

Концептуальные предложения по формированию организационных и методологических подходов к использованию больших данных в статистических системах государств–участников СНГ

Содержание

1.	Общие положения.....	2
2.	Цели, задачи и ожидаемые результаты.....	2
3.	Международные рекомендации и опыт использования больших данных в официальной статистике.....	5
4.	Анализ текущего состояния и потенциала использования больших данных.....	7
4.1.	Текущее состояние в официальной статистике в СНГ.....	7
4.2.	4.2. SWOT-анализ основных видов больших данных.....	19
5.	Источники и виды больших данных.....	26
6.	Подходы к использованию больших данных.....	31
6.1.	Методологические подходы.....	31
6.2.	Организационные подходы.....	38
6.3.	Инфраструктура и кадры.....	41
7.	Нормативно-правовые аспекты.....	46
8.	Этапы реализации.....	49
9.	Заключительные положения.....	51

1. Общие положения

Основание для разработки проекта Концепции формирования организационных и методологических подходов к использованию больших данных в статистических системах государств-участников СНГ (далее – Концепция) - пункт 4.1 Технического задания к Контракту от 14 августа 2025 г. № CIS-SD/25-CONS-2.6 на оказание услуг по подготовке методических материалов, рекомендаций, стандартов, практических руководств, лучших практик по использованию больших данных и искусственного интеллекта в официальной статистике (этап 2025 года).

Концепция является логическим продолжением, расширяющим основные положения Заключительного отчета по Контракту № CIS-SD/24-ICONS-2.8.1.

Работа выполнена в рамках проекта «Развитие статистики СНГ» в соответствии с Распоряжением Правительства РФ от 28.12.2023 №3996-рс и Календарным планом по реализации проекта «Развитие статистики СНГ» на 2025 г., одобренным координационной рабочей группой и Советом руководителей статистических служб государств-участников СНГ от 5 декабря 2024 г.

2. Цели, задачи и ожидаемые результаты

Целью выполнения данной работы является разработка на основе обобщения и анализа современного международного опыта основных концептуальных организационных, методологических и инфраструктурных подходов к организации использования больших данных в официальной статистике государств - участников СНГ (далее – стран СНГ).

Сформулированные в рамках Концепции подходы основываются на характеристике ключевых источников больших данных на примере данных сотовых операторов о населении и данных об использовании сельскохозяйственных угодий с использованием средств спутникового мониторинга и увязаны с Типовой моделью производства статистической информации ООН (далее – ТМПСИ).

Содержащиеся в Концепции выводы и подходы на основе международного опыта были подтверждены в Докладе Комитета экспертов по использованию больших данных и обработке и анализу больших данных для целей официальной статистики на 57-й сессии Статистической комиссии ООН в марте 2026 года.

**Основные общественно значимые цели и результаты (далее – КРІ):
Ускорение, повышение детализации официальной статистики**

КРІ:

2028 – реализованы не менее 3-х пилотных проектов по использованию больших данных в официальной статистике СНГ, обеспечивающих при формировании официальной статистической информации в рамках реализованных пилотных проектов:

- ускорение формирования на 10-20%;
- детализацию до уровня отдельных социально-демографических групп и муниципальных образований.

Единая методология, сопоставимость страновых данных

КРІ:

2028 – разработан проект пакета типовых (модельных) нормативных правовых актов (далее – НПА), в том числе стандартов качества и безопасности, и рекомендаций по использованию больших данных в официальной статистике СНГ;

2028 - разработано не менее 3-х типовых методик по использованию больших данных в официальной статистике, обеспечивающих в рамках реализованных пилотных проектов единство методологии и сопоставимость официальной статистической информации стран СНГ.

Новая полезная информация для органов власти и бизнеса

КРІ:

2028 – на основе разработанных типовых методик по использованию больших данных в официальной статистике обеспечено формирование в рамках реализованных пилотных проектов не менее 10 новых показателей для бизнеса и органов государственного управления, в том числе:

- моментальная (на дату) численность населения;
- въездные - выездные потоки мигрантов/туристов;
- социально-демографический портрет мигранта/туриста;
- новые показатели статистики сельского хозяйства и природопользования.

Модернизация статистического производства на основе использования больших данных

КРІ:

2028 – разработана адаптированная с учетом особенностей национального регулирования в сфере официальной статистики и больших данных в странах

СНГ Типовая модель производства статистической информации. Версия 5.2 (далее – ТМПСИ 5.2);

2028 – подготовлены предложения по масштабированию ТМПСИ на основе использования больших данных в официальной статистике на период до 2030 года не менее чем в 3-х отраслях статистики стран СНГ.

Создание базовых организационных механизмов и ИТ-инфраструктуры

2028 – сформированы базовые международные организационные механизмы и партнерства в рамках реализации Концепции;

2028 – разработана типовая ИТ-архитектура и сформирована базовая международная ИТ-инфраструктура в рамках реализации Концепции.

Повышение статистического потенциала НСС

2028 – разработано не менее 3-х курсов дистанционного обучения работе с большими данными и обеспечено обучение не менее 50-ти специалистов НСС стран СНГ.

2028 – разработаны типовая (модельная) Матрица зрелости (Maturity Matrix) для проведения самооценки НСС готовности к использованию больших данных в официальной статистике и типовая Дорожная карта по повышению статистического потенциала НСС на основе Матрицы зрелости.

Международное лидерство

КРІ:

2028 – специалисты НСС стран СНГ – участвуют во всех ключевых рабочих/экспертных группах по большим данным в рамках ООН (по выбранным направлениям работ);

2028 – примеры лучших практик по использованию больших данных в официальной статистике НСС стран СНГ включены в углубленные обзоры лучших мировых практик в рамках ООН.

Основные потребители

Ключевыми потребителями результатов реализации Концепции будут выступать в странах СНГ:

- НСС и органы власти всех уровней;
- бизнес;
- наука и экспертное сообщество;
- образование;
- население.

3.Международные рекомендации и опыт использования больших данных

За последнее десятилетие Статистическая комиссия ООН, Евростат, национальные статистические службы (далее – НСС) мира существенно продвинулись практически во всех важнейших направлениях, связанных с использованием больших данных в официальной статистике:

1. На международном уровне были разработаны и приняты ключевые нормативно-правовые акты, стандарты, рекомендации, практические руководства, направленные на обеспечение использования больших данных в официальной статистике:

- по совершенствованию статистического законодательства для обеспечения доступа НСС к частным источникам больших данных, обеспечению безопасности и конфиденциальности при использовании больших данных;

- по модернизации типового процесса (модели) производства статистической информации с учетом использования больших данных;

- по методологии использования основных видов больших данных в отраслях официальной статистики;

- по обеспечению качества официальной статистической информации, формируемой с использованием больших данных;

- по повышению статистического потенциала и компетенций НСС и по целому ряду других вопросов, связанных с использованием больших данных в официальной статистике.

2. НСС были реализованы пилотные проекты с основными видами больших данных и сформирована соответствующая экспериментальная статистика.

3. В ряде НСС началось регулярное использование отдельных видов больших данных при формировании официальной статистики. Наиболее активными сферами использования больших данных в официальной статистике стали:

- использование данных сотовых операторов при формировании статистики населения, туризма, транспортных потоков (НСС Италии, Нидерландов, Португалии, Эстонии, Латвии, Австралии, Индонезии и ряда других стран);

- использование данных дистанционного зондирования земли (далее – ДЗЗ) при формировании статистики сельского хозяйства и природопользования (НСС Нидерландов, Италии, Канады, Китая, Южной Кореи, Индонезии, Бразилии, Мексики и ряда других).

4. Многие НСС разработали перспективные планы, программы, стратегии, дорожные карты в этой сфере.
5. Были созданы и эффективно работают международные организационные механизмы, в первую очередь, Комитет экспертов по большим данным и науке о данных для официальной статистики ООН (в настоящее время - Комитет экспертов по использованию больших данных и обработке и анализу больших данных для целей официальной статистики), в рамках которого активно функционируют 9 тематических рабочих групп (далее – ТРГ).
6. Сформировано и активно развивается ключевое инфраструктурное направление в рамках ООН - Глобальная платформа ООН по большим данным для официальной статистики (2017), сформирована и развивается сеть Региональных и Секторальных центров Глобальной платформы (2019-2022).
7. На международном и национальном уровне созданы партнерства, центры компетенций, учебные центры и программы с участием международных организаций, НСС, науки и бизнеса по использованию больших данных в официальной статистике.
8. Сформировано международное профессиональное и экспертное сообщество, организованы и регулярно проводятся World Data Forum, многочисленные международные и национальные конференции, семинары, олимпиады, хакатоны/дататоны и другие профессиональные мероприятия.

Обобщение и анализ международного опыта позволяют сделать вывод о том, что в настоящее время на международном уровне:

- сформулированы базовые подходы к нормативному правовому регулированию, в том числе обеспечению конфиденциальности и качества больших данных при их использовании в официальной статистике;
- подготовлены базовые международные рекомендации и руководства по методологии и практическому использованию основных видов больших данных в официальной статистике;
- сформированы глобальные, региональные и национальные ИТ-платформы, центры компетенций, международные экспертные и рабочие группы по большим данным и науке о данных в официальной статистике;
- сформирована система повышения квалификации кадров НСС.

В целом это обеспечивает на международном уровне переход от проведения экспериментальных расчетов и реализации пилотных проектов к регулярному использованию наиболее проработанных источников больших

данных (в первую очередь, данных сотовых операторов и ДЗЗ) при формировании официальной статистической информации.

При этом в последние годы во всех перечисленных выше направлениях, связанных с использованием больших данных в официальной статистике, серьезно усиливается акцент на использование искусственного интеллекта, науки о данных и новейших технологий сбора, обработки, передачи, хранения, анализа и представления больших данных.

4. Анализ текущего состояния и потенциала использования больших данных

4.1. Текущее состояние в официальной статистике в СНГ

Стратегии развития официальной статистики

В большинстве стран СНГ на национальном уровне разработаны и приняты стратегии развития официальной (государственной) статистики.

Все стратегии (концепции, среднесрочные программы, планы и дорожные карты развития) базируются на основополагающих принципах официальной статистики ООН, уделяют первостепенное внимание обеспечению качества официальной статистики, конфиденциальности и безопасности первичных статистических данных, снижению нагрузки на респондентов, максимальному использованию административных данных, соответствию международным стандартам, модернизации процессов производства статистической информации, повышению статистического потенциала НСС.

При этом цели и задачи стратегий и зафиксированные в них основные направления развития официальной статистики в части использования больших данных существенно зависят, в первую очередь, от наличного статистического потенциала НСС, а также от даты принятия и периода действия соответствующих стратегий.

Наиболее развернутые подходы к использованию больших данных в официальной статистике представлены в стратегиях России и Казахстана, которые были приняты (соответственно) в конце 2024 года и в 2025 году.

Ключевые положения указанных стратегий в отношении использования больших данных в официальной статистике:

Россия

Стратегия развития системы государственной статистики и Росстата до 2030 года (утверждена Распоряжением Правительства Российской Федерации от 30 декабря 2024 г. № 4159-р) (далее – Стратегия):

«Ключевыми задачами по выявлению потребностей в информации, планированию и обеспечению качества статистических данных являются ...использование новых источников данных и внедрение в статистическую практику новых инструментов и алгоритмов».

«...трансформация федерального государственного бюджетного учреждения «Научно-исследовательский институт проблем социально-экономической статистики Федеральной службы государственной статистики» в единый методологический центр - передовой научный центр по разработке опережающей методологии и проведению поисковых и экспериментальных расчетов с использованием новых технологий и новых источников данных».

