



## [Научные статьи]

Сапрыкина Д.И., Козырева М.

*Большие данные в измерениях аудитории цифровых медиа:  
новые возможности для исследований и необходимость  
создания единого измерителя*

# БОЛЬШИЕ ДАННЫЕ В ИЗМЕРЕНИЯХ АУДИТОРИИ ЦИФРОВЫХ МЕДИА: НОВЫЕ ВОЗМОЖНОСТИ ДЛЯ ИССЛЕДОВАНИЙ И НЕОБХОДИМОСТЬ СОЗДАНИЯ ЕДИНОГО ИЗМЕРИТЕЛЯ

### **Сапрыкина Д.И.**

Эксперт проектно-учебной лаборатории  
медиакоммуникаций в образовании  
Национального исследовательского университета  
«Высшая школа экономики» (Москва, Россия)  
[dsaprykina@hse.ru](mailto:dsaprykina@hse.ru)

### **Козырева М.**

Студент бакалаврской программы  
«Медиакоммуникации» Национального  
исследовательского университета «Высшая школа  
экономики» (Москва, Россия)  
[mariiakozyreva@gmail.com](mailto:mariiakozyreva@gmail.com)

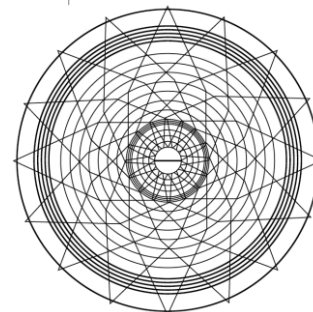
### **Аннотация:**

Целью исследования было определить, как специалисты из медиасферы оценивают возможность создания единого измерителя и формирования альянса для обмена информацией и выработки единых методик измерения и анализа аудитории цифровых медиа при помощи технологий больших данных. В рамках исследования было проведено 10 экспертных полуструктурированных интервью со специалистами, работающими с аудиторными данными или занимающимися их сбором. В частности, в выборку попали представители таких компаний, как Mail.ru, Mediascope, Russia Today, Трансперенси Интернешнл-Р и другие. Данная работа помогла вскрыть противоречие между интересом к унификации показателей к обмену данными и рядом серьезных внешних факторов, влияющих на невозможность реализации на данном этапе развития рынка. Один из основных барьеров для создания единого измерителя и обмена информацией – финансовая незаинтересованность крупных игроков и опасения по поводу конкурентного преимущества. Отдельным вектором для обсуждения является вопрос этики в отношении сбора данных о пользователях. Ожидания от внедрения технологий двойственны: часть экспертов считает, что внедрение новых методов измерения существенно упростит деятельность компании, когда другие полагают, что данные технологии переоценивают, и они существенно не повлияют на методы измерения аудитории.

## [Научные статьи]

Сапрыкина Д.И., Козырева М.

*Большие данные в измерениях аудитории цифровых медиа:  
новые возможности для исследований и необходимость  
создания единого измерителя*



**Ключевые слова:** Большие данные, измерения аудитории, цифровые медиа, метрики, аналитика, Big Data, СМИ, исследования аудитории.

В связи с развитием информационных технологий и информатизацией общества возникала потребность говорить о больших данных, генерирование которых происходит в настоящий момент с рекордной скоростью (Villars, Olofson, Eastwood, 2013). Практически каждое действие, совершаемое индивидом в цифровой среде, оставляет свой след. Большие данные открывают новые возможности для исследований и являются важнейшей частью интернет-маркетинга и ИТ-индустрии (McAfee A. et al., 2012). В то же время внимание к большим данным является актуальным и для медийных исследований, где особое место занимают измерения аудитории СМИ (Полуэхтова, 2016; Вартанов, 2017). В большинстве случаев такие процессы, как принятие решения о производстве и размещение контента, планирование стратегии привлечения целевой аудитории, определение цены на рекламносители, а также размещение рекламы как в электронных, так и в печатных СМИ основываются на аудиторных показателях. В том числе именно на этот тип данных в сегменте изучения медиа «направлены основные исследовательские расходы субъектов практической деятельности в сфере СМИ» (Давыдов, 2007). Подобная «валюта» в виде показателей необходима для формирования отношений между всеми игроками рынка медиаиндустрии. По этой причине метрики должны быть унифицированными, методология исследования быть понятна и признана на институциональном уровне (Kosterich & Napoli, 2015; Napoli, 2011; Webster et al., 2014).

Введя понятие больших данных, считаем необходимым уже сейчас обозначить, что в рамках данной работы за основу было взято следующее определение: большие данные (то же самое «Big Data», «биг дата») – структурированные или неструктурированные непрерывные потоки информации, получаемые из одного или нескольких источников (Provost F., Fawcett, 2013). Говоря о больших данных, мы понимаем, что существуют различные подходы к сбору и обработке данных, но в данной работе осознанно не делаем нам это акцент, перенося его на возможности, которые могут дать большие данные в измерении аудитории цифровых медиа.

Итак, казалось бы, в связи с активным развитием цифровой среды доступно огромное количество информации об аудитории из совершенно разных каналов. И это должно исключительно положительно влиять на медиабизнес. Так, в случае с большими данными представляется возможность решить проблему с фрагментарностью данных. У компаний появляется ресурс проводить «сплошные» исследования – регистрировать все взаимодействия пользователя с контентом, в том числе детально замерять время и модель поведения индивида в цифровой среде (Bilton, 2014; Haile, 2014; Ingram, 2014, 2015; Lowenstein, 2014). К тому же сервера баз данных позволяют связать



## [Научные статьи]

Сапрыкина Д.И., Козырева М.

*Большие данные в измерениях аудитории цифровых медиа:  
новые возможности для исследований и необходимость  
создания единого измерителя*

повторный заход пользователя с его портретом и временем взаимодействия с контентом и типом сайта. Однако существует ограничение в виде не унифицированных методик измерения и анализа собранных данных, что может привести к получению разных показателей и к их несопоставимости.

На данный момент на российском рынке не существует выработанной единой методологии использования технологий больших данных в целях анализа аудитории, отсутствуют унифицированные метрики для оценивания взаимодействия аудитории с контентом разных форматов, нет выработанных практик формирования целостного представления о собственной аудитории. Подобная методологическая неопределенность мешает выстраиванию контентной политики цифровых медиа и, как следствие, влияет на доходы от рекламы. Ситуация осложняется тем, что представители индустрии нередко закрыты для диалога об обмене данными с другими игроками медиарынка, в то время как на зарубежных рынках есть положительный опыт обмена данными об аудитории между СМИ в рамках таких проектов как “Pangaea Alliance” и “Emetriq”. Цель подобного взаимодействия – формирование целостного портрета пользователей и снижения влияния от крупных технологических компаний путем создания альтернативных баз данных.

В связи со сложившейся ситуацией на рынке медиаисследований целью нашего исследования стало провести анализ возможностей применения технологий больших данных в измерениях аудитории цифровых медиа. В частности, оценить возможность создания единого измерителя и формирования альянса.

### **Обзор литературы**

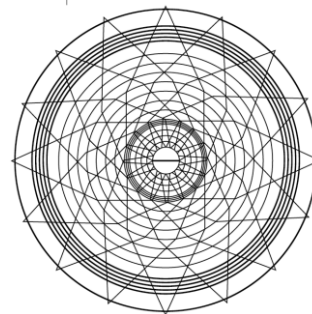
Проведенное исследование сосредоточено в первую очередь на менеджериальном аспекте, упуская технологические особенности сбора и обработки больших данных. Вследствие этого в данном разделе статьи был сделан акцент на существующие результаты внедрения технологий больших данных в исследования аудитории медиа и, в первую очередь, на возможности и ограничения, с которыми сталкиваются компания при использовании этих технологий.

