



## **Европейская экономическая комиссия**

### **Конференция европейских статистиков**

**Семьдесят вторая пленарная сессия**

Женева, 20 и 21 июня 2024 года

Пункт 3 предварительной повестки дня

**Связывание данных между областями и источниками**

### **Углубленный обзор темы связывания данных между областями и источниками**

**Подготовлено Канадой**

#### *Резюме*

В настоящем углубленном обзоре, подготовленном по поручению Бюро Конференции европейских статистиков (КЕС), рассматривается роль национальных статистических управлений (НСУ) в связывании данных между областями и источниками в целях обеспечения информационных потребностей и подготовки статистических выводов по многомерным вопросам. В записке представлен обзор системного подхода к связыванию данных и обобщен опыт НСУ в этой области. В последнем разделе подводятся итоги обсуждения, проведенного Бюро КЕС на совещании в феврале 2024 года.

Конференции предлагается одобрить результаты этого углубленного обзора.



## I. Резюме

1. В феврале 2023 года Бюро Конференции европейских статистиков (КЕС) приняло решение о проведении углубленного обзора на тему «Связывание данных между областями и источниками» под руководством Канады. В настоящем обзоре поднимаются дискуссионные вопросы о путях развития достижений национальных статистических управлений (НСУ) в области связывания данных, в частности использования уже имеющихся данных, новых источников данных, а также научных инструментов и методов для повышения эффективности работы национальных статистических систем (НСС) и предоставления более качественной информации по вопросам политики и многомерным проблемам. Вопросы, находящиеся в центре внимания мировой и национальной политики, такие как пандемия COVID-19, чрезвычайная климатическая ситуация, энергетический кризис и кризис стоимости жизни, свидетельствуют о взаимосвязи экономики, общества и окружающей среды. Кроме того, очевидно, что последствия этих проблем не одинаково ощущаются в разных обществах, и для решения проблемы неравенства требуется более детальная информация. В результате этого директивные органы применяют более целостный подход к рассматриваемым вопросам, чтобы учесть взаимосвязи между различными областями. Это обуславливает спрос на такие аналитические статистические данные НСУ, которые позволяют решать эти междисциплинарные вопросы политики и предоставлять более детальную информацию, охватывающую множество областей.
2. В настоящем обзоре описывается систематический подход к связыванию данных с целью решения стратегических, аналитических и оперативных задач НСУ и различных ведомств. Обзор основан на тематических исследованиях и опросе, посвященном готовности НСУ выступать в качестве координаторов работы по связыванию данных в усложняющихся экосистемах данных. Кроме того, в обзоре уделяется внимание предыдущей работе КЕС в этой области.
3. В ходе анализа тематических исследований и результатов опроса было выявлено несколько тем. Многие касающиеся данных проблемы НСУ были решены путем связывания данных. Вместе с тем очевидно, что связывание данных может быть не только ответом на существующие проблемы (в виде мер реагирования), но и способом использования имеющихся возможностей (в виде упреждающих мер). Во-первых, связывание данных обычно используется для решения проблем, касающихся уменьшения доли ответивших респондентов, отсутствия данных и их качества. Во-вторых, связывание данных используется в качестве ответа на имеющиеся возможности в плане повышения эффективности НСС за счет снижения затрат на обследования, нагрузки по предоставлению ответов и избыточности данных в НСС. В-третьих, связывание данных — это экономически эффективный способ получения данных, характеризующихся большей частотностью и оперативностью, дезагрегированных по отдельным сегментам населения и по географическому признаку и позволяющих выявлять многомерные социальные и экономические явления, которые незаметны в одиночных источниках данных.
4. Отмечаются проблемы, связанные с созданием такой экосистемы данных, которая была бы эффективной и удовлетворяла информационные потребности всех заинтересованных сторон. В основе этих проблем лежит отсутствие организационного звена, которое бы координировало имеющиеся в НСС массивы данных и разрабатывало бы комплексную систему связанных данных. Одной из ключевых рекомендаций настоящего углубленного обзора является необходимость создания пользовательского шлюза, обеспечивающего доступ к связанным данным, а также предоставляющего услуги по работе с ними. В отсутствие координации высок риск фрагментации данных (отсутствие доступа), дублирования данных (избыточность и нагрузка по предоставлению ответов) и несогласованности данных в НСС, а также снижения способности наблюдать за взаимосвязанными явлениями.
5. Существуют также возможности задействовать источники данных, которые НСУ обычно не используют (например, [спутниковые снимки](#) и [данные сканеров](#)), чтобы расширить рамки потенциальных вариантов связывания данных. Однако такие возможности сопряжены с новыми трудностями в плане приобретения этих

источников данных, которые зачастую находятся в частной собственности, и получения общественного одобрения относительно их использования. Для более эффективного распоряжения результатами связывания различных областей можно также задействовать такие новые технологии и инструменты, как искусственный интеллект (ИИ) и машинное обучение (МО), однако необходимо учитывать присущие этим технологиям риски (этические, технологические и инфраструктурные).

## II. Введение

6. Бюро Конференции европейских статистиков (КЕС) регулярно проводит углубленный обзор избранных областей статистики. Целью такого обзора является совершенствование координации статистической деятельности в регионе Европейской экономической комиссии Организации Объединенных Наций (ЕЭК ООН), выявление пробелов или случаев дублирования в работе и рассмотрение новых вопросов. Основное внимание в рамках этих обзоров уделяется стратегическим вопросам, а также освещению стоящих перед статистическими управлениями проблем концептуального и координационного характера.

7. На своем заседании в феврале 2023 года Бюро КЕС выбрало для углубленного рассмотрения тему «Связывание данных между областями и источниками». В настоящем обзоре дается краткое описание международной статистической деятельности в выбранной области, обозначаются возможности и вызовы, а также выносятся рекомендации относительно возможных последующих действий. Настоящий обзор проводится под руководством Канады при участии Польши и других стран, которые предоставили ответы на опрос в ходе подготовки обзора.

8. Связывание данных не является чем-то новым, и НСУ занимаются этим на протяжении многих лет. Однако тип и количество наборов данных, которые связываются между собой, равно как и области, между которыми происходит связывание данных, претерпевают изменения. Это обусловлено необходимостью лучше понять сложные проблемы, требующие принятия политических мер в различных областях. Как отмечается ниже, в нескольких предыдущих углубленных обзорах КЕС обсуждались **способы** связывания данных и рассматривались различные аспекты — от технических вопросов до проблем конфиденциальности и надзора. Настоящий углубленный обзор посвящен новым вопросам: **почему** возникают потребности в связывании данных и **в чем проявляются** вызовы и возможности для перевода связывания данных на новые рубежи. В результате углубленного обзора были выявлены три области для обсуждения на совещании КЕС в феврале:

9. **Во-первых, связывание данных может быть использовано для создания новой системы подготовки НСУ официальной статистики и удовлетворения информационных потребностей.** В рамках НСС имеется огромное количество данных, и у НСУ есть возможность объединить эти данные, чтобы уменьшить зависимость от обследований при подготовке официальной статистики. Традиционные методы сбора данных начинаются с проведения обследований, после чего эти данные связываются с другими источниками, чтобы заполнить пробелы или повысить качество данных. Такой процесс сбора данных можно изменить на противоположный, начав со связывания данных, т. е. использовать подход «сначала связать, потом собрать», поскольку сначала используются уже существующие данные, а затем и лишь по мере необходимости проводятся дополняющие эти данные обследования. Это соответствует принципу «только один раз», при котором одни и те же данные собираются один раз и используются повторно в рамках НСС. Например, чтобы сократить время, затрачиваемое предприятиями на предоставление ответов на обследования, Статистическое управление Канады по мере возможности использует в обследованиях уже предоставленные предприятиями и агрокомпаниями административные данные, такие как налоговые декларации и ведомости начисления заработной платы. Кроме того, были рассмотрены другие заменяющие обследования предприятий источники данных, такие как дистанционное зондирование и прослеживаемость.

