

## МЕТОДЫ И МЕТОДОЛОГИЯ

DOI: 10.14515/monitoring.2020.2.915

### Правильная ссылка на статью:

Лебедев Д. В. Параданные: определение, типы, сбор и возможное применение // Мониторинг общественного мнения: экономические и социальные перемены. 2020. № 2. С. 4—32. <https://doi.org/10.14515/monitoring.2020.2.915>.

### For citation:

Lebedev D.V. (2020) Paradata: definition, types, collection, and possible uses. *Monitoring of Public Opinion: Economic and Social Changes*. No. 2. P. 4—32. <https://doi.org/10.14515/monitoring.2020.2.915>.



### Д. В. Лебедев ПАРАДААННЫЕ: ОПРЕДЕЛЕНИЕ, ТИПЫ, СБОР И ВОЗМОЖНОЕ ПРИМЕНЕНИЕ

ПАРАДААННЫЕ: ОПРЕДЕЛЕНИЕ, ТИПЫ,  
СБОР И ВОЗМОЖНОЕ ПРИМЕНЕНИЕ

PARADATA: DEFINITION, TYPES, COL-  
LECTION, AND POSSIBLE USES

*ЛЕБЕДЕВ Даниил Вадимович — стажер-исследователь Международной лаборатории исследований социальной интеграции, аспирант школы по социологическим наукам, преподаватель кафедры методов сбора и анализа социологической информации, Национальный исследовательский университет «Высшая школа экономики», Москва, Россия*  
E-MAIL: [zenon-daniil@yandex.ru](mailto:zenon-daniil@yandex.ru)  
<https://orcid.org/0000-0001-7246-8022>

*Daniil V. LEBEDEV<sup>1</sup> — Research Assistant at International Laboratory of Social Integration Research; Ph.D. Student (Sociology); Faculty Member, Department of Sociological Research Methods*  
E-MAIL: [zenon-daniil@yandex.ru](mailto:zenon-daniil@yandex.ru)  
<https://orcid.org/0000-0001-7246-8022>

<sup>1</sup> National Research University Higher School of Economics, Moscow, Russia

**Аннотация.** Преимуществом использования компьютеризированных и онлайн-методов сбора данных является в том числе возможность получения

**Abstract.** One of the advantages of computer-assisted and online data collection tools is also an opportunity to obtain a large array of information about the

широкого массива информации о процессе сбора — параданных. В представленном обзоре рассмотрены работы, применяющие параданные в разных исследовательских сферах; сформированы определение, типологии параданных; выделены основные направления применения такого рода данных, их ограничения, перспективы и возможности сбора. В заключении приведены практические рекомендации относительно максимально эффективной работы с параданными, их использования в качестве исследовательского инструмента.

Обзор будет полезен и интересен академическим и маркетинговым исследователям и методологам (наряду с практиками опросов). Прделанная работа будет способствовать лучшему пониманию такого типа данных, возможностей, связанных с их применением, ограничений и перспектив.

**Ключевые слова:** параданные, сбор данных, сбор параданных, наблюдения интервьюеров

**Благодарность.** Автор статьи выражает благодарность Айгуль Маратовне Климовой (Мавлетовой) за ценные содержательные комментарии и советы, способствовавшие повышению качества этой обзорной статьи.

Статья подготовлена при поддержке Российского научного фонда (грант № 177820172).

collection process, i.e. paradata. The article reviews the works employing paradata in various research areas. The author defines and classifies paradata and identifies the basic directions where paradata can be used paying attention to restrictions and collection opportunities. In conclusion, the author gives practical suggestions for the effective use of paradata as a research tool.

The review will be useful and interesting both to academic and marketing researchers and methodologists (along with practitioners). The paper will develop a better understanding of this data type, opportunities, constraints and future prospects related to its uses.

**Keywords:** paradata, data collection, paradata collection, interviewers' observations

**Acknowledgments.** The authors of the paper expresses gratitude to Aygul M.Klimova (Mavletova) for her valuable substantive comments which helped improve the quality of the review.

The article is supported by Russian Science Foundation (grant No. 177820172).

## Введение

В последние три десятилетия в сфере методологии и методов сбора данных широкое распространение получили компьютеризированные и онлайн-способы сбора данных [Durrant, Kreuter, 2013]. В настоящий момент многие исследования используют стратегии смешивания методов сбора данных (mix-mode), одним из которых зачастую является либо онлайн-опрос [например, Kurakin, 2014], либо опрос с применением компьютерных технологий (телефонный, с использованием планшета) [например, Kozuyeva, Kosolapov, Porkin, 2016]. Кроме того, доказанная эффективность подобных способов сбора данных привела к тому, что в некоторых странах компьютеризированные или онлайн-опросы проводятся (или планируются к проведению) как альтернатива личному интервьюированию с бумажной анкетой в случае переписи населения [Reeves et al., 2018; Никитина, 2018].

Зачастую наиболее важными причинами задействования таких методов сбора данных являются: уменьшение стоимости (особенно в долгосрочной перспективе); снижение затрат на логистику (по сравнению с личными интервью с бумажными анкетами) — печать анкет, перевозка, вбивка данных; сокращение времени, требуемого для перевода имеющихся анкет в формат базы данных [Терентьев, Мавлетова, Косолапов, 2018; Лебедев, Богданов, 2019]. Кроме того, очень важным преимуществом стоит считать возможность сбора дополнительных данных, описывающих разные этапы процесса сбора данных (как сторону интервьюера, так и сторону респондента), — параданные [Мавлетова, 2017].

Термин «параданные» был впервые определен М. Купером в 1998 году как «побочные данные о процессе сбора данных, получаемые в результате использования компьютеризированных методов сбора данных» [Couper, 1998; Kreuter, Couper, Lyberg, 2010]. Важно отметить, что методологические исследования, в рамках которых использовались параданные, существовали и до этого (например, сведения о результате установления связи с респондентом употреблялись для определения более эффективного способа отбора респондентов [Eckman, 2013: 108; Fazio, 1990]). Тем не менее, распространение применения подобного типа данных и более комплексное их осмысление наблюдаются именно в последние два десятилетия [Bosnjak, 2001; Kreuter, Casas-Cordero, 2010; McClain et al., 2019], что связывается с широким распространением способов сбора данных, основанных на использовании компьютерных или онлайн-технологий [Olson, Parkhurst, 2013], а также с повышением интереса к анализу процесса сбора данных, выявлению ошибок и изучению возможностей их сокращения [Yan, Olson, 2013].

Исследования с применением параданных показывают их состоятельность в случае изучения различных аспектов методологии исследований, таких как описание эффектов интервьюера [Sharma, 2019], расчет корректировочных коэффициентов с учетом уровня неотчетов — отказов от ответов (nonresponse adjustment) [Lynn, Nicolaas, 2010], контроль процесса отбора респондентов интервьюерами [Караева, 2015], анализ коммуникации между интервьюером и респондентом и опросного инструментария в целом [Ипатова, 2014; Рогозин, Сапонов, 2014], оценка качества данных [Малошонок, Терентьев, 2014] и т. д. Проще говоря, параданные дают возможность заглянуть в процесс сбора информации, понять особенности поведения респондентов, интервьюеров, ситуации

интервью/прохождения интервью в целом [Kreuter, Casas-Cordero, 2010; Sharma, 2019; Sakshaug, 2013]. Параданные — это сведения о процессе сбора данных, которые зачастую применяются для оптимизации и повышения эффективности такого процесса [Kaczmarek, 2008].

Зачастую наиболее часто используемые типы параданных (например, время заполнения анкеты, результаты установления связи с респондентом, оценка интервьюером ситуации интервью) не требуют дополнительных финансовых вложений для их сбора и анализа [Kreuter, Casas-Cordero, 2010; Improving Surveys with Paradata, 2013; Sharma, 2019]. Это особенно справедливо в случае компьютеризированных и онлайн-опросов, для которых во многих платформах, занимающихся таким методом сбора данных, имеются встроенные инструменты для сбора наиболее общих параданных, которые при этом могут быть сведены к индивидуальному уровню каждого респондента и каждого вопроса [McClain et al., 2019; Терентьев, Мавлетова, Косолапов, 2018]<sup>1</sup>. Стоит оговориться, что имеются также типы параданных, которые связаны с более трудоемкими процессами как сбора, так и анализа (этот аспект будет более детально освещен далее в разделе «Специфика работы с параданными») [Callegaro, 2013].

В русскоязычной научной литературе параданные были описаны в нескольких исследованиях [Сидоров, 2011; Ипатова, 2016; Она же, 2014; Рогозин, Сапонов, 2014; Мавлетова, 2017]. Стоит особенно выделить работы Н. Сидорова, А. Ипатовой и А. Мавлетовой. У первого автора дан хороший обзор на 2011 год с рефлексией по поводу определения, типов и примеров применений параданных [Сидоров, 2011]. А. Ипатовой представлены пример и возможности употребления параданных в контексте телефонных опросов [Ипатова, 2014]. В работе А. Мавлетовой показано, каким образом параданные могут быть использованы для адаптации опросного инструментария под различные характеристики респондентов и ситуации прохождения опроса [Мавлетова, 2017]. Эти статьи помогают получить общее представление о параданных и некоторых потенциальных возможностях их применения. Тем не менее, к сожалению, русскоязычная научная литература не дает комплексного описания параданных как инструмента, к которому широко обращаются при методологических исследованиях. В то же время, подобная работа является ключевой для развития согласованных методологических исследований параданных, их анализа, сбора и применения в практике, так как позволяет определить наиболее эффективные подходы, возможности сбора, границы параданных и их возможностей в целом.

Цель нашей обзорной статьи заключается в представлении общей информации о параданных. Структура обзора следующая: определение термина «параданные»; формирование типологии с учетом разных оснований для разделения различных типов параданных; описание сфер использования параданных в исследовательской практике; описание специфики параданных и трудностей их сбора, анализа и интерпретации; анализ перспективных направлений исследований, связанных с разработкой методов анализа и применения параданных в практике и методологической теории; описание возможностей сбора параданных. В заключении

<sup>1</sup> Также см.: Enjoy Survey. URL: <https://enjoysurvey.com/> (дата обращения: 07.10.2019); Survey Solutions. URL: [https://support.mysurvey.solutions/headquarters/export/paradata\\_file\\_format/](https://support.mysurvey.solutions/headquarters/export/paradata_file_format/) (дата обращения: 10.10.2019).

приведен список практических рекомендаций, которые будут полезны исследователям, планирующим использовать параданные на практике.

Обзор стоит рассматривать как отправную точку для академических и маркетинговых исследователей, методологов и профессионалов опросной индустрии, рассматривающих возможности применения параданных, а также их перспективы, ограничения и требования по затратам ресурсов программного обеспечения.

