



**ЭКОНОМИЧЕСКИЙ
И СОЦИАЛЬНЫЙ СОВЕТ**

Distr.
GENERAL

ECE/CES/AC.6/2008/6
3 March 2008

RUSSIAN
Original: ENGLISH

ЕВРОПЕЙСКАЯ ЭКОНОМИЧЕСКАЯ КОМИССИЯ

КОНФЕРЕНЦИЯ ЕВРОПЕЙСКИХ СТАТИСТИКОВ

Совместное совещание ЕЭК ООН/Евростата
по переписям населения и жилищного фонда

Одиннадцатое совещание
Женева, 13-15 мая 2008 года
Пункт 2 предварительной повестки дня

ОБЕСПЕЧЕНИЕ И ОЦЕНКА КАЧЕСТВА ПЕРЕПИСИ

Оценка качества проводимой на основе регистров словенской переписи 2011 года

Записка Статистического управления Республики Словении

Резюме

Использование административных данных в целях снижения издержек и нагрузки получило широкое распространение в современной официальной статистике. В Словении частично на основе административных данных перепись проводилась уже в 2002 году, а следующую перепись в 2011 году планируется проводить полностью на основе регистров. Оценка качества переписи и ее результатов является важной обязанностью статистических управлений. В настоящем документе рассматриваются изменения в концепциях оценки качества при переходе от традиционной переписи к регистровой. Первая часть документа посвящена некоторым общим соображениям, а вторая – планам в отношении оценки качества словенской переписи 2011 года, которая будет опираться на регистры.

I. ВВЕДЕНИЕ

1. Использование административных данных, особенно данных различных регистров, находит все более широкое распространение в различных областях официальной статистики. Административные данные уже давно находят применение в некоторых статистических процессах, например в процессе составления выборок или процедурах калибровки, однако в последние годы сфера их использования стала быстро расширяться, и они превратились в непосредственный источник данных. Конечно же, не стали исключением и переписи населения и жилищного фонда, которые, возможно, являются самым дорогостоящим и обременительным видом статистического обследования.

В последние десять лет различные статистические управления проводят широкомасштабную работу по созданию адекватных условий и разработке эффективных процедур сбора как можно большего объема данных из уже существующих административных источников. Ведущую роль в этой области играют страны Северной Европы, где процесс перехода от обычной методики к переписям, основанным на регистрах, уже практически завершен.

2. Статистическое управление Республики Словении (СУРС) уже давно использует североевропейскую модель широкого и комплексного использования административных данных. Многие административные регистры, которые сегодня находятся в ведении других органов, создавались и в течение некоторого времени велись СУРС. Для СУРС это является большим преимуществом, поскольку оно хорошо знакомо с концептуальными особенностями этих регистров, что, вполне естественно, позволяет ему широко использовать их в различных статистических целях. В отличие от последней проводившейся в Словении переписи населения и жилищного фонда, которая носила комбинированный характер и в которой использовались данные как регистров, так и обычной регистрации на местах, уже сегодня известно, что в основе следующей переписи 2011 года будут лежать исключительно данные регистров. На настоящем этапе сохраняется ряд сложных проблем, которые должны быть решены до проведения такой переписи. В то же время хочется надеяться на то, что стоящие проблемы будут решены и что сформулированные амбициозные планы будут реализованы.

3. Одной из важных обязанностей и функций органов, занимающихся подготовкой и представлением данных, являются регулярный и систематический контроль качества своей работы и представление соответствующей отчетности. Евростат совместно с государствами-членами разработал получившие широкое признание концептуальные рамки оценки и отчетности о качестве данных, в основе которых лежат шесть составляющих качества (актуальность, точность, своевременность и пунктуальность, доступность и ясность, сопоставимость и последовательность), дополняемых такими

компонентами, как издержки и нагрузка, связанные с подготовкой данных. Следует отметить, что эти рамки с заложенными в них концепциями разрабатывались главным образом для целей классических статистических обследований. В последнее время многие специалисты указывают на отсутствие теоретических рамок для систематической оценки качества статистических данных, заимствованных из регистров, и эта проблема потребует своего решения уже в самом ближайшем будущем.

