

DOI: 10.34680/BENEFICIUM.2023.1(46).20-27

УДК 330.564.2:004.9(470)

JEL C8, R1, R2



ОРИГИНАЛЬНАЯ СТАТЬЯ

ПРОЕКТИРОВАНИЕ СТРУКТУРЫ БАЗЫ ДАННЫХ, ХАРАКТЕРИЗУЮЩИХ УРОВЕНЬ ЖИЗНИ НАСЕЛЕНИЯ РЕГИОНОВ РОССИИ И СОЦИАЛЬНО-ЭКОНОМИЧЕСКИЕ ФАКТОРЫ, ЕГО ОПРЕДЕЛЯЮЩИЕ

А.А. Дуркин, Сыктывкарский государственный университет имени Питирима Сорокина, Сыктывкар, Россия

А.Г. Шеломенцев, Югорский государственный университет, Ханты-Мансийск, Россия

К.С. Гончарова, Институт экономики Уральского отделения Российской академии наук, Екатеринбург, Россия

Аннотация. Для возможности анализа, оценки и прогнозирования социально-экономического развития территорий необходимо наличие полной и объективной информации, основу которой должна составлять обобщенная база научно обоснованных показателей, характеризующих, с одной стороны, уровень жизни населения в регионах и в стране в целом, а с другой, – основные социальные и экономические факторы, оказывающие на него значимое влияние. Соответственно, целью данной научной статьи явилась разработка структуры открытой базы данных, характеризующих процессы конвергенции и дивергенции доходов и расходов населения в регионах России, происходящие под влиянием тех или иных социальных и экономических факторов. Проектирование базы данных основывалось на применении методов формализации, алгоритмизации, структуризации и группировки. В результате была разработана структура базы, связывающая данные, характеризующие, с одной стороны, доходы и расходы населения, а с другой, – динамику социально-экономического развития субъектов Российской Федерации. Она позволила: во-первых, сформировать и систематизировать оптимальный набор показателей, охватывающий все существенные стороны изучаемого явления, интегрировав статистическую информацию, размещенную в различных информационных системах; во-вторых, обеспечить решение задачи установления зависимостей между параметрами, характеризующими сложные процессы формирования и дифференциации доходов и расходов населения, а также влияния на эти процессы социальных и экономических факторов. Научная значимость результатов исследования состоит в том, что сформированная база данных посредством своего целевого назначения, а также большей конкретизации предмета (обуславливающего набор анализируемых признаков – показателей) и детализации объекта исследования позволяет с помощью методов математической статистики, системно-динамического моделирования и др. выявлять тенденции и детерминанты дифференциации уровня жизни населения регионов страны. Практическая значимость исследования состоит в оптимизации исследовательского процесса поиска достоверных и репрезентативных для регионов России и анализируемого проблемного поля данных. Помимо этого, результаты обработки данных, хранящихся в базе, позволят определять в рассматриваемом контексте инструменты совершенствования государственной политики регионального развития Российской Федерации.

Ключевые слова: доходы и расходы населения, проектирование базы данных, пространственная дифференциация, региональное развитие, социально-экономическое развитие, уровень жизни

Благодарность: исследование выполнено при финансовой поддержке Российского научного фонда в рамках научного проекта № 22-28-01702 «Экспериментальные траектории процессов пространственной конвергенции и дивергенции доходов населения регионов России в условиях их адаптации к динамичным изменениям».

Для цитирования: Дуркин А.А., Шеломенцев А.Г., Гончарова К.С. Проектирование структуры базы данных, характеризующих уровень жизни населения регионов России и социально-экономические факторы, его определяющие // BENEFICIUM. 2023. № 1(46). С. 20-27. DOI: 10.34680/BENEFICIUM.2023.1(46).20-27

ORIGINAL PAPER

DESIGNING THE STRUCTURE OF THE DATABASE CHARACTERIZING THE STANDART OF LIVING OF THE POPULATION OF RUSSIAN REGIONS AND THE SOCIO-ECONOMIC FACTORS THAT DETERMINE IT

A.A. Durkin, Pitirim Sorokin Syktyvkar State University, Syktyvkar, Russia

A.G. Shelomentsev, Yugra State University, Khanty-Mansiysk, Russia

K.S. Goncharova, Institute of Economics of the Ural Branch of the Russian Academy of Sciences, Yekaterinburg, Russia