10. **Во-вторых, одним из движущих факторов связывания данных из разных областей и источников являются комплексные информационные потребности директивных органов.** Вопросы, находящиеся в центре внимания национальных и глобальных политических повесток, такие как чрезвычайная климатическая ситуация и Цели в области устойчивого развития, показывают, что сферы экономики, общества и окружающей среды взаимосвязаны, и вопросы в этих сферах не должны рассматриваться изолированно. Связывание данных необходимо для понимания взаимосвязанной природы этих проблем и масштабов их влияния. Например, недавно проведенное в Канаде исследование, основанное на связанных данных, показало, что эпидемия COVID-19 привела к изменениям в организации труда, которые повлияли на использование общественного транспорта и выбросы парниковых газов. В других проектах по связыванию данных использовались разрозненные источники данных (например, данные служб здравоохранения, коронеров и системы правосудия и данные о доходах), что позволило получить подробную и детально проработанную информацию об опиоидном кризисе в Канаде, которую невозможно было выявить в отдельных источниках данных. Кроме того, Статистическое управление Канады внедрило Индекс местных условий ведения бизнеса в реальном времени (ИМУВБ-РВ), который связывает данные Реестра предприятий с коммерческими данными («Гугл плейсиз», «Йелп фьюжн» и «Зомато») и данными о дорожном движении для получения периодической (раз в неделю или чаще) и географически детализированной (город и район) статистики с целью мониторинга деловой активности после вызванных COVID-19 экономических перебоев и экономического восстановления. Проще говоря, без связывания данных высок риск фрагментации информации или неполного понимания сложных проблем.

11. **В-третьих, рост доступности новых типов данных и технологии искусственного интеллекта открывают новые возможности для связывания данных и доступа к данным в более широких масштабах, чем когда-либо прежде.** Новые источники данных становятся все более доступными и необходимыми для решения многомерных проблем и получения данных, отсутствующих в традиционных источниках. В недавно опубликованном докладе ЕЭК ООН рассматривается вопрос о том, как можно использовать связывание данных из традиционных источников (например, обследований) с новыми источниками данных (например, данными из социальных сетей или мобильных телефонов) для получения более точных показателей миграции и трансграничной мобильности. Однако при связывании и анализе этих новых типов данных возникают серьезные технические сложности. Группа высокого уровня по модернизации официальной статистики (ГВУ-МОС) уже разработала рамки качества больших данных и рекомендации по созданию партнерств в проектах, касающихся больших данных. Существует множество способов, с помощью которых искусственный интеллект облегчает сбор данных с целью их классификации и построения безопасных и эффективных прогнозов относительно данных, а также повышения значимости анализа данных. Например, Статистическое управление Канады использовало машинное обучение (МО) для решения проблем совместимости (например, различных способов классификации данных), которые существуют среди схожих типов административных данных (например, данных коронеров) из разных юрисдикций. Еще одна область, которую изучает Статистическое управление Канады, — это связывание с сохранением конфиденциальности в случае конфиденциальных данных, при которой связывание и анализ могут проводиться на основе связанных данных в рамках федеративной модели без необходимости переносить данные из одного места в другое. Необходимо продолжить работу над тем, каким образом может быть задействован ИИ для использования больших данных в инновационных целях и получения более полной информации. Кроме того, доступ к данным такого типа и обеспечение общественного одобрения такого масштабного связывания данных, которое может нарушать неприкосновенность частной жизни, сопряжены с более серьезными сложностями в плане курирования данных.

12. НСУ по-прежнему сталкиваются с трудностями в плане связывания данных, которые варьируются от технических вопросов, касающихся осуществимости такой деятельности, до вопросов курирования, относящихся к надзору над ней и ее

общественному одобрению. Главная проблема, которая была выявлена в ходе проведения настоящего углубленного обзора, заключается в необходимости создания пользовательского шлюза для обеспечения доступа к связанным данным и услугам, основанным на связанных данных. Учитывая огромное количество поставщиков и пользователей данных в рамках НСС, возникает логистическая необходимость в департаменте или управлении, отвечающем за создание реестра активов данных в НСС и координирующем деятельность по связыванию данных на основе принципов удобства поиска, доступности, функциональной совместимости и возможности повторного использования данных (ПДФП).

13. Остальная часть настоящего углубленного обзора структурирована следующим образом. В разделе III изложены цели обзора, в котором представлен систематический подход к связыванию данных. В разделе IV представлены тематические исследования и результаты обследования НСУ, посвященного их полномочиям и опыту в сфере связывания данных между областями и источниками. В разделе V кратко описывается работа в сфере связывания данных, которую Бюро КЕС уже завершило. В разделе VI освещаются вопросы и вызовы, выявленные в ходе обзора, а в разделе VII даны рекомендации для дальнейшего обсуждения.

### III. Тематический охват рассматриваемой статистической области

14. В настоящем углубленном обзоре основное внимание уделяется роли НСУ в качестве координатора деятельности по связыванию данных в НСС. В настоящем обзоре связывание данных определяется как связывание учетных записей или компиляция данных из двух или более источников, например, связывание административных источников из разных государственных департаментов или министерств, связывание данных обследований и административных данных, или связывание геопространственных данных и данных обследований<sup>1</sup>. Связывание данных не ограничивается связыванием записей, хотя это наиболее распространенный его тип. Связывание данных может также включать в себя компиляцию информации из разных областей, моделирование и разработку системы показателей. Как будет показано на примерах стран в следующем разделе, связывание данных представляет собой обширную тему, имеющую множество аспектов, включая специфические вопросы, которые могут быть включены в повестку дня различных тематических групп, и горизонтальные вопросы, такие как терминология, разъяснительная работа, общественная приемлемость, методы и информационная работа.

15. В настоящем обзоре основное внимание уделяется причинам, по которым НСУ участвуют в связывании данных между областями и источниками, и подробно не рассматриваются вопросы, касающиеся того, как осуществляется такая деятельность. Последняя тема была рассмотрена в рамках предыдущей работы КЕС, посвященной технологическим, методологическим, управленческим, юридическим и институциональным вызовам, касающимся связывания данных. Настоящий углубленный обзор преследует три цели:

а) во-первых, дать определение систематическому подходу к связыванию данных между областями и источниками и способствовать внедрению такого подхода. Этот подход отвечает рекомендации предыдущего углубленного обзора [КЕС по интеграции данных](#) о необходимости стандартных процессов для облегчения интеграции данных. Систематический подход представляет собой формализованную

<sup>1</sup> Как отмечается в комментариях ЕЭК ООН к настоящему обзору (см. п. 10 документа ECE/CES/BUR/2024/FEB/2/Add.1), терминология является важным аспектом, поскольку в некоторых случаях одни и те же термины могут иметь разное значение в разных странах. Например, в большинстве стран термин «административные данные» относится только к данным, которые хранятся в государственном секторе по оперативным причинам, при этом в некоторых странах этот термин относится ко всем данным, которые собираются в рамках деятельности организации, включая данные, находящиеся в частном владении.

серию шагов, которые можно использовать для ведения проектов по интеграции данных;

b) во-вторых, документально зафиксировать текущие и будущие функции НСУ как координаторов связывания данных в рамках НСС и как поставщиков услуг, основанных на связанных данных. КЕС попросил провести в рамках этого углубленного обзора опрос о деятельности НСУ по связыванию данных. Полученные в ходе опроса результаты включают примеры использования связывания данных для удовлетворения политических, аналитических и оперативных потребностей, а также извлеченные из этой работы уроки;

c) в-третьих, повысить осведомленность о соответствующих результатах недавно проведенной КЕС работы в сфере связывания данных в больших масштабах, такой как [углубленный обзор интеграции данных](#) (2017 год), [углубленный обзор этики данных](#) (2022 год), доклад о [курировании данных](#) (2023 год) и система управления совместимостью данных (в процессе подготовки). В настоящем углубленном обзоре кратко рассматриваются некоторые из этих проектов (раздел V) в контексте вопросов, которые были подняты в ходе опроса и имеют отношение к рекомендациям.

