

Большие данные для целей официальной статистики

Основные соображения: использование
данных мобильной телефонной связи

Доклад выпускается без официального редактирования. Мнения, изложенные в данном документе, не обязательно совпадают с мнением Экономической и социальной комиссии для Азии и Тихого океана (ЭСКАТО) или учреждений Организации Объединенных Наций.

Для получения дополнительной информации просим обращаться:

Отдел статистики

Экономическая и социальная комиссия для Азии и Тихого океана (ЭСКАТО)

Здание Организации Объединенных Наций, Бангкок 10200, Таиланд

Электронная почта: stat.unescap@un.org

Веб-сайт: <https://www.unescap.org/our-work/statistics>

Выражение признательности

Общее руководство работой по подготовке доклада осуществляли г-жа Рейчел Бивен, директор Отдела статистики Экономической и социальной комиссии для Азии и Тихого океана Организации Объединенных Наций (ЭСКАТО) и г-жа Джемма Ван Халдерен, бывший директор Отдела статистики ЭСКАТО. Основными авторами доклада являются г-жа Ирина Берналь (консультант) и г-н Сованнарот Тей (младший статистик Отдела статистики, ЭСКАТО), проводившие работу под руководством г-жи Тани Сейерсен, статистика Отдела статистики.

Вклад в подготовку доклада внесли сотрудники ЭСКАТО и Статистического института Организации Объединенных Наций для Азии и Тихого океана (СИАТО) г-н Афсанех Яздани (статистик) и г-н Кристоф Бонтам (статистик/лектор). Участники состоявшегося в 2021 году Сопсовещания группы экспертов по использованию больших данных для целей официальной статистики предоставили ценные рекомендации, комментарии и замечания относительно общего направления работы и целей настоящего руководства.

При подготовке доклада проведено внешнее коллегиальное рецензирование, учтены исчерпывающие замечания и пожелания группы экспертов, состоящей из сотрудников директивных органов, ученых и специалистов-практиков в области развития, в числе которых г-н Энди Татем (директор, программа WorldPop), г-жа Марианна Джалагония (главный специалист, Служба переписи населения и демографической статистики, Национальное статистическое управление Грузии), г-н Сиу-Мин Там (директор, компания Tam Data Advisory Pty Ltd), г-жа Нилима Лал (специалист по вопросам экономической статистики, Секретариат тихоокеанского сообщества/СТС), г-жа Юнирти Юнирти (Управление по вопросам анализа и развития статистики, Центральное статистическое бюро Индонезии/ЦСБ), г-н Гари Даннет (заместитель главного методиста, Статистическое управление Новой Зеландии), г-н Нгуен Туан Анх (заместитель директора, Департамент сбора информации и применения технологий в области статистики, Главное статистическое управление Вьетнама), г-н Янус Крун (начальник, Отдел статистики, Банк Эстонии) и г-н Сиим Эско (начальник, Отдел продаж и маркетинга, компания «Позитивум»).

СОДЕРЖАНИЕ

Выражение признательности	I
Содержание	II
1. Введение	1
2. Растущая роль НСУ в национальной экосистеме данных	3
3. Основные проблемы, связанные с использованием больших данных для целей официальной статистики	7
4. Законодательная база	10
5. Неприкосновенность личной информации, этические нормы и вопрос доверия	15
6. Обеспечение доступа к большим данным и модели партнерства	18
6.1 Уровни доступа к данным	18
6.2 Модели доступа к данным	19
6.3 Вовлечение других заинтересованных сторон	23
6.4 Проблемы, связанные с заключением соглашений на добровольной и коммерческой основе	24
7. Издержки и проведение закупок	26
8. Производство статистики с использованием больших данных	30
9. Контроль качества	33
10. Обеспечение поддержки на глобальном и региональном уровнях	36
11. Уроки использования данных мобильной телефонной связи для целей официальной статистики	38
11.1 Что такое данные мобильной телефонной связи?	38
11.2 Аспекты обеспечения доступа к ДМП	39
11.3 Дорожная карта проекта по ДМП	40
11.4 Конфиденциальность данных и вопрос общественного доверия	41
11.5 Производство статистики с использованием ДМП	44
11.6 Примеры стран	44

Введение

Стремительный технологический прогресс - причина появления у пользователей официальных данных и статистики новых ожиданий. Органам государственной власти, деловому сообществу и гражданам необходима актуальная информация. Недавняя пандемия COVID-19 укрепила понимание того, что для выработки мер политического реагирования крайне важны своевременные и подробные данные, наряду с этим пандемия выявила узкие места существующих статистических производственных систем. Растет взаимосвязанность современного мира, в котором люди занимаются привычными делами в режиме онлайн и постоянно пользуются мобильным телефоном. Благодаря этому правительства и организации частного сектора получают обширные сведения. Эти данные могут служить цели формирования официальной статистики и выработки своевременных политических мер при условии соблюдения этических норм и сохранения неприкосновенности частной жизни.

Развитие технологий и спрос на актуальную и достоверную информацию заставляют переосмыслить существующие подходы к подготовке официальной статистики, а также роль национальных статистических управлений (НСУ). НСУ, как ожидается, отреагируют на острую потребность в своевременных и полных соответствующих данных, обеспечивая при этом конфиденциальность информации. НСУ вынуждены интегрировать новые источники данных, что требует выстраивания партнерских отношений с новыми заинтересованными сторонами, инвестиций в новую инфраструктуру и развития новых компетенций, а также обновления существующих механизмов управления данными для обеспечения гибкости и оперативности работы с учетом новых потребностей.

Исследование ЭСКАТО по использованию больших данных в целях официальной статистики в странах Азиатско-Тихоокеанского региона выявило существование многих проблем, с которыми сталкиваются НСУ, пытаясь применить большие данные в официальной статистике. Две важнейшие проблемы связаны с механизмами управления данными, в частности, с законодательной базой, часто ограничивающей использование альтернативных источников данных при подготовке официальной статистики, и выстраиванием новых партнерских отношений с компаниями частного сектора. Участники совещания группы экспертов по использованию больших данных в целях официальной статистики, состоявшегося

в августе-сентябре 2021 года, уделили основное внимание этим двум приоритетным вопросам¹. По итогам заседания было рекомендовано разработать руководство по применению больших данных в целях официальной статистики.

Подготовка данного руководства отвечает существующим в регионе потребностям в разработке руководящих принципов в области использования больших данных в официальной статистике, независимо от типа информации. Цель руководства - оказать содействие тем НСУ Азиатско-Тихоокеанского региона, которые стремятся найти возможности для интеграции больших данных в статистические производственные процессы. Другие участники национальных статистических систем как в регионе, так и за его пределами, вероятно, также найдут данное руководство полезным.

Страны находятся на разных уровнях развития, имеют разное институциональное устройство и нормативно-правовую базу, разный потенциал для изучения подходов к использованию больших данных. Данное руководство разработано с учетом передовых практик и примеров стран региона и мира в целом, представляет разные подходы к интеграции больших данных в процессы формирования официальной статистики. В нем делается акцент на некоторых важных мерах, которые могут быть рассмотрены НСУ и государственными органами по мере дальнейших действий в этом направлении.

В руководстве рассматриваются семь основных аспектов, связанных с использованием больших данных в официальной статистике: 1) растущая роль НСУ, 2) основные проблемы применения больших данных в целях официальной статистики, 3) законодательная база, 4) конфиденциальность информации, этические нормы и вопрос доверия, 5) доступ к большим данным и модели партнерства, 6) издержки и проведение закупок и 7) контроль качества. В руководстве представлены примеры формирования официальной статистики на основе разных источников больших данных, дополнительно делаются ссылки на механизмы и инструменты содействия на глобальном и региональном уровнях. Наконец, данное руководство документирует разные подходы к использованию определенного источника данных в официальной статистике, данных мобильной телефонной связи, с использованием конкретных примеров таких стран, как Эстония, Оман, Индонезия и Российская Федерация.

¹ Отдел статистики, Экономическая и социальная комиссия для Азии и Тихого океана (ЭСКАТО), [Доклад совещания группы экспертов по использованию больших данных в целях официальной статистики: управление данными и модели партнерства](#) (Бангкок, 2021).

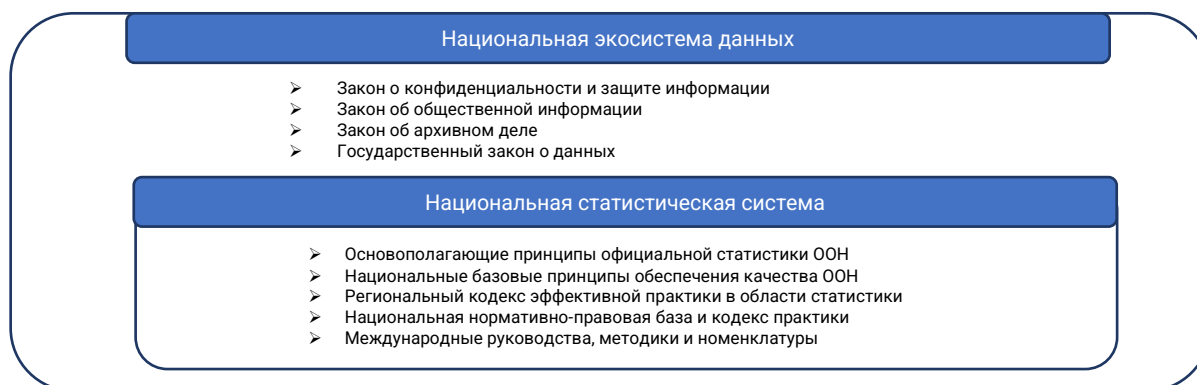
Растущая роль НСУ в национальной экосистеме данных

Роль НСУ в национальных экосистемах данных непрерывно растет в ответ на новые информационные потребности. Во время пандемии COVID-19 очевидным стало то, что для разработки политики требуются точные, актуальные и подробные данные. Более того, меняется роль официальной статистики в жизни общества. Сегодня заинтересованным сторонам необходима более открытая, доступная, своевременная, качественная и полная информация. В ответ на ожидания пользователей НСУ производят соответствующие актуальные данные и статистическую информацию. Компании частного сектора и другие организации владеют собственными большими наборами данных и, как представляется, конкурируют с НСУ, предоставляя директивным органам необходимые сведения. Это означает, что НСУ рискуют ослабить доверие пользователей, если не смогут удовлетворить существующим требованиям.

В рамках действующих статистических бизнес-моделей предстоит решить многочисленные задачи, такие как поиск и использование новых источников данных, создание или разработка новых платформ, обучение сотрудников и обновление существующей нормативно-правовой базы.² НСУ необходим потенциал для того, чтобы интегрировать данные из разных источников, в том числе новых, и разрабатывать новые статистические модели. Надлежит создать новые платформы для безопасного хранения и обработки растущих объемов данных разного типа, что требует адаптации существующих финансовых моделей и политики конфиденциальности. Для работы с новыми источниками данных часто требуются новые профессиональные навыки сотрудников и благоприятные законодательные основы, выстроенные на уважении к неприкосновенности личной информации и этических правилах.

² [Выполнение новой роли национальными статистическими управлениями в период расширенных возможностей](#) (ECE/CES/2020/10).

Диаграмма 1. Национальная статистическая система в рамках национальной экосистемы данных³



Национальные экосистемы данных развиваются на фоне усилий государственных органов по использованию больших данных, машинного обучения, искусственного интеллекта и других технологий в ходе цифровой трансформации в масштабах правительственных структур и экономики в целом. Эти меры направлены на повышение эффективности рабочих процессов и качества предоставления государственных услуг. Роль НСУ меняется по мере того, как их функции трансформируются от производства к распоряжению данными и их хранению. В этой связи НСУ надлежит рассмотреть способы обеспечения качества информации и полного соответствия основополагающим принципам официальной статистики.

В долгосрочной перспективе НСУ, вероятно, рассмотрят следующие вопросы относительно видения своей роли и изменения позиций в национальных экосистемах данных:

- ❖ Какую роль будут играть национальные статистические управления и официальная статистика в жизни общества?
- ❖ Какую роль должны играть НСУ в управлении данными?
- ❖ Какие функции, традиционно возложенные на НСУ, сохранятся, а какие должны быть переданы сторонним исполнителям?
- ❖ Как национальные статистические системы ответят на конкуренцию со стороны частного сектора, привлекая талантливых специалистов?
- ❖ Какие стратегические партнерства должны быть созданы в ответ на новые информационные запросы, какие подходы надлежит выработать в этой связи?
- ❖ Как мы можем обеспечить конфиденциальность информации и сохранить доверие общественности к данным?

³ Gabriel Gamez, *Governance beyond the national statistical system*, презентация в режиме онлайн в рамках сессий Кафе по статистике ЭСКАТО, 16 августа 2021.

Усиление роли НСУ как куратора данных позволит государственным институтам статистики использовать возможности стремительно меняющихся экосистем данных. Усовершенствование механизма распоряжения данными – аспект управления информацией, обеспечивающий формирование, обработку, хранение и соответствие данных целевому назначению в интересах всей экосистемы.⁴ Кураторы данных должны работать во благо экосистемы и сотрудничать, повышая совместными усилиями операционную совместимость данных, совершенствуя производственные процессы, распространяя информационную грамотность и передовой опыт в масштабах экосистемы данных.

Необходимо оказать содействие НСУ, определив новые полномочия в рамках соответствующей нормативной базы. Международное статистическое сообщество рассмотрит возможности учета изменений и новых функций НСУ в последующих версиях Типового закона об официальной статистике (ТЗОС). В ходе 52^й сессии Статистической комиссии Организации Объединенных Наций (СК ООН) Комиссия подчеркнула важность международных классификаций для совершенствования управления данными и роли НСУ как куратора информации.⁵ На основании решений Комиссии была учреждена Рабочая группа по распоряжению данными, уполномоченная провести анализ тематических исследований и разработать концептуальные основы распоряжения данными.⁶

Управление данными развивается в контексте проблем и изменений в сфере больших данных. Перед НСУ стоит необходимость развития тесного сотрудничества как с компаниями частного сектора, так и с другими государственными структурами. Сотрудничество с органами государственной власти должно быть направлено на формирование экосистемы данных, в рамках которой поощряется использование и повторное использование данных при соблюдении этических норм, также обеспечивается доступ и обмен данными внутри и вне правительственных структур. Приоритетным направлением сотрудничества с частным сектором должно стать выявление новых механизмов партнерства в сфере управления данными. Это позволит использовать преимущества данных как общественного блага, предоставит информацию, услуги и создаст новые коммерческие возможности для компаний частного сектора.

Экономическая и социальная комиссия ООН для Европы (ЕЭК) сформулировала 21 рекомендацию по распоряжению данными НСУ. Рекомендации отражают

⁴ Gabriel Gamez, *Governance beyond the national statistical system*, презентация в режиме онлайн в рамках сессий Кафе по статистике ЭСКАТО, 16 августа 2021.

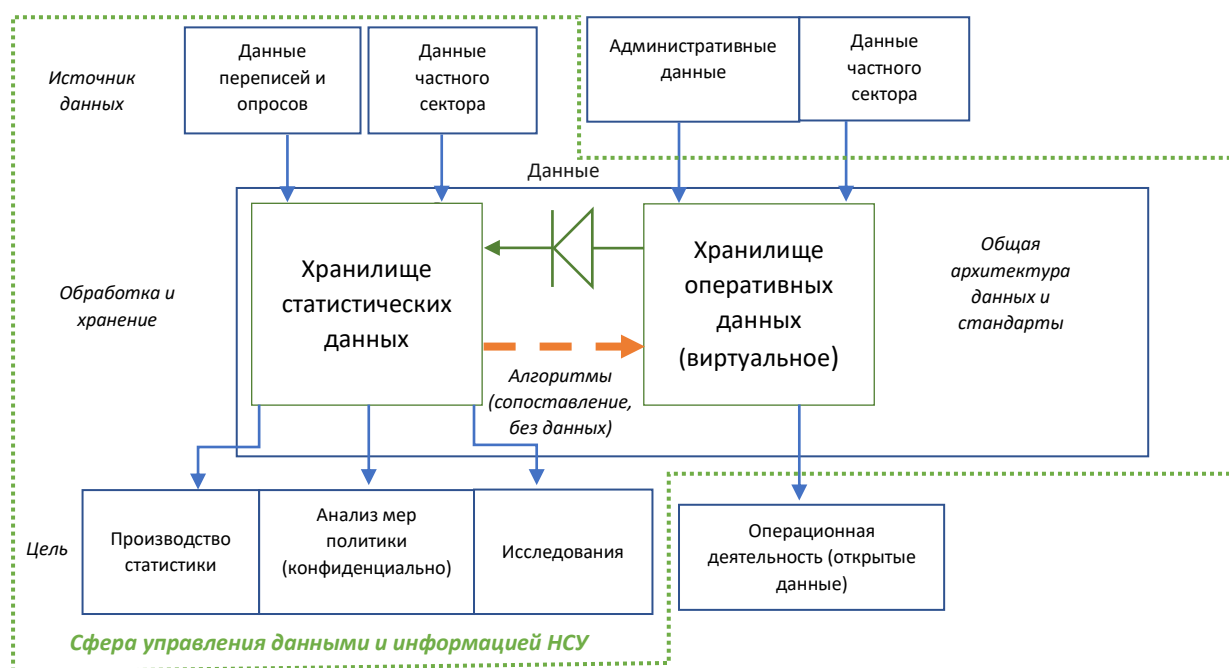
⁵ [Доклад о работе пятьдесят второй сессии](#) (E/2021/24-E/CN.3/2021/30).

