

**ВАЛЕНТИНА СЕРГЕЕВНА БЕРЕЖНАЯ**

Национальный исследовательский университет

«Высшая школа экономики»,

ResearcherID: AAB-7697-2019

ORCID: 0000-0002-8671-1342

e-mail: Vberezhnaya@hse.ru

## ВОПРОСЫ СТАНДАРТИЗАЦИИ ФАКТЧЕКИНГА В ЖУРНАЛИСТИКЕ ДАННЫХ. ТЕОРЕТИЧЕСКИЙ АСПЕКТ

**Аннотация:** Журналистика данных в расследованиях опирается на данные как источник и инструмент раскрытия истории. Принципы работы с данными, их поиска, верификации, анализа и дата-сторителлинга значительно отличаются от традиционных журналистских практик, но в это же время являются органичной частью журналистского материала и сложившихся стандартов и этических норм журналистики. Фактчекинг в журналистике данных при этом часто сводится к математической проверке корректности обработки данных, при этом практически не уделяется внимания множеству других факторов, влияющих на корректность и этичность текста: происхождение данных, методика и мотивация их сбора, корректность интерпретации, контекстуализация результатов анализа, корректности представления данных в визуализациях. В академическом поле изучаются эпистемологические отличия журналистики данных — создание форм собственного знания и его принятия аудиторией под воздействием практик, основанных на данных, и соучастие читателя в создании и верификации дата-материалов. В то же время фактчекинг в области журналистики данных также рассматривается только как технологический процесс сверки вычислений, а не как единая и системная верификация журналистского материала на всех его уровнях, от технического до этического. В этой статье автор попытался заполнить разрыв между академическим полем и наработанных в ньюсрумах практиках

дата-фактчекинга: были исследованы существующие методы фактчекинга в различных редакциях, рассмотрены различные подходы к оценке достоверности дата-материалов, описаны существующие лакуны в дата-фактчекинге. Отсутствие стандартов в этой области ведет как к падению качества материалов, так и к публикации ошибочных сведений или заведомой манипуляции данными с различными целями, поэтому автор предлагает собственный взгляд на фактчекинг в дата-историях как систему последовательной многоуровневой проверки.

**Ключевые слова:** журналистика данных, фактчекинг, воспроизводимость, интерпретация и контекстуализация данных, анализ данных, стандарты

**VALENTINA S. BEREZHNYAYA**

National Research University Higher School of Economics,

Researcher ID: AAB-7697-2019

ORCID: 0000-0002-8671-1342

e-mail: vberezhnaya@hse.ru

## STANDARDIZED FACT CHECKING IN DATA JOURNALISM. A THEORETICAL VIEW

**Abstract.** Data journalism is based on data used both as a source of a story and as a proof for facts stated in journalistic investigations. Core principles of journalistic work are changing under the influence of data: working with data, acquiring datasets, verifying data, analyzing and presenting it in data stories is drastically different from traditional journalistic methods, while simultaneously continuing to be an organic part of journalistic research within the existing framework of journalism standards and ethical requirements. Fact-checking in data journalism is often limited to verifying correct math and analysis methods in data, whereas other factors defining the correctness and ethics of a journalistic product are ignored. Those include assessing the sources of data, methods and reasons of data collection, correctness of interpretation, contextual dependencies of data, correctness of visual representation of data analysis results, etc. Scientists are expanding research into the epistemological differences of data journalism from traditional journalistic practice, noting such distinct features as creating personal knowledge and its acceptance by the audience under the influence of data-driven practices and co-creation and crowd verification of data-based investigations. At

the same time, academic research also focuses on data journalism fact-checking as a mere technological process of revision and comparison of calculations, not as a holistic system of data-story verification on multiple interconnected planes from technology to ethics. In this article, the author tries to fill the existing gap between academic research and actual data fact-checking practices in newsrooms by scrutinizing and evaluating various approaches to data-story fact-checking in a number of media, and consequently defining white spaces in the data fact-checking workflows. Lack of professional standards in the area allows for lower quality of publications, as well as publishing wrongly interpreted or presented data, whether by mistake or by intent. This prompted the author's original view of fact-checking in data journalism as a system of consistent multilevel assessment.

**Keywords:** data journalism, fact checking, reproducibility, interpreting and contextualizing data, data analysis, standards

Одной из основ профессии журналиста является строгая и доскональная проверка фактов, сообщаемых в материале. Во многих редакциях существует практика верификации любого материала перед публикацией либо специальным подразделением или другими журналистами, не принимавшими участия в работе над конкретно этой историей, либо при помощи отдельных организаций, специализирующихся на проверке фактов. Существующие методы и этика проверки фактов (фактчекинга) сформировались за последнее столетие, на их основе выработаны профессиональные стандарты качества журналистских материалов.