16. В обзоре используется информация, полученная из тематических исследований, опроса и предыдущей работы КЕС и показывающая, что НСУ имеют все возможности для того, чтобы играть ведущую роль в координации деятельности по связыванию данных в НСС. В обзоре содержится ряд рекомендаций по вопросам и вызовам, на которые необходимо ответить, чтобы НСУ могли функционировать как пользовательские шлюзы для связанных данных и услуг, основанных на таких данных.

## Систематический подход

17. В настоящем углубленном обзоре рассматривается **необходимость систематического подхода к связыванию данных** для удовлетворения политических, аналитических и оперативных потребностей. Отправной точкой для связывания данных должна быть четкая формулировка соответствующей информационной потребности, а также следование принципам необходимости и пропорциональности (рассмотрены ниже), с тем чтобы связанные данные представляли необходимость для принятия обоснованных решений с учетом этических соображений (Rancourt, 2019). Это согласуется с утверждением о том, что социальная статистика «должна строиться как система, связанная воедино концептуальной основой, которая отражает наиболее передовой исследовательский опыт и наиболее качественную информацию о причинно-следственных связях, а также о связях между мерами политики и результатами» (Fellegi & Wolfson, 1999).

18. Систематический подход к связыванию данных определяется политическими, аналитическими и/или оперативными вопросами или потребностями, основан на консультациях с заинтересованными сторонами и научно обоснованных исследованиях и имеет целенаправленный характер на всех этапах. Как показано ниже, такой подход к связыванию данных предусматривает ряд этапов для обеспечения правильной реализации и соответствия требованиям пользователей.

## Систематический подход к связыванию данных

## Систематический подход к связыванию данных



a) Первым этапом в рамках систематического подхода к связыванию данных является **идентификация проблемы**, например определение того, **какие** информационные потребности, аналитические вопросы или операционные задачи (например, нагрузка на респондентов, отсутствие данных) должны быть решены.

b) Вторым этапом — это рациональное обоснование того, **почему** существует необходимость в связывании данных, например в использовании связывания данных в качестве эффективного решения для получения детализированных данных и/или улучшения качества данных. Этот шаг согласуется с принципами **необходимости и пропорциональности**. **Необходимость** — это принцип, связанный с потребностями в информации, с тем, кому и зачем требуется такая информация. **Пропорциональность** — это принцип получения тех данных, которые необходимы (и не больше, чем необходимо) таким образом, чтобы это соответствовало ожидаемым выгодам от проекта и обеспечивало соблюдение требований к этике, конфиденциальности и прозрачности данных. На данном этапе также проводятся консультации с заинтересованными сторонами, чтобы обеспечить включение основных переменных и адекватный охват подгрупп совокупности или географических единиц.

c) Третий этап посвящен **методам связывания** или **тому, как** связанные данные будут использоваться для удовлетворения потребностей в данных, указанных на первом и втором этапах. Необходимо оценить различные источники данных, чтобы выбрать наиболее полезные из них, исходя из требуемых переменных и уровней дезагрегирования. Ключевыми контрольными точками этого этапа являются выявление методологических проблем увязки данных, подходы к их решению, а также проверка данных для обеспечения достаточного объема выборки и качества данных при соблюдении их конфиденциальности. Когда связывание данных используется для удовлетворения информационных потребностей или ответа на аналитические вопросы, необходим аналитический план, в соответствии с которым будет проводиться анализ и обеспечиваться соотнесение проекта с потребностями заинтересованных сторон.

d) Четвертый этап отвечает на вопрос «**Что в результате?**» — т. е. на этом этапе сообщается о **ключевых преимуществах и результатах** или добавочной стоимости связывания данных, таких как новые сведения о данных, повышение качества данных, снижение нагрузки на респондентов или эффективное использование активов данных в НСС. Эффективное информирование

о преимуществах связывания данных — один из основных аспектов формирования и поддержания общественного одобрения этих проектов. Результаты оперирующих с помощью данных алгоритмов на основе связывания данных должны быть четко интерпретированы и переработаны в ключевые выводы, которые будут служить обоснованием для принятия решений и разработки политики. Это включает в себя обсуждение сильных сторон и ограничений связывания данных.

19. Обратная связь усиливает взаимосвязь между результатами и вновь возникающими потребностями в информации. Постоянное взаимодействие с заинтересованными сторонами (как поставщиками, так и пользователями данных) с самого начала является неотъемлемой частью систематического подхода к связыванию данных, благодаря чему итоговые результаты будут целенаправленно и полноценно удовлетворять информационные потребности, а также способствовать формированию культуры обмена данными в НСС.

20. Подход к связыванию данных и методы его осуществления также зависят от законодательного и институционального контекста и требований к конфиденциальности в каждой стране. В следующем разделе рассматриваются примеры связывания данных в разных странах, которые отвечают различным политическим, аналитическим или операционным потребностям. Эти примеры также будут использованы для выявления вызовов, касающихся различных прикладных аспектов, подходов или методов, которые НСУ используют для извлечения большей пользы из статистических данных.

#### IV. Страновой контекст и практика

21. Как правило, страны, в которых существует более централизованная НСС, находятся в более выгодном положении для статистической деятельности по связыванию данных, чем страны с менее централизованными системами. Одним из основных препятствий для связывания данных из разных источников является несоответствие данных и метаданных из-за отсутствия стандартов или процесса гармонизации. НСС не может функционировать как экосистема без **операционной совместимости** или логистических возможностей для обмена и использования данных из различных источников согласованным образом. Связывание данных между источниками может быть еще более затруднено в децентрализованных системах из-за юридических препятствий для обмена данными или отсутствия приверженности горизонтальным инициативам, а также препятствий, касающихся координации.

22. Страны, в которых статистические системы основаны на реестрах, имеют самую высокую степень централизации, при этом скандинавские страны использовали свои системы, основанные на реестрах, и стали мировыми лидерами в области связывания данных. Системы, основанные на реестрах, опираются на административные данные из разных областей (например, здравоохранение, образование, трудовые ресурсы), которые интегрируются в статистическую систему. Эти системы упрощают процесс интеграции данных, благодаря выполнению ключевых предварительных условий, таких как наличие правовой основы для интеграции данных, обеспечение общественного одобрения массового связывания данных («общественное согласие») и внедрение единой системы идентификации для интеграции данных на уровне единиц из различных источников<sup>2</sup>. В странах, не имеющих национальных реестров, существует больше юрисдикционных препятствий для доступа к данным и техническая необходимость в создании необходимой для связывания данных среды.

23. Один из недостатков централизованных систем заключается в том, что НСУ, как правило, оказываются еще больше отстраненными от обсуждения политики в конкретных областях и могут не обладать опытом в конкретной области, позволяющим создавать статистические и аналитические продукты, учитывающие

<sup>2</sup> Более подробную информацию об этом типе статистической системы см. в докладе ЕЭК ООН «Статистика на основе регистров в Северо-европейских странах».

интересы пользователей (Edmunds, 2005). В соответствии с систематическим подходом, необходимо постоянное консультативное взаимодействие для определения приоритетов и обеспечения того, чтобы результаты связывания данных отвечали информационным потребностям подразделений, входящих в статистическую систему.

24. Связывание данных важно для подготовки статистических данных в разных странах. Например, в Соединенном Королевстве Великобритании и Северной Ирландии связывание данных играет важную роль в подготовке статистических данных в рамках всего правительства, а также в рамках сотрудничества с университетами и другими исследовательскими организациями. В следующих подразделах приводятся примеры связывания данных из Канады, Польши и Российской Федерации, иллюстрирующие преимущества работы НСУ в качестве руководящих информационных органов в сложной экосистеме данных. Кроме того, на эту тему был проведен опрос НСУ, и в следующем разделе представлены результаты анкетирования.

## A. Канада

25. Статистическое управление Канады играет координирующую роль в связывании данных из различных административных источников, находящихся вне сферы ведения управления. В эти функции входит определение информационных потребностей, консультации с поставщиками и пользователями данных, анализ осуществимости, гармонизация данных и обеспечение их качества, а также разработка ИТ-инфраструктуры и институциональных процедур для защиты конфиденциальности респондентов при связывании данных. Примерами крупномасштабных систем связывания, разработанных в Статистическом управлении Канады, являются Система связывания социальных данных (СССД), Система связывания файлов для коммерческих предприятий (ССФК) и Система связывания открытых данных (ССОД). Система связывания данных — это безопасная инфраструктура и система обработки данных, которая связывает деидентифицированные записи на уровне единиц из нескольких файлов административных данных по запросу. Она не представляет собой единый, интегрированный файл данных, поскольку связывание конкретных источников и переменных выполняется только по мере необходимости для утвержденных проектов.