⁶ [Доклад Рабочей группы по распоряжению данными](#) (E/CN.3/2022/5).

многогранность приоритетных для НСУ вопросов и функций в рамках национальной экосистемы данных.⁷ Некоторые члены Европейского союза (ЕС) представили архитектуру управления данными высокого уровня в виде метафоры озера данных (диаграмма 2). Она разделена на две зоны (хранилища): первая – хранилище оперативных данных, где находятся индивидуальные и агрегированные данные из нестатистических источников, и вторая – хранилище статистических данных, где размещена информация, в отношении которой действуют требования статистической конфиденциальности.⁸ Данные могут передаваться из одной зоны в другую, но не наоборот.

Кроме того, ЕЭК учредила Целевую группу по распоряжению данными⁹ для формулировки терминов в области сопровождения данных и управления общедоступной информацией, определения новых функций НСУ в этом контексте.

Диаграмма 2. Архитектура высокого уровня для управления данными в виде метафоры озера данных¹⁰



⁷ [Роль национальных статистических систем в новой экосистеме данных](#) (ECE/CES/2019/25).

⁸ [Выполнение новой роли национальными статистическими управлениями в период расширенных возможностей](#) (ECE/CES/2020/10).

⁹ См. <https://unece.org/statistics/task-force-data-stewardship>.

¹⁰ [Выполнение новой роли национальными статистическими управлениями в период расширенных возможностей](#) (ECE/CES/2020/10).

Проблемы, связанные с использованием больших данных для целей официальной статистики

НСУ сталкиваются с рядом проблем, пытаясь использовать большие данные в производстве официальной статистики. Эти проблемы связаны либо с особенностями больших данных, либо с существующими стандартами производства официальной статистики. Объем, скорость прироста и динамические показатели больших данных отличаются от характеристик традиционных источников, таких как опросы, переписи или информация административных ресурсов. По этой причине необходимы дополнительные процедуры и инфраструктура, обеспечивающие доступ, преобразование, передачу, хранение, обработку и анализ больших данных. Более того, НСУ должны следовать нормам и законодательству в сфере производства официальной статистики, что предполагает проведение тщательных и комплексных проверок и обеспечение контроля качества.

Некоторые НСУ предпринимали попытки проанализировать проблемы, связанные с использованием больших данных в производстве официальной статистики. Итальянский национальный институт статистики выделяет четыре категории вопросов: характеристики, качество, временная зависимость и доступность.¹¹ Центральное статистическое бюро Нидерландов уделяет особое внимание следующим основным аспектам, связанным с использованием больших данных для получения официальной статистики: методология, качество, конфиденциальность и нормативная база, обработка, хранение и передача больших наборов данных, также рассматривает вопрос волатильности.¹² Австралийское бюро статистики определило похожий круг вопросов, включающий достоверность статистического вывода, конфиденциальность и общественное доверие, право собственности на данные и обеспечение доступа, эффективность расчетов и наличие технологической инфраструктуры.¹³

Несмотря на то, что проблематика, обозначенная НСУ, отличается, вопросы имеют немало общего. В настоящем руководстве представлены тематические

¹¹ Monica Scannapieco, Antonino Virgillito и Diego Zardetto, *Placing big data in official statistics: a big challenge?*, доклад в ходе Конференции по использованию новых методов и технологий в статистике, 2013.

¹² Peter Struijs, Barteld Braaksma и Piet JH Daas, *Official statistics and big data*, *Big Data and Society* (April-June 2014).

¹³ Siu-Ming Tam and Frederic Clarke, *Big Data, Official Statistics and Some Initiatives by the Australian Bureau of Statistics*, *International Statistical Review*, vol. 83, No. 3 (December 2015).

исследования, на основе которых рассматривается опыт разных стран по преодолению следующих проблем.

- ❖ **Нормативно-правовая база.** Законы и нормативные акты играют важнейшую роль, регулируя доступ и передачу больших данных НСУ. Невзирая на то, что в ряде стран существуют комплексные правовые системы, обеспечивающие использование традиционных источников данных, правила доступа и обмена большими данными зачастую сформулированы менее четко и в отдельных случаях противоречат друг другу. В подобных ситуациях требуется внесение поправок в законодательство или принятие новых правовых актов для упразднения противоречащих положений и обеспечения гибкости работы систем для интеграции больших данных. Тем не менее, процесс изменения правовых инструментов может оказаться сложным и длительным. НСУ могут обойти этот вопрос стороной, осуществляя инициативы по большим данным, однако в дальнейшем неизбежно придут к пониманию того, что это неотъемлемая часть процесса. Необходим набор комплексных и вместе с тем гибких правовых инструментов, включающих закон о статистике и меморандум о взаимопонимании. Для разработки понадобятся компетенции специалиста, поиск которого потребует усилий со стороны НСУ.
- ❖ **Конфиденциальность, этические нормы и доверие.** В сфере применения больших данных эти аспекты имеют особое значение. Новые технологии производят огромные массивы персональных данных и ставят НСУ перед необходимостью заверить общественность в том, что государственные институты руководствуются этическими принципами и практиками работы с этими источниками данных. Процедуры анонимизации или псевдонимизации данных традиционных источников могут оказаться неэффективными в силу объема и многообразия больших данных. В свою очередь это повышает риски, связанные с обеспечением конфиденциальности, в ходе обмена микроданными больших наборов данных. Принимая во внимание масштаб больших данных, сделаем особый акцент на соблюдении этических норм. Этот вопрос представляет собой особую важность, поскольку, несмотря на действие в ряде стран законов, обеспечивающих доступ НСУ к наборам первичных данных, требуются стандарты этики, определяющие на практике принятие НСУ решений о необходимости запроса наборов первичных данных, что поможет предотвратить неэтичное обращение с персональными данными. Следовательно, перед НСУ еще более остро стоит вопрос обновления стандартов конфиденциальности и этики и повышения наряду с этим прозрачности процессов для обеспечения и сохранения доверия общественности. Текущая работа в рамках создания Методов повышения

конфиденциальности ООН (МПК)¹⁴ и руководства по соблюдению этических норм при использовании машинного обучения в исследованиях и для целей статистики Статистического управления Соединенного Королевства¹⁵ закладывает основы для внедрения ряда стандартов, обеспечивающих конфиденциальность, соблюдение этических норм и доверие.

- ❖ **Доступ к большим данным и модели партнерства.** Большие данные, как правило, производятся компаниями частного сектора и являются побочным продуктом предоставления услуг. Для получения доступа к этим данным требуется не только благоприятный законодательный режим, но и подходящая модель партнерства, тип которой зависит от характера больших данных и готовности частных компаний к предложенному партнерству. Эти компании получают конкурентные преимущества, контролируя распространение данных, что снижает их заинтересованность в обмене большими данными. Если компании дают согласие на распространение больших данных, то, как правило, это небольшие поднаборы или наборы микроданных без детального описания критериев выборки поднаборов. В силу этих ограничивающих факторов важно выбрать приемлемую модель партнерства с частными держателями данных. Модели партнерства отличаются в зависимости от нормативно-правовых требований к частным поставщикам данных, вознаграждения держателей данных (при наличии такового) и услуг по обработке данных частными компаниями.
- ❖ **Издержки и проведение закупок.** Расходы, связанные с большими данными, заметно рознятся в зависимости от сложности внутриорганизационных процессов. НСУ могут столкнуться с издержками уже на этапе получения доступа к данным, поскольку в рамках партнерских соглашений с частными держателями данных, вероятно, будет взиматься плата за предоставление доступа к данным. Издержки могут вырасти более по мере того, как НСУ начнут учитывать затраты на человеческие ресурсы и инфраструктуру передачи, хранения, обработки и анализа больших наборов данных. Принимая во внимание объемы и скорость прироста больших данных, затраты НСУ на передачу, обработку и анализ такого количества информации могут вырасти в разы. Вероятнее всего, частные держатели данных изменят стоимость доступа к данным в будущем. В определенном смысле издержки станут определяющим фактором при принятии НСУ решений относительно самостоятельной обработки данных или приобретения необходимых услуг

¹⁴ Целевая группа по методам сохранения конфиденциальности, [Руководство ООН по использованию компьютерных технологий, обеспечивающих конфиденциальность](#) (апрель 2019).

¹⁵ UK Statistics Authority, [Ethical considerations in the use of Machine Learning for research and statistics](#) (October 2021).

частных держателей данных. Независимо от принятого решения, НСУ должны быть уверены в том, что их системы закупок и соответствующие отделы готовы к переговорам с частными держателями больших данных. В ходе осуществления инициативы по использованию больших данных могут возникнуть серьезные препятствия, если НСУ не создадут команду специалистов по закупкам, способную вести переговоры с частными компаниями о предоставлении доступа к данным.

В этом разделе описываются проблемы, с которыми сталкиваются НСУ, а следующих разделах представлена более подробная информация о потенциальных мерах, которые могут быть приняты для их искоренения. Следующие разделы содержат тематические исследования по мерам НСУ, направленным на решение этих проблем.

Законодательная база

Законодательство в сфере статистики определяет решения, принимаемые странами в чрезвычайных обстоятельствах, когда потребность в актуальной информации становится критической. Во время пандемии COVID-19 страны с развитой законодательной базой в области статистики с меньшими трудностями отреагировали на потребности в данных со стороны директивных органов. Следовательно, прочные законодательные основы, содействующие развитию новых источников данных и технологий при соблюдении этических норм и конфиденциальности, крайне важны для удовлетворения растущих потребностей в своевременных, подробных и точных статистических данных.

Во многих странах Азиатско-Тихоокеанского региона действуют **жесткие и устаревшие законы о статистике, сдерживающие использование новых источников данных**. По большей части эти правовые рамки напрямую ограничивают масштабы производства статистики проведением опросов, переписей и использованием административных источников данных. В ряде случаев законы проработаны детально и являются чрезмерно директивными по своему характеру, что сдерживает инновации и возможности для эксперимента с новыми источниками данных в официальной статистике.

Типовой закон об официальной статистике и общие положения

Многие страны региона уже осуществляют или рассматривают возможности для обновления своего законодательства в сфере статистики. Хорошую основу для этой работы закладывают Руководство по модернизации статистического законодательства¹⁶ и Типовой закон об официальной статистике¹⁷ (ТЗОС), разработанные ЕЭК. ТЗОС, созданный с учетом основополагающих принципов официальной статистики ООН, предусматривает дальнейшую трансформацию и модернизацию национальных статистических систем (НСС). С законом связаны перспективные ожидания стран создать НСС, которых они хотят и которые необходимы. Следующие страны региона приняли ТЗОС: Кыргызстан, Узбекистан, Армения и Мальдивы.

¹⁶ [Руководство по модернизации статистического законодательства](#) (издание Организации Объединенных Наций, 2018).

¹⁷ [Типовой закон об официальной статистике](#) (издание Организации Объединенных Наций, 2016).

ТЗОС содержит важнейшие положения касательно модернизации НСС, в частности, затрагивает следующие аспекты: полномочия главного статистика, механизмы и инструменты координации, источники данных, обеспечение конфиденциальности и связь с другими нормативно-правовыми актами.

Основные положения Типового закона об официальной статистике (ТЗОС)

- Определения и принципы
- Определение рамок национальной статистической системы (НСС): функций национальных статистических управлений (НСУ) и других производителей официальной статистики (ДПОС) при неукоснительном соблюдении принципов
- Роль главного статистика
- Статистический совет/исполнительный комитет
- Механизмы и инструменты координации, программы
- Источники данных, сбор, обмен, обработка, распространение, обеспечение конфиденциальности данных
- Связь с другими законодательными актами и нарушения
- Прочие положения

Регулирование доступа к данным третьих лиц

Рассматривая вопрос об обеспечении доступа НСУ к данным частного сектора, следует отметить, что в ряде стран действует законодательство, четко определяющее и ограничивающее круг источников информации для целей официальной статистики. Это законодательство обходит стороной новые альтернативные источники данных, включая информацию частного сектора. В другой группе стран отсутствуют понятные руководства по обеспечению доступа к этим новым источникам данных, выстраиванию партнерских отношений и выбору надлежащих моделей финансирования.

Законодательство некоторых стран других регионов открыто признает официальную статистику общественным благом и регулирует доступ к альтернативным источникам данных, в том числе к данным частного сектора. В законодательстве таких стран, как Соединенное Королевство (Закон о цифровой экономике, 2017), Эстония (Закон о статистике, 2010) и Оман (Закон о статистике) содержатся положения, обеспечивающие свободный доступ к данным частного сектора и прочим данным в целях официальной статистики. В Европейском союзе на общеправительственном уровне рассматриваются и формулируются условия предоставления доступа к данным в формате коммерческие структуры – государственные организации (Б2Г) для целей включающих, но не ограниченных

подготовкой официальной статистики.¹⁸ Более того, Европейская Комиссия учредила Группу экспертов по обмену данными между частным сектором и правительственными органами.

Во многих странах, в законодательствах которых отсутствуют правовые положения, обеспечивающие общий доступ к альтернативным источникам информации, доступ к специализированным отраслевым данным частного сектора предоставляется регулятором или на основании других форм институционального управления. Зачастую это относится к регулятору телекоммуникационной связи, который управляет доступом к данным оператора мобильной связи, однако известно и о других случаях в других секторах, к примеру, в налогообложении.

Как правило, проблемы связаны с противоречиями положений разных законодательных актов, возникают при получении доступа и обмене данными между государственными учреждениями. Примером может служить ситуация, когда регулятор получает данные, однако законодательство запрещает дальнейшую передачу сведений внутри правительственных структур даже для целей статистики по соображениям конфиденциальности. НСУ надлежит рассмотреть положения ТЗОС, активно поддерживать и участвовать в процессах гармонизации законодательства, прежде всего, в области предоставления доступа к данным для производства официальной статистики.

Руководство по модернизации статистического законодательства поддерживает предоставление прямого и свободного доступа к информации государственного и частного секторов в целях официальной статистики:

Параграф 61. Закон о статистике должен включать положения, обязывающие все национальные и местные органы власти, а также частные организации предоставлять имеющиеся в их распоряжении данные производителям официальной статистики, если это необходимо в статистических целях и соответствует политической конфигурации государственного сектора соответствующей юрисдикции. Данные должны предоставляться бесплатно на уровне детализации, необходимом для формирования официальной статистики (включая идентификаторы), и с метаданными, которые позволяют оценить качество данных. Закон о статистике также должен включать положения, обязывающие держателей административных данных консультироваться с НСУ и при необходимости с другими производителями официальной статистики, если они планируют разработать новый набор данных или провести серьезный

¹⁸ Европейский союз, Совещание экспертов высокого уровня по вопросам обмена данными между коммерческими структурами и государственными органами (Б2Г), [Разработка европейской стратегии по обмену данными между коммерческими и государственными структурами в общественных интересах](#). (Люксембург, 2020).

пересмотр своих методов сбора или обработки данных таким образом, что это может повлиять на данные, предоставляемые для официальной статистики. Следует побуждать руководителей статистических служб к активному содействию дальнейшему совершенствованию административных и частных источников данных, например, путем использования статистических классификаций, особенно в наборах данных, которые могут использоваться в статистических целях.

Меморандум о взаимопонимании (МоВ)

МоВ позволяет эффективно выстраивать отношения с поставщиками данных. Доступ к данным должен регулироваться на основе МоВ и соглашения о сборе данных. Приведем пример Австралийского бюро статистики (АБС), которое применило упрощенную версию МоВ и соглашений о сборе данных под контролем директора по данным АБС и при поддержке отдела по вопросам политики и юридического сопровождения.¹⁹

Минимальный перечень вопросов МоВ с поставщиками данных должен включать следующее:²⁰

1. правовые основы (законы, статистические программы, положения контракта);
2. описание цели передачи данных/получения доступа к данным;
3. подробное описание исследуемых данных, включающее периодичность и современность позиций данных;
4. стандарты качества и отчеты о проверке качества;
5. технические требования;
6. механизмы сотрудничества для повышения соответствия данных статистическим требованиям (консультации и механизмы партнерства);
7. контактные лица.

Прочие законодательные акты

Другие законы и постановления могут оказать влияние на законодательство в сфере статистики и обеспечение доступа НСУ к альтернативным источникам данных. К их числу можно отнести следующее:

¹⁹ Australian Bureau of Statistics (ABS), *Access to Privately Held Data – Australia's Example*, доклад, подготовленный к Совещанию экспертов по модернизации статистического законодательства, Женева, сентябрь 2021.

²⁰ Gabriel Gamez, *Governance beyond the national statistical system*, презентация в режиме онлайн в рамках сессий Кафе по статистике ЭСКАТО, 16 августа 2021.

- Государственный закон о данных
- Закон о доступе к общественной информации
- Закон об архивном деле
- Закон о конфиденциальности и защите информации
- Закон об электронных коммуникациях
- Закон о цифровой экономике или цифровой трансформации
- Различные законодательные акты, регулирующие использование административных и других источников данных (реестры, большие данные, наборы данных частных организаций, искусственный интеллект и прочее)
- Закон об использовании данных в электронной торговле
- Отраслевое законодательство

Необходимо гармонизировать данные законодательные акты и постановления для обеспечения своевременного доступа к любому источнику данных в целях официальной статистики при условии соблюдения этических норм и обеспечения конфиденциальности. НСУ следует принимать активное участие в разработках и проведении консультаций по любым законодательным актам и постановлениям в сфере официальной статистики, регулирующим доступ, использование, повторное использование данных и защиту конфиденциальности государственным и частным сектором.

