



Большие данные для статистики населения и социальной статистики

В этом кратком статистическом обзоре рассматриваются источники больших данных, которые можно использовать для производства статистики населения и социальной статистики, а также представлены страновые примеры по использованию данных мобильных телефонов, данных о продажах, данных наблюдения Земли, социальных сетей и других онлайн-данных для получения статистики, связанной с перемещениями, статистики труда, статистики здравоохранения и показателей ЦУР.

Этот краткий обзор является частью серии выпусков ЭСКАТО по использованию нетрадиционных источников данных для официальной статистики.

Этот статистический обзор подготовлен следующими специалистами ЭСКАТО ООН: Ирина Берналь, консультант Статистического отдела, Джемма Ван Халдерен, директор Статистического отдела, Таня Сейерсен, статистик Статистического отдела; также ценный вклад внесли: Петра Нахмиас, начальник Секции народонаселения и социальной статистики Статистического отдела и Афсане Яздани, статистик Статистического отдела.

Что такое статистика населения и социальная статистика?

Статистика населения и социальная статистика подкрепляют политические решения, основанные на фактах, которые имеют огромное значение для нашего благополучия и качества жизни. От здравоохранения до образования, от занятости до преступности – статистика населения и социальная статистика охватывают все аспекты нашей жизни. И поскольку новые технологии экспоненциально увеличивают объем доступных данных и нашу способность анализировать и распространять их, вполне вероятно, что Национальным статистическим системам (НСС) потребуется производить информацию о все более и более широком круге социальных проблем с более высоким уровнем детализации.

В отличие от экономической статистики и статистики окружающей среды, не существует согласованной на глобальном уровне системы тем или вопросов, которые включались бы в статистику населения и социальную статистику. Статистику населения часто принято называть демографической статистикой и включать в нее статистику естественного движения населения, например, собираемую из систем регистрации актов гражданского состояния и естественного движения населения. На странице сайта Статистического отдела ООН определены четыре темы статистики населения и социальной статистики: гендерная статистика, статистика использования времени, статистика миграции и статистика инвалидности. На региональном уровне Комитет по статистике ЭСКАТО согласовал основной диапазон предметных областей статистики населения и социальной статистики, включающий 11 областей: население; здоровье; доход, богатство и расходы; занятость; образование и обучение; жилье и инфраструктура; информация и связь;

преступность и правосудие; семья и сообщество; культура и досуг; и управление¹.

Какие источники больших данных могут поддержать производство статистики населения и социальной статистики?

Традиционно статистика населения разрабатывается на основе систем регистрации актов гражданского состояния и естественного движения населения, переписей и других обследований лиц и домашних хозяйств. В большинстве стран перепись населения и жилищного фонда проводится каждые 10 лет, в других – нерегулярно. Перепись также используется для производства социальной статистики, но во многих странах ключевым источником данных для этой статистики обычно являются обследования домашних хозяйств. Однако как правило, существует задержка, часто значительная, в выпуске данных и статистических показателей по результатам переписей и обследований, тогда как для разработки политики обычно требуются своевременные сведения. Кроме того, кризисы, такие как последняя пандемия COVID-19, часто требуют данных в реальном времени, особенно о местонахождении, плотности и перемещении населения, при сохранении конфиденциальности и обеспечении защиты персональных данных.

Национальные статистические управления (НСУ) все больше полагаются на использование данных из регистров населения и других административных источников при производстве демографической статистики². Сами по себе, или в сочетании с другими соответствующими административными данными, регистры населения также могут быть полезны в качестве дополнения или замены традиционной переписи в различной степени, могут заменять или дополнять обследования или использоваться в качестве основы выборки.

Объединение административных данных с нетрадиционными источниками может предоставить более своевременные и детализированные данные и даже привести к новым выводам. Многие НСУ в Азиатско-Тихоокеанском регионе все чаще изучают нетрадиционные источники данных для получения дополнительных сведений для обследований и административных данных или для построения социальных и демографических показателей в случае отсутствия важных данных. Быстрое развитие современных технологий и их широкое принятие обществом создают цифровые следы, которые, если их переориентировать, при условии обеспечения защиты конфиденциальности и этичности, могут дать ценную, своевременную и детальную информацию о группах населения и их перемещениях.

Например, **данные мобильных телефонов** используются при производстве статистики, связанной с мобильностью, такой как миграция, ежедневные поездки на работу (маятниковая миграция), урбанизация и туризм. Можно различать данные с высоким разрешением, такие как данные сигнализации, и подробные записи о вызовах (CDR). Данные сигнализации предоставляют информацию об обновлениях местоположения мобильного устройства каждый раз, когда меняется вышка сотовой связи, и данные активного позиционирования, когда мобильный телефон пингуется, а его местоположение определяется с помощью технологии A-GPS. CDR содержат такую информацию, как время, продолжительность, номер телефона исходящего вызова, вызываемый номер и приблизительное местоположение коммуникаций о телефонном звонке или другой операции (например, текстового сообщения), которая проходит через устройство и регистрируется оператором мобильной сети.

¹ UN ESCAP, Core set of population and social statistics for Asia and the Pacific, E/ESCAP/CST(3)/5/Add.1, (Основной набор данных статистики населения и социальной статистики в регионе Азии и Тихого океана) [ссылка](#)

² UNESCAP Stats Brief, October 2020 (Issue no. 26): Population Registers: A Key Resource for Producing Vital Statistics (Регистры населения, основной ресурс для производства статистики естественного движения населения), [ссылка](#)

Целевая группа ООН по использованию данных мобильных телефонов, членами которой являются в частности, Грузия, Индонезия, Филиппины и Республика Корея, кроме других стран, разработала **Руководство по использованию данных мобильных телефонов для целей официальной статистики**³, в котором приводятся примеры использования данных мобильных телефонов для производства статистики населения, в том числе для статистики миграции и маятниковой миграции (ежедневных поездок на работу), для мониторинга занятости, а также дается информация о моделях партнерства с операторами связи⁴.

В качестве альтернативы данным оператора мобильной связи для отслеживания перемещений людей можно использовать **агрегированные геолокационные данные из мобильных приложений**. Согласие пользователя на то, чтобы «делиться» своей геолокацией делает эти данные более привлекательными для использования с целью оценки перемещений людей в тех случаях, когда согласие пользователей представляет собой серьезную проблему.

Другие примеры больших данных, используемых для статистики населения и социальной статистики, включают в себя **онлайн-вакансии** для получения информации о занятости, мобильности рабочей силы и рыночном спросе на рабочую силу, **данные наблюдения Земли**, чтобы дать представление о динамике урбанизации, **данные социальных сетей**, чтобы дать представление о плотности населения, мобильности и аспектах, связанных со здоровьем, а также другие **интернет-данные** с подробными демографическими характеристиками пользователей, которые могут дополнять и уточнять официальную статистику для поддержки разработки государственной политики по широкому кругу вопросов.

Пандемия COVID-19 подчеркнула необходимость своевременных данных о перемещениях населения для понимания моделей мобильности и оценки государственных мер по сдерживанию распространения вируса. Пробел в наличии такой информации побудил правительства и национальные статистические управления исследовать данные частного сектора, в основном от операторов мобильной связи и интернет-компаний, например, инициативу Facebook Data for Good⁴ (карты профилактики заболеваний, карты плотности населения, карты диапазона передвижений, исследование гендерного равенства в домохозяйстве) или разработки Google 'индекс мобильности' и онлайн-порталы вакансий⁵.

Как национальные статистические системы Азиатско-Тихоокеанского региона используют большие данные для производства статистики населения?

Многие исследования и эксперименты в области использования больших данных для целей производства статистики населения и социальной статистики были проведены и задокументированы академическими организациями, гражданским обществом и международными организациями, в том числе в странах Азиатско-Тихоокеанского региона. Однако имеется гораздо меньше задокументированных примеров экспериментов и случаев использования больших данных для производства официальной статистики населения и социальной статистики от НСУ.

В этом статистическом обзоре восполняется этот пробел и приводятся конкретные примеры проектов по использованию больших данных под руководством НСУ. Тематические исследования из восьми

³ UN Global Working Group on Big Data for Official Statistics, Handbook on the Use of Mobile Phone Data for Official Statistics, September 2019 (*Глобальная рабочая группа ООН по использованию больших данных в официальной статистике, Руководство по использованию данных мобильных телефонов для целей официальной статистики, сентябрь 2019 г.*), [ссылка](#)

⁴ Facebook, Data for Good (Facebook, инициатива «Данные во благо»), [ссылка](#)

⁵ Google, COVID-19 Community Mobility Reports (Google, COVID-19: Отчеты о передвижениях людей), [ссылка](#)

стран (Австралия, Индонезия, Грузия, Российская Федерация, Новая Зеландия, Республика Корея, Китай и Бангладеш) демонстрируют опыт НСУ в использовании больших данных для производства статистики населения и социальной статистики. Приводятся примеры из четырех областей статистики: статистика перемещений населения, включая маятниковую миграцию, внутреннюю миграцию и урбанизацию, а также модели перемещений во время COVID-19; статистика труда; статистика здравоохранения и показатели ЦУР. В этих тематических исследованиях изучаются несколько источников больших данных, таких как данные мобильных телефонов, данные о геолокации из мобильных приложений, данные о продажах, онлайн-данные из поисковых систем, социальных сетей, порталов вакансий и данные наблюдения Земли.

Эти тематические исследования дополняются информацией о других региональных и глобальных усилиях, обсуждением различных моделей партнерства в области данных с операторами связи и освещением основных проблем, препятствующих изучению и интеграции источников больших данных в производство статистики населения и социальной статистики.

Аналитическая информация, содержащаяся в этом статистическом обзоре, была получена в результате кабинетных исследований и консультаций с представителями НСУ.

1. Статистика, связанная с перемещением населения

Действия с мобильным телефоном, мобильные приложения с поддержкой геолокации, электронные транзакции и другие электронные или онлайн-действия обычно содержат отметки времени действия/операции и местонахождения пользователя или действия/операции. Таким образом, местоположение можно отслеживать с течением времени, выявляя модели перемещения отдельных пользователей и целых групп населения. Хотя доступ к индивидуальным данным и моделям перемещения без явного согласия пользователей во многих странах считается нарушением неприкосновенности частной жизни, агрегирование данных на определенном уровне, сохраняющее конфиденциальность информации, может выявить модели перемещений, достаточно важные для предоставления информации для разработки политики.

Сведения о перемещении могут предоставить информацию о ряде социальных и экономических тенденций, таких как маятниковая миграция и мобильность рабочей силы, миграция, урбанизация, туризм, а также перемещения после стихийных бедствий и пандемий, например, во время кризиса COVID-19.

В таких странах Азиатско-Тихоокеанского региона, как Австралия, Индонезия, Китай, Грузия, Республика Корея, Бангладеш и Новая Зеландия, НСУ активно изучают альтернативные источники данных для получения социальных и экономических показателей, связанных с перемещением населения, в частности данные от операторов мобильной связи, сторонних агрегаторов данных геолокации из мобильных приложений и данные о продажах. В то время как некоторые из национальных статистических управлений полагались на уже существовавшие партнерские отношения с поставщиками данных для понимания моделей перемещения во время пандемии COVID-19, другие организации, такие как программа a2i (Доступ к информации)⁶ в Бангладеш и Австралийское бюро статистики, разработали новые партнерские отношения с поставщиками данных или соглашения с частными компаниями во время пандемии для получения дополнительных своевременных результатов и реагирования на пандемию.