26. В дополнение к общим связующим системам Статистическое управление Канады также разработало аналитические платформы с возможностью установления связей между источниками и областями для удовлетворения сложных информационных потребностей. В качестве некоторых примеров можно привести Канадскую базу данных динамики работодателей и работников (КБДРР), Канадские когорты переписи населения по вопросам здоровья и окружающей среды (ККПНЗО), а также Платформу динамического анализа в области образования и рынка труда (ПДАОТ). Ниже КБДРР рассматривается как пример такого связывания данных, обеспечивающего преимущества в сложной экосистеме данных.

27. КБДРР содержит сопоставительные сведения о работниках и работодателях на канадском рынке труда, что позволяет проводить исследования по различным темам, включая производительность и мобильность труда, промышленную организацию а также экономическое развитие и рост. КБДРР не является единым, интегрированным набором данных — структура данных представляет собой набор связанных между собой файлов, основанных на уникальных индивидуальных и отраслевых идентификаторах. Связываемые файлы, составляющие КБДРР, получены из административных данных Статистического управления Канады, а также министерств по вопросам доходов, занятости и социального развития и иммиграции.

28. КБДРР разработана таким образом, чтобы защищать данные в соответствии с положениями о конфиденциальности Закона о статистике, Правилами использования административных данных Статистического управления Канады и Директивой по связыванию микроданных, а также следовать принципам необходимости и пропорциональности, сформулированным выше. Все записи на уровне единиц

деидентифицируются и хранятся на защищенном сервере в Статистическом управлении Канады. Данные КБДРР закрыты для общественности из-за их конфиденциального характера. С целью облегчить доступ к данным частичная версия КБДРР (под названием «Аналитические микроданные по предприятиям и работникам») предлагается в исследовательских центрах данных Статистического управления Канады, которые представляют собой защищенные физические платформы, где аккредитованные пользователи данных, связанные с принимающей организацией, имеющие заранее утвержденные аналитические проекты и приравненные к сотрудникам Статистического управления Канады, могут получить доступ к данным.

29. КБДРР создала возможности для стратегических аналитиков и исследователей, позволяющие заполнить информационные пробелы, что способствует достижению политических целей. Например, коренные народы и представители видимых меньшинств<sup>3</sup> являются группами, обозначенными в [Законе Канады о равенстве в сфере занятости](#), но до появления КБДРР не хватало информации для мониторинга их прогресса в отношении владения предприятиями. Этот недостаток информации был признан заинтересованными сторонами препятствием для разработки программ и оказания экономической поддержки группам с целью обеспечения равноправия в сфере занятости (Gueye, Lafrance-Cooke, & Oyargun, 2022), что является потенциальной угрозой для достижения политических целей, связанных с инклюзивностью и многообразием. КБДРР (дополненная связкой с данными переписи населения) используется для восполнения пробелов в информации о владении предприятиями представителями [коренных народов](#) и [чернокожими канадцами](#), позволяя составить портрет их присутствия в канадской экономике, определить их социально-демографические характеристики и провести сравнительный анализ эффективности их предпринимательской деятельности с течением времени.

30. КБДРР также предоставила Статистическому управлению Канады возможность получить многомерную информацию о трудоустройстве канадцев в контексте изменений на рынке труда. Данные КБДРР содержат информацию о влиянии «ресурсного бума» на национальную экономику Канады, показывая, как миграция и поездки на работу из удаленных от места работы районов расщепили выгоды и издержки от этого бума в пределах обширного географического региона (Green et al., 2019). Одно из преимуществ связывания налоговых данных заключается в том, что они позволяют более точно измерить заработную плату тех работников, которые приезжают на работу из других регионов, и тех, которые проживают недалеко от места работы, и в результате лучше понять, как быстрый экономический рост в богатых ресурсами регионах оказывает побочное влияние на заработную плату на рынках труда в других регионах. В других исследованиях Статистического управления Канады КБДРР использовалась для измерения канадской «гиг-экономики» и характеристик гиг-работников, что позволило получить представление о росте нестандартных форм работы (Jeon, Liu, & Ostrovsky, 2021) и о том, как характеристики фирмы влияют на экономическую интеграцию иммигрантов (Ci & Hou, 2017).

## **В. Польша**

31. Статистическое управление Польши разработало Интегрированную систему метаданных (ИСМ), которая представляет собой ИТ-систему для создания статистического реестра и использования данных из различных областей и источников. ИСМ позволяет наблюдать и анализировать явления, которые невозможно наблюдать отдельно ни в одном из источников данных, подключенных к

---

<sup>3</sup> Данный термин используется в настоящем документе, поскольку описывает демографическую категорию, официально закрепленную в Законе о равенстве в сфере занятости, который в настоящее время пересматривается (см. [Консультации по пересмотру Закона о равенстве в сфере занятости](#)). Кроме того, Статистическое управление Канады провело консультации, чтобы определить подходящую терминологию для концепции видимого меньшинства (см. [Консультативная работа по концепции видимого меньшинства](#)).

системе. ИСМ состоит из трех перечней, включающих перечень населения, перечень зданий и квартир, а также перечень предприятий. Для составления перечня населения используется семь источников данных, и одним из ключевых процессов является сбор и объединение данных с использованием национальных личных идентификационных номеров из различных административных источников данных. Помимо основной рамки физических лиц, в ИСМ имеются отдельные тематические блоки (связанные с рамкой совокупности физических лиц), которые содержат предметную информацию, касающуюся всей совокупности или определенной подгруппы, и могут быть использованы для удовлетворения различных информационных потребностей. ИСМ можно использовать для сравнения ежегодных показателей или анализа меняющихся тенденций демографических, социальных и экономических явлений.

32. ИСМ включает в себя:

a) Систему обработки административных реестров (СОАР), в которой осуществляется подготовка, преобразование, проверка, корректировка и интеграция данных, полученных из официальных реестров, информационных систем органов государственного управления и негосударственных информационных систем;

b) Систему качества переменных (СКП), которая используется для осуществления передового анализа качества административных переменных и контролирует изменения в метаданных реестров;

c) Систему статистических операций (ССО), которая используется для создания реестра населения путем интеграции отдельных административных источников;

d) наборы данных домена (НДД), под которыми понимаются уникальные тематические информационные блоки, предметная область которых определяется отчетами ССО, причем предметная область охватывает конкретную тематику, являющуюся основой для наблюдения за социальными, экономическими и пространственными явлениями.

33. Статистическое управление Польши использует ИСМ для заполнения важных и требующих оперативного реагирования информационных пробелов. Во время пандемии COVID-19 возникла проблема с определением объема имеющих в Польше ресурсов медицинских кадров. Статистические исследования проводились на основе отчетов медицинских учреждений. Предоставленные ими данные были агрегированными, а это означало, что если врач или медсестра работали в нескольких местах, то они фигурировали в итоговых данных несколько раз. Еще одним явлением, искажающим данные, был тот факт, что часть медицинского персонала, проживающего вблизи государственной границы, нередко работала в соседней стране. Кроме того, некоторые врачи, а чаще всего медсестры, имеющие лицензию на медицинскую деятельность, работали не в медицинских учреждениях, а в других местах. Благодаря интеграции данных из многих источников (например, из реестра врачей, реестра медсестер, системы социального обеспечения, реестра медицинских учреждений) стало возможным производить очень точные расчеты практически в режиме реального времени. Это позволило определить, сколько врачей и медсестер проживает в Польше, имеется ли у них лицензия на медицинскую деятельность, работают ли они с пациентами и в каких учреждениях. Также удалось выявить демографические характеристики этих лиц и их территориальное распределение.