Неприкосновенность личной информации, этические нормы и вопрос доверия

Обеспечение конфиденциальности и защита персональных данных – аспекты, вызывающие серьезную обеспокоенность при рассмотрении вопроса использования больших данных в целях официальной статистики. Национальные нормативно-правовые базы стран имеют отличия, по этой причине НСУ необходимо перестроиться в рамках этого законодательства и создать надлежащие модели партнерства и доступа к данным с учетом национального контекста.

Выделяют три аспекта обеспечения конфиденциальности данных, каждый из которых регулируется соответствующим законодательством. Аспекты включают: неприкосновенность частной жизни, электронную связь и конфиденциальность статистики.²¹ Неприкосновенность частной жизни пользователей технологий обеспечивается государственным законом о защите данных или законом о конфиденциальности. В Европейском союзе эта сфера регулируется Общим регламентом по защите персональных данных (ОРЗПД). Законоположения в сфере электронной связи определяют основы деятельности пользователей персональных и неперсональных данных и обеспечивают неприкосновенность личной информации в сфере электронных коммуникаций. В Европейском союзе в дополнение к ОРЗПД действует Постановление об уважении неприкосновенности частной жизни и защите персональных данных в электронных сообщениях. Конфиденциальность статистики – один из главных аспектов Основополагающих принципов официальной статистики. Законы о статистике, соответствующие Основополагающим принципам, обеспечивают конфиденциальность статистических данных.

Защита конфиденциальности данных и личной информации зависит не только от законодательства, но и от инфраструктуры доступа и использования больших данных. Чем больше первичных или обезличенных данных попадает за защитный экран владельцев данных, тем выше риски и возможность появления проблем, связанных с утечкой информации. Некоторые страны применяют националистические подходы в сфере данных для того, чтобы конфиденциальная информация не оказалась за пределами страны, при этом данные хранятся на

²¹ Erki Saluveer, *Mobile positioning data and data access*, презентация в ходе Совещания экспертов по модернизации статистического законодательства ЕЭК ООН / ЕАСТ, Позитивум, Эстония, ноябрь 2021.

расположенных на территории страны серверах. По условиям некоторых партнерских соглашений НСУ не получают доступа к первичным данным, им предлагают агрегированные результаты обработки и анализа, проведенных держателем данных на своем сервере.

Примеры стран: Австралия

Рассмотрим пример Австралии в сфере данных государственного сектора. Австралийское бюро статистики (АБС) разработало важный интегрированный ресурс данных в рамках межучрежденческого проекта по интеграции данных (МПИД). МПИД – интегрированная и безопасная система персональных данных по здравоохранению, образованию, государственным выплатам, доходам и налогообложению, занятости и демографии (включая перепись). АБС создало прочные регулятивные основы для реализации проекта, в рамках которых проводится оценка всех мер по интеграции с точки зрения общественной пользы, анализируется и регулируется воздействие на неприкосновенность личной информации.²²

Важно завоевать и сохранить доверие со стороны общественности и избежать ошибочных представлений об использовании личной информации. Более того, важно обеспечить доверие граждан к статистике, производимой на основе источников больших данных. НСУ необходимо приложить максимальные усилия для оправдания ожиданий общества, главный инструмент для достижения этой цели – открытый диалог о том, какая информация и как используется. Заключая новые партнерские соглашения с компаниями частного сектора, НСУ не должны вызывать у граждан ощущение того, что они не придают значения существующим правилам неприкосновенности личной информации. НСУ надлежит разработать и применить прозрачные механизмы управления данными, а также этические нормы. Более того, НСУ должны гарантировать реализацию этих норм и доводить до сведения общественности соответствующую информацию на протяжении всего статистического производственного процесса.

²² ABS, *The Role of The Australian Bureau of Statistics in Data Governance and Stewardship in Australia from a Regulatory Perspective*, доклад, подготовленный к Совещанию экспертов по модернизации статистического законодательства, Женева, сентябрь 2021.

Примеры стран: Соединенное Королевство

Статистическое управление Соединенного Королевства разработало инструмент этической самооценки²³, который создает рамки для оценки этических аспектов проекта в ходе исследовательского цикла. Основой данного инструмента являются шесть этических принципов.

Принцип 1: Использование данных приносит очевидные выгоды пользователям и служит общественному благу.

Принцип 2: Личная информация субъекта данных (лица или организации) находится под защитой, соблюдается конфиденциальность и безопасность хранения информации, а вопрос согласия рассматривается надлежащим образом.

Принцип 3: Учитываются риски и ограничения, связанные с новыми методиками и/или технологиями, в достаточной мере осуществляется человеческий надзор за соответствием используемых методов принятым стандартам добросовестности и качества.

Принцип 4: Использование данных и методов соответствует нормативным требованиям, в частности, законодательства о защите данных, Закона 1998 года о правах человека, Закона 2007 года о службе статистики и регистрации, соответствует обязанностям по соблюдению конфиденциальности в рамках общего права.

Принцип 5: Мнения общественности учитываются в свете используемых данных и ожидаемых выгод от проведения исследования.

Принцип 6: Доступ, использование и передача данных прозрачны, информация доводится до сведения общественности в четкой и доступной форме.

Кроме того, Статистическое управление Соединенного Королевства и Центр прикладной этики данных разработали широкий по своему содержанию контрольный перечень вопросов по этике использования данных третьих лиц в исследованиях и статистике.²⁴ Перечень содержит список вопросов, затрагивающих основные этические аспекты использования данных третьих лиц в исследованиях и статистике.

Другие заинтересованные НСУ имеют свободный доступ к перечню и могут воспользоваться им как руководством. Вопросы ориентированы на исследование

²³ См. <https://uksa.statisticsauthority.gov.uk/the-authority-board/committees/national-statisticians-advisory-committees-and-panels/national-statisticians-data-ethics-advisory-committee/ethics-self-assessment-tool/>.

²⁴ UK Statistics Authority, *Using Data from Third Parties for Research and Statistics: High-Level Ethics Checklist* (October 2021).

следующих аспектов:

1. общественное благо;
2. методики и качество;
3. прозрачность;
4. соответствие законодательству;
5. мнение и участие общественности;
6. конфиденциальность и безопасность данных.

Обеспечение доступа к большим данным и модели партнерства

Национальные статистические управления и государственные органы получают доступ к данным частных организаций разными способами. Отличия определяются **уровнем доступа** к данным и **моделью доступа** к данным. Под уровнем доступа к данным подразумевается доступ НСУ к первичным данным, псевдонимизированным данным или обобщенной информации. Модели доступа к данным определяют условия доступа к сведениям. Наиболее часто выделяют такие типы моделей, как обязательное предоставление данных, предоставление данных на добровольной основе и коммерческие соглашения, при этом существуют и другие механизмы, постепенно появляются новые выигрышные для всех сторон модели, в частности, партнерства. Каждая модель отвечает требованиям национальной конъюнктуры и имеет свои преимущества. Важно понимать, какую роль в ней играет НСУ, и какие недостатки присущи каждой модели.

В соответствии со своей национальной законодательной базой НСУ выбирают наиболее приемлемый тип модели доступа к данным, учитывая существующие потребности. Модели также могут отличаться в зависимости от источника данных. Существуют разные уровни доступа и модели, однако единого универсального подхода нет.

6.1 Уровни доступа к данным

В официальной статистике выделяют два основных уровня доступа к данным частного сектора: доступ к **псевдонимизированной информации** и **обезличенным данным**. Впрочем, в зависимости от источника данных некоторые НСУ могут получить и первичные данные.

Псевдонимизированные данные – это первичные обезличенные микроданные (личность не может быть установлена). НСУ могут получить доступ к псевдонимизированным данным для проведения дальнейшего анализа с использованием собственной инфраструктуры при обеспечении конфиденциальности данных и безопасности. Данный подход максимально расширяет возможности НСУ с точки зрения исследования и анализа данных.

С другой стороны, НСУ могут получить обезличенные агрегированные данные, при этом повторная идентификация первичных данных станет невозможной. Часто НСУ получают обезличенные данные или информацию, которые подготовлены на основе анализа, проведенного с использованием инфраструктуры компании. В подобных случаях НСУ принимают участие в разработке методик для обработки данных компании. Данный подход снижает риски, связанные с возможной утечкой информации, однако ограничивает возможности НСУ с точки зрения исследования и анализа данных.

Примеры стран: Индонезия

Центральное статистическое бюро Индонезии (ЦСБ) заключило коммерческое соглашение с крупнейшим оператором мобильной связи Телкомсел. ЦСБ Индонезии получает доступ к обработанной информации, а не к данным. Тем не менее, в целях разработки и тестирования методик узкому кругу исследователей НСУ предоставляется доступ к доле первичных данных в «песочнице» на сервере компании. По мере завершения создания методики компания запускает алгоритм для обработки всех первичных данных и обменивается результатами с ЦСБ.

6.2 Модели доступа к данным

Обязательное предоставление данных

Обязательное предоставление данных частным сектором может предусматриваться национальным законодательством, в частности, Законом о статистике, Законом об информации, распоряжениями или другими правовыми актами. Подобное законодательство обеспечивает прямой доступ к любому источнику больших данных в общественных целях, к примеру, для официальной статистики или проведения исследований. Проблемы возникают в тех случаях, когда данные имеют трансграничный характер и/или частные держатели информации находятся в другой стране. В этом случае другая модель обеспечения доступа к данным, предоставление данных на добровольной основе или в формате коммерческого соглашения, представляется наиболее приемлемой. Приведем пример в разделе ниже.

Примеры стран: Оман

В Омане статья 19 Закона о статистике и информации, принятого Королевским указом в 2019 году, гласит: *«для выполнения своих задач Национальный центр статистики и информации может запросить любые необходимые данные, информацию и административные документы у частных лиц, правительственных и неправительственных организаций на безвозмездной основе»*.²⁵

Более того, упомянутый Королевский указ (статья 32) предусматривает меры в отношении любой неправительственной организации, которая отказывается предоставить данные и информацию, запрашиваемые в соответствии с положениями настоящего закона и необходимую лицам, осуществляющим деятельность в сфере официальной статистики.

Впрочем, регулирование больших данных может быть затронуто и обеспечено другими государственными стратегиями и законодательными актами, включающими национальную стратегию формирования цифрового правительства, стратегию по развитию цифровой экономики, государственную стратегию развития искусственного интеллекта, национальную стратегию в области данных и другие общегосударственные стратегии или законодательные акты, которые регулируют сферу официальной статистики как одно из направлений использования и применения больших данных. Данное законодательство может описывать процедуры обеспечения доступа и передачи больших данных в соответствии с принципами, лежащими в его основе.

Примеры стран: Соединенное Королевство

В Соединенном Королевстве обеспечение доступа и использование больших данных государственными органами регулируются Законом о цифровой экономике и Национальной стратегией в области данных. Эти документы нацелены на расширение использования больших данных в широком спектре услуг, в официальной статистике и исследованиях, частью которых они являются. Более того, Закон о цифровой экономике обеспечивает право статистических ведомств Соединенного Королевства на получение консультаций и информации о внесенных изменениях со стороны поставщиков данных. Надлежит информировать органы статистики об изменениях методов сбора и обработки информации.

²⁵ Ahmed Al Mufarji, National Centre for Statistics and Information (NCSI), [*Oman's experience in Utilizing Mobile Positioning Data for Official Statistics*](#), презентация в ходе Вебинара по вопросам глобальной сети в режиме онлайн, апрель 2021.

«Глава 7 части 5 Закона о цифровой экономике вносит поправки в Закон 2007 года о службе статистики и регистрации, предоставляя Статистическому управлению Соединенного Королевства (и Управлению национальной статистики как его исполнительному органу) более широкий и беспрепятственный доступ к данным государственного и частного секторов для осуществления официальных функций Статистического управления по производству официальной статистики и проведения статистических исследований».²⁶

Примеры: Европейский союз

Европейская комиссия признает в Европейской стратегии в области данных, что «в настоящее время данные частного сектора недостаточны. Они необходимы государственным органам для совершенствования процессов разработки фактологически обоснованной политики и предоставления государственных услуг, таких как управление мобильностью, расширение охвата официальной статистики, повышение ее актуальности, и, следовательно, повышения обоснованности этих услуг в контексте последних тенденций общественного развития. Рекомендации группы экспертов, учрежденной Комиссией, включают формирование национальных структур для обмена данными между коммерческими и государственными организациями (Б2Г), разработку надлежащих стимулирующих мер для создания культуры обмена данными и изучение возможностей законодательной базы ЕС в области регулирования повторного использования данных частных компаний государством в общественных интересах».²⁷

Европейская комиссия при поддержке экспертов разработала Руководство по передаче данных частного сектора.²⁸

Государственные законы о статистике и национальные стратегии в области данных определяют общий порядок доступа к данным частного сектора, при этом **отраслевое законодательство может регулировать доступ к определенным источникам данных, в определенных целях и при соблюдении особых условий.** Таковы, среди прочего, примеры регуляторов в сфере финансов и телекоммуникаций. Тем не менее, в отдельных случаях секторальное

²⁶ UK Statistics Authority, [Digital Economy Act: Research and Statistics Powers](#).

²⁷ [Сообщение Комиссии в адрес Европейского парламента, Совета, Европейского экономического и социального комитета и Комитета регионов, Европейская стратегия в области данных](#) (COM/2020/66).

²⁸ Европейская комиссия, [Руководство по передаче данных частного сектора](#), 7 июня 2022.

законодательство предоставляет доступ определенным учреждениям или регулятору и запрещает дальнейшую передачу данных, к примеру, НСУ.

Подобные противоречия в положениях законодательства могут препятствовать эффективному обмену данными. Прежде всего, это касается официальной статистики, области, в которой по умолчанию следует придерживаться официальных принципов. Возможны существенные затруднения, если вопросы передачи и повторного использования данных третьей стороной, например, НСУ, строго регламентируются законодательством. Проблемы могут быть связаны и с инфраструктурой, например, в тех случаях, когда регулятор не использует инфраструктуру, отвечающую требованиям национального законодательства о конфиденциальности данных и безопасности.

Предоставление данных на добровольной основе

Доступ к данным частного сектора может предоставляться на добровольной основе, если компании передают информацию государственным структурами по собственному желанию. Эти добровольные меры могут быть частью политики компании в области корпоративной социальной ответственности (КСО) или взаимовыгодного соглашения, в рамках которого компания получает преимущества в результате передачи данных НСУ или другим государственным органам. Выгоды заключаются в экономии времени и ресурсов, если сравнивать с проведением опросов при сборе информации традиционным способом или получением необходимой информации и данных от НСУ в ответ, к примеру, сведений о доле рынка и т.д.

Примеры стран: Австралия

До наступления пандемии COVID-19 АБС с 2012 года проводило сбор данных сканирования четырех крупнейших цепочек супермаркетов. Этот важный источник данных, характеризующий 25% ценовых показателей, использовался для расчета индекса потребительских цен в стране. Супермаркеты предоставляли данные добровольно на основании письменного соглашения между цепочками супермаркетов и АБС. Таков пример бесприигрышного положения. Супермаркеты предпочли обмен данными сканирования личному сбору данных, а АБС получило более подробную информацию о транзакциях в сравнении с той, которая могла бы содержаться в опросах.

Во время пандемии COVID-19 появились новые потребности в данных, фокус внимания был быстро переведен на микроданные частного сектора, прежде всего,

крупных финансовых институтов, предприятий розничной торговли, информацию по электроснабжению и связи.²⁹ Глава Австралийского бюро статистики обратился за поддержкой к главам компаний. Конфиденциальность данных была обеспечена на тот момент законодательством. АБС добилось получения на добровольных началах обезличенных обобщенных микроданных частного сектора на основании письменного обязательства, предусматривающего не только использование данных в агрегированном виде, но и неразглашение информации третьей стороне.

Соглашения вступили в силу спустя два месяца после разразившейся в Австралии эпидемии. АБС согласилось на получение данных в любой форме с использованием любых безопасных технологий, применяемых коммерческими организациями для передачи данных. Открытость и гибкость подходов АБС к формату данных и технологиям передачи позволили получить данные во время пандемии быстро, однако создали проблемы в операционной деятельности. Более того, возникла необходимость в разработке правил, регулирующих вопросы интеллектуальной собственности и финансовой ответственности.

Коммерческие или партнерские соглашения

И коммерческие, и партнерские/взаимные соглашения заключаются на основании договора с частными поставщиками данных. Как правило, они подразумевают выплату финансового вознаграждения в обмен на данные компании, информацию или другие услуги, связанные с данными. При заключении партнерских соглашений составляется договор, обеспечивающий равные преимущества обеих сторон и предусматривающий выплату небольшого вознаграждения поставщикам данных, или его отсутствие. Партнерские/взаимные соглашения обоснованы в тех случаях, когда частные поставщики данных не имеют или предпочитают не вкладывать средства в персонал, необходимый для проведения анализа больших данных при подготовке официальной статистики и информации для выработки политики. Компании могут установить рыночную цену или льготную цену на свои данные и услуги. Некоторые НСУ выбирают коммерческие или партнерские соглашения при отсутствии национального законодательства, регулирующего доступ к данным частного сектора. Другие НСУ неохотно идут на заключение коммерческих соглашений, не желая выплачивать вознаграждение за информацию, которая, по их

²⁹ ABS, [Access to Privately Held Data – Australia's Example](#), документ, подготовленный к Совещанию экспертов по модернизации статистического законодательства, Женева, сентябрь 2021.