### **Определение параданных**

На текущий момент в сфере исследований методологии социальных наук отсутствует единое определение этого термина. Имеющиеся примеры концептуализации варьируются от довольно широких [Kreuter, Couper, Lyberg, 2010; Nicolaas, 2011; Matjasic, Vehovar, Manfreda, 2018] до очень узких [Kreuter, Casas-Cordero, 2010; McClain et al., 2019]. Нередко имеющиеся определения могут даже в какой-то степени противоречить друг другу [Kreuter, Casas-Cordero, 2010; McClain et al., 2019].

Наиболее часто упоминается уже приведенное в нашей статье определение М. Купера, приведенное выше и представленное на конференции Американской социологической ассоциации [Couper, 1998; Sharma, 2019]. Тем не менее, популярность этой дефиниции, скорее, объясняется тем, что именно в этом выступлении термин был впервые упомянут и описан. Развитие исследований с применением параданных на текущий момент указывает, что такое определение упускает значительную часть параданных, используемых для поправки на ошибку неотчетов (например, наблюдения и оценки интервьюеров относительно респондента, домохозяйства, района и ситуации интервью в целом) [Kreuter, Couper, Lyberg, 2010; Couper, Kreuter, 2013].

В этом контексте необходимо выделить наиболее важные характеристики параданных, а также кратко описать их отличие от сходных типов данных: вспомогательных (auxiliary data) и метаданных (metadata). Важно заметить, что нередко исследователи методологии социальных наук пренебрегают обсуждением возможного определения параданных в пользу обсуждения способов их сбора и использования [Lynn, Nicolaas, 2010].

Параданные могут быть описаны как дополнительные сведения, измеряемые в процессе заполнения опроса/интервью [Durrant, Kreuter, 2013; West, 2011; Matjasic, Vehovar, Manfreda, 2018]. Это данные о самом процессе, включающие поведенческие характеристики интервьюера, респондента, описание ситуации прохождения опроса/интервью, в том числе измеряемые с помощью вспомогательных устройств [Kreuter, Casas-Cordero, 2010]. Параданные недоступны перед проведением полевого этапа, но возникают и меняются в этом процессе, при этом они предоставляют возможность увидеть наиболее значимое поведение респондентов и интервьюеров при сборе данных (проведение интервью, заполнение онлайн-анкеты и т. п.) [Kaczmirek, 2008; McClain et al., 2019; Global Guideline: Online Research ESOMAR].

Выделяют несколько важных особенностей параданных. Они фиксируются на каждом этапе сбора данных [Finnamore, 2019]; в подавляющем большинстве случаев являются нереактивными (то есть собираются без ведома или согласия

со стороны респондентов, а также без указания на цели сбора такого рода данных — эта особенность и сопутствующие ей проблемы будут подробнее описаны в разделе «Специфика работы с параданными» нашей статьи) [Callegaro, 2013]; имеют различные уровни детализации и генерализации [Kaczmirek, 2008]; зачастую отличаются для каждого отдельно взятого респондента [Kaczmirek, 2008; Callegaro, 2013].

Параданные неразрывно связаны с метаданными и вспомогательными данными [Kreuter, Couper, Lyberg, 2010]. Нередко параданные описываются как одна из разновидностей метаданных (поведенческие) [Ипатова, 2014; Сидоров, 2011]. Тем не менее, в последние годы зачастую параданные выделяют как отдельную категорию наравне с метаданными [Smith, 2011; Kreuter, 2013; Мавлетова, 2017].

Метаданные описывают решения относительно различных этапов сбора данных, позволяющие понять структуру имеющихся данных [Мавлетова, 2017; Сидоров, 2011]. В отличие от параданных, такие сведения могут быть доступны до начала самого процесса их сбора, не отличаются по респондентам и интервьюерам (поскольку описывают данные макроуровня) и позволяют понять, каким образом были собраны данные (построение выборки, инструментарий, методы сбора данных и т. п.), но не дают возможность анализировать поведение интервьюеров или респондентов и ситуацию каждого отдельно взятого интервью [McClain et al., 2019].

Вспомогательные данные содержат информацию о респондентах или их группах, позволяющую использовать в анализе сведения, на которые не было получено прямого ответа от респондента в процессе интервью, но которые имеются в распоряжении у исследователей (например, связь результатов опроса с административными базами данных) [Kreuter, 2013]. Вспомогательные данные отличаются от параданных тем, что так же, как и метаданные, зачастую существуют до начала процесса их сбора; не обязательно описывают такой процесс (нередко это информация из административных источников о характеристиках респондента, интересующих исследователей); перед началом их сбора можно определить, какие типы параданных собирать дополнительно, тогда как для вспомогательных данных приходится ограничиваться тем, что есть изначально [Smith, 2011].

Обобщая приведенные особенности параданных, стоит отметить, что такой тип сведений описывает поведенческие характеристики респондентов и интервьюеров в процессе интервью/опроса, ситуацию интервью в целом и особенности респондента, домохозяйства или района, в котором проходило интервью, отмечаемые интервьюером.

Выделяются следующие отличительные черты параданных:

- они недоступны перед проведением полевого этапа, но появляются и меняются в процессе сбора данных<sup>2</sup>;
- фиксируются на каждом этапе сбора данных;
- являются нереактивными (не предполагают активного участия респондентов в их сборе и зачастую не указывают отвечающему на то, что подобные сведения собираются и будут каким-то образом использованы в анализе);

<sup>2</sup> Исключение составляют параданные из предыдущих исследований или волн исследований выбранного респондента. Особенно это релевантно в случае панельных исследований.

- имеют различные уровни детализации и генерализации (например, среднее время заполнения всей анкеты, время заполнения анкеты одним респондентом, среднее время заполнения одного вопроса, время заполнения отдельно взятого вопроса одним респондентом), в том числе могут описывать индивидуальный уровень одного респондента и отдельного вопроса.

### Типы параданных

В этом разделе будет представлена типология параданных с учетом наиболее часто выделяемых характеристик и методов, позволяющих собирать такие данные. Важно заметить, что построение типологии усложняется с учетом того факта, что их формат, размер, аналитическая ценность данных и их наличие как таковое различаются как по методам сбора (личное интервьюирование с бумажной анкетой (paper and pencil personal interview — PAPI), с планшетом (computer-assisted personal interview — CAPI), телефонное интервью с помощью компьютера (computer-assisted telephone interview — CATI), онлайн-интервью (computer-assisted web interview — CAWI) и т. д.), так и по типам исследований (срезовые, лонгитюдные, панельные). То же справедливо относительно качества данных: в зависимости от типов их сбора, дизайна и контекста исследования в целом качество различных типов параданных может значительно варьироваться [West, Sinibaldi, 2013].

Прежде всего будут приведены наиболее часто упоминаемые типологии из зарубежной литературы; далее будет представлена обобщающая типология с учетом наиболее упоминаемых в зарубежных исследованиях подтипов параданных; после этого будут указаны ключевые характеристики, позволяющие разделять разные типы параданных.

М. Каллегаро предлагает разбиение на параданные об устройстве и параданные о навигации в опроснике [Callegaro, 2013; Callegaro et al., 2015]. Первые содержат релевантные сведения относительно характеристик компьютера или мобильного устройства, используемого при подключении к опросу (например, строка «пользователь — агент» (user agent string), в которой содержатся сведения о браузере, языке операционной системы и проч. — хороший обзор типов данных и возможностей работы с таким типом данных содержится в [Robmann, Gummer, Kaszmirsek, 2020]). Вторые описывают действия респондента при заполнении опроса (прокручивание страницы, изменение масштаба) и позволяют реконструировать непосредственно процесс указания респондентом ответа на заданный/представленный вопрос.

Имеется также разделение на прямые и непрямые параданные [Lynn, Nicolaas, 2010; Callegaro, 2015; Callegaro, Manfreda, Vehovar, 2015]. Оно исходит из ответа на вопрос, нужны ли дополнительные усилия и затраты временных и денежных ресурсов для получения такой информации. Прямые параданные записываются автоматически в процессе их сбора (например, время начала интервью, движения компьютерной мышью в процессе заполнения опроса), а непрямые требуют отдельных усилий по разработке инструментария и подключения дополнительных считывающих устройств (направление взгляда, размер зрачка и т. п.) [Galesic et al., 2008].

К. МакКлейн с коллегами выделяют четыре подтипа параданных с учетом различных видов исследований (срезовые, лонгитюдные, панельные) [McClain et al., 2019]: параданные из предыдущих исследований, параданные об установлении контакта/связи с респондентом, обращение к анкете со стороны респондента, представление ответа. На первом уровне используются данные об участии респондента в предыдущих волнах исследования или других исследованиях для выделения некоторых ключевых характеристик (например, средняя скорость представления ответов, использованные устройства). Второй уровень содержит имеющиеся сведения о процессе установления контакта/связи с респондентом (факт открытия письма, переход по ссылке и т. д.). Третий уровень отвечает за доступ к самому опроснику (используемое устройство, его характеристики, количество возвратов к анкете и проч.). Четвертый уровень включает информацию о поведении респондента в контексте представления ответов на вопросы (например, длительность ответа, переход на другие окна в браузере, движения мыши).

Важно отметить, что это разделение нестрогое и, скорее, должно восприниматься как разные уровни описываемых процессов. Информация об устройстве может быть использована как на уровне представления ответов, так и на уровне доступа к опросу; количество пропущенных ответов может быть характеристикой респондента и на уровне предыдущих исследований, и на уровне представления ответов на актуальный опрос [McClain et al., 2019: 199—201].

Обобщая имеющиеся типологии параданных и описываемые подтипы [Kreuter, Casas-Cordero, 2010; Smith, 2011; Погозин, Ипатова, 2015; Мавлетова, 2017; Сидоров, 2011; Callegaro, Manfreda, Vehovar, 2015; Kreuter, 2017a; Kaczmirek, 2008; Sharma, 2019; Sirkis, McGovern, 2019; Jans et al., 2019], можно выделить следующие пять типов параданных.

1. *Наблюдения интервьюера* — описание им ключевых особенностей ситуации интервью (факт наличия третьих лиц, вовлеченность респондента (по мнению интервьюера), оценка причины отказа от участия в интервью), характеристик домохозяйства (вероятность наличия детей в домохозяйстве, наличие партнера у респондента), района.
2. *Информация об устройстве, используемом для участия в опросе*, — данные: IP-адрес, тип подключения, браузер, операционная система, разрешение экрана, язык операционной системы, размер окна браузера, наличие в браузере JavaScript и Adobe Flash, их статус (активный/неактивный), cookie-файлы, user agent string, местоположение (GPS-локация), активность окна браузера с опросом.
3. *Данные о поведении в процессе заполнения анкеты (лог-файлы)* — скорость печати, движения и клики мышью, нажатие клавиш, прокручивание страницы, время (длительность интервью, ответы на отдельные вопросы), дата и время начала и окончания интервью, количество и последовательность исправлений, возвращение к предыдущим частям анкеты/вопросам, страница/вопрос, на котором случился разрыв — недозаполнение респондентом анкеты (dropout), предупреждения об ошибках при заполнении анкеты (количество, время, страница появления), последовательность заполнения вопросов, выбора ответов, факт возвращения к заполнению опроса после



выхода из опроса, в случае рандомизации вариантов ответов — их последовательность в каждом отдельном случае, увеличение масштаба.