Abstract. To be able to analyze, assess and forecast socio-economic development of territories, it is necessary to have complete and objective information, the basis of which should be a generalized base of scientifically valid indicators, characterizing, on the one hand, the standard of living in the regions and the country as a whole, and on the other hand, the main social and economic factors that have a significant impact on it. Accordingly, the purpose of this article was to develop a structure of an open data base, characterizing the processes of convergence and divergence of income and expenditures of the population in the regions of Russia, which occur under the influence of various social and economic factors. Design of the database was based on the application of methods of formalization, algorithmization, structuring and grouping. As a result, the structure of the database was developed, it links data characterizing incomes and expenditures of the population, on the one hand, and the dynamics of socio-economic development of the subjects of the Russian Federation, on the other hand. It has enabled us, firstly, to form and systematize an optimum set of indicators covering all essential aspects of the phenomenon under study by integrating statistical information from different information systems and, secondly, to solve the problem of establishing relationships between parameters characterizing complex

processes of formation and differentiation of incomes and expenditures of the population as well as the influence of social and economic factors on these processes. Scientific significance of the results of the research consists in the fact that the formed data base through its target purpose, as well as greater specification of the subject (which determines the set of analyzed attributes – indicators) and specification of the research object allows using methods of mathematical statistics, system dynamic modeling, etc. to identify trends and determinants of differentiation of living standards of the country. The practical significance of the research consists in optimizing the research process of finding reliable and representative data for the regions of Russia and the analyzed problem. In addition, the results of processing the data stored in the database will also allow us to determine the tools for improving the state policy of regional development of the Russian Federation in the context under consideration.

Keywords: income and expenditures of the population, database design, spatial differentiation, regional development, socio-economic development, standard of living

Acknowledgement: the research was supported financially by the Russian Science Foundation within the scientific project № 22-28-01702 "Experimental trajectories of spatial convergence and divergence of incomes of the Russian regions in their adaptation to the dynamic changes".

For citation: Durkin A.A., Shelomentsev A.G., Goncharova K.S. Designing the Structure of the Database Characterizing the Standard of Living of the Population of Russian Regions and the Socio-Economic Factors That Determine It // BENEFICIUM. 2023. Vol. 1(46). Pp. 20-27. (In Russ.). DOI: 10.34680/BENEFICIUM.2023.1(46).20-27

Введение

Обеспечение социальной справедливости, с одной стороны, и поддержание экономического роста экономики, – с другой, являются важнейшими стратегическими задачами (как на региональном, так и на национальном уровнях) правительств современных государств. Значимость решения этих проблем нашла отражение в принятой в 2015 г. всеми государствами – членами ООН Повестке дня в области устойчивого развития на период до 2030 г., включающей 17 целей. В то же время, как показывает практика, социальная справедливость в текущем периоде непосредственно связана с существующим уровнем дифференциации доходов населения, находящимся под влиянием одновременно широкого круга факторов разнородной природы, без анализа и прогнозирования влияния которых решение ее невозможно.

Таким образом, в целях анализа, оценки и прогнозирования социально-экономического развития России необходима полная и объективная информация, основу которой должна составлять база данных показателей, характеризующих: во-первых, дифференциацию доходов населения в регионах и стране в целом; во-вторых, основные социальные и экономические факторы, оказывающие значимое влияние на распределение доходов населения. Это и обусловило актуальность темы настоящего исследования, связанного с разработкой такой базы данных.

Изучение и обобщение существующих баз данных, характеризующих влияние социальных и экономических факторов на дифференциацию доходов населения, позволило выделить следующие их основные виды, отличающиеся: объектами исследования; периодом, за который анализируется информация; факторами, влияние которых изучается; методами, используемыми для обработки исходной информации. Кроме того, следует отметить характер информации, которая представлена как количественными, так и качественными показателями. Создаваемые базы данных представляют научную и практическую ценность в контексте их использования при решении задач анализа, оценки и прогнозирования дифференциации доходов населения с помощью определенного инструментария, предполагающего применение методов математической статистики, математического моделирования, компаративного анализа и др.

Как показал анализ, объектами, описываемыми собираемой информацией, являются отдельные районы, страны [1-6], а также группы стран, обладающих схожими при-

знаками [7-11]. В последнем случае количество стран, доходы населения которых подвергаются исследованию, варьируется от нескольких единиц до нескольких десятков. Так, например, для анализа взаимосвязи размера социального капитала, уровня неравенства доходов и числа смертей от COVID-19 была сформирована база данных социальных показателей по 84 странам мира [11]. В России в проблемном поле настоящего исследования особую ценность представляют базы данных, содержащие информацию о доходах населения на региональном и муниципальном уровнях [12, 13].

Продолжительность периода, за который собирается информация, планируемая для включения в формирующуюся базу данных, также существенно различается: от нескольких лет, например, при анализе неравенства доходов и бремени болезней, вызванных пандемией COVID-19 [9, 14], и десятилетий – при анализе взаимосвязи между заболеваемостью населения и неравенством его доходов [10, 15] до продолжительных периодов, например, с 1950 г., – при оценке трудового потенциала в регионах России [16] и анализе тенденций изменения неравенства доходов населения в отдельных штатах США [1], с 1945 г. – при оценке влияния бюджетного федерализма на неравенство доходов с Швейцарии [6]. Кроме того можно получить доступ к базам данных, которые относятся к далекому историческому периоду, например, созданная на основе персонифицированных массовых источников база, содержащая информацию о населении г. Тамбова в 1896-1917 гг., отражающая социальную стратификацию и мобильность населения российского провинциального города в конце XIX в. – начале XX в. [17], или база данных по первичным материалам первой всеобщей переписи населения Российской империи 1897 г. [18].