## **С. Российская Федерация**

34. Федеральная служба государственной статистики Российской Федерации (Росстат) разработала государственную информационную систему «Цифровая аналитическая платформа предоставления статистических данных» (Платформа). Платформа организована в соответствии со следующими принципами: a) прозрачность процессов сбора и обработки информации; b) использование единой методологии сбора и обработки статистических данных; c) достоверность и непротиворечивость информации системы; и d) однократность предоставления

данных и многократность их использования. Работа Платформы предполагает взаимодействие с информационно-вычислительной системой Росстата, а также с внешними информационными системами. Возможность загрузки, обработки и анализа данных (включая официальную статистическую информацию, административные и первичные статистические данные) из различных источников реализована в подсистемах Платформы. Платформа используется в следующих целях:

a) **Расчет и анализ показателей реализации приоритетных национальных проектов Российской Федерации.** Цель состоит в том, чтобы оценить эффективность проектов, а также выполнение национальных целей развития Российской Федерации. Наборы данных, используемые для расчета этих показателей, конфигурируются автоматически из внешних информационных систем. В системе предусмотрены алгоритмы расчета показателей по утвержденным методикам. Для оценки степени связи между компонентами показателя и их влияния на сам показатель применяется корреляционный анализ.

b) **Подсистема цифровых методологий** автоматизирует процессы формирования, согласования и утверждения методик расчета показателей и преобразования их в структурированные электронные документы. Цель состоит в унификации и систематизации требований к расчету показателей национальных и федеральных проектов, что обеспечивает использование единых подходов, направленных на воспроизводимость и прослеживаемость данных. Эта подсистема предназначена для i) оптимизации процесса гармонизации данных; ii) обеспечения единой точки сбора, расчета и представления значений показателей; iii) оценки прогресса в достижении требуемых значений показателей с учетом влияющих на них факторов; и iv) синхронизации процессов принятия решений в рамках управления проектами и подготовки статистики.

c) **Технологический модуль для оперативной оценки доходов по 10-процентным (децильным) группам.** Модуль направлен на совершенствование процесса формирования и выпуска статистической информации и разработку мер по сокращению бедности. Система аккумулирует данные из различных источников, включая микроданные обследований, итоговые таблицы официальных форм статистического наблюдения, тексты законов и постановлений, административные данные и дополнительную информацию.

d) **Система микромоделирования в области социальной поддержки.** Это микромоделирование прогнозирует изменения ключевых показателей при различных сценариях изменения доходов населения и границы уровня бедности. Система также определяет наиболее эффективные меры социальной поддержки и бюджетные расходы для достижения оптимальных результатов. Система микромоделирования основана на выборочном наблюдении данных о доходах населения. Она позволяет создавать различные сценарии изменения доходов на индивидуальном уровне в зависимости от изменения прожиточного минимума и границы уровня бедности. Специально разработанный интерфейс позволяет пользователю моделировать доходы от социальных пособий и другие компоненты доходов, такие как трудовые или имущественные доходы. Результатом моделирования является набор интерактивных аналитических отчетов, состоящих из рассчитанных показателей (изменение уровня бедности, бюджетных расходов и душевых доходов населения). Также имеется возможность моделирования и прогнозирования на основе статистических данных в различных областях статистики.

## D. Результаты опроса

35. В рамках настоящего углубленного обзора был проведен опрос тех НСУ, которые ранее выразили заинтересованность в предоставлении информации о своем опыте связывания данных между областями и источниками. Этот опрос включал вопросы о а) типе НСС; б) нынешних функциях по связыванию данных; в) будущих функциях по связыванию данных; г) примерах использования связывания данных для удовлетворения информационных потребностей; д) стандартах, инструментах

и инфраструктуре для облегчения связывания данных; f) уроках, извлеченных из работы по связыванию данных; и g) возможностях для международного сотрудничества. Ответы были получены от Венгрии, Италии, Канады, Латвии, Мексики, Нидерландов и Эстонии. Ответы кратко изложены ниже и подробно представлены в документе ECE/CES/2024/INF.1.

36. **Тип НСС (контекст страны).** Страновой контекст является важным аспектом готовности НСУ к связыванию данных между областями и источниками. В ходе опроса респондентов просили предоставить информацию о типе статистической системы, существующей в их стране, включая подробные сведения о степени централизации их систем и роли НСУ в координационной работе и подготовке официальной статистики. В НСС опрошенных стран распространена централизация официальной статистики. Большинство НСУ, принявших участие в опросе, имеют законный доступ к административным данным всей НСС и координируют статистическую деятельность. Однако даже в централизованных системах данные и ответственность за официальную статистику в некоторых областях рассредоточены по различным ведомствам и уровням правительства.

37. **Нынешние функции НСУ.** В анкете респондентов просили описать нынешние функции и обязанности их НСУ в области связывания административных данных из различных источников. Опрошенные НСУ, как правило, играют координирующую роль в увязке данных из разных ведомств, а использование таких связанных данных ограничено статистическими целями и регулируется законами, обеспечивающими конфиденциальность данных. Большинство НСУ разработали необходимую для связывания данных ИТ-инфраструктуру и выполняют требования законов о статистике, чтобы обеспечить конфиденциальность данных и гарантировать, что данные будут использоваться только в статистических целях и без каких-либо злоупотреблений.

38. **Будущие функции НСУ.** В анкете респондентам также задавался вопрос о том, существуют ли дополнительные функции, которые их НСУ могли бы взять на себя сейчас или в будущем и которые позволили бы им перейти от роли поставщиков данных к роли производителей актуальных статистических показателей и многомерной информации. Сообщалось, что НСУ могут взять на себя более важную роль в координации деятельности по связыванию данных в рамках НСС. Расширение функций будет способствовать эффективному использованию ресурсов (активов данных, человеческого капитала и ИТ-инфраструктуры) и обмену информацией в рамках статистической системы. В ответах преобладала мысль о том, что НСУ имеют все шансы стать пользовательскими шлюзами для доступа к связанным данным. НСУ может позиционировать себя как «национальный аналитический и экспертный центр, отвечающий за связывание административных данных и предоставляющий доступ к таким данным» в защищенной среде (Латвия).

39. **Связывание данных для удовлетворения информационных потребностей.** В вопроснике респондентов просили привести примеры информационных потребностей в их стране, которые были удовлетворены путем связывания данных. Большинство НСУ, принявших участие в опросе, использовали связывание данных для решения проблем, касающихся операционных задач, например, для снижения нагрузки на респондентов, уменьшения затрат на проведение обследований, повышения точности статистических данных, а также для решения проблем с точки зрения ошибок охвата и снижения доли ответивших респондентов. Связывание данных, продиктованное имеющимися возможностями, включало получение более своевременной статистики и понимание явлений, которые невозможно наблюдать с помощью одного источника данных. НСУ сообщили, что «связывание данных является наиболее ценной нашей услугой — она быстрее и гибче, чем официальная статистика» (Эстония), и она «обеспечивает множество решений проблемы публикации мгновенной и точной демографической статистики» (Венгрия). «Вероятно, наиболее значимым примером связывания данных является Система социальных баз данных (ССБ)» (Нидерланды). ССБ основывается на имеющихся возможностях, поскольку статистические данные переписи основаны на уже имеющихся в системе данных.

40. **Упрощение связывания данных.** В ходе опроса респондентов спрашивали, существуют ли специальные стандарты, инструменты или инфраструктура (например, в области права, информационных технологий или обработки данных), разработанные их НСУ для упрощения связывания данных между областями и источниками. Опрошенные НСУ а) предприняли шаги для обеспечения совместимости административных данных в своих НСС; б) разработали ИТ-инфраструктуру, необходимую для связывания данных и обмена ими; и в) выстроили процессы связывания административных данных для устранения дублирования и других признаков неэффективности в своих статистических системах. Сюда входит разработка систем **связывания данных** (в виде безопасной инфраструктуры данных и систем обработки данных для связывания деидентифицированных данных), а также программного обеспечения с открытым исходным кодом и экосистемных решений для обмена данными. Кроме того, в НСУ была внедрена ИТ-инфраструктура, позволяющая использовать принцип «только один раз», цель которого — собирать одни и те же данные только один раз, чтобы избежать дублирования и снизить нагрузку по предоставлению ответов.

