

ФОРМИРОВАНИЕ ЭЛЕКТРОННОЙ ДЕМОГРАФИИ КАК ЭФФЕКТИВНОГО ИНСТРУМЕНТА СОЦИАЛЬНЫХ ИССЛЕДОВАНИЙ И МОНИТОРИНГА ДАННЫХ О НАСЕЛЕНИИ¹

Алгулиев Р., Алыгулиев Р., Юсифов Ф., Алекперова И.²

Аннотация

Статья посвящена формированию электронной демографической (э-демографической) системы на основе единого государственного реестра для проведения демографических исследований. В эпоху цифровых технологий возникают новые источники данных для исследования демографического поведения. Данные реестра населения, используемые для демографических исследований, сегодня все же достаточно ограничены. Поэтому создание основанной на едином государственном реестре э-демографической системы для эффективных демографических исследований, а также анализа не только реестра населения, но и других государственных реестров, представляет актуальность. В работе изучается международный опыт использования демографических характеристик данных, накопленных в реестре населения. Практика показывает, что, хотя данные реестра населения в настоящее время применяются в демографических исследованиях, подходов к использованию данных других государственных реестров и созданию э-демографической системы не существует. Авторами предлагается концепция построения э-демографической системы на основе единого государ-

¹ Данная работа выполнена при финансовой поддержке Фонда Развития Науки при Президенте Азербайджанской Республики – Грант № EiF-BGM-4-RFTF-1/2017-21/08/1

² Алгулиев Расим Магамед оглу – академик НАНА, доктор технических наук, директор Института информационных технологий, Национальная академия наук Азербайджана (ИИТ НАНА). Адрес: Азербайджанская Республика, AZ1141, Баку, ул. Б. Вахабзаде, 9А. E-mail: director@iit.science.az

Алыгулиев Рамиз Магамед оглу – член-корреспондент НАНА, доктор технических наук, заведующий отделом, Институт информационных технологий, Национальная академия наук Азербайджана (ИИТ НАНА). Адрес: Азербайджанская Республика, AZ1141, Баку, ул. Б. Вахабзаде, 9А. E-mail: r.aliguliyev@gmail.com

Юсифов Фархад Фирудин оглу – кандидат технических наук, заведующий отделом, Институт информационных технологий, Национальная академия наук Азербайджана (ИИТ НАНА). Адрес: Азербайджанская Республика, AZ1141, Баку, ул. Б. Вахабзаде, 9А. E-mail: farhadyusifov@gmail.com

Алекперова Ирада Явер гызы – кандидат технических наук, главный специалист, Институт информационных технологий, Национальная академия наук Азербайджана (ИИТ НАНА). Адрес: Азербайджанская Республика, AZ1141, Баку, ул. Б. Вахабзаде, 9А. E-mail: airada.09@gmail.com

ственного реестра, действующей независимо от функций э-государства. Представлены преимущества технологий Big Data, OLAP, ETL для анализа демографических данных. Система позволяет лучше понять демографические процессы и способствует проведению более глубоких исследований на разных уровнях – от социальной демографии до географии населения и от миграционных процессов до медицинской демографии.

Ключевые слова: электронное государство; электронная демография; демографические показатели; миграция; электронные реестры; реестр населения.

Введение

Распространение цифровых технологий и стремительный рост числа пользователей интернета привели к формированию и сбору больших объемов демографических данных. Новые источники информации и большие данные представляют важность с точки зрения анализа демографических характеристик и предоставляют широкие возможности. Анализ литературы показывает, что в последние годы проводились многочисленные исследования, важные для выявления демографического поведения, характеристик и изучения новых источников информации (Billari et al. 2013; Poulain & Herm, 2013; Zagheni & Weber, 2015; Yildiz et al. 2017; Zagheni, 2017; Cesare et al. 2018; Gil-Clavel & Zagheni, 2019).

Ряд публикаций посвящен вопросам использования данных, накопленных в реестре населения, в качестве нового источника данных для демографических исследований и статистики (см., например: Андриченко, Мещерякова, 2012; Prins, 2017; Careja & Bevelander, 2018). В целях ведения демографической статистики национальные статистические управление осуществляли ряд инициатив по трансферу демографических данных из реестров населения (Poulain & Herm, 2013; Prins, 2017; Careja & Bevelander, 2018). Хотя история реестров населения восходит к Древнему Китаю, их создание в централизованном порядке началось со второй половины XX в. Реестр населения менее известен и задокументирован, чем другая демографическая информация (акты гражданского состояния, перепись населения, опросы и т.д.). Если обратить внимание на мировую практику, то реестры населения распространены во многих развитых европейских странах (Poulain & Herm, 2013; Prins, 2017; Alburez-Gutierrez et al. 2019). Данные реестра населения в статистических исследованиях имеют большое значение для оценки численности и социально-демографической структуры населения в определенный момент времени, а также оценки и анализа изменений численности населения и его отдельных показателей.

В настоящее время постепенное внедрение и расширение возможностей этих реестров полностью изменило способ использования демографической статистики. В целом, можно отметить, что эффективность реестров населения увеличивается из-за постоянного доступа пользователей к базе данных, и создаются условия для устранения возможных ошибок.

Действующие каналы обратной связи в отношениях между гражданами и государством помогут минимизировать ошибки в будущем.

Несмотря на то, что частично расширены возможности для демографического анализа и реестр предоставляет биографические данные о жизненном цикле человека и событиях, эти данные весьма ограничены. С этой точки зрения анализ других государственных регистраторов, а не только населения, т.е. создание э-демографической системы, является актуальной проблемой. Формирование э-демографической системы позволяет лучше понять демографические процессы, проводить более глубокие исследования на разных уровнях и в разных разделах – от социальной демографии до географии населения и медицинской демографии. Хотя данные, собранные в отдельных реестрах, являются структурированными, для демографических исследований в различных срезах очень важен анализ больших данных.

Практика показывает: несмотря на то, что данные демографических исследований в настоящее время применяются в реестре населения, единой системы использования данных из других государственных реестров нет. В настоящей статье авторы предлагают разработанные ими концептуальные основы формирования такой единой э-демографической системы. Подход основан на создании э-демографической системы на платформе электронного государства (э-государства). Важно учитывать, что могут быть предложены разные модели э-демографии в зависимости от опыта стран, структуры правительства, национальных приоритетов, стратегии построения э-государства и интеграции электронных систем.

В международной практике представлены две модели э-государства; в некоторой литературе их называют западной и восточной моделями (Черемных, Яковлев, 2017). В первой, западной, модели сначала создается платформа э-государства, а все государственные реестры постепенно интегрируются в нее. Эта модель распространена в странах Западной Европы, США, Южной Корее, Эстонии и др. Во второй модели – восточной – отдельные государственные реестры интегрируются, функционируют или переносятся на портал э-государства. Этую модель применяет ряд стран СНГ, в том числе и Азербайджан, развивающиеся страны.

Опыт показывает, что создание второй модели с использованием отдельных платформ и различного программного обеспечения вызывает проблемы: например, проблемы интеграции, безопасности, невозможности использования личных идентификационных номеров для всех реестров. Принимая во внимание опыт работы первой модели, авторы статьи предлагают концептуальную модель э-демографической системы на основе единого государственного реестра. Они полагают, что, когда речь идет о платформе э-государства, в системе следует учитывать все данные, собранные в едином реестре граждан, даже те, которые могут считаться незначительными. Описанная концептуальная модель основана на анализе данных государственного реестра и не связана напрямую с функциями э-государства.

Авторы, рассматривая тему создания э-демографической системы для проведения демографических исследований на платформе э-государства, провели сравнительный анализ международного опыта в области демографических исследований, изучили вопросы интеграции государственных реестров и подходы к использованию новых источников информации.

Для изучения демографических характеристик исследуются структурные принципы применения данных реестра населения, а также системы э-демографии. Показаны возможности применения методов анализа больших данных и технологий OLAP, ETL в демографических исследованиях.

Демографические исследования в цифровой век

Исторически термин «демография» был впервые упомянут в XIX в. и означает науку о народонаселении. Демография как наука тесно связана с другими научными дисциплинами; пересечение этих наук рождает такие направления, как историческая демография, экономическая демография, социальная демография, этническая демография, география населения, медицинская демография, демография семьи, образовательная демография, военная демография, демография миграции и т.д.