Исследователи школы коммуникаций и информации Rutgers выделяют следующие стандарты качества журналистики: медиа должны предоставлять исчерпывающую и релевантную информацию в обществе и мире, информация должна быть объективной (точной, честной, достаточно полной, соответствующей реальности, проверяемой, факты должны быть отличны от мнений), а также беспристрастной (все стороны и интерпретации должны быть представлены равнозначно и непредвзято) [1].

Британская вещательная корпорация (BBC) устанавливает такие этические стандарты [2]: свобода выражения, независимость, следование общественным интересам, непредвзятость и объективность, достоверность, редакционная этичность, защита уязвимых групп населения, защита приватности. С появлением в современной технологической среде принципиально иного источника — данных — практики фактчекинга в целом не изменились и не адаптировались к новой информационной реальности.

Журналистика данных выделилась как отдельное направление журналистики в последнее десятилетие: в основе его лежит работа с данными, их обретение, обработка, анализ, представление в визуальной форме и нарратив, основанный на результатах этой работы с данными. В этом случае данные служат и инструментом для раскрытия определенных историй, и доказательством приводимых фактов, и источником журналистского расследования. Это высокотехнологическое направление журналистики требует знаний в области науки о данных, визуализации данных, часто — программирования. Журналистика данных обязана своим появлением всемирным движением за открытость информации, доступностью и открытостью данных, а также технологиям, которые позволяют хранить и обрабатывать большие объемы данных.

Инструменты и методы журналистики данных все чаще применяются в расследовательской, новостной и объяснительной журналистике как одна из составляющих сторителлинга, базовые умения анализа и визуализации данных становятся не просто востребованным, но обязательным набором навыков современного журналиста. При этом даже в крупных мировых редакциях *не разработан стандарт фактчекинга для дата-расследований*, не существует сложившихся мировых практик в этой области, более того, публичное обсуждение необходимости другого подхода к фактчекингу в части анализа, интерпретации и визуализации данных ведется лишь в очень узком кругу специалистов. Даже Международная сеть журналистов-расследователей в рекомендациях по изучению и использованию журналистики данных не упоминает фактчекинг дата-расследований как один из этапов работы журналиста [3], а многочисленные сервисы фактчекинга, в том числе автоматизированные, направлены на работу с «традиционными» ресурсами и не рассматривают расследования, основанные на данных.

Методы работы с данными в контексте фактчекинга также часто рассматриваются как составляющая вычислительного фактчекинга (computational fact checking) [4] для информации онлайн в широком ее понимании, при этом фактчекинг в журналистике данных часто сводится к корректности технологических процессов сбора и анализа данных.

С точки зрения автора, в эпоху быстрого чтения, массовой и стремительной дигитализации информации, доступности и распространен-

ности данных проверка фактов в дата-материалах становится крайне актуальной проблемой и с позиций технологического анализа данных, и со стороны корректной интерпретации и контекстуализации результатов основанного на данных исследования. Фактчекинг такого рода требует специальных навыков и часто недоступен в силу объективных (недостаток знаний, технологических навыков, технических ресурсов и т.д.) и субъективных (нежелание, недостаток времени, доверие автору, низкий уровень критического мышления и т.д.) причин, что позволяет как допускать ошибки в дата-материалах, так и намеренно манипулировать данными для достижения определенных целей. Разработка стандартов в области фактчекинга в журналистике данных и их открытая публикация, как представляется автору, помогут значительно улучшить прозрачность и качество дата-расследований. Большинство журналистских материалов, основанных на данных, публикуются в цифровой среде и остаются в открытом доступе для просмотра, проверки и верификации пользователями, и распространяются в социальных сетях и личных сообщениях. Такая «долгоиграющая» и доступная природа дата-расследований подразумевает высокую точность работы с фактурой и прозрачность методов работы с данными, а также определенный уровень экспертности в расследуемых темах.

Исследователи Дэвид Черуйот и Рауль Феррер-Конилл [5] отмечают, что даже в условиях, когда данные становятся значительным источником для медиа и других организаций, применяющих инструментарий журналистики в своей работе, в академическом поле наиболее активно рассматриваются сложившиеся практики Запада, а исследования фактчекинга в области данных в медиа практически отсутствуют. По их мнению, нежурналистские организации, например, международные НКО, занимающиеся проверкой фактов на основе данных, заставляют пересматривать и обновлять журналистский дискурс в части практик, основанных на данных. Черуйот и Феррер-Конилл полагают, что эпистемологически журналистика меняется — от оценки фактов, предоставленных другими людьми, и создания собственного знания до форм знания и производства знания, тесно связанных с принятием этого знания аудиторией. По их мнению, это изменение происходит под воздействием систем количественной оценки журналистики и практик, основанных на данных. В логике фактчекинга данные могут стать журналистским источником, который, в

рамках существующего дискурса объективности, означает качественное измерение журналистики. Исследователи говорят о ценности данных как эксплицитном источнике знаний для их последующего воспроизводства и распространения, но не останавливаются на практиках фактчекинга непосредственно в области дата-журналистики, сосредотачивая внимание на алгоритмизированных системах фактчекинга.

