can empower journalists, democracy's watchdogs, in the production of news in the public interest. *Communications of the ACM* 54(10): 66–71.

102. Cohen S., Li C., Yang J., et al. (2011) Computational journalism: A call to arms to database researchers. In: Proceedings of the 5th Biennial Conference on Innovative Data Systems Research, Asilomar, CA, 9–12 January: 148–151.

103. Cottle S. and Ashton M. (1999) From BBC newsroom to BBC newscentre: On changing technology and journalist practices. *Convergence*, 5(3), 22-43.

104. Cox M. (2000) The development of computer-assisted reporting. In: Proceedings of the Newspaper Division of the Association for Education in Journalism and Mass Communication's Southeast Colloquium: Chapel Hill, NC, 17–18 March.

105. Cushion S., Lewis J. and Callaghan R. (2017) Data Journalism, Impartiality And Statistical Claims. *Journalism Practice*, 11(10): 1198-1215.
106. Davenport L., Fico F. and Detwiler M. (2000) Computer-assisted reporting in Michigan daily newspapers: More than a decade of adoption. In: Proceedings of the national convention of the association for education in journalism and mass communication: Phoenix, AZ, 9–12 August.
107. Davenport L., Fico F. and Weinstock D. (1996) Computers in newsrooms of Michigan's newspapers. *Newspaper Research Journal* 17(3–4): 14–28.
108. De Maeyer J., Libert M., Domingo D., et al. (2015) Waiting for data journalism: A qualitative assessment of the anecdotal take-up of data journalism in French-speaking Belgium. *Digital Journalism* 3(3): 432–446.
109. Diakopoulos N. (2011) A functional roadmap for innovation in computational journalism. Available at: <http://www.nickdiakopoulos.com/2011/04/22/a-functional-roadmap-for-innovation-in-computational-journalism/>
110. Diakopoulos N. (2012) Cultivating the landscape of innovation in computational journalism. CUNY Graduate School of Journalism & Tow-Knight Center for Entrepreneurial Journalism. Available at: http://www.nickdiakopoulos.com/wp-content/uploads/2012/05/diakopoulos_whitepaper_systematicinnovation.pdf
111. Diakopoulos N. (2014) Algorithmic Accountability. Journalistic investigation of computational power structures. *Digital Journalism* 3(3): 398-415.
112. Dick M. (2014) Interactive infographics and news values. *Digital Journalism* 2(4): 490–506.
113. European Journalism Centre (ed.) (2010) Data-Driven Journalism: What Is There to Learn? Amsterdam: European Journalism Center. Available at: http://mediapusher.eu/datadrivenjournalism/pdf/ddj_paper_final.pdf

114. Fairfield J. and Shtein H. (2014) Big data, big problems: Emerging issues in the ethics of data science and journalism. *Journal of Mass Media Ethics* 29(1): 38–51.
115. Fan Yang, Ying Roselyn Du (2016) Storytelling in the Age of Big Data: Hong Kong Students' Readiness and Attitude towards Data Journalism. *Asia Pacific Media Educator* 26(2): 148-162.
116. Fecher B., Friesike S. and Hebing M. (2015) What drives academic data sharing? *PLoS ONE* 10(2).
117. Felle T. (2016) Digital watchdogs? Data reporting and the news media's traditional 'fourth estate' function. *Journalism* 17(1): 85–96.
118. Fink K. and Anderson C.W. (2015) Data journalism in the United States: Beyond the 'usual suspects'. *Journalism Studies* 16(4): 467–481.
119. Flew T., Daniel A. and Spurgeon C. L. (2010) The promise of computational journalism. In: *Media, democracy and change: Refereed proceedings of the Australian and New Zealand Communication Association annual conference* (ed K McCallum), Canberra, ACT, Australia, 7–9 July.
120. Flew T., Spurgeon C., Daniel A., et al. (2012) The promise of computational journalism. *Journalism Practice* 6(2): 157–171.
121. Garrison B. (1999) Newspaper size as a factor in use of computer-assisted reporting. *Newspaper Research Journal* 20(3): 72-85.
122. Godler Y. and Reich Z. (2013) How journalists think about facts. *Journalism Studies* 14(1): 94–112.
123. Graham C. (2015) By the Numbers. Data Journalism Projects as a Means of Teaching Political Investigative Reporting. *Asia Pacific Media Educator* 25(2): 247-261.
124. Guthrie G. and Murthy V. (2009) Past, present and possible future developments in human capital accounting: A tribute to Jan-Erik Gröjer. *Journal of Human Resource Costing & Accounting* 13(2): 125–142.