Прежде чем перейти к описанию имеющихся практик, обратимся к концептуализации понятия большие данные, которое носит неоднозначных характер. Употребление словосочетания «большие данные» уже стало общемировой тенденцией. Как только объем разноформатных неструктурированных данных резко увеличился с ростом доступности технологий (Snijders et al., 2012; Das, 2017), инструменты больших данных вышли за пределы ИТ-сообщества. К 2013 году, согласно специальному отчету Gartner, популярность больших данных достигла пика, о чем свидетельствует положение в Gartner Hype Cycle – модель, которая позволяет оценить популярность технологии в публикациях и общественных обсуждениях. Столь широкое

## [Научные статьи]

Сапрыкина Д.И., Козырева М.

*Большие данные в измерениях аудитории цифровых медиа: новые возможности для исследований и необходимость создания единого измерителя*

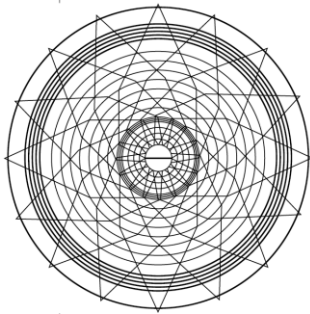


распространение во многие научные и деловые сферы привело к концептуальной неопределенности данного термина. Как следствие, специалисты из разных профессиональных областей под большими данными стали понимать абсолютно разные механизмы: начиная от технических способностей компьютерной системы генерировать, обрабатывать и хранить огромные потоки информации до социального феномена, влияющего на установки и ценности современного общества (Колесниченко, Смородин, Ильин и др., 2015).

IDC определяет большие данные как «новое поколение технологий и стратегий, предназначенных для сбора и анализа больших объемов данных из нескольких гетерогенных источников с высокой скоростью, а также с извлечением из этого процесса экономической ценности» (Woodward, Carter, 2015). В научных источниках термин используют для обозначения массивных, быстро расширяющихся, разнообразных и неструктурированных наборов оцифрованных данных, с которыми сложно работать с помощью традиционных баз данных (Hashem et al., 2015). Этому же определению придерживаются специалисты из McKinsey, акцентируя внимание на том, что определение не является конечным и может трансформироваться в зависимости от индустрии, в которой происходит работа с данными (Manuyika, J., 2011). Как уже было отмечено вначале работы, под большими данными мы понимаем структурированные или неструктурированные непрерывные потоки информации, получаемые из одного или нескольких источников (Provost F., Fawcett, 2013).

Об использовании технологий больших данных для обработки и анализа информации об аудитории сказано в ряде научных работ российских ученых. В частности, это труды С.А. Вартанова, И.А. Полуэхтовой (Вартанов, 2014; Полуэхтова, 2016), но речь в них идет о традиционных медиа (в большей степени о ТВ), поэтому открытыми остаются вопросы о желании, существующих практиках и готовности игроков на рынке цифровых медиа взаимодействовать с такими технологиями. Без ответа остаются вопросы о востребованности единых метрик и измерителя в данной сфере.

Если говорить о сегменте измерения аудитории СМИ, то можно найти не так много источников, где комплексно рассматривалась проблема внедрения и использования технологий больших данных в сегменте измерения аудитории СМИ. Так, под этими технологиями может пониматься возможность перехода выборочных исследований к сплошным. Ричард Маркс в аналитическом докладе «The Big Opportunity: Audience Research Meets Big Data» отмечает, что благодаря серверным данным, социальным медиа, клиентским базам и данным Return Path Data, существует возможность для переведения медиаисследований на следующий уровень. Эти инструменты позволяют, в отличие от выборочных опросов населения, охватить всю генеральную совокупность и получать данные моментально. Такие возможности использования больших данных открывают перспективы как средствам массовой информации, так и рекламодателям,



## [Научные статьи]

Сапрыкина Д.И., Козырева М.

*Большие данные в измерениях аудитории цифровых медиа:  
новые возможности для исследований и необходимость  
создания единого измерителя*

которые могут предлагать свою продукцию, подстраиваясь под те или иные предпочтения клиента в данный момент времени. (Marks, 2013)

Однако технологии больших данных в медиа – не только источник конкурентного преимущества, но и специфический инструментарий. В случае с цифровыми медиа основными стратегиями применения технологий больших данных являются (Martha L. Stone, 2014):

1) Анализ аудитории и поведенческих особенностей взаимодействия с контентом (к примеру, A/B тестирование материалов, которое используют The Huffington Post);

2) Использование инструментов для построения нарратива на открытых данных (оперативная система реагирования на экстренные новости у CNN);

3) Инструменты визуализации в журналистике данных (Global Editors Network выделяют отдельные номинации в GEN's annual Data Journalism Awards за визуализацию данных);

4) Инструменты для управления и поиска видео и аудио материалов, а также анализа социальных сетей и другого контента (аналитика социальных сетей, отраженная GlobalWebIndex);

5) Настройка и формирование таргетинга рекламных кампаний на определенные группы целевых аудиторий;

6) Оценка эффективности взаимодействия с аудиторией (разработка Channel 4 по анализу телевизионной аудитории через регистрацию пользователей и формирование их цельного портрета по разным девайсам);

7) Для обмена информацией и выработки единых методик и метрик для анализа аудитории при помощи технологий больших данных.

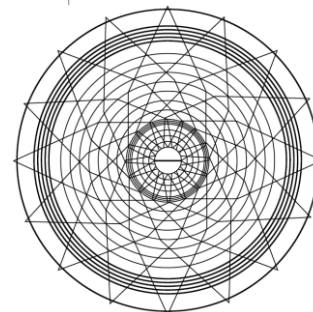
Говоря непосредственно о цифровых медиа, технологии больших данных могут адаптироваться для взаимодействия с массивами неструктурированной разноформатной информации, которая постоянно подвергается тщательной аналитике в формате подхода data fusion. Это постепенно перевело на новый уровень исследования контента и культуры (например, значимыми являются решения компании Habidatum), которые в дальнейшем ложатся в основу новых проектов. Аналогичные разработки уже сейчас позволяют медиакомпаниям получать комплексное понимание аудитории и скорректировать деятельность, а также анализировать не только свой собственный контент, но и тренды среди пользователей или конкурентов (от метаданных изображений и видео, заканчивая алгоритмами анализа визуальной информации по типу исследований лаборатории Льва Мановича (Манович, 2016)).

Lippell Н. предоставляет в своем исследовании три основные сферы, которые могут быть улучшены за счет использования больших данных в СМИ: продукты и услуги, клиенты и поставщики, инфраструктура и процессы (Lippell, 2016). Автор дополнительно выделяет потенциальные движущие силы для развития технологий больших данных. В частности, это быстрое изменение предпочтений аудитории, необходимость анализа расхождений в моделях

## [Научные статьи]

Сапрыкина Д.И., Козырева М.

