

**Европейская экономическая комиссия****Конференция европейских статистиков****Шестьдесят шестая пленарная сессия**

Женева, 18–20 июня 2018 года

Пункт 4 b) предварительной повестки дня

**Интеграция данных для измерения миграции****Руководство по интеграции данных для измерения  
миграции****Записка Целевой группы по интеграции данных для измерения  
миграции***Резюме*

В настоящем документе представлена краткая версия *Руководства по интеграции данных для измерения миграции*. Это руководство было подготовлено Целевой группой в составе представителей Испании (Председатель), Австрии, Италии, Новой Зеландии, Соединенного Королевства, Соединенных Штатов, Швейцарии, Евростата и ЕЭК ООН.

Нынешняя краткая версия Руководства подготовлена для перевода и не включает в себя страновые тематические исследования. Руководство включает в себя обсуждение определения термина «интеграция данных» со ссылкой, в частности, на статистику международной миграции (глава II); обсуждение результатов обследования методов интеграции данных (глава III); резюме результатов тринадцати национальных тематических исследований (глава IV); обсуждение и перечень метаданных (глава V), выводы (глава VI) и ряд общих рекомендаций для стран (глава VII).

Полный текст Руководства (включая страновые тематические исследования) был разослан всем членам Конференции европейских статистиков для проведения электронных консультаций; с ним можно ознакомиться по адресу <http://www.unece.org/index.php?id=47411>. В случае успешного исхода консультаций участникам пленарной сессии КЕС будет предложено одобрить данное Руководство.



## Содержание

	<i>Пункты</i>	<i>Стр.</i>
I. Введение.....	1–10	3
II. Определение термина «интеграция данных».....	11–31	5
A. Общие характеристики.....	11–23	5
B. Рабочее определение .....	24–31	7
III. Обследование методов интеграции данных.....	32–47	9
A. Обзор основных результатов обследования .....	35–40	9
B. Рекомендации .....	41–42	10
C. Резюме результатов.....	43–47	10
IV. Тематические исследования .....	48–49	11
A. Резюме тематических исследований .....	50–70	11
V. Метаданные .....	71–77	14
VI. Заключение .....	78–88	16
VII. Рекомендации .....	89–98	18
A. Улучшить доступ национальных статистических учреждений к административным данным .....	90	18
B. Использовать административные данные для статистики миграции.....	91–92	18
C. Комбинировать данные из различных источников с использованием «сигналов присутствия».....	93	18
D. Следить за качеством интегрированных данных .....	94	19
E. Обеспечивать прозрачность применяемых методов интеграции данных и разрабатывать стандарты .....	95	19
F. Развивать международное сопоставление данных о миграции и обмен ими.....	96–98	19
VIII. Дальнейшая работа .....	99–103	20
A. Потенциал больших данных .....	100	20
B. Продольные данные.....	101–103	20
Ссылки .....		20

## I. Введение

1. В последние годы тема интеграции данных вызывает большой интерес, особенно в условиях роста потребностей в данных при жестком налоговом климате в сочетании с возросшей нагрузкой на респондентов и требованиями в отношении конфиденциальности. В 2016–2017 годах Группа высокого уровня ЕЭК ООН по модернизации официальной статистики (ГВУ-МОС) занималась реализацией проекта в области интеграции данных. Этот проект предусматривал проведение экспериментов по интеграции данных в различных секторах и завершился разработкой практического электронного руководства по интеграции данных для официальной статистики<sup>1</sup>. Евростат поддержал исследования, посвященные методам совершенствования интеграции данных, такие как проект ESSnet в области интеграции результатов обследований и административных данных<sup>2</sup>. Этот проект стал первой попыткой создания общей методологической базы для интеграции различных источников данных.

2. Однако многие страны заинтересованы в интеграции данных непосредственно для целей измерения международной миграции. Учитывая проблемы в области сбора статистики миграции, зачастую полезно собирать данные из различных источников. Интеграция данных позволяет преодолевать затруднения, связанные с их охватом или точностью, которые могут касаться общих данных о численности и потоках мигрантов, и повышать разнообразие данных о миграции, добавляя к имеющимся данным социально-демографические или экономические параметры.

3. На европейском уровне значение интеграции данных для совершенствования статистики миграции было вновь подтверждено в ходе ежегодной конференции генеральных директоров НСУ 2017 года, программа которой предусматривала статистическое совещание на тему «Вопросы перемещения и интеграции населения: статистика миграции»<sup>3</sup>. Будапештский меморандум, который обсуждался на конференции и был одобрен Комитетом Европейской статистической системы, включает в себя список мероприятий, среди которых фигурирует *«содействие выявлению, оценке и внедрению новых методов и источников данных, в частности расширение использования для статистических целей источников административных данных соответствующего качества, обеспечиваемого благодаря постоянной оценке качества, – либо отдельных регистров, увязанных данных их различных административных источников, либо в сочетании с источниками данных, полученных в результате обследований, а также возможностям новых источников данных (например, больших данных)»*.

4. Методы интеграции различных источников административных данных внутри той или иной страны (например, регистры населения и регистрационные записи в сферах здравоохранения, трудовой деятельности, налогообложения или образования) для восполнения информации, которую не удастся получить из других источников (например, по измерению эмиграции для тех, кто не удаляет регистрационные данные), показали в ряде стран успешные результаты. В других случаях источники неадминистративных данных позволяют получать информацию, недополученную из административных источников (например, применение обследований домохозяйств для сбора информации, не включенной в регистры). Большие данные, поступающие от коммунальных и телефонных компаний, в некоторых странах также рассматриваются в качестве дополнительной информации. Другие виды применения интеграции данных могут заключаться в согласовании различных показателей миграции, поступающих из разных источников, в частности оценочных данных по миграционным группам, которые с трудом поддаются учету, таким как нелегальные мигранты или эмигранты.

5. Учитывая важность и злободневность этой темы, в 2014 году на рабочей сессии ЕЭК ООН-Евростата по статистике миграции состоялось обсуждение вопроса об

<sup>1</sup> <https://statswiki.unecce.org/display/DI/Data+Integration+Home>.

<sup>2</sup> [https://ec.europa.eu/eurostat/cros/content/data-integration\\_en](https://ec.europa.eu/eurostat/cros/content/data-integration_en).

<sup>3</sup> 103-я Конференция ГД НСУ (Будапешт, 20–21 сентября 2017 года). Документы и доклады см. по адресу <http://www.ksh.hu/dgins2017/presentations.html>.

использовании различных источников административных данных для измерения международной миграции. Обсуждение было сосредоточено на возможностях и затруднениях при использовании административных данных, особенно с учетом разных уровней развития систем административных данных на территории региона. Были рассмотрены пути улучшения сотрудничества между национальными миграционными службами, статистическими учреждениями, реестровыми службами и другими органами, готовящими административные данные, а также методы интеграции различных источников административных данных для совершенствования измерений миграции. Кроме того, участники отметили потребность в практических рекомендациях и руководстве по вопросам подготовки метаданных для упрощения сопоставлений между оценочными данными по миграции, составленными в разных странах.

6. В заключение участники рабочей сессии рекомендовали продолжать методологическую работу по теме интеграции множественных источников данных для измерения международной миграции, включая источники данных внутри одной страны и между разными странами, и передовой практики в сфере коммуникации между национальными статистическими учреждениями и производителями административных данных.