Следует отметить, что также существенно варьируются количество и состав анализируемых факторов, определяющих уровень жизни населения и описываемых показателями, входящими в базы данных. В качестве основных групп факторов можно назвать: условия жизни населения [19], экологическая ситуация и антропогенное воздействие на окружающую среду территории [3, 8, 20], заболеваемость [10, 14, 15], здоровье населения [21], реализация экономических реформ в условиях переходных экономик [5], потребление энергетических ресурсов [22], обеспеченность природными ресурсами и изменение цен на них [23, 24], доверие населения [11].

Большинство сформированных к настоящему времени

баз данных, характеризующих дифференциацию доходов населения, непосредственно связано с методами анализа исходной информации, обеспечивая простоту доступа к данным и корректность их обработки. При этом используются методы математической статистики и прогнозные модели, опирающихся на показатели, характеризующие анализируемый круг факторов, в частности, интеллектуальный анализ условий жизни населения [19, 25], статистическую обработку панельных данных неравенства доходов и углеродной продуктивности [3], многомерную регрессию Пуассона при анализе взаимосвязи неравенства доходов и смертности от COVID-19 [14], метод Парето при оценке изменений в неравенстве и перераспределении доходов [5], трехуровневую иерархическую модель при исследовании влияния неравенства доходов и других социально-экономических факторов на энергетическую бедность домохозяйств [22], институциональный анализ [6], метаанализ для оценки влияния обеспеченности природными ресурсами на неравенство доходов [23], непараметрический метод оценки долгосрочного неравенства на основе смешанных исторических табличных данных [4], межрегиональное сравнение на базе ASPIRE [26] и др. Особую группу баз данных составляют географические информационные системы (ГИС), позволяющие выполнять пространственный анализ данных [27].

Федеральной службой по интеллектуальной собственности (Роспатент) зарегистрировано несколько баз данных, представляющих научно-прикладной интерес для достижения цели настоящего исследования и характеризующих:

- 1) во-первых, дифференциацию доходов населения в составе более широкого круга показателей качества жизни: «Мониторинг доходов и уровня жизни населения России» (более 85 показателей анализа социально-экономического неравенства, доходов, уровня и качества жизни населения Российской Федерации, федеральных округов, субъектов Российской Федерации)¹; «Информационная база данных по политике доходов и заработной плате» (данные о структуре денежных доходов населения и их использовании в Российской Федерации (1970-2018 гг.) и Республике Саха (Якутия) (1980-2017 гг.), о доходах населения и их дифференциации в Российской Федерации и Республике Саха (Якутия) (2000-2017 гг.))²; «Комплексное обследование занятости, доходов и уровня жизни населения – 2013» (данные о доходах, имущественном положении, состоянии здоровья, материальном и социальном положении, образовательных и трудовых стратегиях различных групп населения России)³; др.;
- 2) во-вторых, факторы, влияющие на доходы насе-

ния: «Система немонетарных показателей уровня жизни населения в Российской Федерации в 2019 г.» (характеристики обеспеченности жилищными условиями и товарами длительного пользования групп домашних хозяйств, сформированных по уровню среднедушевых денежных доходов)⁴; «Информационная база данных по анализу демографических процессов и уровня жизни населения Республики Саха (Якутия)» (данные об уровне жизни населения в разрезе различных типов семей в зависимости от наличия детей и их количества)⁵; «Наиболее значимые показатели, характеризующие состояние муниципальных бюджетов, уровень жизни и уровень доходов населения городских округов АЗ Российской Федерации по основным социальным направлениям (образование, здравоохранение)» (социальные расходы бюджета и уровень жизни населения 23 городских округов арктической зоны, представляющих 7 регионов России)⁶; др.

Таким образом, обзор российской практики формирования баз данных, содержащих информацию для анализа доходов населения и их дифференциации, позволил сделать следующие выводы. Во-первых, можно выделить две основные концепции построения структуры таких баз данных: 1) базы данных, включающие информацию только о доходах населения; 2) базы данных, включающие информацию как о доходах населения, так и о факторах, влияющих на их дифференциацию. Во-вторых, информация баз данных ограничена достаточно узким кругом факторов, влияющих на дифференциацию доходов населения. В-третьих, формирование таких баз данных является основой анализа, оценки, прогнозирования и государственного регулирования социальных и экономических процессов.

Исходя из вышеизложенного, целью данной работы явилась разработка структуры открытой базы данных, характеризующих процессы конвергенции и дивергенции доходов населения в регионах России, происходящие под влиянием социальных и экономических факторов.

