

УДК 81'42:004.738.5

**BIG DATA, OPEN DATA КАК НОВЫЕ ФОРМАТЫ ИНФОРМАЦИИ:
СУЩНОСТЬ, ХАРАКТЕРИСТИКИ,
ОСОБЕННОСТИ ПРИМЕНЕНИЯ В ЖУРНАЛИЗМЕ**

М. Г. Шилина

Национальный исследовательский университет
Высшая школа экономики
департамент интегрированных коммуникаций

В статье исследуются сущность, характеристики новейших форматов цифровых данных, так называемых больших (*big data*) и открытых (*open data*). Автор рассматривает влияние этих типов информации на технологический цикл работы журналиста над материалом, фиксирует появление новых специфических параметров медиакommunikации, усиление ее социоцентричности, что подтверждает выделение дата-журнализма в качестве особого направления журналистики.

Ключевые слова: *цифровая информация, большие данные, открытые данные, журнализм, креативная команда, визуализация, просьюмер, технологический цикл, социоцентричность, дата-журнализм.*

Во втором десятилетии XXI века цифровизация, Интернет определяют особенности социальной коммуникации, в том числе профессиональной. Информационно-коммуникационная революция, по мнению Я. Н. Засурского, приводит к конвергенции как возможности существования и хранения информации на различных источниках в цифровой форме [3, с. 4].

Данные, которые обретают цифровой формат, сегодня повсюду: это «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» [9] и так далее. Данные увеличиваются в объеме экспоненциально (на нецифровую информацию приходится не более 10 процентов, и доля ее сокращается), растут непрерывно, накапливаются, и этот процесс необратим.

Накопление больших объемов цифровой информации привело к появлению в начале 2000-х годов новых ее форматов. Форматы цифровых данных многообразны, для их классификации релевантны два основания: технология обработки, хранения и передачи и характер использования, закрытый или публичный.

В своей статье мы рассмотрим так называемые большие данные (англ. *big data*) и открытые данные (англ. *open data*). Обозначенные феномены отражают не только количественные, но и качественные изменения в функционировании и использовании цифровой информации: большие объемы данных позволяют получать при их анализе более релевантные результаты, выявлять корреляционные связи между явлениями, более точно выстраивать прогностические модели.

Большие и открытые данные привлекают внимание общественности и профессионалов во всех сферах [22]. Для медиа большие данные, по мнению анали-

тиков, являются одним из наиболее значимых параметров функционирования [29], трансформаций [11] и стратегий [28].

Понятия «большие данные» и «открытые данные» пока не имеют общепризнанных дефиниций. Термином «большие данные» (предложен по аналогии с метафорами «большая нефть», «большая вода» в 2008 году редактором журнала *Nature* К. Линчем) обозначают наборы данных, «размер которых выходит за пределы возможностей типичных программных средств сбора, хранения, управления и анализа баз данных» [11] (то есть больше терабайта). Это также «совокупности данных с возможным экспоненциальным ростом, которые слишком велики, слишком неформатированы или слишком неструктурированы для анализа традиционными методами» [6, с. 42].

Характеристики больших данных многообразны, базовыми традиционно называют три (так называемые «три V»): объём (англ. *volume*), скорость прироста, обработки, получения результатов (англ. *velocity*), многообразие типов структурированных и полуструктурированных данных (англ. *variety*) [4].

Концепция открытых данных предполагает свободный доступ и открытость неких наборов больших данных для машиночитаемого использования и публикации, распространения без ограничений авторского права (как правило, это данные из государственных источников). Доступ к данным и их использование контролируются, как правило, государственными и частными организациями через патенты, лицензии, платный доступ и т.п. Сторонники концепции открытых данных полагают, что все данные должны быть общедоступны, ограничения не соответствуют идее гражданского общества. Развитие открытых данных во второй половине 2000-х гг. обусловлено законодательно закрепленными инициативами правительств ведущих стран мира по открытию информации (например, запуск ресурса *Data.gov* в 2009 г. в США).