Отношения между демографией и другими социальными науками позволяют анализировать демографические процессы, понимать причины и прогнозировать развитие общества. Демография исследует закономерности событий и процессов на основе социальных, экономических, культурных, биологических и географических проблем, возникающих в структуре, местоположении, миграции и динамике населения. Другими словами, демография – это междисциплинарная область исследований, изучающая зависимость возрастной структуры, национального и этнического составов, географического положения населения, числа граждан, миграций, количества рождений и смертей от социально-экономических, исторических и других факторов (Борисов, 2001; Mehr, 2017). Связь демографии с другими науками позволила провести более глубокий анализ демографических процессов, их причин и прогнозирование тенденций.

На протяжении многих лет для получения информации о населении, его демографических, экономических и социальных характеристиках проводили переписи (Shryock et al., 1980). Но информация, собранная при переписи населения, имела статистическую значимость только на конкретную дату или период. Многие факторы, такие как миграция, в определенное время не учитывались.

Сегодня развитие цифровых технологий, глобальное распространение интернета, появление интернета вещей, устройств слежения, умных часов и планшетов, популярность социальных сетей радикально изменили образ жизни людей, открыли новые источники персональных данных и новые возможности для решения социальных и экономических вопросов. Научное направление, которое мы можем назвать наукой данных или

цифровой демографией, – это инновационный инструмент для всех наук, дающий возможность обмена знаниями с другими областями науки в сфере больших данных (Zagheni & Weber, 2015; Zagheni, 2017).

Э-демография является ядром э-государства. Она позволяет изучать влияние цифровых технологий на демографическое поведение и привлекать новые источники данных для более тщательного изучения демографических процессов. Следует отметить, что э-демография имеет потенциал для исследований во многих областях: социальная демография, география населения и т.д. С помощью э-демографии можно исследовать различные важные вопросы, например, как социальные сети могут отражаться на поведении людей, как новые технологии воздействуют на отношения между поколениями или как онлайн-знакомства влияют на заключение брака и т.д. (Zagheni & Weber, 2015).

Международный опыт по формированию э-демографии

Развитие интернета и цифровая революция радикально изменили практику исследований. Теперь каждый может собирать данные в разных формах и даже пользоваться приложениями мобильных телефонов для контроля состояния здоровья. Этот процесс, с одной стороны, полностью децентрализован, но с другой – привлечение новых источников, централизованный сбор, хранение и обработка персональных данных должны осуществляться через э-демографическую систему.

Вики-технологии в демографических исследованиях. Точно так же платформа «Википедия» и многие проекты, созданные по технологии вики и основанные на данных профессиональных ученых, служат для получения знаний и являются революционной формой массового сотрудничества или коллективного интеллекта. Вики-технологии позволяют свободно размещать информацию о себе и своих близких на веб-странице, а также хранить информацию о том, кто и когда редактировал информацию. Эта функция помогает отслеживать демографические изменения с течением времени. Например, проводились исследования пожизненных отношений между родителями и детьми, а также супругами, в которых использовались данные, собранные на веб-сайте WikiTree (см.: Fire & Elovici, 2015).

Роль цифровых данных в демографических исследованиях. В литературе существуют различные подходы к применению цифровых данных для исследования демографических характеристик. Например, использование интернет-данных для исследования потенциала рождаемости все еще находится на начальной стадии, но считается, что в ближайшем будущем будут достигнуты обнадеживающие результаты. Работы в этой области проводятся путем изучения систем опросов и демографического поведения, а также посредством мониторинга (Zagheni & Weber, 2015; Zagheni, 2017).

Директор Института демографических исследований имени Макса Планка (*Max Planck Institute for Demographic Research*) Е. Загхени (Zagheni, 2018), характеризуя э-демографию, отметил: «Динамика численности на-

селения, связанная с миграцией, старением и отношениями между поколениями, происходит на фоне так называемой цифровой революции. Глобальное распространение интернета, социальных сетей, смартфонов и всевозможных устройств слежения, а также увеличение вычислительной мощности революционизируют как нашу жизнь, так и социальные науки» (Zagheni, 2018, веб-стр.). Он считает, что э-демография, используя все персональные данные (даже на первый взгляд незначительные), в силах измерять и прогнозировать демографические изменения, а также по оценке последствий цифровой революции определять демографическое поведение граждан.

Система э-демографии даст возможность эффективнее решать проблемы потребностей населения и четко определять стратегические аспекты будущего экономического и социального развития. Включение онлайн-регистрации и мониторинга поможет минимизировать время для принятия управленческих решений, в целом повысить эффективность управления в правительстве и муниципалитетах и сократить время, затрачиваемое на предоставление государственных услуг. Кроме того, э-демографическая система подготовит почву для ликвидации традиционной переписи: отпадет необходимость проводить онлайн-переписи и различные социально-демографические исследования на государственном (макроуровень), региональном (мезоуровень) и индивидуальном уровнях (микроуровень).

Е. Загхени и И. Вебер (Zagheni & Weber, 2015) изучали возможности и проблемы демографических исследований на основе интернет-данных и разработали методику получения неструктурированных данных, генерируемых на платформе социальной сети с применением методов интеллектуального анализа данных (*web data mining*).

Д. Албурез-Гутierrez с соавторами (Alburez-Gutierrez et al., 2019) рассматривают новые источники для исследования населения в цифровую эпоху, подчеркивая важность цифровых и библиометрических баз данных, цифровых следов в социальных сетях и интернета вещей как новых источников для демографических исследований. Они обсуждают технические и этические проблемы, возникающие в связи с этими источниками, а также возможности, предоставляемые для изучения традиционной и цифровой демографической динамики во всем мире.

Во многих странах гражданам рекомендуется для снижения риска использовать различные устройства слежения. Через устройства слежения ежедневная информация о здоровье граждан, часах досуга, времени, проведенном за просмотром телевизора, собирается и передается в базу данных соответствующих организаций (Gangietal, 2012; Redondietal, 2013). Несмотря на то, что цифровые данные, накопленные в устройствах слежения, подвергаются опасности захвата персональных данных, они могут помочь эффективнее определять плотность, уровень благосостояния населения или экономический рост. Они раскрывают атрибуты индивидуального уровня: пол, день рождения, образование, семейные и социальные отношения, интересы. Все это может быть использовано для получения социально-демографических оценок для разных групп населения. Анализ этих цифровых

данных по любой комбинации атрибутов может помочь более детализированно описать демографические тенденции общества, дать оценку демографии в определенных государствах (Качагина, 2016).

Социальные сети и демографические исследования. Использование интернета как источника данных для лучшего понимания демографических процессов является одной из наиболее важных задач, стоящих перед государством. Другая важная задача заключается в том, как с помощью информационных технологий управлять миграционными потоками и интеграцией мигрантов в условиях современных социально-политических процессов и растущей угрозы внезапных кризисов и потрясений. В э-демографии в качестве источника данных для демографических исследований в основном берутся популярные интернет-проекты – *Facebook*, *Google*, *Twitter*, *Yandex* и т.д. На основе данных, собранных в аналитических системах этих проектов, можно уточнить демографические характеристики многих стран. Отдельные компании также эффективно используют демографические данные, собираемые в виртуальных сетях, для повышения коэффициентов конверсии и сбора необходимых характеристик пользователей. Например, практически каждый рекламодатель *Facebook* ориентируется на пользователей по возрасту, доходу, интересам, времени занятости и полу. Эти параметры необходимы для углубленного анализа демографии.

Все большее количество географических данных, доступных в интернете, стало большим преимуществом для исследований миграции. Работы с различными открытыми виртуальными проектами, включая *Flickr*, *Twitter*, проекты *Google* (*Google Latitude*, *Google Maps* и т.д.), *Facebook*, *Wikipedia*, *Yahoo*, полученные с помощью этих проектов виртуальные карты, изображения, видеофайлы дают возможность изучить маршруты путешествий, миграции или информацию о географическом местоположении населения или отдельных групп (Gangietal, 2012; Redondietal, 2013; Качагина, 2016; Zagheni, 2018).

В большинстве случаев традиционные источники данных демографических исследований включают архивы переписей, данные о здоровье населения, статистику заболеваний и т.д. В качестве основного источника цифровых демографических исследований используются веб-браузеры, поисковые запросы, данные социальных сетей, данные, аккумулированные в государственных реестрах (электронные услуги, удовлетворенность граждан, отзывы о взаимоотношениях правительства и граждан и т.д.). Анализ больших данных, собранных в вышеупомянутых источниках, позволит эффективно проводить демографические исследования, получать знания о демографическом поведении населения и прогнозировать социально-демографические процессы. Так, Дж. Гинсберг с соавторами (J.Ginsberg et al., 2008), анализируя веб-запросы в *Google*, пришли к выводу, что эти запросы очень полезны и необходимы для мониторинга симптомов гриппа и выявления потенциальной эпидемии (Ginsberg et al., 2008).