*Большие данные в измерениях аудитории цифровых медиа:  
новые возможности для исследований и необходимость  
создания единого измерителя*



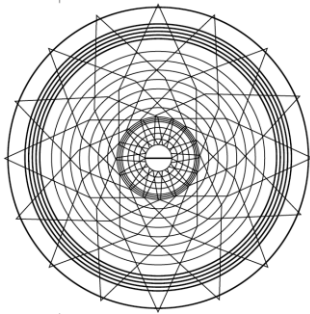
взаимодействия с контентом, а также желание организации разнообразить предоставляемые товары или услуги.

Ключевое отличие работы с большими данными для анализа аудитории цифровых медиа заключается в том, что собранные данные являются поведенческими, они позволяют наблюдать реальное взаимодействие с контентом, а не собирать информацию опосредованно через опросы или иные выборочные исследования. Применяется и подход *single source*, в рамках которого после установки специальной программы собирается информация из одного источника. Так изучают, к примеру, покупки, поведение и медиа-активность одного и того же респондента в сети Интернет с разных девайсов. Зарубежным примером такого подхода служат исследования Nielsen в США, когда рассматривается просмотр контента на разных платформах, таких как ТВ-приставки с подключением к сети, планшеты, ПК и так далее (Brauer, 2011).

Такие мировые гиганты мира медиа как Netflix успешно внедряют аналитические технологии обработки больших данных для оптимизации процесса подбора контента и его индивидуализации. Помогает это и в разработке новых проектов, особенно в прогнозировании будущего спроса среди аудитории (примером служит кейс «Карточный домик», перед созданием которого также была проведена масштабная аналитика предпочтений американских зрителей в течение шести лет). Уже на этапе разработки концепции появляется возможность оценивать риски вложений в контент, а после его создания выстраивать удобную систему поиска и рекомендаций. Такого рода предиктивная аналитика повышает вероятность успеха медиапродукта до 80% (Еременко, 2019).

Кроме того, современные алгоритмы, такие как Twitter Enabled Supplier Status Assessment (TESSA), используются для проведения мониторинга социального медиа (Huang SF., Su CJ., Saballos MBV, 2018). К тому же сама аудитория становится источником данных для создания контента и даже принимает активную роль разработчика алгоритмов для медиа: Hallinan и Striphos точно описали такой случай поиска коллективного решения для Netflix. Краудсорсинг привел к разработке новых прогностических алгоритмов, которые впоследствии стали основой для новой системы классификации контента (Hallinan & Striphos, 2016).

Конечно, технологии больших данных не свободны от ограничений: высокая цена на эффективные технологические решения, легализация и проблемы с конфиденциальностью данных об аудитории, нехватка образованных профессионалов, давление доминирующих игроков (например, Amazon, Google и т. Д.) и зависимость от доступа к развитой широкополосной инфраструктуре – это лишь некоторые ограничения, которые упоминаются как характерные для применения технологий больших данных в СМИ (Lippell, 2016). К тому же большие трудности в анализе аудитории медиа происходят из-за отсутствия единых метрик и понимания того, как работает конкретный инструмент (например, на одностраничных и даже многостраничных ресурсах



## [Научные статьи]

Сапрыкина Д.И., Козырева М.

*Большие данные в измерениях аудитории цифровых медиа:  
новые возможности для исследований и необходимость  
создания единого измерителя*

Яндекс.Метрика и Google Analytics по-разному будут считать так называемые «отказы», которые важны для финальной аналитики по глубине чтения и общих интересов аудитории). В то же время согласно исследованию Циллнер и ее коллег, представители медиа-менеджмента должны иметь «структуру для общих стандартов, четкого нормативного ландшафта и совместной экосистемы» (Zillner et al, 2014), поэтому нельзя упускать из внимания неунифицированность методик измерения.

Важным этапом в измерении российской интернет-аудитории стал проект TNS Web-Index (был запущен в 2006 году и предоставляет данные о всем населении России возрастом от 12 до 64 лет). В рамках этого проекта изучают использование Интернета с компьютеров, ноутбуков, а также аудиторию mobile. Любопытно, что до 2014 года изучался именно десктоп, а потом на имеющейся выборке стали изучать и mobile, что дает комплексное понимание того, какой контент и где предпочитает потреблять аудитория.

Описывая динамику российского рынка больших данных, в 2014 году генеральный директор компании «АлгоМост» Михаил Левиев упоминал о том, что рынок показывает рост не менее 50% в год (Краузова, 2014). Известные российские холдинги, имеющие в своей структуре медиа, также используют большие данные, опираясь на собственные сервисы и разработки. Примером служат Mail.Ru Group (создали систему управления базой данных Tarantool, запустили свою систему веб-аналитику, а позже стали применять большие данные для таргетирования рекламы, поведенческого анализа, обработки спама и многих других задач на платформах холдинга) и «Рамблер» (начинали использовать большие данные для оптимизации поиска, затем внедрили технологии для блокировки спама, таргетирования рекламы, персонализации контента и обработки текстов на разных языках). Технологии Big Data во всем своем многообразии определений и форм в какой-то момент вышли за пределы IT-сообщества и стали настоящим трендом, частой темой в повестке многих конференций для представителей индустрии медиа (например, AdIndex City Conference, Social Media Moscow, HPF и других).

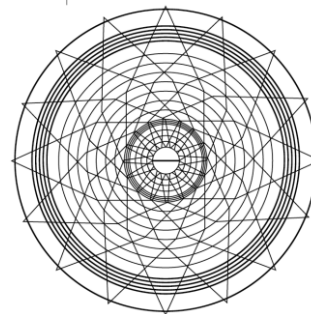
На данный момент медиа – та сфера, в которой компаниям, на первый взгляд, практически нереально унифицировать все форматы информации и соответствующих аудиторных показателей, которые подвергаются анализу: метрики потоковых видео, данные с изображений (с обработкой и без), аудиозаписи, интерактивные таймлайны и карты, данные о глубине просмотра, вовлеченности и многие другие типы информации требуют разработки всё новых технологических решений и создания как минимум единого индустриального глоссария.

Достаточно точную аналитику, и как следствие, построение точных прогнозов, в некоторых случаях позволяет получить комбинация технологий больших данных с дополнительными методами (например, опросами). Таким образом, комплексный анализ разных источников (трекинга, тематического изучения социальных сетей, интерактивных блицев, движения людей и товаров,

## [Научные статьи]

Сапрыкина Д.И., Козырева М.

*Большие данные в измерениях аудитории цифровых медиа:  
новые возможности для исследований и необходимость  
создания единого измерителя*



social listening, видео и так далее) позволяет сформировать цельный профиль пользователя. Однако, при наличии такого опыта, на российском рынке на данный момент нет единых практик и метрик, которые могли бы вывести индустрию медиаизмерений аудитории на новый уровень. Возникает большой вопрос относительно обмена данными между компаниями с целью их обогащения и повышения эффективности. Таким образом, с учетом всего вышеописанного, представленное исследование – только первый шаг к изучению проблемы в российском контексте, и прямой необходимостью является налаживание индустриального диалога, невозможное без участия экспертов-практиков.

### **Методология исследования**

Исследование проводилось в рамках качественной методологии. Для понимания текущей ситуации на рынке медиаисследований мы обратились к экспертным интервью, как основному методу сбора данных. Все интервью проводились по заранее разработанному гайду, состоящему из трех основных блоков:

1) Субъективное понимание эксперта, что такое технологии больших данных и как они применяются в сфере медиаисследований.

2) Опыт применения больших данных в медиаизмерениях на уровне компании

3) Текущая ситуация на рынке медиаисследований в цифровой среде: отношение к вопросу этичности сбора пользовательских данных, к созданию единого измерителя, обмену данными с другими игроками рынка.

