

DOI: 10.19181/socjour.2021.27.3.8422

Т.Ю. ЧЕРКАШИНА^{1,2}

¹ Институт экономики и организации промышленного производства СО РАН, 630090, Новосибирск, проспект Академика Лаврентьева, 17.

² Новосибирский национальный исследовательский государственный университет, 630090, Новосибирск, ул. Пирогова, д. 1, каб. 3338.

СВЕДЕНИЯ О ДОХОДАХ И ИМУЩЕСТВЕ РОССИЙСКИХ РЕГИОНАЛЬНЫХ ПОЛИТИКОВ: ПИЛОТАЖ И ОЦЕНКА КАЧЕСТВА АДМИНИСТРАТИВНЫХ ДАННЫХ

Аннотация. В статье на примере сведений о доходах и имуществе депутатов законодательных органов власти субъектов РФ за 2019 г., собираемых в рамках антикоррупционной деятельности, представлен опыт конвертации нецелевых административных данных в исследовательские. Из всей совокупности сведений этот эмпирический фрагмент был выбран для пилотажа административных данных, включающего оценку возможности интеграции разрозненных фрагментов сведений в единую базу данных, оценку качества данных и их релевантности для решения исследовательских задач, в частности анализа высокодоходных слоев и оценки тенденций к индивидуализации частной собственности. Система индикаторов для оценки качества данных включает их своевременность, доступность, интерпретируемость, достоверность, сравнимость, согласованность, ошибки репрезентации и измерения, релевантность. Для рассматриваемого набора данных ошибки измерения более характерны, чем ошибки репрезентации. В целом статья акцентирует тезисы, что введение в оборот новых нецелевых данных требует их пилотажа, а оценка качества данных становится распределенной как во времени, так и между разными субъектами. Переход от созданных данных к «обнаруженным» смещает функции оценки их качества от исследователя-создателя к исследователю-пользователю. Хотя в рассмотренном случае качество данных отчасти обеспечивается юридическим сопровождением их генерации, преобразование административных данных в исследовательские включает оценку множества измерений качества — от их доступности до унифицированности и точности.

Ключевые слова: административные данные; качество данных; ошибки репрезентации; ошибки измерения; пилотаж административных данных.

Для цитирования: Черкашина Ю.Ю. Сведения о доходах и имуществе российских региональных политиков: пилотаж и оценка качества административных данных // Социологический журнал. 2021. Том 27. № 3. С. 8–34. DOI: 10.19181/socjour.2021.27.3.8422

Социологи все чаще используют данные, которые не были специально созданы и собраны для поиска ответов на исследовательские вопросы. Масштабы обращения к нетрадиционным данным таковы, что ситуацию характеризуют как «взрыв новых форм данных» [32]. Хронологически почти одновременно такими «внешними» для социологов стали «большие данные» и административные данные, возникающие как результат деятельности, связанной с управлением, административными решениями, организационными процессами. Революционность изменений в работе с данными не столько в их размере и количестве, сколько в стремительном распространении новых типов данных [15]. Но если данные, создаваемые в полевых исследованиях, более или менее знакомы и понятны социологам, то вторичный анализ уже существующих данных предполагает оценку их пригодности для решения поставленных задач, а в некоторых случаях — реализацию иной логики: исследовательские вопросы ставятся с учетом содержания и качества имеющихся данных.

В статье представлен опыт оценки качества административных данных, в том числе определения ограничений, которые необходимо учитывать при постановке исследовательских вопросов. Имея многолетний опыт вторичного анализа опросных данных о доходах и собственности российских домохозяйств (Российского мониторинга экономического положения и здоровья населения НИУ «Высшая школа экономики» (RLMS–HSE), Комплексного наблюдения условий жизни населения, Росстат), мы обращаемся к административным данным аналогичного содержания — сведениям о доходе, имуществе и обязательствах имущественного характера государственных служащих, чиновников, депутатов, руководителей организаций, которые ежегодно сообщаются ими в рамках антикоррупционной деятельности.

Нас привлекло уникальное содержание данных. Во-первых, все-российские социально-экономические обследования фиксируют собственность всего домохозяйства, «принадлежащую кому-либо из членов семьи», без детализации, кому именно она принадлежит. Хотя названные административные сведения включают информацию только о собственности и имуществе декларантов и членов их нуклеарных семей (супругов, с которыми они состоят в зарегистрированном браке, и несовершеннолетних детей), а не всего домохозяйства, об объектах собственности сообщается, обладают ли ими члены семьи

на правах индивидуальной, совместной или долевой собственности или только пользуются. Мы имеем уникальную возможность оценить распространенность разных видов собственности на частные активы. Во-вторых, в массовых обследованиях не представлены верхние слои социально-экономической стратификации, и мы предполагаем, что эти административные сведения отчасти компенсируют такой информационный дефицит, хотя ставят перед исследователями сложную задачу идентифицировать слои стратификационной пирамиды, доходной или имущественной, представленные в этих сведениях. В-третьих, это открытые данные, доступные на сайтах органов власти, что отчасти снимает проблему доступа к ним. Неанонимность данных (указаны фамилии декларантов) нам важна в той мере, в какой позволяет дополнить сведения открытыми данными из других источников, например, об образовании, занятости, возрасте и т. д. декларантов, или связать на микроуровне данные за разные годы.

Но эти сведения собраны и открыты отнюдь не с исследовательскими целями. Обратившись к ним, сможем ли мы решить задачи, которые нам интересны? Вся совокупность сведений огромна — их ежегодно с 2014 г. предоставляют сотрудники федеральных министерств и агентств; государственные служащие в федеральных и территориальных органах власти; занятые в органах местного самоуправления; руководители государственных и муниципальных организаций и др. По аналогии с классическими социологическими исследованиями, в которых предполагается пилотажный этап, в ходе которого проверяются адекватность способов и процедур сбора данных, операционализация основных понятий в исследовательских инструментах, при обращении к административным источникам необходим *пилотаж данных*. Для обоснованной оценки реализуемости наших замыслов из всей совокупности был выбран фрагмент — сведения о доходах и имуществе депутатов законодательных органов власти субъектов Федерации за 2019 г. В данном конкретном случае пилотаж данных включает определение возможности конвертировать разрозненные сведения в привычный для статистического анализа файл матричного вида «объект — признак»; оценку качества данных, включая детальный анализ унификации и полноты сведений, собранных в разных субъектах Российской Федерации, и определение потенциала данных для решения намеченных исследовательских задач. Методология оценки качества административных данных подразумевает идентификацию их места среди других типов данных, систематизацию подходов к оценке их качества и адаптацию различных систем оценки к выбранному кейсу.

Методология оценки качества административных данных

Административные данные на «карте» данных. По каким параметрам будет оцениваться качество данных или процесса их генерации, зависит от типа данных. В классификациях данных о населении, с которыми

работают социальные исследователи и статистики, вполне конвенционально выделение первой группы, обозначаемой как созданные, сконструированные, спроектированные (*designed data* [21; 25], “*made*” by researchers [15]). Так как эти данные получают, как правило, в результате проведения опросов, обследований, переписей, то в некоторых классификациях их и обозначают как опросные (*survey*): данные, полученные из ответов индивидов (“*collection of data obtained from responses of individuals*” [14]). Отличительная особенность этого класса данных — они возникают по замыслу и в результате усилий исследователей.

Противоположность им — данные, появление которых обусловлено не исследовательскими, а иными целями и процессами. Так, Р. Гроувз дополняет «сконструированные» данные «естественными» (*organic data*) [21, p. 866–868]; в классификации, предложенной Р. Коннелли с коллегами, созданные данные (*made data*) образуют дихотомию с обнаруженными, «найденными» (*found data*), но то и другое — зонтичные обозначения: первые включают результаты экспериментов и наблюдательных исследований, в том числе опросов и обследований, а вторые — всю совокупность больших данных, в том числе административных [15], хотя последнее не так однозначно. Большой объем административных данных, генерация не для целей исследования, многомерность (включают множество фрагментов, которые при наличии различных идентификаторов могут быть объединены) позволяют, по мнению некоторых исследователей, относить административные данные к большим данным [1; 15; 16]. Но отмечают и различия между ними: административные данные более систематизированы, менее хаотичные и «загрязненные» (*messy*); совокупность, которую они характеризуют, более определена, чем в остальных типах больших данных; концептуальные конструкции, заложенные в эти данные, могут быть более прозрачны. Ожидаемо, что те исследователи, которые считают эти отличия более значимыми, выделяют административные данные в отдельный класс [14; 25].