7. Опираясь на результаты этих дискуссий, в октябре 2015 года Бюро Конференции европейских статистиков (КЕС)<sup>4</sup> учредило Целевую группу по интеграции данных для измерения миграции с целью подготовки свода руководящих принципов и описания передовой практики интеграции различных источников данных для совершенствования измерений иммиграции, эмиграции и чистой миграции. В настоящей публикации представлены результаты работы Целевой группы.

8. В настоящей публикации дается общий обзор этой темы наряду с обзором видов интеграции данных, которые уже используются в различных странах, будь то разрозненные регулярно публикуемые данные или элементы пилотных исследований. На основе этого обзора излагаются также принципы передовой практики, призванные служить рекомендациями по совершенствованию интеграции данных для измерения миграции в различных странах. Опыт отдельных стран систематизирован в этой публикации на основе обследования поставщиков данных по миграции почти в 50 странах, а также на основе более подробных тематических исследований по ряду стран.

9. Вопрос об интеграции данных рассматривается в настоящей публикации путем анализа как *интеграции макроданных*, т. е. сопоставления/статистического моделирования на основе данных, представляющих собой конгломераты (статистику) индивидуальных регистрационных систем, так и *интеграции микроданных*, т. е. интеграции данных на основе увязывания/сопоставления отдельных записей. Наряду с этим анализируются различные уровни дублирования переменных и/или индивидуальных данных из разных источников.

10. Последующие детали, связанные в вышеуказанными понятиями, изложены в виде элемента концептуальной основы и определения термина «интеграция данных» в главе II. Глава III содержит анализ результатов обследования поставщиков данных по миграции. Практические примеры интеграции данных по миграции приведены в рамках углубленных тематических исследований, выполненных в ряде стран. Обобщенный результат этих тематических исследований содержится в главе IV

<sup>4</sup> В состав Конференции европейских статистиков входят национальные статистические организации региона ЕЭК ООН (по странам – членам ЕЭК ООН см. [www.unece.org/oes/nutshell/member\\_states\\_representatives.html](http://www.unece.org/oes/nutshell/member_states_representatives.html)), а также Австралия, Бразилия, Китай, Колумбия, Мексика, Монголия, Новая Зеландия, Республика Корея, Чили и Япония. В ее работе также участвуют крупные международные организации, занимающиеся вопросами статистики в регионе ЕЭК ООН, такие как Статистическая служба Европейской комиссии (Евростат), Организация экономического сотрудничества и развития (ОЭСР), Межгосударственный статистический комитет Содружества Независимых Государств (СНГ-СТАТ), Международная организация труда (МОТ), Международный валютный фонд (МВФ), Всемирная торговая организация (ВТО) и Всемирный банк.

(тематические исследования включены в полную версию настоящего документа, которая издана только на английском языке). Предложение по метаданным, которое предстоит рассмотреть в связи со статистикой международной миграции, представлено в главе V. Наконец, выводы и общие рекомендации, сформулированные на основе итогов обследования и тематических исследований, изложены в главе VI.

## II. Определение термина «интеграция данных»

### A. Общие характеристики

11. Уже давно известно, что измерение миграции является весьма сложной задачей. Хотя в последнее время как в ученых кругах, так и в сфере официальной статистики предпринимались значительные усилия для повышения качества статистики миграции, точность этих данных по-прежнему имеет определенный предел погрешности. Позднее в более широком статистическом контексте внимание обращалось на перспективы, связанные с «интеграцией данных» как с возможным дополнительным методом совершенствования статистики (например, Евростат в 2009 и 2013 годах; ESSnet в 2008, 2011 и 2014 годах; ГВУ-МОС ЕЭК ООН в 2017 году).

12. Интеграция данных зачастую упоминается как один из наиболее эффективных подходов, способствующих обогащению и совершенствованию статистической информации. В последние годы с интеграцией данных был связан целый ряд видов деятельности и проектов, в рамках которых использовались различные определения этого термина. Например, согласно Типовой модели производства статистической информации (ТМПСИ), интеграция данных – это один из методов производства статистической информации, который заключается в объединении данных из одного или нескольких источников. В вышеупомянутом проекте ГВУ-МОС интеграция данных определяется как «метод, при котором данные, поступившие как минимум из двух разных источников, комбинируются в одном наборе данных»<sup>5</sup>. Однако, насколько нам известно, в статистическом сообществе не существует определения интеграции данных, согласованного на международном уровне. Есть целый ряд аспектов, помогающих охарактеризовать интеграцию данных: количество и типология источников данных, методология, своевременность, нагрузка на респондентов и, не в последнюю очередь, цель ее применения.

13. Интуиция подсказывает, что для интеграции требуется наличие как минимум двух исходных материалов. Такими материалами являются наборы данных, т. е. систематизированная информация, взятая из определенных источников данных в рамках статистической операции. Применительно к миграции наборы данных можно распределить по четырем категориям: (данные, источником которых являются) статистические обследования (в том числе сплошные обследования, т. е. переписи), административные реестры, большие данные и географическая информация. Указанные типологии способны сочетаться во всевозможных комбинациях, и эту процедуру можно повторять, формируя наборы данных, которые затем будут интегрированы в другие наборы данных.

14. Метод интеграции наборов данных зависит в первую очередь от того, какой является содержащаяся в них информация – микро- или макроуровневой, а в случае отдельных записей – от наличия единого идентификатора. Для интеграции на микроуровне от этого последнего аспекта зависит выбор между статистическим согласованием и увязкой записей; метод увязки записей обычно применяется при наличии идентификатора во всех интегрируемых наборах данных. При интеграции на макроуровне, т. е. между данными, собранными из отдельных записей, соответствующая процедура проводится в два этапа: во-первых, устранение различий в понятиях и функциональных определениях, а во-вторых – согласование

<sup>5</sup> <https://statswiki.unece.org/pages/viewpage.action?pageId=169018059>.

(«уравновешивание») данных с использованием различных статистических методов или только на основе экспертного мнения.

15. Применительно к данным о миграции индивидуальным идентификатором является персональный идентификационный номер (ПИН). На практике метод микроуровневой интеграции предпочтителен для составления регистров населения – процесса, производной которого зачастую является статистика миграции. Следует отметить, что ПИН – это не единственный возможный единый идентификатор, поскольку при увязке данных, полученных в результате субнационального агрегирования, от административных образований, из файловых записей и т. д., могут использоваться разные идентифицирующие коды.

16. Источники данных, которые используются в качестве исходных материалов для составления регистра населения, могут быть многочисленны, и в некоторых странах со временем были интегрированы дополнительные источники данных, обогащающие статистическую информацию о населении и косвенно – о миграции. Аналогичным образом, хотя обычно в меньшей степени, может обстоять дело с регистрами иностранного населения/registрами иностранцев. Таким образом, интеграцию данных можно рассматривать как расширяющийся процесс, в ходе которого к первым интегрированным наборам данных добавляются другие данные, полученные благодаря приобретенному опыту, совершенствованию методов интеграции и доступу к новым источникам данных.

17. Еще одной отличительной чертой измерения миграции является то, что, учитывая характер этого явления, наблюдающегося между двумя странами (страной приема и страной происхождения, либо страной назначения и страной рождения, либо других переменных для идентификации мигрантов), могут использоваться другие источники данных, кроме национальных. Использование этого характерного международного фактора ограничено в основном обменом агрегированными данными, учитывая серьезную озабоченность соблюдением требований конфиденциальности и национальной безопасности, которыми страны могут объяснять свой отказ на любой запрос об обмене микроданными по международной миграции. Примеры применения этого последнего подхода имеются в северных странах, где международный обмен микроданными используется для повышения качества регистров населения.