Марк Коддингтон [6] обращает внимание, что журналистика данных базируется на четырех столпах открытости — прозрачность, итеративность, изменяемость и соучастие, а распространенность данных меняет устоявшиеся журналистские практики. Основным эпистемологическим отличием журналистики данных от других цифровых методов в журналистике он считает практику соучастия, в которой журналисты видят в аудитории соавторов в поисках правды и морального права. Коддингтон также отмечает, что в журналистике данных отдельной ценностью стала визуализация — одна из ключевых особенностей жанра, предполагающая связанность дизайна с ценностями журналиста. Также он говорит об особенности журналистики данных, в которой данные подчинены журналистским ценностям и нарративу истории, при этом данные исследуются научными методами, а оценка интерпретации и контекстуализации часто остается за пределами знаний журналистов в экспертной среде. При этом исследователь не рефлексировал изменение редакционных практик фактчекинга в связи с появлением данных как одного из значимых источников.

Исследователи Кеннеди, Вебет и Энгебретсен [7, с. 169–182] утверждают, что визуализация данных ассоциируются с такими характеристиками как правда и объективность, что в глазах аудитории делает журналистский материал более заслуживающим доверия. Ученые настаивают, что прозрачность визуализаций должна становиться новой журналистской нормой, но практики создания такой прозрачности пока определяют как «интерпретативная гибкость». Визуализация данных кажется объективным способом представления результатов журналистских расследований, но при этом не является нейтральным взглядом на данные. Для увеличения прозрачности и точности визуализации исследователи предлагают редакциям описывать процесс создания визуализаций и давать ссылки на источники данных, а также представлять в публичном доступе датасеты, на основании которых были созданы визуализации. При этом единой прак-

тики оценки качества данных или их визуализаций авторы не представляют, равно как и не предлагают единого подхода к указанию источников.

Вышесказанное приводит автора статьи к необходимости восполнить разрыв между сложившейся в ньюсрумах практикой журналистики данных и академическим полем, в котором фактчекинг в этом жанре обсуждается редко и лишь частично, не как единая система со своими особенностями, а также не осмысливается эпистемологическая «наследственность» методов верификации и проверки информации в журналистике данных. В этой работе автор изучает существующие редакционные практики и целостность их подхода, а на основе анализа предлагает основы собственной системы фактчекинга в журналистике данных.

Агентство Reuters в Handbook of Journalism [8] вообще не упоминает данные как источник или метод расследований, несмотря на то что регулярно выпускает дата-материалы и сложные расследования, основанные на сборе, анализе и интерпретации данных. Данные как источник, требующий проверки, также целиком опущены в Verification Handbook [9] Европейского центра журналистики [10]. Многие редакции, проводящие дата-расследования, указывают в материалах источники данных и иногда описания датасетов, но не дают информации о методах исследования и фактчекинга в дата-материалах: среди них ProPublica, The New York Times, The Guardian.

Практика оценки достоверности при работе с данными в большинстве СМИ ограничена технологическими аспектами сбора, очистки, анализа данных. Например, Associated Press в открытом документе «Новостные ценности и принципы» [11] ограничивается кратким замечанием о данных: «Данные для историй и визуальных презентаций должны быть проверены на полноту и валидность. Данные должны быть оценены с точки зрения методологии сбора, размера выборки, времени сбора и доступности других данных, которые могут подтвердить или оспорить их. Объединение более одного датасета в презентации должно быть сделано внимательно и прозрачно. Избегайте сравнения процентов и процентных пунктов для малой выборки, включая сырые цифры, необходимые для общего понимания. Мы должны четко отличать корреляции от причинно-следственных связей». В широко признанном пособии по журналистике данных Data Journalism Handbook [12] Европейского центра журналистики

подробно рассматриваются инструменты работы с данными и способы их обработки, но также не затрагивается вопрос фактчекинга таких исследований, авторы не останавливаются на качестве интерпретации результатов анализа данных и помещения их в контекст.