«...В рамках проекта по использованию больших данных планируется: провести актуализацию концепции использования больших данных в системе государственной статистики с учетом международного опыта;

осуществить поиск, анализ и оценку возможных источников больших данных с учетом возможностей их дальнейшего применения в официальной статистике, регулярно актуализировать состав возможных источников больших данных;

провести категоризацию источников больших данных по их применимости для официального статистического учета;

расширить практику и провести доработку инструментов «машинного обучения» для анализа больших данных;

интегрировать большие данные в статистическую методологию с учетом возможности замены традиционных первичных источников статистической информации, а также использовать большие данные для верификации традиционных показателей.

Целевой индикатор к 2030 году - более половины статистических направлений используют большие данные».

Казахстан

План реформирования Бюро национальной статистики на 2025-2027 годы (утвержден Руководителем Бюро национальной статистики Агентства по стратегическому планированию и реформам Республики Казахстан. Астана, 2025):

«1.3. Использование административных и альтернативных источников данных

1.3.1. Заключение соглашений по интеграции БНС с 10 источниками административных и альтернативных данных...

1.3.2. Интеграция ИИС «е-Статистика» с источниками административных и альтернативных данных...

1.3.3. Интеграция ИИС «е-Статистика» с учетной системой «1С-Бухгалтерия» для автоматизированного сбора данных.

1.3.6. Переход на использование альтернативных данных банков второго уровня (БВУ) и административных данных Цифровой карты семьи в расчете среднедушевых номинальных денежных доходов.

1.3.7. Переход на использование альтернативных данных БВУ для ежемесячной оценки уровня доходов и их структуры по источникам.

1.3.8. Переход на использование альтернативных данных БВУ для ежемесячной оценки расходов и их структуры по продовольственным, непродовольственным товарам и услугам.

1.3.9. Переход на использование альтернативных данных БВУ для ежемесячной оценки структуры занятости населения.

1.3.10. Переход на использование данных ИС «Электронные счета-фактуры» КГД МФ для формирования объемов реализации оптовой торговли и индексов цен оптовых продаж, и ОФД по объемам реализации розничной торговли.

1.3.11. Переход на использование данных КГД МФ для формирования статистических показателей структурной статистики.

1.3.12. Интеграция ИИС «е-Статистика» с учетной системой на базе SAP для автоматизированного сбора данных.

2.1. Внедрение архитектуры Централизованного хранилища данных для обработки больших данных

2.1.5. Миграция статистических, административных и альтернативных данных в Центральное хранилище данных.

2.3. Аналитика данных и реализация аналитических проектов

2.3.2. Формирование экспериментальной статистики по туризму на основе данных мобильного позиционирования.

2.3.5. Разработка аналитической карты бизнес-демографии...

4. Совершенствование нормативно-правовой и методологической базы

4.3. Утверждение Правил представления операторами связи и (или) владельцами сетей связи, осуществляющих деятельность на территории Республики Казахстан в уполномоченный орган в области государственной статистики для осуществления статистической деятельности без передачи третьим лицам обезличенных данных об абонентах».

Общий вывод

В целом пример последних принятых стратегий подтверждает высокую степень готовности НСС стран СНГ к реализации проектов по использованию больших данных в официальной статистике.

Важным шагом при разработке стратегий дальнейшего развития может стать проведение предварительной самооценки НСС стран СНГ степени своей институциональной готовности к использованию больших данных на основе рекомендованной ООН типовой модели «Матрицы зрелости по большим данным».

Внедрение Типовой модели производства статистической информации (ТМПСИ)

Все национальные статистические службы (НСС) стран СНГ в той или иной форме ориентируются на ТМПСИ. Однако степень ее внедрения различается.

Беларусь:

Статус: Используется как методологический ориентир для совершенствования процессов и гармонизации с международными стандартами.

Внедрение с 2015 года.

Специфика:

Тесная связь с системой качества: Внедрение ТМПСИ тесно увязано с внедрением СТБ 8.0 и внутренних инструкций по качеству. Каждый этап ТМПСИ сопровождается контрольными точками по качеству.

Казахстан:

Статус: ТМПСИ является официальной концептуальной основой для реинжиниринга статистических бизнес-процессов и развития ИТ-архитектуры.

Активная фаза внедрения началась в 2018-2020 годы в рамках реализации Концепции развития государственной статистики.

Модель и специфика:

Модель: ТМПСИ версии 5.x. Используется как каркас для описания «as-is» и проектирования «to-be» процессов.

Специфика:

1. Интеграция с ИТ: ТМПСИ легла в основу разработки и внедрения Единой платформы сбора и обработки данных (ЕПСОД) Бюро национальной статистики. Эта платформа фактически реализует этапы ТМПСИ в виде цифровых сервисов.

2. Акцент на этапах, связанных с «неопросными» данными: Особое внимание уделяется этапам «Проектирование получения данных» (Design data acquisition) и «Обработка» (Process) для интеграции больших данных.

Узбекистан и Кыргызстан:

Статус: Активное знакомство и пилотное внедрение при поддержке международных доноров (Всемирный банк, ПРООН, Евростат).

Внедрение в рамках проектов, стартовавших после 2020 года.

Специфика:

1. Обучение и наращивание потенциала: Основной фокус — на обучении сотрудников НСС принципам ТМПСИ.

2. Пилотные применения: Модель применяется для редизайна конкретных процессов, например, для организации сбора данных по коротким опросам предприятий или для описания процесса использования данных ДЗЗ.

Россия:

Статус:

Приказом Росстата от 7 декабря 2018 г. N 732 утверждены Методологические положения по организации процессов производства официальной статистической информации.

Де-факто с 2018 года внедрена в ключевых процессах и ИТ-системах Росстата. Принципы модели глубоко интегрированы в методологию и новые технологические платформы.

Модель и специфика:

Модель: Фактически используется логика ТМПСИ, адаптированная к масштабам и структуре Росстата.

Специфика:

1. Отраслевая детализация: Для каждого направления статистики (сельскохозяйственная перепись, экономическая статистика) разработаны детальные технологические схемы, которые являются отражением этапов ТМПСИ.

2. Централизация ИТ: Создание централизованных систем сбора и обработки (например, для экономической переписи) — это прямая реализация принципов ТМПСИ по стандартизации и сквозной автоматизации процессов.

3. Акцент на валидацию и контроль: Этапы «Обработка» и «Анализ» (Process, Analyse) в российской практике чрезвычайно детализированы и формализованы, что соответствует национальной специфике требований к качеству.

Общие вызовы:

Недостаток ИТ-инфраструктуры: Полная реализация ТМПСИ требует наличия у НСС современных гибких ИТ-систем.

Кадры: Нехватка специалистов по сквозным статистическим бизнес-процессам.

Общий вывод:

ТМПСИ в странах СНГ стала практическим инструментом модернизации, который с разной степенью интенсивности используется для повышения эффективности, стандартизации и качества статистического производства.

На последующих этапах необходимо развитие уже реализованной модели ТМПСИ 5.1 до модели ТМПСИ 5.2 с учетом национальной специфики использования больших данных НСС стран СНГ.

Системы мониторинга и контроля качества

Все НСС стран СНГ в той или иной форме используют системы контроля качества, основанные на рамочных принципах ООН. При этом в Беларуси (Белстат), Казахстане (БНС Республики Казахстан), России (Росстат), Узбекистане (Госкомстат Узбекистана) руководства (политики, стандарты) по управлению и контролю качества официальной статистической информации утверждены в явной форме.

Однако указанные системы контроля качества в настоящее время не учитывают специфики формирования официальной статистической информации с использованием больших данных.

Использование данных операторов сотовой связи

Страны СНГ демонстрируют растущий интерес к использованию данных операторов сотовой связи (далее – СО), однако их внедрение в официальную статистику сталкивается с общими вызовами: отсутствие четкой правовой базы, ограниченный ресурсный потенциал статистических служб и вопросы доверия со стороны СО. Наиболее продвинутые проекты реализуются, как правило, при поддержке международных организаций.

Азербайджан

Период: 2021 год - по настоящее время (пилотные исследования).

Проекты при поддержке Всемирного банка.

Цели: Оценка мобильности населения в Баку, анализ транспортных потоков и планирование городской инфраструктуры.

Участники:

Госорганы: Государственный комитет по статистике Азербайджана, Министерство цифрового развития и транспорта.

Операторы: Партнерство с одним или несколькими крупнейшими СО (Azercell, Bакcell, Nar).

Основные результаты:

Проведен анализ загруженности дорог и пассажиропотоков в столице.

Оценена эффективность работы общественного транспорта.

Армения

Период: 2020-2022 гг. (пилот в рамках программы ООН).

Цели: Оперативная оценка экономической активности и региональных различий, особенно в контексте последствий пандемии COVID-19 и чрезвычайных ситуаций.

Участники:

Национальная статслужба: Статистический комитет Армении.

Операторы: Партнерство с ведущим СО Ucom (ранее Viva-MTS).

Международные организации: UN Global Pulse, ПРООН.

Основные результаты:

Разработаны модели для оперативной оценки изменений экономической активности на основе данных о мобильности и использовании сети.

Проанализировано влияние кризисов на внутреннюю миграцию и перемещения населения.

Казахстан

Период: С 2020 года (ряд проектов и инициатив).

Цели: Исследование возможностей использования данных СО для мониторинга туристских потоков, оценки мобильности населения и верификации данных переписи населения.

Участники:

Национальная статслужба: Бюро национальной статистики (БНС) Агентства по стратегическому планированию и реформам (ранее Комитет по статистике МНЭ РК).

Операторы: Проект реализован с сотовым оператором Kcell, ведутся переговоры с другими крупнейшими операторами (KaR-Tel (Beeline) и Tele2).

Область исследования: Туризм, внутренняя миграция, мобильность населения.

Основные результаты:

Разработана методология и проведено сопоставление оценок туристского потока и других показателей статистики туризма на основе данных СО Kcell, статистической отчетности и обследований домашних

хозяйств, подтвердившее достаточно высокую достоверность и точность данных СО.

Кыргызстан

Период: 2021-2023 гг. (пилот при поддержке Всемирного банка).

Цели: Оценка внутренней миграции и мобильности населения, улучшение данных о численности и распределении населения.

Участники:

Национальная статслужба: Национальный статистический комитет КР (Нацстатком).

Операторы: Партнерство с основными СО (Sky Mobile (Beeline), MegaCom, O!).

Международные организации: Всемирный банк.

Основные результаты:

Получены первые оценки масштабов внутренней миграции между регионами.

Апробированы методики расчета наличного населения.

Россия

Период: 2018-2021 годы (пилоты)

Цели: Разработка концепции использования больших данных в официальной статистике, в том числе в части данных СО: оценка туристских потоков, верификация данных переписи населения.

Участники: Росстат, Аналитический центр при Правительстве РФ.

Операторы: Партнерство с крупнейшими СО и агрегаторами данных.

Основные результаты:

Разработка методологических подходов к использованию данных СО в официальной статистике в сфере статистики населения и туризма.

Период: 2021-2023 годы (пилоты на региональном уровне)

Цели: Разработка методологии оценки туристских потоков и верификации данных переписи населения на основе данных СО.

Участники: Росстат, Аналитический центр при Правительстве РФ.

Операторы: Партнерство с крупнейшими СО и агрегаторами данных.

Основные результаты:

Верификация и апробация основных методологических подходов к использованию данных СО в официальной статистике в сфере статистики туризма.

Выработка модели взаимодействия с региональными органами власти в сфере статистики туризма, в том числе в области использования данных СО.