18. Добавление новых источников данных может приводить к изменению своевременности статистической информации, т. е. промежутка времени между контрольным периодом и получением результатов. Хотя по общему правилу своевременность того или иного конкретного продукта из комплексных данных определяется источником данных, имеющим более широкий временной охват, статистическое моделирование может в конечном счете преодолеть эту ситуацию, открывая возможность для представления более своевременных оценочных показателей. По-видимому, снижение уровня своевременности будет приемлемым лишь в тех случаях, когда интеграция данных обеспечивает соответствующее повышение охвата и точности статистических измерений.

19. В качестве одной из возможностей, создаваемых благодаря интеграции данных, иногда упоминается сокращение нагрузки на респондентов. На сегодняшний день большинство видов использования альтернативных источников данных для нужд статистики миграции обеспечивается путем замены одного источника данных другим, с меньшей нагрузкой на респондентов/поставщиков данных, или путем совершенствования статистической операции, применяемой к соответствующему источнику данных. С концептуальной точки зрения этот метод можно рассматривать скорее как повышение эффективности, нежели как расширение спектра имеющихся данных, вследствие чего он не вполне характерен/актуален для интеграции данных.

20. Исходные материалы, взятые из альтернативных источников данных, можно использовать для подтверждения статистической информации из отдельного, официального источника данных. В этом случае применяемый метод можно считать *сопоставлением*, а не *интеграцией*. Он также может рассматриваться в качестве первого шага в процессе постепенной интеграции используемых источников данных,

особенно когда дополнительные источники данных не отвечают требованиям официальной статистики.

21. Еще один случай – это охват каждым источником данных определенной подгруппы населения (мигрантов). Объединение таких данных для получения общей статистической информации можно охарактеризовать как компиляцию, а не как чистую интеграцию, поскольку фактического объединения на уровне отдельной записи не происходит (либо индивидуальные, либо агрегированные образования).

22. Из вышеизложенных соображений следует, что квалифицирующим признаком интеграции данных, бесспорно, является использование множественных наборов данных, но в зависимости от метода их совокупного использования. На практике существуют статистические операции, которые в некоторой степени представляют собой «пограничные» случаи с интеграцией данных, такие как компиляция данных и сопоставление данных. Что касается других признаков, то степень детализации данных имеет значение скорее для выбора методологии, которая будет применена, нежели для идентификации вида деятельности по интеграции данных. Своевременность также нельзя использовать в качестве критерия идентификации, поскольку интеграция данных в принципе способна как сократить, так и удлинить задержку в получении результатов, причем позитивное влияние на сроки можно оказывать и с помощью применения других статистических операций, кроме интеграции. Аналогичным образом, не всякая типовая ссылка на повышение качества статистической информации (включая снижение нагрузки на респондентов) будет полезной, поскольку в принципе каждая статистическая операция должна быть нацелена на достижение такого результата. Очевидно, это не исключает, что метод интеграции данных может обеспечивать подобный позитивный эффект, а лишь означает, что такие ссылки нельзя использовать для выявления отличительных признаков интеграции данных.

23. Результатом мер по интеграции данных должен являться «обогащенный» или «высококачественный» набор данных. Расчет, производимый в рамках типовой матрицы данных  $n \times p$ , где  $n$  – это записи, а  $p$  – переменные, может обеспечить такое обогащение в обоих направлениях, расширяя охват (т. е. увеличивая  $n$ ) и/или объем информации в тех же записях (т. е. добавляя новые переменные в набор  $p$ ). Что касается повышения качества, то оно найдет отражение не в изменениях объема получаемого набора данных, а скорее в его содержании (например, в весах выборочного обследования).

## **В. Рабочее определение**

24. Имеющееся определение термина «интеграция данных» заимствовано из ИТ-среды (ОСДМ, 2009 год) и сформулировано следующим образом: *«процесс комбинирования данных из двух или более источников для подготовки статистической информации»*. В этой же ссылке уточняется, что *«интеграция данных может осуществляться на микроуровне, где она часто квалифицируется как сопоставление, или на макроуровне»*.

25. В основу этого определения заложены исходные материалы (*«двух или более источников»*) и цель (*«для подготовки статистической информации»*). Последний элемент носит довольно типовой характер, поскольку получение статистической информации можно считать итогом любого статистического процесса, не обязательно требующего интеграции данных.

26. Для целей настоящего документа предлагается следующее рабочее определение: *«Интеграция данных – это вид статистической деятельности с использованием двух или более наборов данных, результатом которого является получение единого расширенного и/или более качественного набора данных»*.

27. Интеграция данных может производиться на двух основных уровнях агрегации – микро- и макроинтеграции<sup>6</sup>, которые определяются следующим образом:

а) «Интеграция микроданных» – это интеграция данных на основе увязки/статистического сопоставления отдельных записей с использованием ключевых идентифицируемых переменных.

б) «Интеграция макро данных» – это комбинирование данных на основе совокупностей (статистики) отдельных записей.

28. Ссылка на типовое понятие «статистическая деятельность» позволяет применять на макро- или на микроуровне любую методологию, включая экспертное мнение. Это понятие также охватывает анализ концептуальных различий, являющийся одним из важнейших шагов любого статистического процесса. Прямое упоминание методов статистического согласования и/или увязки записей, хотя они составляют специфику интеграции данных, привело бы к исключению макроуровневой интеграции.

29. Выражение «двух или более источников» заостряет внимание на многоаспектном характере интеграции данных. Оно также указывает на то, что, хотя интеграция завершается достижением единого результата, она представляет собой процедуру, которую необходимо повторять каждый раз, когда искомая цель заключается в достижении такого результата. Иными словами, при регулярной подготовке статистических данных на основе более чем одного набора данных интеграция не является окказиональной, однократной статистической процедурой. «Интегрированный» результат генерируется всякий раз, когда появляются более свежие вводные данные<sup>7</sup>.

30. Употребление термина «наборы данных» вместо «источники данных» указывает на то, что предметом интеграции являются фактические данные, а не те источники данных, из которых они получены. Например, объединение двух имеющихся выборочных обследований в одно всеобъемлющее обследование путем изменения вопросников, плана выборки и т. п. не может рассматриваться в данном контексте как пример интеграции данных. Это также подтверждает вышеизложенную точку зрения, согласно которой интеграция не преобразует источники вводных данных в новые, смешанные источники данных; основное внимание в ходе этой процедуры уделяется результатам, полученным на основе этих источников данных. Таким образом, процедура интеграции носит, пожалуй, столь же систематический характер, что и подготовка статистической информации на основе данных, к которым она применяется.

31. Результатом интеграции данных является прежде всего «единый» набор данных, т. е. новый набор данных, в котором вся информация из наборов вводных данных реорганизована в единообразном виде. Это не просто объединение нескольких наборов данных, поскольку предполагается, что из них надлежащим образом устранены дублирования, концептуальные различия и любые другие проявления необъективности. Это – «расширенный» набор данных, т. е. структурированный новый набор данных, измерения которого не могут быть меньше самых крупных соответствующих измерений отдельных наборов вводных данных, и/или «высококачественный» набор данных, элементы которого были изменены в результате интеграции данных, обеспечившей (по возможности измеряемое) повышение качества статистической информации.

<sup>6</sup> В некоторых контекстах термин «микро-макро» употребляется при комбинировании микроуровневых и макроуровневых данных. Здесь этот подход может рассматриваться как один из случаев интеграции макро данных.

