

## ИСПОЛЬЗОВАНИЕ ПОТЕНЦИАЛА BIG DATA В ИЗМЕРЕНИИ ТЕЛЕВИЗИОННОЙ АУДИТОРИИ

**А.В. Посталовский**

Центр социологических и политических исследований БГУ  
ул. Академическая, 25, 220072, г. Минск, Республика Беларусь  
postalnio@tut.by

**Аннотация.** Представленный материал посвящен описанию концепции реализации проекта «FUSION панель» - многоэтапного процесса слияния (интеграции) «больших данных», получаемых по каналам обратной связи с телевизионных приставок РУП «Белтелеком» с данными пиплметрической телевизионной панели, участвующей в проекте национального медиаизмерения (ТВ панель ЗАО «МедиаИзмеритель»). «Большие данные» содержат объемный массив данных о телесмотрении, однако ТВ-приставки не позволяют говорить о социально-демографических характеристиках зрителей, в то время как ТВ-панель сформирована исключительно в контексте демографии аудитории. Объединение этих двух эмпирических массивов позволит, с одной стороны, уточнить данные пиплметрической панели, а с другой стороны – наделить демографическими признаками аудиторию телевизионных приставок.

**Ключевые слова:** телевидение; панель; fusion data; big data; медиаизмерения.

## USING BIG DATA POTENTIAL IN MEASURING TELEVISION AUDIENCE

**A.V. Postalovsky**

Center for Sociological and Political Research BSU  
st. Akademicheskaya, 25, 220072, Minsk, Belarus  
postalnio@tut.by

The presented material is devoted to the description of the concept of the project "FUSION panel" - a multi-stage process of merging (integration) of "big data" received through feedback channels from TV set-top boxes of RUE "Beltelecom" ("MediaMeter"). Big Data contains a large array of data on television viewing, but set-top boxes do not allow us to talk about the socio-demographic characteristics of viewers, while the TV panel is formed exclusively in the context of audience demographics. Combining these two empirical arrays will make it possible, on the one hand, to clarify the data of the people-metric panel, and, on the other hand, to endow the set-top box audience with demographic characteristics.

**Keywords:** television; panel; fusion data; big data; media measurements.

В современных условиях развития общества приобретают значимость исследования, направленные на фиксацию и измерение рейтинговых показателей аудиовизуальных СМИ. Организация и

проведение медиаизмерений позволяет получить комплексное представление о структуре и содержании тематических предпочтений аудитории телевидения. Получаемые показатели востребованности телевизионных каналов делают возможным построение прогнозной аналитики как в отношении развития самого телевизионного сегмента национального информационного поля, так и медиапространства в целом. В указанных контекстах актуализируются вопросы применения инновационных методик медиаизмерений с целью повышения качества эмпирических данных. Применительно к организации медиаизмерений телевидения, проводимых посредством пассивного измерения в рекрутируемой телевизионной панели домохозяйств для уточнения показателей эмпирического массива возможно формирование FUSION панели.

FUSION панель – это интегрированный массив эмпирических «больших» данных (BIG DATA) телевизионных приставок STB (set-top boxes, STB – ресиверы ТВ-приставок РУП «Белтелеком») и эмпирических показателей телевизионной панели, оснащенной пиплметрическим (PM) устройствами фиксации телесмотрения. Сама по себе FUSION DATA (PM + STB) представляет собой многоэтапный процесс интеграции (слияния) данных STB (set-top boxes, STB – ресиверы ТВ-приставок РУП «Белтелеком») – сплошных объемных данных телесмотрения, получаемых по каналам обратной связи (return path data, RPD) без привязки к социально-демографическому блоку (исследователь не знает, кто в настоящее время сидит у экрана телевизора) и репрезентативных данных, получаемых из телевизионной панели PM, которая наделена социально-демографическими показателями (исследователю доступен социально-демографический профиль аудитории).

Исследовательская реализация процесса интеграции телевизионной BIG DATA и пиплметрической телевизионной панели является стадийным процессом. Включающим в себя следующие этапы.