Период: 2024 год по настоящее время (национальная платформа - эксперимент на федеральном уровне)

Цели: Проведение масштабного эксперимента по созданию единой национальной платформы для сбора, обработки и анализа данных СО и системы сервисов на ее основе для госорганов и других уполномоченных пользователей.

Участники: Минцифры России (координатор, оператор), Росстат (функциональный заказчик сервисов), другие госорганы.

Операторы: Все сотовые операторы: МТС, МегаФон, Билайн, Tele2.

Область исследования: Статистика населения, миграция, туризм.

Узбекистан

Период: 2022 год - по настоящее время (начало исследований).

Цели: Изучение возможностей для оценки туристических потоков и внутренней мобильности.

Участники:

Национальная статслужба: Госкомстат Узбекистана.

Операторы: Начало диалога с СО (Ucell, Beeline Uzbekistan, Mobiuz).

Область исследования: Туризм, мобильность населения.

Основные результаты:

На стадии предварительного анализа и переговоров.

Беларусь, Таджикистан и Туркменистан

Информация о проектах по использованию данных СО в официальной статистике отсутствует.

Общие перспективы

Переход от экспериментальной к официальной статистике: По мере развития методологии и укрепления доверия показатели на основе данных СО будут переходить в разряд официальной статистики.

Интеграция с другими источниками данных: Гибридные показатели, сочетающие данные СО с опросами, спутниковыми снимками и данными социальных сетей, будут давать более полную картину.

Стандартизация: ООН и другие международные организации работают над созданием единых методологических стандартов для обеспечения сопоставимости данных между странами.

Реальное время (Near Real-time): Развитие в сторону «Nowcasting» – предоставления почти мгновенных статистических оценок для поддержки принятия решений (например, во время кризисов).

Расширение сферы применения: Применение для мониторинга ЦУР, например, для показателей, связанных с городами, транспортом и неравенством.

Повышение статистического потенциала: Необходимо повышать компетенции специалистов НСС в сфере работы с большими данными, в том числе, с данными СО в целях их интеграции в официальную статистику.

Использование данных ДЗЗ

Продовольственная и сельскохозяйственная организация ООН (далее – ФАО) является основным методологом и координатором работ по использованию данных ДЗЗ в сельскохозяйственной статистике.

Проекты ФАО в странах СНГ

Основной фокус проектов ФАО в странах Центральной Азии смещен в сторону решения проблем сокращения (деградации) сельскохозяйственных земель, управления водными и пастбищными ресурсами в условиях изменения климата с использованием ДЗЗ как ключевого инструмента для получения объективных данных.

Казахстан

Ключевой проект: Внедрение Системы мониторинга сельскохозяйственных стрессов (ASIS) и оценка пастбищ.

Цель: Повышение устойчивости агропромышленного комплекса к засухам и совершенствование системы управления пастбищами.

Деятельность в части ДЗЗ:

Адаптация методологии ASIS: Казахстанский научно-исследовательский институт экономики агропромышленного комплекса и развития сельских территорий при поддержке ФАО провел работу по адаптации глобальной методологии ASIS к национальным условиям. Это позволяет на регулярной основе получать карты индекса стресса растительности (VHI) для мониторинга засух на пастбищных и пахотных землях.

Оценка состояния пастбищ: С использованием данных ДЗЗ (Landsat, Sentinel-2) и платформы SEPAL проводятся работы по картографированию продуктивности пастбищ, оценке нагрузки на пастбищные экосистемы и выявлению деградированных территорий.

Партнеры: Министерство сельского хозяйства РК, КазНИИ АПК и РСТ.

Результат: Методология внедрена и используется для подготовки аналитических записок и прогнозов. Технологии переданы от ФАО национальному институту.

Кыргызстан

Ключевой проект: Устойчивое управление природными ресурсами в условиях изменения климата. Внедрение инструментов ДЗЗ для оценки земель.

Цель: Создание национального потенциала для мониторинга деградации земель и обоснования планов по адаптации к изменению климата.

Деятельность в части ДЗЗ:

Обучение использованию Collect Earth: ФАО провела серию интенсивных национальных тренингов для специалистов из Минсельхоза, Агентства по охране окружающей среды и научных кругов. Участники обучены проводить оценку земельного покрова и его изменений с помощью Collect Earth на основе снимков высокого разрешения (Google Earth, Bing Maps).

Пилотные проекты по оценке деградации земель: В пилотных районах были проведены оценки, совмещающие полевые исследования и дешифрирование данных ДЗЗ. Это позволило верифицировать карты деградации и оценить масштабы проблемы.

Партнеры: Министерство сельского, водного хозяйства и развития регионов, Государственное агентство охраны окружающей среды и лесного хозяйства.

Результат: Создан пул сертифицированных национальных экспертов, способных самостоятельно проводить мониторинг земельных ресурсов с использованием современных геопространственных методов. Полученные данные используются для отчетности в рамках КБО ООН (например, по ЦУР 15.3.1 «Доля деградированных земель»).

Узбекистан

Ключевой проект: Внедрение климатически оптимизированных подходов в сельское хозяйство и мониторинг засух.

Цель: Повышение продуктивности и устойчивости сельского хозяйства, в частности, в критически важном регионе — Приаралье.

Деятельность в части ДЗЗ:

Мониторинг засух с помощью ASIS: ФАО оказала поддержку Узгидромету в использовании системы ASIS для оперативного мониторинга засушливых явлений на территории страны.

Оценка потенциала пастбищ в Приаралье: Используя данные ДЗЗ, проводятся работы по оценке состояния и продуктивности пастбищ в зоне, подверженной опустыниванию. Это необходимо для разработки планов по устойчивому животноводству.

Использование платформы SEPAL: Национальные специалисты обучены работе в облачной платформе SEPAL для самостоятельного проведения анализа многолетней динамики растительного покрова, классификации землепользования.

Партнеры: Узгидромет, Министерство сельского хозяйства РУ.

Результат: Узбекистан активно интегрирует данные ДЗЗ в свою практику. Информация, полученная с помощью инструментов ФАО, используется для подготовки государственных программ по развитию регионов, подверженных засухам и деградации земель.

Общий вывод по региону:

Во всех странах ключевой акцент делался ФАО на передаче технологий (SEPAL, Collect Earth) и методологии (ASIS), а также на обучении специалистов национальных органов, чтобы они могли самостоятельно и на регулярной основе использовать данные ДЗЗ для принятия управленческих решений в сфере АПК и природопользования.

НСС де-факто не являлись ключевыми партнерами в указанных проектах ФАО в Казахстане и Узбекистане. Проекты были сфокусированы на прикладных задачах оперативного мониторинга (засухи, состояние пастбищ, деградация земель). За эти направления в странах СНГ традиционно отвечают не НСС, а отраслевые министерства (сельского хозяйства) и специализированные агентства (гидромет, экологии).

Целевая аудитория проектов ФАО: Помощь ФАО была направлена именно на государственные органы, которые принимают оперативные управленческие решения в сфере АПК и природопользования. Например, карты засух от Узгидромета нужны Минсельхозу для планирования мероприятий, а не НСС для расчета валового сбора.

Статус методологий ДЗЗ: На момент реализации проектов методы ДЗЗ еще не были в полной мере стандартизированы и легитимизированы для производства официальных статистических данных. Они рассматривались как ценный, но все же вспомогательный или экспериментальный инструмент.

Опосредованное влияние и перспективы интеграции:

1. Стратификация и улучшение рамок выборки: НСС могут использовать карты землепользования, созданные в рамках этих проектов (например, с помощью Collect Earth), для совершенствования основы выборки сельскохозяйственных обследований. Это позволяет более точно определять целевые совокупности и снижать ошибку покрытия.

2. Верификация и оценка качества данных: Данные ДЗЗ могут использоваться НСС для перекрестной проверки (верификации) данных,

поступающих от сельхозпроизводителей или из административных источников. Например, для подтверждения факта использования земельного участка.

3. Подготовка к будущим переписям: Как следует из Программы Всемирной переписи сельского хозяйства 2030 (WCA 2030), ДЗЗ рекомендованы для использования практически на всех этапах переписи. Опыт, накопленный отраслевыми ведомствами, создает национальный потенциал, которым НСС могут воспользоваться при проведении следующей сельскохозяйственной переписи.

4. Расширение показателей ЦУР: Для отчетности по Целям устойчивого развития (например, индикатор 15.3.1 по деградации земель) НСС часто являются координирующим органом. Данные, полученные экологическими и сельскохозяйственными ведомствами с помощью ДЗЗ, могут быть официально использованы НСС для представления в международные организации.

Опыт России

В Российской Федерации Росстатом, начиная с Всероссийской сельскохозяйственной переписи 2016 года, и в дальнейшем в 2021-2024 годах совместно с Институтом космических исследований Земли (ИКИЗ) был реализован ряд масштабных исследовательских и практически-ориентированных проектов по использованию данных ДЗЗ в различных областях сельскохозяйственной статистики.

Реализованные проекты показали безусловную возможность и эффективность использования ДЗЗ в официальной статистике в области сельского хозяйства.

При этом одним из важных результатов указанных исследований стала разработка проекта концепции совершенствования методологии и организации всероссийских сельскохозяйственных переписей (микрпереписей) и текущей сельскохозяйственной статистики на основе инновационных технологий, включая использование данных ДЗЗ.

Проект данной концепции, адаптированный с учетом национальной специфики, может в дальнейшем послужить прототипом при разработке типовой (модельной) концепции использования данных ДЗЗ в статистике сельского хозяйства стран СНГ.

4.2.SWOT-анализ основных видов больших данных

Ниже приведена характеристика ключевых источников больших данных на примере данных сотовых операторов (СО) о населении и данных

об использовании сельскохозяйственных угодий с использованием средств спутникового мониторинга (ДЗЗ).

Характеристика дана под углом зрения возможности использования указанных выше ключевых источников больших данных в официальной статистике с позиции НСС.

В качестве инструмента для систематизации использован принятый в международной практике SWOT-анализ: анализ сильных и слабых сторон (внутренние факторы) и возможностей и угроз (внешние факторы).

SWOT-анализ данных сотовых операторов о населении

Примеры данных:

Данные активного позиционирования (Call Detail Records - CDR).

Параметры: время, длительность, координаты базовой станции исходящего и входящего звонка/SMS, идентификатор абонента (обезличенный).

Данные пассивного позиционирования (Mobile Positioning Data - MPD).

Параметры: пассивные сигналы («rings») телефона к вышкам, время, данные о смене локации вышечного уровня, идентификатор абонента (обезличенный).

Примеры источников:

Операторы сотовой связи (например, MTS, Tele2, Beeline, другие СО). Часто используются специализированные агрегаторы данных СО.

Сильные стороны:

1. Чрезвычайно высокое пространственно-временное разрешение. Данные СО дают информацию о плотности и перемещениях населения с высочайшей пространственной (в ряде случаев, вплоть до квадрата 500x500 метров и менее) и временной (вплоть до почасовой) детализацией, в том числе в режиме, близком к реальному времени.

2. Объективность. Данные СО фиксируют фактическое поведение (местоположение и перемещения) населения, а не декларируемое, что минимизирует смещения, присущие опросам (память, социальная желательность).

3. Высокая репрезентативность. Охват до 90% взрослого населения в большинстве стран обеспечивает беспрецедентный объем данных, снижая ошибку выборки и позволяя анализировать малые группы и территории. Конечно, не все население пользуется мобильными телефонами одинаково (дети, пожилые люди, инвалиды, маргинализированные группы). Это создает систематическое смещение (bias), которое необходимо учитывать и

корректировать, но методология корректировки достаточно хорошо разработана.