Отличие больших данных от открытых в их потенциальной и реальной коммерческой ценности (например, «большие» персональные данные, данные о поведении потребителя, о денежных транзакциях и т.д. всегда могут быть монетизированы), такой формат данных их обладатели раскрывают на особых условиях.

Основным источником больших и открытых данных является Интернет, где массивы данных открыты и накапливаются, где функционируют масштабные поисковые системы *Google*, *Яндекс*, присутствуют глобальные общественные проекты Википедия, *OpenStreetMap* и др. Информационные ресурсы и потоки интегрируются в базы данных, что также увеличивает их объемы. Например, *Semantic Web* и *Linked Data* выступили с инициативой о приведении всех данных к общим форматам, особом способе их публикации, который удобен для извлечения информации, причем сразу из нескольких источников. Трендом в последнее время становится интеграция сервисов, открываются порталы и каталоги открытых данных. В итоге подобных трансформаций источники структурированных данных больших объемов становятся общедоступными.

В рамках данного исследования предложим рабочие определения данных понятий. Большие данные – серия подходов, инструментов и методов обработки структурированных и неструктурированных машиночитаемых данных объемом от одного петабайта для получения воспринимаемых человеком результатов. Открытые данные – данные, находящиеся в свободном доступе для машиночитаемого ис-

пользования, обработки и распространения без ограничений авторского права, лицензий, патентов и иных механизмов контроля.

В использовании больших и открытых данных особое значение приобретает способ их обработки и представления, то есть аналитическая компонента, в первую очередь, профессиональная. Для масс-медиа «накопление данных и неограниченный доступ к цифровой информации становятся важными составляющими выживания журналистики как элемента общественной и экономической сферы» [30].

Новые форматы данных обусловили появление новых форм работы с данными и новых типов материалов, так называемой журналистики данных (англ. *data journalism*). Данный феномен, несмотря на отсутствие общепризнанной дефиниции, получил интенсивное распространение в ведущих зарубежных масс-медиа (с 2006 г. [18]), в России зафиксированы пилотные проекты профессиональных журналистов и журналистики граждан в 2010-х годах [11].

Понятие «журналистика данных», по мнению шведских исследователей Е. Аппельгрена и Г. Нигрена, сложно дефиницировать, поскольку это комбинация компетенций нескольких профессий и навыков [13]. В рамках данной работы используем в качестве рабочего следующее определение: «Журналистика данных – это набор специфических навыков для поиска, анализа, визуализации информации цифровых источников данных с целью формирования интерактивных форматов уникальной подачи авторского аналитического журналистского контента и эффективного взаимодействия СМИ, журналиста с аудиторией; это формат актуального журнализма; формат профессионального журналистского медиатекста, медиаконтента; метод его создания, трансляции, потребления» [11].

Дата-журналистика становится все более востребованной: регулярно публикуются материалы в качественной прессе во всем мире («Гардиан»/ *The Guardian*, «Файнэншл Таймс»/ *Financial Times*, «Нью-Йорк Таймс»/ *The New York Times*, «Чикаго Трибьюн»/ *Chicago Tribune* и др.), для них отведены специальные разделы (например, блог в «Гардиан», *The Guardian Datablog*). Первое учебное пособие по журналистике данных переиздавалось неоднократно, переведено на многие языки, в том числе на русский [7]; ежегодно организация «Глобальная сеть редакторов» (*Global Editors Network*) проводит дата-хакатоны по всему миру; лучшим авторам присуждается международная премия *Data Journalism Award*.

В России данный формат журналистских материалов пока не вошел в регулярную практику СМИ, что выявило анкетирование, проведенное в 2015 году в десяти редакциях ведущих качественных СМИ [10, с. 8]: термин «журналистика данных» оказался известным и использовался журналистом только в газете «Коммерсант *daily*».