В большинстве руководств по работе с административными данными они определяются как информация, возникающая при взаимодействии людей с государственными службами, учреждениями¹; получают из форм, разработанных административными органами в соответствии с законами, другими нормативными актами или регламентами различных социальных программ. Так как один из самых распространенных и давно существующих видов административных данных, особенно в североевропейских странах, — регистры насе-

¹ Например, Administrative Data Research UK. — URL: <https://www.adruk.org/our-mission/administrative-data> или OECD. Short-Term Economic Statistics (STES) Administrative Data: Two Frameworks of Papers. — URL: <http://www.oecd.org/sdd/short-termeconomicstatisticsstesadministrativedatatwoframeworksofpapers.htm> (Дата обращения 31.01.2021.)

ния [32, р. 3], возможно, именно это вызывает «прогосударственное» определение административных данных, хотя такую трактовку можно считать слишком узкой, поскольку административные данные должны включать и корпоративные [23].

От общей ошибки исследования к оценке качества данных. Для социологов наиболее привычной методологией оценки качества количественных данных является концепция общей ошибки исследования (Total Survey Error, TSE), которую считают вкладом социальных наук в формирование подходов, инструментов или исследовательской инфраструктуры, обеспечивающей качество данных [24, р. 87]. Общая ошибка исследования раскладывается на составные элементы, относящиеся к выборке (ошибки покрытия, ошибки выборки, ошибки отказа от участия в исследовании) и с ней не связанные, источниками которых могут быть интервьюер, респондент, анкета, режим интервью [20]. Хотя строго определенного перечня ошибок измерения нет, в целом выделяют ошибки спецификации или операционализации, ошибки регистрации, ошибки неответа (отказы отвечать на отдельный вопрос, пропуски вопросов и т. п.), ошибки измерения (респонденты сообщают неверную информацию, интервьюеры не следуют точно процедуре опроса и т. п.) и ошибки обработки (в редактировании и вводе данных, вменении пропущенных значений, ошибки кодирования ответов на открытые вопросы, ошибки взвешивания и т. п.) [8; 11; 27].

Но любой список ошибок в рамках TSE будет неполным, так как новые ошибки возникают постоянно из-за новых технологий сбора данных, методологических новаций [22]. Кроме неполноты оценки большинства общих ошибок исследования из-за невозможности учесть все их источники, слабое место TSE также в том, что не для каждого оцениваемого параметра может быть предложено эталонное или пороговое значение, чтобы, собственно, определить отклонение и «размер» ошибки. И хотя теория общей ошибки исследования привлекает внимание к декомпозиции ошибок, это не единственный способ их оценивать. С распространением новых типов данных статистика и другие дисциплины развивают концепцию качества данных (информации), во-первых, подразумевающую, что оно может быть достигнуто через «качественные» процессы генерации данных, которые, в свою очередь, зависят от качества организационных процессов (и это сближает концепцию с теорией общей ошибки исследования); во-вторых, набор параметров качества расширяется за счет добавления пунктов, связанных с использованием данных, их оценкой с позиции пользователя [22; 29].

Как и в случае с теорией общей ошибки исследования, различные варианты рамок качества данных скорее задают ориентиры в измерениях качества, чем предлагают однообразный набор этих параметров. Г. Брэкстоун, а вслед за ним С. Келлер с коллегами предлагают шесть измерений качества данных, прежде всего статистических: релевант-

ность (relevance), точность (accuracy); своевременность, быстрота предоставления (timeliness); доступность (accessibility), интерпретируемость (interpretability) и согласованность (coherence) [12; 25]. П. Бимер расширяет рамку качества данных, добавив к перечисленным показателям достоверность, убедительность (credibility), сравнимость (comparability), полноту (completeness) [11].

Но даже за одинаковыми обозначениями измерений качества в разных наборах могут быть отличающиеся определения: если, к примеру, для Бимера точность — это сведение к минимуму общей ошибки исследования (в этом плане предложенную им рамку качества можно рассматривать как вариант «примирения» двух концепций оценки качества данных, когда общая ошибка исследования встроена в больший набор их измерений), то в первом наборе точность понимается шире: «это степень, в которой данные правильно описывают явления, для измерения которых они предназначены. Она обычно характеризуется с точки зрения погрешности статистических оценок и традиционно раскладывается на компоненты смещения (систематическая ошибка) и дисперсии (случайная ошибка). Также может быть описана в терминах основных источников ошибок, которые потенциально могут привести к неточности (например, охват, выборка, отсутствие ответа)» [12, р. 141]. Своевременность подразумевает в одном случае «разницу между временем, когда данные были собраны и когда они стали доступны» [25, р. 96], а в другом — предоставление данных согласно графику [11, р. 819].

Согласованность отражает возможность сводить собранные данные с информацией, полученной из других источников в другое время, другими организациями. Основа согласованности — использование стандартных концепций, классификаций, как и общей методологии для всех обследований. Доступность данных для пользователя включает: «...легкость, с которой может быть установлено существование информации, а также пригодность формы или носителя, через которые информация может быть получена. Для некоторых пользователей стоимость информации также может быть одним из аспектов доступности» [12, р. 141].

Интерпретируемость информации отражает наличие дополнительных сведений и метаданных, необходимых для ее трактовки и использования, «легкость, с которой пользователь может понять, правильно использовать и анализировать данные» [30]. В некоторых рамках качества индикатор такого содержания обозначается как «ясность» [17; 35]. Релевантность информации — это ее соответствие реальным потребностям пользователей. Сравнимость (comparability) подразумевает такое структурирование данных, когда возможны демографические, пространственные и временные сравнения. Полнота означает, что данные достаточно содержательны, чтобы удовлетворить цели ана-

лиза. Достоверные, убедительные данные, в интерпретации Бимера, заслуживают доверия со стороны профессионального сообщества «производителей» таких же данных [11, p. 819].

Если смотреть на качество данных с перспективы применениях тех или иных процедур анализа, аппарата статистического моделирования, то форма распределений переменных-признаков также может стать критерием оценки [5].

Очевидно, что эти измерения качества данных не являются независимыми друг от друга: могут конфликтовать между собой точность и своевременность, согласованность и релевантность/актуальность, так как потребности в сохранении методики ради построения исторических трендов могут не совпадать с потребностью модификации методики в меняющихся условиях. При этом вся совокупность измерений оказывает влияние на релевантность [12, p. 140].

Этот краткий обзор трактовки и параметров качества данных обращает внимание как минимум на две проблемы. Первая из них: кто должен давать оценку качества данных? В рамках теории общей ошибки исследования именно «поставщик» данных знает их свойства и процесс генерации настолько, чтобы оценить отдельные ошибки, связанные с отклонением от идеальной выборки, идеального процесса сбора, обработки и подготовки данных к анализу. Но чем больше субъектов вовлечено не только в производство, но и в использование данных, чем сложнее и многомернее становится рамка качества данных, тем более сложными и многообразными становятся требования к качеству данных, они индивидуализируются с позиции отдельных пользователей, их исследовательских целей. Пользователь все более участвует в оценке качества данных, а в некоторых случаях, наверное, только он и может ее осуществить. Оценка качества данных становится *распределенной* между разными субъектами, вовлеченными в их производство и использование. Другая проблема заключается в том, что процесс совокупной оценки качества данных растягивается во времени по мере ввода данных в оборот, в активное использование. Попытка использовать данные для новых исследовательских целей новыми пользователями может привести к обнаружению новых ошибок и ограничений, которые невидимы или неактуальны при генерации данных. В этом плане оценка качества данных никогда не будет не только полной, но и законченной.

Не все параметры качества данных можно представить через точные значения и отклонения от эталона, но эти измерения и выделение этапов генерации административных данных *организуют* отчеты о качестве данных. Хотя есть ядро индикаторов, присутствующее во многих рамках качества данных, это — гибкие конструкции, дополняемые в зависимости от целей оценки. И какой бы очевидной ни казалась рекомендация, важно фиксировать все, что исследователи делают с данными: «Лучший

способ оценить качество административного источника — это получить доскональные знания об этом источнике, включая основную цель и способы сбора и обработки данных. Четкое понимание источника позволит более точно оценить его сильные и слабые стороны» [35, р. 41]. Руководствуясь этой рекомендацией, представим описание таких административных данных, как сведения о доходах, расходах и имуществе депутатов региональных законодательных органов власти.

**Оценка качества сведений о доходах и имуществе
депутатов региональных законодательных органов власти**

Генерация и открытое размещение сведений о доходах, расходах, об имуществе и обязательствах имущественного характера отдельных категорий лиц и членов их семей регламентируются федеральным законодательством и нормативными документами субъектов Федерации. Они определяют круг лиц, которые должны представлять такие сведения, сроки представления и содержание сведений. Указами Президента РФ утверждена форма справки о доходах, расходах, об имуществе и обязательствах имущественного характера, которую представляют декларанты², и определяется, что из представленной информации подлежит размещению на официальных сайтах федеральных государственных органов, органов государственной власти субъектов РФ и организаций³. Различия в содержании представляемых и размещаемых сведений отражены в таблице 1.