<sup>7</sup> Возможно, выявление концептуальных различий в контексте интеграции данных является исключением из «правила повторений», поскольку такая процедура потребует лишь один раз, если концептуальная основа вводимых данных со временем не изменится.



### III. Обследование методов интеграции данных

32. Чтобы собрать информацию, необходимую для выполнения этой работы, Целевая группа провела обследование методов интеграции данных. Был подготовлен специальный вопросник с целью сбора информации о нынешних методах использования комбинированных или комплексных источников данных для измерения иммиграционных и эмиграционных потоков и для подготовки статистики по группам мигрантов и некоренного населения.

33. В сентябре 2016 года этот вопросник был разослан национальным статистическим учреждениям стран – членов ЕЭК ООН<sup>8</sup>. Ответы на те части обследования, которые касались применяемых ими методов сбора данных, представили 56 стран.

34. Анализ итогов обследования был проведен за период с конца 2016 года по первый квартал 2017 года; его основные результаты обобщены в пунктах ниже.

#### A. Обзор основных результатов обследования

35. Для обеспечения подготовки статистической информации о международных миграционных потоках ключевую роль играет та или иная форма интеграции данных. Там, где это возможно, интеграция данных осуществляется на уровне отдельных записей, однако такой подход приемлем в основном для стран, имеющих полные регистры населения, в которых учитываются выезды своих граждан и въезды иностранцев.

36. Во многих странах фактическое миграционное событие не регистрируется в каком-либо одном источнике данных. Статистику миграционных потоков можно получить только путем объединения составных элементов, в которых учитываются изменения, происходящие в постоянном составе населения вследствие миграции.

37. Статистическая информация о международных миграционных потоках на основе одного источника производится при наличии возможности постоянно вести административный учет миграционных событий (фактических пересечений границы) или при наличии системы учета, основанной на полном регистре населения. Однако в ходе контроля качества и дальнейшей разбивки этой информации может потребоваться интеграция со вспомогательными источниками данных. Эта процедура может происходить в форме корректировки миграционных потоков после сопоставления с имеющейся статистикой, полученной из других стран (зеркальная статистика). Объединение лиц в рамках одного источника может потребоваться для продольного исследования трансграничных перемещений как метода определения статуса мигрантов.

38. В результате обследования было выявлено широкое разнообразие способов применения интеграции данных. В некоторых странах регистры учета граждан могут не обеспечивать увязку на уровне регистрационных единиц с какой-либо отдельной административной процедурой учета данных об иностранных гражданах. В таких случаях регистр населения может использоваться в качестве источника для перекрестной проверки и условного расчета отсутствующих данных.

39. Наиболее распространенными источниками для разработки статистических данных о некоренном населении будут являться перепись населения и регистр населения. Перепись населения может обеспечить информацию о населении в целом, а последующая интеграция с административными данными позволяет обновлять статистические данные о некоренном населении.

<sup>8</sup> Кроме 56 стран – членов ЕЭК ООН, вопросник был также направлен другим странам, регулярно участвующим в работе Конференции европейских статистиков, в том числе Австралии, Колумбии, Мексике, Новой Зеландии, Чили и Японии.

40. В случаях, когда миграционная статистика составляется на основе данных из нескольких источников, некоторые страны указали на необходимость оценки приоритетности источников с точки зрения их своевременности, надежности, последовательности и других критериев качества.

## **В. Рекомендации**

41. Поскольку страны занимаются развитием и стандартизацией регистров наборов статистических данных по социальной сфере и народонаселению, интеграция данных чаще осуществляется на уровне отдельных записей по различным секторам. Одним из неперенных элементов разработки статистики международной миграции наверняка станет уделение повышенного внимания оценке качества интегрированных данных.

42. Все более важным для статистических учреждений становится обеспечение транспарентного представления правил и методов, которые используются в процессе интеграции данных для разработки статистики международных миграционных потоков и других видов статистики международной миграции. Это открывает возможности для разработки в будущем методологических стандартов по способам интеграции данных на уровне национальных статистических учреждений. Отдельные страны всегда будут иметь собственные наборы данных и методы их интеграции, но наряду с этим, скорее всего, возрастет число открытых технических форумов для обсуждения правил и методов, используемых при комбинировании источников для выработки миграционной статистики.

## **С. Резюме результатов**

43. Почти все страны-респонденты занимаются выработкой статистики международных миграционных потоков, причем более половины из этих стран используют несколько источников данных. Одни страны опираются на центральный регистр учета всех сведений о регистрациях и deregistrations жителей, а последующая интеграция источника на уровне отдельных записей с административными наборами данных по эмиграционным/иммиграционным потокам и случаям смерти способствует уточнению информации о миграционных потоках. Другие страны прибегают к интеграции регистров населения с данными, полученными в результате обследований или других видов административного сбора информации о негражданах, проводимых на макроуровне.

44. В целом около трети стран-респондентов (всего 35 ответов) интегрировали источники данных на макроуровне с применением статистического моделирования или другого метода комбинирования агрегированных данных, а несколько меньшая доля стран использует методы интеграции на уровне регистрационных единиц. При этом следует отметить, что несколько стран пользовались другими источниками только для перекрестной проверки и дополнительной информации. Около половины стран-респондентов отметили, что между единицами наблюдения, взятыми из интегрированных источников данных, наблюдается частичное дублирование; у меньшей доли стран при комбинировании источников оказывались идентичные единицы наблюдения; а несколько стран сообщили об использовании взаимоисключающих единиц наблюдения.

45. Менее половины стран готовят статистическую информацию о международных миграционных потоках, используя только один источник. Подготовка статистики по миграционным потокам на основе единственного источника зачастую выполняется в случаях, когда есть возможность постоянно вести сбор административных данных на всех пунктах пересечения границы или когда миграционные потоки могут вычлняться из регистра населения. Еще один, реже встречающийся вариант – это использование данных переписи населения для косвенной оценки миграционных потоков в показателях численности населения.

46. Порядка 90% стран-респондентов представили статистику или собранные данные по некоренному населению. Лишь немногим более половины этих стран

используют более одного источника данных и обычно осуществляют интеграцию данных на уровне регистрационных единиц или на макроуровне. К числу основных видов и признаков используемых источников, будь то одиночные или интегрированные, относились проводимая каждые пять лет перепись населения, регистр населения, обследование или административный сбор данных. Сообщалось, что основной целью интеграции является подготовка статистики некоренного населения.

47. Около 60% стран-респондентов сообщили, что они не вырабатывают другой статистики по мигрантам или некоренному населению. Поэтому несколько стран указали, что интеграция источников данных дает возможность готовить такую статистическую информацию, как данные о поданных и удовлетворенных ходатайствах об убежище, разрешениях на проживание, разрешениях на работу, нелегальных иностранных работниках, социально-экономических характеристиках мигрантов и причинах их расселения.

## IV. Тематические исследования

48. С учетом информации, собранной в ходе обследования методов интеграции данных, Целевая группа выявила ряд стран, накопивших значительный опыт в сфере их интеграции. Этим странам было предложено представить краткие записки с указанием первичных источников данных о миграции, причин считать процесс интеграции данных необходимым, применяемых методов и выгод от интеграции.

49. Записки, полученные от этих стран, представлены в полной версии настоящего документа (имеется только на английском языке) в виде подборки значимых практических примеров интеграции данных по миграции.