4. Возможность для оперативного мониторинга и прогнозирования. Позволяют отслеживать быстро меняющиеся явления (туристские потоки, маятниковую миграцию, пиковые нагрузки, реакцию на чрезвычайные ситуации) с малым временным лагом, что критически важно для принятия управленческих решений.

5. Расширение и обогащение традиционной статистики. Открывают возможности для разработки принципиально новых и существенному уточнению существующих показателей (динамическая плотность населения, реальная мобильность, туристские потоки в индивидуальных средствах размещения), недоступных или трудно реализуемых традиционными методами.

6. Снижение нагрузки на респондентов и затрат на проведение регулярных масштабных полевых обследований.

7. Высокая методологическая проработанность. Наличие большого количества международных рекомендаций, руководств и обзоров по использованию данных сотовых операторов в официальной статистике.

8. Практическая подготовленность. Наличие большого количества реализованных пилотных проектов по использованию данных сотовых операторов в официальной статистике НСС различных стран мира.

Слабые стороны:

1. Методологическая сложность и необходимость валидации. Перевод «сырых» сетевых событий в статистические показатели требует сложных алгоритмов (идентификация дома/работы, распознавание категорий туристов и экскурсантов). Результаты требуют обязательной валидации и сравнения с традиционными данными (переписи, обследования).

2. Дублирование. Многие пользователи услуг сотовых операторов имеют (регистрируют) несколько SIM-карт на одно физическое лицо, что делает необходимым устранение соответствующего дублирования данных о людях (физических лицах). Методология корректировки достаточно хорошо разработана.

3. Зависимость от проникновения сотовой связи, данные одного оператора сотовой связи, как правило, не охватывают все рассматриваемое население, необходима интеграция данных нескольких СО.

4. Высокие требования к ИТ-инфраструктуре и кадрам. Обработка больших объемов данных требует использования мощных вычислительных

систем, специализированного программного обеспечения и высококвалифицированных кадров (data scientists, геоаналитики).

5. Зависимость от внешнего поставщика. НСС не контролируют источник данных. Получение доступа к данным СО, изменения в политике оператора, тарифах, структуре сети или формате данных могут напрямую влиять на качество данных и непрерывность статистического производства.

6. Относительно высокая начальная стоимость. Необходимы значительные инвестиции в пилотные проекты, построение ИТ-инфраструктуры, подготовку персонала.

Возможности:

1. Повышение релевантности и ценности официальной статистики. Возможность предоставлять актуальные, детализированные данные для принятия решений на всех уровнях управления (от городского планирования до национальной безопасности), укрепляя статус НСС.

2. Модернизация ключевых отраслей статистики. Существенное улучшение статистики населения, туризма, миграции, транспорта, рынка труда за счет кардинального повышения оперативности и детализации существующих и внедрения новых, более точных показателей, которые невозможно получить традиционными способами. Внесение вклада по целому ряду показателей в мониторинг ЦУР.

3. Оперативная реакция на новые вызовы. Данные СО стали важнейшим инструментом для мониторинга и оценки последствий кризисов, таких как пандемии, стихийные бедствия и массовые перемещения населения.

4. Высокая экономическая эффективность в перспективе. После внедрения в регулярное статистическое производство стоимость формирования официальных показателей может стать значительно ниже, чем при использовании традиционных обследований.

Угрозы

1. Риски нарушения конфиденциальности и этики. Это главная угроза. Данные о местоположении и перемещениях населения являются высокочувствительными. Любая утечка или возможность идентификации конкретных лиц нанесет очень серьезный ущерб репутации и доверию к НСС. Необходим жесткий контроль и использование строгих процедур обезличивания данных и обеспечения конфиденциальности и защиты информации.

2. Жесткое правовое регулирование. Действующие законы о защите персональных данных и другие могут строго ограничивать или делать

невозможным доступ НСС к данным СО. Отсутствие у многих НСС права на проведение экспериментальных работ и соответствующего финансирования. Необходимо внесение в законодательство соответствующих поправок.

3. Отсутствие стандартизации. Множественность операторов, разные форматы данных, методы обработки у разных операторов затрудняют построение согласованных временных рядов, интеграцию данных различных СО как внутри отдельных стран, так и при проведении международных сопоставлений. Необходима выработка единых стандартов и методологии интеграции данных различных сотовых операторов.

4. Волатильность рынка телекоммуникаций. Слияния и поглощения сотовых операторов, появление новых технологий могут привести к изменению условий сотрудничества или прекращению доступа к данным. Необходим постоянный мониторинг и формирование соответствующей нормативно-правовой и договорной базы.

SWOT-анализ данных спутникового мониторинга (ДЗЗ) об использовании сельскохозяйственных угодий

Примеры данных:

Спектральные снимки поверхности Земли в видимом и инфракрасном диапазоне (Landsat, Sentinel-2), данные радиолокационной съемки (Sentinel-1).

Примеры источников:

Космические агентства (NASA, ESA, Роскосмос), коммерческие провайдеры (Maxar, Planet Labs).

Сильные стороны

1. Высокая оперативность и частота обновления. Спутники обеспечивают регулярный сбор данных (например, Sentinel-1/2 - каждые 5-6 дней), что позволяет перейти от ежегодных/ежеквартальных оценок к мониторингу в режиме, близком к реальному времени (мониторинг состояния посевов, уборки урожая, засух).

2. Полное покрытие и объективность. Данные спутникового мониторинга обеспечивают сплошное покрытие территории страны, устраняя проблемы, связанные с удаленностью и труднодоступностью участков для полевых обследований.

3. Снижение нагрузки на респондентов и затрат на сбор данных. Использование ДЗЗ для стратификации и формирования выборки позволяет оптимизировать полевые работы, сокращая их объем и стоимость, особенно при проведении масштабных мероприятий, таких как сельскохозяйственная перепись.

4. **Согласованность и сопоставимость.** Методы обработки ДЗЗ стандартизированы алгоритмически, что обеспечивает высокую степень сопоставимости данных во времени и между регионами, в отличие от субъективных оценок разных интервьюеров.

5. **Возможность получения новых показателей.** ДЗЗ открывают доступ к статистике, которую невозможно или крайне дорого собрать традиционными методами (динамика деградации земель, картографирование малых водных объектов, мониторинг незаконных вырубок). Внесение вклада по целому ряду показателей в мониторинг ЦУР.

6. **Высокая методологическая проработанность.** Наличие большого количества международных рекомендаций, руководств и обзоров по использованию данных ДЗЗ в официальной статистике.

7. **Практическая подготовленность.** Наличие большого количества реализованных пилотных проектов по использованию данных ДЗЗ в официальной статистике НСС различных стран мира.

8. **Свободный доступ многих НСС к специализированным облачным вычислительным платформам.** Наличие международных решений, таких как SEPAL и Google Earth Engine, позволяет НСС проводить сложный анализ данных ДЗЗ без первоначальных крупных инвестиций в собственную высокопроизводительную ИТ-инфраструктуру.

Слабые стороны

1. **Зависимость от косвенных измерений (прокси-переменных).** ДЗЗ фиксируют физические параметры (спектральную отражательную способность), которые затем через модели связываются с целевыми показателями (урожайность, тип культуры). Точность этих моделей не всегда равна прямой регистрации.

2. **Проблемы с точностью и валидацией.** Точность автоматизированной классификации (например, видов культур) может быть недостаточной для гетерогенных ландшафтов с мелкими полями. Обязательным и ресурсоемким этапом является верификация результатов на основе данных наземных (полевых) обследований.

3. **Относительно высокие пороговые затраты на создание компетенций.** НСС необходимо создать подразделение или нанять специалистов, обладающих уникальной комбинацией знаний в области статистики, геоинформатики, агрономии и программирования.

4. **Недостаточная методологическая проработка отдельных направлений.** Не для всех статистических показателей существуют

утвержденные на международном уровне и валидированные методики расчета на основе ДЗЗ.

5. Зависимость от внешних факторов. Качество данных может ухудшаться из-за облачности (для оптических сенсоров), что требует использования радарных данных и усложняет анализ.

Возможности

1. Стратегическая поддержка со стороны международных организаций. ФАО, ЭСКАТО, Всемирный банк, Евростат активно развивают и финансируют проекты по передаче технологий (SEPAL, Open Foris), предоставляя НСС доступ к мощной ИТ-инфраструктуре и методологической базе бесплатно.

2. Повышение актуальности и глубины анализа. Интеграция ДЗЗ позволяет НСС перейти от описания прошедших событий к оперативному мониторингу и прогнозированию, значительно повышая ценность статистической информации для органов управления.

3. Экономическая эффективность в перспективе. После первоначальных инвестиций в обучение и внедрение, стоимость мониторинга на основе ДЗЗ становится значительно ниже, чем постоянное проведение масштабных полевых обследований.

4. Стимулирование развития «зеленой» статистики. ДЗЗ являются незаменимым источником для расчета показателей ЦУР, связанных с экологией (деградация земель, лесные ресурсы, водные экосистемы), что усиливает роль НСС в глобальной повестке.

5. Развитие национальных космических группировок. Создание или расширение национальных группировок спутников, осуществляющих ДЗЗ, открывает возможности для суверенного доступа к данным, независимого от международных провайдеров, и разработки специализированных продуктов, учитывающих национальную специфику.

Угрозы

1. Правовые и регуляторные барьеры. Отсутствие в нормативном правовом регулировании положений, разрешающих использование данных ДЗЗ для расчета официальных показателей, может блокировать их внедрение.

2. Быстрое устаревание технологий. Динамичное развитие методов машинного обучения и появление новых группировок спутников требуют от НСС постоянного обновления знаний и программного обеспечения, что создает риски для устойчивости процессов.

3. Риск недофинансирования на этапе внедрения. Бюджеты НСС часто жестко фиксированы, и выделение значительных единовременных средств на закупку оборудования, ПО и обучение персонала может быть затруднено.

4. Вопросы качества и репутации. Публикация статистики, рассчитанной с использованием недостаточно валидированных методов ДЗЗ, может привести к серьезным ошибкам и подорвать доверие к НСС как к объективному источнику информации.

5. Волатильность стандартов и зависимость от сложившихся экосистем данных. Постоянное изменение форматов данных, алгоритмов и интерфейсов API облачных платформ (таких как Google Earth Engine) создает операционные риски. НСС рискует оказаться в зависимости от внешней ИТ-инфраструктуры, контроль над которой она не осуществляет.

6. Недостаточная развитость национальной ИТ-инфраструктуры для больших данных. Отсутствие в стране собственных мощных центров обработки данных и хранилищ для работы с петабайтами данных ДЗЗ делает НСС заложником международных облачных решений со всеми сопутствующими рисками (безопасность данных, стоимость в будущем, юридические ограничения).

Общие выводы

В целом приведенный выше анализ использования данных сотовых операторов и данных ДЗЗ в официальной статистике, наряду с активной методологической и практической работой в этих областях международных организаций и НСС, позволяют рассматривать их как наиболее перспективные направления использования больших данных в официальной статистике.

5. Источники и виды больших данных

Определение больших данных

В международной практике сформировалось общепризнанное базовое определение больших данных и их ключевых свойств, которое было зафиксировано в международном стандарте **ISO/IEC 20546:2019 (E) «Информационные технологии. Большие данные. Обзор и словарь»**.

Цитата:

«3.1.2 Большие данные (big data): Большие массивы данных, отличающиеся главным образом такими характеристиками, как объем (volume), разнообразие (variety), скорость обработки (velocity) и/или вариативность (variability), которые требуют использования технологии

масштабирования для эффективного хранения, обработки, управления и анализа...