125. Gynnild A. (2014) Journalism innovation leads to innovation journalism: The impact of computational exploration on changing mindsets. *Journalism* 15(6): 713–730.
126. Hamilton J. T. and Turner F. (2009) Accountability through algorithm: Developing the field of computational journalism. In: Proceedings of the center for advanced study in the behavioral sciences summer workshop, Palo Alto, CA. Available at: <http://web.stanford.edu/~fturner/Hamilton%20Turner%20Acc%20by%20Alg%20Final.pdf>
127. Hannaford L. (2015) Computational journalism in the UK newsroom: Hybrids or specialists? *Journalism Education* 4(1): 6–21.
128. Heravi B. R. (2019) 3Ws of Data Journalism Education. *Journalism Practice*, 13(3): 349-366.
129. Hermida A. and Young M. L. (2017) Finding the Data Unicorn. *Digital Journalism*, 5(2): 159-176.
130. Hermida A., Young M. L. and Fulda J. (2018) What Makes for Great Data Journalism? *Journalism Practice* 12(1): 115-135.
131. Hermida A., Young M. (2019). Data Journalism and the Regeneration of News. London: Routledge. Available at: <https://doi.org/10.4324/9781315163895>
132. Hewett J. (2015) Learning to teach data journalism: Innovation, influence and constraints. *Journalism* 17(1): 119-137.
133. Holovaty A. (2006) A fundamental way newspaper sites need to change. Available at: <http://www.holovaty.com/writing/fundamental-change/>
134. Howard A. B. (2014) The Art and Science of Data-Driven Journalism. New York: Tow Center for Digital Journalism. Available at: <http://towcenter.org/wp-content/uploads/2014/05/TowCenter-Data-Driven-Journalism.pdf>
135. Hullman J., Diakopoulos N., Momeni E., et al. (2015) Content, context, and critique: Commenting on a data visualization blog. In: Proceedings of

the 18th ACM conference on computer supported cooperative work & social computing, Vancouver, BC, Canada, 14–18 March, pp. 1170–1175. New York: ACM.

136. Jenkins J. and Volz Y. (2016) Players and contestation mechanisms in the journalism field: A historical analysis of journalism awards, 1960s to 2000s. *Journalism Studies*. Epub ahead of print. DOI: 10.1080/1461670X.2016.1249008.

137. Johnson J.A. (2014) From open data to information justice. *Ethics and Information Technology* 16(4): 263–274.

138. Jungherr A. (2016) Twitter use in election campaigns: A systematic literature review. *Journal of Information Technology & Politics* 13(1): 72–91.

139. Kaefer F., Roper J. and Sinh P. (2015) A software-assisted qualitative content analysis of news articles: Example and reflections. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* 16(2). Available at: <http://www.qualitative-research.net/index.php/fqs/article/view/2123>

140. Karlsen J. and Stavelin E. (2014) Computational journalism in Norwegian newsrooms. *Journalism Practice* 8(1): 34–48.

141. Knight, M. (2015) Data journalism in the UK: A preliminary analysis of form and content. *Journal of Media Practice* 16(1): 55–72.

142. Lecheler S. and Kruikemeier S. (2016) Re-evaluating journalistic routines in a digital age: A review of research on the use of online sources. *New Media & Society* 18(1): 156–171.

143. Lewis S. C. (2015) Journalism in an era of big data. *Digital Journalism* 3(3): 321–330.

144. Lewis S. C. and Usher N. (2013) Open source and journalism: Toward new frameworks for imagining news innovation. *Media, Culture & Society* 35(5): 602–619.

145. Lewis, S. C. and Usher, N. (2014) Code, collaboration, and the future of journalism: A case study of the hacks/hackers global network. *Digital Journalism* 2(3): 383–393.