41. **Извлеченные уроки.** Респондентов попросили подробно рассказать об уроках, извлеченных из их опыта связывания данных для удовлетворения информационных потребностей и обеспечения операционной совместимости. Выделялись две темы: необходимость уникальных идентификаторов для связывания данных и унифицированные соглашения об обмене данными в рамках всей НСС.

а) **Значение уникальных идентификаторов.** Общий идентификационный код или единый реестр административных данных в целом оказался полезен для подготовки официальной статистики и способствует связыванию данных, что позволяет создавать инновационные продукты, отвечающие специфическим требованиям процессов принятия решений, а также предоставлять более подробные и регулярные данные, полученные из обследований. Отсутствие уникальных идентификаторов создает определенные трудности, и одним из возможных решений являются вероятностные связи, но для проверки этих методик необходимо проделать дополнительную работу.

б) **Необходимость в усовершенствованных процедурах обмена данными.** НСУ отметили, что «получение доступа к административным источникам по-прежнему происходит слишком медленно и сложно» (Италия) и что из-за необходимости заключения соглашений об обмене данными с третьими сторонами существует «риск потерять поток данных» (Мексика). Существующие законодательные рамки недостаточны для своевременного доступа к данным и для доступа к данным из частного сектора, что необходимо «для получения более качественных и разнообразных данных» (Эстония). Усовершенствованные соглашения об обмене данными необходимы для формализации отношений между НСУ и другими ведомствами, а также для получения общественного одобрения связывания административных данных и использования данных из частного сектора. Для снижения проблем с конфиденциальностью при связывании микроданных из разных источников используется технология связывания с сохранением конфиденциальности (СДСК), которая позволяет проводить аналитическую работу с конфиденциальными наборами данных без необходимости вывода данных из хранящих их учреждений (Канада).

42. **Возможности международного сотрудничества.** Респондентов спросили, существует ли какая-либо международная инициатива или программа совместной работы, с которой НСУ могли бы сотрудничать сейчас или в будущем, чтобы повысить эффективность сбора информации из нескольких источников. Респонденты назвали несколько значимых с их точки зрения текущих международных проектов, посвященных техническим аспектам и проблемам связывания данных. В качестве важных текущих проектов международного сотрудничества были отмечены Группа по применению аналитики данных и современных методов ГВУ-МОС и Группа поддержки стандартов.

## V. Смежная работа в рамках Конференции европейских статистиков и Группы высокого уровня по модернизации официальной статистики

43. КЕС и ГВУ-МОС провели несколько мероприятий по теме связывания данных. В настоящем разделе представлен краткий обзор этих мероприятий. Информация в настоящем разделе служит контекстом для обсуждения вопросов, поднятых в углубленном обзоре, и позволяет избежать дублирования предыдущих работ по этой теме.

44. КЕС провела углубленный обзор, в котором описала опыт, полученный в ходе проекта интеграции данных ГВУ-МОС 2016 года. Целью этого проекта было накопление опыта для «выработки общих рекомендаций и руководящих указаний по интеграции данных и соответствующей системы качества». [Углубленный обзор по интеграции данных](#) был представлен на совещании КЕС в феврале 2017 года и был посвящен общему описанию наиболее распространенных типов интеграции данных и экспериментальным проектам на уровне стран с каждым из этих типов. В обзоре также рассматривались такие сложности интеграции данных, как:

а) **правовые и институциональные вопросы** — т. е. законодательство о конфиденциальности, информирование о деятельности для обеспечения общественного одобрения, сотрудничество с поставщиками данных и надзор за проектами по интеграции данных;

б) **управленческие вопросы** — т. е. людские ресурсы и ИТ-инфраструктура, необходимые для интеграции данных, а также организационные процедуры, необходимые для снижения рисков, сопутствующих интеграции данных;

в) **методологические вопросы** — т. е. такие нерешенные проблемы, как отсутствие унифицированных идентификаторов, различия в понятиях и классификациях, используемых для определения и организации данных, а также недостающие данные и ошибки охвата.

45. Важно отметить, что по результатам углубленного анализа было рекомендовано «использовать стандартные процессы, которые являются общими для различных типов интеграции данных, для существенного облегчения интеграции данных». В ходе обзора был составлен контрольный перечень элементов, которые может включать стандартный процесс интеграции данных. Систематический подход, изложенный в разделе III настоящего обзора, основан на этой рекомендации.

46. Целевая группа экспертов НСУ подготовила публикацию [«Руководство по интеграции данных для измерения миграции»](#). Директивным органам, научным работникам и другим заинтересованным сторонам необходимы данные о мигрантах — их количестве, показателях въезда и выезда, характеристиках и интеграции в общество. Эти данные должны быть полными, точными и часто обновляемыми. Не существует единого источника, который мог бы обеспечить такие данные о миграции, но, объединив несколько источников, можно получить необходимую пользователям информацию. В публикации представлен обзор способов интеграции данных, используемых для составления миграционной статистики, на основе опроса поставщиков миграционных данных в более чем 50 странах. Более подробно об интеграции данных в различных национальных контекстах можно узнать из тринадцати тематических исследований. В публикации предложены принципы передовой практики интеграции данных для измерения миграции, представлены методы объединения административных, статистических и других источников данных для составления миграционной статистики.

47. Система управления данными для обеспечения функциональной совместимости (СУДС) была разработана в рамках проекта ГВУ-МОС «Система управления данными для обеспечения функциональной совместимости», который осуществлялся в 2022–2023 годах. Совместимость данных означает возможность обмена и использования информации при минимальном предварительном

взаимодействии или полностью без него. Это также является основой для непрерывного перемещения данных между источниками и преобразования разрозненных массивов данных «в связанную сеть гармонизированных данных и наборов метаданных». СУДС содержит описание основных элементов, необходимых для создания и управления совместимой платформой данных, метаданных и систем. Поскольку в настоящем обзоре основное внимание уделяется связыванию данных из различных источников, рекомендации, полученные относительно СУДС, являются весьма актуальными.

48. В 2023 году [Бюро КЕС провело углубленный обзор этики данных](#). Основной тезис этого обзора заключается в том, что этические соображения, необходимые для связывания данных, шире, чем этические соображения для традиционных данных. В традиционных условиях НСУ в основном уделяют внимание деловой этике и безопасности данных. Для интеграции данных необходим расширенный взгляд на этику данных, ориентированный на общественное одобрение помимо вопросов конфиденциальности и безопасности данных. При проведении обследований и переписей процесс сбора данных является прозрачным, поскольку респонденты знают, какая информация о них собирается и для чего она будет использоваться. Однако это не совсем так в случае с административными данными и связывания различных источников данных. Общественность может не понимать пользы и масштабов связывания данных или не выражать согласия на их использование. Это означает, что НСУ необходимо подумать о том, что нужно сделать, а не просто о том, что можно сделать, чтобы обеспечить общественное одобрение связывания данных, а для этого необходимо разъяснить общественности преимущества такого связывания и быть готовыми отказаться от тех проектов, в которых нет полной ясности относительно дальнейшего использования данных (или возможного неправомерного их использования).

## VI. Вопросы и вызовы

49. Все НСУ, участвовавшие в настоящем обзоре, сообщили, что играют координирующую роль в подготовке официальной статистики и имеют законный доступ к административным данным других государственных ведомств. Однако НСС представляют собой сложные экосистемы данных с многочисленными поставщиками и пользователями данных, что создает вызовы, связанные с доступом к данным и эффективностью работы всей НСС. Ниже кратко рассмотрены темы, которые были выявлены в ходе анализа тематических исследований и результатов опроса НСУ относительно их опыта в плане связывания данных между областями и источниками.