мнению, является общественным благом. При отсутствии законодательного регулирования такой подход может привести к тому, что НСУ полностью лишатся доступа к данным частного сектора.

Важно отметить, что в тех случаях, когда компании требуют оплату за предоставление информационных сервисов и сведений, что не распространяется на первичные и псевдонимизированные данные, они инвестируют в создание надлежащей инфраструктуры анализа данных, средства защиты конфиденциальности и безопасности, а также визуализации данных. Следовательно, они предоставляют услуги с добавленной стоимостью, которые требуют существенных предварительных инвестиций и покрытия расходов. Если компании выбирают формат партнерских/взаимных соглашений, им удастся избежать инвестиционных издержек. Компании возлагают обязанности по анализу данных на НСУ и получают информацию о рынках на основании расчетов и агрегированных данных, подготовленных НСУ. Частные владельцы данных и НСУ должны проанализировать издержки и преимущества в каждом из случаев и совместно принять решение относительно оптимального механизма сотрудничества, удовлетворяющего интересам обеих сторон.

Примеры стран: Индонезия³⁰

При отсутствии законодательства, регулирующего доступ к данным частного сектора, ЦСБ Индонезии заключило коммерческое соглашение с Телкомсел о предоставлении сведений о мобильности. Анализ проводился на основе данных мобильного позиционирования (ДМП), которые не интересовали оператора мобильной связи с коммерческой точки зрения до начала проекта. Однако по мере его реализации Телкомсел пришла к пониманию того, что ДМП имеют коммерческую ценность и открывают возможности для предоставления нового вида услуг как органам государственной власти, так и частному сектору. Телкомсел сделала существенные финансовые вложения в инфраструктуру хранения и анализа данных.

Ограниченный круг исследователей ЦСБ имеет доступ к небольшой выборке ДМП, которая хранится Телкомсел для исследовательских целей и разработки методик. В дальнейшем методика применяется для обработки всех первичных данных и получения статистики, направляемой в ЦСБ, часть информации публикуется как официальная статистика.

³⁰ Titi Kanti Lestari, [Data Partnership for the Use of Mobile Positioning Data for Official Statistics: Indonesia Experience](#), презентация в ходе сессий Кафе по статистике ЭСКАТО в режиме онлайн, 30 августа 2021.

6.3 Вовлечение других заинтересованных сторон

Принимая во внимание такие факторы, как тип источника данных, национальное законодательство и корпоративные правила владельцев данных, НСУ могут сотрудничать с другими заинтересованными сторонами, не только с поставщиками данных. К их числу относятся регулятор, агентство по защите персональных данных, научные учреждения, международные организации, НКО и другие компании частного сектора. Привлечение других партнеров приносит пользу, если содействует укреплению потенциала проекта.

Примеры стран: Соединенное Королевство

Компания «Фейсбук» не передает личные данные государственным органам, однако предоставляет информацию университетам в исследовательских целях. В Соединенном Королевстве УНС сотрудничает с университетами для получения агрегированных данных на основе информации Фейсбук.³¹ Информация передается в рамках программы Использование данных Фейсбук в благих целях³² сотрудникам проекта WorldPop Университета Саутгемптона. Аналитики WorldPop проводят агрегирование, повторное масштабирование и анализ данных, которые получают через Сеть передачи данных о мобильности COVID-19. Эти данные уже применялись в ходе экспериментального анализа по мониторингу изменений плотности населения в прибрежных городах.

В некоторых случаях желательно избегать участия регулятора. Таков случай, в котором Центральный банк, выступая в роли регулирующего органа, может стать одним из потребителей информации. Участие Центрального банка в процессах сбора информации и предоставления доступа к данным его субъектов объясняют нежелание поставщиков данных проводить обмен информацией. НСУ необходимо использовать преимущества своего независимого положения и не привлекать регулятора к участию в подобных проектах.

Известно о многих примерах успешного сотрудничества в регионе между НСУ и национальными космическими агентствами или агентствами графических или картографических работ. Проводилась работа по подготовке спутниковых изображений или геопространственных данных для производства широкого спектра официальной статистической информации.

³¹ См. <https://datasciencecampus.ons.gov.uk/using-facebook-data-to-understand-changing-mobility-patterns/>.

³² См. <https://dataforgood.facebook.com/>.

6.4 Проблемы, связанные с заключением соглашений на добровольной и коммерческой основе

Не смотря на то, что применение каждого подхода может быть обосновано в зависимости от условий разных стран, неизбежно появление определенных проблем. Некоторые могут возникнуть на этапе перехода от экспериментального проекта к масштабной интеграции источников больших данных в статистическое производство. Другие появляются по причине долгосрочной неопределенности, связанной с существующими подходами.

При отсутствии благоприятных законодательных основ, регулирующих доступ к данным альтернативных источников, при заключении добровольных и коммерческих соглашений возникают две основные проблемы. Во-первых, стабильность подобных соглашений ставит под вопрос возможность **непрерывного производства официальной статистики в долгосрочной перспективе. При заключении как добровольных, так и коммерческих соглашений нет уверенности в том, что поставщики данных продолжат** сотрудничество по истечении изначально согласованного срока. В связи с этим возникают вопросы относительно возможностей НСУ непрерывно производить статистику с использованием этого источника данных.

Во-вторых, в долгосрочной перспективе НСУ, возможно, **не смогут контролировать или получать информацию об изменениях в методиках сбора и структуре данных, принятых владельцами данных.** Эта проблема, вероятно, сохранится даже при условии создания благоприятной нормативно-правовой базы, регулирующей доступ к данным. При отсутствии комплексного и долгосрочного подхода к решению этой проблемы со временем она повлияет на качество и сопоставимость данных статистики. Упомянутые вопросы должны быть четко изложены в контрактах и МоВ, которые заключаются с особым условием, обязующим производителей коммерческой информации консультироваться с НСУ перед внесением каких-либо изменений в методики сбора данных.

Независимо от способа доступа к данным или модели сотрудничества, НСУ должны понимать потребности организаций, в которые они обращаются. НСУ следует создать внутреннюю команду специалистов, обладающую необходимыми навыками, обеспечить надлежащую правовую поддержку переговоров о новом сотрудничестве, получить нужные технологии. НСУ должны выступать в роли хороших координаторов и регуляторов, быть способными убедительно представлять решения по взаимовыгодному сотрудничеству.

Издержки и проведение закупок

Запуск проекта по большим данным, как правило, связан с появлением ряда издержек разного характера. При этом необходимые затраты и инвестиции, связанные с проектами по большим данным, зависят от многих факторов. Во-первых, независимо от типа источника данных, большинство государственных институтов и НСУ запускают экспериментальные проекты по большим данным для того, чтобы оценить потенциал и применимость определенного источника данных. Следовательно, важно разделять издержки экспериментальных проектов и затраты, связанные с интеграцией и использованием источников больших данных при производстве официальной статистики, поскольку они заметно отличаются.

Издержки **экспериментальных проектов** по большим данным могут включать **затраты на сбор данных** и **расходы на внешние ресурсы** для выполнения задач подрядчиком или приобретение услуг в рамках партнерства (к примеру, получение доступа и анализ данных), **инфраструктурные издержки**, связанные в основном с хранением и обработкой данных, если анализ проводится организацией самостоятельно, и **расходы на персонал**, если к участию в проекте привлекается специалист. В большинстве пилотных проектов используются небольшие выборки данных, как правило, по одному участку за короткий период времени. При проведении экспериментальных проектов некоторые НСУ получают поддержку со стороны международных организаций в сфере развития, что позволяет существенно ограничить или сократить их расходы.

Масштабное использование больших данных в официальной статистике приводит к расходам другого порядка. Модели доступа к данным и инфраструктура требуют изменений, что становится причиной роста соответствующих затрат. Требуется пересмотр партнерских договоренностей и условий предоставления услуг, а НСУ должны принять решение о том, какие задачи могут быть выполнены самостоятельно, а какие с привлечением внешних организаций. При расширении проектов по большим данным, независимо от типа источника данных, надлежит учитывать следующие факторы, влияющие на инвестиции и издержки:

- ❖ **Модели доступа к данным и партнерских отношений.** В случаях, когда законодательная база обеспечивает свободный доступ к данным частного сектора в целях официальной статистики и проведения исследований или регулятор имеет прямой доступ к данным, НСУ следует рассмотреть

возможности создания новаторских и взаимовыгодных партнерств с частным сектором. Ситуация, в которой предоставляется доступ к выборке данных для эксперимента, отличается от той, когда проводится работа со всей совокупностью демографических данных, представленных несколькими поставщиками одного сектора. Несмотря на нежелание НСУ участвовать в моделях получения данных на коммерческой основе, практические примеры стран Азиатско-Тихоокеанского региона свидетельствуют о том, что НСУ могут достигнуть коммерческих договоренностей с поставщиками данных. Эти договоренности предполагают другой формат сотрудничества в сфере данных, однако являются эффективными с точки зрения затрат, если сравнивать с издержками сбора и анализа данных привычным способом. Многое зависит и от условий рынка, которые определяют структуру ценообразования. Например, если компания является монополистом в сфере данных определенного сектора, она может устанавливать цены на уровне, неприемлемом для бюджета НСУ.

Издержки могут зависеть от типа доступа к данным: доступа к первичным данным (или псевдонимизированным) или агрегированной информации. В случае первичных (или псевдонимизированных данных) затраты на обеспечение доступа могут быть относительно высокими, поскольку НСУ должны принимать во внимание расходы на хранение и анализ первичных данных. Если рассматривать случай агрегированных данных, расходы на обеспечение доступа могут быть ниже в силу относительно невысоких издержек хранения и анализа агрегированных данных. Затраты на мобилизацию и укрепление потенциала кадровых ресурсов сопоставимы с расходами на получение доступа к первичным и агрегированным данным частных компаний.

- ❖ **Инфраструктура.** Когда НСУ планируют провести сбор или получить доступ к большим данным и провести анализ самостоятельно, на первый план выходят вопросы инфраструктуры. Важно отметить, что для разных источников больших данных используются разные инструменты и методы сбора, хранения, очистки, обработки, визуализации и защиты. Решение по инфраструктуре должно обеспечить достаточный объем памяти, аналитический потенциал, возможности для визуализации и защиты безопасности данных. НСУ могут воспользоваться предоставленной инфраструктурой государственных органов, к примеру, правительственными облачными ресурсами, с выделенной или встроенной собственной инфраструктурой. Постепенно в ходе экспериментальных проектов НСУ анализируют собственные потребности. При выборе поставщиков

облачных услуг НСУ должны учитывать любые действующие в стране нормативно-правые акты в сфере локализации данных, которые ограничивают хранение определенных типов данных, как правило, персональной информации, пределами территории страны.

Комитет экспертов ООН по использованию больших данных и обработке и анализу данных для целей официальной статистики (ООН-КЭБД) создал Глобальную платформу³³ или облачную экосистему, поддерживающую совместные усилия по производству официальной статистики на основе новых источников данных и передовых методик. НСУ могут обратиться к ресурсу платформы перед вложением основных инвестиций в инфраструктуру.

- ❖ **Кадровые ресурсы.** НСУ следует вкладывать в ресурсы по повышению профессиональной подготовки сотрудников. Необходимо проведение обучающих мероприятий для имеющегося персонала и/или привлечение новых специалистов. При этом НСУ должны понимать, что испытывают конкуренцию со стороны частного сектора, привлекая талантливых специалистов по большим данным. Как бы то ни было, спектр возможностей для укрепления потенциала на глобальном и региональном уровнях широк: от обучающих курсов по аналитике данных и обработке и распространению больших данных Статистического института для Азии и Тихого океана ООН (СИАТО),³⁴ до учебных программ по определенным источникам больших данных, разработанных целевыми группами ООН-КЭБД,³⁵ мероприятий тематических групп и международных организаций или региональных встреч по обмену знаниями. Известно о многих примерах партнерств между НСУ и университетами и даже организациями частного сектора, в формате которых происходит обмен навыками, проводится обучение и в отдельных случаях предоставляется персонал.

Можно сравнивать затраты на использование больших данных в официальной статистике с издержками традиционного сбора данных и соответствующих методик, однако большие данные чаще всего будут дополнять, а не замещать традиционные источники данных. Этот аспект имеет особое значение для обоснования бюджета. В зависимости от типа источника данных и его значимости в определенной стране, большинство НСУ отмечают существенную минимизацию

³³ См. <https://unstats.un.org/bigdata/un-global-platform.cshtml>.

³⁴ См. https://unsiap.or.jp/e-learning/e_Learning_1.html.

³⁵ См. <https://unstats.un.org/bigdata/task-teams/training/index.cshtml>.

издержек в результате использования альтернативных источников данных. К другим преимуществам данных альтернативных источников относятся высокая частотность, степень детализации и масштаб охвата данными, низкая или отсутствующая административная нагрузка и нагрузка на респондентов. Необходимо учитывать и автоматическую обработку данных при анализе издержек. Источники больших данных и их значение в условиях определенных стран различны, различны и технологии, необходимые для обеспечения доступа, конфиденциальности и безопасности, обработки, анализа и визуализации. Следовательно, оценка стоимости и принятие инвестиционных решений должны проводиться для каждого случая отдельно.

По мере развития моделей доступа к данным и партнерских связей **НСУ столкнутся с необходимостью разработки закупочных процедур**, соответствующих этим моделями. Сегодня многие НСУ не готовы к работе в формате новых партнерств на основе новой операционной модели. Маловероятно, что сотрудники, собирающие полевые данные, обладают необходимой квалификацией для работы в сфере закупок данных частного сектора. Зачастую НСУ не имеют компетентного персонала для проведения переговоров с компаниями частного сектора. Некоторые НСУ создали достаточно крупную команду юристов, обеспечивающую решение растущих и все более разнообразных задач. Появляются разные формы правовых соглашений, при этом многие НСУ не обладают необходимым техническим потенциалом или кадрами для работы в нужном темпе. По этой причине при разработке проектов по большим данным следует рассматривать не только вопросы доступа к данным и инфраструктуры, но и процедуры проведения закупок.

Примеры стран: Сингапур

В Сингапуре агентство Сингстат не имеет потенциала для проведения переговоров с частным сектором. Другое государственное специализированное учреждение Гостех сотрудничает с компаниями частного сектора по вопросам информации и технологий, представляя иные государственные учреждения. Гостех также обсуждает условия доступа к большим данным, которые могут использоваться в операционной деятельности всех правительственных органов, и, более того, в официальной статистике и для предоставления прочих услуг.

Производство статистики с использованием больших данных

В свете существующих проблем очевидными становятся заключения исследований ЭСКАТО по использованию больших данных в официальной статистике, свидетельствующие об экспериментальном характере применения больших данных, которые интегрируются в производство официальной статистики лишь в отдельных случаях. Как правило, большие данные дополняют традиционные виды данных. Важно отметить, что использование больших данных может обогатить официальную статистику. Недавние исследования, включающие разработки ЭСКАТО, подтверждают, что наибольшая польза может быть получена в результате интеграции больших данных в традиционные общепризнанные источники данных в ходе подготовки официальной статистики.

Известно об отдельных случаях, в которых статистическая информация, получаемая только на основе больших данных, была практически идентична той, что производится на основе опросов. Тем не менее, надлежит приложить большие усилия с области стандартизации процессов сбора, обработки и обеспечения качества данных, прежде чем приступить к масштабному использованию больших данных как самостоятельного источника при подготовке официальной статистики. Необходимо детально рассмотреть вопросы обоснованности использования больших данных в официальной статистике и соответствия концепций и определений. Недостатки больших данных во многом связаны с проблемами охвата, к примеру, в тех случаях, когда представлена не целевая группа населения, или имеются погрешности измерений, в частности, при обработке информации, внесенной вручную.

С помощью больших данных можно обогатить разные виды официальной статистической информации. Впрочем, возможности использования отдельных источников данных должны рассматриваться в контексте определенных стран. Вероятно, использование определенного цифрового источника данных в статистическом производстве будет иметь обоснование в одной стране, но не в другой. Важно понимать, что разные качественные характеристики статистики, формируемой на основе больших данных, свидетельствует об экспериментальном характере их применения. Полученные статистические данные будут отвечать требованиям в актуальной информации только в случае соответствия минимальным

пороговым значениям. Производители статистической информации, работающие с большими данными, должны сделать оговорки или напрямую донести до пользователей информацию о качестве статистики. Необходимо поставить пользователей в известность о возможных последствиях использования данной статистикой информации.

Далее представлен неполный перечень потенциальных направлений использования разных источников больших данных для целей официальной статистики. НСУ Азиатско-Тихоокеанского региона проводят экспериментальную работу или изучают возможности для опытов с этими источниками данных.³⁶

- ❖ **Данные наблюдения Земли и геопространственные данные** – информация, поступающая со спутников, самолетов и БПЛА (беспилотных летательных аппаратов, так называемых дронов). Пример - изображения освещенности в ночное время. Может использоваться при оценке распределения населения и расчета демографических показателей, изучении экономической активности, проведении статистических исследований по охране окружающей среды и сельскому хозяйству (счет СЭЭУ, оценка площади лесов и сельскохозяйственная перепись).
- ❖ **Данные сканирования** – подробная информация о продажах потребительских товаров, которая получается путем сканирования штрих-кодов отдельных товаров на кассовом терминале (КТ) в розничных магазинах.³⁷ Сведения широко используются для формирования статистики цен, в частности, расчета индекса потребительских цен (ИПЦ).
- ❖ **Информация о ценах в режиме онлайн** – сведения в электронном виде, получаемые с веб-сайтов компаний розничной торговли. Данные часто используются для подготовки статистики цен, к примеру, при расчетах ИПЦ или индекса цен на рынке жилья (ИЦРЖ). В силу своей доступности полученные в Интернете данные торговых площадок применяются чаще, чем данные сканирования, однако используются с определенными существенными ограничениями и с учетом существующих в странах условий.