Появление и интенсивное развитие в журналистике проектов, основанных на новейших форматах цифровой информации, формирует научную проблему, которая связана с профессиональным поиском, обработкой, представлением и потреблением больших и открытых данных. Журналистские проекты всегда реализуются в междисциплинарной области, на границе между гуманитарными, социальными и техническими науками; однако сегодня с усложнением технических средств, виртуализацией коммуникации при их создании в обязательном порядке используются компьютерные знания и научные данные, журналистская деятельность приобретает все более интегрированный характер [8; 11]. Для исследователя базовым становится

ся вопрос о том, создают ли новые форматы цифровой информации новые характеристики журналистики. Изучение свойств журнализма в парадигме больших и открытых данных в рамках ограниченного объема данной статьи сфокусируем на технологическом цикле работы над материалом.

Тема больших и открытых данных, журналистики данных, их применения в медиа является новой, обладает высоким дискуссионным потенциалом. Различные аспекты дата-журнализма отражены в работах зарубежных практиков и исследователей Е. Аппельгрена [13], Э. Головатого [18], С. Коэн [14; 15], М. Лоренца [19; 20], К. Ниппарда [23], С. Роджерса [25], Е. Сирккунен [27] и др.; российских ученых [11; 12] и аналитиков И. Бегтина [2], И. Радченко [24], Л. Черняка [9]. В «Пособии по журналистике данных» [7] представлен корпус первых зарубежных практик в данной сфере.

Молодые российские исследователи Д. В. Бабыничева и А. Г. Шилина при анализе практик ведущих британских и американских изданий в 2013–2015 гг. отмечают следующие характеристики использования новых форматов цифровой информации [10]. Материал (проект) дата-журнализма представляет собой интеграцию онлайн-журналистского текста, основанного на анализе больших массивов данных, визуализации этих данных в различных форматах (таблицы, диаграммы, инфографика, мультимедиа). Проект, как правило, постоянно обновляется в онлайн-режиме в процессе появления новых данных или по мере подключения к проекту новых читателей. В разработке проекта участвует креативная команда, в которую входят журналист, программист, статистик, аналитик, дизайнер, фотограф, инфограф и другие специалисты, которые нужны для формирования интегрированного текста дата-журнализма. Вербальный контент в большинстве случаев выполняет функцию разъяснения значимости данных, то есть базовой становится именно информация, которая содержится в *big & open data*; все более существенное значение приобретает визуальная компонента (графический дизайн, мэппинг данных, интерактивная графика и др.), которая начинает доминировать [10]. Функции материала дата-журналистики таковы: информационная, функция критики и публичного контроля над деятельностью властей, регулирующая, организационная, функция общественного участия, архивная функция [1, с. 98].

Отметим, что информация, которую используют журналисты, – различные виды структурированных и неструктурированных данных (базы данных, данные Интернета и др.), – является открытой, инструментарий журналиста распространяется под открытой лицензией с открытым кодом. То есть процесс получения, обработки, анализа, визуализации, представления данных выполняется в рамках именно концепции открытых данных, которая означает свободный доступ к большим массивам данных и к аналитическим инструментам их обработки.

Рассмотрим основные характеристики и особенности использования больших и открытых данных, сфокусировав внимание на технологическом цикле создания журналистского материала. Технологический цикл традиционно (по Г. В. Лазутиной [5]) включает творческую и техническую компоненты, хронологически включает создание первичной, предварительной концепции (до сбора материала) и окончательной (после сбора материала); работа с информацией осуществляется в парадигмах первичного сбора информации (методами наблюдения, эксперимента, интервью и т. д.) и дальнейшего анализа и интерпретации данных.