При разработке документов, регламентирующих размещение сведений на сайтах органов государственной власти субъектов РФ, рекомендовано руководствоваться упомянутым Указом Президента РФ от 08.07.2013 № 613, но рекомендация не подразумевает единообразное воспроизводство правил размещения сведений о доходах и имуществе во всех субъектах Федерации [3]; возможно, этим и объясняется многообразие форматов размещения сведений на сайтах региональных законодательных органов власти.

² Указ Президента Российской Федерации от 23.06.2014 г. № 460 «Об утверждении формы справки о доходах, расходах, об имуществе и обязательствах имущественного характера и внесении изменений в некоторые акты Президента Российской Федерации». — URL: <http://www.kremlin.ru/acts/bank/38589> (Дата обращения 16.02.2021.)

³ Указ Президента Российской Федерации от 08.07.2013 № 613 «Вопросы противодействия коррупции» (вместе с «Порядком размещения сведений о доходах, расходах, об имуществе и обязательствах имущественного характера отдельных категорий лиц и членов их семей на официальных сайтах федеральных государственных органов, органов государственной власти субъектов Российской Федерации и организаций и предоставления этих сведений общероссийским средствам массовой информации для опубликования»). — URL: http://www.consultant.ru/document/cons_doc_LAW_148924/ (Дата обращения 17.02.2021.)

Таблица 1

Содержание представляемых и размещаемых на сайтах органов власти и организаций сведений о доходах, расходах, об имуществе и обязательствах имущественного характера отдельных категорий лиц и членов их семей

Разделы сведений	Представляемые сведения	Размещаемые сведения
доходы		
по видам и величинам	√	
общая сумма	√	√
расходы на приобретение земельных участков, иной недвижимости, транспортных средств, ценных бумаг	√	√
имущество (объекты собственности)	√	√
транспортные средства	√	√
счета в банках и иных кредитных организациях (наименование банка, вид и валюта счета, дата открытия, остаток средств на счете, поступления за отчетный период)	√	
ценные бумаги, в том числе акции и иное участие в коммерческих организациях и фондах	√	
обязательства имущественного характера (объекты, находящиеся в пользовании)	√	√
обязательства финансового характера (в качестве кредитора или должника, с указанием размера финансовых обязательств, оснований возникновения)	√	

Своевременность, быстрота представления — временной промежуток между датой, когда данные стали доступны, и периодом, который эти данные описывают. Депутаты законодательных органов власти представляют сведения ежегодно, отчетная дата для сведений, включенных в нашу пилотажную базу данных, — 31 декабря 2019 г. Сведения о доходах и имуществе должны быть представлены до 30 апреля года, следующего за отчетным. При скачивании файлов со сведениями с официальных сайтов законодательных органов власти субъектов РФ в октябре 2020 г. мы столкнулись с «последствиями» выборов в региональные парламенты в сентябре 2020 г.: в одном из регионов для нового депутатского корпуса был создан новый сайт органа власти, на который не перенесены сведения о доходах и имуществе депутатов прежнего созыва. Нужно также учитывать, что в течение 30 дней после представления сведений депутаты могут внести уточнения, которые, как правило, публикуются отдельными файлами. В реальности самая поздняя дата размещения уточненных сведений — ноябрь 2020 г.,

и есть уточнения сведений за 2018 г. и ранее, размещенные только в 2020 г. Поэтому оптимальный временной коридор для сбора сведений о доходах и имуществе депутатов органов власти субъектов РФ — примерно через месяц после даты их представления, но до единого дня голосования в сентябре, с отслеживанием внесенных уточнений.

Доступность — насколько легко данные могут быть обнаружены и получены из мест хранения. Включает пригодность форм, в которых существуют данные, средства распространения, доступность метаданных.

Совокупность сведений о доходах и имуществе депутатов региональных парламентов можно собрать либо с официальных сайтов каждого органа власти, либо с сайта «Декларатор», на котором они уже аккумулированы⁴. Для лучшего понимания ситуации в пилотажном проекте мы реализовали первый вариант сбора сведений, но были вынуждены в ряде случаев обратиться к «Декларатору» для поиска страниц со сведениями на сайтах органов власти и непосредственно за сведениями.

В большинстве случаев при размещении сведений о доходах и имуществе депутатов на сайтах региональных парламентов ключевыми являются разделы «Противодействие коррупции» или о составе депутатского корпуса. Для 78 регионов мы смогли найти эти сведения, начиная поиск с главных страниц. На сайтах трех региональных органов власти раздел со сведениями был найден только по ссылке, указанной на «Деклараторе», маршруты до этих страниц с главной страницы сайта реконструировать не смогли. И сведения о депутатах четырех парламентов мы были вынуждены скачивать непосредственно с «Декларатора». Делали это в том случае, когда даже по указанной ссылке сведения отсутствовали: были только за 2018 г. и ранее; только для служащих аппарата законодательного органа власти, но не депутатов; не могли быть скачаны из-за технической ошибки; отсутствовали из-за смены состава депутатов после выборов в сентябре 2020 г. и запуска нового сайта. Таким образом, установлено, что при переходе на новую версию сайта или выборов нового состава депутатов законодательного органа власти данные о прежних депутатах могут быть как перемещены в архивные разделы сайта, так и «пропасть» из публичного пространства.

Разнообразие мест размещения сведений повторяется в разнообразии способов и форматов размещения. В 2019 г. они были следующими: файлы со сводной таблицей о доходах и имуществе всех депу-

⁴ «Декларатор» — «постоянно пополняемая база данных о российских публичных должностных лицах: депутатах, чиновниках, судьях, представителях региональной и муниципальной власти, сотрудниках госкорпораций, госкомпаний, бюджетных учреждений и прочих организаций». — URL: <https://declarator.org/>

татов и членов их семей в разных форматах (59 субъектов Федерации); сводные таблицы со сведениями на странице сайта органа власти (4 региона); наборы файлов разных форматов со сведениями о доходах и имуществе отдельно каждого депутата и членов его семьи, размещенные на одной интернет-странице (13 субъектов Федерации); индивидуальные файлы или таблицы на отдельных вкладках на персональных страницах депутатов (9 регионов). Последний вариант — самый трудоемкий при сборе первичных сведений.

Интерпретируемость — наличие дополнительных сведений и метаданных, необходимых для корректной интерпретации данных и результатов анализа. Содержательные особенности запрашиваемых у должностных лиц и депутатов сведений раскрываются в ежегодно обновляемых методических рекомендациях Министерства труда и социальной защиты⁵. Основные моменты, касающиеся сведений, размещаемых в открытом доступе:

— все сведения представляются с учетом семейного положения, в котором находился депутат по состоянию на отчетную дату (31 декабря 2019 г.);

— доход трактуется шире, чем в статье 41 Налогового кодекса РФ, и в Рекомендациях все виды поступлений, которые должны быть учтены, перечислены. Полученные доходы, в том числе по основному месту работы, указываются без вычета налога на доходы физических лиц. Также перечислены виды поступлений, не включаемые в доход;

— раздел о расходах заполняется, если в отчетный период депутаты или члены их семей приобретали земельные участки, недвижимость, транспортные средства и расходы по этим сделкам превышали общий доход депутата и его супруги (супруга) за три последних года, отчетный (2019 г.) и два предыдущих;

— при заполнении раздела об объектах собственности указываются все объекты недвижимости, принадлежащие депутату, его супруге (супругу) и (или) несовершеннолетним детям на праве собственности, независимо от того, когда они были приобретены, в каком регионе РФ или в каком государстве зарегистрированы. Также указываются объекты недвижимости, принадлежащие гражданам, зарегистрированным в качестве индивидуальных предпринимателей. Для всех объектов недвижимости должны быть указаны вид собственности (индивидуальная, совместная или долевая с указанием долей) и общий размер (площадь) объекта;

⁵ Письмо Минтруда России от 27.12.2019 № 18-2/10/В-11200 «Методические рекомендации по вопросам представления сведений о доходах, расходах, об имуществе и обязательствах имущественного характера и заполнения соответствующей формы справки в 2020 году (за отчетный 2019 год)». — URL: http://www.consultant.ru/document/cons_doc_LAW_342265/ (Дата обращения 17.02.2021.)

— в разделе о транспортных средствах (легковых и грузовых автомобилях, мототранспортных средствах, сельскохозяйственной технике, водном и воздушном транспорте, иных транспортных средствах) о каждой единице техники указываются марка, модель, год изготовления, вид собственности и место регистрации;

— в разделе об имущественных обязательствах указывается недвижимое имущество, находящееся во временном пользовании депутатов, его супруги (супруга), несовершеннолетних детей. Также отражаются сведения о недвижимом имуществе, находящемся в пользовании депутата, зарегистрированного в качестве индивидуального предпринимателя. В сведениях об имуществе, используемом одним из супругов, не требуется указывать все объекты собственности другого супруга, если нет их фактического использования и они уже указаны среди объектов собственности.