В силу особенностей общения с экспертным сообществом мы придерживались полуструктурированного типа интервью, что позволило подстраиваться индивидуально под каждого информанта, учитывая опыт и его специфику работы.

В соответствии с целью и задачами исследования для поиска информантов были выдвинуты следующие критерии. Во-первых, было необходимо, чтобы человек обладал экспертным знанием по тематике исследования. Для этого на первоначальном этапе мы искали людей, которые имеют публикации в научных журналах или в различных нишевых СМИ на тему методов измерения аудитории. Также мы обратились к рекомендациям коллег, занимающихся медиаисследованиями. В результате методом «снежного кома» в выборку попали сотрудники «поисковиков» (Яндекс), «СМИ» (Russia Today, Дискурс, Новая газета), «поставщики технологических решений» (CleverData), «агентства» (Zenith Russia), «паблишеры» (Hearst Shkulev Media), «просветительские проекты» (АНО «Трансперенси Интернешнл», АНО «Информационная культура»). Стоит отметить, что данный метод налагает ограничения на трактовку результатов исследования – экспертные интервью в данном случае позволяют делать ограниченный вывод о конкретных компаниях, но не дают возможности





## [Научные статьи]

Сапрыкина Д.И., Козырева М.

*Большие данные в измерениях аудитории цифровых медиа:  
новые возможности для исследований и необходимость  
создания единого измерителя*

экстраполировать результаты на весь российский рынок медиаизмерений. Тем не менее, благодаря нашим критериям отбора эксперты представляют крупнейшие компании на российском рынке и обладают опытом работы не менее 3 лет, что позволяет нам расценивать их суждения как максимально релевантные.

Отметим, что интервью были проведены в период с сентября по декабрь 2018 года., что может наложить ограничения на актуальность на момент публикации данных. Всего было проведено 10 экспертных интервью, 8 из которых по телефону. Средняя продолжительность интервью составила 15-20 минут. Два информанта ответили на вопросы в письменной форме. Все эксперты были осведомлены о целях и задачах исследования, предупреждены о записи на диктофон, а также согласились на публикацию их интервью без упоминания должностей и компаний.

### **Результаты исследования**

Как мы уже не раз отмечали, употребление понятия Big Data является модным трендом. Однако на сегодняшний день нет однозначного понимания, что относить к этому феномену в сегменте медиаизмерений. Исходя из этого, одной из задач исследования являлось определить, что под этим термином понимают специалисты из медиаиндустрии, работающие с аудиторными данными.

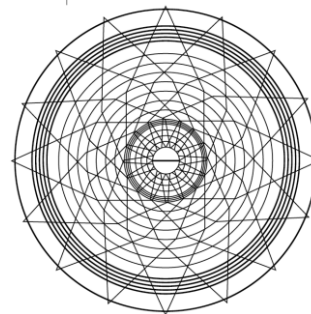
Из общих составляющих - практически все выделяют среди основных характеристик необходимость машинной обработки и большой объем информации. При этом, внутри самого понятия некоторые эксперты выделяют разные составляющие (к примеру, большие данные как тип исследования, большие данные как технология, как феномен с точки зрения общественности, как реклама и таргетинг). В большинстве случаев к использованию технологий больших данных респонденты относят создание алгоритмов и таргетирование. Если говорить о СМИ, то, безусловно, отмечают постепенное развитие data-журналистики и влияние этого на контент и сами редакции. Применение для аудиторного анализа пока не столь развито и в большей степени остается интуитивным (либо скорее качественным в том плане, что некоторые СМИ предпочитают проводить интервью и фокус-группы со своей аудиторией, изредка сочетая это с количественными методами и обработкой большого массива данных под эту задачу).

*“У нас к этому немного спустя рукава относятся, мне кажется, это не очень правильно, потому что, когда ты знаешь аудиторию, ты лучше контент делаешь. Мне кажется, более крупные организации, где больше ресурса могут потратить деньги и время, потратить на анализ, а у нас более интуитивное восприятие или личное знакомство с читателями, хотя было бы здорово анализировать аудиторию с большими данными.*

## [Научные статьи]

Сапрыкина Д.И., Козырева М.

*Большие данные в измерениях аудитории цифровых медиа: новые возможности для исследований и необходимость создания единого измерителя*



*Положительно (отношусь к технологиям больших данных), это главный мировой тренд в журналистике. Особенно в расследовательской журналистике. Ну, если они раньше годами обрабатывали данные для расследования, то сейчас всё больше приходят к такой обработке. Необходимо популяризировать, чтобы было больше людей в редакции, которые умеют с этим работать. Может даже отдельный отдел людей или хотя бы получение дополнительных компетенций. Мне кажется, это необходимо, чтобы и новостники имели минимальный набор техник и знаний, и более серьезные отделы, например расследования и экономика” (интервью 9, СМИ, опыт работы более 3 лет).*

Стоит подчеркнуть, что присутствует существенная разница в видении того, что является инструментарием в данном случае, так как далеко не все респонденты воспринимают однозначно такие доминирующие решения как Google Analytics, Яндекс.Метрика, Ampliflr, Targetings и другие (имеются существенные разногласия на почве того, что упомянутые выше инструменты предоставляют уже так или иначе обработанные данные, которые уже структурированы, в отличие от «классического» необработанного и многообразного формата).

*“Мы комбинируем данные об аудитории, так как у Яндекса есть WebVisor, а у Google более удобные маркеры, которые срабатывают автоматически по тегам. Это разный функционал инструментария. Данные Google и Яндекс расходятся немного, нужны оба, чтобы лучше понимать и ту, и ту, и чтобы понимать, насколько всё точно.*

*Если бы были одни и те же метрики, это было бы куда более удобно. Мы используем Ampliflr для постинга в соцсети, и там тоже есть свои аналитические метрики о времени публикации и о реакциях аудитории. Маркетологам, аналитикам и медиаменеджерам приходится использовать большое количество инструментов и метрик” (интервью 7, самиздат, опыт работы более 5 лет).*

Со стороны представителей информационных проектов и СМИ также был запрос на изначальные, необработанные данные, так как с уже готовой статистикой можно делать выводы и готовить материалы, но при этом нет возможности самостоятельно разобраться, что откуда взято и какие тезисы еще можно было бы извлечь об аудитории или о феномене. Говоря о проблемах, связанных с данным комплексом технологий, выделяют в основном нехватку квалифицированных кадров, отсутствие единых метрик и трудности в сборе качественных данных для дальнейшей обработки. *“Если исходные данные некачественные, на них создаются некачественные проекты. Например, мы организовываем хакатон по данным ДТП на основе открытых государственных данных. Выкладываем данные ДТП, кто-то делает карту ДТП и пытается понять, где чаще всего происходят аварии. А потом оказывается, что в море. Потому что сотрудник в ведомстве перепутал цифру ширины и долготы, неправильно занес. То есть данные испорчены, и этот набор данных сильно влияет. Таких багов много. Если мы такое находим, то мы пишем в ведомство и говорим исправить. На самом деле, если подытожить, то это человеческий фактор. Когда*



## [Научные статьи]

Сапрыкина Д.И., Козырева М.