146. Lewis S. C. and Westlund O. (2015) Big Data and Journalism: Epistemology, expertise, economics, and ethics. *Digital Journalism*. 3(3): Journalism in an Era of Big Data: Cases, Concepts, and Critiques.
147. Lewis N. P. and Al Nashmi E. (2019) Data Journalism in the Arab Region: Role Conflict Exposed. *Digital Journalism*. DOI: 10.1080/21670811.2019.1617041
148. Lorenz M., Kayser-Bril N., McGhee, G. (2011) Media companies must become trusted data hubs. Owni.eu. Available at: <http://www.owni.eu/2011/28/02/media-companies-must-become-trusted-data-hubs-catering-to-the-trust-market/>
149. Loosen W., Reimer J. and Schmidt F. (2015) When data become news: A content analysis of data journalism pieces. In: Proceedings of the future of journalism 2015 conference, Cardiff, 10–11 September.
150. Loosen W., Reimer J. and De Silva-Schmidt, F. (2017) Data-driven reporting: An on-going (r)evolution? An analysis of projects nominated for the Data Journalism Awards 2013–2016. *Journalism*. Available at: <https://doi.org/10.1177/1464884917735691>
151. Lopez P. (2009) GROBID: Combining automatic bibliographic data recognition and term extraction for scholarship publications. In: Agosti, M., Borbinha, J., Kapidakis, S., et al. (eds) *Research and Advanced Technology for Digital Libraries*. Berlin: Springer: 473–474.
152. Lugo-Ocando J. and Brandão R. F. (2015) Stabbing news: Articulating crime statistics in the newsroom. *Journalism Practice* 10: 715–726.
153. De Maeyer J., Libert M., Domingo D., Heinderyckx F. & Le Cam F. (2015) Waiting for Data Journalism. A qualitative assessment of the anecdotal take-up of data journalism in French-speaking Belgium. *Digital Journalism* 3(3): 432-446.
154. Maireder A., Ausserhofer J., Schumann C., et al. (eds) (2015) *Digitale Methoden in der Kommunikationswissenschaft*. Berlin: *Digital Communication Research*. Available at: <https://doi.org/10.17174/dcr.v2.0>

155. Manyika J. et al. (2011) Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute, June, 2011. McKinsey (9 August 2011).
156. Markham A. (2013) Remix cultures, remix methods: Reframing qualitative inquiry for social media contexts. In: Denzin NK and Giardina MD (eds) *Global Dimensions of Qualitative Inquiry*. Walnut Creek, CA: Left Coast Press: 63–81.
157. Mayernik M. S. (2017) Open data: Accountability and transparency. *Big Data & Society*. Available at: <https://doi.org/10.1177/2053951717718853>.
158. Nguyen A. and Lugo-Ocando J. (2016) The state of data and statistics in journalism and journalism education: issues and debates. *Journalism* 17(1): 3–17.
159. Neuberger C., Nuernbergk C. and Rischke M. (2007) Weblogs und Journalismus: Konkurrenz, Ergänzung oder Integration? *Media Perspektiven* 2007(2): 96–112.
160. Nippard C. (2011) Data visualizations emerge across newsrooms, online publishers. *Deutsche Welle*. Available at: <http://www.dw-world.de>.
161. Örnebring H. (2010) Technology and journalism-as-labour: Historical perspectives. *Journalism*, 11(1): 57-74.
162. Ojo A. and Heravi B. (2018) Patterns in Award Winning Data Storytelling. *Digital Journalism*, 6(6): 693-718.
163. Palomo B., Teruel L. and Blanco-Castilla E. (2019) Data Journalism Projects Based on User-Generated Content. How La Nacion Data Transforms Active Audience into Staff. *Digital Journalism*. DOI: 10.1080/21670811.2019.1626257
164. Parasie S. (2015) Data-driven revelation? Epistemological tensions in investigative journalism in the age of ‘big data’. *Digital Journalism* 3(3): 364–380.
165. Parasie S. and Dagiral E. (2013) Data-driven journalism and the public good: ‘Computer-assisted reporters’ and ‘programmer-journalists’ in Chicago. *New Media & Society* 15(6): 853–871.