1-й этап: видоизменение объема данных STB - удаление феномена «false positives» («STB on — TV off») – «Кэппинг» длительных сессий телесмотрения (вероятностное предположение, что телевизор выключен, а приставка работает, например, удаляются сессии свыше 300 минут непрерывного просмотра – обрезка со 301-й минуты либо обрезка чрезмерно «Долгой» сессии целиком либо обрезка «начала» и «конца»), обрезка телесмотрения юридических лиц и сельской местности. «Кэппинговый» массив данных в контексте географии должен полностью повторять географию ТВ-панели

2-й этап: Сравнительный анализ показателей STB и PM (выявление общих тенденций и эмпирических расхождений, разработка возможного

понижающего коэффициента STB по показателям Reach, Share, Rtg - основные показатели ТВ-приставок, у РМ гораздо больше метрик). В целом, объемы телесмотрения STB выше, чем у РМ. Кроме того, вырастают показатели тематических или узкопрофильных телеканалов.

3-й этап: Наделение STB социально-демографическими характеристиками панели. Реализация указанного этапа предполагает три возможных направления:

А) Возможный факторный анализ (метод главных компонент) с последующим выделением паттернов (трендов) телесмотрения (например, для 24 домохозяйств панели характерен 5-ти часовой среднесуточный объем телесмотрения, при этом основной временной массив просмотра приходится на период 20.00-23.30, в рамках которого зрители смотрят 1-2 телеканала с акцентом на художественных фильмах или сериалах). Сформированный эмпирический портрет группы домохозяйств берется в качестве переменной в софте обработки информации, которая задается для поиска аналогичного тренда телесмотрения в массиве BIG DATA. В рамках проведенного поиска уточняется точное количество ТВ-приставок, которые максимально точно соответствуют сформированному паттерну телевизионного поведения. Например, 24 домохозяйствам с заданным трендом просмотра соответствует 32693 ТВ приставки. Соответственно, усредненные социально-демографические показатели 24 домохозяйств накладываются вероятностным методом («Приписываются») 32693 ТВ приставкам.

Б) В настоящее время в панели представлено 600 домохозяйств – уникальных единиц исследования с индивидуальным трендом телесмотрения. В качестве основы берется индивидуально каждое домохозяйство и его тренд телесмотрения (среднесуточный объем, количество просмотренных телеканалов, тематическая направленность контента) и сопоставляется с аналогичным трендом в STB. Например, домохозяйству № 1, в котором проживает 3 + человека (с пожилыми) с указанным трендом телесмотрения соответствует 9433 ТВ-приставки. Таким образом, 9433 ТВ-приставки наделяются социально-демографической характеристикой, что это количество приставок – домохозяйства 3 + человека (с пожилыми)

В) В третьем варианте интеграции данных за основу берется сформированная в рамках установочных исследований панельная матрица – количество домохозяйств с определенными контрольными параметрами рекрута в столице, городах 100000 + и городах менее 100000. Для каждой ячейки панельной матрицы формируется в софте обработки информации паттерн (тренд) телесмотрения, который распространяется на ТВ-приставки, соответствующие трендам указанных ячеек. Распределение

внутри ячейки матрицы производится пропорционально значениям, установленным Национальным статистическим комитетом.

Таким образом, в заключении необходимо отметить следующее. Социально-демографический блок, включая сбор информации, последующая валидация, обработка и получение данных с привязкой к возрастным характеристикам и иным классификациям аудитории выступает задачей ТВ-панели. Уточнение данных, в частности, увеличение показателей охвата, доли, данных малых телеканалов, пропуск пустых значений (нулей) – выступает задачей Big DATA. Социально-демографический блок панели накладывается на определенное количество ТВ-приставок (присваивается по умолчанию в программе обработки информации), суточный просмотр которых соответствует трендам и объемам телесмотрения конкретным домохозяйствам. В результате реализации заявленной исследовательской стратегии уточняются рейтинговые показатели ведущих телевизионных каналов и появляются дополнительные данные для нишевых или узкотематических телеканалов.