В отличие от большинства эксплуатируемых в настоящее время баз данных, проектируемая авторами база данных интегрирует широкий перечень показателей, подобранных и обоснованных в соответствии с тематикой проводимого исследования (выявление ключевых социально-экономических факторов, оказывающих значимое влияние на уровень жизни регионов России), собранных из различных источников и имеющих пространственную детализацию – муниципальное образование (зачастую в базах данных содержится информация, агрегированная по субъек-

¹ Свидетельство о государственной регистрации базы данных № 2021621841 Российская Федерация. Мониторинг доходов и уровня жизни населения России: № 2021621734: заявл. 24.08.2021: опублик. 01.09.2021 / В.Н. Бобков, А.А. Гулюгина, Е.В. Одинцова [и др.]; заявитель ФГБУН «Федеральный научно-исследовательский социологический центр Российской академии наук».

² Свидетельство о государственной регистрации базы данных № 2019622477 Российская Федерация. Информационная база данных по политике доходов и заработной плате: № 2019622338: заявл. 07.12.2019: опублик. 25.12.2019 / К.Ю. Постникова; заявитель ФГАОВ ВО «Северо-Восточный федеральный университет имени М.К. Аммосова».

³ Свидетельство о государственной регистрации базы данных № 2018621279 Российская Федерация. Комплексное обследование занятости, доходов и уровня жизни населения – 2013: № 2018620848: заявл. 22.06.2018: опублик. 14.08.2018 / Т.М. Малева, Е.М. Аврамова, Д.М. Логинов [и др.]; заявитель ФГБОУ ВО «Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации».

⁴ Свидетельство о государственной регистрации базы данных № 2022620291 Российская Федерация. Система немонетарных показателей уровня жизни населения РФ в 2019 г.: № 2022620090: заявл. 18.01.2022: опублик. 04.02.2022 / А.С. Аброскин, Н.А. Аброскина; заявитель ФГБОУ ВО «Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации».

⁵ Свидетельство о государственной регистрации базы данных № 2019622497 Российская Федерация. Информационная база данных по анализу демографических процессов и уровня жизни населения Республики Саха (Якутия): № 2019622440: заявл. 18.12.2019: опублик. 26.12.2019 / С.А. Сукнева, К.Ю. Постникова; заявитель ФГАОВ ВО «Северо-Восточный федеральный университет имени М.К. Аммосова».

⁶ Свидетельство о государственной регистрации базы данных № 2018621190 Российская Федерация. Наиболее значимые показатели, характеризующие состояние муниципальных бюджетов, уровень жизни и уровень доходов населения городских округов АЗ РФ по основным социальным направлениям (образование, здравоохранение): № 2018620900: заявл. 29.06.2018: опублик. 06.08.2018 / А.Н. Чапарина, Е.Е. Емельянова; заявитель ФГБУН Федеральный исследовательский центр «Кольский научный центр Российской академии наук» (ФИЦ КНЦ РАН).

там Российской Федерации, что не позволяет учитывать уровень внутрирегиональной социально-экономической дифференциации). Кроме того, разрабатываемая база данных ориентирована на создание так называемого «дружественного интерфейса», позволяющего работать с ней исследователям с различным уровнем владения персональным компьютером (в том числе без знания языков программирования для составления соответствующих запросов).

Проектирование базы данных представляет собой сложный процесс, заключающийся в переходе от неформального описания рассматриваемой задачи к формальному терминологическому представлению объектов предметной области и их зависимостей. При этом полученная база данных должна обеспечивать хранение и возможность получения данных по запросам, сокращение их избыточности, целостности.

Выделяют следующие этапы процесса проектирования базы данных:

- 1) системный анализ предметной области;
- 2) инфологическое проектирование;
- 3) выбор системы управления базой данных (СУБД);
- 4) даталогическое проектирование;
- 5) физическое проектирование.

На первом, важнейшем, этапе производится подробное исследование предметной области, определяются основные объекты, свойства которых будут использоваться при разработке базы данных, и существующие между ними связи.

Этап инфологического (концептуального) проектирования заключается в создании логической структуры данных. Все требования к базе и объекты предметной области формализуются и записываются на языке некоторой семантической модели. Одним из наиболее удобных и часто используемых инструментов унифицированного представления данных является предложенная в 1976 г. П. Ченом (P. Chen) модель «сущность-связь» (Entity Relationship, ER-модель). Она позволяет выделить ключевые сущности и обозначить связи, которые могут устанавливаться между этими сущностями; при этом является наиболее общей, позволяя в дальнейшем перейти к прочим моделям данных (иерархическая, сетевая, реляционная, объектная). При определении связей между сущностями следует придерживаться трех типов: «один-к-одному (1 : 1)», «один-ко-многим (1 : M)», «многие-ко-многим (M : M)». На этом же этапе для сущностей выбирается ключ, под которым подразумевается элемент данных, уникально идентифицирующий отдельный экземпляр некоторой сущности.