Однако, несмотря на то, что мнения многих ученых о полезности использования веб-запросов в социальных исследованиях совпадают, время

от времени результаты анализов веб-запросов приводят аналитиков к неверным оценкам ситуации. Обычно это происходит, когда меняется содержание отношений между поиском, новостями и поведением (Ginsberg et al., 2008; Lazer, 2014). Данные из социальных сетей используются для выявления географического и экономического положения граждан, природной среды, в которой они живут, и для других целей. Основываясь на этой информации, социально-экономическое положение населения и тенденции миграции изучаются индивидуально и на государственном уровне (De Choudhury et al., 2010; Ferrari & Mamei, 2011).

Одно из основных направлений демографических исследований – изучение миграции. Интернет-ресурсы значительно расширили доступ к информации о географическом местоположении. Маршруты путешествий и краткосрочные смены места жительства устанавливаются по разным источникам, включая изображения и информацию о географическом местоположении на *Flickr*, *Twitter* и *Google Latitude* (Ferrari & Mamei, 2011; Ferrari et al., 2011; Noulas et al., 2011).

Ф. Биллари и соавторы (2013) (Billari et al., 2013) в своих исследованиях доказали, что, используя запросы «беременность» или «роды», можно с помощью *Google* прогнозировать потенциал рождаемости и даты рождения на несколько месяцев вперед. Наиболее значимым результатом этого исследования является то, что, наряду с традиционными источниками, использование данных из новых источников, таких как веб-поисковики, может увеличить точность прогнозирования демографической модели (см.: Billari et al., 2013).

В работе Д. Йилдиз и соавторов (Yildiz et al., 2017) исследована проблема использования данных *Twitter* для демографических исследований. Авторы постарались ответить на вопросы, как данные из социальных сетей могут охватить большое количество людей и какова лучшая практика для оценки демографической информации из данных *Twitter*. Они провели эксперимент, в ходе которого было задействовано 979 992 единицы твитов географических местоположений, отправленных 22 356 уникальными пользователями в Юго-Восточной Англии. Демографические характеристики пользователей *Twitter* оценивались с помощью краудсорсинговой платформы *CrowdFlower* и программного обеспечения для распознавания изображений *Face++*.

Полученные данные о возрасте, поле и географическом местоположении позволили исследователям обобщить результаты *Twitter*, применить их для всего населения. В исследовании Cesare et al. (2018) изучаются преимущества и недостатки использования цифровых следов в демографических исследованиях. Рассматриваются методологические возможности для решения проблем, связанных с цифровым анализом данных следа. Показывается, что существует большой потенциал для обнаружения важных сведений из неструктурированных данных большого объема с применением научных методов анализа данных, а также анализа цифровых следов подгрупп населения с использованием арсенала традиционных демографических методов.

В работе С. Гиль-Клавель и Е. Загхени (Gil-Clavel & Zagheni, 2019) исследовалась демография пользователей *Facebook* по всему миру. Для этого были взяты данные с рекламной платформы *Facebook*. Изучалось, как пользователи *Facebook* различаются по возрасту и полу, как размер сетей дружбы варьируется по этим параметрам, а также были определены демографические характеристики отдельных подгрупп пользователей социальных сетей. Полученные авторами данные позволяют по-новому взглянуть на гендерное неравенство в онлайн-среде, на некоторые нюансы демографических различий при внедрении и использовании цифровых технологий.

Большие данные в демографических исследованиях. Сегодня сотовая связь, сетевые данные помогают в определении перемещений внутри страны или причин миграции, а также в выявлении закономерностей демографических процессов в регионе. В последние годы использование больших данных в демографических исследованиях возросло (см.: Billari & Zagheni, 2017). В будущем интеллектуальный анализ больших данных, собранных в среде э-государства, государственных реестрах и базах данных отдельных компаний, обеспечит целый ряд значимых и более глубоких демографических исследований (Blumenstock, 2012; Deville et al., 2014).

Система э-демографии в среде э-государства содержит и объединяет информацию о метеорологических, демографических процессах, индивидуальные профили, сведения об обеспечении репродуктивного здоровья населения, о контроле за снижением заболеваемости и смертности, данные об условиях жизни населения, проведении непрерывной переписи, об исследовании миграции, о развитии кадрового и научного потенциалов в сфере э-государства. В контексте демографии устойчивость и развитие общества означают оптимальный рост населения и выбор наилучшего метода управления обществом. Демография может оцениваться по странам, регионам и областям. С этой точки зрения демографические исследования могут проводиться на государственном, региональном и индивидуальном уровнях (Ferrari & Mamei, 2011; Ferrari et al., 2011; Noulas et al., 2011; Качагина, 2016).

Электронные реестры и формирование э-демографии. Формирование национального э-государства в различных странах основано на реформировании всей системы государственного управления. Основная цель – соответствовать государственному управлению и повышению эффективности государственного управления за счет применения ИКТ, так как государственная администрация хочет иметь информационные системы, отвечающие таким требованиям, как прозрачность, гибкость, конкурентоспособность, ответственность за результаты, активное взаимодействие с гражданским обществом. Естественно, что при использовании государственных электронных ресурсов (э-ресурсов) учитываются и другие факторы, влияющие на весь характер общественно-политических, социальных, экономических, культурных и демографических отношений. Одним из наиболее важных компонентов э-государства является э-демография. Исследование проблем, связанных с демографической ситуацией, сбор

демографических показателей и работа по формированию индикаторов играют важную роль в повышении эффективности услуг э-государства.

Э-демография, будучи одной из составляющих демографической науки о законах воспроизведения населения, изучает то, как биологические характеристики и поведение людей влияют на развитие общества. Получение точной информации о демографической ситуации в соседних государствах – важная задача в сфере информационной безопасности каждого государства. При анализе демографической ситуации учитывается демографическая структура населения, т.е. его распределение по группам таких признаков, как национальность, возраст, пол, семейное положение, а также уровни старения населения, рождаемости, смертности и т.д. Направления исследований в э-демографии показаны на Рисунке 1.

Рисунок 1

Направления исследований в э-демографии



Работы по созданию реестра населения в разных странах были начаты еще в 1970-х гг. Хотя отдельные информационные системы и были созданы, но по разным причинам, а главное, из-за отсутствия необходимого уровня



информационных технологий стало невозможным создание единого реестра (UN Global Pulse's projects; Андриченко, Мещерякова, 2012).

С 1960-х гг. централизованные реестры населения используются в скандинавских странах, а с 1990-х гг. – во многих Прибалтийских странах и странах Центральной Европы (см: Poulain & Herm, 2013). Реестр населения – это прежде всего административный инструмент, и уже затем инструмент сбора статистических данных.

В европейских странах был предпринят ряд оригинальных инициатив по трансферу национальными статистическими системами управления демографических данных из административных источников, особенно из реестров населения. Таким образом, реестр населения вместо отдельного анализа демографических событий по одному человеку позволяет связать все данные и получить доступ к статистике для всего населения. Например, если рассмотреть опыт стран Северной Европы и Прибалтики, удостоверение личности родителей будет считаться одним из регистрационных данных при условии, что они указаны в вашем свидетельстве о рождении. Вся эта информация может быть использована для идентификации ядер семейств и для формирования статистических таблиц, которые их разъясняют. В Таблице 1 приведены персональные характеристики, которые передаются из реестра населения в статистические управление в некоторых европейских странах (Poulain & Herm, 2013).

Таблица 1

Индивидуальные характеристики, перенесенные из реестра населения в национальные статистические управление

	Личный идентификатор (PIN)	Рождение	Пол	Гражданство	Данные об этническом происхождении	Образование	Прочие идентификаторы		
							Супруг (а)	Отец	Мать
Дания	+	+	+	+	-	-	+	+	+
Эстония	+	+	+	+	+	+	+	+	+
Венгрия	-	+	+	+	-	-	-	-	-
Латвия	+	+	+	+	+	-	+	+	+
Нидерланды	+	+	+	+	-	-	+	+	+
Норвегия	+	+	+	+	-	-	+	+	+
Польша	+	+	+	+	-	+	-	-	-
Словения	+	+	+	+	-	-	-	-	-
Испания	-	+	+	+	-	-	-	-	-
Швеция	+	+	+	+	-	-	+	+	+

Источник: Составлена авторами.