4. *Параданные об интервьюерах и усилиях в процессе сбора данных (level-of-effort)* — результаты дополнительных попыток установки контакта/связи с респондентом, путь интервьюера в процессе сбора данных (последовательность GPS-локаций), данные о контакте/связи с респондентом (call record data — время, дата, количество попыток контакта, описание/оценка характеристик процесса), история контактов/попыток связи с респондентом (данные о процессе связи с респондентом (contact paradata — успешность, количество, время)), характеристики интервьюеров.
5. *Вербальные и физиологические параданные* — аудиозапись интервью (комментарии и вопросы респондентов), речевые характеристики (скорость, паузы, напряженность голоса), движения взгляда, размер зрачков.

Последний тип не полностью отвечает ключевым принципам определения параданных, которое было приведено выше, относительно отдельных подтипов (особенно нереактивность в случае замера направления взгляда и размера зрачка). В этом случае можно оговориться, что, как правило, подобный тип параданных используется в экспериментальном (лабораторном) контексте и может считаться подходящим под цели методологических исследований, поскольку также описывает поведение респондентов в процессе заполнения опроса.

Приведенный список параданных не является конечным, но в нем указаны переменные (подтипы), относящиеся к различным типам параданных, обладающие наибольшей аналитической ценностью. Основанием для их объединения являлась встречаемость в исследовательской литературе.

Далее будут представлены несколько базовых характеристик параданных, выделяющих различные места сбора (сторона сервера/сторона пользователя), содержательные аспекты данных (устройство/навигация), этапы сбора данных (участие в предыдущих исследованиях/установление связи с респондентов/обращение к анкете/представление ответа), специфику сбора (прямые/непрямые), а также уровни агрегации (исследование/секция/ вопрос/действие).

Одной из важнейших характеристик параданных является место их сбора. Параданные могут собираться на стороне сервера (server-side) и на стороне клиента (пользователя или устройства) (client-side) [Heerwegh, 2003; Matjasic, Vehovar, Manfreda, 2018; Сидоров, 2011; Kaczmirek, 2008; Callegaro, 2013].

Серверная сторона сохраняет параданные, получаемые сервером, через который исследователям передаются данные из «поля». Такие параданные описывают детали посещения респондентом или интервьюером отдельных страниц онлайн-опросника/процесса заполнения опроса на планшете. Формат получаемых сведений в таком случае стандартизован для каждой страницы опроса и описывает результаты действия на странице. Ценность сведений, собираемых таким способом, выше, когда на странице меньше вопросов — в идеале один, тогда полученная информация может быть использована для описания процесса представления ответа на отдельно взятый вопрос.

Сторона пользователя собирает информацию (зачастую с помощью JavaScript) на самом устройстве и дает более широкий выбор подтипов параданных, кото-

рые могут быть записаны (движения мыши, нажатия клавиш, различные уровни затрачиваемого на опрос времени и т. д.), нежели сторона сервера. Параданные, собираемые на стороне устройства, описывают все события, в том числе внутри страницы. При этом количество получаемых данных возрастает, так же как и сложность их структуры и, соответственно, трудоемкость их обработки и анализа. В целом параданные, собираемые на стороне клиента (устройства/пользователя), отличают их точность, количество собираемой информации и тот факт, что этот тип параданных содержит больше информации о процессе заполнения опроса. Кроме того, в случае данных стороны клиента требуется наличие JavaScript на странице, тогда как данные стороны сервера собираются вне зависимости от характеристик устройства респондента и не требуют дополнительных затрат на программное обеспечение. Относительно оценивания длительности заполнения отдельных вопросов и опроса в целом параданные стороны клиента дают более точные сведения, так как данные стороны сервера включают в это время длительность передачи данных с устройства на сервер. Хотя эта разница редко превышает три-четыре секунды [Kaczmirek, 2008; Callegaro, 2013], в больших опросах это может значительно влиять на последующие результаты анализа параданных.

Дополнительным преимуществом параданных, получаемых со стороны пользователя, является тот факт, что при покидании опроса до его окончания данные серверной стороны о заполнении последней страницы не сохранятся, тогда как сторона клиента сразу запоминает любые полученные сведения вне зависимости от того, продолжит ли респондент опрос.

М. Каллегаро добавляет к этой характеристике также параданные, собранные с помощью приложений, указывая на то, что информация, собранная таким образом, также может быть расценена как параданные, но требует более активных действий по ее сбору со стороны респондентов (как минимум, отпадает нереактивность, поскольку респондентам необходимо самим установить приложение, предварительно ознакомившись с описанием целей сбора и последующего анализа подобных данных) [Callegaro, 2015]. С нашей точки зрения, описанные данные не отвечают критериям представленного выше определения параданных и, следовательно, их надо относить не к параданным, а, скорее, ко вспомогательным данным.

Важной характеристикой параданных являются различные уровни агрегации получаемых сведений (уровень исследования, секции, вопроса, действия) [Kaczmirek, 2008; Olson, Parkhurst, 2013]. Первый уровень относится к наиболее объемному типу параданных, которые записывают каждое отдельно взятое событие (движение мышью, нажатие клавиш и т. д.). Второй уровень содержит информацию одного типа для отдельного респондента, но по нескольким переменным (например, количество изменений вариантов ответов для каждого вопроса, средний уровень размера зрачка для каждого вопроса). На третьем уровне этот процесс происходит относительно либо всех переменных для одного респондента (длительность интервью, местоположение и проч.), либо одной переменной по всем респондентам (доля пропущенных данных для одной переменной и др.). На четвертом уровне происходит обобщение параданных по всем переменным и всем респондентам (среднее время заполнения опросника, доля ответивших и т. п.). Количество получаемых данных, доступных для анализа, снижается при

повышении уровня, тогда как совместно с этим повышается уровень агрегации [Kaczmirek, 2008: 63—65].

### Направления применения параданных

Относительная новизна, продолжающееся интенсивное развитие технологий и возможностей замера данных разного типа приводят к тому, что список типов параданных и направлений их возможного использования постоянно расширяется. В этом разделе я сконцентрируюсь на наиболее популярных исследовательских направлениях применения параданных с целью показать широту возможностей, которые несут их сбор и анализ.

Существуют две основных стратегии работы с параданными: их использование после завершения сбора данных и анализ в процессе [Kreuter, 2017b]. На текущий момент в большинстве случаев параданные применяются после окончания процесса сбора данных [Kreuter, Couper, Lyberg, 2010].

Чаще всего параданные употребляются в качестве основных индикаторов различных типов ошибок и смещений в рамках теории общей ошибки опроса (Total Survey Error framework — TSE) [Корытникова, 2018; McClain et al., 2019; Караева, 2015; Погозин, Сапонов, 2014; Сидоров, 2011; Kreuter, 2017a]:

- ошибка покрытия может быть оценена с помощью общих параданных относительно времени и места проведения интервью/опроса [Kreuter, Casas-Cordero, 2010; Eckman, 2013];
- параданные о местоположении и перемещении интервьюера в процессе сбора данных могут быть использованы для оценки ошибки выборки [West, 2011; Wagner, Olson, Edgar, 2017; Choumert-Nkolo et al., 2019; Elevelt et al., 2019] наравне с более простыми типами параданных (уровень достижимости, кооперации) [Караева, 2015];
- наблюдения интервьюеров и параданные о процессе установления контакта/связи с респондентом могут быть полезны для изучения ошибки неотчетов (nonresponse error) и для расчета погрешности корректировки данных [Kreuter, 2017a; Kreuter, 2017b; Nicolaas, 2011];
- данные о нажатиях на клавиатуру, движениях и нажатиях мышью, а также о поведении респондента в процессе заполнения опроса в целом могут быть использованы для оценки ошибки измерения (measurement error) и валидности данных [Kreuter, 2017a; Nicolaas, 2011; Smith, 2011; Lynn, Nicolaas, 2010; Mavletova, Couper, Lebedev, 2018].

Согласно приведенной выше типологии К. МакКлейн с коллегами выделяют такие соответствия параданных возможному применению в контексте теории общей ошибки опроса [McClain et al., 2019: 199—201]: параданные из предыдущих исследований или волн исследования могут быть использованы в контексте изучения ошибки измерения и ошибки неотчетов; параданные этапа установления связи/контакта с респондентом подходят для оценки ошибки покрытия (coverage error) и ошибки неотчетов; данные относительно доступа к опросному инструментарию могут быть полезны в контексте изучения ошибки измерения и неотчетов; на этапе непосредственного участия респондента в опросе/интервью получаемые параданные могут пригодиться для оценки ошибки измерения.

Вне теории общей ошибки опроса параданные также имеют широкие исследовательские перспективы. Они позволяют оптимизировать процесс сбора данных, определить наиболее подходящие день недели и время дня для связи/установления контакта с респондентом [Мавлетова, 2017; Сидоров, 2011; Kreuter, 2017a; Durrant, Kreuter, 2013; Durrant, D'Arrigo, Steele, 2011], что может потенциально повысить уровень отклика или более точно определить вероятность отклика отдельных респондентов [Biemer, Chen, Wang, 2013]. Кроме того, сведения такого рода (например, длительность ответов на отдельные вопросы) могут быть основой для предоставления респондентам помощи или другой дополнительной информации, способной повысить качество получаемых данных [Kreuter, 2017b; Conrad et al., 2017].

Изучение поведения респондентов в процессе заполнения опроса/участия в интервью также предполагает использование параданных. В фокусе изучения могут находиться различные вопросы, решение которых зачастую приводит к изменениям в программном обеспечении опросных платформы в целях повышения качества получаемых данных. Это, например, изменение ответов, взаимодействие с активными элементами различных типов вопросов, изменение вариантов ответов [Yan, Olson, 2013]. Возможно также использование данных о движении респондента в процессе заполнения опроса для оценки уровня вовлеченности с помощью данных со смартфонов [Höhne, Revilla, Schlosser, 2019]. Кроме того, интересен пример М. Босняка, который на основе имеющихся параданных выделил семь типов респондентов на основе их поведения при заполнении опросника [Bosnjak, 2001; Bosnjak, Tuten, 2001; Callegaro, 2013].