Кроме использования данных мобильных телефонов, существуют и другие источники данных, которые могут помочь определить модели мобильности людей. Например, опыт Статистического управления Новой Зеландии раскрывает потенциал данных о продажах для отображения моделей

⁶ a2i (Доступ к информации) — это специальная программа, реализуемая Отделом ИКТ Министерства ИКТ Бангладеш и поддерживаемая ПРООН, которая стимулирует инновации в области государственных услуг, ориентированных на граждан, и доступ к информации государственного сектора.

мобильности, особенно после стихийных бедствий, а Австралийское бюро статистики и Национальное бюро статистики Китая использовали агрегированные данные о геолокации из мобильных приложений, чтобы понять модели передвижений во время пандемии COVID-19 и тренды урбанизации.

В следующих тематических исследованиях рассматривается опыт отдельных стран в получении доступа и использовании больших данных для статистики, связанной с перемещением. В этом разделе рассматривается использование данных мобильных телефонов, данных геолокации мобильных приложений и данных наблюдения Земли для получения статистических данных о маятниковой миграции, внутренней миграции, урбанизации и перемещениях, связанных с COVID-19.

1.1 Статистика маятниковой миграции

Статистика маятниковой миграции включает информацию о поездках людей на работу и обратно, такую как расстояние, продолжительность поездки и виды транспорта. Разработчики политики и планировщики используют эту статистику для принятия решений о транспортной инфраструктуре. В то время как традиционно статистические данные о маятниковой миграции разрабатываются на основе данных обследований, НСУ постепенно тестируют источники больших данных. Данные мобильных телефонов и транспортные данные являются двумя наиболее часто используемыми источниками для статистики маятниковой миграции.

Статистика маятниковой миграции с использованием данных мобильного позиционирования в Индонезии

Центральное статистическое управление Индонезии является одним из самых активных статистических управлений в регионе в изучении потенциала различных источников больших данных для целей официальной статистики. В области статистики населения ЦСУ Индонезии осуществило несколько пилотных проектов с использованием данных мобильного позиционирования.

Рисунок 1. Использование данных мобильного позиционирования Telkomsel для оценки показателей перемещения людей в Индонезии



Источник: Центральное статистическое управление Индонезии

Центральное статистическое управление Индонезии использует данные мобильного позиционирования в двух областях. Во-первых, это статистика перемещения населения, такая как статистика въездов в страну на границе, статистика внутреннего туризма, национальная статистика туризма, статистика регулярных поездок на работу (маятниковая миграция) и статистика временного перемещения лиц. Во-вторых, статистические данные в пунктах наблюдения, такие как количество посетителей Азиатских игр 2018 г. и Ежегодного мероприятия МВФ 2018 г., статистика расширения столичного региона Большого Бандунга в 2019 г. и статистика посещений восьми «самых приоритетных» туристических направлений, о возрождении которых после пандемии Министерство туризма приняло решение в 2020 г.

При 355 миллионах активных пользователей мобильными телефонами на 268 миллионов жителей Индонезии, что эквивалентно 132% уровню проникновения мобильной связи, данные мобильных телефонов могут служить альтернативным источником данных, дополняющим и уточняющим существующую статистику населения.

Чтобы изучить многообещающий потенциал данных мобильных сетей, Статистическое управление Индонезии заключило партнерское соглашение с Telkomsel, крупнейшим оператором мобильной связи в стране с более чем 170 миллионами клиентов, для разработки статистики перемещений и миграции. Мобильные данные от каждого абонента получают на основе информации об активности телефона в форме подробной записи вызовов (CDR) и сервиса определения местоположения (LBS), например, подробной записи интернет-протокола (IPDR), обновления области местоположения, событий передачи обслуживания, включений устройства и т. д. Собранные данные используются для формирования информации о продолжительности поездки, расстоянии поездки между регионами и видах транспорта, что, в свою очередь, позволяет различать поездки на работу, внутренний туризм и миграцию, как показано на рисунке 1.

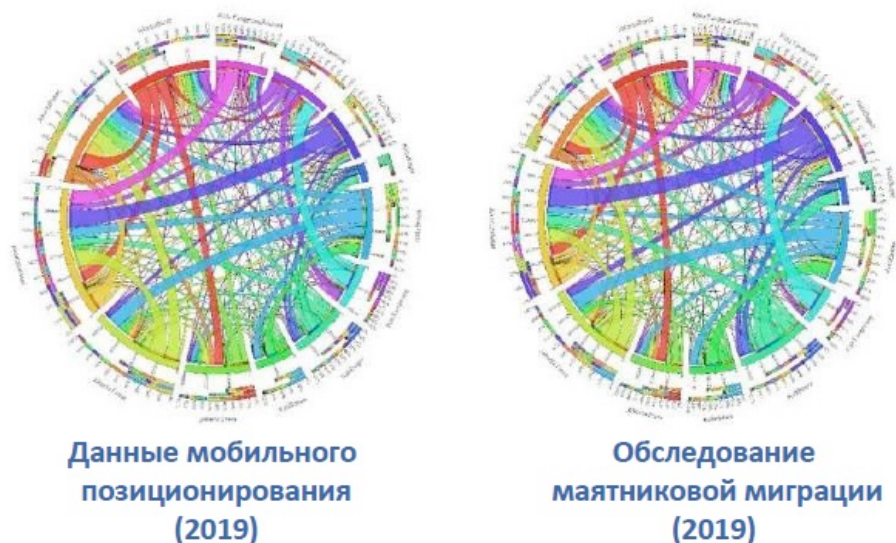
Обработка данных основана на методе «песочницы». Данные мобильного позиционирования хранятся и обрабатываются в системе Telkomsel без передачи данных в статистическое управление. Однако оно может получить доступ к 5% выборке данных мобильного позиционирования в базе данных, используя право доступа от оператора мобильной связи, для исследования данных и разработки алгоритмов. На основе этих данных Центральное статистическое управление дает определение и концепцию потока, определяя лиц, совершающих ежедневные поездки на работу, туристов и мигрантов. Затем оператор мобильной связи разрабатывает и применяет сценарий (скрипт) ко всему набору данных и предоставляет результаты табуляции в Центральное статистическое управление. При поддержке компании Positium методология и алгоритмы были усовершенствованы с течением времени после принятия дополнительных мер для обеспечения качества⁷. Дневники путешествий добровольцев, подтверждающие перемещения, использовались для проверки результатов данных мобильного позиционирования.

По сравнению с результатами традиционных обследований с целью получения данных о маятниковой миграции, данные мобильного позиционирования обеспечили похожие результаты, (как показано на рисунке 2), но с рядом преимуществ. Использование данных мобильного позиционирования оказывается экономически выгодным, так как требует лишь одной восьмой бюджета традиционного обследования и приводит к увеличению охвата, поскольку обеспечивает большую степень детализации на уровне муниципалитетов и подрайонов, чем традиционное обследование. Использование данных мобильного позиционирования также предоставляет более частые, своевременные и актуальные статистические данные и может развертываться ежемесячно, а

⁷ Siim Esko, “Before working with MPD” (*Прежде чем работать с данными мобильного позиционирования*) презентация на семинаре по использованию данных мобильных телефонов, 11 июня 2019 г., [ссылка](#)

не раз в четыре года, и только в разных городских агломерациях⁸.

Рисунок 2. Результаты использования данных мобильного позиционирования по сравнению с результатами традиционного обследования в Индонезии



Источник: Центральное статистическое управление Индонезии

ЦСУ Индонезии также использовало данные мобильного позиционирования для улучшения статистики перемещения населения на уровне подрайонов⁹. В 2019 г. оно ввело понятие статистического столичного района (MSA) в качестве альтернативы определению городских районов (МА), используемому Министерством аграрных дел и территориального планирования. Данные мобильного позиционирования использовались в пилотной области Чекунган Бандунг для получения показателей маятниковой миграции между столичным городом и прилегающими районами на уровне подрайона в области Чекунган Бандунг¹⁰. Среди преимуществ использования данных мобильного позиционирования, выявленных в этом пилотном проекте, можно назвать более широкий охват и уровень детализации данных на уровне подрайонов, снижение затрат по сравнению с традиционными обследованиями, своевременность и увеличенную частоту получения данных¹¹.

Среди основных проблем, которые стоят перед ЦСУ Индонезии при использовании данных мобильного позиционирования, можно назвать необходимость укрепления доверия и сотрудничества с частным сектором, а также обеспечение доступа к данным и защиту конфиденциальности. Как и любое статистическое управление, Центральное статистическое

⁸ Edi Setiawan, "Big data for social and demographic statistics," (*Большие данные для социальной и демографической статистики*), презентация на мероприятии ЭСКАТО «Кафе для статистиков» 2 ноября 2020 г., [ссылка](#)

⁹ Noviyanti, Isnaeni, Prabawa, Panca D., Sari, Dwi Puspita, Koswara, Ade, Lestari, Titi Kanti, Fahyuananto, M. Hanif, Setiawan, Edi, Towards big data as official statistics: Case study of the use of mobile positioning data to delineate metropolitan areas in Indonesia (*На пути к использованию больших данных в официальной статистике. Тематическое исследование использования данных мобильного позиционирования для разграничения городских агломераций в Индонезии*), Statistical Journal of the IAOS (*Статистический журнал IAOS*), vol. 36, no. 4, pp. 943-954, 2020, [ссылка](#)

¹⁰ Panca D. Prabawa¹, Hamim T. Soblia, Yudi F. Amin, Winida Albertha, Edi Setiawan, "The Use of Mobile Positioning Data (MPD) to Delineate Metropolitan Area in Indonesia: Case Study in Cekungan Bandung" (*Использование данных мобильного позиционирования для разграничения городских агломераций в Индонезии: тематическое исследование в Чекунган Бандунг*), Азиатско-тихоокеанская неделя по статистике, июнь 2020 г., [ссылка](#)

¹¹ Edi Setiawan, presentation "Big data for social and demographic statistics," (*презентация «Большие данные для социальной и демографической статистики»*), мероприятие ЭСКАТО «Кафе для статистиков» 2 ноября, 2020г., [ссылка](#)

управление Индонезии стремится получить доступ к «сырым» или первичным данным для анализа. С другой стороны, оператор мобильной связи соблюдает правила и принципы корпоративной конфиденциальности в отсутствие национального закона, регулирующего обмен большими данными между частными компаниями и органами государственного управления. В такой ситуации ЦСУ начало развивать доверительные отношения с оператором мобильной связи, подписав коммерческий контракт и Меморандум о взаимопонимании, тесно сотрудничая с ним по вопросам разработки методологии и позволив специалистам по данным из ЦСУ получать доступ и обрабатывать выборочные данные в сетевой «песочнице». Устойчивость доступа к данным остается проблемой, поскольку контракт с оператором мобильной связи пересматривается на ежегодной основе, и на этот конкретный источник данных должен выделяться годовой бюджет. Тем не менее, при поддержке Министерства планирования, ЦСУ изучает возможность сотрудничества с другими операторами мобильной связи.