³⁶ Более подробная информация представлена в следующих источниках:

- а. ЭСКАТО, [Большие данные для экономической статистики](#), аналитическая записка по статистике, выпуск № 28 (март 2021).
- б. ЭСКАТО, [Большие данные для демографической и социальной статистики](#), аналитическая записка по статистике, выпуск № 29 (апрель 2021).
- в. ЭСКАТО, [Большие данные для статистики окружающей среды и сельского хозяйства](#), рабочий документ, выпуск № 13 (апрель 2021).

³⁷ Организация экономического сотрудничества и развития (ОЭСР), Справочник статистических терминов, [Данные сканирования](#), 11 июля 2005.

- ❖ **Данные мобильной телефонной связи** – данные мобильного отслеживания или геолокации, предоставляемые операторами мобильной связи (ОМС). Данные включают детализацию звонков (ДЗ) и сервисы по определению местоположения абонента (LBS). Сведения используются для формирования широкого спектра статистических данных, в частности, в туризме (въездном, выездном, внутреннем), в демографии, исследованиях миграции, внутригородской ежедневной миграции, транспортных потоков и сезонных перемещений рабочей силы.
- ❖ **Данные умных счетчиков** – сведения о потреблении и производстве электроэнергии отдельных домохозяйств или компаний. Могут дополнять другие виды статистической информации, к примеру, статистические данные о потреблении энергии коммерческими организациями, сезонности в туризме и воздействии на окружающую среду.
- ❖ **Информация о финансовых транзакциях** – сведения о денежных платежах или операциях, предоставленные либо Центральным банком, либо налоговой инспекцией, или поступающие напрямую от платежных операторов. Используются для оценки общего состояния экономики, существующих тенденций и миграционных процессов.
- ❖ **Данные автоматической идентификационной системы (АИС)** или информация о движении судов. Применяются для оценки деятельности человека на море и при проведении предварительных расчетов показателей морской торговли.
- ❖ **Информация социальных сетей** – информация веб-платформ социальных сетей, веб-порталов по трудоустройству, сетевых информационных агентств. Используется преимущественно на экспериментальной основе для анализа настроений, оценки мобильности населения, показателей занятости, при постановке задач и мониторинге мер реагирования в случаях стихийных бедствий и кризисных ситуаций в здравоохранении.

Источники больших данных также могут использоваться для мониторинга и проведения отчетности по **показателям ЦУР**.³⁸ В настоящее время данные наблюдения Земли и геопространственные данные – наиболее хорошо изученный тип больших данных, который широко применяется для непосредственной оценки свыше десятка показателей ЦУР. Устройства Интернета вещей и мобильная

³⁸ ЭСКАТО представила подборку примеров стран, использовавших нетрадиционные источники данных для достижения показателей ЦУР. См. ЭСКАТО, [Использование больших данных для достижения показателей ЦУР](#), рабочие документы, выпуск №12 (январь 2021).

телефонная связь – еще один источник информации, который активно осваивается в целях изучения информационного общества, вспомогательных счетов и подготовки статистики общественного транспорта и умных городов.

Контроль качества

Как и в случае работы с традиционными источниками данных, необходимо создать рамочные основы для обеспечения качества больших данных прежде, чем приступать к использованию этой информации. В силу своей специфики нетрадиционные источники информации, такие как большие данные, требуют дополнительного изучения, поэтому НСУ понадобится проверка качества при производстве официальной статистики в соответствии с установленными принципами обеспечения качества. Такова общепринятая практика, установленная, к примеру, ЕЭК³⁹, Австралийским бюро статистики⁴⁰, Статистическим управлением Канады⁴¹, Статистическим управлением Нидерландов⁴².

Руководство Организации Объединенных Наций по национальным базовым принципам обеспечения качества официальной статистики (НБПОК ООН) Статистического отдела ООН закладывает основы для осуществления контроля качества информации источников данных, в том числе больших данных, используемых в производстве официальной статистики. Добавим, что осуществление мер НБПОК ООН подтверждает приверженность НСУ основополагающим принципам официальной статистики ООН. Положения руководства периодически пересматриваются. Текущая версия документа посвящена вопросам качества и включает возможные варианты обеспечения качества «других» источников данных, включающих большие данные.

НБПОК ООН - комплекс ключевых и дополнительных рекомендаций по осуществлению НСУ ОПОС на национальном уровне. С учетом особенностей, связанных с обеспечением качества, в Руководстве представлены 19 принципов и требований по их реализации, которые сгруппированы по принципу четырех уровней. Принимая во внимание задачи, поставленные в настоящем руководстве, мы рассматриваем только принципы обеспечения качества, которые имеют отношение к большим данным. Далее приводится неполный перечень основных принципов и элементов обеспечения качества больших данных.

³⁹ Европейская экономическая комиссия (ЕЭК), [Предложенные базовые принципы обеспечения качества больших данных](#), декабрь 2014.

⁴⁰ ABS, [The ABS data quality framework](#), 23 December 2015.

⁴¹ Statistics Canada, [Guidelines for ensuring data quality](#), 4 December 2019.

⁴² Barteld Braaksma and Kees Zeelenberg, [Big data in official statistics](#) (January 2020).

Таблица 1. Соответствующие принципы и элементы качества НБПОК ООН, имеющие отношение к большим данным

Принцип НБПОК ООН	Требования/элементы качества, которые необходимо учитывать *	Комментарий
Уровень А: Управление статистической системой		
Принцип 1: Координирование национальной статистической системы	Наличие отдела, который рассматривает возможность использования и содействует применению административных и других источников данных в рамках НСС (1.2, 10.1, 11.5)	Использование административных и других источников данных связано с особыми проблемами, решение которых возможно на основе обмена передовым опытом
Принцип 2: Управление взаимодействием с пользователями данных, поставщиками данных и другими заинтересованными сторонами	Наличие нормативно-правовых актов, обеспечивающих необходимый доступ к административным и другим источникам данных (2.5, 2.6)	Очень часто использование административных и других источников данных затруднено по причине ограниченного доступа
	Наличие соглашений о сотрудничестве с поставщиками данных (2.6)	Использование других источников данных может быть более эффективным, если налажено сотрудничество с поставщиком данных
Принцип 3: Управление статистическими стандартами	Национальный статистический орган следит за соблюдением статистических стандартов (3.2)	Использование стандартных понятий, определений и классификаций упрощает использование и интегрирование данных
Уровень В: Управление институциональной средой		
Принцип 5: Обеспечение непредвзятости и объективности	Источники данных и методология выбираются на основе объективных критериев (5.3)	Данные из других источников нужно использовать только в том случае, если они были отобраны и составлены в соответствии с профессиональными стандартами
Принцип 6: Обеспечение прозрачности	Условия, на которых используются другие источники данных, размещаются в открытом доступе (6.1)	Пользователям должна быть предоставлена возможность получить информацию о том, какие другие источники данных используются и как именно они используются
Принцип 7: Обеспечение конфиденциальности статистической информации и безопасности данных	Наличие специальных процедур для обеспечения безопасности связанных данных (7.2, 7.3, 7.5)	Часто использование данных из нескольких источников требует связывания записей
Уровень С: Управление статистическими процессами		
Принцип 10: Обеспечение эффективной методологии	Наличие методологии использования административных данных и данных из других источников (10.1, 10.3)	Данные из административных и других источников должны использоваться в соответствии с установленной методологией

	Наличие методологии связывания данных их других источников (10.1, 10.3)	Связывание данных из нескольких источников должно проводиться в соответствии с установленной методологией
Принцип 11: Обеспечение эффективности затрат	Данные интегрируются централизованно силами отдела интегрирования данных (11.6)	Интегрировать данные должны опытные и компетентные сотрудники
Принцип 12: Обеспечение надлежащих статистических процедур	Процедуры кодирования, исправления, внесения и проч. оформляются в письменном виде (12.2)	Шаги, связанные с использованием в статистических целях административных данных и данных из других источников, необходимо оформлять в письменном виде
	Процедуры связывания данных из разных источников разработаны, задокументированы и доступны (12.2)	Изменения в процедурах связывания данных из разных источников может поставить под угрозу единообразие результатов с течением времени
Уровень D: Управление статистическими продуктами		
Принцип14: Обеспечение применимости	В ответ на растущие информационные потребности общества статистика разрабатывается на основе новых источников данных (14.3)	При работе с любыми источниками данных очень важно учитывать новые потребности пользователей
Принцип 15: Обеспечение точности и надежности	Ошибки, не связанные с выборкой, сопровождаются пояснениями и соответствующими расчетами (если возможно) (15.1, 15.2)	Не связанные с выборкой ошибки могут серьезно повлиять на достоверность статистических продуктов, их сложно выразить количественно
Принцип 18: Обеспечение взаимосвязанности и сравнимости	Результаты сравниваются с другой имеющейся информацией и со сведениями за другие периоды времени (18.2)	При использовании административных данных и информации других источников необходимо обращать особое внимание на взаимосвязанность и сравнимость данных, так как первоначально они собираются не в статистических целях
Принцип 19: Управление метаданными	Предоставляются метаданные, связанные со всеми аспектами использования разных источников данных (19.2)	Статистика всегда должна сопровождаться метаданными, позволяющими пользователю оценить предлагаемую статистику

**Числовые значения в скобках указывают на определенные показатели/элементы качества в соответствии с НБПОК ООН. Более подробная информация представлена в [Приложении БПОК ООН](#).*

Важно отметить, что НБПОК ООН разработаны не для того, чтобы заменить существующие национальные принципы обеспечения качества, но как дополнительный справочно-информационный ресурс для контроля качества информации, не охваченной существующими правилами. По мере расширения

использования больших данных будут появляться новые версии руководств по обеспечению качества данных, включая НБПОК ООН, нацеленные на совершенствование механизмов управления качеством данных и изучение новых возможностей больших данных в производстве официальной статистики. Непрерывная работа НСУ по созданию собственных базовых принципов обеспечения качества больших данных способствует расширению масштабов их использования. Непрерывно ведется совместная работа НСУ и партнеров по развитию по ряду направлений: обмен полученным опытом и ускорение прогресса в области использования новых источников данных, восполнение пробелов в данных и производство официальной статистики с должным учетом принципов инклюзивности и устойчивости.

Обеспечение поддержки на глобальном и региональном уровнях

На глобальном и региональном уровнях проектам по большим данным оказывается широкая поддержка. На глобальном уровне **Комитет экспертов ООН по использованию больших данных и инструментария науки о данных для целей официальной статистики (КЭБА ООН)** изучает преимущества и недостатки больших данных, рассматривает возможности мониторинга и проведения отчетности по целям в области устойчивого развития (ЦУР). В КЭБА ООН представлено 31 государство-участник и 16 международных организаций, сотрудничающих в формате разных групп. КЭБА ООН включает восемь целевых групп: [Группу по данным АИС, большим данным и ЦУР, данным наблюдения за Землей, Группы по измерению доступа в сельской местности, по данным мобильной телефонной связи, методам защиты информации, использованию данных сканирующих устройств, по учебной подготовке, повышению квалификации и развитию потенциала](#) и учрежденная недавно совместная целевая группа, подотчетная КЭБА ООН и Сети специалистов по экономической статистике ООН, по глобальному содействию получению данных их источников в частном секторе.

Известно и о других тематических группах глобального уровня, таких как [Международная рабочая группа по статистике цен](#), или Оттавская группа, которая изучает возможности использования больших данных для оценки изменений стоимости, или [Группа по наблюдениям за Землей](#), работа которой направлена на увеличение объема, повышение доступности и расширение использования открытых данных наблюдения за Землей.

КЭБА ООН управляет [Глобальной платформой ООН](#). В ее формате создана облачная экосистема для содействия международному сотрудничеству в сфере развития официальной статистики на основе новых источников данных и передовых методик. Платформа – это среда для сотрудничества представителей статистического сообщества всего мира, позволяющая проводить обмен знаниями, информацией и методиками. Платформа предоставляет участникам сообщества возможности для поиска, использования и публикации информации о методах и алгоритмах в облачных ресурсах, в результате чего создается библиотека статистических методов и алгоритмов. Платформа позволяет проводить исследование данных, разработку аналитических решений и безопасных приложений. Глобальная платформа ООН

имеет **четыре физических региональных центра** в Китае, Руанде, Бразилии и Объединенных Арабских Эмиратах, которые осуществляют сотрудничество и разработку новых технологий для применения источников больших данных и методик.

На **региональном уровне** ЭСКАТО проводит активную работу, оказывая статистическому сообществу в Азиатско-Тихоокеанском регионе содействие по обмену знаниями в области больших данных в рамках сессий [Кафе по статистике](#), посвященных большим данным, встреч [Сообщества специалистов-практиков в области интеграции данных](#). ЭСКАТО разрабатывает информационные материалы, такие как [аналитические записки по статистике](#), [рабочие документы](#) и технические руководства. Проведено [Совещание группы экспертов по использованию больших данных в целях официальной статистики](#) и учебные мероприятия СИАТО.

Другие региональные комиссии ООН активно изучают вопрос использования больших данных для целей официальной статистики. Например, ЕЭК подготовила многочисленные издания и создала целевые группы, в числе которых [Руководство по обеспечению качества больших данных](#) и [Целевая группа по распоряжению данными](#).

Уроки использования данных мобильной телефонной связи в целях официальной статистики

В данном разделе предлагается краткая информация о данных сотовой связи/операторов мобильной связи (ОМС), параметрах и сферах применения этих сведений, подобраны конкретные примеры стран по использованию данных мобильной телефонной связи в официальной статистике и экспериментальных проектах.⁴³

11.1 Что такое данные мобильной телефонной связи?

Данные мобильной телефонной связи, используемые в официальной статистике, часто называют данными мобильного позиционирования (ДМП). Ценность этих сведений состоит в определении местоположения или элементе локализации. Данные мобильного позиционирования создают возможности для того, чтобы проанализировать и понять различное поведение людей на основе пространственно-временных оценок.

Выделяют два основных метода сбора данных мобильного позиционирования: **активное отслеживание и пассивное отслеживание**.⁴⁴

Для получения данных активного отслеживания направляется запрос на определение местоположения мобильного устройства, данные пассивного отслеживания, имеющие отношение к моменту прошлого, формируются без запроса. Известно о двух основных методах определения местоположения телефона с помощью технологий активного отслеживания: использование сетевой инфраструктуры ОМС или данных о местоположении, поступающих в режиме реального времени через приложения мобильных устройств, таких как карты и навигационные программы, социальные сети, приложения для обмена информацией

⁴³ Глобальная рабочая группа ООН по использованию больших данных в целях официальной статистики, [Руководство по использованию данных мобильной телефонной связи для целей официальной статистики](#) (сентябрь 2019).

⁴⁴ Margus Tiru, Overview of the Sources and Challenges of Mobile Positioning Data for Statistics, доклад, представленный в ходе Международной конференции по использованию больших данных в целях официальной статистики, Пекин, Китай, октябрь 2014.

о местоположении и геолокационные сервисы. Данные активного отслеживания поступают через глобальную систему навигации и определения положения GPS или A-GPS, беспроводную сеть (WiFi) или из баз данных о местоположении антенн мобильной связи. Определение местоположения через приложения, как правило, проводится с согласия пользователя.

Данные пассивного мобильного отслеживания обычно получают из реестров и баз данных сетевой инфраструктуры ОМС. Пример – детализация звонков (ДЗ) и сведения установленных на мобильном устройстве приложений. В основном информация о местоположении собирается и хранится ОМС для внутреннего пользования, в частности, для начислений оплаты или маркетинговых целей. Многие установленные на телефоне приложения собирают данные отслеживания, главным образом, для обеспечения работы самого приложения. При этом ОМС не имеет отношения к этим сведениям, они хранятся либо на устройстве, либо на сервере поставщика приложения.

11.2 Аспекты обеспечения доступа к ДМП

От ряда аспектов обеспечения доступа к ДМП зависят качественные показатели и эффективность проекта.⁴⁵ К их числу относятся:

Главные аспекты	Уровни доступа к данным для подготовки официальной статистики. Как правило, этот аспект связан с первичными/псевдонимизированными данными или агрегированным результатами.
	Модели доступа к данным. Существуют три основные модели доступа: обязательная, добровольная и коммерческая.
Дополнительные аспекты	Основные правительственные организации. Например, независимые профессиональные учреждения, такие как НСУ.
	Официальный/неофициальный продукт. Определяет соответствие разработанного показателя (показателей) международным стандартам.
	Сфера(-ы) использования ДМП. Определяет

⁴⁵ Erki Saluveer, *Mobile positioning data and data access*, презентация в ходе Совещания экспертов по модернизации статистического законодательства ЕЭК ООН/ЕАСТ, Позитиум, Эстония, ноябрь 2021.