Первые же проекты, в которых журналисты использовали новые форматы данных, больших и открытых, показали отличия данного направления журнализма по всем базовым параметрам технологического цикла. Так, журналисты начали определять тему проекта, исходя из уже имеющегося набора данных (а не наоборот) и использовать читателей как активных соавторов для их обработки; цели проекта изменялись в зависимости от особенностей аналитики данных. В частности, британская ежедневная газета «Телеграф»/The Telegraph, изучив отчеты членов парламента о расходах, обнаружила махинации с налогами, траты представительских средств на личные нужды и т.д. Парламент в итоге открыл еще около 460 тысяч страниц отчетов, которые затем уже газета «Гардиан» разместила на своем сайте как базу данных, и 33 тысячи читателей приняли участие в их анализе. Данный пример стал уже хрестоматийным и показывает характерное изменение классической модели медиакоммуникации от объект-субъектной к субъект-субъектной, при этом автор становится коллективным (творческо-техническая команда: журналист, программист, дизайнер, фотограф, SMM-менеджер); равноправным соавтором становится активная аудитория просьюмеров; обязательным становится наличие заинтересованности аудитории в теме проекта и предварительное выявление этого интереса журналистами, следовательно, изменяются методы работы с информацией, технологический и хронологический циклы создания материала.

Большие и открытые данные формируют принципиально новые форматы журналистских материалов, для создания которых также требуется обновление параметров технологического цикла. Так, один из первых примеров американского проекта журналистики, основанного на данных, – формирование интернет-ресурса для информирования и обеспечения безопасности граждан Чикаго на основе алгоритмов сбора и анализа информации с сайтов администрации, полиции и других служб города [18], – начался в рамках работы газеты как «ЧикагоКрайм» (ChicagoCrime), а затем стал полноценным агрегатором новостей «ЭвэриБлок» (EveryBlock). То есть журналистский проект, благодаря точному «попаданию» в востребованную аудиторией тему, трансформировался в социально значимый самостоятельно функционирующий ресурс, трансформируя СМИ из медиацентричных в социоцентричные.

Растущая благодаря использованию новых типов данных социализация масс-медиа, расширение аудитории просьюмеров – и соответственно трансформации модели и технологического цикла работы над материалом – обретает не только креативную, но и техническую базу. Так, в зарубежных изданиях разрабатываются собственные оригинальные технологии подачи материала на основе использования новых форматов данных и взаимодействия с аудиторией. Например, проекты британской газеты «Гардиан» «Даталог»/The Guardian Datablog) и «Датастор»/Datastore) предоставляют просьюмерам программу, которая делает возможным создание собственных визуализаций данных; американский проект One Million Dollar Block предлагает данные городских бюджетов с понятной визуализацией для любого квартала города и др.

Анализ актуальных практик дата-журнализма ведущих изданий США и Великобритании в 2013–2015 гг. [1; 10] фиксирует следующие их характеристики: новые форматы данных расширяют возможности фактографической базы журнали-

стики; использование больших объемов данных позволяет журналисту увеличить достоверность публикации; открытые данные дают возможность прямого доступа аудитории к массивам официальной открытой информации и работы с ней; варианты различной аналитической трактовки баз данных позволяют использовать одну и ту же информацию многократно [10]; дата-журнализм все еще сохраняет в себе признаки журналистики мнений, а не факта из-за сохранения фактора персональной интерпретации данных [1].

Каков же технологический цикл работы с новыми форматами данных в журнализме? Журналисты и исследователи фиксируют несколько подходов. Так, практики одного из ведущих в сфере дата-журнализма изданий – газеты «Гардиан» – описывают шесть этапов создания дата-проекта. На первом этапе журналист ищет данные, релевантные теме, формирует предварительную концепцию, организует первичный сбор данных из открытых источников (данные проходят первичную проверку, оцениваются с точки зрения актуальности, репрезентативности и др.). На втором этапе происходит формирование базы актуальных «сырых» данных, формируется основная концепция материала, идет поиск тех корреляций в базе, которые могут в результате анализа данных «выстрелить», стать интересными для читателей. На всех этапах к работе подключаются программисты и аналитики, которые работают с базой данных проекта и производят поэтапно их техническую обработку (например, переформатируют данные, создают унифицированный классификатор). Затем производится очистка, форматирование данных аналитиком (дешифровка, создание аналитических таблиц, выделение ключевых сюжетных линий и т. д.). В итоге на основе технически обработанных данных создается журналистский сюжет и начинается окончательный этап: выбор формы представления итоговых данных. Формат визуального представления данных зависит от сюжета, возможностей команды и ресурсов (технических, креативных и др.).