Отслеживание динамики перемещения населения с использованием мобильного позиционирования и транспортных данных в Республике Корея

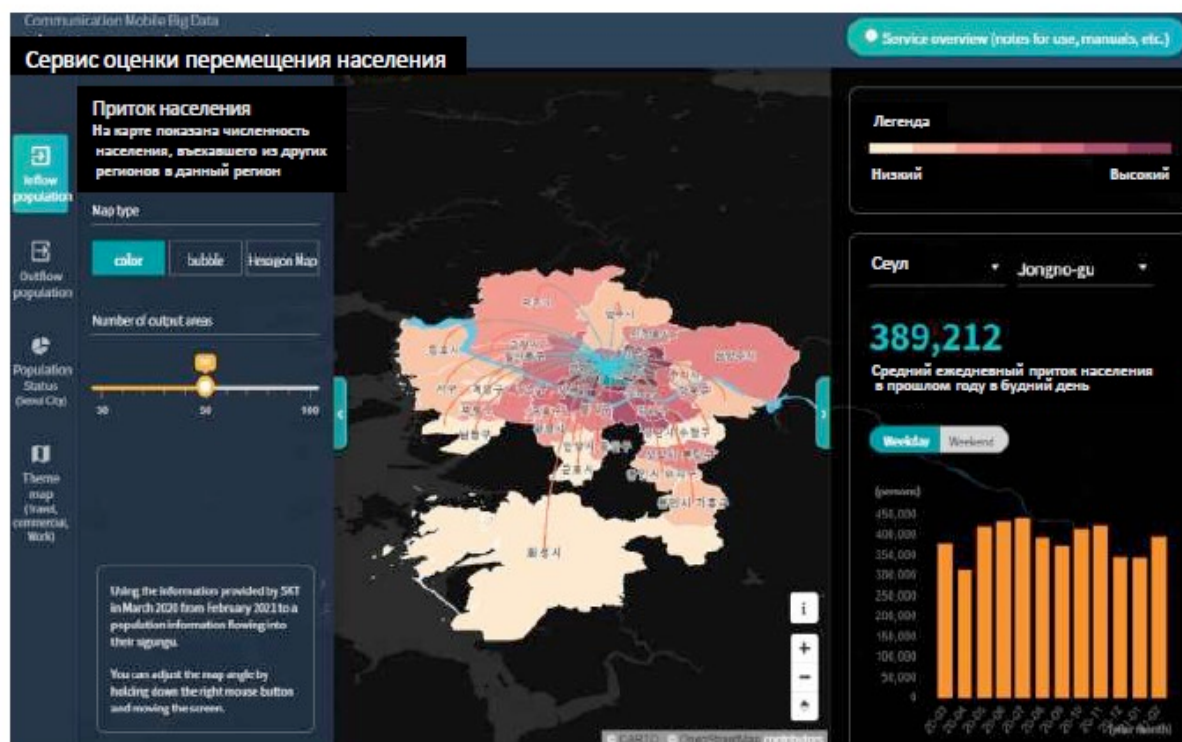
Национальное статистическое управление Республики Корея (НСУ Кореи) провело свой первый анализ перемещения населения на основе данных мобильных телефонов в 2017 г. НСУ Кореи образовало партнерство с корейским оператором связи – КТ, подписав меморандум о взаимопонимании. Анализ проводился на данных НСУ Кореи и КТ на Платформе анализа больших данных КТ в одном из филиалов компании. Этот пилотный проект привел к установлению институциональных стандартов анализа данных мобильных телефонов.

С тех пор нормативно-правовая база претерпела изменения: были приняты три новых закона, связанных с данными. Однако еще предстоит установить четкие руководящие принципы в отношении интеграции различных источников данных, в том числе из частного сектора.

После первого пилотного проекта в мае 2019 г. между НСУ Кореи и SK Telecom (SKT) – еще одним корейским оператором связи – был подписан меморандум о взаимопонимании. Из-за высокой конкуренции на рынке телекоммуникаций НСУ Кореи могло сотрудничать только с одним оператором мобильной связи для проведения исследования, и в качестве партнера был выбран оператор SKT. НСУ Кореи проводит совместные исследования с SKT в соответствии с более строгими правилами защиты конфиденциальности данных, предоставляет рекомендации по статистическому анализу и выполняет валидацию данных, в то время как SKT проводит анализ данных и делится результатами со своим партнером. Партнеры не обмениваются никакими другими данными, кроме результатов исследований.

НСУ Кореи осуществило серию пилотных проектов по анализу перемещения групп населения на основе данных мобильного позиционирования, проведя картирование перемещения работников, а также времени, затрачиваемого на поездки на работу и домой, и времени, проведенного на работе. На основе этих данных НСУ предоставляет сервис визуализации данных о перемещениях населения (притоки и оттоки) на 3D-картах по каждому городу, округу и району в стране в группировке по будним дням, выходным дням и по месяцам. На карте также показана оценка фактической численности жителей Сеула, полученная с использованием данных мобильного позиционирования, а также число зарегистрированных жителей на основе административной статистики. Сервис был расширен за счет включения демографического анализа для каждого делового района Сеула.

Рисунок 3. Оценка притока и оттока населения с использованием данных мобильных телефонов в Республике Корея



Источник: HCU Кореи, <http://data.kostat.go.kr>

Помимо работы с данными мобильного позиционирования, HCU Кореи сотрудничает с государственной организацией, хранящей данные о транспорте, в рамках проекта, который начнется в 2021 г. HCU будет применять ту же модель партнерства, что и с SKT. Точно так же HCU предоставит темы для анализа, а конкретная методология и визуализация выходных данных и методы обработки данных будут рассмотрены государственными и частными организациями.

1.2 Внутренняя миграция и урбанизация

Внутренняя миграция определяется как изменение лицом обычного места проживания в пределах страны. На нее может влиять целый ряд причин, таких как поиск лучших экономических возможностей, бегство от насилия или политических репрессий, стихийные бедствия и многие другие. Многие источники больших данных могут свидетельствовать о наличии внутренней миграции. Например, веб-сайты профессиональных связей и контактов или социальные сети, где пользователи обновляют свое местоположение, могут предоставлять информацию о перемещении пользователя внутри страны или за ее пределы. Однако эти источники данных в основном изучаются международными партнерами по развитию. С другой стороны, данные мобильных телефонов остаются наиболее изученным источником больших данных для оценки миграции. Другой пример — данные о продажах, которые могут отражать внутреннюю миграцию после стихийных бедствий.

Оценка численности внутренних мигрантов с использованием данных мобильного позиционирования в Грузии

Как и Центральное статистическое управление Индонезии, **Национальное статистическое управление Грузии** (HCU Грузии) занимается изучением использования данных мобильных телефонов для разработки статистики перемещений населения и туризма. HCU Грузии проводит обследования туризма, но информации о внутренней миграции не хватает, а регистр населения еще

не создан. Использование данных мобильных телефонов для получения миграционной статистики является частью более крупного проекта, запущенного в 2018 г. в партнерстве с СОООН, МСЭ, Национальной комиссией по коммуникациям Грузии (GNCC) – регулятором электросвязи, и компанией Positium – эстонской компанией, специализирующейся на статистике туризма, статистике миграции и статистике информационного общества.¹²

В отличие от Центрального статистического управления Индонезии и нескольких других НСУ, которые напрямую сотрудничали с операторами мобильной связи для получения доступа к данным или информации, НСУ Грузии сотрудничало с регулирующим органом – Национальной комиссией по коммуникациям Грузии (GNCC). GNCC договорилась о доступе к данным трех операторов мобильной связи: Magticom, Geocell и Mobitel, указав формат и время для запросов данных. GNCC получает бесплатный доступ к анонимизированным данным мобильных телефонов в соответствии с законодательством. Кроме того, Служба по защите персональных данных проводит мероприятия по анонимизации, хранению и предоставлению доступа к данным. НСУ Грузии не имеет доступа к первичным данным. Вместо этого оно сотрудничает с GNCC для разработки методологии, которая будет применяться к анонимизированным данным, которые GNCC собирает у операторов мобильной связи. Хотя методология была разработана, ее тестирование на реальных данных оператора мобильной связи было приостановлено до тех пор, пока GNCC не решит проблемы ИТ-инфраструктуры, связанные с хранением данных, и до момента полного соответствия национальному законодательству. Проект все еще продолжается, и он выявляет недостатки, связанные с технологией и соблюдением законодательства, которые могут возникнуть.

Оценка перемещений населения, связанных с природными катастрофами, с использованием данных о продажах в Новой Зеландии

После землетрясения в Каикоуре в Новой Зеландии в ноябре 2016 г. Статистическое управление Новой Зеландии изучило данные о налоге с продаж для оценки перемещений людей¹³. На основе налоговых данных статистическое управление оценило рост и снижение продаж за период с октября по ноябрь 2015 г. и с октября по ноябрь 2016 г. Результаты анализа позволили получить представление о рассредоточении людей после землетрясения в Каикоуре в ноябре 2016 г., что проявилось в транзакционной нагрузке по разным маршрутам по всей Новой Зеландии.

¹² Paata Shavishvili, presentation “Cooperation with UNSD in the field of migration statistics”, 2019 (*презентация «Сотрудничество с СОООН в области статистики миграции», 2019 г.*), [ссылка](#)

¹³ Gary Dunnet, presentation “Improving economic statistics with big data” (*презентация «Улучшение экономической статистики при помощи использования больших данных»*), мероприятие ЭСКАТО «Кафе для статистиков» 30 ноября 2020 г., [ссылка](#)

Рисунок 4. Рост и снижение продаж в период октябрь-ноябрь 2015 г. и в тот же период в 2016 г., указывающие на перемещение населения после землетрясения в Каикоре



Источник: Статистическое управление Новой Зеландии

Оценка характеристик проживающего населения и мониторинг миграции и урбанизации в Китае

Национальное бюро статистики Китая (НБС Китая) сотрудничает с частными компаниями при проведении исследований и при изучении применения методов больших данных для статистики населения. Поскольку в середине 2019 г. в Китае было зарегистрировано 1,59 миллиарда мобильных пользователей, данные мобильных телефонов могут способствовать динамическому мониторингу перемещений населения.

НБС Китая использует данные сигнализации и данные мобильных приложений в мобильных телефонах только для оценки миграции населения и мониторинга в некоторых регионах¹⁴. Данные мобильных приложений собираются с помощью комплекта для разработки программного обеспечения (SDK), который собирает данные пользователей, то есть информацию о местоположении устройства и о времени. Хотя оба источника данных предоставляют информацию о геолокации, они имеют некоторые различия, связанные с зоной покрытия, режимом определения местоположения и характеристиками позиционирования.

Например, данные сигнализации собираются с базовых станций, а данные мобильных приложений — по Wi-Fi и GPS, что приводит к широкому охвату с небольшой разницей по областям для данных сигнализации и узкому охвату с большими различиями по областям между городской и сельской местностью для данных мобильных приложений.

Данные мобильной сигнализации обеспечивают обновление местоположения мобильных

¹⁴ Департамент статистики населения и занятости, Национальное бюро статистики Китая, презентация «Применение больших данных в демографической статистике», октябрь, 2019 г. [ссылка](#)

пользователей каждый раз, когда меняется вышка сотовой связи. Эти данные используются для мониторинга времени проживания, местоположения и продолжительности проживания мобильных пользователей в каждом административном районе через регулярные промежутки времени. НБС анализирует увеличение и уменьшение числа мобильных пользователей по сравнению с базисным периодом.