Анализ публикаций результатов проведенных исследований показывает, что основной целью реестра населения является получение точной идентификации каждого человека. Реестр населения также предоставляет агрегированные данные о населении, позволяет осуществлять и регулировать в более широком спектре политику государственного управления и планирования. Кроме того, если система допускает непрерывную регистрацию изменений адреса в режиме онлайн, то она может считаться основным источником информации о миграции.

В принципе, функционирование и важность системы не подлежит сомнению после устранения начальных ошибок и упущений в реестрах (любая административная ошибка может быть исправлена без строгого соблюдения норм закона) и проблем общения между администраторами реестра и специалистами, проводящими статистический анализ. С другой стороны, как местные, так и государственные органы заинтересованы в статусе лиц, зарегистрированных в реестре населения. Как правило, одним из самых больших недостатков в реестрах населения и самым слабым компонентом является неспособность населения своевременно предоставлять обновленную информацию.

Например, в нидерландской практике в целях обеспечения обновления и поддержания реестров в актуальном состоянии, а также повышения ответственности граждан, на своевременно не предоставивших информацию в местные органы исполнительной власти налагается штраф в размере 325 евро (Prince, 2017). Однако пробелы допускаются в любом случае, независимо от уделенного внимания и проведенных проверок, и иногда этот процесс периодически повторяется из-за ошибок отдельных лиц или соответствующих органов. Практика показывает, что ошибки чаще встречаются в динамических личных данных, а в документированной информации (место и дата рождения, семейное положение и т.д.) риск допущения ошибок ограничен.

В настоящее время во многих развитых странах власти работают над созданием социальных реестров. В последние годы активизировались работы по формированию единого регистра населения. Например, в Российской Федерации к 2025 г. будет создан единый реестр населения (Чудиновских, 2018; Реестр населения, 2018). Информация обо всех людях, живущих в стране, будет собираться и обновляться в режиме онлайн в реестре населения. Во время переписи населения к 2020 г. граждане смогут свободно заполнять свои анкеты и предоставлять их на портал государственных услуг. С момента рождения человека вся информация о нем будет включена в реестр. В целом реестр будет иметь всю информацию, доступную властям, и обмен информацией между государственными органами и гражданами будет осуществляться в режиме онлайн (Реестр населения, 2018).

Применение реестра населения требует использования индивидуального идентификационного номера (*Personal Identification Number, PIN*) для физических лиц. Во многих случаях эти *PIN* определяются до внедрения реестра. Основная цель применения единого кодового номера состоя-

ла в том, чтобы устраниить дублирование при расчете, улучшить координацию между различными государственными реестрами и усовершенствовать процедуры сбора налогов. В настоящее время уникальные идентификационные номера требуются во всех странах, в которых имеется реестр населения, и этот номер используется для обновления анонимных персональных данных.

С помощью современных цифровых технологий для точного определения демографической ситуации и прогнозирования используются не только данные из архивов переписи, но также данные из социальных сетей, в которых собираются персональные данные граждан (архивы медицинских учреждений, банков, страховых компаний и т.д.). Основные источники э-демографических исследований включают поисковые запросы через веб-браузеры, данные, собранные в государственных реестрах и об э-услугах, отзывы о взаимоотношениях правительства и граждан, удовлетворенность граждан и т.д. Анализ этих больших данных, агрегированных в вышеупомянутых источниках, позволит сформировать новые идеи о демографическом поведении, завершить существующие исследования и получить новые знания в области демографии.

Общая характеристика э-демографической системы

Исторически государственные органы были монополистами в сборе данных, переписи, управлении, административных записях и анализе информации о социально-экономическом положении населения. Основным преимуществом демографов в скандинавских странах, Бельгии и Нидерландах было использование индивидуальных регистрационных данных в демографических исследованиях для расширения и подключения индивидуальных регистраций и других реестров (Lyngstad & Skardhamar, 2011).

Масштаб использования больших данных для демографических исследований растет. Для демографического благополучия, реализации эффективной демографической политики нужно в первую очередь развивать исследования по демографии, применяя современные информационные технологии, которые вместе формируют среду э-демографии. Среда э-демографии позволит проводить мониторинг статистики населения и анализировать различные демографические процессы на интегрированных электронных реестрах (э-реестрах), входящих в портал э-государства. Для управления демографическими процессами важно сформировать э-демографическую среду посредством современных информационных технологий. Формирование э-демографии поможет решить следующие вопросы:

1. Формирование интегрированной базы данных для э-государства.
2. Формирование демографических показателей для интеграционных и миграционных процессов.
3. Модернизация законодательства в области миграции в соответствии с международными нормами и подготовка миграционного реестра в соответствии с требованиями э-демографической системы.

4. Оценка демографической ситуации, системы мониторинга и прогнозирования с подробным анализом.
5. Анализ демографических процессов для изучения демографических характеристик, обеспечения гендерного баланса и формирования индивидуального профиля.
6. Оценка и мониторинг демографической ситуации, детальный анализ систем прогнозирования.
7. Защита экологического баланса и окружающей среды.
8. Расширение системы цифровой медицины, улучшение работы и укрепление материально-технической базы учреждений здравоохранения и т.д.

Для эффективного решения вышеперечисленных задач необходимо разработать единую национальную э-демографическую систему. Формирование э-демографической системы поможет построению единого информационного демографического пространства, создаст дополнительные возможности для анализа демографических процессов и тем самым будет способствовать обеспечению репродуктивного здоровья населения, улучшению условий его жизни, укреплению института семьи, решению миграционных вопросов и развитию кадрового и научного потенциалов в демографической сфере.

Для прогнозирования демографических процессов часто применяются подходы, основанные на моделировании (Courgeau et al., 2016). При этом основное внимание уделяется моделированию конкретных групп населения и развитию мощных связей с эмпирическими данными из социальных сетей (Henderson, 2012; Weber & State, 2017). Моделирование в э-демографии может помочь обойти это ограничение, упрощая использование вычислительных моделей для генерации сценариев. Моделирование может применяться также для изучения пространства параметров, в котором они работают, а также для того, как различные сценарии могут влиять на поведение индивидуума и отдельных групп.

КОНЦЕПТУАЛЬНАЯ МОДЕЛЬ Э-ДЕМОГРАФИЧЕСКОЙ СИСТЕМЫ

Реализация э-демографической политики считается неотъемлемой частью системы э-государства и требует создания систем анализа демографических процессов и реестров для оценки существующих демографических ситуаций.

Развитие э-государства гарантирует, что традиционный процесс переписи будет заменен интеллектуальной системой, основанной на большой базе данных, собранных в государственных реестрах. Онлайн-регистрация и управление мониторингом помогут минимизировать время принятия решений, что в целом позволит правительствам и муниципалитетам повысить эффективность управления и сократить время, затрачиваемое на предоставление государственных услуг. Такая система также имеет ре-

шающее значение для управления демографическими процессами, принятия решений и прогнозирования.

Отметим, что различные э-демографические модели могут быть предложены в зависимости от стратегий построения э-государства стран. В статье предлагается концептуальное построение э-демографической системы на основе единого государственного реестра. Мы считаем, что если речь идет о платформе э-государства, тогда все государственные реестры должны быть интегрированы, и в этом случае все данные, собранные в едином реестре, будут играть важную роль в демографических исследованиях.

Если мы рассмотрим архитектуру э-демографической системы на примере Азербайджана, то увидим, что э-государство более совместимо с восточной моделью. Учитывая, что в настоящее время в стране действует портал э-государства, в него постепенно интегрируются отдельные государственные реестры. В этом случае после процесса интеграции государственных реестров или формирования единого государственного реестра может быть предложено создание э-демографической системы на его основе.

Концептуальная модель системы э-демографии показана на Рисунке 2. Информационной базой данных для анализа демографии служат следующие данные:

1. Данные переписи населения.
2. Данные из реестров разных учреждений (банковские отчеты, учетные листы, истории болезни и т.д.).
3. Данные из аналитических систем, проводящих статистический учет и анализ демографических событий, таких как рождение, браки и разводы, смерти, миграции.
4. Данные из социальных сетей.
5. Данные из специальных систем слежения и др.

В соответствии со стратегией долгосрочного развития, э-реестры будут созданы и интегрированы в э-демографическую систему для контроля гендерного баланса и комплексного развития людских ресурсов.