Международная фактчекинг-сеть (IFCN — International Fact-Checking Network) [13] Пойнтеровского института разработала собственный кодекс принципов [14] для верификации информации в медиа, он включает пять принципов, каждый из которых описывается пятью или шестью требованиями. Мы выделили из этого числа только те обязательства для участников этого сообщества, которые, по мнению автора статьи, применимы, в том числе к данным:

- одинаковые стандарты фактчекинга применяются для проверки всех фактов вне зависимости от стороны, представившей их;
- выводы строятся на основании полученных свидетельств, а не свидетельства выбираются для поддержки выводов;
- возможность независимой проверки фактов читателями — участники сообщества предоставляют максимум информации и ссылки на источники, достаточные для воспроизведения результатов расследования, в случаях, если это не компрометирует личную безопасность и приватность;
- прозрачность методологии, используемой в выборке, поиске, обработке фактов и фактчекинге.

При этом данные как источник в принципе не рассматриваются в стандартах IFCN, соответственно, фактчекинг анализа и представления данных не упоминается. Часть принципов и обязательств описаны достаточно широко, чтобы можно было применить их к дата-расследованиям, при этом ключевыми в предложенной рамке становятся проверка источников, прозрачность методологии и возможность репликации действий журналиста для получения тех же результатов. Вопросы качества данных, корректности их интерпретации, соответствия данных реальности не охвачены.

Пол Брэдшоу, один из пионеров дата-журналистики и преподаватель факультетов журналистики университетов в Бирмингеме и Лондоне, полагает, что одним из ключевых факторов оценки качества дата-материала должна быть и его *этичность* [15]: точность представления данных

зависит от их корректной контекстуализации. В зависимости от контекста данные могут быть по-разному визуализированы, интерпретированы, представлены или даже вовсе не использованы и не опубликованы. Например, российские государственные данные о количестве сирот использовать в качестве основы журналистского расследования невозможно: такие данные представляют государственные структуры [16], но специфика сбора этих данных такова, что, по сути, лишает их смысла — нет единой методики учета для разных регионов, рекомендации по сбору этих данных не соблюдаются владельцами датасетов, а проверить целостность и достоверность фактически невозможно. Такие данные, хоть и считаются открытыми, дают искаженное представление о реальности, а значит, не могут использоваться в журналистском расследовании ни с точки зрения профессиональных стандартов качества журналистики, ни с точки зрения этики. Данные, например, о количестве детей-инвалидов в России [17] представлены ответственными ведомствами (Министерством труда и социальной защиты, Департаментом по делам инвалидов) отрывочно, что также значительно деформирует представление о реальном положении дел — в этом случае журналисту надо собирать и анализировать данные других ведомств: Министерства финансов, Пенсионного фонда, Министерства здравоохранения, Министерства просвещения, где представлены другие данные о детях инвалидах: начисления пенсий, особые условия образования и т.д. Сбор косвенных данных в этом случае дает более полную и соответствующую действительности картину, но требует более внимательного исследования данных, корректных корреляций межведомственных данных, а также экспертных знаний в этой области, позволяющих дать корректную оценку безликим цифрам и увидеть за ними живую человеческую историю и суть существующей социальной проблемы.

В российском издании «Новая газета» [18] процесс фактчекинга зависит от типа материала. В дата-отделе газеты общее правило состоит в том, чтобы найти как можно больше разных датасетов с данными по интересующей теме и составить наиболее полный датасет, исправить возможные неточности (убрать повторы, собрать разные части в единый набор данных), а также получить информацию от экспертов в предметной области о качестве и достоверности источников.

Издание «Важные истории» [19] практикует модель фактчекинга, предложенную OCCRP [20]: автор материала расставляет сноски с доказательствами для каждого факта; фактчекер, другой дата-журналист редакции, проверяет их. В этом случае фактчекинг происходит, по сути, дважды: сначала автором, который при расстановке ссылок может увидеть неточности и исправить их, а затем коллегой-журналистом. В процессе проверки оценивается корректность каждого факта и утверждения в тексте, если у фактчекера возникают вопросы, они решаются с автором текста; если автор и фактчекер не пришли к единому мнению, в тексте оставляют комментарий для редактора. Проверяются все источники, расчеты, даты, имена, названия, соответствие цитат записям бесед с экспертами и героями публикаций. Так как материал проверяется дата-журналистом, снижается вероятность ошибок в расчетах и методах анализа данных.

Издание «проект.» [21] подходит к верификации материалов, основанных на данных, иначе: после редактуры текст вместе с данными, расчетами и ссылками на источники поступает к фактчекеру — другому журналисту редакции, чаще всего не специализирующемуся на дата-расследованиях, который каждому предложению в тексте должен найти доказательство. Если таковое не было найдено, либо обнаружилось ошибки, фактчекер просит уточнений у автора. В этом случае фактчекер ищет доказательства самостоятельно, что позволяет снизить уровень предвзятости и обратить внимание на детали и другие источники, которые автор мог не заметить.