	необходимость использования ДМП для формирования выбранных показателей.
	Создатель методики и алгоритма. Определяет соответствие требованиям качества. Желательно участие НСУ или другого производителя официальной статистики.
	Число участвующих ОМС. Характеризует долю в общем объеме рынка и полноту данных.
	Место обработки данных. Зависит от национального законодательства и соглашений между заинтересованными сторонами. Обработка может проводиться ОМС, регулятором, НСУ или другим государственным органом.
	Участие третьих сторон. Может включать заинтересованные стороны из числа правительственных и неправительственных организаций, таких как научные учреждения и частные компании-партнеры.

При реализации проектов в странах применялись следующие комбинации уровней доступа и моделей по использованию ДМП. Наибольшее распространение получили схемы первичные данные – обязательная модель доступа и агрегированные данные – коммерческая модель, заметно меньшее – схема агрегированные данные – доступ на добровольной основе.

		Уровень доступа	
		Первичные данные	Агрегированные данные
Модель доступа	Обязательная		
	Коммерческая		
	Добровольная		

11.3 Дорожная карта проекта по ДМП

В самом начале проекта по ДМП необходимо сформировать ясное представление о модели внедрения. Следует обеспечить сопровождение модели внедрения надлежащей методологией. Впрочем, использование единых методологических

концепций в каждой из стран невозможно, к тому же, они претерпевают изменения с течением времени даже в отдельно взятой стране.⁴⁶

С учетом широкого опыта НСУ по использованию ДМП в экспериментальных исследованиях или официальной статистике разработана следующая Дорожная карта реализации проекта по ДМП.⁴⁷

❖ **Анализ текущей ситуации и предпроектная деятельность**

- Определение и конкретизация целей проекта
- Определение показателей эффективности
- Оценка национальной нормативно-правовой базы
- Выбор бизнес-модели
- Выбор подходящих партнеров

❖ **Доступ к данным**

- Подготовка к проведению экспериментального проекта
- Проведение консультаций с ОМС и/или регуляторами
- Выбор модели партнерства по использованию данных и подписание необходимых соглашений и контрактов
- Разработка и настройка параметров среды внешнего размещения
- Предварительная проверка данных (обеспечение качества)

❖ **Экспериментальный проект**

- Проведение оценки на всех этапах статистического производства: сбор, обработка, анализ и передача данных при обеспечении надлежащего качества

❖ **Реализация проекта**

- Широкое использование ДМП в статистическом производстве

11.4 Конфиденциальность данных и вопрос общественного доверия

Данные о местоположении являются персональной информацией и могут раскрыть важные сведения личного характера частных пользователей. Доступ к персональным данным и их обработка подвергаются тщательной регламентации.

⁴⁶ Jaanus Kroon, *Mobile Phone Data and Credit Card Data: Experience for 10 years of Public Private Partnership*, документ, представленный в ходе 62^{го} Всемирного конгресса по статистике МСИ, Куала-Лумпур, август 2019.

⁴⁷ Подготовлено на основе презентации Erki Saluveer, *Mobile positioning data and data access* в ходе Совещания экспертов по модернизации статистического законодательства ЕЭК ООН/ЕАСТ, Позитум, Эстония, ноябрь 2021.

Операторы мобильной связи руководствуются строгими корпоративными правилами в отношении защиты личной информации, к тому же ОМС не заинтересованы в распространении этих сведений. Модели сотрудничества и законодательство меняют динамику обмена данными, позволяя достигнуть баланса между необходимостью обеспечения конфиденциальности и получения общественных благ на основе анализа данных.

Известно о существовании многих инструментов и подходов, защищающих конфиденциальную личную информацию. Более десяти лет эстонская компания «Позитиум» оказывает поддержку НСУ и ОМС по вопросам использования ДМП в целях официальной статистики при условии сохранения конфиденциальности и в соответствии с национальным законодательством стран, в интересах которых проводится работа. Кроме того, компания занимается разработкой вспомогательных методик и рамочных основ. Позитиум в сотрудничестве со Статистическим отделом ООН (СОООН) и Международным союзом электросвязи (МСЭ) разработала пять руководящих принципов, обеспечивающих доверие общественности при использовании данных мобильных операторов в целях разработки политики,⁴⁸ а именно:

1. необходимость и соразмерность;
2. профессиональная независимость;
3. защита конфиденциальности;
4. приверженность качеству;
5. сопоставимость на международном уровне.

Евростат сотрудничал с другой эстонской компанией «Кибернетика» для создания обосновывающего концепцию решения по безопасной обработке данных ОМС частной компанией в целях производства официальной статистики. В ходе проекта приоритетом стала выработка решений по обеспечению конфиденциальности входных данных и создание безопасной среды выполнения при изолировании технических средств.⁴⁹ На диаграмме 3 представлена новая предложенная схема организации рабочего процесса.

Шведская некоммерческая организация «Флоумайндер», специализирующаяся на анализе данных о мобильности, спутниковых изображениях и административной информации в гуманитарных целях и в интересах международного развития,

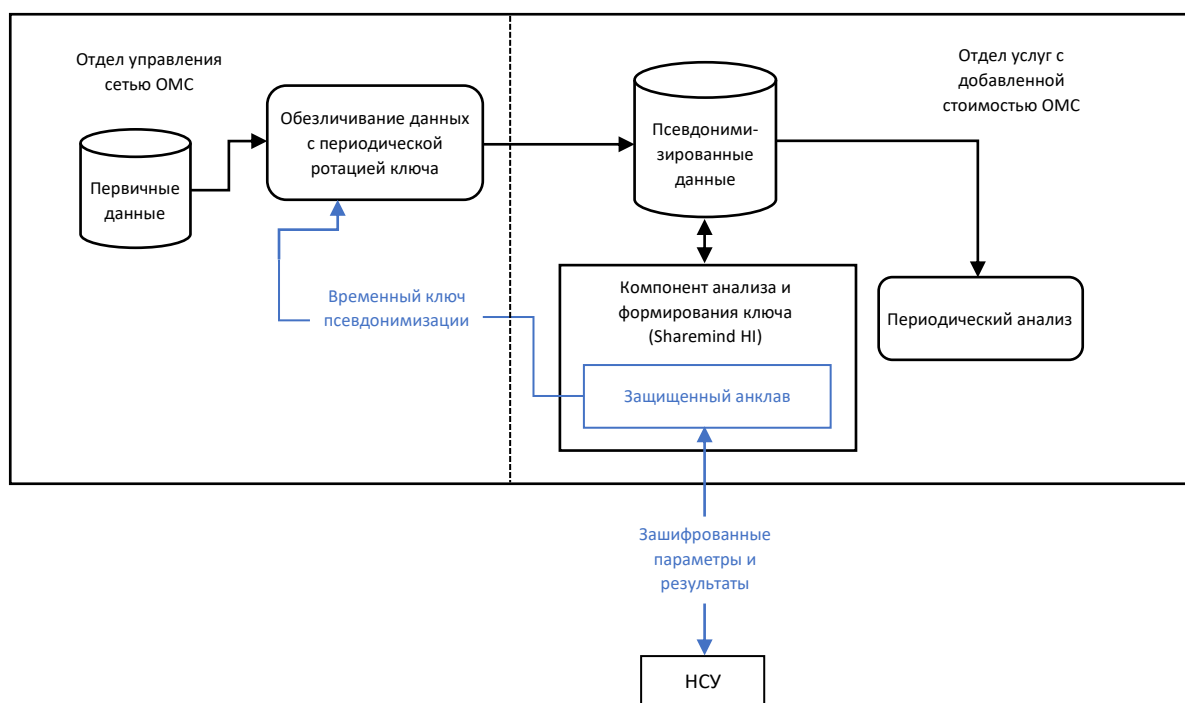
⁴⁸ Ronald Jansen and others, *Guiding principles to maintain public trust in the use of mobile operator data for policy purposes*, *Data and Policy*, 3, e24 (October 2021).

⁴⁹ Fabio Ricciato, *A proof-of-concept solution for secure processing of mobile network operator data for official statistics*, документ, представленный на Совещании экспертов по обеспечению конфиденциальности статистической информации, Польша, декабрь 2021.

разработала концепцию ОПАЛ⁵⁰ (открытый алгоритм). Компания сосредоточила усилия на партнерских связях с операторами мобильной связи.⁵¹ На диаграмме 4 представлена блок-схема для решения Флоукит фонда «Флоуимайндер» по обработке первичных данных детализации вызовов в пределах защитной системы ОМС. Результаты формируются на платформе ОПАЛ, при этом соблюдаются существующие в стране этические нормы и требования законодательства.

Оба решения основаны на анализе данных, который проводится в пределах защитной системы ОМС, и предполагают передачу агрегированных результатов НСУ.

Диаграмма 3. Рабочий процесс внедрения безопасных технологических решений частных компаний для использования ДМП в целях официальной статистики⁵²



⁵⁰ Разработка ОПАЛ проводилась с 2017 года MIT Media Lab в сотрудничестве с Имперским колледжем Лондона, Orange Group, Telefonica Group, Всемирным экономическим форумом и организацией Data-Pop Alliance при поддержке Французского агентства развития (ФАР) и Фонда Уильяма и Флоры Хьюлетт.

⁵¹ Более подробная информация может быть получена по ссылке: [Программа ОПАЛ фонда «Флоуимайндер» для данных мобильных операторов](#)

⁵² Fabio Ricciato, A proof-of-concept solution for secure processing of mobile network operator data for official statistics, документ, представленный в ходе Совещания экспертов по обеспечению конфиденциальности статистической информации, Польша, декабрь 2021.

Диаграмма 4. Использование экосистемы ОПАЛ фонда «Флоумайндер» для обработки данных мобильного оператора⁵³



11.5 Производство статистики с использованием ДМП

Следующие виды официальной и экспериментальной статистики были получены НСУ и другими производителями официальной статистики, такими как Центральный банк Эстонии, на основе ДМП:

- статистика туризма и мероприятий;
- демографическая статистика;
- миграционная статистика;
- статистика внутригородской ежедневной миграции;
- статистические данные транспортных потоков;
- статистика занятости по приграничным работникам и сезонной рабочей силе.

ДМП - официальный вид статистической информации в Эстонии и Индонезии. Другие страны успешно провели эксперименты по использованию ДМП в официальной статистике и рассматривают возможности интеграции этого источника данных в официальную статистику.

Главные преимущества ДМП – эффективность с точки зрения затрат по сравнению регулярными обследованиями и интервью, своевременность, полнота и охват данных. Данные могут быть использованы для разработки новых и не

⁵³ Более подробная информация может быть получена по ссылке: [Программа ОПАЛ фонда «Флоумайндер» для данных мобильных операторов](#)

использовавшихся ранее показателей в таких областях, как туризм, транспорт, демографическая статистика и т.д.

11.6 Примеры стран

Практика стран по использованию ДМП отличается от случая к случаю и в определенном смысле зависит от различий в установленных моделях партнерства. Отчасти отличия также определяются национальной нормативно-правовой базой. На примере нескольких стран региона мы рассмотрим разный практический опыт и существующие подходы.

Далее приводятся примеры четырех стран: Эстонии, Омана, Индонезии и Российской Федерации. Примеры позволяют рассмотреть опыт стран по пяти аспектам:

1. обоснование проекта;
2. основные заинтересованные стороны, модель партнерства по использованию данных и поток данных;
3. нормативно-правовые основы проекта;
4. статистические показатели и обеспечение качества;
5. проблемы и перспективы.

Эстония

Обоснование проекта

Эстония – страна Европейского союза, население которой составляет 1,3 миллиона человек. Департамент статистики и Центральный банк Эстонии (Eesti Pank) являются производителями официальной статистики. Банк Эстонии осуществляет подготовку статистики денежно-кредитного и финансового сектора, статистики внешнего сектора, статистики финансового счета и т.д.

Два важных события предшествовали принятию решения об использовании данных мобильной телефонной связи для расчетов трансграничной статистики. Во-первых, Эстония присоединилась к Шенгенской зоне в конце 2007 года, в результате был отменен пограничный контроль, регистрация и сбор сведений о пересечении границ в семидесяти пяти и более процентов случаев. Во-вторых, Департамент статистики Эстонии столкнулся с существенным сокращением бюджета и передал полномочия по подготовке трансграничной статистики, которая является одной из статей платежного баланса, Банку Эстонии.

Стремительный рост числа путешествий по всему миру содействовал обогащению традиционных источников данных за счет сведений трансграничной статистики. Банк Эстонии вынужден был обратиться к новаторским источникам данных и методам для расчета платежного баланса. Так было принято решение об использовании данных мобильной телефонной связи.

Основные заинтересованные стороны, модели партнерства и поток данных

В 2008 году Банк Эстонии начал сотрудничество с исследователями Тартуского университета и ООО «Позитуим» по разработке новой методологии сбора данных для получения трансграничной статистики на основе данных мобильной телефонной связи. Работа над проектом объединила экономистов-статистиков, специалистов по социально-экономической географии и аналитиков данных.

Обработка данных проводилась внешним исполнителем ООО «Позитуим» на основе договора о государственных закупках. Договор сроком на три года был заключен в 2009, 2012, 2015 и 2018 годах.

Методы исследования входящего туристического потока были согласованы в 2008-2009 годах, а исходящего – в 2009-2010 годах. Завершенные методики были пересмотрены и дополнены в 2015 году.⁵⁴

Центральный банк использовал Типовую модель производства статистической информации (ТМСИ), в рамках которой выстраивался рабочий процесс проекта. Описание функциональных обязанностей Позитиум и Центрального банка представлено в таблице 2. Обработывались обезличенные данные.

Таблица 2. Распределение функциональных обязанностей между Позитиум и Банком Эстонии⁵⁵

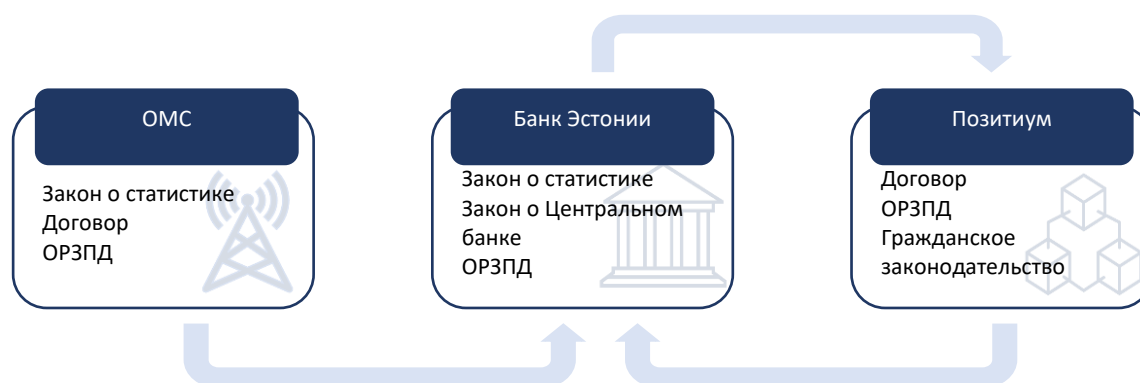
Задачи ТМСИ / партнер	Банк Эстонии	Позитиум
Постановка задач		
<i>Разработка методологии</i>		
Разработка		
Построение		
Сбор данных		
Обработка данных		
Анализ		
Передача		
Оценка		
<i>Калибровка</i>		
<i>Редактирование – разработка и построение</i>		
<i>Комментарии по улучшению – разработка и построение</i>		

Диаграмма 5 иллюстрирует передачу данных в контексте законодательства, регулирующего деятельность заинтересованных сторон.

⁵⁴ Jaanus Kroon, *Mobile Phone Data and Credit Card Data: Experience for 10 years of Public Private Partnership*, доклад в ходе 62^{го} Всемирного конгресса по статистике МСИ, Куала-Лумпур, август 2019.

⁵⁵ Jaanus Kroon, *Mobile Phone Data and Credit Card Data: Experience for 10 years of Public Private Partnership*, доклад в ходе 62^{го} Всемирного конгресса по статистике МСИ, Куала-Лумпур, август 2019.

Диаграмма 5. Поток данных мобильного отслеживания между заинтересованными сторонами Эстонии



Нормативно-правовая база проекта ⁵⁶

Управление данными в рамках проекта регулируется Законом о статистике и Законом о Центральном банке. В соответствии с Законом о статистике компании предоставляют данные для целей официальной статистики на безвозмездной основе. На основании этого положения Центральный банк ежемесячно получает данные мобильной телефонной связи. Обработка и защита личных данных проводится согласно положениям ОРЗПД.

В своей работе Банк Эстонии руководствуется Кодексом практики в области официальной статистики, Кодексом практики европейской статистики, кодексами практики в области государственной статистики МВФ и ЕЦБ.

Банк Эстонии заключил контрактные соглашения с ОМС и Позитиум.

В соответствии с действующим законодательством допускается использование метода пассивного отслеживания. Более подробные сведения, связанные с местоположением, могут быть получены на основе доступа к данным активного отслеживания, однако сбор информации о местоположении запрещается европейским законодательством.

Статистические показатели и контроль качества

Банк Эстонии использует ДМП для подготовки трансграничной статистики, сведения которой формируются в ходе предоставления туристических услуг. Туристические услуги – составляющая счета текущих операций платежного баланса. Диаграмма 6

⁵⁶ Jaanus Kroon, Mobile positioning data for official statistics: Eesti Pank experience, презентация в ходе 32^я встречи в режиме онлайн Сообщества практиков по интеграции данных, сентябрь 2021.