4. В настоящем документе высказываются некоторые соображения относительно оценки качества данных следующей (в основу которой, хотелось бы надеяться, лягут исключительно данные регистров) переписи и представления соответствующей отчетности. В отличие от отчета о качестве словенской переписи населения и жилищного фонда 2002 года (который был размещен в открытом доступе на сайте СУРС), в основу которого в той или иной степени легли классические принципы отчетности, в ходе оценки и публикации данных о качестве переписи 2011 года акцент, без сомнения, сместится в сторону других аспектов качества.

5. В первой части документа излагаются некоторые общие соображения относительно оценки качества полученной на основе регистров статистики, а вторая часть в основном посвящена существующим в Словении реалиям. Поскольку СУРС только начало подготовку к проведению следующей переписи, принципы оценки качества и представления отчетности также находятся на этапе проработки и, безусловно, станут предметом более детального изучения.

II. ОЦЕНКА КАЧЕСТВА: ОТ КЛАССИЧЕСКОЙ СТАТИСТИКИ К СТАТИСТИКЕ, ОСНОВАННОЙ НА РЕГИСТРАХ

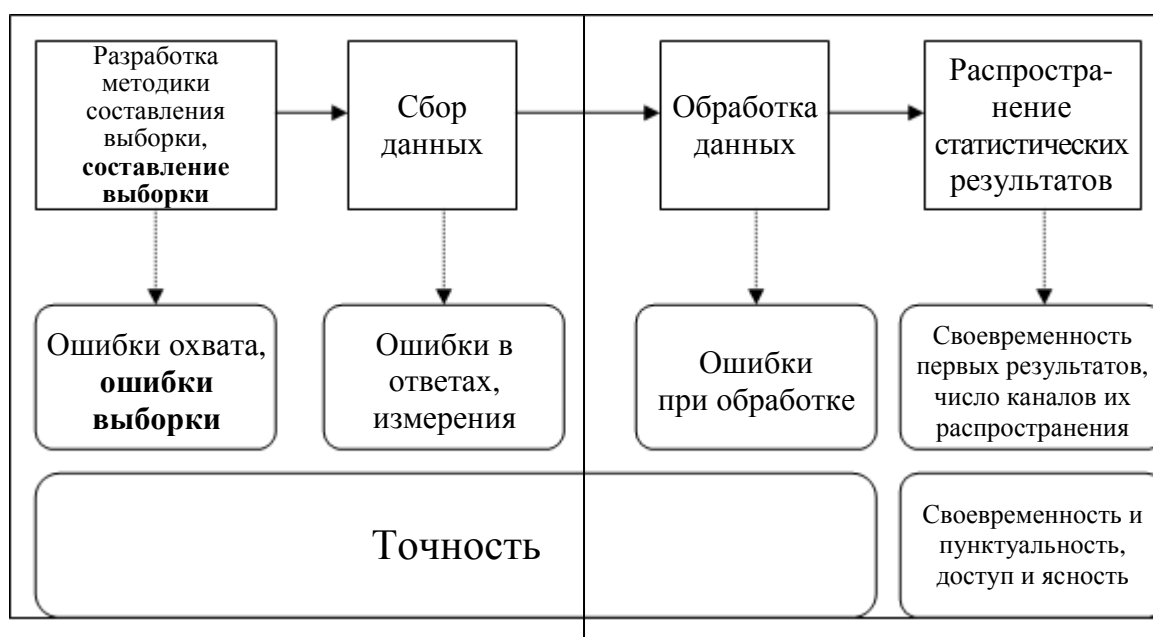
6. В современных моделях оценки качества статистический продукт рассматривается как результат сложного статистического процесса, и, оценивая качество статистического продукта, крайне важно контролировать и оценивать не только конечный продукт, но и его различные составляющие. Общие рамки оценки качества, которые в последние десять лет получили широкое признание в европейской статистической системе, представляют собой мощный инструмент проведения стандартной и последовательной деятельности в этой области.