3.1.13 вариативность данных (data variability): Изменения в скорости передачи, формате или структуре, семантике или качестве массива данных...

3.1.14 разнообразие данных (data variety): Диапазон форматов, логических моделей, временных шкал и семантики массива данных...

3.1.15 скорость обработки данных (data velocity): Скорость потока, с которой данные создаются, передаются, сохраняются, анализируются или визуализируются.

3.1.16 достоверность данных (data veracity): Полнота и/или точность данных...

3.1.17 изменчивость данных (data volatility): Характеристика данных, относящаяся к скорости их изменения с течением времени...

3.1.18 объем данных (data volume): Количественная характеристика данных, влияющая на выбор ресурсов для вычислений и хранения, а также на управление данными в процессе обработки...

4.2.1 Объем данных. Объем данных представляет собой определенное количество данных, доступных для анализа с целью извлечения полезной информации. Одним из основных факторов развития технологий обработки больших данных стал рост объемов данных, как следствие интернет-активности.

4.2.2 Скорость обработки данных. Скорость обработки данных – это скорость потока создания, хранения, анализа и визуализации данных. Скорость обработки больших данных означает необходимость обработки большого количества данных за короткий промежуток времени. В качестве примеров работы с данными с высокой скоростью обработки обычно приводят средства обработки потоковых данных.

4.2.3 Разнообразие данных. Свойство разнообразия данных отражает необходимость анализа данных разных типов из различных предметных областей. Как правило, проблема разнообразия данных решалась посредством их преобразования или проведения предварительного анализа с целью выявления свойств, позволяющих объединить их с другими данными. Более широкий диапазон форматов данных, логических моделей, временных шкал и семантики, которые предпочтительно использовать при аналитике, усложняет интеграцию разнообразных данных. В качестве средства, способствующего интеграции, все чаще используются метаданные. Одним из результатов влияния разнообразия на большие данные является необходимость представления семантики данных в машиночитаемом виде.

4.2.4 Вариативность данных. Вариативность данных означает изменения в скорости передачи данных, их формате/структуре, семантике и/или качестве, которые влияют на поддерживаемое приложение, аналитику или проблему. Влияние вариативности может заключаться в необходимости проведения реорганизации архитектур, интерфейсов, методов обработки/алгоритмов, интеграции/слияния, хранения, применимости или использования данных. В дополнение следует отметить, что вариативность объемов данных подразумевает необходимость увеличения или уменьшения виртуализированных ресурсов для эффективного управления дополнительной нагрузкой на обработку»¹.

Перечисленные выше ключевые свойства больших данных играют решающую роль в определении как основных методологических подходов, так и информационно-технологических (далее – ИТ), организационно-управленческих и иных решений по использованию больших данных в официальной статистике.

Приведенные выше базовые определения больших данных и их свойств вошли также в Национальный стандарт Российской Федерации - ГОСТ Р ИСО/МЭК 20546-2021 «Информационные технологии. Большие данные. Обзор и словарь» (ISO/IEC 20546:2019, IDT), введенный в действие с 1 ноября 2021 года, а также в аналогичные национальные стандарты: СТБ ИСО/МЭК 20546-2020 Республики Беларусь; СТ РК ISO/IEC 20546-2022 Республики Казахстан; O'zMSt ISO/IEC 20546:2025 (ISO/IEC 20546:2019, IDT) Узбекистана. Представляется целесообразным в рамках Концепции также основываться на данном международном определении.

Необходимо отметить, что в России, Беларуси, Казахстане и Узбекистане приняты также более широкие пакеты национальных стандартов в области больших данных, искусственного интеллекта, качества данных и кибербезопасности, соответствующие международным стандартам ISO/IEC, а также профессиональные и образовательные стандарты в соответствующих областях (в том числе в рамках более широких профессий и образовательных дисциплин).

Виды больших данных

Единая, универсальная и строгая классификация, аналогичная, например, ОКВЭД, для больших данных в международной практике

¹ ISO/IEC 20546:2019 (E) «Информационные технологии. Большие данные. Обзор и словарь».

отсутствует. При этом международными организациями разработан ряд типологий, основанных на источниках и природе больших данных.

Подход ООН

ЕЭК ООН классифицирует большие данные по источникам их возникновения. Эта типология является наиболее общей и широко признанной в международном статистическом сообществе:

1. Данные, генерируемые людьми (Human-Sourced). Данные, являющиеся побочным продуктом человеческой деятельности в цифровой среде.

2. Данные, генерируемые процессами (Process-Mediated). Данные о транзакциях и событиях, генерируемые в различных административных и бизнес ИТ системах.

3. Данные, генерируемые устройствами (Device-Mediated). Автоматически генерируемые данные от датчиков и устройств, работающих без прямого вмешательства человека.

Вместе с тем данная типология, в силу своей максимальной обобщенности, является мало применимой на практике.

Перечень основных видов больших данных в рамках приведенной выше верхнеуровневой классификации видов больших данных по источникам их возникновения ЕЭК ООН²:

«1. Социальные сети (информация, генерируемая людьми)

1100. Социальные сети: Facebook, Twitter и др.

1200. Блоги и комментарии

1300. Личные документы

1400. Изображения: Instagram, Flickr, Picasa и др.

1500. Видео: Youtube и др.

1600. Интернет-запросы

1700. Контент мобильных данных: текстовые сообщения

1800. Карты, составляемые пользователями

1900. Электронная почта

2. Традиционные бизнес-системы (данные, опосредованные процессами)

21. Данные, производимые государственными органами

2110. Медицинские записи

22. Данные, производимые бизнесом

² Classification of Types of Big Data. Statistics Division UNDESA. Meeting of Expert Group on International Statistical Classifications. New-York, 19-22 May 2015.

- 2210. Коммерческие транзакции
- 2220. Банковские/биржевые записи
- 2230. Электронная коммерция
- 2240. Кредитные карты

3. Интернет вещей (машинно-генерируемые данные)

- 31. Данные датчиков
 - 311. Стационарные датчики
 - 3111. Домашняя автоматизация
 - 3112. Погодные датчики/датчики загрязнения
 - 3113. Дорожные датчики/веб-камеры
 - 3114. Научные датчики
 - 3115. Видео/изображения для безопасности/наблюдения
 - 312. Мобильные датчики (отслеживание)
 - 3121. Местоположение мобильного телефона
 - 3122. Автотранспорт
 - 3123. Спутниковые снимки
- 32. Данные из компьютерных систем
 - 3210. Журналы (логи)
 - 3220. Веб-журналы»

В дальнейшем в перечень рассматриваемых видов/источников больших данных ЕЭК ООН были также включены данные Системы автоматической идентификации судов (далее - АИС), по своему типу относящиеся к классификационной подгруппе «312. Мобильные датчики (отслеживание)».

Подход Евростата

В отличие от ЕЭК ООН Евростат фокусируется на прагматичном (прикладном) подходе, основанном на выделении конкретных, наиболее перспективных с точки зрения официальной статистики источников больших данных:

1. Данные мобильных телефонов (Mobile Phone Data). Данные о звонках (CDR) и перемещениях (MPD) абонентов. Сферы применения: статистика населения, туризма, миграции, городское планирование, мониторинг ЦУР.

2. Данные сканирования (Scanner Data). Данные о продажах розничных сетей (сканирование штрих и QR кодов). Сферы применения:

индексы потребительских цен (ИПЦ), анализ потребительского поведения, статистика розничной торговли.

3. Спутниковые снимки (Satellite Images). Данные дистанционного зондирования Земли (далее – ДЗЗ). Сферы применения: сельскохозяйственная статистика, природопользование, экологический мониторинг, мониторинг ЦУР и др.

4. Веб-данные (Web Data). Данные, собираемые с веб-сайтов. Сферы применения: статистика рынка труда (вакансии), статистика розничных цен Интернет-магазинов, транспортных компаний, цен на жилье и др.

5. Данные социальных сетей (Social Media Data). Данные различных социальных сервисов Интернет-платформ (текстовые сообщения, аудио, фото, видео материалы, геометки). Сферы применения: анализ потребительских настроений и намерений, потребительского спроса и др.

6. Данные датчиков (Internet of Things - IoT). Данные дорожных камер видеонаблюдения, GPS-треки общественного транспорта, данные АИС о передвижении морских судов, данные датчиков слежения за состоянием окружающей среды. Сферы применения: статистика транспорта, туризма, мобильности населения, экологии, мониторинг ЦУР, городское планирование и др.

Приведенная выше группировка видов больших данных Евростата отражает основные направления работ по использованию больших данных в официальной статистике, что позволяет обеспечить их практическую организацию и соответствующую концентрацию ресурсов НСС в рамках Европейской статистической системы (далее – ЕСС).

Необходимо отметить, что аналогичный практический подход был также реализован в ООН при формировании тематических рабочих групп в рамках Комитета экспертов по большим данным и науке о данных для официальной статистики ООН.

6. Подходы к использованию больших данных

6.1. Методологические подходы

В области методологических подходов на международном уровне преобладающим стал переход от подготовки общих методологических рекомендаций и обзоров к разработке практических руководств по

использованию конкретных видов больших данных при формировании официальной статистики.

При этом первостепенное внимание уделяется:

- адаптации к использованию больших данных разработанных ранее типовых моделей, руководств и рекомендаций;
- стандартизации, обеспечению интероперабельности как различных видов и источников больших данных между собой, так и с данными из традиционных, в том числе административных источников;
- использованию науки о данных и искусственного интеллекта.

В целом большие данные представляют собой один из возможных источников данных для формирования официальной статистики, обладающий набором специфических характеристик, рассмотренных в предыдущих разделах.

Главными методологическими принципами при использовании больших данных, как и любых других источников данных, для формирования официальной статистики, являются:

- соблюдение основополагающих принципов официальной статистики, принятых Резолюцией Генеральной Ассамблеи ООН 68/261 29 января 2014 года (далее – ОПОС);
- соблюдение требований и подходов к обеспечению качества официальной статистики, сформулированных в руководстве ООН «National Quality Assurance Framework Manual for Official Statistics», версия 2.0, Нью-Йорк, 2019;
- соблюдение требований и подходов к обеспечению конфиденциальности при формировании официальной статистической информации, сформулированных в руководстве ООН «Guide on Privacy-Enhancing Technologies for Official Statistics», Нью-Йорк, 2023;
- реализация основных методологических подходов в рамках Типовой модели производства статистической информации ООН, версия 5.2, Нью-Йорк, 2025 (ТМПСИ 5.2);
- а также соответствие общим подходам, сформулированным в руководстве ООН «Handbook on Management and Organization of National Statistical Systems», 4-е издание, версия 2022/А, Нью-Йорк, 2022.

Основополагающие принципы официальной статистики (ОПОС)

В соответствии с ОПОС:

- официальные статистические данные должны «удовлетворять требованиям практической полезности» (Принцип 1 - значимость);
- при этом НСС должны:

- в соответствии с «научными принципами» принимать решения «о методах и процедурах сбора, обработки, хранения и представления статистических данных» (Принцип 2 – профессиональные стандарты, научные принципы);

- «представлять информацию в соответствии с научными стандартами в отношении источников статистических данных, статистических методов и процедур» (Принцип 3 – подотчетность и транспарентность);

- «выбирать источник с учетом соображений качества, оперативности, издержек и бремени, ложащегося на респондентов» (Принцип 5 – источники официальной статистики);

- обеспечивать «строго конфиденциальный характер» индивидуальных данных и использовать их «исключительно в статистических целях» (Принцип 6 – конфиденциальность);

- использовать «международные концепции, классификации и методы» (Принцип 9 – использование международных стандартов).