Адаптивный дизайн — приспособление элементов опросного инструментария и процесса сбора данных под характеристики респондентов (вероятность участия) и определение оптимального времени окончания различных этапов сбора данных (в том числе введение дополнительного метода сбора данных) [Мавлетова, 2017; Kreuter, 2017a; Callegaro, 2014; Kirgis, Lepkowski, 2013]. В этом контексте могут быть использованы различные типы параданных (наблюдения интервьюеров, сведения об установлении связи/контакта с респондентом, поведение в рамках заполнения опроса), в том числе из предыдущих исследований/волн панельного исследования. Адаптация может проходить как для отдельных респондентов/домохозяйств, так и для отдельных труднодоступных групп в целях повышения уровня отклика и снижения уровня недозаполнения опроса [Мавлетова, 2017; West, 2011].

Назовем другие возможные направления использования параданных.

- Это может быть *контроль качества собираемых данных*, в том числе на предмет фальсификаций или фабрикаций со стороны интервьюеров [Рогозин, Ипатова, 2015; Laflamme, Maydan, Miller, 2008; Ипатова, 2014; Dajani, Marquette, 2015] и для оценки эффекта интервьюера [Sharma, 2019].
- *Корректирование опросного инструментария по ходу сбора данных* может происходить в результате анализа комментариев интервьюеров, анализа последней страницы перед уходом из опроса [Ипатова, 2016; Она же, 2014; Мавлетова, 2017; Kaczmirek, 2008]. Особенно это релевантно для панельных и лонгитюдных обследований, когда инструментарий значительно

не меняется, но его можно подкорректировать для получения информации более высокого качества.

- Возможна оценка качества дизайна и методологии исследования — значения достижимости, кооперации, отказов позволяют сравнивать исследования с разными дизайнами и проведенные в разных контекстах. [Звоновский, Соловьева, 2017; Караева, 2015; Matjasic, Vehovar, Manfreda, 2018; Stieger, Reips, 2010].
- Доступ к анализу коммуникации между интервьюером и респондентом в процессе сбора данных/контакта возможен благодаря анализу замечаний, комментариев интервьюеров и аудиозаписей интервью [Ипатова, Рогозин, 2014; Рогозин, Сапонов, 2014; Kreuter, Couper, Lyberg, 2010; Olson, Parkhurst, 2013]. Важно заметить, что в случае использования CATI как метода сбора данных может происходить одномоментный контроль со стороны супервайзера (одновременно со сбором данных), тогда как в других типах исследований с аудиозаписью коммуникации (CATI, CASI<sup>3</sup>) такая возможность отсутствует и прослушивание доступно только после окончания процесса сбора данных [Olson, Parkhurst, 2013].
- Параданные могут быть использованы в качестве дополнения к содержательным данным. Например, наблюдения интервьюеров и оценка степени кооперации и откровенности респондентов можно учитывать при расчете корректирующих коэффициентов [Ипатова, Рогозин, 2014; Рогозин, Сапонов, 2014; Nicolaas, 2011; Kreuter, Couper, Lyberg, 2010; Groves et al., 2007].
- С помощью параданных возможна оценка когнитивной нагрузки, возникающей у респондента или интервьюера при заполнении опросного инструментария. Для этого могут быть использованы показатели затрачиваемого времени на вопрос [Matjasic, Vehovar, Manfreda, 2018; Kaczmirek, 2008], размеров зрачков [Девятко, Лебедев, 2017; Девятко, 2018], изменения вариантов ответа [Nöhne, Schlosser, Krebs, 2017], движения мыши [Horwitz et al., 2017].

### **Специфика работы с параданными: проблемы и ограничения**

Работа с параданными не лишена ограничений. В этом разделе будут описаны наиболее важные проблемные области и преграды на пути сбора, анализа и интерпретации параданных.

Значимое ограничение при сборе параданных проявляется при использовании комбинированных количественных методов (например, PAPI и CAPI в случае РМЭЗ НИУ-ВШЭ [Kozyreva, Kosolapov, Popkin, 2016; Лебедев, Богданов, 2019]). При этом некоторые типы параданных могут измеряться по-разному в случае отдельных методов сбора данных, что, несомненно, связано с появлением ошибки измерения. Так, длительность опроса может оцениваться интервьюером самостоятельно при опросе с бумажной анкетой, тогда как при использовании планшета эти данные будут собираться автоматически и будут более точными, при этом и те, и другие параданные в итоговой базе данных будут содержаться в одной и той же переменной [Kreuter, 2017b; McClain et al., 2019].

<sup>3</sup> Самостоятельное заполнение опросника с использованием планшетного компьютера (computer-assisted self-interview).

Качество параданных (валидность, надежность, доля пропущенных данных) значительно отличается от одного подтипа к другому и зависит от типа, используемого метода и контекста исследования (время, тематика, характеристики района, домохозяйства, интервьюера, респондента) [West, Sinibaldi, 2013; Мавлетова, 2017; West, Kreuter, 2013; Nicolaas, 2011].

Специалистами отмечается также наличие некоторых ошибок измерения относительно отдельных подтипов параданных. Например, время реакции на представленный вопрос может варьироваться в зависимости от браузера, операционной системы, используемого устройства. В этом случае разница может быть минимальной (несколько миллисекунд), тем не менее, ее все равно нужно принимать во внимание [Stieger, Reips, 2010; Мавлетова, 2017; Kreuter, Casas-Cordero, 2010; West, 2011; West, Sinibaldi, 2013].

Возможно также появление пропущенных данных, что может значительно затруднять анализ и интерпретацию параданных [Kreuter, 2017b; Smith, 2011; Yan, Olson, 2013]. Некоторые типы параданных могут меняться на протяжении всего процесса заполнения одним респондентом отдельного опроса (например, продолжение заполнения опроса с другого устройства) и в случае данных стороны сервера не могут быть записаны полностью для одного респондента [Callegaro, 2013].

Комментарии и наблюдения интервьюеров являются одними из самых популярных типов параданных, при этом качество их измерения сильно подвержено субъективности интервьюеров, а сами они сложно стандартизируемы [Ипатова, 2016; Olson, Parkhurst, 2013]. К тому же, необходимость дополнительного замера отдельных параданных интервьюерами (заметки, комментарии, субъективные оценки), помимо ведения коммуникативной ситуации интервью и непосредственно общения с респондентом может приводить к повышенной когнитивной нагрузке на интервьюера. Результатом этого может стать снижение качества проведения интервью, получаемых данных и параданных [Мавлетова, 2017; Девятко, Лебедев, 2017; Smith, 2011].

Важная особенность параданных заключается в том, что большинство их типов может быть собрано без дополнительных затрат (как временных, так и денежных), а это может являться расслабляющим фактором для исследователей. В такой ситуации продумывание плана работы с параданными и их анализа может откладываться на этапы, наступающие после сбора данных. Тем не менее, ученые отмечают, что все измеряемые параданные должны собираться с заранее определенной целью и проработанной структурой анализа, чтобы быть действительно полезными [Ипатова, 2016; McClain et al., 2019; Heerwegh, 2003; Olson, Parkhurst, 2013]. Особенно это важно в контексте более сложных типов параданных (движения мыши, переход к другому окну браузера, нажатия клавиш), поскольку для их сбора зачастую требуется использование дополнительного программного обеспечения.

До сих пор нет явного консенсуса относительно этической стороны работы с параданными. С одной стороны, правовые и этические требования указывают на необходимость уведомлять респондентов о том, что их действия записываются [Global Guideline: Online Research ESOMAR; Сидоров, 2011; Kaczmirek, 2008]. С другой стороны, в таком случае возрастает вероятность отказа респондентов от записи параданных либо от участия в исследовании в целом [McClain et al.,

2019]. По результатам серии экспериментов с представлением виньеток уровень согласия на запись параданных варьировался от 59,5% в Нидерландах до 65,6 и 69,3% в США [Couper, Singer, 2013; Callegaro, 2015; Callegaro, 2014].

Параданные также варьируются по степени анонимности и опасности с точки зрения сохранности частной информации (например, время заполнения опросника, движения мышью, местоположение и аудиозапись интервью). По этой причине зачастую исследователи публикуют результаты и базы данных без добавления параданных, что затрудняет активное развитие сферы и дискуссии относительно разных аспектов сбора, анализа и интерпретации параданных [Nicolaas, 2011; Lynn, Nicolaas, 2010; Eckman, 2013].

В данный момент нет четкого ответа о том, какие данные можно собирать без согласия респондентов. Однако М. Купер предполагает, что вопрос сензитивности таких данных снимается, если они используются в контексте улучшения инструментария без сведения к индивидуальному уровню каждого отдельно взятого респондента [Couper, 2019].

Говоря о таких типах параданных, как местоположение интервью, аудиозапись и путь интервьюера в процессе сбора данных, нельзя не отметить тот факт, что их сбор неразрывно связан с проблемами корректирования собственного поведения со стороны респондентов и со стороны интервьюеров («Хоторнский эффект»). Кроме того, запись такого рода данных может отталкивать от участия в исследовании не только респондентов, но и интервьюеров. При этом дополнительной проблемой будут являться высокие затраты на обработку и интерпретацию подобных данных (особенно с учетом отсутствия устоявшихся стандартов и конвенций в сфере использования параданных) [Nicolaas, 2011; Sharma, 2019]. Несмотря на частое утверждение, что параданные являются бесплатным дополнением к получаемым данным (особенно в случае онлайн- и компьютеризированных опросов) [Olson, Parkhurst, 2013; Sharma, 2019], не стоит упускать из виду ресурсы, затрачиваемые на их обработку и анализ [Nicolaas, 2011; Wagner, Olson, Edgar, 2017; Elevelt et al., 2019]. Это важно в случае данных, собираемых на стороне устройства пользователя, которые имеют сложный формат и требуют дополнительных усилий для их конвертации (движения мыши, направление взгляда, перемещения интервьюера в процессе сбора данных) [Kreuter, 2017b; Heerwegh, 2003].

Работа с параданными, которые используются как индикаторы качества получаемых сведений, нередко требует определения допустимых значений. При этом имеющиеся конвенции могут не нести серьезных теоретических предпосылок. Так, М. Матяшич с коллегами показали, что зачастую, выбирая уровень допустимых значений по длительности заполнения анкеты, исследователи руководствуются общими представлениями (5% самых медленных и самых быстрых, все, кто находится в трех стандартных отклонениях от среднего значения, и т.д.), которые не основаны на проведенных исследованиях, а, скорее, отражают неподкрепленные исследованиями личные представления [Matjašič, Vehovar, Manfreda, 2018]. Обоснованный подход к выделению крайне допустимых значений может базироваться в том числе на типах вопросов, вариантах ответа, а также с учетом общего контекста исследования [Kaczmirek, 2008] на скорости чтения вопросов с учетом количества слов и вариантов ответов [Matjašič, Vehovar, Manfreda, 2018].

Важно понимать, что среди параданных, как и среди показателей качества данных, отсутствует один наиболее широкий и обладающий самой высокой описательной силой относительно поведения интервьюера и респондента тип. Важно пользоваться разными уровнями имеющихся данных, что позволит видеть картину (коммуникацию между интервьюером и респондентом, их поведение, ситуацию интервью в целом) с разных сторон.