Выбор СУБД осуществляется на основе многих параметров, отражающих требования к проектируемой базе данных, а также возможности и ограничения конкретной СУБД: стоимость использования, удобство и надежность, функциональные возможности, др.

На этапе даталогического (логического) проектирования разрабатывается совокупность схем отношений, которые описывают объекты предметной области и связи между ними. При этом каждая из определенных в концептуальной схеме сущностей отображается в таблицу, которая является одним отношением. Здесь следует учитывать ограничения, накладываемые выбранной СУБД. На этом же этапе необходимо разработать набор правил, позволяющих обеспечить в любой момент времени релевантность данных (непротиворечивость, удовлетворяемость, правдоподобность).

Физическое проектирование заключается в реализации

логической модели в выбранной СУБД и системе хранения. На этом этапе решаются вопросы физического размещения данных, обеспечения доступа к ним, безопасности и сохранности.

Для формирования базы данных доходов и расходов населения и факторов, их определяющих, использованы результаты обследований, проводимых Федеральной службой государственной статистики (Росстат) и представленные на ее официальном сайте (rosstat.gov.ru), а также открытые первичные информационно-аналитические ресурсы профильных министерств и ведомств, научно-исследовательских институтов, национальных исследовательских университетов, др. Так, важнейшим источником информации относительно уровня и условий жизни различных групп населения являются результаты выборочного обследования бюджетов домашних хозяйств. Оно ежеквартально проводится органами государственной статистики во всех субъектах Российской Федерации по выборочному методу, охватывает около 49 тыс. домохозяйств и строится на принципе их добровольного участия. Генеральную совокупность при отборе составляют все типы домашних хозяйств, кроме коллективных. Обследование по своему характеру является многоцелевым, и его данные служат основой для анализа широкого круга социальных, экономических и иных вопросов.

Результаты и их обсуждение

Проектируемая база данных призвана сделать возможными анализ и оценку процессов конвергенции и дивергенции пространственной дифференциации доходов населения, а также определение комплекса факторов, оказывающих влияние на эти процессы.

Исходные данные, с одной стороны, описывают структуру доходов и расходов домашних хозяйств с их распределением по регионам России, с другой, – основные социально-экономические показатели развития регионов. Таким образом, наиболее крупным объектом, рассматриваемым на концептуальном этапе проектирования базы данных, является «Регион». Далее для каждого региона исследуются размещенные на его территории домохозяйства, которые, как и категории их показателей, также входят в структуру объектов базы данных. Помимо этого, для каждого региона рассматриваются основные факторы, которые могут оказывать влияние на пространственную дифференциацию доходов его населения (рис. 1).



Рис. 1. Структура объектов базы данных / Fig. 1. Structure of database objects

Источник: составлено авторами / Source: compiled by the authors

В качестве даталогической модели базы данных выбрана реляционная модель. Ее основное отличие от иных моделей в том, что все данные организованы в

виде таблиц, а все операции с ними (данными) представляют собой операции с этими таблицами. Реляционная модель обеспечивает наглядность представления данных и удобство работы с ними. Кроме того, модель в настоящее время наиболее распространена и фактически является стандартом, по которому строятся базы

данных; к тому же на эту модель ориентированно большинство современных СУБД.

Для изображения логической структуры разрабатываемой базы данных построена диаграмма в нотации Мартина (Crow's Foot) (рис. 2).



Рис. 2. Логическая схема базы данных / Fig. 2. Logic circuit of the database

Условные обозначения: ВК – внешний ключ (Foreign Key), ПК – первичный ключ (Primary Key)

Источник: составлено авторами / Source: compiled by the authors

Центральной таблицей в представленной структуре базы данных является таблица «Регион». В нее вносятся все субъекты Российской Федерации с учетом сложности административно-территориального устройства некоторых из них. Так, например, для Тюменской области в таблице присутствуют отдельные строки для области в ее общих границах, а также отдельно для Ханты-Мансийского автономного округа, Ямало-Ненецкого автономного округа и Тюменской области без автономных округов.

Перейти к информации по обследованию бюджетов домашних хозяйств регионов позволяет таблица «Домохозяйство». Каждая запись в последней представляет набор данных, включающий период наблюдения, код территории, на которой размещается домохозяйство, тип местности (городская или сельская) и номер бюджета, идентифицирующий наблюдаемый объект на данной территории в данный период времени.

С каждым уникальным домохозяйством связаны наблюдения в нескольких таблицах; их наименование и содержание (фрагмент) приведены в табл. 1.