Процесс обработки данных в системе э-демографии состоит из четырех этапов (как показано на Рисунке 2):

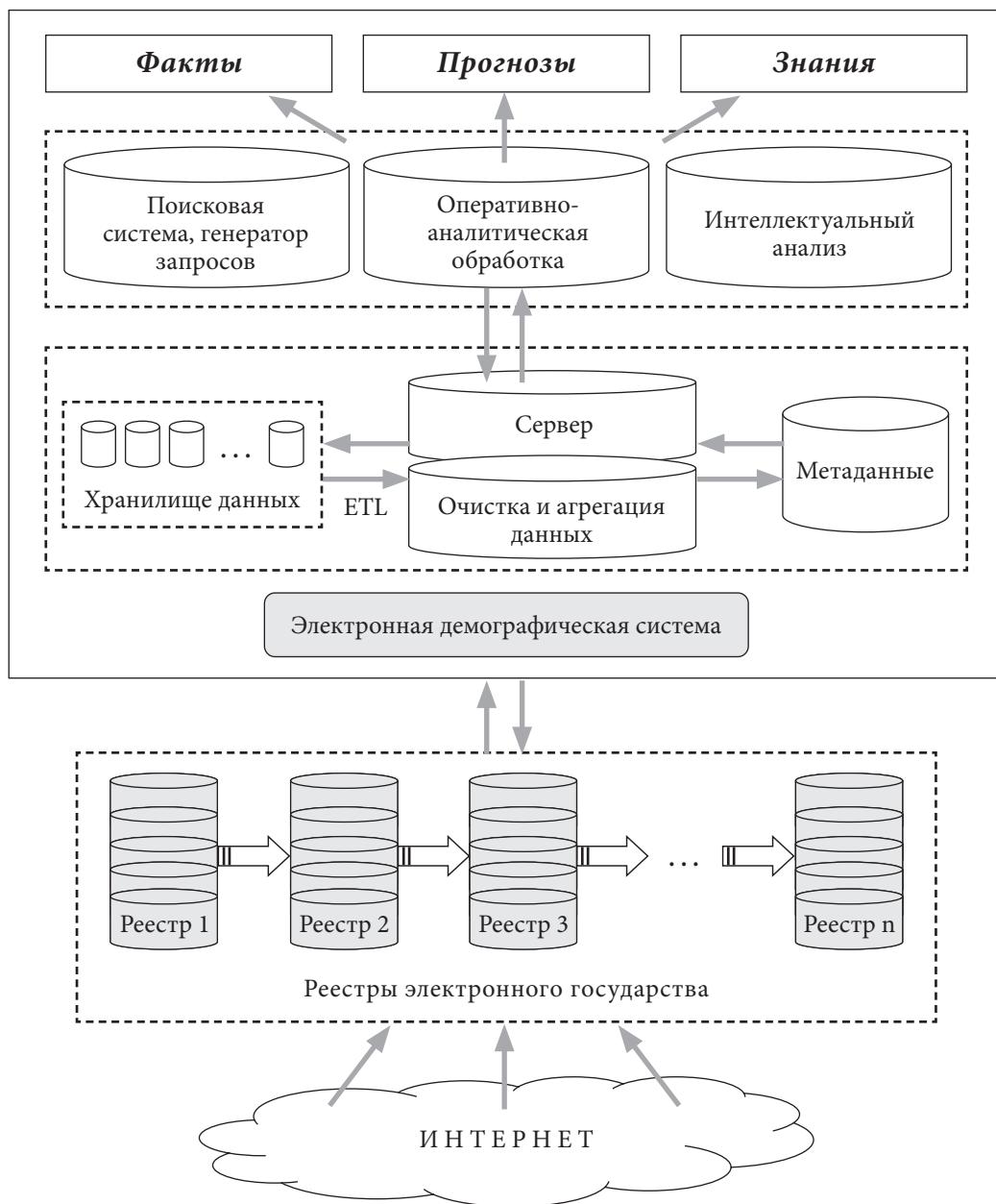
1. Определение реестров данных и сбор данных.
2. Формирование кубов.
3. Анализ данных.
4. Извлечение данных для принятия решений.

Данные анализируются в разных реестрах на основе PIN, и таким образом формируется облако каждого человека. Отдельные облака могут быть объединены в соответствии с различными критериями (образование, медицина, миграция и т.д.). Данные в реестрах – большие данные (*big data*) и PIN – позволят связать значительное количество записей и обнаружить их взаимосвязь.

Система PIN была впервые введена в скандинавских странах: Швеции (1947), Норвегии (1961), Финляндии (1964) (статистика на основе регистров, 2008; Чудиновских, 2018).

Рисунок 2

Концептуальная модель системы э-демографии



Один из ключевых факторов, позволяющих статистически использовать данные, собранные в государственном реестре, – наличие единой системы идентификации или единого идентификационного номера для различных источников и ресурсов. В отсутствие единой системы идентификации трудно интегрировать и объединять различные реестры. В связи с этим можно проводить статистический анализ на основе данных, собранных в реестрах. В качестве минимального требования имеется в виду наличие единого идентификационного номера для отдельных реестров.

Единая система регистрации населения имеет большое практическое значение. Реестры данных считаются эффективным средством сбора и мониторинга информации о населении. Основным преимуществом таких реестров является то, что они не имеют конкретной цели и, следовательно, предоставляют информацию обо всех жителях, проживающих в стране, без исключения. Согласно Заявлению ООН по населению и социальной статистике, в 2008 г. в странах Северной Европы личный идентификационный номер стал применяться почти во всех реестрах, используемых для статистики (Чудиновских, 2018). Для других реестров, в том числе коммерческого, адресного, жилья и т.д., была применена единая система идентификации. В настоящее время применяемые технологии и существующая система позволяют в кратчайшие сроки координировать все доступные государственные реестры. Как отмечается в докладе ООН, одной из причин успешного развития административных регистраций в странах Северной Европы стало то, что население этих стран было небольшим, компактным и однородным. В крупных странах создание административных реестров на государственном уровне казалось более проблематичным из-за технических, организационных и других ограничений. В настоящее время платформа э-государства, которая успешно внедряется в различных странах, имеет широкие возможности и инструменты для интеграции реестров, создания единой инфраструктуры (Статистика на основе регистров, 2008; Чудиновских, 2018).

Методы анализа больших данных

Для демографических анализов могут быть использованы различные методы, такие как:

- *статистический анализ данных.* Применение математических и статистических методов для анализа демографических данных. В настоящее время эта область широко распространена и используется большинством научного сообщества;
- *анализ графов.* Изучение графов, которые представляют собой математические структуры, используемые для моделирования парных отношений между объектами. Граф содержит вершины, представляющие объекты, а также ребра, которые изображают отношения между узлами (Diestel, 2000);
- *интеллектуальный анализ данных.* Это процесс обнаружения скрытой информации (знания) и значимых структур из больших баз данных. В основном используются алгоритмы кластеризации, классификации и прогнозирования. В частности, это может помочь обнаружить, описать и предсказать связи или тенденции между данными (Poltavtseva, Pechenkin, 2017);
- *OLAP (Online Analytical Processing).* Это технология, использующая информацию в хранилищах данных. OLAP позволяет проводить многомерный анализ данных путем построения кубов. Через эти кубы метод обеспечивает визуализацию и быстрый анализ больших данных для принятия решений (Chaudhuri, Dayal, 1997).

Объем, многообразие и скорость обработки данных порождают спрос на аналитическую онлайн-обработку данных – OLAP. OLAP – это система анализа многомерных данных, которая обеспечивает быстрый анализ для принятия решений. Хранилища данных (Data Warehouse) и OLAP являются одними из устоявшихся технологий для обработки и анализа структурированных данных, но с быстрым распространением информационных сетей требования к этим технологиям непрерывно возрастают. Сегодня OLAP используется в управлении разнородными информационными сетями. Система может быть полезна для мониторинга, просмотра и анализа разных типов веб-контента и структуры сетей в режиме онлайн (Loudcher et al., 2015).

Для решения задачи требуется построить OLAP-куб. Предварительная подготовка аналитических кубов позволяет оператору, принимающему решение, сосредоточить внимание на аналитическом процессе с кубом: поворот, детализация, свертывание. Ячейки куба являются аналитическими мерами. Мера куба представляет собой числовую характеристику анализируемого процесса, набор измерений образует ось куба. Количество и содержание кубов зависят от количества решаемых задач. Такой подход приводит к иррациональной группировке мер и измерений. В то же время существует ряд проблем, связанных с анализом неструктурированных больших данных (Big Data), взятых из отдельных источников. Для загрузки из разных источников и консолидации больших данных предлагается в системах э-демографии применять технологию ETL (Extract, Transform, Load – извлечение, преобразование, загрузка). ETL выполняет следующие задачи (см.: Gutiérrez-Batista et al., 2015):

1. Извлечение данных. На этом этапе данные собираются из разных реестров и затем копируются в хранилище данных.
2. Преобразование данных в соответствующую форму. Данные консолидируются.
3. Загрузка данных. ETL выполняет перенос данных из реестров в хранилище данных и для слияния данных.