*Большие данные в измерениях аудитории цифровых медиа:  
новые возможности для исследований и необходимость  
создания единого измерителя*

*включается такая сложная технология и человек недостаточно квалифицирован для этого, происходит ДТП в море” (интервью 5, образовательный проект, стаж более 5 лет).*

Также заметна логика того, что данные технологии так или иначе помещены в контекст общественности: отношение людей к обработке информации (о них самих в том числе) становится отдельным лейтмотивом в обсуждении сбора данных пользователей и подписчиков. *“Мы исходим из позиции, что, если человек выкладывает данные в интернет, он знает, что эти данные будут открыты... И я не вижу основания, по которому мы не можем этим пользоваться. Другое дело, что это довольно слабое доказательство, потому что кто угодно может создать условно копию меня и написать там всё, что угодно. Расследование, которое основано исключительно на таких данных, довольно слабое, поэтому мы стараемся брать два или три доказательства, которые говорят об одном и том же.” (интервью 1, образовательный проект, опыт работы более 3 лет).*

Любопытно отметить, что представители СМИ и технологических компаний расходятся в своем восприятии того, есть ли единогласие по поводу определения больших данных и ряда технологий по обработке, хранению и применению, которые к ним применяются, или таковое отсутствует. Важным аспектом является отношение к популяризации данного комплекса технологий, в частности в медиа.

При оценке влияния возможностей сбора информации при помощи технологий больших данных на развитие рынка и на личную жизнь пользователя, часть экспертов осознает, что существуют аспекты, где можно как навредить, так и принести пользу компании или личной жизни индивида. *«Популяризация больших данных – это для кого-то хорошо, для кого-то плохо. Панацеи нет, зависит от человека. Для меня это хорошо, хотя на оффлайн пока не влияет» (интервью 2, поисковик, стаж более 5 лет).*

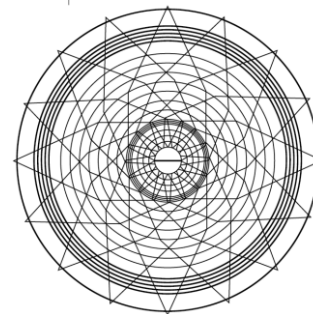
Многие эксперты разделяют отношение личное и отношение профессиональное. Для индустрии такой прогресс в возможностях сбора и анализа разнородных данных, на их взгляд, безусловно хорошо и полезно. А вот манипуляции с информацией с позиции себя, как отдельного пользователя вызывают скорее негативное отношение: *«Для меня как для пользователя нравится, что показывают релевантную рекламу. Паранойи не испытываю, но ситуация как Cambridge Analytica неприятна, как и риск того, что мной будут манипулировать» (интервью 2, поисковик, стаж более 5 лет).*

Также стоит отметить, что ожидания от технологий, связанных с обработкой больших данных, формулируются в двух основных ипостасях: как то, что существенно упростит деятельность компании, и как то, что явно излишне переоценивают и не применяют для задач, связанных с анализом аудитории. *«На мой взгляд, на сегодняшний день для анализа аудитории результаты big data не применяются. Единственное, кто может применить для анализа, это*

## [Научные статьи]

Сапрыкина Д.И., Козырева М.

*Большие данные в измерениях аудитории цифровых медиа: новые возможности для исследований и необходимость создания единого измерителя*



*большие компании(...) Любые большие издатели могут собирать информацию о пользователях, но возможность сбора таких данных сильно ограничена» (интервью 8, издатель, стаж более 7 лет).*

Вопрос о наличии единых метрик и о создании единого измерителя в целом вызывает достаточно много сомнений и разногласий. Большую дискуссию порождает вопрос о том, есть ли необходимость в едином измерителе для цифровой среды, так как на данный момент интернет-реклама по бюджетам обогнала рекламу на ТВ и оценивается объемом в 203 миллиарда рублей по данным АКАР. Соответственно, оценка больших данных об аудитории и единообразии метрик в рамках унифицированного инструментария, по нашим предположениям, должны вызывать скорее одобрение как со стороны представителей рекламных агентств, так и со стороны издателей и СМИ. Однако на практике отношение со стороны специалистов медиаотрасли оказывается не столь однозначным: *«Единый измеритель был бы прогрессом, но это тоже никому не нужно по причине финансов и влияния», «Все экосистемы замкнутые, Google и Яндекс. Это из-за конкуренции, там конкуренция в сфере продуктов, сервиса. Для всеобщего мира в мире и на рынке – это не интересно. Может и здорово было бы иметь единый верификатор, единая система для применения, и чтобы всё промерить, но такого нету, им финансово и политически с точки зрения влияния и власти — это не выгодно. Если Mail.ru и Яндекс согласятся верифицироваться по единому стандарту, кто-то начнет проигрывать с точки зрения качества данных. В рамках этой модели и атрибуции кто-то кому-то проиграет и клиентские деньги перетекают от одного к другому», - объясняет один из наших интервьюируемых (интервью 10, рекламное агентство, стаж более 5 лет).*

Также эксперты отмечают, что создание единого измерителя и возможность открыто обмениваться информацией выглядят привлекательными инициативами, тем не менее, это маловероятно воплотится в реальность в ближайшее время. Объемной темой для дискуссии кажется вопрос этики в отношении сбора данных о пользователях в целом: *«Открытые данные – без проблем. Важна осознанность выкладывания информации, у нас это выбор человека и более глубокое понимание этого придет в ближайшее время. В обезличенном формате тоже нормально собирать данные. Парсить закрытую информацию – уже нет» (интервью 10, рекламное агентство, стаж более 5 лет).*

Общие ожидания лежат в плоскости оптимизации таргетингов и того, как алгоритмы смогут существенно сокращать временные затраты на обработку больших объемов информации, хотя и на этот счет среди экспертов есть тенденция отмечать несколько наивные представления по поводу возможностей технологий. В целом, на взгляд опрошенных нами экспертов, в России еще не пришли к единому взгляду на данные технологии, несмотря на активный интерес к ним: *«У нас не разобрались, что такое big data, поэтому сейчас есть смешные истории, когда обработка 1000 строк подается как работа с big data, лишь бы попасть в тренд» (интервью 9, СМИ, более 3 лет).*



## [Научные статьи]

Сапрыкина Д.И., Козырева М.

*Большие данные в измерениях аудитории цифровых медиа:  
новые возможности для исследований и необходимость  
создания единого измерителя*

Отдельно отметим, что такие медиа, как радио, пресса и, пожалуй, в наибольшей степени телевидение исторически в силу более раннего появления уже прошли определенный путь в отношении формализации измерительной системы и циркуляции информации об аудитории внутри своей сферы. Конечно, есть ряд нюансов (к примеру, прессу до сих пор корректно измерять достаточно трудно и регулирование тиражей становится темой для дискуссий даже сейчас. Это влияет на рекламные бюджеты, которые продолжают сокращаться, существенно оказывая воздействие на финальную цену продукта. Однако даже в сравнении с этим Интернет представляется куда более сложной для измерения сферой со своей технологической и методологической спецификой.

### **Выводы**

На данный момент для СМИ в цифровой среде нет единого измерителя, который позволял бы анализировать аудиторию на достаточно глубоком уровне и унифицировать показатели и метрики. Безусловно, для российских СМИ вопрос анализа аудитории с помощью технологий больших данных сейчас находится на стадии обсуждения. Не все игроки на рынке находят это целесообразным и разделяют позицию иностранных коллег в том, что стоит создавать обмен информацией между паблишерами с целью обеспечения независимости от крупных корпораций, а также обогащения собственных данных об аудитории.