### A. Резюме тематических исследований

50. В тематических исследованиях отражен широкий спектр ситуаций, которые существенно отличаются друг от друга с точки зрения имеющихся источников данных для статистики миграции, ограничений этих источников и применяемых подходов к интеграции данных. Примеры, которые приведены в этих тематических исследованиях, не нужно считать вполне репрезентативной выборкой подходов к интеграции данных, разработанных в различных странах. Кроме того, в процессе оценки опыта этого спектра стран следует учитывать, что национальные условия могут значительно различаться в плане количества важных аспектов и что какой-либо подход к интеграции данных, успешно примененный в одной стране, не обязательно обеспечит оптимальное решение другим странам. Несмотря на эти ограничения, анализ представленных тематических исследований позволяет выделить ряд основных аспектов, которые могут быть полезны для других стран.

51. Что касается *стран, имеющих регистр населения*, то в некоторых из них (в том числе в Австрии, Испании, Италии и Латвии) *этот регистр является главным источником данных о миграции*. В этих странах интеграция данных проводится прежде всего для повышения качества данных. Данные из других регистров могут использоваться, в частности, для корректировок на отсутствующие показатели deregистрации эмигрантов или для выявления умерших. В ряде стран интеграция данных используется также для обеспечения дополнительной информации по переменным, не включенным в первичные источники.

52. В некоторых странах применяется практика, основанная на *«сигналах присутствия»* (в ряде стран они называются «признаками жизни») лиц в разного рода регистрах. Например, в *Австрии* анализ сигналов присутствия в других регистрах проводится для расчета оценок числа лиц, которые фигурируют в регистре населения, но в стране больше не проживают. Данные из Управления социального обеспечения Австрии используются для выявления умерших, а из ряда других административных источников – для корректировок на отсутствующие показатели deregистрации эмигрантов.

53. В *Италии*, где в регистр населения могут вноситься только те лица, которые легально проживают в стране, анализ сигналов присутствия в различных источниках позволяет уточнять оценочные данные по международной миграции посредством выявления лиц, которые включены в регистр населения, но, вероятно, эмигрировали, а также лиц, которые иммигрировали, но не фигурируют в регистрах населения.

54. В *Латвии* анализ сигналов присутствия проводится с использованием различных регистров, включая регистры доходов, социального страхования, образования, здравоохранения и занятости. Оценочные показатели эмиграции рассчитываются также с помощью зеркальной статистики, поступающей из стран назначения. Кроме того, качество статистики миграции оценивается с использованием результатов обследований домохозяйств, включая ОРС и ОДУЖ-ЕС.

55. В *Испании* национальный регистр населения («padrón») находится в ведении национального статистического института INE, и в него должны заноситься все проживающие в стране лица независимо от их юридического статуса. При проведении в 2011 году переписи населения зарегистрированное население было увязано с многими другими административными регистрами с целью выявления лиц, чье присутствие в стране вызывает сомнения, а также для выявления умерших. Что касается сомнительных записей, то благодаря методам интеграции данных на основе административных данных и выборочному обследованию, проведенному в связи этой переписью, удалось оценить число лиц, которые фигурировали в регистре, фактически не проживая в стране.

56. В некоторых странах (включая Венгрию, Израиль, Нидерланды и Швейцарию) статистика миграции производится с применением *систем, где регистр населения является одним из нескольких источников данных*, причем не обязательно самым важным из них. В *Венгрии* используется довольно сложная система регистров, в которой различные источники дополняют друг друга. Показатели иммиграции и эмиграции рассчитываются путем интеграции данных из этого разнообразия источников, в частности из Управления по вопросам иммиграции и убежища и Центрального регистра населения, находящихся в ведении Министерства внутренних дел. Интеграция данных дает возможность выработать качественную статистику миграции и получать новую информацию в дополнение к той, которая поступила из первичных источников.

57. В *Израиле* данные о миграционных потоках и численности населения поступают главным образом из системы пограничной регистрации и передаются Управлением по вопросам народонаселения и иммиграции (которое также ведет регистр населения) в распоряжение Центрального статистического бюро. Эти данные объединяются с другими административными записями (в том числе по секторам налогообложения, образования и т. д.) и с результатами переписи населения и других статистических обследований, чтобы на их основе рассчитывались более качественные оценочные показатели и добавлялись переменные из других источников, связанные с теми же лицами или событиями.

58. В *Нидерландах* статистика миграции (как и вся социально-демографическая статистика) основана на системе баз данных социальной статистики, являющейся результатом интеграции данных из более чем 50 административных регистров. Считается, что зарегистрировано – лишь за несколькими исключениями – практически все проживающее в стране население. Что касается эмиграции, то около трети выезжающих из страны лиц не регистрируются. Когда органы власти обнаруживают отсутствие какого-либо лица, они проводят расследование, и, если разыскать это лицо не удается, оно может быть зарегистрировано как эмигрировавшее в неизвестную страну.

59. В *Швейцарии* к данным по миграции имеют отношение три основных регистра: 1) регистры местного населения, ведущиеся муниципальными властями; 2) федеральный регистр законно проживающих иностранцев, имеющих вид на жительство; и 3) особый федеральный регистр иностранных дипломатов, сотрудников консульств и международных организаций, членов их семей и аналогичных категорий лиц. В стране проводится несложная процедура интеграции данных, состоящая в

сопоставлении данных федерального уровня с данными, собранными на местном уровне. Кроме того, некоторые переменные, имеющиеся только в регистре законно проживающих иностранцев, добавляются к записям из регистра местного населения, имеющим отношение к тем же наблюдениям.

60. Что касается *стран, не имеющих регистра населения*, то в них к основным источникам данных о миграции относятся пассажирские формуляры, заполняемые на границе, паспортные данные, визы и виды на жительство. В странах, где ведутся регистры населения, интеграция данных производится с использованием различных источников в целях повышения качества данных и получения более полной информации. В частности, интеграция данных может способствовать уточнению сведений по иммигрантам в части их фактического места и даты размещения, которые могут отличаться от заявленных в момент прибытия.

61. В *Австралии* данные о въезде и выезде компилируются Австралийским статистическим бюро на основе информации, поступающей из различных систем обработки данных, включая пассажирские формуляры, паспортные данные и визовую информацию. Данные увязываются с помощью персонального идентификатора, который позволяет АСБ создавать личные учетные записи для проезжающих. Наборы данных по мигрантам также увязываются с данными переписей и с налоговой информацией.

62. В *Канаде* административные данные по иммигрантам и непостоянным жителям поступают в статистическое управление этой страны. Поскольку сведения о провинции, планируемой для проживания, и дате прибытия в нее не всегда соответствуют фактическому месту и дате размещения, производится интеграция данных с использованием других источников, таких как налоговые записи. Наряду с этим обеспечивается увязка записей с использованием данных из других источников (таких, как перепись населения Канады, Национальное обследование домохозяйств, Управление Соединенных Штатов по статистике миграции), позволяющая рассчитывать качественные оценочные показатели и получать более полную информацию по социально-экономическим характеристикам мигрантов.

63. В *Новой Зеландии* графические пассажирские формуляры считываются пограничными электронными системами учета передвижений и регистрации пассажиров, обеспечивая поступление дополнительной информации и проверку данных. Кроме того, в случае проезжающих, заявивших о намерении остаться в стране или выехать из нее на длительный период, путевые и личные пассажирские данные увязываются на срок 16 месяцев до контрольного месяца, чтобы на момент совершения поездки подтвердить статус соответствующего лица как жителя Новой Зеландии.