иллюстрирует то, что трансграничная статистика является компонентом туристических услуг.

Диаграмма 6. Трансграничная статистика и составляющие туристических услуг⁵⁷



При подготовке трансграничной статистики рассматриваются следующие категории:

- путешественники – посетители на один день;
- туристы – посетители, которые останавливаются на ночь;
- транзитные путешественники – посетители на один день, которые останавливаются менее чем на 4 часа, не считая пребывание в аэропортах (портах);
- рабочие мигранты и студенты – лица, пребывающие длительно сроком более 91 дня в течение последних 12 месяцев;
- трансграничные маятниковые мигранты – лица, пребывающие менее одного дня, ведется учет числа и продолжительности поездок с остановкой на ночь.

Для валидации используются следующие альтернативные источники данных:

- Выездное пересечение границы
 - Статистические данные платежных карт (Банк Эстонии)
 - Статистические данные аэропортов (портов)
 - Обследования домохозяйств (Департамент статистики Эстонии)
- Въездное пересечение границы
 - Пересечение границ Эстонии за исключением Шенгенской зоны (пункты полиции и пограничного контроля)
 - Статистические данные платежных карт (Банк Эстонии)
 - Статистические данные аэропортов (портов)
 - Обследования по средствам размещения (Департамент статистики Эстонии)

⁵⁷ Подготовлено на основе презентации Jaanus Kroon, Mobile positioning data for official statistics: Eesti Pank experience в ходе 32^й встречи в режиме онлайн Сообщества практиков по интеграции данных, сентябрь 2021.

В ходе валидации Банк Эстонии визуализирует данные, проводит анализ и сопоставление с проверочными данными, находит отклоняющиеся значения и корректирует их. Наиболее релевантные временные ряды для проверки корректности информации получают на основе статистических данных платежных карт.

Данные используются для формирования ежемесячной и квартальной статистики внешнего сектора и размещаются отдельно в секции международной статистики туризма на официальной странице веб-сайта Банка Эстонии ежеквартально с 2012 года. Соответствующие задачи, аналогичные мерам Банка Эстонии по использованию статистики, также сформулированы в эстонской Национальной программе в области статистики.

Банк Эстонии разработал *Методологические основы производства национальной статистики туризма на основе данных мобильного позиционирования*.⁵⁸

Преимущества и недостатки

Центральный банк Эстонии отмечает, что такие характеристики, как своевременность, сопоставимость, полнота охвата и эффективность с точки зрения затрат, являются главными преимуществами использования данных мобильной телефонной связи при подготовке трансграничной статистики. Данные мобильной телефонной связи дают возможность сбора информации в более детальной разбивке по регионам или странам проживания, обеспечивают лучший охват сведений о посетителях на один день и поездках с ночевкой в сравнении со статистикой по средствам размещения. Это достигается благодаря покрытию данных о поездках лиц, которые останавливаются в незарегистрированных средствах размещения. ДМП могут быть беспрепятственно получены из баз данных мобильных телефонных операторов, что отменяет необходимость проведения регулярных обследований, неэффективных с экономической точки зрения. Использование этих данных не повышает административной нагрузки, более того, обработка ДМП проводится автоматически.

Банк Эстонии выявил существование ряда проблем, среди которых необходимость пересмотра моделей в соответствии с изменениями экономической среды и сферы использования мобильных телефонов. Следует учитывать вопросы излишнего и недостаточного покрытия, процессы на рынке телекоммуникаций и роуминга, изменения в телефонной связи и появление новых источников шума. Особое беспокойство вызывает проблема конфиденциальности данных, прежде всего,

⁵⁸ См. <https://sciencedirect.com/science/article/pii/S0160738320300396?via%3Dihub>.

вопрос обеспечения доступа к большим данным на основании ОРЗПД и закона о связи, которые в настоящее время запрещают использование данных активного отслеживания и сигнализации. С помощью этих данных может быть получена более подробная информация о местоположении.

Обоснование проекта

Население Омана составляет 5,1 миллиона человек, при этом число действующих абонентских договоров на предоставление услуг мобильной сотовой связи приблизилось к 6,2 миллиона (показатель проникновения равен 123%).⁶⁰ Приминая во внимание столь высокий уровень проникновения сотовой связи и наличие благоприятной нормативно-правовой базы, Национальный центр статистики и информации (НЦСИ) рассмотрел перспективы использования ДМП и оценил возможности полномасштабного внедрения статистики, основанной на ДМП. Возникла необходимость разработки концепции, обосновывающей использование ДМП. Эти меры соответствовали решению более широкой задачи повышения информированности общества благодаря использованию новых источников данных в статистике, при проведении мониторинга, анализа и принятии решений.

Основные заинтересованные стороны проекта, модели партнерства и поток данных⁶¹

Проект Омана по использованию данных мобильной телефонной связи объединил несколько партнерских организаций. Национальный центр статистики и информации (НЦСИ) осуществляет общее руководство в рамках работы над проектом. Управление по телекоммуникациям (УТ) согласует порядок предоставления данных ОМС. Агентство информационных технологий (АИТ) технически управляет правительственными облачными ресурсами (правительственным облаком). Научно-исследовательский совет Омана (НИС) и Университет имени султана Кабуса (УСК) обеспечивают научную поддержку и консультации ученого сообщества в рамках проекта. Все перечисленные партнерские организации входят в состав национальной рабочей группы.

Среди других партнеров два оператора мобильной связи (ОМС): Омантел и Уреду. На каждую из этих двух компаний, предоставляющих телекоммуникационные услуги, приходится пятидесятипроцентная доля рынка. Исполнителями проекта являются Позитиум и Нортал.

⁵⁹ Ahmed Al Mufarji, National Centre for Statistics and Information (NCSI), *Oman's experience in Utilizing Mobile Positioning Data for Official Statistics*, презентация в режиме онлайн в ходе Вебинара по вопросам глобальной сети, апрель 2021.

⁶⁰ Последние данные Всемирного банка (2020) по численности населения и числу абонентов сотовой связи. Показатель проникновения – соотношение числа абонентов сотовой связи и общей численности населения.

⁶¹ Ahmed Al Mufarji, National Centre for Statistics and Information (NCSI), *Mobile Positioning Data Pilot Project for Official Statistics in Oman*, презентация в ходе Международной конференции по использованию больших данных в целях официальной статистики, Дублин, Ирландия, август 2016.

На начальном этапе проекта применялась модель, в рамках которой компания сохраняла первичные данные, а алгоритмы использовались для их обработки и получения агрегированных результатов. ОМС вынуждены были предоставить существенные ресурсы и инфраструктуру. В дальнейшем было принято решение о переносе первичных данных за пределы оператора связи и внесении их в централизованную базу. Благодаря этим мерам, НСУ получили больший контроль над данными, могли проводить проверки качества, выполнять методологические сценарии, дезагрегировать данные по выбранным показателям и т.д. Первичные данные ОМС поступали в централизованную базу НЦСИ. Вся вспомогательная информационно-технологическая инфраструктура была создана за пределами ОМС.

Нормативно-правовая база проекта

Действующее законодательство обеспечило успех проекта в Омане. Важную роль сыграли следующие правовые документы:

- ❖ **Королевский указ 2014 года, регулирующий деятельность НЦСИ.**
Предоставляет НЦСИ широкие полномочия в сфере сбора и хранения официальной статистики, которая производится Центром и другими правительственными и неправительственными организациями.
- ❖ **Закон о статистке и информации, принятый Королевским указом 2019 года.**
Статья 19 предусматривает право НЦСИ *«запрашивать любые необходимые данные, информацию и административные документы у частных лиц, правительственных и неправительственных организаций на безвозмездной основе»* для достижения своих задач. При этом статья 32 предусматривает меры в отношении любой неправительственной организации, которая отказывается предоставить данные и информацию, запрашиваемую в соответствии с положениями настоящего закона и необходимую лицам, осуществляющим деятельность в сфере официальной статистики.

НЦСИ подписал Меморандум о взаимопонимании (MoV) с Омантел и Уреду.

Согласно законодательным актам о связи и соответствующим внесенным поправками данные не могут передаваться за пределы страны. Поскольку проводится сбор данных двух операторов, УТ обязуется *«создать условия,*

обеспечивающие безопасность персональных данных бенефициаров, и гарантировать конфиденциальность и неприкосновенность личной информации». ⁶²

Статистические показатели и контроль качества

НЦСИ получает от ОМС обезличенные первичные данные, такие как подробный отчет о вызовах (ДВ) или детализации данных (ДД) и информацию о номере сотовой связи. Любые сведения, указывающие на владельца, удаляются и маркируются перед изъятием с сервера ОМС. Демографическая информация об абонентах, полученная от ОМС, не представляет ценности, пока не будет откорректирована и дополнена ОМС. В Омане несколько членов семьи регистрируют номер на имя главы семьи, которым является мужчина. Таким образом, в восьмидесяти процентах случаев абонентами являются мужчины, что создает неверное представление о существующей динамике пользователей в разбивке по полу.

В результате реализации проекта НЦСИ подготовил статистическую информацию за период с января по июнь 2020 года на национальном уровне, на уровне провинций и вилайатов в пяти областях. Направления включали демографическую статистику с расчетом якорных точек, статистику туризма, въездного и внутреннего, формирование данных о временном населении и мобильности, в том числе о наличном населении и маятниковой миграции. Более подробная информация о статистике на основе ДМП представлена далее.

- ❖ **Демографическая статистика** с расчетом якорных точек по числу жилищ, числу мест работы, числу вторых жилищ.
- ❖ **Статистика туризма**, включающая показатели въездного и выездного туризма:
 - агрегирование по времени (день, неделя, месяц);
 - пространственное агрегирование (страна, провинция, область, сеть);
 - по стране происхождения/детализация вызовов;
 - по типу пребывания (с остановкой на ночь, отправление в тот же день);
 - по продолжительности пребывания;
 - по месту назначения (главное, второстепенное, транзитное);
 - число посещений;
 - число уникальных посетителей;
 - число прибытий/проживающих гостей/убытий;

⁶² Ahmed Al Mufarji, National Centre for Statistics and Information (NCSI), Mobile Positioning Data Pilot Project for Official Statistics in Oman, презентация в ходе Международной конференции по использованию больших данных в целях официальной статистики, Дублин, Ирландия, август 2016.

- количество проведенных ночей;
- средняя продолжительность проживания (в часах).

❖ **Временное население и мобильность** – информация о наличном населении и маятниковой миграции. Маятниковая миграция – новый показатель, расчет которого ранее не проводился. Показатель включает следующее:

- матрицы «точка отправления из дома или места работы в точку назначения»;
- категории посетителей (местные жители, работающие лица, постоянные посетители, иностранные/отечественные туристы, путешествующие транзитом);
- агрегирование по времени (дневное время, ночное время, сутки).

В ходе проекта использовались следующие справочные данные:

- | | |
|---|---|
| • информация об административно-территориальном делении | • демография населения |
| • сведения государственной компьютерной сети Омана | • данные реестра народонаселения |
| • данные дорожной сети | • статистика внутреннего туризма |
| • данные о земном покрове | • статистика въездного туризма |
| • данные строительного сектора | • статистика выездного туризма |
| • данные переписи населения | • данные дорожных петлевых детекторов |
| • статистика маятниковой миграции по результатам переписи | • сведения о маятниковой миграции и миграционные данные |
| • база данных точек интереса | • данные мобильной телефонной связи – доля рынка |

Позитиум и НЦСИ руководствовались Базовыми принципами обеспечения качества при работе с большими данными ЕЭК и Базовыми принципами обеспечения качества Евростата, адаптировав положения этих документов в контексте использования ДМП и условий стран.^{63, 64} Компании самостоятельно проводили оценку обработанных данных. На основе комплексного подхода был проведен анализ качества данных. Качество данных оценивалось с учетом четырех основных показателей: характеристики абонента, время, место и страна происхождения;

⁶³ ЕЭК, [Предложенные базовые принципы обеспечения качества больших данных](#), декабрь 2014.

⁶⁴ Европейская статистическая система, [Базовые принципы обеспечения качества Европейской статистической системы, версия 2.0](#) (2019)

дополнительно были выделены еще три характеристики: источник, метаданные и данные. Отметим, что перед использованием качество ДМП может оцениваться на основании 29 параметров.

Оценка качества и проведение калибровочных испытаний неотделимы друг от друга. Важным источником калибровочных данных является электронная перепись. Тем не менее, использование этих данных в статистике туризма связано с определенными сложностями, поскольку отсутствуют эталонные данные по туризму.

Преимущества и недостатки

НЦСИ удалось добиться заметных результатов от вложения инвестиций по ряду показателей. Центр обеспечил более быструю обработку и производство важных статистических показателей с временным и пространственным охватом, смог повысить точность данных без дополнительной нагрузки на респондентов. ДМП применяются в разных областях, в связи с чем НЦСИ провел разработку новых показателей и разбивок информации по категориям, не использовавшимся ранее. Кроме того, ДМП подтвердили свою обоснованность с экономической точки зрения в сравнении с принятыми методами того же порядка. Приведем пример. В Омане не производится статистика транспортного сектора о происхождении-назначении. НЦСИ тратит на проведение обследования в одном регионе 260000 долларов США, в стране 11 регионов. Использование ДМП позволило бы получить результаты быстрее с большим охватом и заметно меньшими издержками.

Индонезия

Обоснование проекта

При общей численности населения, составляющей 268 миллионов человек, в Индонезии зарегистрировано 355 миллионов действующих телефонных номеров. 91% взрослого населения пользуется мобильным телефоном, при этом все номера сотовых телефонов проходят обязательную регистрацию с внесением единого идентификационного номера, указанного на удостоверяющем личность документе.⁶⁵ Возможности использования данных мобильной телефонной связи в официальной статистике в Индонезии очевидны.

Принимая во внимание передовые международные практики и выводы Технико-экономического обоснования использования данных мобильного отслеживания в статистике туризма, разработанного Евростатом, учитывая приведенные выше цифры,⁶⁶ Центральное статистическое бюро Индонезии приступило к исследованию данных мобильной телефонной связи в целях повышения качества официальной статистики. Приоритетные вопросы включали точность и своевременность данных, охват наблюдением и эффективность работы региональных статистических управлений.

Основные заинтересованные стороны проекта, модели партнерств и поток данных⁶⁷

В реализации проекта по ДМП в Индонезии участвует ряд заинтересованных сторон, в числе которых Центральное статистическое бюро Индонезии и Министерство туризма, представляющие органы государственной власти, компания «Телкомсел», выступающая в роли поставщика данных, и компания «Позитиум» - в роли партнера по обработке данных и консультанта.

Центральное статистическое бюро Индонезии впервые обратилось в Телкомсел, крупнейший оператор мобильной связи в Индонезии, число клиентов которого превышает 170 миллионов человек, в 2016 году. ЦСБ Индонезии и Телкомсел подписали MoU и коммерческое соглашение.

До начала сотрудничества с ЦСБ Телкомсел не занималась хранением данных сигнализации, поскольку эта информация не представляла для компании интереса с коммерческой точки зрения и требовала больших мощностей для хранения. Однако

⁶⁵ Edi Setiawan, [Big data for social and demographic statistics](#), презентация в ходе сессий Кафе по статистике ЭСКАТО в режиме онлайн, 2 ноября 2020.

⁶⁶ Европейская комиссия, [Технико-экономическое обоснование использования данных мобильного отслеживания в статистике туризма](#) (Люксембург, июнь 2014).

⁶⁷ Ирина Берналь и Таня Сейерсен, [Большие данные для экономической статистики](#), аналитическая записка по статистике, выпуск № 28 (март 2021).

перед Телкомсел открылись коммерческие возможности, и компания сделала финансовые вложения в инфраструктуру хранения и анализа данных. Сотрудничество с органами статистики и применение методологических концепций позволило Телкомсел развить новый сегмент рынка, связанный с информацией о мобильности. Компания разработала для коммерческих организаций и государственных органов специализированные решения, такие как анализ моделей мобильности.⁶⁸

В 2017 году Министерство туризма и ЦСБ начали сотрудничество с Позитиум. Позитиум было поручено провести анализ данных и процессов ОМС, обеспечив прозрачность. Позитиум создала автоматизированную систему обработки ДМП в статистике туризма, физически расположенную на территории Телкомсел.

Обработка данных проводится по методу «песочницы». ДМП хранятся и обрабатываются системами Телкомсел без передачи какой-либо информации ЦСБ. При этом ЦСБ проводит исследования и разработки в «песочнице» Телкомсел, где имеет доступ к пятипроцентной выборке данных абонентов со скрытой личной информацией. Данные обрабатываются ЦСБ только в рамках условий контракта. ЦСБ исследует данные и пишет сценарии на основании полученных результатов. Телкомсел адаптирует сценарии в своей среде и выполняет их для всех пользовательских данных. Результат - агрегированная информация, личная информация может быть получена только по согласию абонентов. ЦСБ получает отчеты в виде матрицы источник-назначение в соответствии с установленным форматом.