Убедительность, достоверность. Проверка достоверности сведений о доходах и имуществе, представляемых депутатами региональных парламентов, регулируется юридическими нормами. В каждом субъекте Федерации нормативными документами определяются подразделения в органах власти (как правило, это кадровые), в которые представляют сведения и которые в случае необходимости осуществляют их проверку, а также основания для проверки достоверности. Для обеспечения достоверности представляемых сведений декларантам рекомендуется заполнять справку на основании правоустанавливающих и иных подтверждающих официальных документов [2; 3]. Нарушения в представлении сведений обычно связаны с небрежностью заполнения справок, незнанием и непониманием требований к их заполнению [7]. Проблемы при заполнении, как правило, возникают из-за трактовки доходов, при указании доходов, получаемых от вкладов, и остатков на счетах, преимущественно из-за отсутствия единой формы документа, выдаваемого банками [4]. Но возможные здесь ошибки «невидимы» для исследователей, так как детализация этих и других доходов не публикуется в открытом доступе: в размещенных на сайтах сведениях указана общая сумма доходов.

В зарубежных странах накоплен обширный опыт сравнения опросных и административных данных о доходах, показывающий, что административные данные обычно содержат более высокие значения доходов, чем сообщаемые этими же людьми в опросах, и с увеличением доходов выше вероятность занижить сумму, сообщаемую в опросе [6; 10; 26; 36]. Учитывая всю процедуру сбора сведений о доходах и имуществе депутатов и членов их семей, опыт сравнения опросных и административных данных, полагаем, что рассматриваемые нами данные более достоверны, чем полученные в массовых опросах домохозяйств.

Сравнимость — *возможны ли демографические, пространственные сравнения.* Содержание размещаемых сведений позволяет сравнивать характеристики доходов и имущества депутатов-одиночек или супружеских пар из разных регионов; возможно сравнение доходов и имущества супругов и в целом мужчин и женщин из данной совокупности супружеских домохозяйств. Потенциально при внесении в массив дополнительных данных (возраст, характеристики основной занятости, членство в политических партиях и др.) возможны другие сравнения, а при создании массивов аналогичных данных для других групп лиц, предоставляющих сведения, возможны сравнения разных категорий декларантов.

Согласованность (сопоставимость) — *возможность сравнивать декларационные данные с данными, полученными из других источников.* Такими «другими источниками» выступают прежде всего выборочные обследования населения и домохозяйств. Какие возможности и ограничения для сопоставлений мы видим?

— Даже проживая в расширенной семье, со взрослыми детьми или с родственниками старших поколений, лица, предоставляющие сведения, не включают их доходы и имущество в свой декларационный пакет. В этом плане «единица наблюдения» — супружеская пара или отдельный индивид с несовершеннолетними детьми или без них — отличается от привычных единиц наблюдения в социологии благосостояния, которыми являются индивид или домохозяйство. Но организация данных (наличие номеров-идентификаторов супругов) во всероссийских обследованиях домохозяйств, проводимых, к примеру, Росстатом, позволяет конструировать такие единицы наблюдения и сопоставлять результаты анализа данных из двух разных источников — опросных и административных.

— В большинстве обследований населения фиксируется ежемесячный доход или полученный в течение 30 дней перед обследованием. Из известных нам есть только одно обследование — Выборочное наблюдение доходов населения и участия в социальных программах (ВНДН), проводимое ежегодно Росстатом⁶, — в котором как и в декларациях депутатов, собирают данные о доходе, полученном в течение года.

— Другое ограничение на сопоставление связано с измерением дохода: в обследованиях, как правило, фиксируют данные о располагаемом доходе, в декларационных сведениях — о доходе до вычета налогов.

⁶ Итоги выборочного наблюдения доходов населения и участия в социальных программах / Росстат. Федеральные статистические наблюдения по социально-демографическим проблемам. — URL: https://rosstat.gov.ru/itog_inspect

Источники ошибок данных. Мы не даем оценку тому, насколько полученные значения близки истинным, а фиксируем основные источники ошибок, которые потенциально приводят к неточности и неполноте данных. Для каждого этапа генерации данных и их преобразования в файл «объект – признак» — на этапе заполнения сведений о доходе и имуществе, при размещении сведений на сайтах законодательных органов власти, при сборе этих сведений исследователями и формировании файла данных — характерны специфические ошибки, но проблематично однозначно разделить источники ошибок двух первых этапов. Полагаем, что контроль сведений, их открытая публикация и возможность проверки снижают на этапе представления недостоверность и неполноту, если не фактическую, то юридическую.

Совокупность депутатов региональных парламентов и их семей не является репрезентативной для населения страны (в классическом понимании репрезентативности эмпирического объекта), но «репрезентирует сама себя»: мы работаем со сплошной выборкой регионального депутатского корпуса по состоянию на конец 2019 г. Тем не менее ее формирование не обошлось без *ошибок репрезентации*:

— отсутствие единиц наблюдения из-за нормативных ограничений — в одном из регионов (Нижегородская обл.) размещаются сведения о доходах и имуществе только депутатов, исполняющих свои обязанности на постоянной основе (руководители органа власти, комитетов), поэтому в открытых источниках опубликованы сведения о 15 членах Законодательного собрания из 50;

— депутаты могут по тем или иным причинам досрочно выбыть из депутатского корпуса, поэтому на отчетную дату численность декларантов может быть меньше законодательно определенной. В то же время на сайтах некоторых законодательных органов размещены сведения о доходах и имуществе бывших депутатов действующего созыва, сложивших полномочия. В целом для 37 регионов численность депутатов, сведения о доходах и имуществе которых размещены, совпадает с нормативным размером депутатского корпуса; в пяти регионах размещены сведения о большем количестве депутатов, чем предусмотрено; количество опубликованных деклараций меньше нормативной численности депутатов в 43 субъектах Федерации. Разница между фактической и нормативной численностью для отдельных регионов не превышает четырех человек (в большую и в меньшую сторону), за исключением Нижегородской области. При должной совокупной численности депутатского корпуса региональных законодательных органов власти, составляющей 3983 человека, выборка включает 3896 наблюдений (97,8% от ожидаемого количества).

Мы считали *ошибкой измерения* ситуацию явного отклонения от «Методических рекомендаций» Минтруда или такие пропуски дан-

ных, которые нам не удавалось ликвидировать при определенных допущениях и вменении значений переменных. В отношении объектов недвижимости идеальным по полноте будет описание с указанием для каждого из них вида собственности (индивидуальная, совместная или долевая), размера доли декларанта в праве долевой собственности, размера объекта (площадь, протяженность или указание, что определение площади для данного объекта не предусмотрено), страны расположения недвижимости. Лишь в отношении последней переменной не обнаружено пропусков данных. В целом различия в описании собственности депутатов региональных законодательных органов власти и членов их семей отражены в таблице 2. Полное описание параметров собственности на объекты недвижимости представлено в сведениях депутатов 52 региональных законодательных органов власти. В ситуации, когда для объекта недвижимости, а также транспортных средств не указан вид собственности, мы вменяли значение «индивидуальная».

Рекомендации Минтруда о предоставлении сведений о доходе и имуществе подразумевают, что жилье, в котором зарегистрирован декларант, отражается либо как объект собственности, либо как объект пользования. При отсутствии собственности невключение в декларацию характеристик жилья в пользовании приводит к *ошибке неответа*, причем чаще отсутствуют сведения о недвижимом имуществе супругов (1,3%), чем самих депутатов (0,9%). Ошибки неответа также возникают при пропуске части сведений о каком-либо из объектов в собственности или пользовании — 1,9% наблюдений.

Среди других ошибок измерения отметим *неоднозначность в трактовке сведений о каком-либо из объектов в собственности или пользовании*, что фактически означает пропуск данных (1,6% наблюдений); по 0,2% наблюдений содержат ошибки несогласованности сведений (например, для объектов в совместной собственности указываются доли в праве) и опечатки, некорректность сведений.

В целом названные ошибки измерения и репрезентации, зафиксированные при вводе данных и программными средствами в уже созданном файле, присутствуют в 6% наблюдений. Использование сведений прошлых лет для исправления этих ошибок возможно, но не во всех случаях. Во-первых, для некоторых депутатов 2019 г. — первый в исполнении полномочий, и сведения за прошлые годы отсутствуют. Во-вторых, как показывает обращение в неоднозначных ситуациях к декларациям прошлых лет, названные ошибки воспроизводятся (наблюдается автокорреляция ошибок административных данных [31]), то есть размещаемые сведения об имуществе копируются из прошлогодних с ошибками и с изменением некоторых данных, например, о доходе.