Исследователь М. Лоренц считает, что последовательность создания проекта на основе больших и открытых данных такова: получение данных, фильтрация (обработка) данных, визуализация данных, создание истории на основе данных [20].

Технологический цикл, по Б. Фраю, состоит из четырех этапов [16]. На первом этапе происходит выбор источников данных, их получение и компьютерный анализ (используется инструментарий информатики). Затем следует извлечение нужных данных и их фильтрация (с использованием математических и статистических методов работы с данными). На третьем этапе к работе над проектом подключается специалист по визуализации (веб-дизайнер, инфограф, графический дизайнер, фотограф), который представляет концепт визуализации. Программист «очищает» данные, создает интерактивный дизайн. На завершающем этапе проект открывают для читателей.

Общая логика формирования технологического цикла работе над дата-проектом в целом сходна: выбор данных и работа с ними (фильтрация «сырых» данных, поиск корреляций, очистка данных аналитика), и только затем – формирование истории, визуальной концепции, их реализация (представление читателям).

Более подробно проработанный (и апробированный на практике) технологический цикл создания дата-контента (журналистского и визуального материала) описывает А. Г. Шилина [10, с. 24].

Выбор темы проекта.
 Выбор цели проекта.
 Описание целевой аудитории проекта.
 Поиск и анализ данных.
 Объединение и интерпретация данных.
 Написание журналистского текстового материала на основе анализа данных.

Выбор инструмента визуализации данных.

Визуализация.

Тестирование и оценка эффективности проекта.

Отметим отличия от описанных выше циклов работы над журналистским текстом в пунктах 3, 5, 7, 9, то есть изменения затрагивают даже по формальным признакам почти половину технологического цикла. Наличие выделенных нами позиций отражает обязательное повышенное внимание к запросам аудитории, именно в этом и заключается особенность, социоцентричность дата-журнализма, которую обеспечивают большие и открытые данные.

Новаторские отличия технологического цикла дата-журнализма очевидны: данные диктуют тему, определяют целеполагание, особенности сторителлинга, расширение субъектности – числа авторов за счет привлечения технических специалистов и просьюмеров, необходимость работы достаточно большой креативной команды (журналист, программист, аналитик, дизайнер и т. д.) и наличия в индустрии новых профессиональных компетенций и профессий (программист, веб-аналитик). Изменяются технологии и технологический цикл работы с информацией и подготовки материала. Большие и открытые данные определяют необходимость новых способов представления вербальной и визуальной информации, их интеграции, а также интерактивность проекта. Отметим, что определяющим фактором эффективности дата-проекта становится не только высокий уровень профессионализма, но и качество внутрикомандной коммуникации.

Таким образом, анализ сущности, характеристик, параметров применения больших и открытых данных как форматов информации в журнализме позволяет зафиксировать появление новых его свойств. Выявленные особенности определяют необходимость создания релевантной методологии и методов профессиональной деятельности.

Список литературы

1. Бабыничева Д. В. Открытые данные в медиакоммуникации. М.: НИУ ВШЭ, 2015. 95 с.
2. Бегтин И. Готовы ли мы к журналистике данных? [Электронный ресурс] // Полит.ру. 2013. URL: http://polit.ru/article/2013/04/29/data_journalism. (Дата обращения: 10.09.2015.)
3. Засурский Я. Н. Коммуникация в обществе знаний. Российская журналистика: 2008–2013. М.: МедиаМир, 2013. 296 с.
4. Канаракус К. Машина Больших Данных. [Электронный ресурс] // Сети. 2011. № 4. URL: <http://www.osp.ru/nets/2011/04/13010802/>. (Дата обращения: 10.09.2015.)
5. Лазутина Г. В. Основы творческой деятельности журналиста. М.: МедиМир, 2008. 240 с.