64. В *Соединенном Королевстве* имеется целый ряд источников официальной статистики, предоставляющих информацию по различным аспектам потоков международной миграции и численности мигрантов. Применительно к миграционным потокам главным источником является обследование международных пассажиропотоков (ОМП) – многоцелевое выборочное обследование, в ходе которого собирается информация от пассажиров, въезжающих в страну или выезжающих из нее. Поскольку при проведении ОМП не учитываются изменения в намерениях пассажиров, для расчета долгосрочных оценок международной миграции проводится интеграция данных с использованием информации, полученной в результате ОМП и ОРС, административных данных по иммиграции Министерства внутренних дел и оценочных данных о потоках на сухопутных границах Северной Ирландии (которые не охвачены ОМП).

65. В *Соединенных Штатах* Бюро переписей составляет ежегодные оценки международных миграционных потоков, объединяя различные субкомпоненты, получаемые из различных источников данных. В отношении некоторых субкомпонентов проводится ограниченная интеграция (на совокупном уровне) с использованием ряда иностранных и административных данных, в том числе результатов Обследования американского общества, зарубежных переписей и данных о перемещениях военнослужащих, полученных от Министерства обороны.

Интеграцию данных по международной миграции на макроуровне Бюро переписей не проводит, хотя и осуществило ряд экспериментов в этом направлении.

66. Если обобщить результаты тематических исследований, то они показывают, что практически все страны применяют интеграцию данных для повышения качества статистики миграции и для обеспечения дополнительной информации сверх той, которая поступает из первичных источников.

67. В странах, где ведется регистр населения, интеграция данных – хотя и посредством анализа «сигналов присутствия» – может обеспечить выявление: 1) лиц, которые эмигрировали, но по-прежнему числятся в регистре населения; 2) лиц, которые иммигрировали, но не внесены в регистр населения; и 3) умерших лиц.

68. В странах, не имеющих регистра населения, интеграция данных может способствовать повышению качества информации об иммигрантах в части их фактического места и даты расселения, которая нередко отличается от заявленной при въезде в страну.

69. В тематических исследованиях было приведено множество примеров интеграции данных, проводимой на микро- или макроуровне. Например, в Венгрии и Канаде различные источники увязываются на микроуровне. Такая микроинтеграция осуществляется путем увязывания административных источников данных об иностранцах как между собой, так и с регистром населения, чтобы не допускать дублирования и обеспечивать дополнительные переменные. С другой стороны, интересный пример макроинтеграции был обнаружен в Соединенном Королевстве, где проведение пограничного обследования пассажиров комбинируется путем калибровки с другим источником (обследованием рабочей силы) для разбивки прибывающих потоков по территориальному признаку.

70. Как и ожидалось, для интеграции данных используется широкий спектр видов источников, включая различные административные или статистические регистры, другие административные данные (в том числе данные пограничного контроля) и статистические обследования (в частности, результаты переписей). Некоторые страны для уточнения оценки масштабов своей эмиграции используют «зеркальную статистику» из стран назначения. Это происходит главным образом на совокупном уровне, хотя имеются примеры международного обмена микроданными, осуществляемого, в частности, между северными странами (не включены в тематические исследования) или ограниченного соседними странами.

## V. Метаданные

71. В процессе сопоставления оценок миграции из разных стран очень важно иметь доступ к всеобъемлющим метаданным, чтобы подробно проанализировать совместимость этих данных.

72. В идеале метаданные можно распределить по трем категориям: первая касается непосредственно каждого источника данных, описания их основных признаков и способов преобразования их наборов данных перед процедурой интеграции; ко второй относятся способы интеграции наборов данных, которые генерированы из перечисленных источников данных; а третья категория опирается на определенные измерения качества данных, прежде всего для подтверждения роста их информационного наполнения, достигнутого благодаря интеграции. Эта информация должна предоставляться по каждому соответствующему виду статистики миграции, т. е. по численности мигрантов, миграционным потокам и любой из их рассматриваемых подгрупп.

73. Первоочередной информацией для представления является перечень источников данных, которые используются для выработки интегрированных данных. В таком перечне каждый источник данных могут сопровождать следующие метаданные:

- a) название;
- b) типология (административный регистр, выборочное обследование, перепись, большие данные и т. п.);
- c) владелец (включая статистическое учреждение другой страны);
- d) правила доступа/предоставления данных (если владельцем является не национальное НСУ);
- e) контрольная группа населения;
- f) контрольная дата/контрольный период данных;
- g) частота обновления;
- h) своевременность (т. е. интервал между датой наличия данных и их контрольной датой/контрольным периодом);
- i) уровень детализации первоначальных данных (микро, макро);
- j) уровень детализации набора данных, используемого для интеграции (микро, макро);
- k) описание любого метода, применяемого для преобразования первоначального набора данных в исходный материал для интеграции данных;
- l) размеры набора данных (n x p);
- m) переменные, содержащиеся в источнике данных (или подробное описание, если их несколько);
- n) переменные, содержащиеся в источниках данных, которые отображены для интеграции (возможно, подраздел предыдущего пункта);
- o) переменные, используемые для интеграции этого источника данных с другим/другими (возможно, подраздел предыдущего пункта);
- p) наличие ПИН (в принципе, этот аспект включен в предыдущий пункт, однако, учитывая актуальность этого аспекта, его, вероятно, следует упомянуть отдельно) и, в соответствующих случаях, его основные признаки.

74. Поскольку процесс интеграции осуществляется поэтапно, для каждого этапа, т. е. для каждой пары интегрируемых наборов данных, можно указывать следующую информацию:

- a) номер этапа;
- b) применяемые наборы данных (используя названия или числа из перечня отдельных источников данных);
- c) частота проведения процедуры интеграции;
- d) переменные, используемые для интеграции (если применимо);
- e) описание методологии интеграции наборов данных;
- f) доля дублируемых данных;
- g) основные проблемы/трудности при интеграции этих наборов данных, такие как сообщения об отсутствии увязки, методах и результатах оценки доли ложно-положительных увязок;
- h) размеры (n x p) полученного интегрированного набора данных.

75. В некоторых случаях, особенно когда статистика миграции вырабатывается на основе регистров населения, число источников данных может быть сравнительно высоким, и вследствие этого процедура сообщения метаданных может стать громоздкой. В таких случаях может быть целесообразным подготовить всего один общий отчет о функционировании системы. При этом следует учитывать, что для пользователей может иметь значение четкое понимание того, как производятся

окончательные данные в рамках внешне сложной системы анализа миграции. Сложность не должна препятствовать транспарентности.

76. Третья часть метаданных, требующая большой дальнейшей работы, может выстраиваться на основе следующего первоначального набора количественных измерений качества, направленного на выявление позитивного эффекта интеграции данных:

a) разница в количестве записей/численности охватываемого населения между окончательным интегрированным набором данных и наименьшим отдельным набором данных:  $n^* - \min(n)$ ;

b) разница в количестве переменных между окончательным интегрированным набором данных и наименьшим отдельным набором данных:  $p^* - \min(p)$ ;

c) разница в своевременности между окончательным интегрированным набором данных и наименее своевременным отдельным набором данных:  $t^* - \max(t)$ ;

d) разница в количестве записей/численности охватываемого населения между окончательным интегрированным набором данных и наибольшим отдельным набором данных:  $n^* - \max(n)$ ;

e) разница в количестве переменных между окончательным интегрированным набором данных и наибольшим отдельным набором данных:  $p^* - \max(p)$ ;

f) разница в своевременности между окончательным интегрированным набором данных и наиболее своевременным отдельным набором данных:  $t^* - \min(t)$ ;

g) повышение качества оценок, такое как сокращение доли вариативности оценочных данных или любое количественное измерение повышения точности.