ETL может работать как с временными, так и постоянными данными, в зависимости от варианта использования. Данные, предназначенные для анализа, используются для построения OLAP-кубов. Благодаря стандартизации к технологии хранилища данных можно подключать интерфейсные инструменты различных реестров, что удобно для всестороннего анализа больших объемов данных, генерируемых через интернет. Хранилище данных содержит данные в агрегированном виде, т.е. исходные данные агрегируют до определенного уровня детализации. Но анализ данных в режиме реального времени накладывает дополнительные требования на предлагаемую систему. Представленная архитектура многоуровневой системы дает возможность за счет оптимизации загрузки новых данных обеспечить анализ в реальном времени. Хотя анализируемые данные довольно динамичны, сам процесс анализа интерактивный и должен учитывать неограниченные сценарии «что, если». Это включает изменение отношений данных или обновление значений для определения, как различные изменения повлияют на демографию.

Э-демографическая система на платформе э-государства может быть разделена на различные секторы, такие как социальное страхование, медицина, миграция и т.д. Так можно анализировать демографические процессы эффективнее и быстрее. Со временем после полной интеграции в платформу э-государства э-реестры будут преобразованы в виртуальные облака. Следует отметить, что э-демографическая система будет сформирована на платформе э-государства и наряду с онлайн-переписью обеспечит условия для проведения различных социально-демографических исследований: точного определения статистики населения, онлайн-мониторинга, углубленного анализа демографических процессов и выявления демографических проблем в обществе.

Демографическая безопасность общества и развитие человеческих ресурсов – ключевые стратегические аспекты любого государства. Создание э-демографических систем и механизмов для достижения прогнозирования человеческих ресурсов и выявления рисков в формировании э-государства поможет построить научно обоснованную и эффективную э-государственную систему и определить стратегические направления экономического и социального развития страны.

Заключение

Реализация эффективной демографической политики в стране воспринимается как неотъемлемая часть системы э-государства и требует эффективного использования реестров и систем для оценки, анализа и принятия действенных решений в существующей демографической ситуации. На демографические процессы, помимо миграции, рождаемости или смертности, также влияют разные социально-экономические ситуации. Авторами определены перспективные направления исследований в области демографии и предложена концептуальная модель э-демографической системы для управления демографическими процессами.

Создание единого регистра населения позволит отменить традиционные переписи, осуществлять онлайн-перепись и мониторинг населения на основе анализа данных, собранных в государственном реестре. Можно отметить, что, если речь идет о платформе э-государства, все государственные реестры должны быть интегрированы, и все данные, собранные в реестре, сыграют важную роль в демографических исследованиях. В рассматриваемом подходе э-демографическая система формируется на основе единого государственного реестра и действует независимо от функций э-государства.

Демографическая безопасность и развитие человеческих ресурсов – ключевые стратегические аспекты развития стран. Э-демографическая система как основной компонент государства позволит точно определить статистику населения, проводить онлайн-мониторинги, глубокий анализ демографических процессов и выявлять проблемы, связанные с демографией. Создание э-демографических систем и механизмов для достижения целей при прогнозировании человеческих ресурсов и выявлении рисков в процессе развития

э-государства поможет сформировать всеобъемлющую, научно обоснованную и эффективную систему мер, определить стратегические направления будущего экономического и социального развития. Изучение демографических процессов, формирование систем э-реестра и комплексных условий для стабилизации уровня развития населения можно эффективно использовать в системах принятия решений по обеспечению общественного прогресса и социально-экономического развития страны, тем самым способствуя увеличению численности и качества трудоспособного населения. В статье описаны преимущества применения технологий Big Data, OLAP, ETL для анализа демографических данных. Учитывая актуальность темы, с целью проведения различных глубоких демографических анализов интеллектуальный анализ больших данных, собранных в социальных сетях и государственных реестрах, будет рассмотрен в дальнейших исследованиях.

ЛИТЕРАТУРА

1. Андриченко Л.В., Мещерякова М.А. Информационные регистры как эффективное средство сбора и мониторинга данных о населении // Журнал российского права. – 2012. – № 8. – С. 16–40.
2. Борисов В.А. Демография. – М.: Издательский дом NOTABENE, 2001. URL: http://www.sociologos.ru/upload/File/Methods/Demography_Borisov.pdf (дата обращения: 15.03.2019).
3. Качагина О.В. Основы демографии: основы теории и практические задания: Учебное пособие. – Ульяновск: УлГУ, 2016.
4. Реестр населения. В России будет создан реестр населения // Российская газета. – Федеральный выпуск № 122 (7585). URL: <https://rg.ru/2018/06/06/v-rossii-budet-sozdan-reestr-naseleniia.html> (дата обращения: 26.03.2019).
5. Статистика на основе регистров. Статистика на основе регистров в североевропейских странах. Обзор передовых методик с уделением основного внимания статистике населения и социальной статистике / ООН, Европ. экон. комис. – Нью-Йорк; Женева: ООН, 2008. – VI.
6. Черемных В.Ю., Яковлев Л.С. Электронное правительство: модели и перспективы // Вестник Поволжского института управления. – 2017. – Т. 17. – № 1. – С. 68–74.
7. Чудиновских О. Большие данные и статистика миграции // Вопросы статистики. – 2018. – Т. 25. – № 2. – С. 48–56.
8. Alburez-Gutierrez D., Aref S., Gil-Clavel S. and et al. Demography in the Digital Era: New Data Sources for Population Research. In: Arbia G., Peluso S., Pini A., Rivellini G. (eds.), Book of short Papers SIS2019. Pearson. 2019. DOI: <https://osf.io/preprints/socarxiv/24jp7/>
9. Billari F., D'Amuri F., Marcucci J. Forecasting births using google. Annual Meeting of the Population Association of America, New Orleans, LA.17. 2013.

10. Billari F., Zagheni E. Big data and population processes: a revolution? In: Petrucci A., Verde R. (eds.) *Statistics and Data Science: new challenges, new generations*. Proceedings of the Conference of the Italian Statistical Society. Firenze University Press. 28–30 June, Florence (Italy). 2017. P. 167–178.
11. Blumenstock J.E. Inferring patterns of internal migration from mobile phone call records: evidence from Rwanda // *Information Technology for Development*. 2012. Vol. 18. No. 2. P. 107–125.
12. Careja R., Bevelander P. Using population registers for migration and integration research: examples from Denmark and Sweden // *Comparative Migration Studies*. 2018. Vol. 6. No. 1. P. 6–19.
13. Cesare N., Lee H., McCormick T. and et al. Promises and Pitfalls of Using Digital Traces for Demographic Research // *Demography*. 2018. Vol. 55. No. 5. P. 1979–1999.
14. Chaudhuri S., Dayal U. An overview of data warehousing and OLAP technology // *ACM SIGMOD*. 1997. Vol. 26. No. 1. P. 65–74.
15. Courgeau B., Bijak J., Franck R., Silverman E. Model-Based Demography: Towards a Research Agenda // *Agent-Based Modelling in Population Studies*. 2016. Vol. 48. P. 29–51.
16. De Choudhury M., Feldman M., Amer-Yahia S., Golbandi N., Lempel R., Yu. C. Automatic construction of travel itineraries using social breadcrumbs. In: *Proceedings of the 21st ACM Conference on Hypertext and Hypermedia*. ACM. 2010. P. 35–44.
17. Deville P., Linard C., Martin S., Gilbert M., Stevens F.R., Gaughan A.E., Blondel V.D., Tattem A.J. Dynamic population mapping using mobile phone data. In: *Proceedings of the National Academy of Sciences*. 2014. Vol. 111. No. 45. P. 15888–15893.
18. Diestel R. *Graph Theory*. Springer-Verlag, New York. 2000. URL: www.esi2.us.es/~mbilbao/pdffiles/DiestelGT.pdf (дата обращения: 11.03.2019).
19. Emilio Zagheni is new MPIDR director. 2018. URL: www.demogr.mpg.de/en/news_press/news/news/emilio_zagheni_is_new_mpидr_director_5556.htm (дата обращения: 08.03.2019).
20. Ferrari L., Mamei M. Discovering daily routines from google latitude with topic models. In: *Proceedings of the International Conference on Pervasive Computing and Communications Workshops*, IEEE. 2011. P. 432–437.
21. Ferrari L., Rosi A., Mamei M., Zambonelli F. Extracting urban patterns from location-based social networks. In: *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks*. ACM. 2011. P. 9–16.
22. Fire M., Elovici Y. Data Mining of Online Genealogy Datasets for Revealing Lifespan Patterns in Human Population // *ACM Transactions on Intelligent Systems and Technology*. 2015. Vol. 6. Issue 2. P. 1–24.
23. Gangi R.R., Rajesh N.B., Sudhakar N.P., Raviteja B., Rammohanarao K. Tracking objects, using RFID and wireless sensor networks // *International Journal of Engineering Science & Advanced Technology*. 2012. Vol. 2. Issue 3. P. 513–517.
24. Gil-Clavel S., Zagheni E. Demographic Differentials in Facebook Usage around the World, In: *Proceedings of the Thirteenth International AAAI Conference on Web and Social Media (ICWSM 2019)*. 2019. P. 647–650.
25. Ginsberg J., Mohebbi M.H., Patel R.S., Brammer L., Smolinski M.S., Brilliant L. Detecting influenza epidemics using search engine query data // *Nature*. 2008. Vol. 457. No. 7232. P. 1012–1014.