Дата-журналистка Винни де Джонг [22] предлагает такой чеклист для проверки корректности работы с данными:

1. Проверьте даты, орфографию и дубликаты, выпадающие значения. Помните, что статистическая значимость — еще не новость. Убедитесь, что тренды представлены в диахронии. Убедитесь, что данные соотносятся с реальностью.
2. Во время работы ведите дневник данных, где описываете последовательность действий и сами действия с данными. Вы должны быть способны воспроизвести свои вычисления.
3. Опишите методы работы — читатель должен понимать, как автор обнаружил в данных историю.
4. Ведите файл со сносками: каждый факт снабжайте номером, для

каждого номера собирайте информацию: откуда этот факт вам известен, источник, доказательство. Исправьте все ошибки, найденные в этом файле.

Предложенная методика также больше фокусируется на технической точности в сборе и обработке данных, хотя здесь уже встречается упоминание необходимости их контекстуализировать, при этом методика не рассматривает оценку качества данных и корректность их представления как этапы фактчекинга.

В отличие от многих других журналистских жанров, дата-расследования должны быть воспроизводимыми: любой человек, повторив те же действия с тем же набором данных, должен прийти к таким же выводам, к которым пришел журналист. Это, по мнению автора, *является одним из ключевых показателей качества дата-журналистской работы*. Большинство дата-журналистских расследований основываются на открытых данных, которые доступны для анализа любому пользователю сети. Кроме того, наука о данных как точная дисциплина предполагает абсолютную воспроизводимость результата. Сочетание этих факторов с основополагающий принципом достоверности журналистских материалов приводит автора к выводу, что любое расследование, сделанное в интересах общества на основании открытых источников, должны быть проверяемо и достоверно, и в этом случае достоверность анализа будет подтверждена его воспроизводимостью. Возможность проверки представленной информации также легла в основу принципов IFCN. Поэтому точное пошаговое логирование всех действий, произведенных с данными, становится необходимым условием как в работе дата-журналиста, так и в фактчекинге в редакции и волонтерами. Такой лог позволяет также отследить корректность действий и использования разных методов анализа на всех этапах сбора, очистки, анализа и других операций с данными и вычислить ошибки в них как во время самопроверки, так и во время фактчекинга. Хранение всех версий данных позволит быстро вернуться к точке, где произошла ошибка или неточность, и далее провести анализ корректно.

*Надежность и репутация источника данных также должны оцениваться в процессе фактчекинга*. Если источник данных вызывает сомнения, в том числе в части мотивации сбора и предоставления данных,

обязательно получить комментарий эксперта в исследуемой области: зачем были собраны данные, на какие вопросы они могут ответить, какие ограничения и исключения могут быть в этих данных.

*Методология сбора исходных данных также должна быть изучена и учтена при анализе, поскольку она влияет на методы обработки и анализа данных.* Например, с 15 января 2020 года по нынешний день параметры подсчета заболевших коронавирусом COVID-19 в Китае менялись уже семь раз [23]. Это значит, что визуализация тренда, основанного на этих данных как одинаковых, будет заведомо ошибочной, хотя источник — официальное статистическое ведомство крупного государства — может считаться надежным.

С точки зрения автора этой статьи, *корректность интерпретации данных также должна оцениваться во время процедур факт-чекинга в дата-историях*, поскольку часто одни и те же данные могут быть использованы для поддержания двух противоречивых точек зрения. Например, данные о количестве женщин в правительстве могут быть истолкованы и как «число женщин во власти удвоилось», и как «женщины недостаточно представлены во властных структурах», если за год количество женщин-парламентариев изменилось от 20 до 40 при общей численности парламента в 250 человек. Контекст, в котором существуют данные, методы и цели их сбора играют важную роль в истории, и также должны быть верифицированы при фактчекинге.

Ни одна из перечисленных выше практик фактчекинга не предлагает этапа оценки качества визуализаций — от необходимости таковых до корректности и уместности их исполнения. Такая верификация позволит избежать намеренных или случайных искажений в представлении данных и, как следствие, формирования неверного представления о предмете у публики. Таким часто встречающимся примером данных, представленных некорректно и искажающих представление о реальности, можно назвать столбиковые диаграммы в новостных программах телеканала «Россия», где визуализации не предлагают точки отсчета, но отображают только различие показателей, исключив или уменьшив кодифицируемые в такой диаграмме значения. Автор не предлагает оценивать творческие решения в области визуализации, но настаивает на применении объективных метрик оценки качества визуализации. Такие предлагает, например,

А. Богачев [24]: точность представления данных, корректность выбора типа графика, корректность указания единиц исчисления и точек отсчета, корректное отображение соотношений величин, корректность написания легенды и т.д.

В результате изучения практик дата-фактчекинга в редакциях автор сформулировал несколько рекомендаций по верификации в журналистике данных, которые могут быть внедрены в редакциях как специализирующихся на дата-расследованиях, так и тех, кто иногда прибегает к инструментарию журналистики данных.