Таблица 2

Варианты описания параметров прав собственности на объекты недвижимости в сведениях о доходах и имуществе депутатов региональных законодательных органов власти

Параметры описания	Комбинации параметров описания прав собственности на объекты недвижимости									
Есть объекты, для которых вид собственности указан как «индивидуальная»										
да	√					√	√	√		
нет		√	√	√					√	√
Есть объекты в долевой собственности										
да, с определением «долевая»	√	√				√	√		√	√
да, с указанием только размера доли				√	√			√		
Есть объекты, для которых вид собственности определен как «совместная»										
да	√	√	√				√	√		√
нет				√	√				√	
Для объектов в долевой собственности указан размер доли в праве собственности										
да	√	√	√	√	√			√	√	
нет							√			√
Количество субъектов Федерации с такой комбинацией параметров прав собственности на объекты недвижимости в сведениях о доходе и имуществе депутатов законодательных органов власти	52	11	12	5	1	1	1	1	1	1

Релевантность содержания данных для исследовательских целей

На этапе пилотажа коротко представим эмпирические распределения нескольких переменных из созданного файла данных, демонстрирующие их аналитический потенциал.

Как отмечено выше, личные доходы депутатов и их супругов в размещенных сведениях — это доход до вычета налогов, тогда как в Выборочном наблюдении доходов населения и участия в социальных программах Росстата, единственном обследовании, в котором денежные поступления

также фиксируются за год, доход домохозяйства — это располагаемый доход. Тем не менее, если в выборке ВНДН выбрать домохозяйства, имеющие в своем составе одну супружескую пару, при всех методических различиях можно сопоставить распределения суммы личных доходов депутатов и их супругов с располагаемым доходом отобранных домохозяйств ВНДН. Но даже такое грубое, приблизительное сопоставление показывает, что по сравнению с данными массовых обследований домохозяйств административные данные охватывают более высокодоходные слои (см. рис.)⁷, что дает основания для дальнейшей работы со сведениями о доходах и имуществе с возможным увеличением набора данных за счет других категорий государственных служащих и политиков.

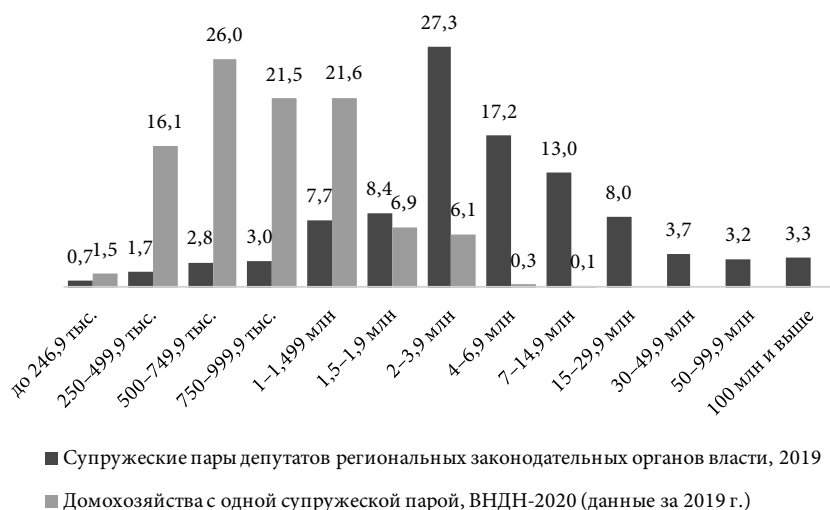


Рис. Сопоставление распределений (в %) суммарных личных доходов в супружеских парах депутатов региональных законодательных органов власти и совокупных доходов домохозяйств с одной супружеской парой, ВНДН

Уникальность рассматриваемых административных данных в том, что они содержат информацию о видах собственности на недвижимость и транспортные средства, потенциально давая возможность оценить индивидуализацию или «общность» в правах собственности [18; 28]. В таблице 3 представлены сведения о жилой недвижимости из деклараций депутатов (без учета собственности их супругов). Из всех видов собственности на жилье наиболее распространена индивидуальная, а совместная, преимущественно «унаследованная» из приватизации 1990-х гг., является миноритарной. Возможно, по этой при-

⁷ Обследование проведено в январе 2020 г., отражены данные о доходах членов домохозяйств за 2019 г.

чине, а не только из-за ошибок при размещении сведений, в наборах деклараций семи субъектов Федерации нет объектов недвижимости, находящихся в совместной собственности (см. табл. 2).

Таблица 3

Распространенность объектов жилой недвижимости разных видов собственности у депутатов региональных законодательных органов власти, 2019 г.

Объекты жилой недвижимости в собственности	Все депутаты региональных законодательных органов власти	Депутаты 52 законодательных органов власти, в сведениях которых содержится <i>полное</i> описание объектов недвижимости	Депутаты 33 законодательных органов власти, в сведениях которых содержится <i>неполное</i> описание объектов недвижимости
Количество депутатов, имеющих объекты в собственности (множественный выбор), %			
Квартиры в собственности			
индивидуальной	44,2	44,2	44,2
совместной	7,5	9,5	4,5
долевой	21,7	21,2	22,3
Жилые дома в собственности			
индивидуальной	33,4	33,1	33,8
совместной	2,3	2,8	1,6
долевой	7,2	7,5	6,8
Количество объектов, находящихся в собственности депутатов, %			
Квартиры в собственности			
индивидуальной	66,5	64,5	69,4
совместной	8,7	11,4	4,7
долевой	24,8	24,1	25,9
Всего:	100,0	100,0	100,0
Жилые дома в собственности			
индивидуальной	80,6	80,1	81,5
совместной	4,5	5,5	3,0
долевой	14,8	14,4	15,5
Всего:	100,0	100,0	100,0

Сравнение двух наборов деклараций, отличающихся полнотой сведений о параметрах частной собственности, позволяет предполагать, что реализованная процедура вменения пропущенных данных о виде собственности снижает показатели распространенности объектов жилой недвижимости, находящейся в совместной собственности. Хотя в целом соотношение распространенности разных видов

собственности не меняется, такое вменение потенциально может быть источником ошибок данных на этапе их обработки.

Обсуждение

Введение в оборот нецелевых административных данных подразумевает пилотаж данных, методология которого включает оценку их качества и аналитического потенциала для решения исследовательских задач. В данном случае работа с количественными административными данными подразумевала, во-первых, их подготовку — структурирование, кодирование, создание переменных, то есть непосредственно формирование файлов данных в виде привычных матриц «объект — признак». Во-вторых, мы работали с сопутствующими документами, чтобы понять содержание данных, характеристики «объектов», о которых они собраны, и их «признаков». В-третьих, мы дали оценку качества этих данных, понимая, что некоторые аспекты качества могут проявиться только в процессе работы с ними: «...исследователям, возможно, придется начинать исследовательский проект, не обладая полнотой знания о данных, которые они используют. Хотя эта работа может быть сложной для первопроходцев, эту проблему можно преодолеть, если исследователи задокументируют сведения о характеристиках, преимуществах и недостатках конкретных ресурсов административных данных и поделятся ими» [15, р. 8].

Выполненный анализ качества данных развивает тезис о том, что фундаментально неверно предполагать, будто некоторые данные содержат ошибки, в то время как другие — нет. Административным данным также свойственны ошибки [9; 24; 31], хотя прямой перенос терминологии общей ошибки исследования, принятой в социологических опросах и обследованиях, на другой тип данных не всегда оказывается релевантным: «...необходимо больше усилий для концептуализации компонентов ошибок в административных данных и стандартизации терминологии в разных контекстах» [19, р. 194].

Большинство ошибок рассмотренных административных данных вызвано неунифицированностью содержания размещаемых сведений. Но это тот случай, когда исходные сведения, а не данные создаются не для исследовательских целей, и наблюдаемое многообразие способов их размещения и представления характеристик объектов собственности вполне может удовлетворять изначальным, административным, целям. Отчасти зафиксированное разнообразие обусловлено выбором эмпирического объекта для пилотажа: каждому территориальному законодательному органу власти *рекомендовано* руководствоваться нормами, обязательными для федеральных органов власти. Можно предполагать, что сведения о доходе и имуществе государственных служащих из федеральных министерств и ведомств будут более однообразные по форме, хотя в описании набора данных «Сведения о декларированных доходах государственных гражданских служащих центральных аппаратов феде-

ральных министерств РФ», подготовленного Центром перспективных управленческих решений в рамках проекта «Инфраструктура научно-исследовательских данных», также отмечаются ошибки, но иного рода: произвольные формулировки названий должностей служащих, фрагментарность сведений о доходах служащих, занимающих неруководящие должности⁸.