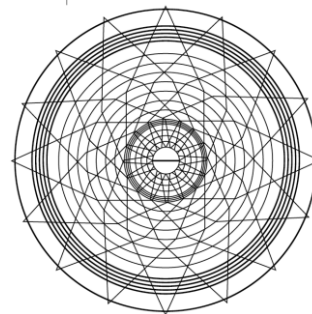
Нельзя сказать, что все понимают необходимость глубокого изучения портрета аудитории в принципе и связь этого процесса с медиапланированием и привлечением рекламных бюджетов. Тем не менее, существует интерес ключевых игроков к созданию единого измерителя и формированию альянса для обмена информацией и выработки единых методик и метрик для измерения и анализа аудитории при помощи технологий больших данных. Серьезными барьерами для этого являются финансовая незаинтересованность наиболее крупных игроков и их опасения по поводу потери конкурентного преимущества.

Подобная инициатива – создание единого измерителя для Интернета – уже фигурировала в индустриальных дискуссиях. Однако не было единогласия по этому поводу как по этическим причинам, так и по финансовым. Когда речь идет о рекламных бюджетах, которые сейчас активно переходят в цифровую среду, большие компании и паблишеры отказываются от обмена данными, не желая терять информационное конкурентное преимущество и прибыль, с этим связанную. Вместо этого доминирующей стратегией представляется наличие большого количества мелких компаний, которые выдвигают свои решения (и тогда у паблишеров возникает вопрос аудита, верификации таких данных и их непредвзятости), а также явное давление со стороны платформ-гигантов, которые обладают наибольшим объемом данных и при этом являются закрытыми системами. Поэтому, на наш взгляд, речь о создании единого измерителя и даже о крупном альянсе для обмена большими данными об

## [Научные статьи]

Сапрыкина Д.И., Козырева М.

*Большие данные в измерениях аудитории цифровых медиа: новые возможности для исследований и необходимость создания единого измерителя*



аудиторных показателях (не в формате уже готовой статистики, как подчеркивали наши эксперты, а в формате raw data) пока не идет, хотя интерес к этому наблюдается.

Как мы уже упоминали, существенным ограничением для трактовки результатов является тот факт, что экспертное мнение относится непосредственно к компаниям и личному опыту каждого респондента. Поэтому экстраполировать ответы и личные оценки на все компании нет возможности. Тем не менее, проведенное исследование позволяет оценить восприятие ситуации достаточно комплексно. Данная работа помогла вскрыть противоречие между интересом к унификации показателей к обмену данными и рядом серьезных внешних факторов, влияющих на невозможность реализации этого на данном этапе развития рынка. Тема обмена аудиторными данными и приведение их к единому знаменателю, на наш взгляд, продолжит фигурировать в индустриальной повестке и потребует ресурсов, открытости, а также общей готовности к изменению качества контента и перераспределению рекламных бюджетов.

## БИБЛИОГРАФИЯ

Вартанов С. А. (2014). Big data в измерении телевизионной аудитории: методы очистки и обработки данных цифровых stb. Теория и практика медиарекламных исследований. Выпуск 4. С. 151–167.

Вартанов, С. А. (2017). Большие данные в онлайн-СМИ: подходы и стратегии использования. Медиаскоп, (4). [электронный ресурс]. URL:<http://www.mediascope.ru/2375>

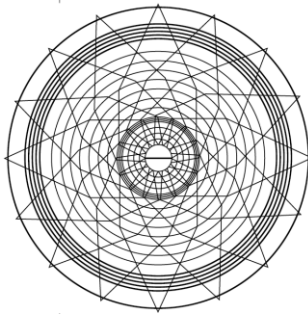
Вартанов, С. А. (2017). Телевизионные измерения в эпоху Big Data: концепции и примеры. Вестник Московского университета. Серия 10. Журналистика, (3).

Давыдов, С. Г. (2007). Методологические основы измерений аудитории СМИ. ОГТРК "Ямал-Регион" в информационном пространстве ЯНАО. Салехард-Надым: ОГТРК "Ямал-Регион", НИЦ "Горизонт-М", 7.

Еременко К. Работа с данными в любой сфере: Как выйти на новый уровень, используя аналитику. Пер. с англ. — М.: Альпина Паблшер, 2019.

Колесниченко О.Ю., Смородин Г.Н., Ильин И.В. и др. Аналитика big data. Новые возможности для понимания глобальных процессов // Security Analysis Bulletin [электронный ресурс]. URL:<http://securityanalysisbulletin.com/api-sociology/10/73/globalistika-2015-analitika-big-data-novye-vozmozhnosti-dlya-pon.htm>

Краузова, Е. (2014). Мастера по добыче данных Российский стартап объединяет экспертов датамайнинга. РБК. [электронный ресурс]. URL:<https://www.rbc.ru/newspaper/2014/04/21/56bee5189a7947299f72d1fc>



## [Научные статьи]

Сапрыкина Д.И., Козырева М.

*Большие данные в измерениях аудитории цифровых медиа:  
новые возможности для исследований и необходимость  
создания единого измерителя*

Официальный сайт компании Mediascope. [электронный ресурс].  
[URL:https://mediascope.net/services/media/media-audience/internet/information/](https://mediascope.net/services/media/media-audience/internet/information/)

Полуэхтова, И. А. (2016). Исследования аудитории и медиапотребления в цифровой среде: методологические и практические проблемы. Медиаскоп, (4). [электронный ресурс]. [URL:http://www.mediascope.ru/2199](http://www.mediascope.ru/2199)

Bilton, R. (2014, June 23). Can attention replace the pageview? Digiday.

Brauer, J. (2011). People First: A User-Centric Hybrid Online Audience Measurement Model. Nielsen, p.9. [URL:https://www.nielsen.com/wp-content/uploads/sites/3/2019/04/nielsen-hybrid-methodology.pdf](https://www.nielsen.com/wp-content/uploads/sites/3/2019/04/nielsen-hybrid-methodology.pdf)

Das, R. (2017). Audiences: A decade of transformations – reflections from the CEDAR network on emerging directions in audience analysis. Media, Culture & Society, 39(8), pp. 1257-1267. <https://doi.org/10.1177/0163443717717632>

Haile, T. (2014). What you think you know about the web is wrong.

Hashem I. A. T. et al. (2015). The rise of “big data” on cloud computing: Review and open research issues. Information Systems, 47. pp. 98-115. <https://doi.org/10.1016/j.is.2014.07.006>

Hallinan, B., & Striphas, T. (2016). Recommended for you: the Netflix Prize and the production of algorithmic culture. New Media & Society, 18(1), pp. 117–137. <https://doi.org/10.1177/1461444814538646>

Huang, S.F., Su, C.J., & Saballos, M.B.V. (2018). Social Media Big Data Analysis for Global Sourcing Realization. In: Bhatia S., Mishra K., Tiwari S., Singh V. (eds) Advances in Computer and Computational Sciences. Advances in Intelligent Systems and Computing, 554, pp. 251-256. [https://doi.org/10.1007/978-981-10-3773-3\\_24](https://doi.org/10.1007/978-981-10-3773-3_24)

Ingram, M. (2014). Chartbeat gets certified to measure attention, tries to move advertising away from clicks and pageviews.