Таблица 1 / Table 1

Содержание таблиц базы данных (фрагмент) / Contents of Database Tables (Excerpt)

Название таблицы / Table Name	Содержание таблицы / Table Contents	Содержание группы показателей / Contents of the Indicator Group
FC	Основные расчетные показатели	Доходы, расходы, располагаемые ресурсы домохозяйства
FL	Данные о членах домохозяйств	Тип населенного пункта, в котором расположено домохозяйство, размер домохозяйства, семейное положение членов домохозяйства, статус основной занятости членов домохозяйства, льготы членов домохозяйства (льготная категория, полученные льготы и др.) и проч.
FO	Данные об уровне образования членов домохозяйств	Уровень достигнутого образования, уровень получаемого образования (вид образования, тип учебного заведения и проч.)
FG	Данные о жилищных условиях домохозяйств	Жилая площадь, число комнат, принадлежность (форма собственности) жилого помещения, тип жилого помещения, наличие элементов инженерной инфраструктуры, общая и специальная доступность и проч.
FN	Данные о доходах домохозяйств	Денежный доход на одного члена домохозяйства, структура доходов по источникам (трудовая и предпринимательская деятельность, трансферты, проч.), виды социальных трансфертов домохозяйства
FH	Данные о расходах домохозяйств на жилье	Расходы на оплату жилищных и коммунальных услуг, перечень расходов по платежному документу и проч.
FQ	Данные о расходах домохозяйств по направлениям потребления и стоимости натуральных поступлений продуктов питания и непродовольственных товаров и услуг	Стоимость всех видов потребленных продовольственных и непродовольственных товаров, поступивших в домохозяйство без оплаты
FZ	Данные о предметах длительного пользования домохозяйств	Наличие и количество в домохозяйстве предметов длительного пользования

Источник: составлено авторами с использованием шифров исследования Росстата [28] / Source: compiled by the authors based on Rosstat survey codes [28]

Следует отметить, что часть таблиц, описывающих показатели домохозяйств, кроме связи с вышестоящим объектом имеют и связи между собой. Так, в таблице «FL» каждому члену домохозяйства присваивается уникальный номер. На этот номер ссылаются таблицы «FO» (информация об образовании) и «FN» (информация о доходах). Также учитывается, что, начиная с 2016 г. (в связи с изменением методики формирования базы данных исследования), сведения по двум указанным таблицам не содержат прямого указания на лицо, описывая общее состояние показателя в домохозяйстве. Исходя из этого, атрибут «Код лица» не является обязательным.

В каждой таблице представлен широкий перечень показателей, подробно раскрывающих ее содержание (см. табл. 1, на рис. 2 они объединены под именем «[Показатели]»).

Помимо данных по обследованию домохозяйств для каждого субъекта Российской Федерации, представленного в блоке «Регион», приведен перечень основных показателей его социально-экономического развития. Аргументы, перечисленные на рис. 2 в таблице «Факторы», представляют собой обобщенное описание основных характеристик. Так, например, аргумент «[Труд]» следует понимать как набор показателей, включающий среднегодовую численность занятых и безработных, среднегодовую численность официально зарегистрированных безработных, среднегодовую численность работников организаций и государственных служащих, социально-демографический состав занятых, среднемесячную номинальную заработную плату, выпуск квалифицированных рабочих, служащих, специалистов, бакалавров и др., индекс производительности труда и проч.

При проектировании логической структуры базы данных также учитывались требования нормализации данных. В связи с необходимостью проведения в дальнейшем исследовательских вычислений с данными таблиц обеспече-

но, чтобы ячейки не содержали списков, а только отдельные значения, что соответствует первой нормальной форме (1NF). Также исключено дублирование данных и уменьшен объем таблиц путем вынесения информации, идентифицирующей домохозяйство, в отдельную таблицу.

Полученная структура связывает домохозяйства и их основные характеристики с факторами социально-экономического развития регионов, позволяя определить влияние последних на пространственную дифференциацию доходов населения. При необходимости рассмотрения дополнительных показателей полученная структура базы данных позволяет легко добавить новые данные и связать их с существующими.

Заключение

Спроектированная авторами структура базы данных, характеризующих уровень жизни населения регионов России и социально-экономические факторы, его определяющие, позволяет обобщить и систематизировать статистические количественные и качественные показатели по ключевым группам, интегрируя информацию, представленную в настоящее время в различных информационных системах. Состав данных обуславливает применение междисциплинарного подхода к сбору и первичной обработке информации, а также последующему выбору методов ее анализа. Это позволяет ставить и решать задачи установления зависимостей между параметрами, характеризующими сложные процессы дифференциации доходов населения и социальными и экономическими факторами, влияющими на их динамику. Сформированная база данных создает широкие возможности для применения методов математической статистики, системно-динамического моделирования и других методов оценки динамики показателей уровня жизни населения, а также прогнозирования воздействия на них широкого круга факторов.