1. Хранение исходных данных с указанием их источников, дат обращения и скачивания наборов данных. Разумно также сохранять скриншоты веб-страниц, откуда скачивались файлы, поскольку не всегда владельцы данных выполняют требования по открытому и доступному хранению исходных данных и удаляют их по разным причинам, от незнания правил обращения с открытыми данными до попыток скрыть информацию.

2. В случаях, когда данные собирались редакцией (опросы, скрейпинг — сбор данных заданного типа с сайта или сайтов специальными программами и т.д.), необходимо подробно описать методологию и инструментарий сбора данных.

3. Ведение подробного журнала работы с данными. Существует несколько опций, доступных редакциям. Одна из них — в программах обработки данных, например, Excel, Google Sheets, OpenRefine и других, необходимо включать автоматическое логирование всех действий. Другой опцией может послужить открытое программное обеспечение Jupyter, веб-приложение Jupyter Notebook позволяет в одном документе логировать программный код, уравнения, визуализации, текст. Важно, что Jupyter Notebooks — одновременно человекочитаемые и машиночитаемые документы. В последнее время сочетание этих подходов все чаще используется в индустрии, например, в BuzzFeed и The New York Times.

4. Необходимо хранить все версии данных — это позволит во время фактчекинга обнаружить, на каком этапе могли произойти ошибки, а также позволит вернуться к предыдущим версиям и гарантировать целостность и сохранность данных.

5. Подробное описание методологии анализа данных. Для опреде-

ленных типов данных, предоставляемых по единому образцу, собранных по единой методике, методологии исследования должны быть стандартизированы, чтобы исключить предвзятость в отдельных случаях. Это верно, например, для регулярных исследований статистических показателей, типовых отчетов и т.д. Для всех исследований, требующих нестандартных подходов, методология должна быть детально описана.

6. Экспертная оценка корректности интерпретации результата и его контекстуализации профессионалом в исследуемой области с сохранением комментариев эксперта.

7. Коллегиальная проверка, в которой участвуют дата-журналисты, не занимавшиеся материалом, с возможностью сохранения комментариев и внесенных по результатам такой проверки правок.

8. Оценка качества визуализации: насколько корректно визуализация представляет данные, отражает ли она результат расследования, не искажает ли соотношения, отображает ли все необходимые значения, корректно указывает величины и т.д. Автор полагает, система оценки качества визуализации данных должна быть разработана отдельно и служить своего рода чек-листом для журналистов и дизайнеров при создании визуализаций, инфографики, геоинфографики.

Частично эти практики уже существуют в крупных изданиях. Некоторые медиа, например, BuzzFeed, иногда публикуют свои Jupyter Notebooks на промышленных платформах — такая открытость расследований должна показать непредвзятость и объективность издания, но в то же время избирательность публикации ноутбуков вызывает обоснованные этические вопросы. Отдельные издания, например, «Новая газета» в своих дата-расследованиях обязательно указывают методологию обработки данных («Как мы считали») и ограничения исследования — это важный репутационный элемент, который дает возможность не аффилированным с редакцией исследователям проверить выводы, полученные путем анализа данных, что соответствует рекомендациям IFCN и OCCRP. Ссылки на источники стали де факто стандартом в журналистике данных, в то время как редакций, так подробно описывающих методы работы с данными в ходе журналистского расследования, меньшинство.