77. Метаданные могут обогащаться качественной оценкой эффекта от интеграции, такой как отзывы в связи с улучшением первоначальных источников данных и уменьшение нагрузки на респондентов, по возможности подкрепленной конкретными ссылками/примерами, а также общим выводом о процедуре интеграции.

## VI. Заключение

78. Статистика миграции – это, по-видимому, наиболее трудный элемент социальной статистики, причем не только с практической, но и с концептуальной точки зрения. Несмотря на имеющиеся на международном уровне определения понятия «мигрант», установить его точный охват не всегда легко. С другой стороны, формы миграции становятся все более разнообразными, а источники, дающие частичные измерения миграционных процессов, весьма многочисленны.

79. Задача статистических учреждений состоит в том, чтобы объединять данные из разных источников в надежную комплексную информацию о миграции. В 13 рассмотренных тематических исследованиях показаны различные способы и источники данных, которые используются странами для совершенствования их миграционной статистики.

80. Целевая группа провела обследование с целью оценки потенциала большого числа различных источников для выработки миграционной статистики. Согласно полученным ответам, большинство стран (31 из 56 ответов) используют *более одного источника, чтобы генерировать статистику миграции*. Применяемые подходы неодинаковы, поскольку обстоятельства и источники данных весьма вариативны в зависимости от страны. Общая тенденция задействовать больше административных источников для измерения миграции прослеживается как в странах, чья статистика народонаселения опирается на классические методы (обследования, сплошные переписи), так и в странах, ведущих статистику на основе регистров населения.



81. Выбор источников для получения данных о миграции нередко зависит от *географии*. Страны, которые являются довольно изолированными или имеют небольшое число пунктов въезда и выезда (например, Новая Зеландия и Мальта), могут вырабатывать статистику миграции на основе совсем иных источников, нежели другие страны с многочисленными пунктами доступа или без контрольно-пропускных пунктов, как страны Шенгенской зоны в Европе.

82. Еще одним важным ограничивающим фактором при исследовании статистики миграции является *разнообразие определений* мигрантов в различных административных источниках. Во многих случаях при подготовке статистических данных о миграции используется не стандартный срок фактического или планируемого пребывания продолжительностью 12 месяцев, а применяются другие концепции (например, другие периоды времени). Кроме того, в одной и той же стране различные административные источники могут давать информацию, основанную на разных концепциях.

83. В странах Европейского союза статистика миграции призвана *полностью отражать изменения в численности мигрантов*. Во многих других случаях перед статистикой миграции не ставится задача обеспечения полного соответствия между потоками и численностью. Кроме того, многие страны рассчитывают отдельные показатели по долгосрочным мигрантам, просителям убежища, непостоянному иностранному населению, краткосрочным мигрантам и т. д.

84. В сфере интеграции данных имеется ряд соответствующих примеров комбинирования данных *как на макро-, так и на микроуровнях*. Что касается микроинтеграции, или объединения источников на уровне отдельных записей, то есть несколько примеров, способных проиллюстрировать эту тенденцию на перспективу. Такое комбинирование записей может приводить к случаям дублирования, которые устраняются в ходе процессов корректирования и калибровки. Эта интеграция проводится для совершенствования измерения численности мигрантов, а также для выработки более полной статистики миграционных потоков.

85. Основной недостаток регистров населения заключается в том, что с их помощью *трудно выявлять случаи эмиграции*. Если какое-либо лицо не информирует органы власти о своем выезде из страны, административный регистр следует обновлять посредством административных процедур, которые не всегда носят простой или незамедлительный характер; в результате в регистрах населения могут фигурировать лица, которые больше не проживают в стране.

86. При комбинировании источников могут обнаруживаться *«сигналы присутствия»*. Так, например, регистр населения может увязываться с налоговой документацией или файлами социального обеспечения, чтобы обеспечивать высокую степень уверенности относительно присутствия соответствующих лиц в стране. Это особенно важно при комбинировании различных административных записей в таких странах, как Латвия или Австрия, где при оценке проживающего населения на основе сведений из различных административных регистров применяются логистические регрессии. Анализ «сигналов присутствия» также может способствовать выявлению лиц, которые иммигрировали, но при этом не внесены в регистр населения (в Италии), и умерших лиц (в Австрии и Испании).

87. Что касается статистики миграционных потоков, то есть ряд примеров интеграции в строгом смысле этого термина, т. е. комбинирования источников для получения единой базы данных. Например, в Венгрии и Канаде различные источники увязаны на микроуровне. Такая микроинтеграция проводится путем увязывания информации об иностранцах между различными административными источниками и регистром населения, чтобы не допускать дублирований и обеспечивать дополнительные переменные.

88. Одной из инициатив, которая начинает предъявляться пользователям и представляется весьма многообещающей, является *обмен данными индивидуального уровня* между странами для статистических целей. Главная предпосылка для измерения миграционных потоков заключается в том, что они воздействуют на две страны – на страну происхождения и на страну назначения. Поэтому логично полагать,

что обмен статистическими данными позволит улучшить статистику этого явления. Обмен индивидуальными данными для статистических целей наверняка получит дальнейшее развитие, хотя для этого потребуются преодолевать препятствия, причем не только правового или технического характера. Удачный пример такого обмена показывают северные страны.

## **VII. Рекомендации**

89. На основе результатов работы Целевой группы и содержания настоящего документа в этом разделе приводится ряд рекомендаций для национальных статистических учреждений. Эти рекомендации считаются актуальными для всех стран, независимо от их нынешней ситуации с точки зрения наличия и использования административных данных для статистики миграции.

### **A. Улучшить доступ национальных статистических учреждений к административным данным**

90. В странах, где административные записи недоступны или лишь частично доступны статистическим учреждениям для подготовки официальной статистики, одним из приоритетов является принятие мер по расширению доступа к административным записям и расширения их использования статистическим учреждением. Для этого может потребоваться поддержка на политическом уровне, а в некоторых случаях – внесение изменений в законодательство.

### **B. Использовать административные данные для статистики миграции**

91. Многие административные источники содержат информацию, которую можно использовать в сфере статистики миграции. К ним относятся регистры и базы данных о населении, визах, паспортах, пограничном контроле, разрешениях на проживание и на работу, просителях убежища и беженцах. Рекомендуется рассмотреть потенциал всех этих административных источников с точки зрения статистики миграции и использовать их при подготовке статистической информации по миграции в той мере, в какой они подходят для этой цели. Для этого потребуется сотрудничество с профильными министерствами, учреждениями и административными органами, занимающимися подготовкой данных.

92. Рекомендуется выработать различные статистические продукты, содержащие информацию о численности мигрантов, а также о масштабах иммиграции и эмиграции, охватывая при этом разные виды миграционных потоков. В некоторых случаях различные виды статистики могут не обеспечивать полностью согласованных данных. При этом их опубликование может помочь в формировании общей картины и в осознании того факта, что отдельный продукт не дает всей необходимой информации о международной миграции. Простое опубликование данных из различных источников – это первый механизм интеграции информации по мигрантам.