### **Перспективные направления исследований**

В этом разделе будут представлены перспективные направления исследований и работы с параданными. Этот раздел отличается от части с описанием имеющихся направлений использования параданных тем, что здесь описываются, скорее, планируемые направления их использования, которые пока не были воплощены или по каким-либо причинам ограничены в использовании в настоящее время.

На данный момент в опросной индустрии параданные активно применяются для контроля качества сбора данных с их комбинацией в определенные мониторинговые системы, которые позволяют наблюдать динамику выбранных параметров (в том числе параданных) в процессе проведения полевого этапа [Mohadjer, Edwards, 2018; Kreuter, Casas-Cordero, 2010; Kirgis, Lepkowski, 2013]. В этом контексте важно разрабатывать комплексный подход и добавлять больше типов параданных, учитывая, что они могут быть полезны в контексте оптимизации процесса сбора данных, корректировки инструментария в процессе исследования, мониторинга интервьюеров и их выбора для проведения дополнительного обучения (например, при работе с планшетом, при коммуникации с респондентами).

Важной перспективой употребления параданных является представление индивидуальных стратегий достижения респондентов (установления связи/контакта) и адаптации опросного инструментария в случае панельных исследований [McClain et al., 2019; Callegaro, 2013]. Данные и параданные из прошлых волн могут быть очень полезными для снижения вероятности выбывания респондента из панели, а также для роста удовлетворенности участием в исследовании.

Параданные могут пригодиться также для сравнения ошибки покрытия между различными методами сбора данных и для оценки сопоставимости различных методов сбора данных при использовании стратегии комбинирования количественных методов сбора данных (mix-mode) [McClain et al., 2019]. Для этого очень полезны данные об установлении контакта/связи с респондентом. В случае онлайн-опроса такими параданными можно считать уведомления об открытии письма с приглашением и факт перехода по ссылке в письме.

С учетом ограничений, связанных с наблюдениями и комментариями интервьюеров, требуется работа по стандартизации и разработке детального инструментария для сбора такого типа параданных [Smith, 2011]. В перспективе обдуманное выделение наиболее важных характеристик поведения респондентов, ситуации интервью, домохозяйства и района помогут снизить нагрузку на интервьюеров, при этом повысив ценность получаемых данных.

Ученые отмечают недостаток исследований, связанных с использованием параданных для снижения ошибки измерения и корректировки на подобные смещения в данных [Yan, Olson, 2013].



Важным направлением исследований, связанных с применением параданных, является анализ того, каким образом отличаются поведение респондентов, интервьюеров и ситуация интервью в целом в случаях, когда параданные собираются, и в случаях, когда их сбор не происходит («Хоторнский эффект») [West, 2011].

Качество параданных до сих пор значительно варьируется в зависимости от используемых методов сбора данных, контекста исследования, характеристик интервьюеров и устройств для работы респондентов. Соответственно, требуется изучить оценку качества параданных и влияющих факторов [West, Sinibaldi, 2013].

Важно также работать с самими параданными, улучшать процесс их сбора, выделять новые релевантные типы и разрабатывать более эффективный инструментарий анализа и интерпретации параданных (особенно в контексте обучения супервайзеров и руководителей исследований работе с параданными) [Kreuter, Casas-Cordero, 2010; Nicolaas, 2011; Kreuter, Couper, Lyberg, 2010]. В том числе сюда можно отнести идеи использования более сложных методов анализа данных (например, Байесовские методы анализа данных) [Sharma, 2019]. Перспективным направлением можно считать работу с параданными процесса кодирования или вбивки информации. Потенциально такие параданные могут позволить оптимизировать процесс и иметь возможность его оценивать (например, выделять более надежных кодировщиков) [Lynn, Nicolaas, 2010].

## **Сбор параданных**

Важным моментом в процессе работы с параданными является их сбор. Выделяют несколько возможных путей получения сведений такого рода: создание своего либо использование имеющегося программного обеспечения, применение функционала опросных программ/специализированных онлайн-платформ для сбора данных по интернету [Stieger, Reips, 2010].

В 2012 году В. Веховар с коллегами в ходе масштабного исследования проверили, какие параданные записываются различными видами программного обеспечения, занимающимися сбором данных онлайн. Из 143 представивших демоверсии компаний (всего было приглашено 365) 30 % подавали информацию по общему времени заполнения опроса, 4 % представляли информацию относительно длительности заполнения отдельных страниц и только 1 % позволяли измерять движения по опросу (возврат к предыдущим вопросам) [Callegaro, Manfreda, Vehovar, 2015; Vehovar et al., 2012: 63].

На данный момент есть основания полагать: рост интереса к параданным и перспективам их использования привел к тому, что большая доля опросных платформ дает возможность замера как минимум основных типов параданных (время прохождения всего опроса, дата и время начала, время заполнения отдельных страниц). Тем не менее, для более глубоких типов данных, описывающих поведение во время заполнения опроса (движения мыши, нажатия клавиш, прокручивание страницы), все еще требуется дополнительное программное обеспечение.

Цель этого раздела — описать, какие существуют программные решения для разных типов параданных. Важно заметить, что указанный тут список не является полным и направлен, скорее, на создание представления о том, какие есть про-

граммные решения для сбора разных типов параданных и какие ресурсы внедрения программного обеспечения несет использование тех или иных типов параданных.

Одним из первых видов подобного программного обеспечения является компонент JavaScript — Client-side paradata (CSP) [Heerwegh, 2003]<sup>4</sup>. Этот компонент, по сути, являющийся программным скриптом, может дополнить любое онлайн-исследование и собирать следующие данные (для всех действий указывается также время их совершения): открытие гиперссылок, взаимодействие с элементами ввода ответов/различными типами вопросов и завершение опроса.

Недавно Й. Хёне и С. Шлоссер представили дополненную версию кода CSP на базе JavaScript — SurveyFocus/Embedded client-side paradata (ECSP)<sup>5</sup> [Höhne, Schlosser, 2018; Höhne, Schlosser, Krebs, 2017; Schlosser, 2018; McClain et al., 2019]. Она позволяет записывать такие параданные, как количество миллисекунд, прошедших с загрузки страницы до перехода к следующей, нажатия и движения мыши, прокручивание страницы, нажатие на кнопки клавиатуры, размер экрана в пикселях, масштабирование страницы с опросом, активность страницы веб-браузера с опросом, а также определять время, на которое респондент покидал окно с опросом [Höhne, Schlosser, Krebs, 2017].

Другим примером является программа PageFocus<sup>6</sup> на JavaScript, которая указывает количество, время и длительность переключения со страницы с опросом на другую, а также время возвращения [Diedenhofen, Musch, 2017].

Имеются возможности сбора параданных и с помощью дополнительных плагинов к существующему программному обеспечению. Например, к платформе OpenSesame<sup>7</sup> [Mathôt, Schreij, Theeuwes, 2012] — универсальному графическому конструктору экспериментов. Дополнительный плагин Mousetrap позволяет замерять различные характеристики использования мыши респондентом (движения, нажатия, в том числе относительно элементов на странице) и вводимые значения [Horwitz et al., 2017]. Важно заметить, что в этом случае разработанность основного программного обеспечения снимает сложности работы с такого рода данными (это может быть сделано без наличия развитых умений программирования). Кроме того, возможен прямой экспорт полученных данных в язык программирования R (с использованием пакета mousetrap), что также упрощает работу с этим типом данных [Kieslich, Henninger, 2017]<sup>8</sup>.

Существует также программное обеспечение Universal client-side paradata (UCSP) [Kaczmirek, 2008]<sup>9</sup>, которое позволяет наблюдать все нажатия мыши (вне зависимости от того, на какой элемент опроса кликнул респондент), записывать их

<sup>4</sup> Детальное описание см.: [Heerwegh, 2003].

<sup>5</sup> Более детальную информацию о ECSP, а также сам код и способ его внедрения в онлайн-опрос см.: [Schlosser, 2016].

<sup>6</sup> Доступ к коду и другой документации возможен на сайте: URL: <https://github.com/deboerk/PageFocus/> (дата обращения: 03.04.2020).

<sup>7</sup> Доступ для скачивания имеется на сайте: URL: <http://osdoc.cogsci.nl/> (дата обращения: 03.04.2020).

<sup>8</sup> Mousetrap является бесплатным программным обеспечением и доступно для скачивания на сайте: URL: <https://github.com/pascalkieslich/mousetrap-os> (дата обращения: 03.04.2020). Пример использования программного обеспечения mousetrap приведен на сайте: URL: <https://github.com/pascalkieslich/mousetrap-os#examples> (дата обращения: 03.04.2020).

<sup>9</sup> Доступ и примеры получаемых параданных можно найти на сайте Л. Качмирека: URL: <http://www.kaczmirek.de/ucsp/ucsp.html> (дата обращения: 03.04.2020). Также детальное описание см. в работе: [Kaczmirek, 2008].

количество, факт изменений выбранного варианта ответа и количество нажатий вне активных полей. Преимуществом UCSF является то, что для его использования необходимо лишь раз добавить в исследование код, который впоследствии будет замерять все нажатия мыши.

Й. К. Хёне с коллегами использовали инструмент SurveyMotion для замера передвижений и изменения положения смартфона в процессе прохождения онлайн-опроса респондентами. Для этого применялись данные акселерометров, встроенных в смартфоны [Höhne, Revilla, Schlosser, 2019]. В последствии эти данные можно использовать как показатель вовлеченности респондентов в процесс заполнения опроса (опрос на ходу может приводить к снижению качества получаемых данных в отличие от заполнения без движений). Сам инструмент написан на языке Java-Script и доступен в статье [Ibid.: 15].

Более комплексный подход наблюдается в случае программного обеспечения User-Action Tracer (UAT). Этот код позволяет сохранять информацию относительно использованного браузера, операционной системы, размера и масштаба монитора, времени всех действий в процессе заполнения опроса, все нажатия мыши (в том числе двойные), выборы вариантов ответа в различных типах вопросов, завершение опроса, вводимый текст и нажатые клавиши, а также позицию, в которой находился курсор в каждые полсекунды прохождения опроса [Stieger, Reips, 2010; Olson, Parkhurst, 2013]<sup>10</sup>. Тем не менее, важно заметить, что получаемый в результате текст с записанными параданными требует последующей конвертации (в итоге были проанализированы 336 262 действия всех респондентов). С. Стигер и У. Райпс использовали для этого отдельный скрипт Perl [Stieger, Reips, 2010].

Работа с поведенческим кодированием аудиозаписей интервью (определение изменения напряженности голоса, заинтересованности, комментариев и вопросов со стороны респондентов) возможна с помощью программы Sequence Viewer [Olson, Parkhurst, 2013]<sup>11</sup>.