Нераспространенная практика архивации интернет-сайтов может привести к потере сведений о доходах и имуществе депутатского корпуса, ставшего бывшим не только в результате новых выборов, но и при реорганизации органов власти. Так, на сайте Министерства науки и высшего образования Российской Федерации, созданного в мае 2018 г. при преобразовании Министерства образования и науки, размещены сведения о доходах и имуществе государственных служащих только за 2019 г., без архива сведений сотрудников «предыдущей версии» министерства⁹.

Таким образом, работа со сведениями о доходах и имуществе других категорий лиц, не только депутатов региональных законодательных органов власти, может обнаружить другие ошибки, иной их масштаб, иные особенности данных. В созданном наборе данных ошибки репрезентации менее остры по сравнению с ошибками измерения, и неполнота сведений ограничила возможные исследовательские задачи. Так, в административных данных только 18 регионов указано назначение земельных участков; детализация имущественных обязательств (пользуются ли имуществом на правах аренды, фактического использования и т. п.) содержится в сведениях о 17 регионах. От анализа этих аспектов имущественной дифференциации мы пока отказываемся. Сравнение доходов и имущественной обеспеченности депутатов, занимающих разные позиции в структуре парламента (руководители органов власти и их заместители, руководители комитетов или «рядовые» депутаты), потребует внесения в файл данных дополнительной информации, так как непосредственно в сведениях она указана только в 42 регионах из 85.

Заключение

Мы работали с видимой частью «айсберга». От нас скрыты как процессы генерации административных данных, так и их часть: на сайтах органов власти и организаций не размещаются сведения о средствах на банковских счетах и финансовых активах депутатов. Но социальные практики и отношения, проявившиеся в административно созданной антикоррупционной открытости, интересны не меньше,

⁸ Сведения о декларированных доходах государственных гражданских служащих центральных аппаратов федеральных министерств РФ. — URL: <https://data-in.ru/data-catalog/datasets/150/> (Дата обращения 30.03.2021.)

⁹ Министерство науки и высшего образования Российской Федерации. Сведения о доходах, расходах, об имуществе и обязательствах имущественного характера. — URL: https://minobrnauki.gov.ru/open_ministry/anticorr/propertyinfo/ (Дата обращения 16.04.2021.)

чем создаваемые административные базы данных, и могут быть предметом отдельных исследований.

Жесткое правоприменение, зачастую в условиях неопределенности норм, приводило к лишению депутатских полномочий¹⁰, но и сами депутаты, особенно муниципального уровня, отказывались исполнять полномочия, если они сопряжены с трудностями сбора банковских справок для сведений и несоразмерным наказанием за неточности в декларациях [4]. Постепенно Совет Федерации был вынужден смягчить наказание за незначительные ошибки в сведениях.

Антикоррупционная открытость сведений о доходах и имуществе публичных должностных лиц и политиков — широко распространенная практика, исследования которой не всегда демонстрируют консистентные результаты. Так, на примере законодательных органов штатов США показано, что чем строже в законодательстве штатов нормы о финансовой открытости, тем ниже в их составе доля бизнесменов [33]. Но сравнение нескольких составов парламентов в десяти странах до и после введения норм раскрытия финансовой информации скорее отрицает влияние таких институциональных изменений на профессиональный состав избираемых органов власти [37]. Сравнение состава кандидатов более 40 тыс. муниципальных выборов в России в период 2009–2017 гг., на середину которого пришелся Указ Президента РФ о представлении кандидатами и депутатами сведений о доходах и имуществе, показывает сокращение доли повторно выдвигаемых кандидатов [34]. В целом предметом исследований становятся влияние институтов прозрачности на отбор политиков определенного «качества», их решения баллотироваться, принимаемые через взвешивание индивидуальных выгод и издержек от раскрытия финансовой информации, с учетом конфликтов интересов [13].

Какими бы ни были эмпирические результаты исследований в разных странах, при работе со сведениями о доходах и имуществе государственных служащих и политиков нельзя не учитывать, что институциональные условия генерации административных данных могут менять состав тех, о ком собираются эти данные.

ЛИТЕРАТУРА

1. *Джанек Л., Крейтер Ф., Берг М. и др.* Отчет ААРОР о больших данных: 12 февраля 2015 // Американская ассоциация исследователей общественного мнения / Пер. с англ. Д. Рогозина, А. Ипатовой, Е. Вьюговской; Предисл. Д. Рогозина. М., 2015 [электронный ресурс]. Дата обращения 31.01.2021. URL: https://book.wciom.ru/knigi/knigi/otchyot_aaopor_o_bolshih_dannyh/

¹⁰ Муниципальные депутаты получили послабление по декларациям // Законодательное Собрание Ростовской области. 29.07.2019. — URL: https://zsgro.ru/press_center/news/95/21176/ (Дата обращения 02.04.2021.)

2. *Ищунин А.Г.* Проведение проверок полноты и достоверности сведений о доходах, расходах, об имуществе и обязательствах имущественного характера, предоставленных государственными гражданскими служащими Курской области // Государство и общество: вчера, сегодня, завтра. 2018. № 3. С. 24–32.
3. *Кабанов П.А.* Антикоррупционное процессуальное законодательство субъектов Российской Федерации: формирование и содержание // Актуальные проблемы экономики и права. 2018. Т. 12. № 2. С. 300–333. DOI: 10.21202/1993-047X.12.2018.2.300-333
4. *Табакова И.Н.* Актуальные вопросы, возникающие при представлении сведений о доходах, расходах, об имуществе и обязательствах имущественного характера // Актуальные вопросы противодействия коррупции в сфере образования: Сб. науч. ст. по материалам IV Сибирского антикоррупционного форума, Красноярск, 26–27 декабря 2018 г. / Отв. ред. И.А. Дамм, Е.А. Акунченко. Красноярск: Сиб. федер. ун-т, 2019. С. 207–213.
5. *Цивинская А.О., Губа К.С.* Мониторинг эффективности образовательных организаций как источник данных о российском высшем образовании // Университетское управление: практика и анализ. 2020. Т. 24. № 2. С. 121–130. DOI: 10.15826/umpa.2020.02.018
6. *Черкашина Т.Ю.* Измерение доходов населения: варианты оценки смещения // Вопросы экономики. 2020. № 1. С. 127–144. DOI: 10.32609/0042-8736-2020-1-127-144
7. *Чернышева Л.А.* Дисциплинарная ответственность государственных служащих за предоставление недостоверных или неполных сведений о доходах и расходах // КриминалистЪ. 2015. № 2. С. 44–47.
8. *Чуриков А.В.* Основы построения выборки для социологических исследований. М.: Ин-т фонда «Общественное мнение», 2020. — 240 с.
9. *Abowd J.M., Stinson M.H.* Estimating measurement error in annual job earnings: A comparison of survey and administrative data // Review of Economics and Statistics. 2013. Vol. 95. No. 5. P. 1451–1467. DOI: 10.1162/REST_a_00352
10. *Angel S., Heuberger R., Lamei N.* Differences between household income from surveys and registers and how these affect the poverty headcount: Evidence from the Austrian SILC // Social Indicators Research. 2018. Vol. 138. No. 2. P. 575–603. DOI: 10.1007/s11205-017-1672-7
11. *Biemer P.* Total Survey Error: Design, Implementation, and Evaluation // Public Opinion Quarterly. 2010. Vol. 74. Iss. 5. P. 817–848. DOI: 10.1093/poq/nfq058
12. *Brackstone G.* Managing data quality in a statistical agency // Survey Methodology. 1999. Vol. 25. No. 2. P. 139–150.
13. *Braendle T.* Do Institutions affect Citizens's Election into Politics? // Journal of Economic Surveys. 2016. Vol. 30. No. 2. P. 205–227. DOI: 10.1111/joes.12098
14. *Citro C.F.* From multiple modes for surveys to multiple data sources for estimates // Survey Methodology. Vol. 40. Iss. 2. 1 December 2014. P. 137–161.
15. *Connelly R., Playford C.J., Gayle V., Dibben C.* The role of administrative data in the big data revolution in social science research // Social Science Research. September 2016. Vol. 59. P. 1–12. DOI: 10.1016/j.ssresearch.2016.04.015
16. *Couper M.P.* New Developments in Survey Data Collection // Annual Review of Sociology. 2017. Vol. 43. No. 1. P. 121–145. DOI: 10.1146/annurev-soc-060116-053613