НБС использует геолокационные данные из мобильных приложений пользователей для динамического мониторинга миграции населения и урбанизации. На основе данных мобильных приложений НБС Китая оценивает общую численность населения и моделирует прогноз численности населения. Количество мобильных устройств служит основой для расчета, при этом отфильтровываются случаи наличия нескольких устройств у одного человека, учитывается доля рынка приложений и доля пожилых людей без смартфонов, а затем рассчитывается доля взрослых и детей. Эти данные также используются для подсчета ежемесячного постоянного населения и анализа потока пользователей между определенным муниципалитетом в Китае и различными провинциями и городами. Также на основе геолокационных данных и границ городских и сельских районов анализируются распределение и перемещение городского и сельского населения. На этом примере мы видим, что НБС использует большие данные для динамического мониторинга миграции населения по регионам страны, а также выявления тенденций урбанизации.

1.3 Оценка моделей перемещений во время пандемии COVID-19

Реагирование на пандемию COVID-19 потребовало наличия своевременных и дезагрегированных данных. Несколько правительств и национальных статистических управлений в регионе, в таких странах как Индонезия, Республика Корея и Новая Зеландия, опирались на существующие партнерские отношения с частным сектором, в первую очередь с операторами мобильной связи, для анализа перемещений населения и прогнозирования распространения вируса и очагов инфекции. Другие, например, программа a2i (Доступ к информации) в Бангладеш, договорились о доступе к данным оператора мобильной связи в начале пандемии. Иногда статистические управления изучали другие данные, связанные с местоположением, например Австралийское бюро статистики использовало агрегированные геолокационные данные из приложений на мобильных устройствах.

Понимание движения населения до и после COVID-19 в Республике Корея и Новой Зеландии с помощью данных мобильных телефонов

Национальное статистическое управление Республики Корея еженедельно проводит *анализ движения населения до и после COVID-19* с февраля 2020 г. на основе данных оператора мобильной связи¹⁵. Смещение в сторону пользователей SKT (телекоммуникационный оператор, с которым НСУ Кореи сотрудничает в этом исследовании, и чья доля на внутреннем рынке мобильной связи составляет около 42%) при анализе перемещений населения было скорректировано с использованием данных переписи путем применения взвешивания по регионам, полу и возрасту. Был проведен анализ потоков населения по полу, возрасту, типу местности, региону и рабочим/выходным дням недели для выявления изменений в перемещениях населения, вызванных COVID-19.

Data Ventures, дочерняя компания **Статистического управления Новой Зеландии**, сотрудничает с крупными телекоммуникационными компаниями страны и получает агрегированные и анонимизированные данные мобильных телефонов для анализа. В ситуации пандемии COVID-19 компания Data Ventures подготовила *Доклад о влиянии COVID-19 на население центральных деловых*

¹⁵ [Ссылка](#)

районов (CBD) местных советов¹⁶. В докладе представлены изменения в перемещении населения для восьми центральных деловых районов по сравнению с предыдущим месяцем и до COVID-19. Компания также регулярно публикует *отчет о населении в условиях пандемии COVID-19*¹⁷ и делится сценарием для оценки степени перемещения внутри региональных советов или по таким категориям, как рекреация, розничная торговля и рабочие места, на основе анонимизированных данных мобильных телефонов в общедоступном хранилище¹⁸. Эти телекоммуникационные данные также попали в статистику международных поездок, подготовленную Статистическим управлением Новой Зеландии, что стало частью дополнительной информации в период пандемии COVID-19 об оценках нахождения по регионам международных посетителей в Новой Зеландии. Предполагаемая разбивка числа международных посетителей по регионам публикуется еженедельно в рамках регулярного предварительного выпуска.

Данные из интернета для оценки перемещений, связанных с COVID-19, в Индонезии

Центральное статистическое управление Индонезии использовало несколько онлайн-источников данных, таких как Отчет о мобильности сообщества (Community Mobility Report) компании Google, отчеты о тенденциях мобильности (Mobility Trend Report) компании Apple, систему отслеживания полетов и информацию о ночном освещении, для оценки изменений в перемещениях населения и экономической активности во время пандемии COVID-19.

Изменения в перемещении населения отслеживались с помощью отчета Google о мобильности сообщества. Данные Google о перемещениях предоставили информацию о новых моделях мобильности на работе и дома, а также в сфере розничной торговли и отдыха, продуктовых магазинов и аптек, парков и транзитных перевозок, а также об изменениях по сравнению с уровнем до пандемии. Тепловая карта мобильности населения для секторов розничной торговли и рекреации представлена на рисунке 5. Активность воздушного транспорта отслеживалась с помощью системы отслеживания полетов, которая зафиксировала существенное снижение активности воздушного транспорта в пяти самых загруженных аэропортах во время пандемии.

Также данные мобильного позиционирования использовались для оценки внутренней миграции и ежедневных поездок на работу до и после пандемии COVID-19¹⁹, определяя тенденции мобильности в мегаполисах во время пандемии.

Использование агрегированных данных геолокации из мобильных приложений для понимания изменений в перемещениях населения, деловой активности и поездках в Мельбурне (Австралия)

В отличие от национальных статистических управлений, использующих данные мобильных телефонов, такие как данные сигнализации или подробные записи вызовов (CDR), для отслеживания перемещений населения во время пандемии COVID-19, **Австралийское бюро статистики (АБС)** собирало геолокационные данные мобильных устройств от поставщиков агрегированных деидентифицированных данных о мобильности, чтобы понять изменения в перемещениях населения, деловой активности и количестве поездок. Цель состояла в том, чтобы определить области тесных контактов и измерить пешеходный трафик между двумя интересующими точками, включая предприятия, чтобы понять уровень активности.

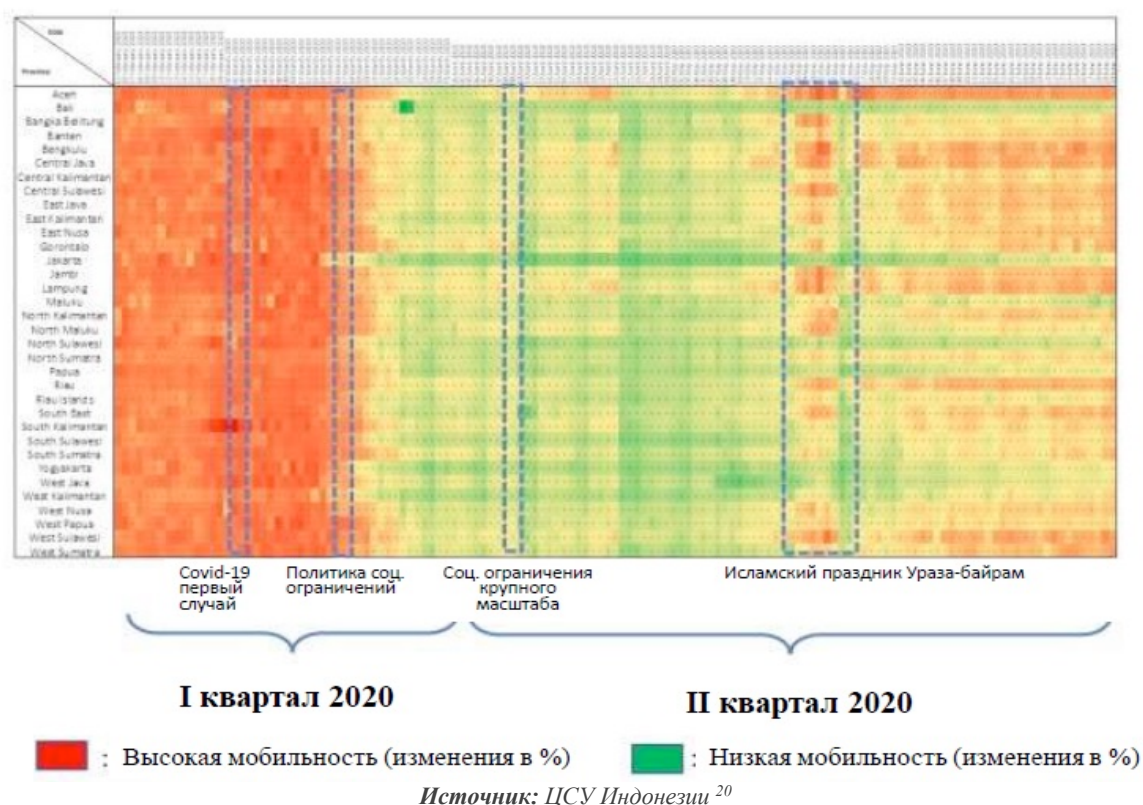
¹⁶ Data Ventures, "COVID-19 Impact on Local Councils' CBD Population," (*Влияние COVID-19 на население центральных деловых районов (CBD) местных советов*), October 12, 2020, [ссылка](#)

¹⁷ Data Ventures, COVID-19 Population Report, 6 July 2020, (*Отчет о населении в условиях пандемии COVID-19, 6 июля, 2020 г.*) [ссылка](#)

¹⁸ GitHub, dativenturesnz/mobility-index, Индекс мобильности, [ссылка](#)

¹⁹ [ссылка](#)

Рисунок 5. Тепловая карта перемещения людей между 34 провинциями для целей розничной торговли и рекреации в Индонезии



Анонимизированные выборочные данные для Мельбурна (Австралия) были получены по коммерческому соглашению с местной частной компанией, которая собирала такие данные. Хотя данные были агрегированы и не позволяли отслеживать местоположение отдельных лиц, они давали ценную информацию о перемещениях населения во время пандемии. Хотя АБС имело опыт анализа агрегированных и анонимизированных данных CDR при выполнении пилотного проекта 2016 г., в ситуации COVID-19 оно предпочло иметь доступ к анонимизированным данным о перемещениях, которые были переданы и собраны с явного согласия пользователей. АБС также изучило другие источники данных из государственных реестров и частного сектора для получения информации для целей социальной и экономической статистики в ситуации с COVID-19 и открыто информировало о различных источниках, которые были использованы²¹.

Выявление зон высокого риска COVID-19 на основе плотности населения и мобильности в Бангладеш

В Бангладеш группа a2i возглавляет работу по инновациям в области данных в органах государственного управления и делает это во время пандемии COVID-19. Признавая потребность в своевременных данных, привязанных к местности, для быстрой и целенаправленной выработки ответных мер и отслеживания заражения во время пандемии, a2i обратилась за новыми данными к частному сектору и государственным учреждениям. Были собраны данные о синдромах, о которых сообщали сами инфицированные лица, и проанализированы подробные записи вызовов (CDR) от крупнейшего оператора мобильной связи в Бангладеш, имеющего более 64 миллионов абонентов,