Для сбора наблюдений интервьюеров можно использовать инструментарий PEDAksi, представленный П. Линном в 2003 году. Анкета содержит вопросы для оценки интервьюерами различных характеристик домохозяйства (тип жилища, наличие домофона, тип входа в дом, количество этажей в здании, наличие охраны) и района (состояние зданий вокруг, регион, плотность населения, доля различных социальных групп в районе), где проживает респондент, отказавшийся от участия в исследовании [Lynn, 2003; Ипатова, 2016; Сидоров, 2011].

Дж. Синибальди с коллегами также использовали отдельный инструментарий, в котором просили интервьюеров оценивать тип жилища респондентов, наличие работающих взрослых, этнический состав домохозяйства респондента, предполагаемое количество детей в возрасте до 16 лет в домохозяйстве и их наличие в целом. Причем такие вопросы применялись для дополнения данных, получаемых от респондентов на случай их отказа от ответа на подобные прямые вопросы [Sinibaldi, Durrant, Kreuter, 2013].

<sup>10</sup> Доступ к коду возможен по личному обращению к Стигеру, адрес электронной почты указан в источнике [Stieger, Reips, 2010: 1].

<sup>11</sup> Скачивание программы доступно на сайте: URL: <http://www.sequenceviewer.nl/> (дата обращения: 03.04.2020).

В Российском мониторинге экономического положения и здоровья населения НИУ ВШЭ (РМЭЗ НИУ ВШЭ) в конце каждого типа вопросников (детский, взрослый, опрос домохозяйства) интервьюерам необходимо заполнить короткий список вопросов относительно нахождения третьих лиц (членов семьи или посторонних людей) рядом при проведении интервью, отношения респондента к интервью, уровень понимания вопросов респондентом, его поведения, сообразительности, искренности, открытости, а также надежности полученной информации<sup>12</sup>.

Кроме того, стоит привести примеры программного обеспечения для СAPI, которые позволяют собирать параданные

Blaise сохраняет значительное количество видов параданных — длительность опроса, длительность различных сессий заполнения опроса/опросов, строку user agent string и др. [Wood, Piskorowski, Williams, 2016; Olson, Parkhurst, 2013].

Важным преимуществом программного обеспечения Survey Solutions Всемирного банка для проведения СAPI-исследований является, помимо записи базовых параданных (время, длительность), запись местоположения интервью (в том числе есть возможность замерять GPS-локации в начале и в конце интервью)<sup>13</sup>.

## Заключение

Параданные, несомненно, представляют большую ценность и открывают значимые перспективы для методологических исследований и практики опросных компаний. Особенно это актуально в современном контексте повышения доли исследований, проводимых онлайн и/или с использованием компьютерных технологий.

Целью этого обзора было показать, что такое параданные, каковы их типы, способы их сбора, возможности применения, ограничения и перспективные направления применения.

Параданные описывают поведенческие характеристики респондентов и интервьюеров в процессе интервью/прохождения опроса, ситуацию интервью в целом и особенности респондента, домохозяйства или района, в котором проходило интервью, отмечаемые интервьюером. Отличительными чертами параданных являются: недоступность перед проведением полевого этапа, но возникновение и изменения в процессе сбора данных (исключение — параданные из предыдущих исследований/волн исследований); фиксация на каждом этапе сбора данных; нереактивность; наличие различных уровней детализации и генерализации.

Было выделено пять типов параданных: 1) наблюдения интервьюеров; 2) информация об устройстве; 3) поведенческие особенности, касающиеся заполнения опроса; 4) характеристики интервьюеров и установления контакта/связи с респондентом; 5) вербальные и физиологические свойства, проявляющиеся при заполнении опроса/участии в интервью.

Далее показано, насколько широкий набор исследовательских задач связан с потенциальным использованием параданных. Затем выделены ограничения

<sup>12</sup> Вопросники // РМЭЗ НИУ-ВШЭ. URL: <https://www.hse.ru/rims/question> (дата обращения: 10.10.2019).

<sup>13</sup> См.: Survey Solutions. URL: [https://support.mysurvey.solutions/headquarters/export/paradata\\_file\\_format/](https://support.mysurvey.solutions/headquarters/export/paradata_file_format/) (accessed: 10.10.2019).

работы с таким типом данных, описаны перспективные направления развития исследований и способов использования параданных. Приведены инструментальные примеры решений сбора параданных различных типов.

В соответствии с выделенными ограничениями и перспективами использования параданных разработаны следующие практические рекомендации работы с таким типом данных.

- Не бояться использовать параданные — их анализ открывает широкий набор перспектив по улучшению дизайна исследований, инструментария, корректировки данных и т. д. При этом даже для не самых удобных с точки зрения формата выгрузки типов параданных имеются решения (программное обеспечение, пакеты в R, Python — языки программирования), позволяющие работать с этими типами данных исследователям, у которых нет умения и/или опыта работы с подобными форматами данных.
- Применять запись параданных на стороне пользователя/устройства, которая позволяет иметь более широкий выбор типов параданных.
- Заранее продумывать стратегию работы с параданными (какие типы будут собираться, индикаторами чего они будут выступать и какое дополнительное программное обеспечение потребуется для их сбора).
- Аккуратно и внимательно работать с параданными, особенно в контексте их анализа и интерпретации, а также выбора допустимых значений, которые могут повлиять на последующие выводы.
- Предоставлять интервьюерам возможность записывать комментарии и наблюдения относительно процесса сбора данных. Такие сведения могут быть очень ценными как для качества инструментария и процесса сбора данных [Lynn, 2003; Maitland, Casas-Cordero, Kreuter, 2009], так и для вовлеченности и заинтересованности интервьюеров [Ипатова, 2016].
- Не перегружать интервьюеров дополнительным инструментарием, который требуется заполнить, чтобы не повышать когнитивную нагрузку. Делать процесс представления комментариев и оценок респондента/домохозяйства/района наиболее интуитивно понятным и когнитивно не затратным [Kreuter, Couper, Lyberg, 2010].
- Заранее продумывать способы унифицированного сбора параданных одного типа при использовании разных методов сбора данных или при их комбинировании (mix-mode).
- Применять возможности работы с параданными во время сбора данных (например, просить исследователей быть более внимательными к ответам респондентов, заполняющих анкету слишком быстро [Conrad et al., 2017]).
- Собирать и использовать разные типы и уровни параданных, чтобы иметь возможность описывать различные аспекты и этапы процесса сбора данных [Sharma, 2019].
- При предоставлении общего доступа к результатам исследований добавлять в базу данных такие параданные, которые не могут быть сведены к респондентам и раскрывать личную информацию (местоположение, аудиозаписи). Это способствует более активному развитию сферы работы с параданными, в том числе с помощью возможностей метаанализа.

## Список литературы (References)

Девятко И. Ф. Разработка подхода к количественной мультимодальной оценке когнитивной нагрузки интервьюеров: результаты пилотного квазиэксперимента // Вестник Российского университета дружбы народов. Серия: Социология. 2018. Т. 18. № 4. С. 627—637.

Devyatko I. F. (2018) Development of an Approach to Quantitative Multimodal Assessment of the Cognitive Load of Interviewers: Results of a Pilot Quasi-Experiment. *Bulletin of the RUDN, Russia. Series: Sociology*. Vol. 18. No. 4. P. 627—637. (In Russ.)

Девятко И. Ф., Лебедев Д. В. Глазами интервьюера, глазами респондента: контуры нового подхода к оценке когнитивной нагрузки при проведении опроса // Мониторинг общественного мнения: экономические и социальные перемены. 2017. № 5. С. 1—19. <https://doi.org/10.14515/monitoring.2017.5.01>.

Deviatko I. F., Lebedev D. V. (2017) Through the Eyes of the Interviewer, through the Eyes of the Respondent: Outlining a New Approach Towards the Assessment of Cognitive Load During the Interview. *Monitoring of Public Opinion: Economic and Social Changes*. No. 5. P. 1—19. (In Russ.) <https://doi.org/10.14515/monitoring.2017.5.01>.

Звоновский В. Б., Соловьёва Ю. В. Адаптация системы конечных кодов AAPOR к российской практике CATI-опросов // Мониторинг общественного мнения: экономические и социальные перемены. 2017. № 3. С. 59—72. <https://doi.org/10.14515/monitoring.2017.3.06>.

Zvonovski V. B., Solovieva Yu. V. (2017) Adaptation of the AAPOR Final Code System for CATI Technique in Russia. *Monitoring of Public Opinion: Economic and Social Changes*. No. 3. P. 59—72. (In Russ.) <https://doi.org/10.14515/monitoring.2017.3.06>.

Ипатова А. А. Использование параданных в анализе телефонных опросов // Телескоп: журнал социологических и маркетинговых исследований. 2014. № 6. С. 34—41.

Ipatova A. A. (2014) The Use of Paradata in the Analysis of Telephone Surveys. *Telescope: a Journal of Sociological and Marketing Research*. No. 6. P. 34—41. (In Russ.)

Ипатова А. А. Опыт анализа комментариев интервьюеров к телефонным интервью // Мониторинг общественного мнения: экономические и социальные перемены. 2016. № 5. С. 64—76. <http://dx.doi.org/10.14515/monitoring.2016.5.05>.

Ipatova A. A. (2016) The Experience in the Analysis of the Interviewers Comments to Telephone Interviews. *Monitoring of Public Opinion: Economic and Social Changes*. No. 5. P. 64—76. (In Russ.) <http://dx.doi.org/10.14515/monitoring.2016.5.05>.

Ипатова А. А., Rogozin D. M. Условия коммуникативного успеха в стандартизованном телефонном интервью // Социологический журнал. 2014. № 1. P. 21—53. <https://doi.org/10.19181/socjour.2014.1.480>.

Ipatova A. A., Rogozin D. M. (2014) A formula to communicative success in the structured telephone interview. *Sociological Journal*. No. 1. P. 21—53. <https://doi.org/10.19181/socjour.2014.1.480>. (In Russ.)

Караева О. С. «Поиски лучшей репрезентации»: сравнение телефонных и квартирных опросных методик в социально-политических исследованиях (методический эксперимент) // Вестник общественного мнения. Данные. Анализ. Дискуссии. 2015. № 1 (119). С. 116—128.

Karaeva O. S. (2015) «Searches for a Better Representation»: a Comparison of Telephone and Apartment Survey Techniques in Socio-Political Research (Methodological Experiment). *Bulletin of Public Opinion. Data. Analysis. Discussions*. No. 1 (119). P. 116—128. (In Russ.)

Корытникова Н. В. Параметры проверки и контроля качества онлайн-опроса с использованием параданных // Мониторинг общественного мнения: экономические и социальные перемены. 2018. № 3. С. 65—77. <https://doi.org/10.14515/monitoring.2018.3.04>.

Korytnikova N. V. (2018) Use of Paradata in Quality Control of Online-Survey. *Monitoring of Public Opinion: Economic and Social Changes*. No. 3. P. 65—77. (In Russ.) <https://doi.org/10.14515/monitoring.2018.3.04>.