## ЛИТЕРАТУРА

1. Lacy S., Rosenstiel T. Defining and Measuring Quality Journalism // Rutgers. School of Communication and Information: сайт.  
URL: <http://mpii.rutgers.edu/wp-content/uploads/sites/129/2015/04/Defining-and-Measuring-Quality-Journalism.pdf> (дата обращения: 28.05.2020).
2. BBC. Editorial Guidelines // BBC: сайт.  
URL: <https://www.bbc.co.uk/editorialguidelines/guidelines> (дата обращения: 28.05.2020).
3. Data Journalism // Global Investigative Journalism Network: сайт.  
URL: <https://helpdesk.gijn.org/support/solutions/articles/14000036505-data-journalism> (дата обращения: 25.04.2020).
4. Computational Fact Checking: A Content Management Perspective / Cazalens S., Leblay J., Lamarre P., Manolescu I., Tannier X. // Proceedings of the VLDB Endowment. 2018. Vol. 11. Issue 12. DOI: 10.14778/3229863.3229880.
5. Cheruiyot D., Ferrer-Conill R. "Fact-Checking Africa": Epistemologies, Data, and the Expansion of Journalistic Discourse // Digital Journalism. 2018. Vol. 6. Issue 8. P. 964–975.
6. Coddington M. Clarifying Journalism's Quantitative Turn: A Typology for Evaluating Data Journalism, Computational Journalism, and Computer-Assisted Reporting // Digital Journalism. 2014. Vol. 3. Issue 3. P. 331–348.
7. Henedy H., Weber W., Engerbretsen M. Data visualization and transparency in the news // Data Visualization in Society / Ed. by M. Engerbretsen, H. Kennedy. Amsterdam: Amsterdam University Press, 2020. P. 169–185.
8. Reuters Handbook of Journalism: сайт.  
URL: [http://handbook.reuters.com/index.php?title=Main\\_Page](http://handbook.reuters.com/index.php?title=Main_Page) (дата обращения: 08.06.2020).
9. Verification Handbook: сайт.  
URL: <http://verificationhandbook.com/> (дата обращения: 10.06.2020).
10. European Journalism Centre: сайт.  
URL: <https://ejc.net/> (дата обращения: 10.06.2020).
11. The Associated Press Statement of News Values And Principles // The Associated Press: сайт.  
URL: <https://www.ap.org/about/news-values-and-principles/downloads/ap-news-values-and-principles.pdf> (дата обращения: 08.06.2020).
12. The Data Journalism Handbook 2: Towards a Critical Data Practice: сайт.  
URL: <https://datajournalism.com/read/handbook/two> (дата обращения: 08.06.2020).

13. Poynter. The International Fact-Checking Network // Poynter: сайт.  
URL: <https://www.poynter.org/ifcn/> (дата обращения: 09.06.2020).
14. IFCN Code of Principles: сайт. URL: <https://ifcncodeofprinciples.poynter.org/> (дата обращения: 09.06.2020).
15. Bradshaw P. Data Journalism // Ethics for Digital Journalists: Emerging Best Practices / Ed. by L. Zion, D. Craig. N.-J, London: Routledge, 2015. P. 202–219.
16. Численность детей-сирот, детей, оставшихся без попечения родителей в стационарных учреждениях социального обслуживания // ЕМИСС. Государственная статистика: сайт.  
URL: <https://fedstat.ru/indicator/41601> (дата обращения: 12.06.2020).
17. Численность впервые признанных инвалидами по категории «ребенок-инвалид» // ЕМИСС. Государственная статистика: сайт.  
URL: <https://fedstat.ru/indicator/41623> (дата обращения: 12.06.2020).
18. Новая газета: сайт.  
URL: <https://novayagazeta.ru/> (дата обращения: 12.06.2020).
19. Важные истории: сайт.  
URL: <https://www.istories.media/> (дата обращения: 12.06.2020).
20. Organized Crime and Corruption Reporting Project: сайт.  
URL: <https://www.occrp.org/en> (дата обращения: 11.06.2020).
21. Проект: сайт.  
URL: <https://www.proekt.media> (дата обращения: 11.06.20).
22. De Jong W. How Not To Be Wrong // Global Investigative Journalism Network: сайт.  
URL: <https://gijn.org/2017/05/25/how-not-to-be-wrong/> (дата обращения: 10.06.2020).
23. China Says It's Beating Coronavirus. But Can We Believe Its Numbers? // Time: сайт.  
URL: <https://time.com/5813628/china-coronavirus-statistics-wuhan/> (дата обращения: 11.06.2020).
24. Богачёв А. А. Графики, которые убеждают всех. Москва: АСТ, 2020. 280 с.

## REFERENCES

1. Lacy S., Rosenstiel T. Defining and Measuring Quality Journalism. *Rutgers. School of Communication and Information*.  
URL: <http://mpii.rutgers.edu/wp-content/uploads/sites/129/2015/04/Defining-and-Measuring-Quality-Journalism.pdf> (accessed 28.05.2020).
2. BBC. Editorial Guidelines. *BBC*.

URL: <https://www.bbc.co.uk/editorialguidelines/guidelines> (accessed 28.05.2020).

3. Data Journalism. *Global Investigative Journalism Network*.

URL: <https://helpdesk.gijn.org/support/solutions/articles/14000036505-data-journalism> (accessed 25.04.2020).

4. Computational Fact Checking: A Content Management Perspective. Cazalens S., Leblay J., Lamarre P., Manolescu I., Tannier X. *Proceedings of the VLDB Endowment*. 2018. 11(12). DOI: 10.14778/3229863.3229880.

5. Cheruiyot D., Ferrer-Conill R. "Fact-Checking Africa": Epistemologies, Data, and the Expansion of Journalistic Discourse. *Digital Journalism*. 2018. 6(8), pp. 964–975.

6. Coddington M. Clarifying Journalism's Quantitative Turn: A Typology for Evaluating Data Journalism, Computational Journalism, and Computer-Assisted Reporting. *Digital Journalism*. 2014. 3(3), pp. 331–348.