Оценка качества официальных статистических данных при использовании больших данных

Традиционная основа:

ООН: «National Quality Assurance Frameworks (NQAF) Manual», версия 2.0, 2019 г. (NQAF 2.0).

Евростат: «Quality Assurance Framework of the European Statistical System (ESS QAF)», версия 2.0 (2022 г.).

Ключевое изменение: Эти рамочные руководства были созданы для данных, формируемых в контролируемых процессах (обследования, регистры).

В целях учета специфики оценки качества при использовании больших данных ООН был разработан соответствующий модуль в рамках NQAF 2.0 «Module for Quality Assurance when using Administrative and Other Data Sources to Produce Official Statistics», 2025 (подготовлен Subgroup on Administrative and Other Data Sources of the UN Expert Group on National Quality Assurance Frameworks/EG-NQAF).

Характеристика специфики оценки по компонентам качества

Трансформация основных классических компонент качества в контексте больших данных:

Точность и надежность (Accuracy and Reliability)

Специфика больших данных:

Смещение (Bias): Из случайной ошибки превращается в систематическое смещение. Данные СО не включают детей и стариков,

данные соцсетей смещены в сторону молодежи. Задача — не устранить, а измерить и скорректировать это смещение.

Погрешность: Источником становится не только выборка, но и алгоритм обработки. Точность модели классификации (F1-score) становится новой мерой статистической погрешности.

Требования руководств: ESS QAF 2.0 (Раздел 3.3) требует оценки и документирования всех источников погрешности. Для больших данных это означает обязательное вычисление и публикацию метрик качества алгоритмов (precision, recall) и оценок смещения.

Своевременность и пунктуальность (Timeliness and Punctuality)

Специфика больших данных: Главное преимущество. Позволяет перейти к near-real-time статистике. Однако возникает компромисс «скорость или точность»: первая, быстрая оценка может быть менее точной.

Требования руководств: NQAF 2.0 (п. 2.4.1) требует оптимизации временного лага. Для больших данных это означает создание многоуровневой системы распространения: сначала быстрые оперативные индикаторы, затем проверенные и откалиброванные данные.

Сопоставимость (Comparability)

Специфика больших данных: Частые «разломы сопоставимости» при смене алгоритма или источника данных. Обновление модели ML может сделать временные ряды несопоставимыми.

Требования руководств: ESS QAF 2.0 (Раздел 3.5) требует предупреждения пользователей о разрывах в рядах. Для больших данных это означает ведение параллельных рядов при обновлении методологии и четкое версионирование алгоритмов.

Когерентность (Coherence)

Специфика больших данных: Когерентность не достигается сама собой, а является результатом сложной интеграции. Например, данные СО о населении должны быть согласованы с данными переписи и регистров.

Требования руководств: NQAF 2.0 (п. 2.6.1) требует обеспечения согласованности между различными статистическими продуктами. Для больших данных это требует разработки методов калибровки и статистического взвешивания для согласования с традиционными источниками.

Новые критические компоненты качества больших данных

Помимо адаптации старых, появляются новые компоненты, без которых обеспечение качества невозможно:

Надежность источника и прозрачность методологии (Source Credibility & Methodological Transparency)

Описание: Источник данных, надежность источника (например, оператора связи или платформы соцсетей), описание использованного для получения результата алгоритма машинного обучения (далее – ML) и процесса получения результата.

Отражение в руководствах: ESS QAF 2.0 прямо указывает на необходимость прозрачности методов и управления метаданными. Это требует «Методологических паспортов», описывающих весь конвейер данных, а не только финальные результаты.

Воспроизводимость и стабильность алгоритма (Reproducibility & Algorithmic Stability)

Описание: Приведет ли повторный запуск того же алгоритма с теми же данными к тем же результатам? Не «сойдет ли алгоритм с ума» при поступлении новых данных?

Отражение в руководствах: Это прямое требование принципа научности основополагающих принципов официальной статистики ООН. Реализуется через версионирование кода, контейнеризацию (Docker) и сквозное документирование всех параметров модели.

Этичность и конфиденциальность (Ethics & Privacy)

Описание: Качество статистики, полученной с нарушением этических норм и прав граждан, является «псевдокачеством». Конфиденциальность — неотъемлемая часть качества.

Отражение в руководствах: NQAF 2.0 (Принцип 8) и ESS QAF 2.0 содержат строгие требования по защите данных. Для больших данных это означает использование дифференциальной приватности, синтетических данных и методов безопасной аналитики (Secure Multi-Party Computation).

Руководство ООН по технологиям обеспечения конфиденциальности для официальной статистики (UN «Guide on Privacy-Enhancing Technologies for Official Statistics»)

Базовые требования, подходы и технологии обеспечения конфиденциальности при формировании официальной статистической информации с использованием больших данных подробно рассмотрены в руководстве ООН «Guide on Privacy-Enhancing Technologies for Official Statistics», Нью-Йорк, 2023 (PETs Guide).

Типовая модель производства статистической информации ООН

Реализация всех основных методологических подходов (ОПОС, NQAF, PETs и др.) в отношении использования больших данных при формировании

официальной статистики в соответствии с международными рекомендациями должно осуществляться в рамках Типовой модели производства статистической информации ООН, версия 5.2, Нью-Йорк, 2025 (ТМПСИ 5.2).

Это подтверждается также общими подходами и рекомендациями, сформулированными в руководстве ООН «Handbook on Management and Organization of National Statistical Systems», 4-е издание, версия 2022/А, Нью-Йорк, 2022.

Отражение специфики использования больших данных в рамках ТМПСИ 5.2:

Этап 1. Определение потребностей

1. Определение «практической полезности» использования больших данных, что означает необходимость выявления общественной (управленческой) потребности:
 - a. в статистическом изучении нового явления (новых аспектов ранее изучавшегося явления) или
 - b. в получении более оперативной и/или более детализированной статистической информации о ранее изучавшемся явлении или
 - c. в повышении качества и/или снижении стоимости ранее формировавшейся статистической информации о том или ином явлении.

Результатом является определение целей и задач планируемого статистического исследования (объекта и предмета статистического наблюдения).

Этап 2. Проектирование

2. Определение необходимой и достаточной для достижения поставленных целей системы статистических показателей и их основных параметров (периодичность, разрезность, компоненты, единицы измерения и др.).
3. Обоснование выбора конкретного вида больших данных для разработки сформированной системы статистических показателей, включая:
 - a. отсутствие правовых ограничений;
 - b. стабильность выбранного источника больших данных;
 - c. доступ к источнику соответствующего вида больших данных или способ получения данных от выбранного источника;
 - d. репрезентативность и качество данных выбранного источника;
 - e. уровень методологической проработанности и практического опыта использования выбранного источника (наличие международных и/или национальных рекомендаций, стандартов, руководств, обзоров лучших практик);

- f. возможность интеграции с большими данными из других источников и с данными из традиционных источников, в том числе из административных источников;
 - g. экономическую эффективность использования выбранного источника;
 - h. наличие альтернативных источников, сравнительный анализ, в том числе нагрузки на респондентов;
 - i. наличие необходимых финансовых ресурсов, ИТ-инфраструктуры, программного обеспечения, компетенций.
4. Разработка методов интеграции данных из различных источников больших данных (при необходимости).

Этап 6. Анализ (после получения и обработки больших данных)

5. Валидацию рассчитанных с использованием больших данных показателей официальной статистики, включая их сравнительный анализ с показателями на основе данных традиционных источников.

Этап 8. Оценка

6. Оценка качества фактически сформированных с использованием больших данных официальных показателей статистики.
7. Оценка фактической экономической эффективности использования выбранного источника, оптимальности спроектированной системы статистических показателей и процессов их формирования, реальной нагрузки на респондентов.

Принятие решения о внедрении процесса в регулярное статистическое производство или, при необходимости, принятие решения о корректировке разработанных ранее методологических подходов.

Общий вывод:

Использование больших данных не отменяет существующие оценки качества, включая конфиденциальность, но требует их серьезного углубления. НСС должны научиться измерять и документировать не только погрешность измерения, но и погрешность алгоритма, смещение источника и стабильность ИТ-конвейера. Качество в условиях использования больших данных — это качество не только результата, но и всего процесса, от источника до алгоритма.

Наука о данных и искусственный интеллект

Важным направлением развития как в сфере методологических подходов, так и в международной статистической практике в последние годы стал акцент на науку о данных и на использование искусственного интеллекта (далее – ИИ).

Вопросы активного использования ИИ во всех сферах экономики и общественной жизни в последнее десятилетие стоят в центре внимания всего мирового сообщества и национальных правительств. В большинстве стран, в том числе в большинстве стран СНГ, приняты и активно реализуются (в том числе в рамках стратегий цифровизации/информатизации) национальные стратегии (концепции, программы) развития ИИ.

Соответствующие задачи стоят и перед международным сообществом официальных статистиков. Использование технологий ИИ неразрывно связано с развитием науки о данных и использованием больших данных в официальной статистике. Именно всестороннее использование ИИ, науки о данных и больших данных в официальной статистике легло в основу нового расширенного мандата Комитета экспертов по использованию больших данных и обработке и анализу данных для целей официальной статистики, одобренного на 56-й сессии Статистической комиссии ООН в марте 2025 года, и положено в основу Доклада Комитета экспертов по использованию больших данных и обработке и анализу больших данных для целей официальной статистики на 57-й сессии Статистической комиссии ООН в марте 2026 года.

В этом направлении работают также Сеть лидеров в области науки о данных, созданная Статистической комиссией ООН в 2022 году и Глобальная платформа ООН по большим данным для официальной статистики, реализующие совместно с Евростатом и НСС целый ряд инициатив в области использования ИИ и науки о данных в официальной статистике. Широкое развитие получили также сетевые формы организации работы экспертных и научных сообществ и международного сотрудничества с участием НСС и международных организаций.

6.2. Организационные подходы

На международном уровне еще в 2013 году в ООН был создан важнейший организационный механизм, обеспечивший координацию работ по использованию больших данных в официальной статистике. В настоящее время в рамках Комитета экспертов по использованию больших данных и обработке и анализу больших данных для целей официальной статистики ООН³ созданы и активно работают девять ТРГ по следующим наиболее значимым направлениям:

³ Ранее именовался: в 2013-2014 году - Глобальная рабочая группа по большим данным для официальной статистики ООН, в 2014-2021 году - Группа экспертов высокого уровня по большим данным и

- данные сотовых операторов;
- данные дистанционного зондирования Земли;
- данные сканирования;
- большие данные для мониторинга ЦУР;
- данные Автоматизированных систем идентификации судов;
- сельская доступность (предполагается объединение с ЦУР);
- доступ к данным в частной собственности;
- технологии повышения конфиденциальности данных;
- обучение, компетенции, повышение статистического потенциала.

Перечисленные выше ТРГ по своей сути являются глобальными центрами компетенций и осуществляют на уровне ООН координацию работ и в сотрудничестве с другими организациями разработку международных рекомендаций и руководств по соответствующим направлениям.

Важную роль в этой работе играют также отраслевые и региональные комиссии и другие организации в системе ООН (ЕЭК ООН, ЭСКАТО, ФАО и другие).

Аналогичная работа ведется в рамках Европейской статистической системы (далее – ЕСС).

На национальном уровне НСС в партнерстве с международными организациями, бизнесом и наукой созданы и успешно функционируют многочисленные центры компетенций, центры экспериментальной статистики (экспериментальные «песочницы»), учебные центры в области использования больших данных в официальной статистике.

Обобщение международного опыта позволяет сформулировать предложения по подходу к формированию стартовой и целевой моделей организационного механизма реализации Концепции.