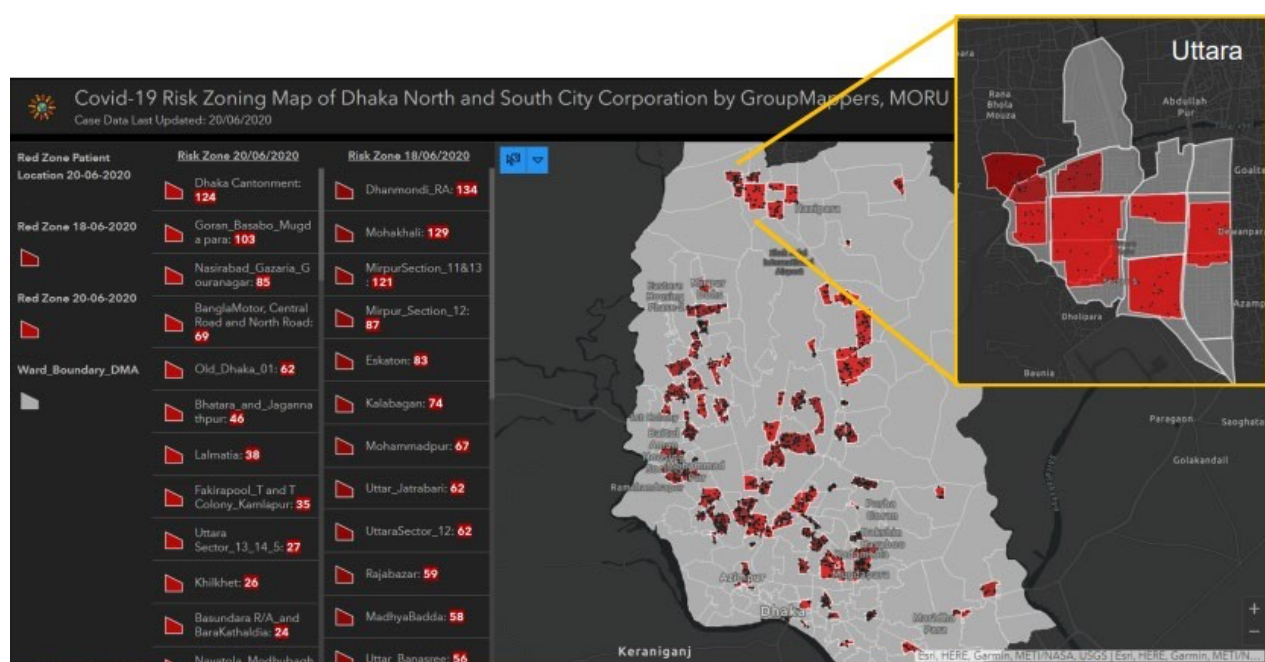
²⁰ Rendra Achyunda Anugrah Putra, Silvia Arini, Measuring the Economics of a Pandemic: How People Mobility depict Economics: AsEvidence of People's Mobility Data towards Economic Activities (*Измерение экономики пандемии: как мобильность людей отражает экономику: данные о мобильности людей в качестве свидетельств об экономической деятельности*, [ссылка](#))

²¹ Австралийское бюро статистики, "АБС отвечает на вызовы COVID-19", [ссылка](#)

чтобы отслеживать распространение эпидемии почти в режиме реального времени²².

Собранные со слов пациентов данные о синдроме (кашель, лихорадка и одышка), а также информация о контактах с кем-то с симптомами или с положительным результатом теста на COVID-19 были собраны из нескольких источников: 1) номер горячей линии, оснащенный интерактивным голосовым ответом (IVR) система Национальной телефонной линии и горячей линии здравоохранения, 2) несколько интернет-приложений и мобильных приложений и 3) система обмена сообщениями на основе неструктурированных дополнительных служебных данных (USSD). Ответы были геолоцированы на основе ближайшей вышки сотовой связи респондента и картированы на уровне *упазилы* (подрайона). Министерство электросвязи упростило доступ к анонимизированным данным от оператора мобильной связи.

Рисунок 6. Карта зонирования COVID-19 для Дакки, Бангладеш, с указанием горячих точек



Источники: Ramiz Uddin, ПРООН

Для анализа данных была создана Целевая группа, в которую вошли представители Министерства здравоохранения и a2i. Специалисты по данным и эпидемиологи из нескольких ведущих университетов оказали поддержку Целевой группе в создании карт с классификацией зон риска по всей стране на уровне районов и *упазил* (подрайонов). Несмотря на успех в доступе к данным и создании карт рисков в рекордно короткие сроки, группа a2i осознает необходимость увеличения внутренних мощностей, улучшения управления данными и безопасности данных, защиты конфиденциальности, а также усиления партнерства в области данных.

2. Статистика труда

Статистика труда охватывает статистические данные о рабочих местах и безработице, и в дополнение к оценке предложения рабочей силы она предоставляет данные об условиях труда, бедности и неравенстве, конкурентоспособности и производственных отношениях. На глобальном уровне Международная организация труда является ведущим источником статистических данных о

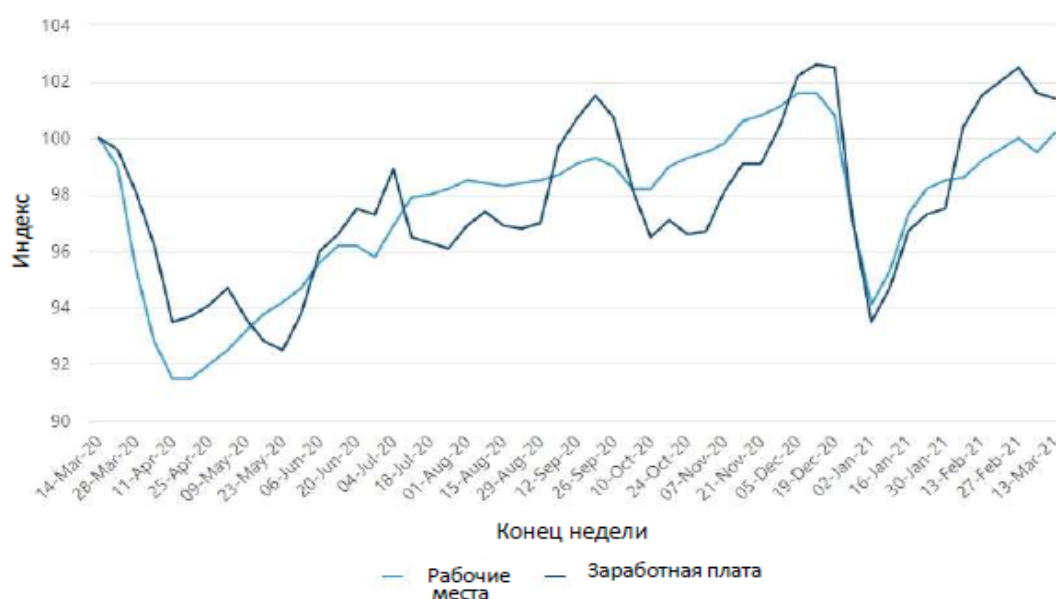
²² Ramiz Uddin, "Achieving the SDGs in a time of COVID-19: Bangladesh Perspective" (презентация «Достижение ЦУР во времена COVID-19: взгляд Бангладеш»), 6-я международная конференция по использованию больших данных для официальной статистики, [ссылка](#)

труде. На страновом уровне НСУ обычно собирают статистику труда на основе отчетов работодателей, данных налоговых органов, переписи населения и обследований домашних хозяйств и предприятий. Во время пандемии COVID-19 некоторые НСУ использовали как административные, так и интернет-данные, чтобы понять динамику трудовых ресурсов во время кризиса.

Оценка рабочих мест и оплаты труда в Австралии во время COVID-19 с использованием данных о заработной плате

Австралийское бюро статистики (АБС) имеет доступ к данным частного сектора, таким как данные о финансовых транзакциях Австралийской банковской ассоциации, чтобы оценить экономическое влияние COVID-19 на экономику. Тем не менее, чтобы понять динамику рабочих мест и заработной платы за тот же период, АБС изучило административные большие данные из системы Single Touch Payroll (STP) Австралийского налогового управления (АТО). Поскольку кризис требовал быстрой и надежной информации о событиях на рынке труда, АБС совместно с Налоговой службой Австралии разработали экспресс-индекс, представленный на рис. 7, который в настоящее время совершенствуется. Этот набор данных предоставляет новую оперативную информацию об изменениях в общем количестве рабочих мест и общей сумме заработной платы, выплачиваемой всеми предприятиями-работодателями, которые отчитываются перед налоговой службой через систему STP. Доступ к этому источнику данных был ускорен во время пандемии. Однако после кризиса процедуры обмена данными потребуют формализации.

Рисунок 7. Рабочие места и общая сумма заработной платы в Австралии, проиндексировано к неделе, закончившейся 14 марта 2020



Источник: Бюро статистики Австралии

Оценка показателей занятости с использованием онлайн-данных в Индонезии

Во время пандемии COVID-19 Центральное статистическое управление Индонезии проанализировало данные онлайн-объявлений о вакансиях на портале вакансий www.jobs.id, чтобы оценить ситуацию с занятостью. Онлайн-данные показали снижение количества онлайн-объявлений о вакансиях во всех секторах в период с января по май 2020 г.

3. Статистика здравоохранения

Статистика здравоохранения включает информацию о смертности, заболеваемости, системах здравоохранения, охвате услугами здравоохранения и факторах риска. Традиционно медицинские статистические данные собираются из медицинских карт, систем эпидемиологического надзора, больниц и обследований домохозяйств. Хотя большие данные уже широко используются в системе здравоохранения для продвижения исследований, определения новых лекарств и улучшения лечения и диагностики пациентов, существует также потенциал использования больших данных для получения статистики здравоохранения. В регионе Центральное статистическое управление Индонезии запустило пилотный проект по изучению использования данных социальных сетей для получения статистики психического здоровья.

Использование данных социальных сетей для получения статистики психического здоровья в Индонезии

Центральное статистическое управление Индонезии (ЦСУ Индонезии) использовало неструктурированные данные Twitter для эксперимента по получению статистики психического здоровья на основе данных из социальных сетей²³. С 1 января по 30 апреля 2020 г. ЦСУ осуществило пилотный проект, собирая данные через Twitter API. Вручную были определены ключевые слова, связанные с психическим здоровьем, и сообщения в Twitter разделены на четыре категории: депрессия, стресс, генерализованное тревожное расстройство и распространенные психические расстройства. Затем к новым данным были применены алгоритмы машинного обучения. Поскольку пилотный проект совпал с началом пандемии COVID-19, результаты показали всплеск количества сообщений в Twitter, содержащих интересующие ключевые слова. В рамках этого пилотного проекта ЦСУ удалось определить потенциально релевантный контент для показателей психического здоровья. В качестве следующего шага ЦСУ рассматривает возможность изучения других платформ социальных сетей, таких как Facebook, YouTube и Instagram, для более глубокого анализа эмоциональной окраски сообщений.

4. Показатели ЦУР

Многие НСУ в регионе изучают источники больших данных для восполнения пробелов в данных официальной статистики для мониторинга показателей ЦУР. В то время как ЮНЕП выделила 93 официальных показателя ЦУР, связанных с окружающей средой, авторам не известно о каком-либо аналогичном процессе выделения показателей ЦУР, связанных с населением и социальной сферой, или показателей ЦУР, связанных с экономикой. Определение тематических или предметных показателей ЦУР позволяет проанализировать, какие источники больших данных обладают наибольшим потенциалом. Например, было определено, что геопространственная информация и данные наблюдения Земли обладают наибольшим потенциалом для показателей ЦУР, связанных с окружающей средой.