17. Eurostat. Methodological Documents — Definition of Quality in Statistics // Working Group “Assessment of quality in statistics”. Sixth meeting. Luxembourg, 2–3 October 2003 [online]. Accessed 10.02.2021. URL: <https://ec.europa.eu/eurostat/documents/64157/4373735/02-ESS-quality-definition.pdf>
18. *Frémeaux N., Leturcq M.* Inequalities and the individualization of wealth // *Journal of Public Economics*. 2020. Vol. 184, April. Article number 104145. DOI: 10.1016/j.jpubeco.2020.104145
19. *Groen J.A.* Sources of Error in Survey and Administrative Data: The Importance of Reporting Procedures // *Journal of Official Statistics*. 2012. Vol. 28. No. 2. P. 173–198.
20. *Groves R.M.* Research on Survey Data Quality // *Public Opinion Quarterly*. 1987. Vol. 51. Iss. 4. Winter 1987. P. s156–s172. DOI:10.1093/poq/51.4_part_2.s156
21. *Groves R.M.* Three Eras of Survey Research // *Public Opinion Quarterly*. 2011. Vol. 75. Iss. 5. P. 861–871. DOI: 10.1093/poq/nfr057
22. *Groves R.M., Lyberg L.* Total Survey Error: Past, Present, and Future // *Public Opinion Quarterly*. 2010. Vol. 74. Iss. 5. P. 849–879. DOI: 10.1093/poq/nfq065
23. *Hand D.J.* Statistical Challenges of Administrative and Transaction Data // *Journal of the Royal Statistical Society*. 2018. Vol. 181. Iss. 3. P. 555–605. DOI: 10.1111/rssa.12315
24. *Kapteyn A., Ypma J.Y.* Measurement Error and Misclassification: A Comparison of Survey and Administrative Data // *Journal of Labor Economics*. 2007. Vol. 25. No. 3. P. 513–551. DOI: 10.1086/513298
25. *Keller S., Korkmaz G., Orr M., Schroeder A., Shipp S.* The Evolution of Data Quality: Understanding the Transdisciplinary Origins of Data Quality Concepts and Approaches // *Annual Review of Statistics and Its Application*. Vol. 4. 7 March 2017. P. 85–108. DOI: 10.1146/annurev-statistics-060116-054114
26. *Kim C., Tamborini C.R.* Response error in earnings: An analysis of the survey of income and program participation matched with administrative data // *Sociological Methods & Research*. 2014. Vol. 43. No. 1. P. 39–72. DOI: 10.1177/0049124112460371
27. *Lavrakas P.J.* Presidential Address: Applying a Total Error Perspective for Improving Research Quality in the Social, Behavioral, and Marketing Sciences // *Public Opinion Quarterly*. 2013. Vol. 77. Iss. 3. P. 831–850. DOI: 10.1093/poq/nft033
28. *Lersch P.M., Vidal S.* My house or our home? Transitions into sole home ownership in British couples // *Demographic Research*. 2016. Vol. 35. Iss. 1. P. 139–166. DOI: 10.4054/DemRes.2016.35.6
29. *Lyberg L.* Survey Quality // *Survey Methodology*. 2012. Vol. 38. No. 2. P. 107–130.
30. OECD. Quality Framework and Guidelines for OECD Statistical Activities, Version 2011/1 [online]. Accessed 10.02.2021. URL: <https://www.oecd.org/sdd/qualityframeworkforoecdstatisticalactivities.htm>
31. *Pavlopoulos D., Vermunt J.K.* Measuring temporary employment. Do survey or register data tell the truth? // *Survey Methodology*. 2015. Vol. 41. Iss. 1. P. 197–214.
32. *Playford C.J., Gayle V., Connelly R., Gray A.J.* Administrative social science data: The challenge of reproducible research // *Big Data & Society*. 2016. December. P. 1–13. DOI: 10.1177/2053951716684143

33. *Rosenson B.A.* The Impact of Ethics Laws on Legislative Recruitment and the Occupational Composition of State Legislatures // *Political Research Quarterly*. 2006. Vol. 59. No. 4. P. 619–627. DOI: 10.1177/106591290605900411
34. *Szakonyi D.* Indecent Disclosures: Anti-Corruption Reforms and Political Selection // SSRN. March 27, 2020. DOI: 10.2139/ssrn.3101123
35. UNECE. Using administrative secondary sources for official statistics. A handbook of principles and practices. United Nations Economic Commission for Europe. 2011 [online]. Accessed 07.12.2020. URL: <https://unstats.un.org/unsd/EconStatKB/KnowledgebaseArticle10349.aspx>
36. *Valet P., Adriaans J., Liebig S.* Comparing survey data and administrative records on gross earnings: nonreporting, misreporting, interviewer presence and earnings inequality // *Quality and Quantity*. 2019. Vol. 53. No 1. P. 471–491. DOI: 10.1007/s11135-018-0764-z
37. *van Aaken A., Voigt S.* Do individual disclosure rules for parliamentarians improve government effectiveness? // *Economics of Governance*. 2011. Vol. 12. Iss. 4. P. 301–324. DOI: 10.1007/s10101-011-0100-8

СВЕДЕНИЯ ОБ АВТОРЕ

Черкашина Татьяна Юрьевна — кандидат социологических наук, доцент; ведущий научный сотрудник, Институт экономики и организации промышленного производства СО РАН; заведующая кафедрой общей социологии, Новосибирский национальный исследовательский государственный университет.

Телефон: +7 (383) 363-42-63. **Электронная почта:** touch@nsu.ru

Дата поступления: 21.04.2021.

SOTSIOLOGICHESKIY ZHURNAL = SOCIOLOGICAL JOURNAL. 2021.
VOL. 27. No. 3. P. 8–34. DOI: 10.19181/socjour.2021.27.3.8422

Research Article

TATYANA YU. CHERKASHINA^{1,2}

¹ Institute of Economics and Industrial Engineering, SB RAS.

17, Academician Lavrentiev Avenue, 630090, Novosibirsk, Russian Federation.

² Novosibirsk State University.

1, Pirogova str., 630090, Novosibirsk, Russian Federation.

RUSSIAN REGIONAL POLITICIANS' INCOME AND PROPERTY DECLARATIONS: A PILOT STUDY AND QUALITY ASSESSMENT OF ADMINISTRATIVE DATA

Abstract. The article presents the experience of converting non-targeted administrative data into research data, using as an example data on the income and property of deputies from local legislative bodies of the Russian Federation for 2019, collected as part of anti-corruption operations. This particular empirical fragment was selected for the pilot study of administrative data, which includes assessing the possibility of integrating scattered fragments of information into a single database, assessing quality of data and their relevance

for solving research problems, particularly analysis of high-income strata and the apparent trends towards individualization of private property. The system of indicators for assessing data quality includes their timeliness, availability, interpretability, reliability, comparability, coherence, errors of representation and measurement, and relevance. In the case of the data set in question, measurement errors are more common than representation errors. Overall the article emphasizes the notion that introducing new non-target data into circulation requires their preliminary testing, while data quality assessment becomes distributed both in time and between different subjects. The transition from created data to “obtained” data shifts the functions of evaluating its quality from the researcher-creator to the researcher-user. And though in this case data quality is in part ensured by the legal support for their production, the transformation of administrative data into research data involves assessing a variety of quality measurements — from availability to uniformity and accuracy.

Keywords: administrative data; data quality; representation errors; measurement errors; pilot study of administrative data.

For citation: Cherkashina, T.Yu. Russian Regional Politicians’ Income and Property Declarations: a Pilot Study and Quality Assessment of Administrative Data. *Sotsiologicheskii Zhurnal = Sociological Journal*. 2021. Vol. 27. No. 3. P. 8–34. DOI: 10.19181/socjour.2021.27.3.8422

REFERENCES

1. Japec L., Kreuter F., Berg M., Biemer P., Decker P., Lampe C., Lane J., O’Neil C., Usher A. Otchet AAPOR o bol’shikh dannykh: 12 fevralya 2015. [AAPOR Report on Big Data: February 12, 2015. Prepared for AAPOR Council by the Task Force.] Transl. from Eng. by D. Rogozin, A. Ipatova, E. V’yugovskaya; Pref. by D. Rogozin. Moscow: VCIOM publ., 2015 // *Russian Public Opinion Research Center (OAO “VCIOM”)*. Accessed 31.01.2021. URL: https://book.wciom.ru/knigi/knigi/otchyot_aapor_o_bolshih_dannykh/ DOI: 10.19181/socjour.2016.22.2.4262 (In Russ.)
2. Ishunin A.G. Conducting Checks of Completeness and Reliability of Data on Income, Expenses, about the Property and Liabilities of Property Nature provided by the Government Civil Servants in Kursk Region. *Gosudarstvo i obshchestvo: vchera, segodnya, zavtra*. 2018. No. 3. P. 24–32. DOI: 10.7256/2409-7136.2017.5.22713 (In Russ.)
3. Kabanov P.A. Anti-corruption procedural legislation of the Russian Federation subjects: formation and content. *Aktual’nye problemy ekonomiki i prava*. 2018. Vol. 12. No. 2. P. 300–333. DOI: 10.21202/1993-047X.12.2018.2.300-333 (In Russ.)
4. Tabakova I.N. Current Issues arising when submitting of Data on Income, Expenses, about the Property and Liabilities of Property Nature. *Aktual’nye voprosy protivodeistviya korruptsii v sfere obrazovaniya: Sb. nauch. st. po materialam IV Sibirskogo antikorrupcionnogo foruma, Krasnoyarsk, 26–27 dekabrya 2018 g.* [Topical Issues of Combating Corruption in the Education: Coll. of scientific papers. IV Siberian Anti-Corruption Forum, Krasnoyarsk, 26–27 December 2018.] Ed. by I.A. Damm, E.A. Akunchenko. Krasnoyarsk: Sib. feder. un-t publ., 2019. P. 207–213. (In Russ.)
5. Tsivinskaya A.O., Guba K.S. The Survey of HEIs Performance as a Data Source on Higher Education in Russia. *Universitetskoe upravlenie: praktika i analiz*. 2020. Vol. 24. No. 2. P. 121–130. DOI: 10.15826/umpa.2020.02.018 (In Russ.)
6. Cherkashina T.Yu. Measurement of population income: Variants of estimating biases. *Voprosy ekonomiki*. 2020. No. 1. P. 127–144. DOI: 10.32609/0042-8736-2020-1-127-144 (In Russ.)
7. Chernysheva L.A. Disciplinary Responsibility of Public Servants for Submission of False or Insufficient Information about Incomes and Expenditures. *Kriminalist*. 2015. No. 2. P. 44–47. (In Russ.)