166. Pavlik V. (2000) The impact of technology on journalism. *Journalism Studies* 1(2): 229–237.
167. Pilhofer A. (2010) Programmer-Journalist? Hacker-Journalist? Our Identity Crisis. Media Shift – Idea Lab. Available at: <http://www.pbs.org/idealab/2010/04/programmer-journalist-hacker-journalist-our-identity-crisis107>
168. Porlezza C. (2016) Data Journalism in Italy. Between closed data, networked journalism and ethical conflicts. IAMCR Conference – Leicester, UK. Available at: https://prezi.com/p_gvprilp_uo/accountability-and-transparency-in-data-journalism-the-case-of-italy/
169. Porlezza C. and Splendore S. (2019) From Open Journalism to Closed Data: Data Journalism in Italy. *Digital Journalism*. DOI: 10.1080/21670811.2019.1657778
170. Powers M. (2012) In Forms That Are Familiar and Yet-to-Be Invented: American Journalism and the Discourse of Technologically Specific Work. *Journal of Communication Inquiry*, 36 (1): 24–43.
171. Radchenko I. and Sakoyan A. (2014) The view on open data and data journalism: Cases, educational resources and current trends. In: Ignatov D.I., Khachay M.Y., Panchenko A., et al. (eds) *Analysis of Images, Social Networks and Texts*. Cham: Springer, pp. 47–54.
172. Rogers S. (2011) Data journalism at the Guardian: what is it and how do we do it? // *The Guardian*, – 28 Jul 2011. Available at: <https://www.theguardian.com/news/Datablog/2011/jul/28/data-journalism>
173. Rogers S. (2008) Turning official figures into understandable graphics, at the press of a button // *The Guardian*, – 18 Dec 2008. Available at: <http://www.theguardian.com/help/insideguardian/2008/dec/18/unemploymentdata>
174. Rogers S., Schwabish, J. and Bowers, D. (2017) The state of data journalism in 2017 // *Google Blog*. Available at: <https://www.blog.google/topics/journalism-news/data-journalism-2017/>
<https://newslab.withgoogle.com/assets/docs/data-journalism-in-2017.pdf>

175. Royal C. (2012) The journalist as programmer: A case study of the New York Times Interactive News Technology department. *ISOJ Journal* 2(1).
176. Royal C. and Blasingame D. (2015) Data journalism: An explication. *ISOJ* 5(1): 24–46.
177. Schudson M. (1989). The sociology of news production. *Media, culture and society*, 11(3): 263-282.
178. Schudson M. (2005) Four approaches to the sociology of news. In: Curran, J. and Gurevitch, M. (eds), *Mass Media and Society*, 4th edn. London: Hodder Arnold: 172–197.
179. Schudson M. (2010) Political observatories, databases & news in the emerging ecology of public information. *Daedalus* 139(2): 100–109.
180. Segel E. and Heer J. (2010) Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics* 16(6): 1139–1148.
181. Sirkkunen E., Aitamurto T. and Lehtonen P. (2011) Trends in Data Journalism. Finnish Strategic Centre for Science, Technology and Innovation in the field of ICT. TIVIT.
182. Smit G., de Haan Y. and Buijs L. (2014) Visualizing news: Make it work. *Digital Journalism* 2(3): 344–354.
183. Stalph F. (2017) Classifying Data Journalism. A content analysis of daily data-driven stories. *Journalism Practice*, 12(10): 1332-1350.
184. Stalph F. and Borges-Rey E. (2018) Data Journalism Sustainability. *Digital Journalism*, 6(8): 1078-1089.
185. Stavelin E. (2013) Computational journalism: When journalism meets programming. PhD Dissertation, University of Bergen, Bergen.
186. Steensen S. and Ahva L. (2015) Theories of journalism in a digital age. *Journalism Practice* 9(1): 1–18.
187. Shilina A. (2017) Data journalism and ethics: a Russian approach. *Russian Journal of Communication*, 9(3): 299-300.