Kosterich, A. & Napoli, P. (2015). Reconfiguring the Audience Commodity: The Institutionalization of Social TV Analytics as Market Information Regime. Rutgers University, New Brunswick, NJ, USA. 17(3), pp. 254-271. <https://doi.org/10.1177/1527476415597480>

Lippell, H. (2016). Big Data in the Media and Entertainment Sectors. In: Cavanillas J., Curry E., Wahlster W. (eds) New Horizons for a Data-Driven Economy. Springer, Cham, 2016. [https://doi.org/10.1007/978-3-319-21569-3\\_14](https://doi.org/10.1007/978-3-319-21569-3_14)

Lowenstein, F. (2014). Why you should pay attention to upworthy measuring engagement in “attention minutes”. Columbia Journalism Review.

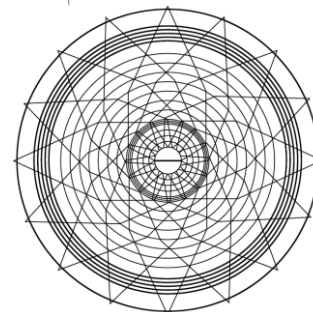
Manyika, J., Chui, M., Brown, B., et al. (2011) Big Data The Next Frontier for Innovation, Competition, and Productivity. McKinsey Global Institute.

Manovich L. (2016). Instagram and contemporary image. 2016.

## [Научные статьи]

Сапрыкина Д.И., Козырева М.

*Большие данные в измерениях аудитории цифровых медиа:  
новые возможности для исследований и необходимость  
создания единого измерителя*



Marks, R. (2013). The Big Opportunity: Audience Research Meets Big Data. Report for the IPA, 2013.

McAfee A. et al. Big data: the management revolution //Harvard business review. 2012. T. 90. N°. 10. P. 60-68.

Nelson, J. & Webster, J. (2016) Audience Currencies in the Age of Big Data, International Journal on Media Management, 18:1, 9-24.

Napoli, P. (2011). Audience Evolution: New Technologies and the Transformation of Media Audiences. New York: Columbia University Press.

Napoli, P. (2012). Audience Evolution and the Future of Audience Research. International Journal on Media Management 14(2): Audience Research and Media Management. <https://doi.org/10.1080/14241277.2012.675753>

Nelson, J. L. (2019). The next media regime: The pursuit of 'audience engagement' in journalism. Journalism. <https://doi.org/10.1177/1464884919862375>

Official page of Emetriq. Über emetriq. [URL:https://www.emetriq.com/](https://www.emetriq.com/)

Official page of Pangaea Alliance [URL:https://www.pangaeaalliance.com/](https://www.pangaeaalliance.com/)

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. Big data, 1(1), pp. 51-59. <https://doi.org/10.1089/big.2013.1508>

Webster, J. G., & Nelson, J. L. (2015). Accounting for attention minutes. Digital Content Next. [URL:https://digitalcontentnext.org/blog/2015/05/11/accounting-for-attention-minutes-as-a-currency/](https://digitalcontentnext.org/blog/2015/05/11/accounting-for-attention-minutes-as-a-currency/)

Woodward, A. & Carter, P. (2015). Analyze the future. Big Data. Innovation in. EMEA in 2015. An IDC White Paper sponsored by SAP and Intel. Business Evolution.

Villars R. L., Olofson C. W., Eastwood M. (2011). Big data: What it is and why you should care. White Paper, IDC. 2011.

Zillner, S., Neururer, S., Munne´, R., Lippell, H., Vilela, L., Prieto, E., Hussain K., & Rusitschka S. (2014). D2.4.2 Final version of sectors roadmap. Public deliverable of the EU-Project BIG (318062; ICT-2011.4.4)

Snijders, C., Matzat, U., & Reips, U. (2012). Big Data: big gaps of knowledge in the field of internet science. International Journal of Internet Science 7(1), pp. 1–5.

Stone, M. (2014). Big Data for Media. Report of Reuters Institute For The Study Of Journalism, University of Oxford, 2014.





## [Научные статьи]

Сапрыкина Д.И., Козырева М.

*Большие данные в измерениях аудитории цифровых медиа:  
новые возможности для исследований и необходимость  
создания единого измерителя*

# BIG DATA TECHNOLOGIES IN RESEARCH AND ANALYSIS OF DIGITAL MEDIA AUDIENCE

### **Saprykina D.I.**

Expert at Laboratory for Mediacommunications in Education,  
HSE (Moscow, Russia)  
[dsaprykina@hse.ru](mailto:dsaprykina@hse.ru)

### **Kozyreva M.**

Student of Bachelor's Programme 'Media Communications' at  
HSE (Moscow, Russia)  
[mariiakozyreva@gmail.com](mailto:mariiakozyreva@gmail.com)

### **Abstract:**

The aim of the study was to determine the media experts' attitude towards the single measurement system creation and data alliance formation, and towards the development of the common measurement methods for digital media audience analysis using the Big Data technologies.

As part of the study, 10 semi-structured expert interviews were conducted with specialists who works with audience data or engaged in the data collection. In particular, the sample included representatives of such companies as Mail.ru, Russia Today, Transparency International-R and others.

This work helped to reveal the contradiction between the interest in standardizing audience currencies for data exchange and a number of serious external factors affecting the impossibility of exchange implementation at this stage of market development. One of the main barriers to creating a single measurement system and sharing information is the financial disinterest of major players and concerns about competitive advantage. A separate vector for discussion is the issue of ethics regarding the collection of user data. Expectations from the Big Data technologies are controversial: some experts believe that the introduction of new measurement methods will greatly simplify the company's operations, while others state that these technologies are overestimated and will not significantly affect audience measurement methods.

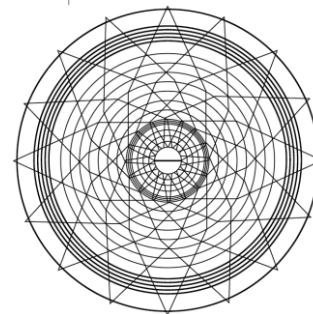
**Keywords:** Big Data, audience measurements, digital media, metrics, analytics, audience research

## REFERENCES

## [Научные статьи]

Сапрыкина Д.И., Козырева М.

*Большие данные в измерениях аудитории цифровых медиа:  
новые возможности для исследований и необходимость  
создания единого измерителя*



Bilton, R. (2014, June 23). Can attention replace the pageview? Digiday.

Brauer, J. (2011). People First: A User-Centric Hybrid Online Audience Measurement Model. Nielsen, p.9. [URL:https://www.nielsen.com/wp-content/uploads/sites/3/2019/04/nielsen-hybrid-methodology.pdf](https://www.nielsen.com/wp-content/uploads/sites/3/2019/04/nielsen-hybrid-methodology.pdf)

Das, R. (2017). Audiences: A decade of transformations – reflections from the CEDAR network on emerging directions in audience analysis. Media, Culture & Society, 39(8), pp. 1257-1267. <https://doi.org/10.1177/0163443717717632>

Davydov. S. G. (2007). Metodologicheskiye osnovy izmereniy auditorii SMI. OGTRK “Yamal-Region” v informatsionnom prostranstve YaNAO. Salekhard-Nadym: OGTRK; “Yamal-Region”; Gorizont. 7.

Eremenko K. Rabota s dannymi v lyuboy sfere: Kak vyiti na novyy uroven. ispolzuya analitiku. Per. s angl. — M.: Alpina Pablsher. 2019.

Haile, T. (2014). What you think you know about the web is wrong.