Лебедев Д. В., Богданов М. Б. Переход с PAPI на CAPI: опыт интервьюеров и характеристики, влияющие на их ожидания // Мониторинг общественного мнения: экономические и социальные перемены. 2019. № 4. С. 43—67. <https://doi.org/10.14515/monitoring.2019.4.03>.

Lebedev D. V., Bogdanov M. B. (2019) Transition from PAPI to CAPI: Interviewers' Experience and Factors Influencing Their Expectations. *Monitoring of Public Opinion: Economic and Social Changes*. No. 4. P. 43—67. (In Russ.) <https://doi.org/10.14515/monitoring.2019.4.03>.

Мавлетова А. М. Использование параданных в опросах для корректировки и оптимизации полевых работ в адаптивном дизайне // Мониторинг общественного мнения: экономические и социальные перемены. 2017. № 5. С. 105—119. <https://doi.org/10.14515/monitoring.2017.5.07>.

Mavletova A. M. (2017) Using Survey Paradata for the Fieldwork Corrections and Optimization in Adaptive Design. *Monitoring of Public Opinion: Economic and Social Changes*. No. 5. P. 105—119. <https://doi.org/10.14515/monitoring.2017.5.07>. (In Russ.)

Малошонок Н. Г., Терентьев Е. А. Влияние дизайна анкеты на качество данных в онлайн-опросах студентов // Мониторинг общественного мнения: экономические и социальные перемены. 2014. № 6. С. 15—26. <https://doi.org/10.14515/monitoring.2014.6.02>.

Maloshonok N. G., Terentyev E. A. (2014) Effects of questionnaire design on data quality in student online surveys. *Monitoring of Public Opinion: Economic and Social Changes*. No. 6. P. 15—26. (In Russ.) <https://doi.org/10.14515/monitoring.2014.6.02>.

Никитина С. Ю. О подготовке и проведении пробной переписи населения 2018 года // Вопросы статистики. 2018. Т. 25. № 4. С. 22—26.

Nikitina S. Yu. (2018) On the Preparation and Conduct of a Trial Population Census in 2018. *Questions of Statistics*. Vol. 25. No. 4. P. 22—26. (In Russ.)

Рогозин Д. М., Ипатова А. А. Контроль качества социальных обследований. М., 2015. URL: <ftp://w82.ranepa.ru/rnp/ppaper/020915.pdf> (дата обращения: 13.08.2017).  
Rogozin D. M., Ipatova A. A. (2015) Control Over the Quality of Social Surveys. Moscow. URL: <ftp://w82.ranepa.ru/rnp/ppaper/020915.pdf> (accessed: 13.08.2017). (In Russ.)

Рогозин Д. М., Сапонов Д. И. Корпус ошибок автоматизированного телефонного опроса // *Полития*. 2014. № 4 (75). С. 145—162.

Rogozin D. M., Saponov D. I. (2014) Corps of Errors in an Automated Telephone Survey. *Politeia*. No. 4 (75). P. 145—162. (In Russ.)

Сидоров Н. Использование параданных в социологических исследованиях // *Социология: теория, методы, маркетинг*. 2011. № 4. С. 198—208.

Sidorov N. (2011) The Use of Parades in Sociological Research. *Sociology: Theory, Methods, Marketing*. No. 4. P. 198—208. (In Russ.)

Терентьев Е. А., Мавлетова А. М., Косолапов А. М. Интервьюирование с помощью компьютерных технологий в лонгитюдных обследованиях домохозяйств // *Мониторинг общественного мнения: экономические и социальные перемены*. 2018. № 3 (145). С. 47—64. <https://doi.org/10.14515/monitoring.2018.3.03>.

Terentyev E. A., Mavletova A. M., Kosolapov A. M. (2018) Computer-assisted personal interviewing for longitudinal household studies. *Public Opinion Monitoring: Economic and Social Changes*. No. 3 (145). P. 47—64. (In Russ.) <https://doi.org/10.14515/monitoring.2018.3.03>.

Biemer P. P., Chen P., Wang K. (2013) Using Level-of-Effort Paradata in Non-Response Adjustments with Application to Field Surveys. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. Vol. 176. No. 1. P. 147—168. <https://doi.org/10.1111/j.1467-985x.2012.01058.x>.

Bosnjak M. (2001) Participation in Non-Restricted Web-Surveys: a Typology and Explanatory Model for Item-Nonresponse. Lengerich: Pabst Science Publishers.

Bosnjak M., Tuten T. L. (2001) Classifying Response Behaviors in Web-Based Surveys. *Journal of Computer-Mediated Communication*. Vol. 6. Iss. 3. P. JCMC636. <https://doi.org/10.1111/j.1083-6101.2001.tb00124.x>.

Callegaro M. (2013) Paradata in Web Surveys. In: Kreuter F. (ed.) *Improving Surveys with Paradata: Analytic Uses of Process Information*. Hoboken, New Jersey: John Wiley & Sons. P. 261—279. <https://doi.org/10.1002/9781118596869.ch11>.

Callegaro M. (2015) Using Paradata to Better Interpret Online Experiments and Non-Experiments. *Centre for Experimental Social Sciences*. 24 May. URL: <https://cess-nuffield.nuff.ox.ac.uk/> (accessed: 10.10.2019).

Callegaro M. (2014) Using Paradata to Improve the Quality of Web Surveys: Some Examples and Applications. University of Southampton. URL: [https://www.southampton.ac.uk/assets/sharepoint/groupsite/Administration/SitePublisher-document-store/Documents/Callegaro\\_Paradata\\_Web\\_Survey\\_RSS1.pdf](https://www.southampton.ac.uk/assets/sharepoint/groupsite/Administration/SitePublisher-document-store/Documents/Callegaro_Paradata_Web_Survey_RSS1.pdf) (accessed: 10.10.2019).



Callegaro M., Manfreda K. L., Vehovar V. (2015) *Web Survey Methodology*. London: Sage. P. 318.

Choumert-Nkolo J., Cust H., Taylor C. (2019) Using Paradata to Collect Better Survey Data: Evidence from a Household Survey in Tanzania. *Review of Development Economics*. Vol. 23. No. 2. P. 598—618. <https://doi.org/10.1111/rode.12583>.

Conrad F., Tourangeau R., Couper M., Zhang C. (2017) Reducing Speeding in Web Surveys by Providing Immediate Feedback. *Survey Research Methods*. Vol. 11. No. 1. P. 45—61.

Couper M. P. (1998) A Measuring Survey Quality in a CASIC Environment. *Proceedings of the Survey Research Methods Section, ASA. Achieving Quality in Surveys*. Aug. P. 41—49. URL: <https://www.amstat.org/ASA/Membership/Sections-and-Interest-Groups.aspx>.

Couper M. (2019) Practical Challenges in Web Survey Methodology. *Web Survey Day. University of Ljubljana*. 25 Sept.

Couper M. P., Kreuter F. (2013) Using Paradata to Explore Item Level Response Times in Surveys. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. Vol. 176. No. 1. P. 271—286. <https://doi.org/10.1111/j.1467-985x.2012.01041.x>.

Couper M. P., Singer E. (2013) Informed Consent for Web Paradata Use. *Survey Research Methods*. Vol. 7. No. 1. P. 57.

Dajani N., Marquette R. Q. (2015) Reinterview Detection and Prevention at Census: New Initiatives. URL: [http://washstat.org/presentations/20150609/dajani\\_marquette-revised.pdf](http://washstat.org/presentations/20150609/dajani_marquette-revised.pdf) (accessed: 9.04.2019).

Diedenhofen B., Musch J. (2017) PageFocus: Using Paradata to Detect and Prevent Cheating on Online Achievement Tests. *Behavior Research Methods*. Vol. 49. No. 4. P. 1444—1459. <https://doi.org/10.3758/s13428-016-0800-7>.

Durrant G. B., D'Arrigo J., Steele F. (2011) Using Paradata to Predict Best Times of Contact, Conditioning on Household and Interviewer Influences. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. Vol. 174. No. 4. P. 1029—1049. <https://doi.org/10.1111/j.1467-985x.2011.00715.x>.

Durrant G., Kreuter F. (2013) Editorial: the Use of Paradata in Social Survey Research. *Journal of the Royal Statistical Society*. Vol. 176. P. 1—3. <https://doi.org/10.1111/j.1467-985x.2012.01082.x>.

Eckman S. (2013) Paradata for Coverage Research. In: Kreuter (Hg.). *Kontaktformular*. URL: [http://www.europeansocialsurvey.org/docs/round6/fieldwork/germany/ESS6\\_contact\\_form\\_DE.pdf](http://www.europeansocialsurvey.org/docs/round6/fieldwork/germany/ESS6_contact_form_DE.pdf) (accessed: 3.07.2014). P. 97—116.

Elevelt A., Bernasco W., Lugtig P., Ruiter S., Toepoel V. (2019) Where You at? Using GPS Locations in an Electronic Time Use Diary Study to Derive Functional Locations. *Social Science Computer Review*. Sept. 26. P. 1—18. <https://doi.org/10.1177/0894439319877872>.

- Fazio R. H. (1990) *A Practical Guide to the Use of Response Latency in Social Psychological Research*. London: SAGE Publications.
- Finnamore J. (2019) The Role of Survey Paradata in a Federal Statistical Agency's Commitment to Quality. *FedCASIC WorkshoP*. Apr. 16.
- Galesic M., Tourangeau R., Couper M., Conrad F. (2008) Eye-Tracking Data: New Insights on Response Order Effects and Other Cognitive Shortcuts in Survey Responding. *Public Opinion Quarterly*. Vol. 72. No. 5. P. 892—913. <https://doi.org/10.1093/poq/nfn059>.
- Global Guideline: Online Research ESOMAR. URL: <https://www.esomar.org/what-we-do/code-guidelines/esomargrbn-online-research-guideline> (accessed: 1.04.2020).
- Groves R. M., Wagner J., Peytcheva E. (2007) Use of Interviewer Judgments about Attributes of Selected Respondents in Post-Survey Adjustment for Unit Nonresponse: an Illustration with the National Survey of Family Growth. *Proceedings of the Survey Research Methods Section, ASA*. P. 3428—3431.
- Heerwegh D. (2003) Explaining Response Latencies and Changing Answers Using Client-Side Paradata from a Web Survey. *Social Science Computer Review*. Vol. 21. No. 3. P. 360—373. <https://doi.org/10.1177/0894439303253985>.
- Höhne J. K., Revilla M., Schlosser S. (2019) Motion Instructions in Surveys: Compliance, Acceleration, and Response Quality. *International Journal of Market Research*. June 27. P. 1—15. <https://doi.org/10.1177/1470785319858587>.
- Höhne J. K., Schlosser S. (2018) Investigating the Adequacy of Response Time Outlier Definitions in Computer-Based Web Surveys Using Paradata SurveyFocus. *Social Science Computer Review*. Vol. 36. No. 3. P. 369—378. <https://doi.org/10.1177/0894439317710450>.
- Höhne J. K., Schlosser S., Krebs D. (2017) Investigating Cognitive Effort and Response Quality of Question Formats in Web Surveys Using Paradata. *Field Methods*. Vol. 29. No. 4. P. 365—382. <https://doi.org/10.1177/1525822x17710640>.
- Horwitz R., Henninger F., Brockhaus S., Kieslich P. J. (2017) Learning from Mouse Movements: Improving Questionnaire and Respondents' User Experience through Passive Data Collection. *IAB-Discussion Paper*. No. 34. P. 1—27.
- Jans M. et al. (2019) Analyzing and Using Web Paradata to Understand and Improve the Data Collection Process. *FedCASIC WorkshoP*. Apr. 16.
- Kaczmarek L. (2008) *Human-Survey Interaction: Usability and Nonresponse in Online Surveys*: Doctoral Dissertation. Mannheim: Universität Mannheim.
- Kieslich P. J., Henninger F. (2017) Mousetrap: an Integrated, Open-Source Mouse-Tracking Package. *Behavior Research Methods*. Vol. 49. No. 5. P. 1652—1667. <https://doi.org/10.3758/s13428-017-0900-z>.
- Kirgis N. G., Lepkowski J. M. (2013) Design and Management Strategies for Paradata-Driven Responsive Design: Illustrations from the 2006—2010 National Survey of Family Growth. In: Kreuter F. (ed.) *Improving Surveys with Paradata: Analytic Uses of*