### **C. Комбинировать данные из различных источников с использованием «сигналов присутствия»**

93. В странах, где административные данные доступны для целей статистики, комбинирование источников с использованием «сигналов присутствия» представляет собой весьма перспективный метод. Общей чертой этих разных подходов является использование ряда источников для оценки вероятности проживания в стране тех или иных лиц на основе их присутствия в соответствующих источниках. Такое комбинирование данных из разных источников позволяет повысить качество оценок нынешнего населения и численности иностранцев.

#### **D. Следить за качеством интегрированных данных**

94. Поскольку страны занимаются развитием и стандартизацией регистров наборов статистических данных по социальной сфере и народонаселению, весьма вероятно, что интеграция данных будет осуществляться на уровне регистрационных единиц с привлечением нескольких источников. Одним из неперенных элементов разработки статистики международной миграции наверняка станет уделение повышенного внимания оценке качества интегрированных данных.

#### **E. Обеспечивать прозрачность применяемых методов интеграции данных и разрабатывать стандарты**

95. Все более важным для статистических учреждений становится обеспечение транспарентного представления правил и методов, которые используются в процессе интеграции данных для разработки статистики международных миграционных потоков и других видов статистики международной миграции. Это открывает возможности для разработки в будущем методологических стандартов по способам интеграции данных на уровне национальных статистических учреждений. Отдельные страны всегда будут иметь собственные наборы данных и методы их интеграции, но наряду с этим, скорее всего, возрастет число открытых технических форумов для обсуждения правил и методов, используемых при комбинировании источников для выработки миграционной статистики.

#### **F. Развивать международное сопоставление данных о миграции и обмен ими**

96. Еще одним способом совершенствования статистики можно считать международное сопоставление и обмен данными по миграции, некоренному населению и иностранцам. В административных источниках масштабы эмиграции обычно недооцениваются из-за отсутствия у эмигрантов заинтересованности регистрировать такое событие, как выезд из страны. Однако эти события могут отражаться в иммиграционных данных тех стран, куда отправляются эмигранты. Соответствующую проверку можно проводить с использованием так называемой «зеркальной статистики», которая показывает миграционный поток из страны А в страну Б под углом зрения эмиграционной статистики страны А и иммиграционной статистики страны Б. Кроме того, данные из стран назначения могут предоставлять информацию о количестве и характеристиках эмигрантов и граждан той или иной страны происхождения. Поэтому обмен данными с другими странами может быть очень полезен для осмысления миграционных процессов и совершенствования их систем измерения миграции.

97. Весьма эффективным подспорьем в этой связи являются международные источники данных. Что касается Европы, то Евростат публикует вполне подробные данные об иммиграции и эмиграции, которые делают возможным сопоставление между странами Европейского союза. В своей информационной базе (базе данных) по статистике миграции ЕЭК ООН публикует такие данные из стран Восточной Европы, Кавказа и Центральной Азии.

98. Прорыва в этом направлении можно достичь в случае, если страны будут обмениваться базами данных с увязкой на уровне индивидуальных данных. Уже имеются реальные прецеденты такого взаимодействия, например среди северных стран. На сегодняшний день маловероятно, что обмены данными индивидуального уровня получат всеобщее распространение, однако возможна реализация ряда инициатив между странами ввиду интенсивности их обменов или иных обстоятельств. Успешное проведение таких обменов могло бы заинтересовать других в том, чтобы последовать этому примеру.

## VIII. Дальнейшая работа

99. На международном уровне дальнейшую работу можно было бы сосредоточить на рассмотрении потенциальных возможностей использования больших данных для разработки миграционной статистики и на возможных путях включения анализа продольных данных в ежегодную статистику миграции и дополнения имеющихся поперечных измерений.

### A. Потенциал больших данных

100. Новые, нетрадиционные источники данных – такие как мобильные телефоны, кредитные карты и социальные сети, в целом известные как большие данные, – могли бы стать полезными для статистики миграции. В настоящее время доступ статистических учреждений к таким данным весьма затруднен, поскольку в большинстве случаев они являются частной собственностью. В ближайшие годы нужно принять меры по обеспечению доступа к таким источникам и найти способы их использования в целях официальной статистики. В одиночку эта информация, вероятно, окажется недостаточной и необъективной, но после интеграции с традиционными источниками она позволила бы уточнить оценочные данные по миграции. Будет важно поделиться этими появляющимися методами на международном уровне, чтобы поддержать первые шаги стран по пути использования потенциала таких данных для статистики миграции.

### B. Продольные данные

101. Возможно, будущее интеграции данных – это не только повышение качества измерения миграционных потоков. Интеграция данных может обеспечивать наличие новых данных о мигрантах. Перспектива увязки баз данных о населении (или результатов обследований) с регистрационными данными, в частности, в сферах занятости, социального обеспечения, налогообложения, здравоохранения и образования позволит обеспечивать в будущем более обширную и качественную информацию об условиях жизни или социальной интеграции мигрантов.

102. Показательным примером является информация о продольных исследованиях. Официальная статистика почти полностью сосредоточена на предоставлении поперечной информации о мигрантах. Например, практически не имеется сведений о продолжительности миграционных эпизодов (долго ли тот или иной мигрант проживает в стране приема) или об интервалах, имеющих большое значение при исследовании миграционных потоков: периоде времени между прибытием в страну назначения и получением работы, либо между потерей работы и эмиграцией. Продольные данные необходимы для рассмотрения вопроса о том, улучшаются или ухудшаются со временем социально-экономические последствия миграции. Они также необходимы для осмысления семейной миграции и таких новых явлений, как круговая миграция.

103. Обмен передовой практикой в области создания наборов данных продольного характера и их использования в регулярной статистике миграции станет подспорьем в работе НСУ. Последующим шагом может являться разработка рекомендаций о том, как обеспечить согласование такой статистики в международном масштабе.

## Ссылки

ESSnet (2008): Project on Integration of Survey and Administrative Data. Можно ознакомиться по адресу: [https://ec.europa.eu/eurostat/cros/content/isad-0\\_en](https://ec.europa.eu/eurostat/cros/content/isad-0_en).

ESSnet (2011): Project on Data Integration. Можно ознакомиться по адресу: [https://ec.europa.eu/eurostat/cros/content/data-integration\\_en](https://ec.europa.eu/eurostat/cros/content/data-integration_en).

ESSnet (2014): Project on Handbook on Methodology of Modern Business Statistics ('Memobust handbook'), Module on Macro Integration. Можно ознакомиться по адресу: [https://ec.europa.eu/eurostat/cros/content/macro-integration\\_en](https://ec.europa.eu/eurostat/cros/content/macro-integration_en).

Eurostat (2009): "Insights on Data Integration Methodologies". Proceedings of the ESSnet-ISAD workshop, Vienna, 29–30 May 2008. Eurostat Methodologies and Working Papers. Можно ознакомиться по адресу: <http://ec.europa.eu/eurostat/en/web/products-statistical-working-papers/-/KS-RA-09-005>.

Eurostat (2013): "Statistical matching: a model based approach for data integration". Eurostat Methodologies and Working Papers. Можно ознакомиться по адресу: <http://ec.europa.eu/eurostat/en/web/products-statistical-working-papers/-/KS-RA-13-020>.

SDMX (2009): "SDMX Content-Oriented Guidelines – Annex 4: Metadata Common Vocabulary". Можно ознакомиться по адресу: <http://www.sdmx.org>.

Группа высокого уровня ЕЭК ООН по модернизации официальной статистики (2017 год): «Углубленный анализ темы "Интеграция данных"». Документ ECE/CES/2017/8, подготовленный для сессии Конференции европейских статистиков, состоявшейся 19–21 июня 2017 года.

---