Стартовая модель

На начальном этапе представляется возможным формирование при Статкомитете СНГ:

- Консультационного комитета по использованию больших данных в официальной статистике стран СНГ на уровне заместителей руководителей НСС стран СНГ;

- формирование экспертной комиссии (рабочей группы) по использованию данных сотовых операторов в официальной статистике.

При этом:

- ключевые решения и документы по использованию больших данных в официальной статистике будут обсуждаться и утверждаться на Совете руководителей НСС стран СНГ и/или Статкомитетом СНГ (в зависимости от характера соответствующих документов);

- реализация пилотных проектов будет осуществляться исполнителями по соответствующим направлениям работ с большими данными на конкурсной основе.

Основной функционал Консультационного комитета:

- рассмотрение планов работы Статкомитета СНГ и проектов стратегических документов в области использования больших данных в официальной статистике;

- формирование экспертных комиссий (рабочих групп) по направлениям работ по использованию больших данных (на первом этапе – данных сотовых операторов) в официальной статистике.

Основной функционал экспертных комиссий (рабочих групп):

- разработка планов работы Статкомитета СНГ в области использования больших данных в официальной статистике;

- предварительное рассмотрение проектов технических заданий и отчетов о выполнении работ (услуг) в сфере использования больших данных в официальной статистике, выполняемых на конкурсной основе по договорам (контрактам) с Статкомитетом СНГ.

Целевая модель

Целевая модель может быть сформирована на последующих этапах реализации Концепции (в случае принятия соответствующего решения и наличия необходимого финансирования) и включать дополнительно создание международного центра компетенций по использованию больших данных в официальной статистике при Статкомитете СНГ.

В настоящее время процесс формирования центров компетенций в области использования больших данных в официальной статистике в странах СНГ находится в начальной стадии.

Так, например, в России функционирует большое количество разнообразных государственных и частных научно-исследовательских, экспертных, учебных, бизнес и иных центров (самостоятельных компаний, подразделений различных учреждений и организаций и т.п.), обладающих достаточно высокой компетентностью в различных областях использования больших данных. Некоторые из них имеют также положительный опыт сотрудничества с органами государственной статистики в лице Росстата.

Росстатом на базе НИИ статистики Росстата в настоящее время осуществляется формирование национального центра компетенций по использованию больших данных в официальной статистике (с учетом инициатив Минцифры России по созданию общенациональной платформы по работе с большими данными).

Задачи по формированию национальных центров компетенций стоят также перед НСС других стран СНГ в целях сохранения и упрочения лидирующей роли НСС в общей системе управления данными своих стран.

6.3. Инфраструктура и кадры

ИТ-инфраструктура

На международном уровне происходит активное формирование и развитие как национальной, так и наднациональной ИТ-инфраструктуры для работы с большими данными.

В рамках ООН ключевым инфраструктурным проектом является создание и развитие Глобальной платформы ООН по большим данным для официальной статистики (2017) (далее – Глобальная платформа) и формирование сети Региональных центров Глобальной платформы (2019-2022) (далее – Региональные центры).

Под эгидой Глобальной платформы (штаб-квартира в Великобритании) к настоящему времени созданы 5 Региональных центров в Китае, Бразилии, Индонезии, Объединенных Арабских Эмиратах и Руанде. В 2024 году Региональный центр в Китае был преобразован в Глобальный (секторальный)⁴ центр, еще один Глобальный (секторальный) центр создан в Испании.

Глобальная платформа предоставляет своим участникам ИТ-ресурсы на основе облачных технологий, всестороннюю методическую, алгоритмическую, консультационную и учебную поддержку. Аналогичные функции в региональном масштабе выполняют ее Региональные центры.

Наряду с Глобальной платформой, серьезные инициативы по развитию наднациональной ИТ-инфраструктуры реализуются в Европейском союзе в рамках проекта ESSnet, в соответствии с которым создана и развивается система общеевропейских информационных пространств, исследовательских инициатив и центров компетенций в области использования больших

⁴ Терминология в этой области на международном уровне еще окончательно не устоялась, поэтому используются обе формулировки: как Глобальный центр, так и Секторальный центр.

данных, науки о данных и искусственного интеллекта в официальной статистике.

На национальном уровне в современной международной практике сложился общепризнанный подход к пониманию наличия современной, безопасной, гибкой и масштабируемой ИТ-инфраструктуры как стратегической, а не технической, необходимости для НСС, обеспечивающей удовлетворение быстро растущих потребностей в более своевременной и детализированной официальной статистике на основе использования больших данных, огромные объемы, высокая скорость и изменчивость которых являются их базовыми характеристиками.

В современных условиях в основе формирования требований к развитию ИТ-инфраструктуры НСС в соответствии с международными рекомендациями лежат:

1. Признание ИТ-инфраструктуры стратегическим активом. Инвестиции в ИТ-инфраструктуру обеспечивают способность НСС оставаться конкурентоспособными в условиях быстрых технологических изменений.
2. Приоритетность масштабируемости, гибкости, безопасности и совместимости. Принятие модульных, облачных решений и общих стандартов гарантирует, что ИТ-инфраструктура может эффективно адаптироваться и масштабироваться. В ряде случаев рекомендуется использование технологий с открытым исходным кодом.
3. Разработка и принятие стратегий и дорожных карт модернизации ИТ-инфраструктуры, соответствующих их институциональной готовности, имеющимся ресурсам и национальным приоритетам. Модернизация ИТ-инфраструктуры должна носить непрерывный характер и быть тесно увязана с наращиванием соответствующих компетенций НСС.

В целом в соответствии с рекомендациями ООН⁵ при выработке стратегий (политик) развития ИТ НСС рекомендуется различать: инфраструктуру данных, архитектуру данных и ИТ-инфраструктуру.

Архитектура данных относится к наиболее высокому уровню, который определяет стандарты данных, структуры, управление метаданными и правила интероперабельности между системами, определяет, как данные

⁵ Big Data for Official Statistics: Strategic Considerations and Recommendations on Data Infrastructure and Governance (ESCAP, SD/WP/118/July 2025); Technical Document to Modernize IT architecture in NSO/NSS using open-source technology stack (UN Data4Now, 6/18/2025).

должны храниться и интегрироваться, обеспечивая согласованную работу различных частей инфраструктуры.

Инфраструктура данных относится к интегрированной системе технологий, людей, процессов, политик и принципов управления, которая поддерживает весь жизненный цикл данных, от сбора и хранения до анализа и распространения и охватывает физические компоненты (серверы, системы хранения, сетевое оборудование, облачные платформы), цифровые инструменты (включая базы данных, конвейеры данных, аналитическое программное обеспечение) и механизмы управления, предназначенные для обеспечения качества данных и конфиденциальности.

ИТ-инфраструктура включает физические и виртуальные аппаратные и программные системы, необходимые для поддержки операций с данными. В соответствии с рекомендациями ООН типовые технические компоненты современной архитектуры данных для работы с большими данными включают:

- конвейеры данных (Data Pipelines) - автоматизируют поток данных от источника к хранилищу, преобразованию и анализу;
- облачное хранилище (Cloud Storage) - обеспечивает гибкое и масштабируемое хранилище в размещаемой среде;
- API - обеспечивают безопасный, структурированный обмен данными между системами;
- модели ИИ/МО (AI/ML Models) - поддерживают аналитику, классификацию, прогнозирование и обнаружение трендов;
- потоковая передача данных (Data Streaming) - обрабатывает данные в реальном времени по мере их генерации;
- Kubernetes - управляет развертыванием и масштабированием приложений в контейнерах;
- облачные вычисления (Cloud Computing) - обеспечивают доступ по требованию к вычислительной мощности и сервисам;
- аналитика в реальном времени (Real-time Analytics) - предоставляет результаты анализа по мере поступления данных.

Основные компоненты инфраструктуры данных для НСС в соответствии с структурой жизненного цикла данных (базовыми этапами ТМПСИ) включают:

Сбор данных: Получение данных из традиционных источников, административных данных, а также геопространственных данных, данных ДЗЗ, сотовых операторов и др. Указанные источники могут быть доступны

различными способами, включая API, прямые подключения к базам данных, ручную передачу или протоколы передачи файлов (FTP).

Потребности в инфраструктуре на этом этапе включают безопасные механизмы доступа, надежные системы для импорта данных в различных форматах и с разной частотой, а также соответствующие механизмы управления.

Прием данных (Data Ingestion): После получения данные должны быть загружены в ИТ-системы НСС. На этапе приема применяются такие процессы, как: Extract-Transform-Load (ETL), Extract-Load-Transform (ELT), пакетный и потоковый прием, а также контроль версионности. На данном этапе также должны применяться базовые правила проверки для обеспечения того, что данные отформатированы, структурированы и готовы к внутреннему использованию.

Подготовка данных (Data Preparation): На этом этапе принятые данные подготавливаются для анализа - очищаются, интегрируются, гармонизируются и хранятся в структурированных средах, таких как: хранилища данных, озера данных или системы управления метаданными (MDM). Здесь также проводятся такие процессы, как анонимизация и др. Данный этап является критическим, особенно при объединении нескольких источников или работе с конфиденциальными данными. Это требует безопасной и масштабируемой инфраструктуры, а также процессов, которые обеспечивают прослеживаемость, воспроизводимость и соответствие политикам управления данными.

Вычисления с данными (Data Computation): Аналитическое ядро, включающее как традиционные методы, так и машинное обучение, обработку изображений и другие методы науки о данных. Обычно используемые инструменты: R, Python, QGIS, SPSS, Jupyter notebooks и др. Вычислительная инфраструктура может включать облачные платформы или локальные системы, каждая из которых имеет свои преимущества с точки зрения масштабируемости, стоимости и соответствия требованиям. Такие системы как Hadoop или Spark, могут использоваться для поддержки распределенной или параллельной обработки, особенно при работе с большими объемами данных.

Принятие решений на этом этапе должно учитывать долгосрочную устойчивость, рентабельность и соответствие национальным требованиям безопасности и защиты данных. Должны быть предусмотрены механизмы управления - контроль доступа, процедуры валидации моделей и др.

Представление данных (Data Presentation): Заключительный этап — распространение результатов. Может производиться в форме публикаций, дашбордов, порталов открытых данных, API и загружаемых наборов данных. Инфраструктура на этом этапе должна поддерживать своевременный, безопасный и удобный доступ к статистической информации для различных категорий пользователей.

«Оркестрация» (Orchestration): Ключевым сквозным элементом является «оркестрация» - комплекс систем и процедур, которые координируют рабочие процессы на протяжении всего жизненного цикла данных. Включает планирование задач, управление ресурсами, мониторинг, автоматизацию, обработку ошибок и др.

Конкретные требования к ИТ-инфраструктуре НСС на национальном уровне в высокой степени зависят от следующих факторов:

- национальной политики в области построения ИТ-инфраструктуры для работы с большими данными (создание единой национальной платформы или децентрализованная модель);

- выбора НСС перечисленных выше моделей работы с большими данными (получение исходных данных и их самостоятельная обработка – наиболее ресурсоемкий вариант, требующий развертывания полномасштабной ИТ-инфраструктуры на стороне НСС; получение доступа к данным; передача алгоритмов обработки; взаимодействие с доверенными третьими сторонами – удостоверенными агрегаторами данных);

- целей и задач НСС по использованию больших данных, видов больших данных и их источников, необходимых для реализации поставленных целей;

- наличной ИТ-инфраструктуры, финансовых ресурсов и компетенций НСС по работе с большими данными и современной ИТ-инфраструктурой.