6. Моррисон А. Поиск эффективных инструментов работы с большими данными [Электронный ресурс] // Технологический прогноз. 2010. Вып. 3. С. 41–50. URL: <http://ru.scribd.com/doc/83309742/>. (Дата обращения: 12.09.2015.)
7. Пособие по журналистике данных / Под ред. Д. Грей. М.: РИА Новости, 2013. 240 с.
8. Севтунова Е. А. Дата-журналистика в системе медиакоммуникаций: особенности интеграции контента. М.: НИУ ВШЭ, 2015. 95 с.
9. Черняк Л. Большие Данные – новая теория и практика. [Электронный ресурс] // Открытые системы. 2011. № 10. URL: <http://www.osp.ru/os/2011/10/13010990/>. (Дата обращения: 10.09.2015.)
10. Шилина А. Г. Дата-журналистика в качественной прессе России: опыт идентификации. М.: Моск. гос. ун-т, 2015. 110 с.
11. Шилина М. Г. *Data Journalism* – дата-журналистика, журналистика метаданных как новый формат медиакоммуникации: к вопросу формирования исследовательских подходов [Электронный ресурс] // Медиаскоп. 2013. Вып. 1. URL: <http://www.mediascope.ru/node/1263>. (Дата обращения: 10.09.2015.)
12. Шилина М. Г. *Big data, open data, linked data*, метаданные в PR: актуальные модели трансформации теории и практики [Электронный ресурс] // Медиаскоп. 2014. Вып. 1. URL: <http://www.mediascope.ru/node/1486>. (Дата обращения: 10.09.2015.)
13. Appelgren E., Nygren G. Data journalism in Sweden – opportunities and challenges. A case study of Brottspejl at Sveriges Television (STV). [Электронный ресурс] // Proceedings of the digital humanities congress. 2012. URL: <http://www.hrionline.ac.uk/openbook/chapter/dhc2012-appelgren>. (Дата обращения: 10.09.2015.)
14. Cohen S., Hamilton J. T., Turner F. Computational journalism // Communications of the ACM. 2011. Vol. 54. No. 10. P. 66–71.
15. Cohen S. Research Chat: Sarah Cohen of the New York Times on the state of data journalism and what reporters need to know [Электронный ресурс] // Harvard Kennedy School. Shorenstein Center on Media, Politics and Public Policy, 2014. URL: <http://www.journalistsresource.org/skills/reporting/research-chat-new-york-times-Sarah-Cohen-state-data-journalism-what-reporters-need-know>. (Дата обращения: 10.09.2015.)
16. Fry B. Visualizing data: exploring and explaining data with the processing environment. Beijing etc.: O'Reilly, 2007. 367 p.
17. Gray J. The data journalism handbook. How journalists can use data to improve the news // Beijing etc.: O'Reilly, 2012. 182 p.
18. Holovaty A. A fundamental way newspaper sites need to change. [Электронный ресурс]. URL: <http://www.holovaty.com/writing/fundamental-change/>. (Дата обращения: 10.09.2015.)
19. Lorenz M. Information architect. 2013. [Электронный ресурс]. URL: <http://www.mirkolorenz.com/?id=6>. (Дата обращения: 10.09.2015.)
20. Lorenz M., Kayser-Bril N., McGhee G. Media companies must become trusted data hubs [Электронный ресурс] // European Journalism Centre. URL: <http://ejc.net/magazine/article/media-companies-must-become-trusted-data-hubs#.Vg-PTu3dX9aR>. (Дата обращения: 10.09.2015.)