### A. Нынешние и будущие функции национальных статистических управлений

50. **Текущая деятельность по связыванию данных в НСУ направлена на эффективное использование существующих в НСС данных для получения актуальной и точной информации.** Увеличение затрат на проведение обследований и снижение доли ответивших респондентов являются вызовом для многих НСУ, и это происходит на фоне растущего спроса на оперативные, дезагрегированные и многомерные данные. НСУ, принявшие участие в настоящем обзоре, указали следующие практические причины для связывания данных, которые подтверждают преимущества интеграции данных, описанные в предыдущем [углубленном обзоре КЕС](#) по этой теме:

а) обычное связывание данных, ориентированное на решение конкретных задач, основано на принципе «только один раз», при котором административные данные интегрируются в обследования для замены переменных (или даже для полной замены обследований, как в случае с переписями на основе регистров), чтобы решить проблему дублирования данных в рамках НСС, что снижает затраты на проведение обследований и нагрузку на респондентов. Связывание административных данных также используется при построении рамок выборки для обследований и в целях

решения проблем с качеством данных в виде уменьшения доли ответивших и ошибок охвата;

b) связывание данных, ориентированное на использование имеющихся возможностей, проводится в следующих целях:

i) использование непрерывного потока административных данных в качестве ресурса для подготовки более регулярных оценок, более тщательного мониторинга тенденций и реагирования на такие кризисы, как пандемия COVID-19;

ii) использование большого объема административных данных для расширения охвата малочисленных и труднодоступных групп населения, удовлетворения потребностей в данных для дезагрегированной статистики и наблюдения за явлениями, которые незаметны в отдельных источниках данных.

**51. Национальные статистические управления имеют все возможности для работы в качестве пользовательских шлюзов с целью обеспечения доступа к связанным данным и основанным на таких данных услугам.** Респонденты согласились с тем, что НСУ следует назначить четко определенную координирующую роль из-за сложностей связывания данных в НСС, включающих множество сторон и характеризующихся ростом объема административных данных. Заключение соглашений об обмене данными на двусторонней основе стало нецелесообразным, при этом людские и ИТ-ресурсы, необходимые для правильной увязки данных, требуют слишком больших затрат, умножающихся в несколько раз в рамках всей НСС. НСУ обладают необходимыми компетенциями для создания инфраструктуры, методологий, процедур и регламентов, необходимых для связывания данных, обеспечивая при этом соблюдение принципов ПДФП и этики данных.

## **В. Вызовы для национальных статистических управлений**

**52. Одним из препятствий для связывания данных между областями и источниками является отсутствие четко определенных ролей и упорядоченного процесса обмена данными.** Наличие централизованной статистической системы и законодательства, позволяющего НСУ получать доступ к административным данным из разных частей НСС, является недостаточным условием для того, чтобы НСУ стали пользовательскими шлюзами для доступа к связанным данным и основанным на таких данных услугам:

a) хотя НСУ, ответившие на проведенный для настоящего обзора опрос, как правило, имели законный доступ к административным данным, у них было меньше контроля над тем, как эти данные собираются, и над структурой данных, что может снизить операционную совместимость и качество данных;

b) одним из часто упоминаемых препятствий для связывания данных между внешними источниками является отсутствие уникальных идентификаторов, при этом использование вероятностного связывания обсуждалось как временный подход, который может ухудшить качество связывания данных;

c) с тем чтобы не повторять эту работу для каждой базы данных в системе, необходим единый процесс передачи, преобразования и связывания данных;

d) не всегда существуют единые стандарты в отношении сбора данных, структуры данных, методологий, идентификаторов, терминов и определений, но они необходимы для обеспечения функциональной совместимости и сокращения времени и ресурсов, необходимых для связывания данных;

e) процесс получения административных данных из других ведомств зачастую является медленным и обременительным.

**53. Риски, связанные с несовершенством стандартов обмена данными, заключаются в увеличении временных затрат и перебоях в потоках данных.** Единые стандарты данных необходимы для того, чтобы расширить возможности

использования данных для будущих и текущих оперативных нужд, а также для целей, отличных от тех, для которых эти данные были изначально собраны. Совершенствование стандартов и гармонизация данных (включая метаданные) из различных областей и источников облегчает связывание данных, а также помогает тем пользователям, которые могут быть экспертами в своих областях, но не разбираются в других.

54. **Существует спрос на предоставление связанных данных предприятиям и неправительственным организациям, но это сопряжено с дополнительными трудностями.** Предоставление данных сторонам, не относящимся к государственному сектору, является принципиально новым этапом. Предприятия все чаще запрашивают у НСУ данные и информационные услуги, но предоставление данных предприятиям порождает новые, требующие решения вопросы, связанные с общественным одобрением, этикой данных и конфиденциальностью. Существует множество случаев, когда предоставление данных частному сектору или неправительственным организациям запрещено. Учитывая эту асимметрию, НСУ необходимо найти правильный баланс в вопросах связывания данных, а также найти способы совершенствования процесса предоставления данных как части более масштабного процесса оптимизации и модернизации статистической работы.

55. **Растет понимание того, что для удовлетворения информационных потребностей необходимы большие данные и данные, находящиеся в частном владении.** НСУ необходимо получить доступ к данным, собранным коммерческими структурами и с помощью информационно-коммуникационных технологий (например, веб-скрейпинга), чтобы расширить возможности официальной статистики по реагированию на экономические вопросы в частном секторе и мониторингу социальных вопросов, не охваченных административными данными или данными обследований. КЕС завершила [углубленный обзор по вопросам сотрудничества с поставщиками данных из частного сектора](#), в котором описаны вызовы и уроки, извлеченные НСУ на данный момент в этой области.

56. **Использование существующих и новых данных из множества источников для построения показателей может привести к разрастанию числа показателей.** НСУ могут взять на себя роль по стандартизации или гармонизации показателей, чтобы обеспечить их единообразие и сопоставимость между системами показателей. Если НСУ возьмут на себя эту роль, это также позволит разрабатывать показатели на основе одних и тех же подмножеств совокупности, что не всегда возможно при использовании несвязанных данных.

57. Как отмечалось в описанных выше вызовах, в дополнение к культурным изменениям, которые должны быть проведены в НСУ, необходимы и важны ответные меры со стороны держателей административных и других источников данных, чтобы позволить НСУ превратиться из поставщиков данных в разработчиков необходимых статистических показателей, обеспечивающих аналитическую информацию.

## VII. Выводы и рекомендации

58. В настоящем углубленном обзоре рассматривается, как НСУ используют связывание данных для решения операционных проблем (таких, как снижение доли ответивших респондентов) и удовлетворения информационных потребностей, одновременно преодолевая вызовы в плане предоставления данных. В обзоре также обсуждаются перспективы того, как НСУ могут перейти от предоставления официальной статистики к выполнению дополнительных функций пользовательского шлюза для доступа к связанным данным и поставщиков аналитической информации о многомерных явлениях. Обзор основан на тематических исследованиях, опросе НСУ и изучении предыдущих работ КЕС по вопросам интеграции данных и смежным темам.

59. Выносятся следующие рекомендации:

а) **Для проектов по связыванию данных необходим систематический подход.** Связывание данных является одним из ключевых методов, который может быть использован для удовлетворения сложных информационных потребностей, однако для его применения требуется систематический подход, обеспечивающий эффективную реализацию и политическую актуальность. Этот подход должен i) определяться политическими, аналитическими и/или оперативными вопросами и потребностями; ii) основываться на результатах предыдущей работы; iii) носить целенаправленный характер на всех этапах, то есть определяться либо существующими проблемами (в виде мер реагирования), либо имеющимися возможностями (в виде упреждающих мер); и iv) предусматривать постоянное консультативное взаимодействие с заинтересованными сторонами для поддержки сотрудничества в области обмена данными и стандартов, а также обеспечения соответствия связанных данных интересам пользователей.

б) В основе систематического подхода лежит оценка существующих и альтернативных источников данных с целью выбора наиболее полезных источников в зависимости от переменных и требуемого уровня дезагрегирования. **Необходимо также изменить подход к сбору данных.** НСУ следует сначала изучить варианты использования существующих источников данных, а затем оценить возможность использования других дополнительных источников данных из разных источников или областей.