Подготовка кадров

В современной международной практике общепризнана необходимость первостепенного внимания к постоянному повышению квалификации статистических кадров в области использования больших данных и науки о данных при формировании официальной статистики.

В этой области функционируют многочисленные международные и национальные учебные центры и программы, разработано большое число международных и национальных рекомендаций, практических руководств, учебных курсов, в том числе находящихся в открытом доступе, в первую очередь, для дистанционного обучения.

Вопросы повышения статистического потенциала НСС находятся в центре внимания всех международных организаций. НСС рекомендуется регулярно оценивать уровень компетенций своих сотрудников и обеспечивать их регулярное обучение, подготовку и переподготовку.

7. Нормативно-правовые аспекты

Наиболее перспективной на международном уровне признана модель нормативного правового регулирования, непосредственно закрепляющая на национальном уровне право НСС на прямой доступ или получение, как правило, предварительно обезличенных данных, находящихся в частной собственности, в целях формирования официальной статистики. Данная норма не должна противоречить или ограничиваться другими, уже действующими на национальном уровне нормами.

Вторым важнейшим направлением совершенствования регулирования стало нормативное правовое закрепление возможности реализации различных моделей работы НСС с большими данными (получение самих данных, получение доступа к данным, передача алгоритмов обработки и наиболее перспективная модель - взаимодействие с доверенными третьими сторонами – удостоверенными агрегаторами данных).

Важнейшее внимание в рамках нормативного правового регулирования продолжает уделяться обеспечению безопасности и конфиденциальности официальной статистики, формируемой с использованием больших данных.

В странах СНГ регулирование использования больших данных в официальной статистике осуществляется преимущественно в рамках законов об официальной (государственной) статистике (далее – законы о статистике) и законов о персональных данных (аналогичных законах).

В настоящее время все национальные законы о статистике в странах СНГ базируются на основополагающих принципах официальной статистики ООН, в соответствии с которыми обеспечивают на национальном уровне объективность, прозрачность, высокое качество, конфиденциальность, соответствие международным стандартам и другие требования к официальной статистике.

При этом в большинстве стран СНГ непосредственно в национальных законах о статистике за НСС в той или иной форме закреплён «государственный мандат» на получение (как правило, на безвозмездной основе) помимо первичных статистических данных и административных данных, также любых (иных) данных, необходимых в целях производства официальной статистической информации. В России соответствующая норма

в отношении получения иных данных в явном виде не сформулирована, в Казахстане ограничивается данными ДЗЗ.

Конкретные формы и возможности реализации указанного выше «государственного мандата» на национальном уровне зависят от выбранной модели взаимодействия НСС с поставщиками больших данных, в первую очередь, с сотовыми операторами.

При этом в большинстве стран СНГ, несмотря на наличие у НСС соответствующего «мандата», основной формой взаимодействия в настоящее время выступает «добровольное партнерство» на основе соглашений о взаимодействии (в том числе на коммерческой основе).

В России, в отсутствие указанной нормы, на федеральном уровне начато создание единой общенациональной платформы (ГосТех) для работы с большими данными (в качестве эксперимента – с данными сотовых операторов). В рамках данного проекта все СО будут обязаны предоставлять предварительно обезличенные данные пассивного позиционирования своих абонентов, а платформа предоставлять всем аккредитованным пользователям сервисы для работы с данными и обеспечивать безопасность и конфиденциальность данных. При этом Минцифры России выступает оператором платформы, Росстат - функциональным заказчиком (и главным пользователем) сервисов по статистике населения и статистике туризма.

По итогам реализации проекта в 2026-2027 годах будет принято решение о его дальнейшем развитии, что, безусловно, окажет решающее влияние на все аспекты использования данных сотовых операторов в российской официальной статистике.

Стандарты по большим данным

В странах СНГ ситуация неоднородна. В рамках Межгосударственного совета по стандартизации (МГС) страны СНГ работают над принятием единых стандартов, часто основанных на адаптации ISO. Национальные институты стандартизации стран СНГ, как правило, также ориентируются на прямые переводы и внедрение стандартов ISO/IEC.

Основное внимание уделяется не столько архитектурным стандартам, сколько техническим регламентам и законам, касающимся защиты персональных данных и информационной безопасности, что создает правовой контекст для работы с большими данными.

Стандарты по большим данным в Республике Беларусь

В Беларуси был разработан и введен в действие собственный стандарт.

Официальное название на русском: СТБ ИСО/МЭК 20546-2020 «Информационные технологии. Большие данные. Обзор и словарь»

Реквизиты: Дата введения: 2021-04-01. Является полным аналогом международного стандарта ISO/IEC 20546:2019.

Краткая характеристика: Как и его международный и российский аналоги, этот стандарт устанавливает основополагающие термины и определения, что создает единую терминологическую базу для дальнейшего развития направления в стране.

Стандарты по большим данным в России

В России работа ведется в рамках Технического комитета по стандартизации ТК 001 «Искусственный интеллект» при Росстандарте. Помимо адаптации стандартов ISO, разрабатываются также собственные стандарты.

Принятые стандарты (адаптированные международные)

1. ГОСТ Р ИСО/МЭК 20546-2021 «Информационные технологии. Большие данные. Обзор и словарь»

2. Серия ГОСТ Р, аутентичная серии ISO/IEC 20547 (Эталонная модель больших данных)

2.1. Часть 1: Фундамент и процессы

Официальное название: ГОСТ Р ИСО/МЭК 20547-1-2023 «Информационные технологии. Эталонная модель больших данных. Часть 1. Структура и процесс для эталонной архитектуры»

Реквизиты: Дата введения: 2024-07-01. Заменяет собой предварительный стандарт ПНСТ 468-2021.

2.2. Часть 2: Варианты использования

Официальное название: ГОСТ Р ИСО/МЭК 20547-2-2023 «Информационные технологии. Эталонная модель больших данных. Часть 2. Варианты использования и производные требования»

Реквизиты: Дата введения: 2024-07-01.

2.3. Часть 3: Эталонная архитектура (ядро)

Официальное название: ГОСТ Р 56888-2023/ISO/IEC 20547-3:2020 «Информационные технологии. Эталонная модель больших данных. Часть 3. Эталонная архитектура»

Реквизиты: Дата введения: 2024-01-01.

2.4. Часть 4: Безопасность и конфиденциальность

Официальное название: ГОСТ Р ИСО/МЭК 20547-4-2023 «Информационные технологии. Эталонная модель больших данных. Часть 4. Безопасность и защита конфиденциальности»

Реквизиты: Дата введения: 2024-07-01.

2.5. Часть 5: Дорожная карта стандартов

Официальное название: ГОСТ Р ИСО/МЭК 20547-5-2023 «Информационные технологии. Эталонная модель больших данных. Часть 5. Дорожная карта стандартов»

Реквизиты: Дата введения: 2024-07-01.

Другие ключевые принятые стандарты

Стандарт по процессу аналитики данных (аутентичный международному)

Официальное название: ГОСТ Р ИСО/МЭК 24668-2023 «Информационные технологии. Большие данные. Процесс анализа данных»

Реквизиты: Дата введения: 2024-07-01.

Стандарт на техническое задание (оригинальный российский)

Официальное название: ГОСТ Р 57799-2023 «Информационные технологии. Большие данные. Техническое задание на создание системы больших данных»

Реквизиты: Дата введения: 2024-07-01.

Характеристика: Это важный оригинальный российский стандарт, аналога которому в линейке ISO на данный момент нет. Он устанавливает требования к структуре и содержанию технического задания (ТЗ) на создание систем больших данных. Стандарт ориентирован на практическое применение.

Проекты стандартов (находятся в разработке)

Помимо принятых, ряд проектов национальных стандартов (ПНСТ) находятся на стадии обсуждения и утверждения:

ПНСТ 714-2023/ISO/IEC 20547-2:2020 «...Эталонная модель больших данных. Часть 2. Варианты использования и производные требования» (Проект аутентичного перевода).

ПНСТ 715-2023/ISO/IEC 20547-4:2023 «...Эталонная модель больших данных. Часть 4. Безопасность и защита конфиденциальности» (Проект аутентичного перевода).

ПНСТ 716-2023/ISO/IEC 24668:2022 «...Большие данные. Процесс анализа данных» (Проект аутентичного перевода).

Общие выводы

Пакет стандартов на базе ISO/IEC является одним из самых передовых и полных в мире и создает серьезную основу для реализации проектов НСС стран СНГ по использованию больших данных.

8. Этапы реализации

Основные этапы реализации Концепции в 2026-2028 годах (при наличии необходимого финансирования и принятия соответствующих решений):

2026 год:

- сформированы международный Консультационный комитет и экспертная комиссия (рабочая группа) по использованию больших данных в официальной статистике при Статкомитете СНГ;

- разработаны типовые ТЗ на реализацию пилотных проектов по использованию больших данных в официальной статистике (на основе данных сотовых операторов);

- реализован пилотный проект по использованию в официальной статистике данных сотовых операторов (статистика населения);

- создана информационная система для обучения специалистов НСС работе с большими данными на основе реализованных пилотных проектов;

2027 год:

- разработаны типовые ТЗ на реализацию пилотных проектов по использованию больших данных в официальной статистике (по выбранным направлениям работ);

- реализованы пилотные проекты по использованию в официальной статистике данных сотовых операторов (по выбранным направлениям);

- на основе разработанной концепции использования больших данных в официальной статистике и с учетом результатов реализации указанных выше пилотных проектов определен состав и начата разработка пакета типовых (модельных) НПА, рекомендаций, методик по использованию больших данных в официальной статистике;

- сформирован базовый учебный курс по большим данным для специалистов НСС и органов власти;

- обеспечено дистанционное обучение (не менее 20 специалистов НСС и органов власти) и наполнение базы знаний и наборов больших данных;

- проведен международный семинар по использованию больших данных в официальной статистике по итогам реализации Концепции в 2026-2027 годах.

2028 год:

- реализованы все КРІ проекта;

- сформирован международный Центр компетенций по использованию больших данных в официальной статистике при Статкомитете СНГ;

- разработаны типовые ТЗ на реализацию пилотных проектов по использованию больших данных в официальной статистике (по выбранным направлениям работ);

- реализованы пилотные проекты по использованию в официальной статистике больших данных (по выбранным направлениям работ);
- разработан пакет типовых (модельных) НПА, рекомендаций, методик использования больших данных в официальной статистике на основе результатов реализации пилотных проектов в 2026-2027 годах;
- обеспечено дистанционное обучение (не менее 30 специалистов НСС и органов власти) и наполнение базы знаний и наборов больших данных, ведение Интернет-ресурса проекта;
- проведена международная конференция по использованию больших данных в официальной статистике по итогам реализации Концепции в 2027-2028 годах.

9. Заключительные положения

Главные результаты реализации Концепции для органов власти стран СНГ, Статкомитета СНГ и НСС:

- функционирующая организационная структура (Консультационный комитет, экспертные комиссии (рабочие группы), международный Центр компетенций);
- истории успеха - результаты реализации пилотных проектов;
- сопоставимая методология, ускорение и детализация официальных статистических данных, новые показатели для бизнеса и управления;
- рабочий апробированный инструментарий по использованию больших данных в официальной статистике (пакет типовых (модельных) НПА, рекомендаций, методик, ТЗ, алгоритмов);
- готовая (работающая) ИТ-инфраструктура для обучения и работы с большими данными;
- рабочий инструментарий для обучения работе по использованию больших данных в официальной статистике;
- подготовленный современный кадровый потенциал - обученные специалисты НСС и органов власти СНГ;
- возможность масштабирования проекта на БРИКС+ и другие объединения и страны;
- высокий международный авторитет в статистическом сообществе.