- Process Information*. Hoboken, New Jersey: John Wiley & Sons. P. 121—144. <https://doi.org/10.1002/9781118596869.ch6>.
- Kozyreva P., Kosolapov M., Popkin B. M. (2016) Data Resource Profile: the Russia Longitudinal Monitoring Survey — Higher School of Economics (RLMS-HSE) Phase II: Monitoring the Economic and Health Situation in Russia, 1994—2013. *International Journal of Epidemiology*. Vol. 45. No. 2. P. 395—401. <https://doi.org/10.1093/ije/dyv357>.
- Kreuter F. (ed.). (2013) *Improving Surveys with Paradata: Analytic Uses of Process Information*. Hoboken, New Jersey: John Wiley & Sons. Vol. 581. <https://doi.org/10.1002/9781118596869>.
- Kreuter F. (2013) *Improving Surveys with Paradata: Introduction*. In: Kreuter F. (ed.) *Improving Surveys with Paradata: Analytic Uses of Process Information*. Hoboken, New Jersey: John Wiley & Sons. P. 1—9. <https://doi.org/10.1002/9781118596869.ch1>.
- Kreuter F. (2017a) Getting the Most Out of Paradata. In: Vannette D., Krosnick J. (eds.) *The Palgrave Handbook of Survey Research*. Cham: Palgrave Macmillan. P. 193—198. [https://doi.org/10.1007/978-3-319-54395-6\\_24](https://doi.org/10.1007/978-3-319-54395-6_24).
- Kreuter F. (2017b) Paradata. In: Vannette D., Krosnick J. (eds.) *The Palgrave Handbook of Survey Research*. Springer. P. 529—535. [https://doi.org/10.1007/978-3-319-54395-6\\_61](https://doi.org/10.1007/978-3-319-54395-6_61).
- Kreuter F., Casas-Cordero C. (2010) Paradata. Working Paper Series des Rates für Sozial-und Wirtschaftsdaten. No. 136. P. 1—16.
- Kreuter F., Couper M., Lyberg L. (2010) The Use of Paradata to Monitor and Manage Survey Data Collection. In: *Proceedings of the Joint Statistical Meetings, American Statistical Association*. Alexandria: American Statistical Association. P. 282—296.
- Kurakin D. (2014) Russian Longitudinal Panel Study of Educational and Occupational Trajectories: Building Culturally-Sensitive Research Framework (Working Papers of Institute of Education). Moscow: HSE Publishing House.
- Laflamme F., Maydan M., Miller A. (2008) Using Paradata to Actively Manage Data Collection Survey Process. In: *Proceedings of the Survey Research Methods Section, American Statistical Association*. Alexandria: American Statistical Association. P. 630—637.
- Lynn P. (2003) PEDAKSI: Methodology for Collecting Data about Survey Non-Respondents. *Quality and Quantity*. Vol. 37. No. 3. P. 239—261.
- Lynn P., Nicolaas G. (2010) Making Good Use of Survey Paradata. *Survey Practice*. Vol. 3. No. 2. P. 1—5. <https://doi.org/10.29115/sp-2010-0010>.
- Maitland A., Casas-Cordero C., Kreuter F. (2009) An Evaluation of Nonresponse Bias Using Paradata from a Health Survey. In: *Proceedings of the Section on Survey Research Methods of the American Statistical Association*. Alexandria: American Statistical Association. P. 370—378.

- Mathôt S., Schreij D., Theeuwes J. (2012) OpenSesame: an Open-Source, Graphical Experiment Builder for the Social Sciences. *Behavior Research Methods*. Vol. 44. No. 2. P. 314—324. <https://doi.org/10.3758/s13428-011-0168-7>.
- Matjašič M., Vehovar V., Manfreda K. (2018) Web Survey Paradata on Response Time Outliers: a Systematic Literature Review. *Metodološki Zvezki*. Vol. 15. No. 1. P. 23—41.
- Mavletova A., Couper M. P., Lebedev D. (2018) Grid and Item-by-Item Formats in PC and Mobile Web Surveys. *Social Science Computer Review*. Vol. 36. No. 6. P. 647—668. <https://doi.org/10.1177/0894439317735307>.
- McClain C.A., Couper M. P., Hupp A. L., Keusch F., Peterson G., Piskorowski A. D., West B. T. (2019) A Typology of Web Survey Paradata for Assessing Total Survey Error. *Social Science Computer Review*. Vol. 37. No. 2. P. 196—213. <https://doi.org/10.1177/0894439318759670>.
- Mohadjer L., Edwards B. (2018) Paradata and Dashboards in PIAAC. *Quality Assurance in Education*. Vol. 26. No. 2. P. 263—277. <https://doi.org/10.1108/qae-06-2017-0031>.
- Nicolaus G. (2011) Survey Paradata: a Review. ESRC National Centre for Research Methods Review Paper. URL: [http://eprints.ncrm.ac.uk/1719/1/Nicolaas\\_review\\_paper\\_jan11.pdf](http://eprints.ncrm.ac.uk/1719/1/Nicolaas_review_paper_jan11.pdf) (accessed: 3.04.2020).
- Olson K., Parkhurst B. (2013) Collecting Paradata for Measurement Error Evaluations. In: Kreuter F. (ed.) *Improving Surveys with Paradata: Analytic Uses of Process Information*. Hoboken, New Jersey: John Wiley & Sons. Vol. 580. P. 43. <https://doi.org/10.1002/9781118596869.ch3>.
- Reeves R., Horwitz R., Misra J., Hernandez-Viver A. (2018) Optimizing a Web Survey Instrument Using Paradata Measures. In: *United States Census Bureau*. URL: <https://www.census.gov/content/dam/Census/newsroom/press-kits/2018/aapor/aapor-presentation-paradata.pdf> (accessed: 11.10.2019).
- Robmann J., Gummer T., Kaczmirek L. (2020) Working with User Agent Strings in Stata: the PARSEUAS command. *Journal of Statistical Software*. Vol. 92. No. 1. P. 1—16. <https://doi.org/10.18637/jss.v092.c01>.
- Sakshaug J. W. (2013) Using Paradata to Study Response to Within-Survey Requests. *Improving Surveys with Paradata*. P. 171—190. <https://doi.org/10.1002/9781118596869.ch8>.
- Schlosser S. (2018) ECSP — Embedded Client Side Paradata. *Zenodo*. P. 1—32.
- Sharma S. (2019) Paradata, Interviewing Quality, and Interviewer Effects: Dissertation. Maryland: University of Maryland, College Park.
- Sinibaldi J., Durrant G. B., Kreuter F. (2013) Evaluating the Measurement Error of Interviewer Observed Paradata. *Public Opinion Quarterly*. Vol. 77. No. S1. P. 173—193. <https://doi.org/10.1093/poq/nfs062>.
- Sirkis R., McGovern P. (2019) Analyzing and Using Web Paradata to Understand and Improve the Data Collection Process. *FedCASIC WorkshoP*. Apr. 16.

Smith T. W. (2011) The Report of the International Workshop on Using Multi-Level Data from Sample Frames, Auxiliary Databases, Paradata and Related Sources to Detect and Adjust for Nonresponse Bias in Surveys. *International Journal of Public Opinion Research*. Vol. 23. No. 3. P. 389—402. <https://doi.org/10.1093/ijpor/edr035>.

Stieger S., Reips U. D. (2010) What are Participants Doing While Filling in an Online Questionnaire: a Paradata Collection Tool and an Empirical Study. *Computers in Human Behavior*. Vol. 26. No. 6. P. 1488—1495. <https://doi.org/10.1016/j.chb.2010.05.013>.

Vehovar V., Čehovin G., Kavčič L., Lenar J. (2012) Survey Software Features Overview. URL: [www.websm.org/uploadi/editor/1362408700WebSM\\_WSS\\_report\\_b\\_v04.pdf](http://www.websm.org/uploadi/editor/1362408700WebSM_WSS_report_b_v04.pdf) (accessed: 10.10.2019).

Wagner J., Olson K., Edgar M. (2017) Assessing Potential Errors in Level-of-Effort Paradata Using GPS Data. *Survey Research Methods*. Vol. 11. No. 3. P. 219—233.

West B. T. (2011) Paradata in Survey Research. *Survey Practice*. Vol. 4. No. 4. P. 1—8.

West B. T., Kreuter F. (2013) Factors Affecting the Accuracy of Interviewer Observations: Evidence from the National Survey of Family Growth. *Public Opinion Quarterly*. Vol. 77. No. 2. P. 522—548. <https://doi.org/10.1093/poq/nft016>.

West B. T., Sinibaldi J. (2013) The Quality of Paradata: a Literature Review. In: Kreuter F. (ed.) *Improving Surveys with Paradata: Analytic Uses of Process Information*. Hoboken, New Jersey: John Wiley & Sons. P. 339—359. <https://doi.org/10.1002/9781118596869.ch14>.

Wood L., Piskorowski A., Williams J. (2016) Transforming Survey Paradata. In: *17th International Blaise Users Group Conference*. Amsterdam.

Yan T., Olson K. (2013) Analyzing Paradata to Investigate Measurement Error. In: Kreuter F. (ed.) *Improving Surveys with Paradata: Analytic Uses of Process Information*. Hoboken, New Jersey: John Wiley & Sons. P. 73—96. <https://doi.org/10.1002/9781118596869.ch4>.