в) **Необходимо, чтобы НСУ выполняли официально закрепленную координирующую функцию по связыванию данных.** Законный доступ к административным данным в рамках всей НСС не устраняет препятствий для доступа к данным из разных источников, что сказывается на услугах, которые могут предоставлять НСУ, и удовлетворении информационных потребностей в рамках всей НСС. Предоставление НСУ возможности работать в качестве пользовательского шлюза для связанных данных имеет ряд преимуществ. Такой шлюз может стать центром доступа к связанным данным для всей НСС и позволит избежать повторного связывания данных. Кроме того, это позволяет обеспечить соблюдение принципа «только один раз» и эффективно использовать активы данных в рамках НСС, а также обеспечить единообразие связывания и уменьшить число разнородных систем показателей.

г) **Следует рассмотреть возможность разработки «дорожной карты», которая поможет НСУ, стремящимся связывать данные в различных областях, предоставлять директивным органам более качественную информацию для решения многомерных социальных проблем.** Эта «дорожная карта» должна содержать 1) рекомендации относительно системы управления, которая должна быть создана для осуществления расширенного связывания данных между областями; 2) примеры типов данных в различных областях, которые традиционно не используются и которые могут быть полезны и связаны между собой для информационной поддержки мер политики; и 3) конкретные страновые примеры многомерного связывания данных в различных областях политики для иллюстрации связей между областями.

д) **Для разработки такой «дорожной карты» можно использовать результаты предыдущей работы КЕС по интеграции данных и смежным вопросам.** Результаты опроса, проведенного для настоящего углубленного обзора, указывают на технические проблемы функциональной совместимости и проблемы курирования данных, касающиеся соглашений о совместном использовании данных и обеспечения общественного одобрения связывания данных. Респонденты опроса, проведенного для настоящего углубленного обзора, рекомендовали Группу по применению аналитики данных и современных методов ГВУ-МОС и Группу поддержки стандартов в качестве важных проектов международного сотрудничества по техническим аспектам связывания данных между областями и источниками. Предыдущая работа КЕС по вопросам функциональной совместимости, этики данных и курирования данных также является ориентиром для оптимизации обмена данными в рамках НСС и обеспечения общественного одобрения проектов по интеграции

данных. Вместо того чтобы создавать новую группу для разработки «дорожной карты», можно поручить эту работу существующим группам.

## **VIII. Обсуждение в Бюро Конференции европейских статистиков**

60. В феврале 2024 года Бюро провело углубленный обзор вопросов связывания данных между областями и источниками на основе подготовленного Канадой документа и замечаний ЕЭК ООН. Были высказаны следующие замечания:

a) в документе представлен весьма качественный обзор вопросов по теме связывания данных, с акцентом на стратегические и управленческие вопросы: рассматривается необходимость связывания данных и соответствующие вызовы и возможности. Предложение о разработке «дорожной карты» станет конструктивным шагом вперед в международной работе над этой темой;

b) связывание данных играет ключевую роль в превращении НСУ из поставщиков данных в разработчиков актуальных статистических показателей и аналитики в ответ на растущую потребность в многомерной статистической информации. Этот вопрос носит фундаментальный характер и требует изменения культуры не только от НСУ, но и от держателей административных и прочих источников данных. Связывание данных не является чем-то новым, но тип и количество наборов данных, которые могут быть связаны между собой, делают возможным трансформационные изменения. Этот вопрос будет оставаться стратегически важным еще долгие годы;

c) необходим системный подход, чтобы изменить менталитет, а не просто реагировать на запросы. НСУ должны иметь официально оформленные функции по связыванию данных и могли бы стать пользовательскими шлюзами, координирующими увязку данных в рамках своих национальных статистических систем;

d) уровень социальной приемлемости связывания данных в странах различен: в одних странах оно воспринимается положительно, а в других вызывает опасения. В тех странах, где нет системы уникальных идентификаторов или реестра населения, осуществление связывания данных представляет собой серьезную проблему. Это еще раз подчеркивает важность информационной работы, прозрачности и партнерства;

e) НСУ может быть поручено связывание данных из разных источников, но при этом им не разрешается передавать связанные данные другим учреждениям по юридическим причинам или по причине отсутствия общественного одобрения. Такая асимметрия может создавать трудности для других государственных органов. Институциональные отношения между НСУ и другими государственными органами должны быть правильно сбалансированы;

f) с тем чтобы облегчить связывание данных необходимо повысить уровень гармонизации метаданных, терминологии и классификаций. Например, термин «административные данные» имеет разное значение в разных странах;

g) связывание данных сопряжено с предоставлением общего доступа к данным. Этим вопросам следует уделять одинаковое внимание, руководствуясь требованиями пользователей. Горизонтальные вопросы и вопросы, связанные со спросом, должны рассматриваться одновременно;

h) связывание данных позволяет создавать интегрированные данные по различным областям. Это увеличивает потребность в поддержке пользователей, которые могут быть экспертами в своих областях, но не разбираться в других;

i) связывание данных представляет собой обширную тему, имеющую множество аспектов, включая горизонтальные вопросы, такие как терминология, разъяснительная работа, общественная приемлемость, методы и информационная

работа, а также относящиеся к конкретной области вопросы, которые могут быть включены в повестку дня различных тематических групп;

j) вместо того чтобы создавать новую группу, обсуждение горизонтальных вопросов и разработку «дорожной карты» следует рассмотреть в рамках Группы высокого уровня по модернизации официальной статистики (ГВУ-МОС) и ее групп по применению аналитики данных и современных методов, а также по поддержке стандартов;

k) можно подготовить небольшой вопросник, позволяющий определить, как разные страны осуществляют связывание данных. На этой основе можно подготовить тематические исследования и собрать примеры;

l) в Соединенном Королевстве связывание данных играет важную роль в подготовке статистических данных в рамках всего правительства, а также в рамках сотрудничества с университетами и научных исследований. Группа по связыванию данных Управления национальной статистики (УНС) хотела бы принять участие в дальнейшей работе над этой темой;

m) в разработке «дорожной карты» готов принять участие Статкомитет СНГ.

61. Бюро сформулировало следующие выводы:

a) ГВУ-МОС рассмотрит возможность включения горизонтальных вопросов связывания данных в свою рабочую программу и в повестку дня своих групп, а именно, групп по применению аналитики данных и современных методов и по поддержке стандартов, по мере возможности и во всех случаях, когда эти вопросы относятся к мандатам соответствующих групп;

b) Бюро предлагает ГВУ-МОС, основываясь на результатах углубленного обзора, разработать «дорожную карту» по вопросам связывания данных;

c) вопросы связывания данных между источниками и областями следует включить в программу работы тематических групп в рамках КЕС и при необходимости в повестку дня совещаний экспертов. Это может предполагать сбор примеров из разных стран для последующего совместного ознакомления и распространения;

d) в ближайшие годы Бюро будет следить за прогрессом в этой области.

## Справочная литература

Ci, W., & Hou, F. (2017). [Immigrants' initial firm allocation and earnings growth](#). *Canadian Studies in Population*, 44(1–2), 42–58.

Conference of European Statisticians Task Force. (2023). [Data stewardship and the role of national statistical offices in the new data ecosystem](#).

Edmunds, R. (2005). [Models of statistical systems](#). Paris: OECD.

Fellegi, I., & Wolfson, M. (1999). [Towards systems of social statistics – Some principles and their application in Statistics Canada](#). *Journal of Official Statistics*, 15(3), 373–393.

Green, D., Morissette, R., Sand, B. M., & Snoddy, I. (2019). [Economy-wide spillovers from booms: Long-distance commuting and the spread of wage effects](#). *Journal of Labor Economics*, 37(S2), S643–S687.

Gueye, B., Lafrance-Cooke, A., & Oyarzun, J. (2022). [Identifying Indigenous business owners and Indigenous-owned businesses](#). *Analytical Studies: Methods and References*, No. 045 Statistics Canada, Catalogue no. 11-633-X.

Jeon, S-H., Liu, H., & Ostrovsky, Y. (2021). [Measuring the gig economy in Canada using administrative data](#). *Canadian Journal of Economics*, 54(4), 1638–1666.

Rancourt, E. (2019). [The scientific approach as a transparency enabler throughout the data life-cycle](#). *Statistical Journal of the IAOS*, 35(4), 549–558.

---