21. Manyika J. et al. Big data: The next frontier for innovation, competition, and productivity. 2011 [Электронный ресурс]. URL: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation. (Дата обращения: 10.09.2015.)
22. Mayer-Schönberger V., Cukier K. Big Data: A Revolution that will transform how we live, work, and think. London: John Murray, 2013. 242 p.
23. Nippard C. Data visualizations emerge across newsrooms, online publishers. [Электронный ресурс] // Deutsche Welle, 2011. URL: <http://www.dw-world.de>. (Дата обращения: 10.09.2015.)
24. Radchenko I. The view on open data and data journalism: cases, educational resources and current trends. Analysis of images, social networks and texts [Электронный ресурс] // Communications in computer and information science. 2014. Vol. 436. URL: http://link.springer.com/chapter/10.1007/978-3-319-12580-0_4. (Дата обращения: 10.09.2015.)
25. Rogers S. Data Journalism at the Guardian: what is it and how we do it? [Электронный ресурс]. URL: <http://www.theguardian.com/news/datablig/2011/jul28/data-journalism>. (Дата обращения: 10.09.2015.)
26. Rogers S. Facts are sacred: the power of data [Электронный ресурс] // Faber&Faber. 2013. URL: <http://www.bookdepository.com/Facts-are-Sacred-Simon-Rogers/9780571301614>. (Дата обращения: 10.09.2015.)
27. Sirkkunen E., Aitamurto T., Lehtonen, P. Trends in data journalism [Электронный ресурс] // Finnish Strategic Centre for Science, Technology and Innovation in the field of ICT. URL: http://virtual.vtt.fi/virtual/nextmedia/Deliverables-2011/D3.2.1.2.B_Hyperlocal_Trends_In%20Data_Journalism.pdf. (Дата обращения: 10.09.2015.)
28. The role of data repositories in reproducible research [Электронный ресурс] // Yale ISPS. URL: <http://isps.yale.edu/news/blog/2013/07/the-role-of-data-repositories-in-reproducible-research>. (Дата обращения: 10.09.2015.)
29. Thibodeau P. Gartner's Top 10 IT challenges include exiting baby boomers, Big Data [Электронный ресурс] // Computerworld. 2011. October 18. URL: <http://www.computerworld.com/article/2497348/>. (Дата обращения: 10.09.2015.)
30. Van der Haak B., Parks M., Castells M. The future of journalism: networked journalism [Электронный ресурс] // International journal of communication. 2012. Vol. 6. URL: <http://ijoc.org/index.php/ijoc/article/view/1750/832>. (Дата обращения: 10.09.2015.)

**BIG DATA, OPEN DATA AS NEW INFORMATION FORMATS:
ESSENCE, FEATURES, APPLICATION IN JOURNALISM**

M. G. Shilina

National Research University Higher School of Economics
the Department of integrated communications

The paper examines the nature and characteristics of new formats of digital data, the so called big (big data) and open (open data) formats. The author examines the impact of these types of information on technological aspects of the journalistic process, and

outlines new specific parameters of media communication which becomes more and more socially oriented, all this confirming the emergence of data-journalism as a special direction of journalism.

Keywords: *digital information, big data, open data, journalism, creative team, visualization, prosumer, work cycle, social orientation, data-journalism.*

Об авторе:

ШИЛИНА Марина Григорьевна – доктор филологических наук, доцент департамента интегрированных коммуникаций факультета коммуникации, медиа и дизайна НИУ ВШЭ (101000, Москва, ул. Мясницкая, 20), e-mail: marina.shilina@gmail.com.

About the author:

SHILINA Marina Grigorevna – doctor of Philological Sciences, associate Professor of the Department of integrated communications of the faculty of communication, media and design HSE (101000, Moscow, Myasnitskaya str., 20), e-mail: marina.shilina@gmail.com.