Телкомсел и ЦСБ разрабатывают методологию совместно. Постепенно при участии Позитиум совершенствуются методики и алгоритмы, принимаются дополнительные меры по обеспечению качества.⁶⁹

С 2019 года данные **трансграничной статистики** основываются на результатах новой системы, в рамках которой туристические определения видов поездок применяются к первичным данным. ЦСБ Индонезии использует Типовую модель производства статистической информации (ТМСИ) и распределяет функциональные обязанности по подготовке трансграничной статистики туризма между заинтересованными сторонами следующим образом:

⁶⁸ Mia Melinda, *Unravel Potential Partnership using Telco Big Data*, презентация в ходе сессий Кафе по статистике ЭСКАТО в режиме онлайн, 30 августа 2021.

⁶⁹ ЭСКАТО, *Большие данные для демографической и социальной статистики*, аналитическая записка по статистике, выпуск № 29 (апрель 2021).

Таблица 3. Распределение функциональных обязанностей в рамках проекта по трансграничной статистике туризма⁷⁰

Задачи ТМСИ / партнер	ЦСБ Индонезии	Позитиум	Телкомсел
Постановка задач			
Разработка			
Построение			
Сбор данных			
Обработка данных			
Анализ			
Передача данных			
Оценка			

ЦСБ при участии Телкомсел запустило еще один проект в сфере **статистики маятниковой миграции и внутреннего туризма**. В рамках этого проекта ЦСБ занималось подготовкой концепций обработки данных, на Телкомсел были возложены обязанности по их реализации. Позитиум обеспечила общее руководство в ходе работы ЦСБ и Телкомсел. С течением времени сотрудничество между участниками проекта расширялось с распределением обязанностей следующим образом:

Таблица 4. Распределение функциональных обязанностей в рамках проекта по статистике внутреннего туризма⁷¹

Задачи ТМСИ / партнер	ЦСБ Индонезия	Позитиум	Телкомсел
Постановка задач			
Разработка			
Построение			
Сбор данных			
Обработка данных			
Анализ			
Передача данных			
Оценка			

ЦСБ не имеет опыта работы с пространственно-временными моделями ДМП и доступа к первичным данным, зависит от продолжения сотрудничества с компаниями частного сектора по методическому обеспечению и обработке данных.

⁷⁰ Ирина Берналь и Таня Сейерсен, [Большие данные для экономической статистики](#), аналитическая записка по статистике, выпуск № 28 (март 2021).

⁷¹ Там же.

ЦСБ Индонезии заключило с компанией «Позитиум» в 2019 году долгосрочный Меморандум о взаимопонимании, направленный на содействие сотрудничеству в исследовательской и проектной деятельности, а также укрепление потенциала.

Нормативно-правовая база проекта⁷²

Ряд законов и нормативных актов обеспечивают правовые основы проектов по ДМП, в частности, Закон о статистике №16 1997 года и Постановление Министерства связи и информационных технологий №20 2016 года «О защите персональных данных в электронной системе». В сфере обеспечения конфиденциальности ЦСБ Индонезии руководствуется Законом о защите персональных данных, ОМС следует внутренним правилам защиты личной информации и руководящим принципам корпорации.

Доступ к ДМП обеспечивается на основании МоВ и соглашения о неразглашении информации (СНИ) между головной телекоммуникационной компанией (государственным предприятием), Министерством туризма и Позитиум. Заключается годовой контракт с Телкомсел и долгосрочный контракт с Позитиум.

Статистические показатели и контроль качества⁷³

В ходе проектов ЦСБ Индонезии использовало такие виды информации, как **детализация вызовов (ДВ)** (мобильные звонки и СМС), и **данные местоположения абонента (LBS)** (подробная запись Интернет-протокола, обновление информации о местоположении, события переключения, включение устройства и т.д.).

ЦСБ Индонезии использует ДМП для производства широкого спектра статистической информации. Некоторые виды статистики сохраняют экспериментальный характер, другие являются официальной информацией, подготовленной на основе ДМП. В рамках этих проектов конфиденциальность обеспечивается благодаря псевдонимизации и К-анонимности.⁷⁴ Следующие виды статистики формируются с использованием ДМП:

⁷² Titi Kanti Lestari, [Mobile phone data for official statistics: Data access, privacy, regulation](#), презентация в ходе сессий Кафе по статистике ЭСКАТО в режиме онлайн, 5 апреля 2021.

⁷³ Edi Setiawan, [Big data for social and demographic statistics](#), презентация в рамках сессий Кафе по статистике ЭСКАТО в режиме онлайн, 2 ноября 2020.

⁷⁴ К-анонимность – метод сохранения конфиденциальности, часто применяемый при использовании больших данных. Основан на принципе деидентификации атрибутов (квазиидентификаторов) всех записей набора данных. Отдельная запись не может быть определена среди других «к-1» записей набора данных. Более подробная информация представлена в [информационном документе Совещания экспертов ЕЭК ООН по обеспечению конфиденциальности статистических данных](#)

- ❖ **Статистика туризма.** ЦСБ формирует статистику внутреннего туризма и трансграничную статистику на основе ДМП.⁷⁵
- ❖ **Анализ данных мероприятий.** ДМП использовались для оценки числа участников двух мероприятий: Азиатских игр и конференции МВФ.⁷⁶
- ❖ **Статистика маятниковой миграции.** ЦСБ на экспериментальной основе формировало статистику маятниковой миграции с использованием ДМП. Хотя показатели ДМП оказались аналогичными данным традиционных обследований, этот источник обнаружил большую экономическую эффективность. Сведения собирались ежемесячно, а не раз в четыре года и только в нескольких городских агломерациях, как при подготовке традиционных обследований. Оценка данных проводилась с большей детализацией, к примеру, на уровне городов и округов.
- ❖ **Столичная статистическая область.** В 2019 году ЦСБ проводило оценки по столичной статистической области (ССО). Это понятие стало альтернативным критерием определения географического региона и заменило термин «столичный регион» (СР), используемый Министерством сельского хозяйства и территориального планирования. ЦСБ использовало ДМП для расчета показателей передвижения населения между главным городом и прилегающими территориями на уровне района в округе Чежунган Бадунг. Это служило цели определения площади столичного региона.⁷⁷
- ❖ **Показатели ЦУР.** ЦСБ сотрудничает с МСЭ и компанией «Позитивум» по использованию ДМП для достижения показателя ЦУР 9.с.1 (Доля населения, охваченного сетью мобильной связи). На основе ДМП проводилась оценка показателя 8.9.1. Вспомогательный счет туризма и непосредственный вклад туризма в ВВП.
- ❖ **Оценка последствий пандемии COVID-19.** ЦСБ использовало ДМП для оценки численности пассажиров пригородного транспорта до и после наступления пандемии COVID-19.

⁷⁵ Titi Kanti Lestari and others, [Indonesia's Experience of using Signaling Mobile Positioning Data for Official Tourism Statistics](#) (2018).

⁷⁶ Statistics Indonesia, [The Use of Mobile Positioning Data for Official Statistics: Indonesia's Examples](#), более подробная информация об практическом опыте Индонезии предлагается по ссылке: Winida Albertha, [Indonesia's Experience on International Migration Statistics](#), презентация в ходе Совещания группы экспертов ООН по совершенствованию данных миграции, Нью-Йорк, США, 1-3 июля 2019, и Statistics Indonesia, [The Use of Mobile Positioning Data, Indonesia's Experiences](#), презентация в ходе Международной конференции по использованию больших данных в целях официальной статистики, Гуанчжоу, Китай, 16-18 октября 2019.

⁷⁷ Edi Setiawan, [Big data for social and demographic statistics](#), презентация в ходе сессий Кафе по статистике ЭСКАТО в режиме онлайн, 2 ноября 2020.

Для обеспечения качества ЦСБ руководствуется положениями Базовых принципов обеспечения качества ЦСБ (при проведении переписей, обследований и сборе административных данных), Базовыми принципами обеспечения качества ООН и Базовыми принципами обеспечения качества при работе с большими данными ЕЭК. Более того, бюро проводит контроль качества на уровне поступления информации, ее обработки и получения конечных результатов. На уровне поступления проверяются недостающие или отсутствующие данные, корректность меток времени и повторяющиеся записи. На следующем этапе выявляются ошибки обработки данных и наслоения. При получении конечных результатов ведется поиск аномалий, определяется соответствие другим данным путем калибровки и сопоставления.⁷⁸

Компания «Позитиум» и ЦСБ разработали новое методологическое пособие и принципы контроля качества⁷⁹, на основе которых регулировалось производство статистики туризма в 2018 году. Работа была завершена в ходе дальнейших дискуссий по методологии с ЦСБ, рассматривался вопрос укрепления потенциала.

Перспективы и проблемы

ЦСБ Индонезии осознает важность использования ДМП в целях официальной статистики. Определенные преимущества включают уменьшение бюджетных расходов, снижение рабочей нагрузки по сбору полевых данных, более высокий уровень дезагрегации данных по времени и пространству, своевременность данных.

Существующее распределение функциональных обязанностей в рамках проектов по ДМП позволяет ЦСБ сочетать собственные знания о потребностях пользователей и понимание определений официальной статистики с опытом компаний частного сектора по использованию методик, охватывающих все аспекты данных, и применению определений для данных. Тем не менее, на первом этапе ЦСБ Индонезии вынуждено было полагаться на партнеров частного сектора при использовании методик и создании системы обработки данных мобильного отслеживания, поскольку ЦСБ не имело доступа к первичным данным.

Одной из главных проблем, возникающих в рамках существующей модели партнерства, является обеспечение постоянного доступа к данным. Условия контракта с ОМС обсуждаются ежегодно и требуют выделения бюджетных средств каждый год. Контракт не предусматривает возобновления и обеспечения непрерывного доступа к данным. Более того, отсутствует законодательство,

⁷⁸ Titi Kanti Lestari, [Mobile phone data for official statistics: Data access, privacy, regulation](#), презентация в ходе сессий Кафе по статистике ЭСКАТО в режиме онлайн, 5 апреля 2021.

⁷⁹ Там же.

регулирующее подобный доступ к информации. Как бы то ни было, ЦСБ при содействии Министерства планирования ведет поиск возможностей для сотрудничества с другим ОМС. Объемы обрабатываемых данных высоки и достигают 144 ТБ в год.

Среди других проблем, заслуживающих внимания, охват данных. Несмотря на то, что Телкомсел является одним из крупнейших ОМС в Индонезии, компания не обслуживает всего населения страны. Это имеет особое значение с точки зрения устранения погрешностей выборки при производстве статистики на основе больших наборов данных. Как было отмечено ранее, ЦСБ рассматривает возможности сотрудничества с другими ОМС. По мере вступления в силу законодательного акта, регулирующего доступ к данным, ЦСБ сможет с большей точностью рассчитывать расходы бюджета на получение доступа к данным, развитие партнерских связей со всеми ОМС и обеспечение постоянного доступа к данным.

Российская Федерация

Обоснование проекта

Росстат заинтересован в разработке концепции использования больших данных альтернативных источников в официальной государственной статистке и реализует два экспериментальных проекта. Первый нацелен на создание алгоритмов расчета статистических показателей на основании данных ОМС в регионах, где проходят эксперименты. Второй - на получение геоаналитических данных ОМС о численности, плотности и миграции населения в целях дополнения первичных данных и результатов всероссийской переписи населения 2020 года.

В ходе экспериментальных проектов Росстат ставит перед собой ряд задач: снижение нагрузки на респондентов, устранение случаев дублирования показателей, если информация поступает из разных источников, объединение данных в единой базе и разработка новых показателей в области туризма.

Основные заинтересованные стороны, модели партнерства по использованию данных и поток данных

В рамках контрактного соглашения Росстат сотрудничал с Центром экономики инфраструктуры (ЦЭИ) по вопросам подготовки экспериментальной статистики на основе данных ОМС.

Для целей проекта ЦЭИ получил данные трех операторов мобильной связи: МТС, Т2 Мобайл и Тинькофф Мобайл. Информация для двух проектов была предоставлена на основании коммерческого соглашения. Исследовательская организация выполняла функции сбора, обработки данных и расчета показателей.

Полученные данные были частично обработаны и оператором мобильной связи, и привлеченной организацией, анонимизированы по абонентам и агрегированы по каждому представляющему интерес отрезку времени. ОМС провел обработку первичных данных. Каждый оператор имеет собственные алгоритмы машинного обучения, доступ к которым закрыт. ЦЭИ создал алгоритмы и методологию для расчета показателей на основе информации ОМС.

Росстат лишен доступа к первичным данным операторов мобильной связи, однако предоставил со своей стороны рекомендации по разработке методик расчета показателей. Росстат получил данные в виде аналитических таблиц.

Несмотря на отсутствие сложностей с получением данных для целей экспериментальных проектов, доступ к данным ОМС на постоянной основе ограничивается финансовыми условиями и требованиями законодательства.

Нормативно-правовая база проектов

Национальная нормативно-правовая база, содействующая использованию данных мобильной телефонной связи, отсутствует.

Несмотря на внесение недавних поправок (июнь 2021 года) в Федеральный закон «Об официальном статистическом учете и системе государственной статистики в Российской Федерации» от 2007 года, предусматривающих включение «иных источников данных» в перечень источников данных, используемых для производства официальной статистики, новое положение не затронуло данные операторов мобильной связи.

Обеспечение доступа к данным и защита неприкосновенности личной информации регулируются Законом о персональных данных, который гласит (параграф 6, статья 1.9): *«обработка персональных данных осуществляется в статистических или иных исследовательских целях [...] при условии обязательного обезличивания персональных данных», «оператор осуществляет обработку персональных данных для статистических или иных исследовательских целей, для осуществления профессиональной деятельности журналиста либо научной, литературной или иной творческой деятельности, если при этом не нарушаются права и законные интересы субъекта персональных данных».*⁸⁰

В статистике туризма нормативно-правовая база, обеспечивающая осуществление экспериментальных проектов по большим данным, включает Стратегию по развитию туризма в Российской Федерации на период до 2035 года (утверждена Распоряжением Правительства Российской Федерации от 20 сентября 2019 года № 2129-р, глава 15 «Развитие системы сбора, обработки и анализа статистических данных») и Распоряжение Правительства Российской Федерации от 23 марта 2018 года № 482-р (в редакции от 28 мая 2020 года) «Об утверждении плана мероприятий («дорожной карты») по совершенствованию законодательства и устранению административных барьеров в целях обеспечения реализации Национальной технологической инициативы по направлению «Технет» (передовые производственные технологии)».

Статистические показатели

В результате первого пилотного проекта Росстат получил экспериментальные расчеты по двум регионам от одного ОМС. Сведения включали туристические

⁸⁰ Российская Федерация, [Федеральный закон № 153 от 27.07.2006 «О персональных данных»](#), в редакции от 02.07.2021 (на русском языке)

показатели. Второй экспериментальный проект был запущен в одном из регионов для оценки туристского потока на основе данных четырех ОМС. В целом проекты ставили своей задачей формирование демографической, миграционной статистики и статистики туризма. По результатам была получена экспериментальная статистика. Сформированы следующие виды статистических показателей:

Демографическая и миграционная статистика:

- численность постоянного населения;
- численность временного населения;
- численность населения, использующего второе жилье;
- численность маятниковых трудовых мигрантов;
- численность лиц, работающих вахтовым методом;
- характеристики указанных категорий населения: пол, возраст, гражданство;
- население в разбивке по субъектам Российской Федерации и муниципальным образованиям;
- скорректированные оценки численности населения в разбивке по субъектам Российской Федерации и муниципальным образованиям;
- численность населения субъектов Российской Федерации, работающего в других субъектах Российской Федерации;
- численность населения населенных пунктов, работающего в других муниципальных образованиях;
- оценка потоков маятниковой миграции в агломерациях;
- динамика потоков маятниковых трудовых мигрантов по субъектам Российской Федерации;
- динамика потоков маятниковых трудовых мигрантов по муниципальным образованиям;
- оценка численности постоянного населения;
- оценка численности вынужденных переселенцев.

Статистика туризма:

- оценка численности российских и иностранных туристов;
- число туристических поездок российских и иностранных туристов в регион исследования;
- средний срок пребывания российских и иностранных туристов и экскурсантов в регионе исследования;
- характеристики российских и иностранных туристов и экскурсантов: пол, возраст, уровень дохода, срок пребывания на территории, частота посещения

территории, количество посещенных населенных пунктов (показатели агрегируются на региональном уровне);

- оценка численности российских и иностранных туристов в средствах размещения неформального сектора;
- оценка численности посетителей туристических достопримечательностей.

Следующие справочные данные также были использованы в экспериментальных проектах. Представлены по категориям статистической информации:

Демографическая статистика:

- Общероссийский классификатор территорий муниципальных образований;
- первичные административные данные о естественном движении и миграции населения.

Статистика туризма:

- Федеральный перечень туристских объектов Ростуризма;
- сайты для бронирования гостиниц (например, Hotellook);
- сайты для самостоятельной аренды жилья (например, Sutochno.ru);
- веб-ресурсы для покупки авиабилетов (например, АО «Транспортная клиринговая палата»);
- сайты для поиска и бронирования авиабилетов (например, Авиасейлс);
- сайты для покупки железнодорожных билетов (например, информация ОАО «Федеральная пассажирская компания», предоставляемая ОАО «Российские железные дороги»);
- Реестр объектов культурного наследия Министерства культуры России.

И оператор мобильной связи, и привлеченная организация осуществляли контроль качества данных. Общая оценка качества в рамках проекта проводилась на основе Технического задания к государственному контракту на оказание услуг, обеспечивающих доступ к геоаналитическим данным по численности, плотности и миграции населения в муниципальных образованиях и субъектах Российской Федерации, с целью дополнения результатов общероссийской переписи населения 2020 года.