26. Gutiérrez-Batista K., Campaña J.R., Vila M.-A., Martin-Bautista M.J. Building a contextual dimension for OLAP using textual data from social networks // Expert Systems with Applications. 2018. Vol. 93. Issue C. P. 118–133.
27. Henderson J.V., Storeygard A., Weil D.N. Measuring economic growth from outer space // The American Economic Review. 2012. Vol. 102. No. 2. P. 994–1028.
28. Lazer D.M., Kennedy R., King G., Vespignani A. The parable of google flu: traps in big data analysis // Science. 2014. Vol. 343. No. 6176. P. 1203–1205.
29. Loudcher S., Jakawat W., Soriano Morales E.P., Favre C. Combining OLAP and information networks for bibliographic data analysis: a survey // Scientometrics. 2015. Vol. 103. Issue 2. P. 471–487.
30. Lyngstad T.H., Skardhamar T. Nordic register data and their untapped potential for criminological knowledge // Crime and Justice. 2011. Vol. 40. No. 1. P. 613–645.
31. Mehr H. Artificial Intelligence for Citizen Services and Government. Cambridge, Harvard Kennedy School. 2017. URL: http://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf (дата обращения: 29.03.2019).
32. Noulas A., Scellato S., Mascolo C., Pontil M. An empirical study of geographic user activity patterns in foursquare. In: Proceedings of the International AAAI Conference on Web and Social Media (ICWSM). 2011. Vol. 11. P. 570–573.
33. Poltavtseva M. A., Pechenkin, A. I. Intelligent data analysis in decision support systems for penetration tests // Automatic control and computer sciences. 2017. Vol. 51. Issue 8. P. 985–991.
34. Poulain M., Herm A. Central population registers as a source of demographic statistics in Europe // Population. 2013. Vol. 68. No. 2. P. 183–212. DOI: <https://doi.org/10.3917/popu.1302.0215>
35. Prins K. Population register data, basis for the Netherlands Population Statistics. Statistics Netherlands, Hague. 2017. URL: www.cbs.nl/-/media/_pdf/2017/38/population-register-data.pdf (дата обращения: 26.10.2019).
36. Redondi A., Chirico M., Borsani L., Cesana M., Tagliasacchi M. An integrated system based on wireless sensor networks for patient monitoring, localization and tracking // Ad Hoc Networks. 2013. Vol. 11. Issue 1. P. 39–53.
37. Shryock H.S., Siegel J.S., Larmon E.A. The Methods and Materials of Demography. Department of Commerce. Bureau of the Census. 1980.
38. United Nations Global Pulse's projects. URL: www.unglobalpulse.org/projects (дата обращения: 12.03.2019).
39. Weber I., State B. Digital Demography. In: Processing of the International World Wide Web Conference Committee (IW3C2). April 3–7. Perth, Australia. 2017. P. 935–939.
40. Yildiz D., Munson J., Vitali A. and et al. Using Twitter data for demographic research // Demographic Research. 2017. Vol. 37. Article 46. P. 1477–1514.
41. Zagheni E. Data Science, Demography and Social Media: Challenges and Opportunities. 2017. URL: <https://pdfs.semanticscholar.org/presentation/> (дата обращения: 05.03.2019).
42. Zagheni E., Weber I. Demographic research with non-representative internet data // International Journal of Manpower. 2015. Vol. 36. No. 1. P. 13–25.

DEVELOPING ELECTRONIC DEMOGRAPHY AS AN EFFECTIVE TOOL FOR SOCIAL RESEARCH AND MONITORING POPULATION DATA

Rasim M. Alguliyev

Academician of ANAS, Doctor of Technical Sciences,
Director of the Institute of Information Technologies,
Azerbaijan National Academy of Sciences.
Address: 9A, B. Vahabzade str., AZ1141 Baku, Azerbaijan
E-mail: director@iit.science.az

Ramiz M. Aliguliyev

Corresponding member of ANAS, Doctor of Technical Sciences,
Head of the Department, Institute of Information Technologies,
Azerbaijan National Academy of Sciences.
Address: 9A, B. Vahabzade str., AZ1141 Baku, Azerbaijan
E-mail: r.aliguliyev@gmail.com

Farhad F. Yusifov

Associate Professor, Head of the Department,
Institute of Information Technologies,
Azerbaijan National Academy of Sciences.
Address: 9A, B. Vahabzade str., AZ1141 Baku, Azerbaijan
E-mail: farhadyusifov@gmail.com

Irada Y. Alekperova

Ph.d. (in Technical Sciences), Chief Specialist, Institute of Information Technologies, Azerbaijan National Academy of Sciences.
Address: 9A, B. Vahabzade str., Baku, AZ1141, Azerbaijan
E-mail: airada.09@gmail.com

Abstract

The article is devoted to developing an electronic demographic (e-demographic) system based on a single state registry for conducting a demographic research. In the digital age, new data sources for the study of demographic behavior are emerging. Currently, population registry data used for the study of population is still quite limited. Therefore, the creation of the e-demographic system based on a single state registry for effective demographic research, as well as for the analysis of not only the population registry, but also other state registries, is of great relevance. The article examines the international experience of using the demographic characteristics of the data accumulated in the population registry. The practice shows that although population registry data are currently used in population studies, there are no approaches to the use of data from other state registries and to the development of an e-demographic system. The article proposes the concept of forming an e-demographic system based on a single state registry which operates independently of the functions of the e-government. The article outlines the benefits of using Big Data, OLAP, ETL technologies to analyze the demographic data. The development of the system allows better understanding the demographic processes and enables in-depth research at different levels from social demography to population geography and from migration processes to medical demography.

Keywords: electronic government; electronic demography; demographic indicators; migration; electronic registries; population registers.

Citation: Alguliyev, R.M., Aliguliyev, R.M., Yusifov, F.F. & Alekperova, I.Y. (2019). Formirovanie elektronnoi demografii kak effektivnogo instrumenta sotsialnykh issledovanii i monitoringa dannykh o naselenii [Developing Electronic Demography as an Effective Tool for Social Research and Monitoring Population Data]. *Public Administration Issue*, no 4, pp. 61–86 (in Russian).