Ниже приведены примеры использования больших данных для двух показателей ЦУР, которые можно рассматривать как показатели, имеющие отношение к статистике населения и социальной статистике, а дополнительную информацию об использовании больших данных для всех показателей ЦУР можно найти в рабочем документе ЭСКАТО «Большие данные для ЦУР — страновые примеры

²³ Fitri Andri Astuti, Nurul Aunun Nisa, Harnessing Social Media Data to Measuring Mental Health Statistics (Использование данных социальных сетей для получения статистики психического здоровья), Азиатско-тихоокеанская неделя по статистике, 2020, [ссылка](#)

составления показателей ЦУР с использованием нетрадиционных источников данных» (*Big data for the SDGs – country examples in compiling SDG indicators using non-traditional data sources*)²⁴

Показатели ЦУР 9.1.1 и 7.1.1 – Использование геопространственных данных для отслеживания доступа населения к всесезонным дорогам на Филиппинах и в Монголии и доступа населения к электроэнергии в Монголии

Индекс доступа к сельским районам (RAI) измеряет долю людей, которые имеют доступ к всесезонной дороге в пределах 2 километров пешком, и указывает на степень доступа для людей, живущих в сельской местности. Для расчета показателя 9.1.1 ЦУР (доля сельского населения страны, проживающего в пределах 2 км от всесезонной дороги) для региона Давао **Статистическое управление Филиппин** объединило административные карты региона с картой населения с координатной сеткой из WorldPop при помощи программного продукта QGIS²⁵. Расчет показателя продемонстрировал, что лишь немногим более половины сельского населения региона Давао проживало в пределах 2 км от всесезонной дороги в 2015 г. Тем не менее, существует разительное отличие в доступности сельской местности в регионе Давао, особенно между севером и югом, как показано на рисунке 8. Хотя использование геопространственной информации и данных наблюдений Земли дает многообещающие результаты, эта методология имеет некоторые недостатки и нуждается в дальнейшем совершенствовании. Например, на Филиппинах нет официального определения «всесезонной» дороги. Кроме того, в методологии учитывается горизонтальное расстояние от всесезонной дороги, без учета высоты и наличия водной поверхности, которая может препятствовать доступу к дороге. Целевая группа ООН по измерению доступа к сельским районам поддерживает статистическое сообщество, совершенствуя методологии и инструменты для сбора необходимых данных для оценки индекса доступа к сельским районам (RAI)²⁶.

Рисунок 8. Наложение карты сельского населения и карты 2-км радиуса от всесезонной дороги (слева) для получения Индекса доступа к сельским районам (RAI) для региона Давао на Филиппинах (справа).



Источник: Статистическое управление Филиппин

Как и Статистическое управление Филиппин, **Национальное статистическое управление Монголии** также экспериментировало с оценкой показателя ЦУР 9.1.1 с помощью программного

²⁴ UN ESCAP, Working Paper Series (SD/WP/12/January 2021): Big data for the SDGs – country examples in compiling SDG indicators using non-traditional data sources (*Большие данные для ЦУР- страновые примеры получения показателей ЦУР с использованием нетрадиционных источников данных*), [ссылка](#)

²⁵ Mae Abigail O. Miralles, Joy Angiela H. Garraez, presentation « Measuring the Rural Access Index (SDG 9.1.1) for the Philippines, International Seminar on the Use of Big Data for Official Statistics, October 2019, Hangzhou, China, «Измерение индекса доступа в сельском районе (ЦУР 9.1.1) для Филиппин», Международный семинар по использованию больших данных для официальной статистики, октябрь 2019 г., Ханчжоу, Китай, [ссылка](#)

²⁶ Целевая группа ООН по измерению доступа к сельским районам, [ссылка](#)

продукта ArcGIS Desktop²⁷. Основной проблемой в случае с Монголией было отсутствие данных о дорогах. Следуя аналогичному подходу, НСУ также экспериментировало с оценкой показателя ЦУР 7.1.1 (доля населения, имеющего доступ к электроэнергии). Оба показателя ЦУР измеряют доступ населения к основным услугам.

Показатель ЦУР 17.19.2 - Использование данных наблюдения Земли для оценки доли населения и жилых помещений в отдаленных и труднодоступных районах Малайзии.

Департамент статистики Малайзии начал стратегическое сотрудничество с Малазийским космическим агентством (MySA) для определения доли населения и жилых помещений в отдаленных и недоступных районах с использованием данных наблюдения Земли. На начальном этапе были опубликованы ГИС-слои счетных участков и зданий для 3 штатов, еще 3 штата находятся в разработке. В качестве вклада в расчеты показателя Малазийское космическое агентство для определения доли населения и жилых помещений в отдаленных и недоступных районах предоставило последние изображения с разрешением 1,5 м и 0,5 м, сделанные каждые 14 дней, которые обновляются чаще, чем карты Google, как показано на рисунке 9. Одна из самых больших проблем использования данных наблюдений Земли в этом проекте – это облака и деревья, закрывающие вид на здания.

Рисунок 9. Спутниковый снимок Малазийского космического агентства (MySA Map) и карт Google (Google maps)



Источник: Департамент статистики Малайзии

Вклад партнеров по развитию

Несколько партнеров по развитию и НПО поддерживают правительства Азиатско-Тихоокеанского региона в разработке политики и реагировании на кризисы при помощи использования источников больших данных. Обладая техническими знаниями, технологиями, финансовыми ресурсами и вступая в партнерские отношения с держателями данных, эти участники начали использовать различные источники больших данных – от данных операторов мобильной связи до спутниковых изображений, онлайн-карт и данных социальных сетей – для получения оценок численности

²⁷ Bolortuya Jambaldorj, presentation “Earth Observation Data in the Development of Official Statistics for SDGs in Mongolia” (*Данные наблюдения Земли для развития официальной статистики для ЦУР в Монголии*), презентация на симпозиуме Азиатско-Океанской группы по наблюдениям за Землей (AOGEO), март 2021 г., [ссылка](#)

населения или составления карт и отслеживания перемещений населения после стихийных бедствий или, совсем недавно, оценки передвижений людей во время пандемии COVID-19.

В рамках Инициативы Генерального секретаря ООН «Глобальный пульс» (UN Global Pulse), ее региональный инновационный центр – лаборатория «Глобальный пульс» в Джакарте (**Pulse Lab Jakarta**) изучал возможности как данных мобильных телефонов, так и данных социальных сетей с помощью ряда пилотных проектов в странах Азиатско-Тихоокеанского региона для оценки мобильности, миграции и вынужденного перемещения населения, а также выявления в социальных сетях тенденций дискриминации женщин на рабочем месте. Благодаря партнерству с Digicel лаборатория **Pulse Lab Jakarta** подготовила основанные на фактических данных оценки уязвимости населения в случае стихийных бедствий и отследила движение населения во время циклона «Донна» в Вануату²⁸ и циклона «Гита» на Самоа²⁹. Также были нанесены на карту перемещения населения после землетрясения в Папуа-Новой Гвинее³⁰, и составлены карты перемещения в реальном времени после землетрясения в Палу (Индонезия) путем объединения данных из Facebook Geoinsights и данных операторов мобильной связи³¹. В ситуации пандемии COVID-19 лаборатория **Pulse Lab Jakarta** использовала данные статистики потенциала села (PODES) и карты плотности населения Facebook, чтобы определить потенциальные очаги распространения вируса на Западной Яве (Индонезия)³². В Турции осуществлялось сотрудничество инициативы ООН «Глобальный пульс» с УВКБ ООН, чтобы понять и измерить интеграцию беженцев через призму пространственной и социальной сегрегации с использованием данных записей о вызовах (CDR) от мобильного оператора Turk Telekom³³.

Фонд Flowminder с 2010 г. использует возможности данных мобильных телефонов для реагирования на стихийные бедствия, а также для целей социально-экономического анализа и точного определения эпидемиологической ситуации. Используя анонимизированные данные CDR от операторов мобильной связи, Flowminder провел картирование моделей миграции из-за изменения климата в Бангладеш, пострадавшей от циклона Махасен в 2013 г.³⁴, и предоставил ключевую информацию о самом крупном вынужденном перемещении людей после землетрясения в районе

²⁸ Idzalika R., Amin I., Riyadi Y., Hodge G., Lee J. G. (2019). Understanding Aggregate Human Behaviour Changes in Response to a Natural Disaster in Vanuatu via Mobile Network Data Analysis (*Понимание совокупных изменений человеческого поведения в ответ на стихийное бедствие в Вануату с помощью анализа данных мобильной сети*), [ссылка](#)

²⁹ Pulse Lab Jakarta, “Informing National Statistics and Managing Disaster Risks.” (*Предоставление информации для национальной статистики и управления рисками бедствий*) The Medium, October 2, 2019, [ссылка](#)

³⁰ Prahara P., Zahara A., Rheza M., Khaefi M. R., Alkarisya D., Riyadi Y., Idzalika R., Hodge G. (2019). Comparing population displacement estimates from mobile network data and other sources (*Сравнение оценок вынужденного перемещения населения по данным мобильных сетей и другим источникам*), [ссылка](#)

³¹ Pulse Lab Jakarta, “Real Time Data for Faster Decision Making in Times of Crisis” (*Данные в реальном времени для более быстрого принятия решений во время кризиса*) 2019, [ссылка](#)

³² Pulse Lab Jakarta, “Alternative Use of Traditional Data in Times of COVID-19 (*Другое использование традиционных данных во времена COVID-19*) The Medium, September 28, 2020, [ссылка](#)

³³ Using Mobile phone call detail records to understand refugee integration in Turkey.” (*Использование данных мобильных телефонов для понимания интеграции беженцев в Турции*), Migration Data Portal, 2020, [ссылка](#)

³⁴ Xin Lu, David J. Wrathall, Pål Roe Sundsøy, Md. Nadiruzzaman, Erik Wetter, Asif Iqbal, Taimur Qureshi, Andrew Tatem, Geoffrey Canright, Kenth Engø-Monsen, Linus Bengtsson, Unveiling hidden migration and mobility patterns in climate stressed regions: A longitudinal study of six million anonymous mobile phone users in Bangladesh (*Выявление скрытых моделей миграции и мобильности в регионах с неблагоприятным климатом: лонгитюдное исследование шести миллионов анонимных пользователей мобильных телефонов в Бангладеш*), Global Environmental Change, Volume 38, 2016, Pages 1-7, ISSN 0959-3780, [ссылка](#)

Горкха в Непале в 2015 г.³⁵ Flowminder также сотрудничал с ЮНФПА, WorldPop и Национальным статистическим управлением Афганистана для интеграции данных ГИС, спутниковых изображений и данных обследований (социально-демографического и экономического обследования (SDS) и переписи) для получения оценок численности населения в стране (Афганистан), где последняя перепись проводилась в 1979 г.³⁶

В области гендерной статистики фонд Flowminder недавно разработал динамическое картирование с высоким разрешением с разбивкой по полу, объединив традиционные источники данных обследований с цифровыми данными (включая геолокационные данные обследований, спутниковые изображения и данные операторов мобильной связи), чтобы обеспечить лучшее понимание жизни женщин и девочек в Непале³⁷.