188. Splendore S., Di Salvo P., Eberwein T., Groenhart H., Kus M., Porlezza C. (2015) Educational strategies in data journalism: A comparative study of six European countries. *Journalism* 17(1): 138-152.
189. Tabary C., Provost A.-M. and Trottier, A. (2016) Data journalism's actors, practices and skills: A case study from Quebec. *Journalism* 17(1): 66–84.
190. Tandoc E. C. Jr and Oh S.-K. (2015) Small departures, big continuities? Norms, values, and routines in *The Guardian's big data journalism*. *Journalism Studies*. Epub ahead of print 5 November. DOI: 10.1080/1461670X.2015.1104260.
191. Thibodeau P. Gartner's Top 10 IT challenges include exiting baby boomers, Big Data. Computerworld, 2011. Available at: <http://www.gartner.com/technology/home.jsp>
192. Tung-Mou Yang, Jin Lo, and Jing Shiang (2015) To open or not to open? Determinants of open government data. *Journal of Information Science*, 41(5): 596 – 612.
193. Usher N. (2016) *Interactive Journalism: Hackers, Data, and Code*. Urbana, IL: University of Illinois Press.
194. Uskali T. and Appelgren E. (2015) The third Nordic data journalism conference: Call for papers. Available at: <http://www.noda2016.fi/callforpapers.html>
195. Uskali T. and Kuutti H. (2015) Models and streams of data journalism. *The Journal of Media Innovations* 2(1): 77–88.
196. Valeeva A. (2017) *Open Data in a Closed Political System: Open Data Investigative Journalism in Russia*. Reuters Institute for the Study of Journalism. University of Oxford. Available at: http://reutersinstitute.politics.ox.ac.uk/sites/default/files/2017-10/Valeeva_Open%20data%20and%20investigative%20journalism%20in%20Russia.pdf
197. Van Dijck J. (2014) Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society* 12(2): 197–208.

198. Van der Haak, Bregtje, Parks, M. and Castells, M. The Future of Journalism: Networked Journalism. *International Journal of Communication*. Web, 2014. Available at: <http://ijoc.org/index.php/ijoc/article/view/1750/832>
199. Weber W. and Rall H. (2012) Data visualization in online journalism and its implications for the production process. In: Proceedings of the 16th international conference on information visualization (IV), 11–13 July, pp. 349–356. Montpellier: IEEE.
200. Weber W. and Rall H. (2013) ‘We are journalists’: Production practices, attitudes and a case study of the New York Times newsroom. In: Weber, W., Burmester, M. and Tille, R. (eds) *Interaktive Infografiken*. Berlin: Springer: 161–172.
201. Weiss A. and Domingo D. (2010) Innovations processes in online newsrooms as actor-networks and community of practice. *New Media & Society* 12(7): 1156–1171.
202. Weinacht S. and Spiller R. (2014) Datenjournalismus in Deutschland: Eine explorative Untersuchung zu Rollenbildern von Datenjournalisten. *Publizistik* 59(4): 411–433.
203. Wright S. and Doyle K. (2019) The Evolution of Data Journalism: A Case Study of Australia. *Journalism Studies*, 20(13): 1811-1827.
204. Wuchty S., Jones B. F. and Uzzi B. (2007) The increasing dominance of teams in production of knowledge. *Science* 316(5827): 1036–1039.
205. Yarnall L., Johnson J. T., Rinne L. and Ranney, M. A. (2008) How Post-secondary Journalism Educators Teach Advanced CAR Data Analysis Skills in the Digital Age. *Journalism & Mass Communication Educator* 63(2): 146-164.
206. Young M. L. and Hermida A. (2015) From Mr. and Mrs. Outlier to central tendencies: Computational journalism and crime reporting at the Los Angeles Times. *Digital Journalism* 3(3): 381–397.
207. Zamith R. (2019) Transparency, Interactivity, Diversity, and Information Provenance in Everyday Data Journalism, *Digital Journalism*, 7(4): 470-489.

208. Zanchelli M. and Crucianelli, S. (2012) Integrating data journalism into newsrooms. International Center for Journalists. Available at http://www.icfj.org/sites/default/files/integrating%20data%20journalism-english_0.pdf

209. Zhang S. and Feng J. (2019) A Step Forward? Exploring the diffusion of data journalism as journalistic innovations in China. *Journalism Studies*, 20(9): 1281-1300.

210. Zuiderwijk, A., Shinde, R. and Janssen, M. (2018) Investigating the attainment of open government data objectives: Is there a mismatch between objectives and results? *International Review of Administrative Sciences*. Available at:

<http://journals.sagepub.com/doi/full/10.1177/0020852317739115#articleCitationDownloadContainer>