8. Churikov A.V. *Osnovy postroeniya vyborki dlya sotsiologicheskikh issledovaniy*. [Fundamentals of sampling design for sociological research.] Moscow: In-t fonda "Obshchestvennoe mnenie" publ., 2020. 240 p. (In Russ.)
9. Abowd J.M., Stinson M.H. Estimating measurement error in annual job earnings: A comparison of survey and administrative data. *Review of Economics and Statistics*. 2013. Vol. 95. No. 5. P. 1451–1467. DOI: 10.1162/REST_a_00352.
10. Angel S., Heuberger R., Lamei N. Differences between household income from surveys and registers and how these affect the poverty headcount: Evidence from the Austrian SILC. *Social Indicators Research*. 2018. Vol. 138. No. 2. P. 575–603. DOI: 10.1007/s11205-017-1672-7
11. Biemer P. Total Survey Error: Design, Implementation, and Evaluation. *Public Opinion Quarterly*. 2010. Vol. 74. Iss. 5. P. 817–848. DOI: 10.1093/poq/nfq058
12. Brackstone G. Managing data quality in a statistical agency. *Survey Methodology*. 1999. Vol. 25. No. 2. P. 139–150.
13. Braendle T. Do Institutions affect Citizens's Election into Politics? *Journal of Economic Surveys*. 2016. Vol. 30. No. 2. P. 205–227. DOI: 10.1111/joes.12098
14. Citro C.F. From multiple modes for surveys to multiple data sources for estimates. *Survey Methodology*. Vol. 40. Iss. 2. 1 December 2014. P. 137–161.
15. Connelly R., Playford C.J., Gayle V., Dibben C. The role of administrative data in the big data revolution in social science research. *Social Science Research*. September 2016. Vol. 59. P. 1–12. DOI: 10.1016/j.ssresearch.2016.04.015
16. Couper M.P. New Developments in Survey Data Collection. *Annual Review of Sociology*. 2017. Vol. 43. No. 1. P. 121–145. DOI: 10.1146/annurev-soc-060116-053613
17. Eurostat. Methodological Documents — Definition of Quality in Statistics. *Working Group "Assessment of quality in statistics". Sixth meeting. Luxembourg, 2–3 October 2003*. Accessed 10.02.2021. URL: <https://ec.europa.eu/eurostat/documents/64157/4373735/02-ESS-quality-definition.pdf>
18. Frémeaux N., Leturcq M. Inequalities and the individualization of wealth. *Journal of Public Economics*. 2020. Vol. 184, April. Article number 104145. DOI: 10.1016/j.jpubeco.2020.104145
19. Groen J.A. Sources of Error in Survey and Administrative Data: The Importance of Reporting Procedures. *Journal of Official Statistics*. 2012. Vol. 28. No. 2. P. 173–198.
20. Groves R.M. Research on Survey Data Quality. *Public Opinion Quarterly*. 1987. Vol. 51. Iss. 4. Winter 1987. P. s156–s172. DOI:10.1093/poq/51.4_part_2.s156
21. Groves R.M. Three Eras of Survey Research. *Public Opinion Quarterly*. 2011. Vol. 75. Iss. 5. P. 861–871. DOI: 10.1093/poq/nfr057
22. Groves R.M., Lyberg L. Total Survey Error: Past, Present, and Future. *Public Opinion Quarterly*. 2010. Vol. 74. Iss. 5. P. 849–879. DOI: 10.1093/poq/nfq065
23. Hand D.J. Statistical Challenges of Administrative and Transaction Data. *Journal of the Royal Statistical Society*. 2018. Vol. 181. Iss. 3. P. 555–605. DOI: 10.1111/rssa.12315
24. Kapteyn A., Ypma J.Y. Measurement Error and Misclassification: A Comparison of Survey and Administrative Data. *Journal of Labor Economics*. 2007. Vol. 25. No. 3. P. 513–551. DOI: 10.1086/513298
25. Keller S., Korkmaz G., Orr M., Schroeder A., Shipp S. The Evolution of Data Quality: Understanding the Transdisciplinary Origins of Data Quality Concepts and Approaches. *Annual Review of Statistics and Its Application*. Vol. 4. 7 March 2017. P. 85–108. DOI: 10.1146/annurev-statistics-060116-054114
26. Kim C., Tamborini C.R. Response error in earnings: An analysis of the survey of income and program participation matched with administrative data. *Sociological Methods & Research*. 2014. Vol. 43. No. 1. P. 39–72. DOI: 10.1177/0049124112460371
27. Lavrakas P.J. Presidential Address: Applying a Total Error Perspective for Improving Research Quality in the Social, Behavioral, and Marketing Sciences. *Public Opinion Quarterly*. 2013. Vol. 77. Iss. 3. P. 831–850. DOI: 10.1093/poq/nft033

28. Lersch P.M., Vidal S. My house or our home? Transitions into sole home ownership in British couples. *Demographic Research*. 2016. Vol. 35. Iss. 1. P. 139–166. DOI: 10.4054/DemRes.2016.35.6.
29. Lyberg L. Survey Quality. *Survey Methodology*. 2012. Vol. 38. No. 2. P. 107–130.
30. OECD. *Quality Framework and Guidelines for OECD Statistical Activities, Version 2011/1*. Accessed 10.02.2021. URL: <https://www.oecd.org/sdd/qualityframeworkforoeecdstatisticalactivities.htm>
31. Pavlopoulos D., Vermunt J.K. Measuring temporary employment. Do survey or register data tell the truth? *Survey Methodology*. 2015. Vol. 41. Iss. 1. P. 197–214.
32. Playford C.J., Gayle V., Connelly R., Gray A.J. Administrative social science data: The challenge of reproducible research. *Big Data & Society*. 2016. December. P. 1–13. DOI: 10.1177/2053951716684143
33. Rosenson B.A. The Impact of Ethics Laws on Legislative Recruitment and the Occupational Composition of State Legislatures. *Political Research Quarterly*. 2006. Vol. 59. No. 4. P. 619–627. DOI: 10.1177/106591290605900411
34. Szakonyi D. Indecent Disclosures: Anti-Corruption Reforms and Political Selection. *SSRN*. March 27, 2020. DOI: 10.2139/ssrn.3101123
35. UNECE. *Using administrative secondary sources for official statistics. A handbook of principles and practices*. United Nations Economic Commission for Europe. 2011. Accessed 07.12.2020. URL: <https://unstats.un.org/unsd/EconStatKB/KnowledgebaseArticle10349.aspx>
36. Valet P., Adriaans J., Liebig S. Comparing survey data and administrative records on gross earnings: nonreporting, misreporting, interviewer presence and earnings inequality. *Quality and Quantity*. 2019. Vol. 53. No 1. P. 471–491. DOI: 10.1007/s11135-018-0764-z
37. van Aaken A., Voigt S. Do individual disclosure rules for parliamentarians improve government effectiveness? *Economics of Governance*. 2011. Vol. 12. Iss. 4. P. 301–324. DOI: 10.1007/s10101-011-0100-8

INFORMATION ABOUT THE AUTHOR

Tatyana Yu. Cherkashina — Candidate of Sociological Sciences, Leading Researcher, Institute of Economics and Industrial Engineering, SB RAS; Head, Department of General Sociology, Novosibirsk State University.

Phone: +7 (383) 363-42-63. **Email:** touch@nsu.ru

Received: 21.04.2021.