Вклад авторов

Концептуализация, в том числе определение целей и задач – А.Г. Шеломенцев, К.С. Гончарова; определение методов исследования – А.Г. Шеломенцев, К.С. Гончарова, А.А. Дуркин; сбор и анализ данных, а также курирование метаданными – А.Г. Шеломенцев, К.С. Гончарова, А.А. Дуркин; проектирование элементов программного обеспечения – А.А. Дуркин; создание визуализации – А.А. Дуркин; оформление ключевых результатов исследования, формулирование и научное обоснование выводов – А.Г. Шеломенцев, К.С. Гончарова, А.А. Дуркин.

Библиография / References

- [1] Wallace M., Hyde A., Vachon T. States of inequality: Politics, labor, and rising income inequality in the U.S. States since 1950 // *Research in Social Stratification and Mobility*. 2022. Vol. 78. P. 100677. DOI: 10.1016/j.rssm.2022.100677
- [2] Hu Zh. The effect of income inequality on human capital inequality: Evidence from China // *Structural Change and Economic Dynamics*. 2021. Vol. 58(4). Pp. 471-489. DOI: 10.1016/j.strueco.2021.06.015
- [3] Du M., He L., Zhao M. [et al.]. Examining the relations of income inequality and carbon productivity: A panel data analysis // *Sustainable Production and Consumption*. 2022. Vol. 31(14). Pp. 249-262. DOI: 10.1016/j.spc.2022.01.027
- [4] Aaberge R., Atkinson A.B., Modalsli J. Estimating long-run income inequality from mixed tabular data: Empirical evidence from Norway, 1875-2017 // *Journal of Public Economics*. 2020. Vol. 187(4). P. 10496. DOI: 10.1016/j.jpubeco.2020.104196
- [5] Brzezinski M., Myck M., Najsztub M. Sharing the gains of transition: Evaluating changes in income inequality and redistribution in Poland using combined survey and tax return data // *European Journal of Political Economy*. 2022. Vol. 73(3). P. 102121. DOI: 10.1016/j.ejpoleco.2021.102121
- [6] Feld L.P., Frey C., Schaltegger C.A., Schmid L.A. Fiscal federalism and income inequality: An empirical analysis for Switzerland // *Journal of Economic Behavior and Organization*. 2021. Vol. 185. Pp. 463-494. DOI: 10.1016/j.jebo.2021.02.028
- [7] Gorshkov A.A. «Great discrepancy»: the growth of inequality in household income distribution and redistribution in the countries of the Organization for Economic Co-Operation and Development (OECD) (Russia, Moscow) // *Problems of Modern Economics*. 2016. Vol. 1(57). Pp. 47-52. (In Russ.).
- [8] Wang Q., Yang T., Li R. Does income inequality reshape the environmental Kuznets curve (EKC) hypothesis? A nonlinear panel data analysis // *Environmental Re-search*. 2023. Vol. 216. P. 114575. DOI: 10.1016/j.envres.2022.114575
- [9] Su D., Alshehri K., Pagán J. Income inequality and the disease burden of COVID-19: Survival analysis of data from 74 countries // *Preventive Medicine Reports*. 2022. Vol. 27. P. 101828. DOI: 10.1016/j.pmedr.2022.101828
- [10] Anindya K., Marthias T., Vellakkal S. [et al.]. Socioeconomic inequalities in effective service coverage for reproductive, maternal, newborn, and child health: a comparative analysis of 39 low-income and middle-income countries // *EclinicalMedicine*. 2021. Vol. 40. P. 101103. DOI: 10.1016/j.eclinm.2021.101103
- [11] Elgar F.J., Stefaniak A., Wohl M.J.A. The trouble with trust: Time-series analysis of social capital, income inequality, and COVID-19 deaths in 84 countries // *Social Science and Medicine*. 2020. P. 113365. DOI: 10.1016/j.socscimed.2020.113365
- [12] Bogomolov E.V. Russian household financial behavior characteristics // *Ekonomika. Nalogi. Pravo* [Economics, Taxes & Law]. 2020. Vol. 13(1). Pp. 49-59. (In Russ.). DOI: 10.26794/1999-849X-2020-13-1-49-59
- [13] Bychkova S.G. Statistical analysis of income differentiation: international and regional comparisons // *Vestnik Universiteta*. 2015. Vol. 5. Pp. 131-139. (In Russ.).
- [14] Sepulveda E.R., Brooker A.S. Income inequality and COVID-19 mortality: Age-stratified analysis of 22 OECD countries // *SSM – Population Health*. 2021. Vol. 16. Pp. 1-6. DOI: 10.1016/j.ssmph.2021.100904
- [15] Abila D.B., Wasukira S.B., Ainembabazi P. [et al.]. Socioeconomic inequalities in prostate cancer screening in low- and middle-income countries: an analysis of the demographic and health surveys between 2010 and 2019 // *Journal of Cancer Policy*. 2022. Vol. 34. DOI: 10.1016/j.jcpcp.2022.100360
- [16] Popov A.V. Use of information and analytical system of monitoring labor potential of territories for regulating labor behavior of population // *Vestnik NSUEM*. 2016. Vol. 1. Pp. 189-199. (In Russ.).
- [17] Strekalova N.V. Databases on the problems of studying social stratification and mobility of urban elites and the middle strata of Tambov in the late 19th – early 20th centuries: sources, structure, technology of complex analysis // *Istoricheskaya Informatika* [Historical Informatics]. 2021. Vol. 4(38). Pp. 50-66. (In Russ.). DOI: 10.7256/2585-7797.2021.4.36987
- [18] Bryukhanova E.A., Nezhentseva N.V., Chekryzhova O.I., Ivanov D.N. A database built on the first 1897 Russian Empire population census data: structure and analysis // *Istoricheskaya Informatika* [Historical Informatics]. 2020. Vol. 1(31). Pp. 20-33. (In Russ.). DOI: 10.7256/2585-7797.2020.1.32387
- [19] Seifert J.W. Data mining and the search for security: Challenges for connecting the dots and databases // *Government Information Quarterly*. 2004. Vol. 21(4). Pp. 461-480. DOI: 10.1016/j.giq.2004.08.006
- [20] Qiao K., Dowell G. Environmental concerns, income inequality, and purchase of environmentally-friendly products: A longitudinal study of U.S. counties (2010-2017) // *Research Policy*. 2022. Vol. 51(4). P. 104443. DOI: 10.1016/j.respol.2021.104443
- [21] Bogomolova T.Yu., Tapilina V.S. Financial behavior of household in Russia in the mid-90s // *Economics of Contemporary Russia*. 1998. Vol. 4. Pp. 58-69. (In Russ.).
- [22] Igawa M., Managi S. Energy poverty and income inequality: An economic analysis of 37 countries // *Applied Energy*. 2022. Vol. 306(B). P. 118076. DOI: 10.1016/j.apenergy.2021.118076
- [23] Seabri M., Dachraoui H. Natural resources and income inequality: A meta-analytic review // *Resources Policy*. 2021. Vol. 74. P. 102315. DOI: 10.1016/j.resourpol.2021.102315
- [24] Mohtadi S., Castells-Quintana D. The distributional dimension of the resource curse: Commodity price shocks and income inequality // *Structural Change and Economic Dynamics*. 2021. Vol. 59(1). Pp. 63-78. DOI: 10.1016/j.strueco.2021.08.002
- [25] Lee J.S., Jun S.P. Privacy-preserving data mining for open government data from heterogeneous sources // *Government Information Quarterly*. 2021. Vol. 38(1). P. 101544. DOI: 10.1016/j.giq.2020.101544
- [26] Litvinova V.V., Nagernyak M.A., Kirillova M.N. The Atlas of social protection indicators of resilience and equity: opportunities for interregional comparisons // *Financial Journal*. 2017. Vol. 5(39). Pp. 33-46. (In Russ.).
- [27] Dyshlyuk S.S., Utrobina E.S. Creating and loading data base for statistical accounting of regional economic characteristics in the reference-analytical GIS tool // *Interexpo Geo-Siberia*. 2012. Vol. 1(1-2). Pp. 111-116. (In Russ.).
- [28] Sample Survey of Household Income and Expenditures (2021). ROSSTAT. (In Russ.). URL: <https://obdx.gks.ru/> (accessed on 22.12.2022).