REFERENCES

1. Andrichenko, L.V. & Meshcheryakova, M.A. (2012). Informatsionnye registry kak effektivnoe sredstvo sbora i monitoringa dannykh o naselenii [Information Registers as an Effective Tool of Collecting and Monitoring Population Data]. *Journal of Russian Law*, no 8, pp. 16–40.
2. Borisov, V.A. (2001). *Demografiya* [Demography]. Moscow: NOTABENE Publishing House. Available at: http://www.sociologos.ru/upload/File/Methods/Demography_Borisov.pdf (accessed: 15 March 2019).
3. Kachagina, O.V. (2016). *Osnovy demografii: osnovy teorii i prakticheskie zadaniya: Uchebnoe posobie* [Basics of Demography: Theoretical Framework and Practical Tasks: Tutorial]. Ulyanovsk: UlSU.
4. *Reestr naseleniya. V Rossii budet sozdan reestr naseleniya* [Population Register. In Russia, a Population Register Will Be Created]. *Rossiyskaya Gazeta – Federal Issue*, no 122 (7585). Available at: <https://rg.ru/2018/06/06/v-rossii-budet-sozdan-reestr-naseleniiia.html> (accessed: 26 March 2019).
5. UN, Geneva (2008). *Statistika na osnove registrov. Statistika na osnove registrov v severoевропейских странах. Obzor peredovykh metodik s udelemiem osnovnogo vnimaniya statistike naseleniya i sotsial'noi statistike* [Register Based Statistics. Register Based Statistics in Northern European Countries. Review of Best Practices with a Focus on Population Statistics and Social Statistics].
6. Cheremnykh, V.Yu. & Yakovlev, L.S. (2017). Elektronnoe pravitel'stvo: modeli i perspektivy [E-Government: Models and Prospects]. *Bulletin of the Volga Region Institute of Administration*, vol. 17, no 1, pp. 68–74.
7. Chudinovskikh, O. (2018). Bol'shie dannye i statistika migratsii [Big data and migration statistics]. *Statistical Issues*, vol. 25, no 2, pp. 48–56.
8. Alburez-Gutierrez, D., Aref, S., Gil-Clavel, S. et al. (2019). Demography in the Digital Era: New Data Sources for Population Research. In: Arbia G., Peluso S., Pini A., Rivellini G. (eds.) *Book of short Papers SIS2019*. Pearson. Available at: DOI: <https://osf.io/preprints/socarxiv/24jp7/> (accessed: 15 November 2019).
9. Billari, F., D'Amuri, F. & Marcucci, J. (2013). *Forecasting births using google*. Annual Meeting of the Population Association of America, New Orleans, LA.17.
10. Billari, F. & Zagheni, E. (2017). Big Data and Population Processes: A Revolution? In: Petrucci A., Verde R. (eds.) *Statistics and Data Science: New Challenges, New Generations. Proceedings of the Conference of the Italian Statistical Society*. Firenze University Press. 28–30 June, Florence (Italy), pp. 167–178.

11. Blumenstock, J.E. (2012). Inferring Patterns of Internal Migration from Mobile Phone Call Records: Evidence from Rwanda. *Information Technology for Development*, vol. 18, no 2, pp. 107–125.
12. Careja, R. & Bevelander, P. (2018). Using Population Registers for Migration and Integration Research: Examples from Denmark and Sweden. *Comparative Migration Studies*, vol. 6, no 1, pp. 6–19.
13. Cesare, N., Lee, H., McCormick, T. et al. (2018). Promises and Pitfalls of Using Digital Traces for Demographic Research. *Demography*, vol. 55, no 5, pp. 1979–1999.
14. Chaudhuri, S. & Dayal, U. (1997). An Overview of Data Warehousing and OLAP Technology. *ACM SIGMOD*, vol. 26, no 1, pp. 65–74.
15. Courgeau, B., Bijak, J., Franck, R. & Silverman, E. (2016). Model-Based Demography: Towards a Research Agenda. *Agent-Based Modelling in Population Studies*, no 48, pp. 29–51.
16. De Choudhury, M., Feldman, M., Amer-Yahia, S., Golbandi, N., Lempel, R. & Yu, C. (2010). Automatic Construction of Travel Itineraries Using Social Breadcrumbs. In: *Proceedings of the 21st ACM Conference on Hypertext and Hypermedia*. ACM. pp. 35–44.
17. Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F.R., Gaughan, A.E., Blondel, V.D. & Tatem, A.J. (2014). Dynamic Population Mapping Using Mobile Phone Data. In: *Proceedings of the National Academy of Sciences*, vol. 111, no 45, pp. 15888–15893.
18. Diestel, R. (2000). *Graph Theory*. New York: Springer-Verlag. Available at: www.esi2.us.es/~mbilbao/pdffiles/DiestelGT.pdf (accessed: 11 March 2019).
19. Emilio Zagheni is New MPIDR Director (2018). Available at: www.demogr.mpg.de/en/news_press/news/news/emilio_zagheni_is_new_mpidr_director_5556.htm (accessed: 08 March 2019).
20. Ferrari, L. & Mamei, M. (2011). Discovering Daily Routines from Google Latitude with Topic Models. In: *Proceedings of the International Conference on Pervasive Computing and Communications Workshops*, IEEE, pp. 432–437.
21. Ferrari, L., Rosi, A., Mamei, M. & Zambonelli, F. (2011). Extracting urban patterns from location-based social networks. In: *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks*, ACM, pp. 9–16.
22. Fire, M. & Elovici, Y. (2015). Data Mining of Online Genealogy Datasets for Revealing Lifespan Patterns in Human Population. *ACM Transactions on Intelligent Systems and Technology*, vol. 6, no 2, pp. 1–24.
23. Gangi, R.R., Rajesh, N.B., Sudhakar, N.P., Raviteja, B. & Rammohanarao, K. (2012). Tracking Objects, using RFID and Wireless Sensor Networks. *International Journal of Engineering Science & Advanced Technology*, vol. 2, no 3, pp. 513–517.
24. Gil-Clavel, S. & Zagheni, E. (2019). Demographic Differentials in Facebook Usage around the World, In: *Proceedings of the Thirteenth International AAAI Conference on Web and Social Media (ICWSM 2019)*, pp. 647–650.
25. Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S. & Brilliant, L. (2008). Detecting Influenza Epidemics Using Search Engine Query Data. *Nature*, vol. 457, no 7232, pp. 1012–1014.
26. Gutiérrez-Batista, K., Campaña, J.R., Vila, M.-A. & Martin-Bautista, M.J. (2018). Building a Contextual Dimension for OLAP Using Textual Data from Social Networks. *Expert Systems with Applications*, vol. 93, Issue C, pp. 118–133.

27. Henderson, J.V., Storeygard, A. & Weil, D.N. (2012). Measuring Economic Growth from Outer Space. *The American Economic Review*, vol. 102, no 2, pp. 994–1028.
28. Lazer, D.M., Kennedy, R., King, G. & Vespignani, A. (2014). The Parable of Google Flu: Traps in Big Data Analysis. *Science*, vol. 343, no 6176, pp. 1203–1205.
29. Loudcher, S., Jakawat, W., Soriano Morales, E.P. & Favre, C. (2015). Combining OLAP and Information Networks for Bibliographic Data Analysis: A Survey. *Scientometrics*, vol. 103, no 2, pp. 471–487.
30. Lyngstad, T.H. & Skardhamar, T. (2011). Nordic Register Data and Their Untapped Potential for Criminological Knowledge. *Crime and Justice*, vol. 40, no 1, pp. 613–645.
31. Mehr, H. (2017). Artificial Intelligence for Citizen Services and Government. Cambridge: Harvard Kennedy School. Available at: http://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf (accessed: 29 March 2019).
32. Noulas, A., Scellato, S., Mascolo, C. & Pontil, M. (2011). An Empirical Study of Geographic User Activity Patterns in Foursquare. In: *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM)*, vol. 11, pp. 570–573.
33. Poltavtseva, M. A. & Pechenkin, A.I. (2017). Intelligent Data Analysis in Decision Support Systems for Penetration Tests. *Automatic Control and Computer Sciences*, vol. 51, no 8, pp. 985–991.
34. Poulain, M. & Herm, A. (2013). Central Population Registers as a Source of Demographic Statistics in Europe. *Population*, vol. 68, no 2, pp. 183–212. Available at: DOI: <https://doi.org/10.3917/popu.1302.0215> (accessed: 26 October 2019).
35. Prins, K. (2017). Population Register Data, basis for the Netherlands Population Statistics. *Statistics Netherlands, Hague*. Available at: www.cbs.nl/-/media/_pdf/2017/38/population-register-data.pdf (accessed: 26 October 2019).
36. Redondi, A., Chirico, M., Borsani, L., Cesana, M. & Tagliasacchi, M. (2013). An Integrated System Based on Wireless Sensor Networks for Patient Monitoring, Localization And Tracking. *Ad Hoc Networks*, vol. 11, no 1, pp. 39–53.
37. Shryock, H.S., Siegel, J.S. & Larmon, E.A. (1980). *The Methods and Materials of Demography*. Department of Commerce, Bureau of the Census.
38. United Nations Global Pulse's projects. Available at: www.unglobalpulse.org/projects (accessed: 12 March 2019).
39. Weber, I. & State, B. (2017). Digital Demography. In: *Processing of the International World Wide Web Conference Committee (IW3C2)*. April 3–7, Perth, Australia, pp. 935–939.
40. Yildiz, D., Munson, J., Vitali, A. et al. (2017). Using Twitter data for demographic research, *Demographic Research*, no 37, Article 46, pp. 1477–1514.
41. Zagheni, E. (2017). *Data Science, Demography and Social Media: Challenges and Opportunities*. Available at: <https://pdfs.semanticscholar.org/presentation/> (accessed: 05 March 2019).
42. Zagheni, E. & Weber, I. (2015). Demographic Research with Non-Representative Internet Data. *International Journal of Manpower*, vol. 36, no 1, pp. 13–25.