7. Henneidy H., Weber W., Engerbretsen M. Data visualization and transparency in the news. *Data Visualization in Society*. Ed. by M. Engerbretsen, H. Kennedy. Amsterdam: Amsterdam University Press, 2020, pp. 169–185.

8. *Reuters Handbook of Journalism*.

URL: [http://handbook.reuters.com/index.php?title=Main\\_Page](http://handbook.reuters.com/index.php?title=Main_Page) (accessed 08.06.2020).

9. *Verification Handbook*.

URL: <http://verificationhandbook.com/> (accessed 10.06.2020).

10. *European Journalism Centre*.

URL: <https://ejc.net/> (accessed 10.06.2020).

11. The Associated Press Statement of News Values And Principles. *The Associated Press*.

URL: <https://www.ap.org/about/news-values-and-principles/downloads/ap-news-values-and-principles.pdf> (accessed 08.06.2020).

12. *The Data Journalism Handbook 2: Towards a Critical Data Practice*.

URL: <https://datajournalism.com/read/handbook/two> (accessed 08.06.2020).

13. Poynter. The International Fact-Checking Network. *Poynter*.

URL: <https://www.poynter.org/ifcn/> (accessed 09.06.2020).

14. *IFCN Code of Principles*.

URL: <https://ifcncodeofprinciples.poynter.org/> (accessed 09.06.2020).

15. Bradshaw P. Data Journalism. *Ethics for Digital Journalists: Emerging Best Practices*. Ed. by L. Zion, D. Craig. N.-J. London: Routledge, 2015, pp. 202–219.

16. EMISS [Unified Interdepartmental Statistical Information System]. *Chislennost' detei-sirot, detei, ostavshikhnya bez popecheniya roditelei v statsionarnykh*

*uchrezhdeniyakh sotsial'nogo obsluzhivaniya* [Number of orphans and children left without parental care in residential social service institutions].

URL: <https://fedstat.ru/indicator/41601> (accessed 12.06.2020). (In Russ.)

17. EMISS [Unified Interdepartmental Statistical Information System]. *Chislennost' v pervye priznannykh invalidami po kategorii "rebenok-invalid"* [Number of persons first-time recognized as disabled in the category of "disabled child"].

URL: <https://fedstat.ru/indicator/41623> (accessed 12.06.2020). (In Russ.)

18. *Novaya gazeta* [New newspaper].

URL: <https://novyagazeta.ru/> (accessed 12.06.2020). (In Russ.)

19. *Vazhnye istorii* [Important stories].

URL: <https://www.istories.media/> (accessed 12.06.2020). (In Russ.)

20. *Organized Crime and Corruption Reporting Project*.

URL: <https://www.occrp.org/en> (accessed 11.06.2020).

21. *Proekt* [Project].

URL: <https://www.proekt.media> (accessed 11.06.20). (In Russ.)

22. De Jong W. How Not To Be Wrong. *Global Investigative Journalism Network*.

URL: <https://gijn.org/2017/05/25/how-not-to-be-wrong/> (accessed 10.06.2020).

23. China Says It's Beating Coronavirus. But Can We Believe Its Numbers? *Time*.

URL: <https://time.com/5813628/china-coronavirus-statistics-wuhan/> (accessed 11.06.2020).

24. Bogachev A. A. *Grafiki, kotorye ubezhdayut vsekh* [Charts that convince everyone]. Moscow: AST, 2020. 280 p. (In Russ.)

#### СВЕДЕНИЯ ОБ АВТОРЕ

##### **ВАЛЕНТИНА СЕРГЕЕВНА БЕРЕЖНАЯ**

доцент Департамента медиа,  
Академический руководитель магистерской  
образовательной программы «Журналистика данных»,  
директор Многофункционального инновационного  
телевизионного технологического центра,  
факультет Коммуникаций, медиа и дизайна,  
Национальный исследовательский университет  
«Высшая школа экономики»,  
109028, Москва, Хитровский переулок, д. 2/8, стр.5.

**Researcher ID: AAB-7697-2019**

**ORCID: 0000-0002-8671-1342**

e-mail: vberezhnaya@hse.ru

#### ABOUT THE AUTHOR

##### **VALENTINA S. BEREZHNYAYA**

Associate Professor at the School of Media,  
Academic Supervisor of Data Journalism  
Master's Educational Program,  
Director of the Multifunction Innovative Television Technological Center,  
Faculty of Communication, Media, and Design,  
National Research University Higher School of Economics,  
korpus 5 (P), 2–8, Khitrovsky pereulok, Moscow, 109028

**Researcher ID: AAB-7697-2019**

**ORCID: 0000-0002-8671-1342**

e-mail: vberezhnaya@hse.ru