Базирующаяся на Шри-Ланке НПО LIRNEasia также активно изучает потенциал больших данных для целей развития. Используя данные мобильных телефонов, LIRNEasia оценила плотность и движение населения на Шри-Ланке³⁸. В сотрудничестве с группой исследователей, занимающихся пространственными данными в Токийском университете³⁹, LIRNEasia предоставила информацию управлению городского планирования в Коломбо, используя данные мобильного оператора.⁴⁰

Многие организации активно занимаются решением проблем конфиденциальности в эпоху больших данных. Группа Организации Объединенных Наций по устойчивому развитию выпустила документ «Конфиденциальность, этика и защита: рекомендации по использованию больших данных для достижения целей Повестки дня на период до 2030 года»⁴¹ в то время как GSMA, работающая с правительствами, регулирующими органами и представителями индустрии мобильной связи, содействует внедрению ответственных методов защиты конфиденциальности и разработала набор принципов и требований к конфиденциальности данных⁴². Другие организации также разработали руководящие принципы и программное обеспечение для анонимизации данных и получения

³⁵ Monitoring population movements post disaster: 2015 Nepal Earthquakes (Мониторинг движения населения после катастрофы: землетрясение 2015 г. в Непале), Flowminder, [ссылка](#)

³⁶ Presentation “High-resolution population mapping in Afghanistan”, Remote Sensing Technology Complementing Official Statistics, February 2, 2017 (Презентация «Картирование населения в Афганистане с высоким разрешением. Технология удаленного зондирования для дополнения официальной статистики»), [ссылка](#) и Donna Clarke, presentation “High-resolution population mapping with incomplete data: Applications in Afghanistan,” (презентация «Картирование с высоким разрешением с неполными данными: применение в Афганистане»), Всемирный конгрессе по геопространственной информации Организации Объединенных Наций, [ссылка](#)

³⁷ “Towards High-Resolution Sex-Disaggregated Dynamic Mapping,” (На пути к динамическому картированию с высоким разрешением с разбивкой по полу), 2019, [ссылка](#)

³⁸ LIRNEasia, presentation “Mobile Network Big Data for Urban Development” (презентация «Большие данные мобильных сетей для целей городского развития»), September 2019, [ссылка](#) и Rohan Samarajiva, Sriganesh Lokanathan, Kaushalya Madhawa, Gabriel Kreindler, Danaja Maldeniya, “Big Data to Improve Urban Planning” (Большие данные для совершенствования городского планирования), Review of Urban Affairs, May 20, 2015, Vol I. NO 22, [ссылка](#)

³⁹ Spatial Data Commons, [ссылка](#)

⁴⁰ Lokanathan, Sriganesh & Kreindler, Gabriel & de Silva, Nisansa & Miyauchi, Yuhei & Dhananjaya, Dedunu. (2014). Using Mobile Network Big Data for Informing Transportation and Urban Planning in Colombo (Использование больших данных из мобильных сетей для целей транспорта и городского планирования в Коломбо). SSRN Electronic Journal. 10.2139/ssrn.2526642

⁴¹ United Nations Sustainable Development Group, “Data Privacy, Ethics and Protection: Guidance Note on Big Data for Achievement of the 2030 Agenda,” (Группа ООН по устойчивому развитию «Конфиденциальность, этика и защита: рекомендации по использованию больших данных для достижения целей Повестки дня на период до 2030 года» (2017)) [ссылка](#)

⁴² GSMA, Data Protection & Privacy (Защита данных и конфиденциальность), [ссылка](#)

агрегированных статистических данных. Например, для статистики, полученной на основе данных мобильных телефонов, Flowminder предоставил рекомендации по созданию показателей мобильности⁴³ и агрегированных данных на основе данных CDR⁴⁴ с защитой конфиденциальности в соответствии с регламентом GDPR. Компания также разработала FlowKit⁴⁵, который расширяет возможности анализа данных CDR за пределы выставления счетов, допуская их использование для целей реагирования на стихийные бедствия, точного описания эпидемиологической ситуации, транспорта и мобильности с учетом защиты конфиденциальности. Университет Токио разработал программное обеспечение Mobipack⁴⁶ с открытым исходным кодом, позволяющее анализировать и визуализировать мобильность с использованием деидентифицированных телекоммуникационных данных. А Positium, компания, находящаяся в Эстонии и поддерживающая НСУ в Европе, а также в Грузии и Индонезии, разработала инструментарий Positium Data Mediator⁴⁷.

На глобальном уровне Комитет экспертов по использованию больших данных и инструментария науки о данных для целей официальной статистики ООН (UN-CEBD) включает в себя несколько тематических целевых групп⁴⁸, занимающихся конкретными источниками больших данных, в состав которых входят представители нескольких стран Азиатско-Тихоокеанского региона. Эти тематические целевые группы разрабатывают справочники и методологии и проводят обучение на глобальном и региональном уровнях по использованию различных источников больших данных в официальной статистике. Кроме того, UN-CEBD разработал цифровую платформу *Big Data Marketplace*, демонстрирующую проекты больших данных (данные мобильных телефонов, данные социальных сетей, спутниковые снимки, данные дорожных датчиков и т. д.), имеющие отношение к статистике населения и социальной статистике, включая проекты по оценкам дневной численности населения, ежедневной миграции, мобильности, индексу счастья, спутниковому счету здравоохранения, и многие другие.

Помимо UN-CEBD и его тематических целевых групп, отдельные организации по развитию, которые также являются членами этих целевых групп, изучают потенциал источников больших данных, при этом большая часть этих усилий сосредоточена на изучении мобильности и миграции людей. Например, **Центр анализа данных о глобальной миграции (GMDAC)**⁴⁹ Международной организации по миграции (МОМ) изучает потенциал больших данных для понимания глобальной миграции. По мнению МОМ, источники больших данных, используемые в исследованиях, связанных с миграцией, можно разделить на три категории: *данные мобильных телефонов* — подробные записи о вызовах или мобильные денежные переводы; *интернет-данные* — социальные сети и геолокация активности в социальных сетях, IP-адреса из электронной почты; и *сенсорные данные* — данные наблюдения Земли. GMDAC исследует потенциал больших данных для статистики миграции и собирает примеры со всего мира в справочник по инновациям в области данных (Data Innovation

⁴³ Flowminder, “Enabling Mobile Network Operators to produce aggregates from CDR data,” (*Предоставление операторам мобильной связи возможности формировать агрегированные данные CDR*), [ссылка](#)

⁴⁴ Flowminder, “Enabling analysts to produce mobility indicators from CDR aggregates,” (*Предоставление аналитикам возможности получать показатели мобильности на основе агрегированных данных CDR*) [ссылка](#)

⁴⁵ FlowKit, [ссылка](#)

⁴⁶ Spatial Data Commons, Аналитический инструментарий – Mobipack, [ссылка](#)

⁴⁷ Positium, инструментарий Positium Data Mediator, [ссылка](#)

⁴⁸ Комитет экспертов по использованию больших данных и инструментария науки о данных для целей официальной статистики ООН (UN-CEBD) включает в себя следующие целевые группы: данные АИС; большие данные и ЦУР; данные наблюдения Земли; измерение доступа к сельским районам; данные мобильных телефонов; методы сохранения конфиденциальности; данные сканеров; обучение, компетенции и развитие потенциала. Дополнительная информация представлена на сайте, [ссылка](#)

⁴⁹ Центр анализа данных о глобальной миграции, [ссылка](#)

Directory)⁵⁰ на портале данных о миграции. Кроме того, возглавляемая МОМ инициатива «Большие данные для миграции» (BD4M)⁵¹ направлена на ускорение ответственного и этичного использования новых источников данных и инновационных методологий для предоставления информации для разработки миграционной политики и программ.

Еще одна глобальная инициатива — Партнерство в области использования данных в целях развития⁵². Эта инициатива возглавляется Всемирным банком и адресована в первую очередь организациям, занимающимся вопросами развития. Это партнерство объединяет международные организации и 25 партнеров по работе с данными, включая такие компании, как Google, Esri, LinkedIn, Twitter, Waze, которые обмениваются данными для проектов в области развития и гуманитарной сферы. У этого партнерства есть проекты в нескольких странах Азиатско-Тихоокеанского региона. Некоторые из последних проектов сосредоточены на изучении перемещений людей во время пандемии COVID-19, например, анализ пространственной доступности медицинских учреждений в Индонезии и на Филиппинах с использованием данных Facebook и Mapbox⁵³ и отслеживание второй волны COVID-19 во Вьетнаме с использованием данных о перемещениях из Facebook⁵⁴.

Проблемы использования источников больших данных для целей статистики населения и социальной статистики

Источники больших данных предоставляют значительные преимущества для производства статистики населения и социальной статистики: увеличение охвата, повышение уровня детализации, улучшение периодичности и своевременности, а также снижение затрат. Однако сохраняется множество проблем.

Существуют две особые проблемы, связанные с использованием источников больших данных для статистики населения и социальной статистики, которые не так очевидны, как для экономической статистики и статистики окружающей среды. Первая большая проблема связана с единицей сбора данных: статистика населения и социальная статистика в основном построены на данных об индивидах, и чаще всего регулируются национальными законами о защите конфиденциальности информации в дополнение к законам о статистике. Вторая проблема заключается в важности статистики населения для большого количества аспектов государственного управления, таких как представительство в парламенте, механизмы социальной защиты и распределение финансовых средств между федеральными, региональными и местными органами власти. Поэтому надежность и качество статистики населения имеют первостепенное значение для органов государственного управления, общества и экономики.

За некоторыми исключениями, большинство случаев использования больших данных в статистике населения и социальной статистике касаются статистики населения и остаются экспериментальными по своему характеру. Успешные пилотные проекты по работе с большими данными не обязательно приводят к интеграции источников данных для статистического производства по целому ряду причин, не в последнюю очередь потому, что долгосрочный доступ к данным для интеграции требует других обязательств и механизмов.

⁵⁰ Центр анализа данных о глобальной миграции, Справочник по инновациям в области данных, [ссылка](#)

⁵¹ Альянс по вопросам больших данных по миграции (BD4M), [ссылка](#)

⁵² Партнерство в области использования данных в целях развития, [ссылка](#)

⁵³ Партнерство в области использования данных в целях развития, «Facebook, Mapbox: анализ пространственной доступности учреждений здравоохранения в Индонезии и на Филиппинах», [ссылка](#)

⁵⁴ Партнерство в области использования данных в целях развития, «Отслеживание второй волны COVID-19 во Вьетнаме с использованием данных о перемещениях из Facebook», [ссылка](#)

Тем не менее, несмотря на особенности и различия между источниками больших данных, было выявлено несколько общих проблем, связанных с использованием больших данных для целей официальной статистики. Они в основном относятся к 1) юридическому регулированию производства официальной статистики и условиям доступа к нетрадиционным источникам данных, 2) доступу к данным и переговорам о партнерстве, связанным с условиями доступа, защитой конфиденциальности и непрерывностью поставок, и 3) внутренним техническим, финансовым и технологическим ресурсам статистических организаций и национальной статистической системы в целом. Эти проблемы актуальны и для других областей статистики.

Юридические ограничения часто называют препятствием для использования больших данных в официальной статистике. Страны, сталкивающиеся с этой проблемой, ссылаются на нормативно-правовую базу, которая ограничивает виды источников данных, которые могут быть использованы для производства официальной статистики. Однако можно надеяться, что в рамках пилотных проектов можно будет исследовать новые источники данных на предмет их тестирования и обоснования интеграции альтернативных источников данных там, где это уместно и необходимо.

Конфиденциальность данных также является важной проблемой, которую необходимо решать при рассмотрении использования источников больших данных для производства официальной статистики, особенно персональных данных. Национальные законодательные рамки в отношении конфиденциальности информации и защиты персональных данных различны в разных странах Азиатско-Тихоокеанского региона и во всем мире. Кроме того, несколько НСУ в регионе отмечают отсутствие нормативно-правовой базы, обязывающей частные компании делиться своими данными с государственным сектором для подготовки статистики и предоставления информации для разработки государственной политики, или для регулирования доступа к данным из частного сектора.

В то время как в некоторых странах производство статистики регулируется на законодательном уровне, а это означает, что любые изменения в производстве данных и в основных источниках данных (например, в административных источниках или источниках данных в частном секторе) должны пройти юридическую проверку, в других странах статистическое законодательство может быть более общим и охватывать и другие источники данных. Индонезия, например, в дополнение к производству статистики на основе переписи населения, обследований и административных данных, указывает на возможность *«других способов сбора данных в зависимости от развития науки и техники»*⁵⁵.