Hallinan, B., & Striphas, T. (2016). Recommended for you: the Netflix Prize and the production of algorithmic culture. New Media & Society, 18(1), pp. 117–137. <https://doi.org/10.1177/1461444814538646>

Hashem I. A. T. et al. (2015). The rise of “big data” on cloud computing: Review and open research issues. Information Systems, 47. pp. 98-115. <https://doi.org/10.1016/j.is.2014.07.006>

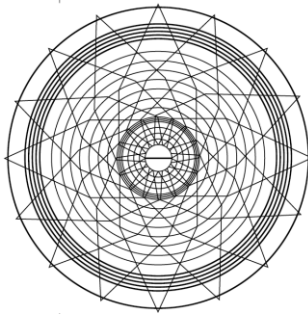
Huang, S.F., Su, C.J., & Saballos, M.B.V. (2018). Social Media Big Data Analysis for Global Sourcing Realization. In: Bhatia S., Mishra K., Tiwari S., Singh V. (eds) Advances in Computer and Computational Sciences. Advances in Intelligent Systems and Computing, 554, pp. 251-256. [https://doi.org/10.1007/978-981-10-3773-3\\_24](https://doi.org/10.1007/978-981-10-3773-3_24)

Ingram, M. (2014). Chartbeat gets certified to measure attention, tries to move advertising away from clicks and pageviews.

Kolesnichenko O.Yu.. Smorodin G.N.. Ilin I.V. i dr. Analitika big data. Novyye vozmozhnosti dlya ponimaniya globalnykh protsessov // Security Analysis Bulletin [elektronnyy resurs]. [URL:http://securityanalysisbulletin.com/api-sociology/10/73/globalistika-2015-analitika-big-data-novyye-vozmozhnosti-dlya-pon.htm](http://securityanalysisbulletin.com/api-sociology/10/73/globalistika-2015-analitika-big-data-novyye-vozmozhnosti-dlya-pon.htm)

Kosterich, A. & Napoli, P. (2015). Reconfiguring the Audience Commodity: The Institutionalization of Social TV Analytics as Market Information Regime. Rutgers University, New Brunswick, NJ, USA. 17(3), pp. 254-271. <https://doi.org/10.1177/1527476415597480>

Krauzova. E. (2014). Mastera po dobyche dannykh Rossiyskiy startap obyedinyayet ekspertov datamayninga. RBK. [elektronnyy resurs]. [URL:https://www.rbc.ru/newspaper/2014/04/21/56bee5189a7947299f72d1fc](https://www.rbc.ru/newspaper/2014/04/21/56bee5189a7947299f72d1fc)



## [Научные статьи]

Сапрыкина Д.И., Козырева М.

*Большие данные в измерениях аудитории цифровых медиа:  
новые возможности для исследований и необходимость  
создания единого измерителя*

Lippell, H. (2016). Big Data in the Media and Entertainment Sectors. In: Cavanillas J., Curry E., Wahlster W. (eds) New Horizons for a Data-Driven Economy. Springer, Cham, 2016. [https://doi.org/10.1007/978-3-319-21569-3\\_14](https://doi.org/10.1007/978-3-319-21569-3_14)

Lowenstein, F. (2014). Why you should pay attention to upworthy measuring engagement in “attention minutes”. Columbia Journalism Review.

Manovich L. (2016). Instagram and contemporary image. 2016.

Manyika, J., Chui, M., Brown, B., et al. (2011) Big Data The Next Frontier for Innovation, Competition, and Productivity. McKinsey Global Institute.

Marks, R. (2013). The Big Opportunity: Audience Research Meets Big Data. Report for the IPA, 2013.

McAfee A. et al. Big data: the management revolution //Harvard business review. 2012. T. 90. N°. 10. P. 60-68.

Napoli, P. (2011). Audience Evolution: New Technologies and the Transformation of Media Audiences. New York: Columbia University Press.

Napoli, P. (2012). Audience Evolution and the Future of Audience Research. International Journal on Media Management 14(2): Audience Research and Media Management. <https://doi.org/10.1080/14241277.2012.675753>

Nelson, J. & Webster, J. (2016) Audience Currencies in the Age of Big Data, International Journal on Media Management, 18:1, 9-24.

Nelson, J. L. (2019). The next media regime: The pursuit of ‘audience engagement’ in journalism. Journalism. <https://doi.org/10.1177/1464884919862375>

Official page of Emetriq. Über emetriq. URL:<https://www.emetriq.com/>

Official page of Pangaea Alliance. URL:<https://www.pangaeaalliance.com/>

Ofitsialnyy sayt kompanii Mediascope. [elektronnyy resurs]. URL:<https://mediascope.net/services/media/media-audience/internet/information/>

Poluekhtova. I. A. (2016). Issledovaniya auditorii i mediapotrebleniya v tsifrovoy srede: metodologicheskiye i prakticheskiye problemy. Mediaskop. (4). [elektronnyy resurs]. URL:<http://www.mediascope.ru/2199>

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. Big data, 1(1), pp. 51-59. <https://doi.org/10.1089/big.2013.1508>

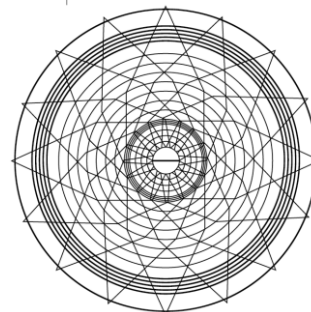
Snijders, C., Matzat, U., & Reips, U. (2012). Big Data: big gaps of knowledge in the field of internet science. International Journal of Internet Science 7(1), pp. 1–5.

Stone, M. (2014). Big Data for Media. Report of Reuters Institute For The Study Of Journalism, University of Oxford, 2014.

## [Научные статьи]

Сапрыкина Д.И., Козырева М.

*Большие данные в измерениях аудитории цифровых медиа:  
новые возможности для исследований и необходимость  
создания единого измерителя*



Vartanov S. A. (2014). Big data v izmerenii televizionnoy auditorii: metody ochistki i obrabotki dannykh tsifrovyykh stb. Teoriya i praktika mediareklamnykh issledovaniy. Vypusk 4. S. 151–167.

Vartanov. S. A. (2017). Bolshiye dannyye v onlayn-SMI: podkhody i strategii ispolzovaniya. Mediaskop. (4). [elektronnyy resurs].  
[URL:http://www.mediascope.ru/2375](http://www.mediascope.ru/2375)

Vartanov. S. A. (2017). Televizionnyye izmereniya v epokhu Big Data: kontseptsii i primery. Vestnik Moskovskogo universiteta. Seriya 10. Zhurnalistika. (3).

Villars R. L., Olofson C. W., Eastwood M. (2011). Big data: What it is and why you should care. White Paper, IDC. 2011.

Webster, J. G., & Nelson, J. L. (2015). Accounting for attention minutes. Digital Content Next. [URL:https://digitalcontentnext.org/blog/2015/05/11/accounting-for-attention-minutes-as-a-currency/](https://digitalcontentnext.org/blog/2015/05/11/accounting-for-attention-minutes-as-a-currency/)

Woodward, A. & Carter, P. (2015). Analyze the future. Big Data. Innovation in. EMEA in 2015. An IDC White Paper sponsored by SAP and Intel. Business Evolution.

Zillner, S., Neururer, S., Munne´, R., Lippell, H., Vilela, L., Prieto, E., Hussain K., & Rusitschka S. (2014). D2.4.2 Final version of sectors roadmap. Public deliverable of the EU-Project BIG (318062; ICT-2011.4.4)