Информация об авторах / About the Authors

Анатолий Альбертович Дуркин – аспирант, Сыктывкарский государственный университет имени Питирима Сорокина, Сыктывкар, Россия / **Anatoliy A. Durkin** – Graduate Student, Pitirim Sorokin Syktyvkar State University, Syktyvkar, Russia
E-mail: anatoliy.durkin@mail.ru

Андрей Геннадьевич Шеломенцев – д-р экон. наук, профессор; профессор, Югорский государственный университет, Ханты-Мансийск, Россия / **Andrei G. Shelomentsev** – Dr. Sci. (Economics), Professor; Professor, Yugra State University, Khanty-Mansiysk, Russia

E-mail: a.shelom@gmail.ru

SPIN РИНЦ 9931-5778

ORCID 0000-0003-1904-9587

Scopus Author 56288580900

Researcher ID M-8644-2016

Ксения Сергеевна Гончарова – канд. экон. наук; научный сотрудник, Институт экономики Уральского отделения РАН, Екатеринбург, Россия / **Ksenia S. Goncharova** – Cand. Sci. (Economics); research associate, Institute of Economics of the Ural Branch of the Russian Academy of Sciences, Yekaterinburg, Russia

E-mail: kseniya.gon4arowa@gmail.com

SPIN РИНЦ 6154-9862

ORCID 0000-0003-2381-3322

Scopus Author 57219669945

Researcher ID K-4838-2018

Дата поступления статьи: 29 декабря 2022
Принято решение о публикации: 28 марта 2023

Received: December 29, 2022
Accepted: March 28, 2023