Опыт стран региона показывает, что статистические организации следуют **различным моделям партнерства для доступа к данным частного сектора**. В примере с данными операторов мобильной связи такие организации, как статистические управления Индонезии, Республики Корея, Новой Зеландии и программа a2i в Бангладеш, договариваются о доступе напрямую с поставщиками данных. В других странах, таких как Грузия, доступ к данным оператора мобильной связи согласовывается косвенно через национальный регулирующий орган, который предписывает обмен данными в частном секторе, по крайней мере, для конкретных проектов, посредством регулятивных процедур. Другие модели включают получение данных от сторонних поставщиков, как в случае с Австралийским бюро статистики, и геолокационных данных из мобильных приложений. В случае Новой Зеландии был предпринят дополнительный шаг – создание дочерней компании статистического управления, которая называется Data Ventures. Эти модели партнерства различаются и основаны на национальной специфике стран, их нормативно-правовой базе, целях проекта, внутренних ресурсах и удобстве.

⁵⁵ Ali Said, Центральное статистическое управление Индонезии, презентация на мероприятии «Кафе для статистиков» 17 августа 2020 г. [ссылка](#)

Неотъемлемой частью переговоров и соглашений о доступе к данным являются **условия доступа**. В приведенных выше примерах одни статистические организации получали данные бесплатно, тогда как другие делали это на основе коммерческого соглашения. Национальное статистическое управление Грузии и Центральное статистическое управление Индонезии, выполняя аналогичные проекты с данными мобильных операторов, получали доступ по-разному: НСУ Грузии имело бесплатный доступ, предоставленный положением закона, в то время как ЦСУ Индонезии получало его по коммерческому соглашению. Опыт стран региона показывает, что большинство статистических управлений получают данные, связанные с передвижением людей, на коммерческой основе. Эти условия также следует анализировать на этапе разработки проекта, будь то пилотный проект или масштабирование.

Стоимость **данных частного сектора** для официальной статистики остается предметом активных дискуссий. Некоторые участники статистического сообщества утверждают, что официальная статистика является общественным благом и что данные для ее производства должны быть получены бесплатно или по цене не выше предельных затрат. Другие считают, что компании несут расходы, связанные с хранением и обработкой данных, такими как анонимизация и агрегирование для того, чтобы предоставлять данные в соответствии с внутренними и национальными процессами и правилами защиты конфиденциальности данных, и что эти расходы должны быть покрыты. Кроме того, учитывая новый характер источников больших данных, некоторые компании, занимающиеся сбором данных, строят новые бизнес-модели в связи со сбором этих данных.

Еще один элемент, который следует учитывать, — это доступ к **данным частного сектора** по сравнению с доступом к результатам аналитической обработки. Представленные тематические исследования подчеркивают эти различия.

В то время как некоторые НСО покупают выборочные агрегированные данные, охватывающие короткий период времени и конкретную географическую область, для проверки релевантности источника данных или просто передают на аутсорсинг сбор и анализ данных, другие договариваются о прямом доступе с поставщиками данных или национальными регулирующими органами, создавая партнерства и совместно разрабатывая методологии анализа данных. В некоторых случаях НСУ получают доступ к анонимизированным и агрегированным данным, которые они анализируют внутри своей организации, тогда как в других случаях они могут получить доступ только к результатам, полученным на основе применения разработанных ими методологий и алгоритмов. В одном случае анализ проводится на «сырых» или первичных данных, а в других случаях на сильно агрегированных данных.

Интеграция источников и методов больших данных для производства официальной статистики требует инвестиций, особенно для наращивания внутреннего потенциала, для развития необходимой ИТ-инфраструктуры для сбора, хранения, обработки и анализа данных, а также для обеспечения непрерывного доступа к данным, особенно когда данные приобретаются на коммерческой основе. В то время как некоторые страны региона, например Австралия, Новая Зеландия, Китай, Грузия, Малайзия, постепенно создают свою внутреннюю технологическую инфраструктуру для удовлетворения требований использования больших данных и искусственного интеллекта, другие страны признали свою технологическую неподготовленность, среди прочих проблем, к тому, чтобы начать изучение альтернативных источников данных для официальной статистики. Инвестиции следует сопоставлять с затратами на традиционные методы сбора данных и сопутствующими расходами. Компромисс возможен, особенно в связи с тем, что большие данные могут дополнять некоторые традиционные методы сбора данных, предоставляя более своевременные и детализированные данные для принятия более информированных и основанных на фактических данных решений.

Рекомендации

Опираясь на успешный опыт и уроки, извлеченные некоторыми НСУ в регионе, можно дать несколько рекомендаций НСУ, которые либо вступают на путь использования больших данных, либо уже экспериментируют с большими данными для производства официальной статистики населения и социальной статистики.

В то время как некоторые НСУ могут сталкиваться с проблемами, связанными с ограничительной правовой базой при производстве статистики, они все же могут **определить приоритетные области использования больших данных в статистике населения и социальной статистике** и осуществлять **небольшие пилотные проекты**, которые в случае успеха могут создать основу для непрерывного производства и нормативно-правовых изменений в случае необходимости.

Создание доверительных партнерских отношений с частным сектором для получения доступа к данным и проведения экспериментов является еще одним важным вопросом, который необходимо рассмотреть. По мере роста интереса НСУ к данным мобильных операторов для целей производства статистики населения и социальной статистики, доступ к этим данным требует длительных переговоров. Поэтому НСУ рекомендуется **вести переговоры с частным сектором на раннем этапе, чтобы изучить возможности партнерства** и понять проблемы.

Регулирующим органам следует рассмотреть возможность разработки руководства по доступу к частным данным для статистических организаций и систем. В отсутствие четких указаний по доступу к данным, их обмену и использованию частные компании могут не захотеть делиться данными или могут ссылаться на строгие корпоративные процедуры защиты данных, препятствующие обмену.

При разработке и реализации пилотных проектов по работе с большими данными НСУ рекомендуется **обращаться за поддержкой и рекомендациями к гражданскому обществу, специализированным частным компаниям или академическим кругам**. Развитие партнерских отношений с академическими кругами также может быть полезным, особенно когда НСУ имеют ограниченные внутренние возможности для обработки и анализа больших данных. UN-CEBD и его целевые группы являются важным источником поддержки, знаний и возможностей для сотрудничества.

Для решения вопросов, связанных с согласием, конфиденциальностью и безопасностью, НСУ должны **сотрудничать с национальными агентствами по защите персональных данных на этапе разработки любого проекта**. Сотрудничество с этими органами поможет снизить риски, связанные с использованием персональных данных, и координировать сотрудничество с поставщиками данных. Кроме того, коммуникация и прозрачность в отношении того, как персональные данные используются и защищаются с помощью методов обеспечения конфиденциальности и безопасности данных, важны для укрепления доверия со стороны широкой общественности. НСУ следует рассмотреть возможность открытого диалога с гражданами по вопросам согласия и защиты персональных данных, особенно при рассмотрении вопроса об интеграции административных данных и данных частного сектора в перепись или в период переписи.

Для повышения осведомленности об использовании больших данных в статистике населения и в социальной статистике НСУ также рекомендуется **документировать свой опыт и активно делиться им** на региональных и глобальных форумах. Независимо от их опыта работы с большими данными, НСУ могут также полагаться на уже имеющийся объем информации. В связи с этим Статистический отдел ЭСКАТО активно поддерживает в регионе обмен знаниями об использовании больших данных для официальной статистики посредством целенаправленных усилий, таких как «Кафе для статистиков» (Stats Café), ориентированных на большие данные для официальной

статистики, Сообщество практиков по интеграции данных (DI CoP), Краткие обзоры по статистике (Stats Briefs), рабочие документы и блоги, посвященные примерам использования больших данных в официальной статистике.

Заключение

В Азиатско-Тихоокеанском регионе статистические организации активно экспериментируют с использованием источников больших данных для производства статистики перемещений населения и урбанизации, но в меньшей степени для других аспектов статистики населения, таких как рождаемость и смертность, или социальной статистики. Данные операторов мобильных сетей являются наиболее изученной областью для использования в статистике пространственной мобильности населения. Другие источники данных, такие как данные Facebook и Google, а также данные общественного транспорта, также изучаются, но в меньшем масштабе и в основном международными партнерами по развитию.

Тематические исследования, представленные в этом кратком обзоре, содержат конкретные примеры того, как НСУ в регионе интегрируют большие данные для производства статистики населения и социальной статистики. Тематические исследования иллюстрируют различные используемые источники данных, методы, типы партнерств и подходы.

Они также иллюстрируют возможности использования больших данных и возникающие проблемы.

Несмотря на множество примеров, большинство проектов по использованию больших данных в статистике населения и в социальной статистике остаются на экспериментальной стадии. Однако можно отметить несколько исключений, как в примере с использованием данных мобильных телефонов для статистики мобильности населения и туризма в Индонезии.

Независимо от того, проводят ли статистические управления эксперименты или только рассматривают возможность использования больших данных, они сталкиваются с целым рядом проблем, которые необходимо решить. К ним относятся правовые рамки, регулирующие производство статистических данных, доступ к данным и конфиденциальность данных; партнерские отношения с частными компаниями; наращивание внутреннего технического и технологического потенциала и финансовых ресурсов. В этом кратком обзоре представлены наиболее распространенные проблемы, рассказано, как некоторые НСУ в регионе реагируют на них, а также даны некоторые рекомендации по инициированию работы с большими данными или продвижению вперед.

Пандемия COVID-19 послужила стимулом для дальнейшего изучения больших данных для своевременного и частого получения результатов, особенно в области статистики мобильности населения. Пандемия выявила важные пробелы в традиционной статистике и показала ценность нетрадиционных данных для получения своевременных результатов. Чтобы работа НСУ оставалась актуальной, они должны дополнительно изучить возможности, которые источники и методы больших данных могут предоставить для производства официальной статистики населения и социальной статистики.

Узнать больше о работе ЭСКАТО в области развития статистики можно на сайте:

<http://www.unescap.org/our-work/statistics>

Предыдущие выпуски Stats Brief:

<http://www.unescap.org/resource-series/stats-brief>

Контакт: stat.unescap@un.org



ⁱ Дополнительные материалы для этого статистического обзора были получены от следующих специалистов: Гэри Даннет, заместитель главного методиста Статистического управления Новой Зеландии; Джисук Юн, директор отдела больших данных и статистики, Национальное статистическое управление Республики Корея; Янг Ран Ким, заместитель директора отдела больших данных и статистики, Национальное статистическое управление Республики Корея; Рамиз Уддин, руководитель экспериментального отдела Accelerator Lab, отделение ПРООН в Бангладеш; Сангита Пол, специалист по данным, a2i, Бангладеш; Андерс Холмберг, главный методолог и генеральный директор отдела методологии Австралийского бюро статистики; Рик Кларк, директор отдела новых данных и методов Австралийского бюро статистики; Шорена Тисклаури, главный специалист отдела методологии и управления качеством Национальной службы статистики Грузии; Али Саид Директор отдела статистического анализа и развития Центрального статистического управления Индонезии; Сетиа Прамана Специалист по большим данным Центрального статистического управления Индонезии; Либэнь Хоу, заместитель директора отдела по общим вопросам Департамента статистического проектирования и управления Национального бюро статистики Китая. В написании обзора также помогли обсуждения и презентации во время мероприятия «Кафе для статистиков Азиатско-Тихоокеанского региона», состоявшегося 2 ноября 2020 г.