7. Данная деятельность опирается на две важнейшие концепции: стандартные составляющие качества и стандартные показатели качества для каждой такой составляющей. Эти концепции могут быть задействованы в разных моделях, каждая из которых предполагает определенную степень упрощения. Для целей настоящего документа шесть составляющих качества вначале разбиваются на две части. В первую

часть входят составляющие, отражающие качество самого продукта, а во вторую – составляющие, в большей степени ориентированные на качество статистического процесса.

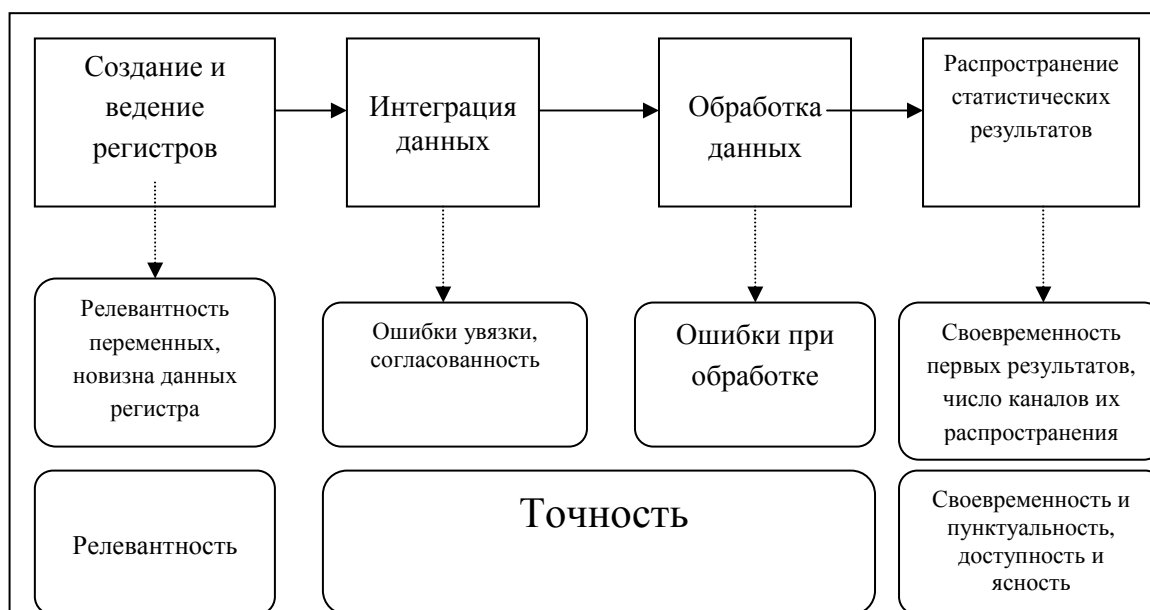
8. К характеризующим "продукт" составляющим относятся релевантность, согласованность и сопоставимость. Их важнейшей отличительной особенностью является то, что для их оценки недостаточно лишь наблюдения за статистическими процессами. Необходима также информация об общих условиях, например о потребностях пользователей, их удовлетворении, согласованности с некоторыми другими видами статистических данных и т.д. Еще одной их особенностью является то, что количественно оценить качество зачастую довольно сложно и что на основе этих составляющих разработано немного широко используемых в реальной жизни показателей качества.

9. К характеризующим "процесс" составляющим относятся точность; своевременность и пунктуальность; доступность и ясность. Эти составляющие можно более или менее точно оценить в ходе мониторинга разных этапов статистического процесса. "Классический" процесс обследования можно разбить на четыре следующих друг за другом этапа. Этот процесс отражен на (упрощенной) диаграмме, и для каждого из его этапов указаны наиболее типичные показатели, используемые при оценке качества.



10. Не будет больших различий, если применять эту модель к выборочному обследованию или к переписи. Единственным общим отличием является то, что выделенные жирным шрифтом компоненты не применимы к переписи, а все остальные могут применяться в обоих видах обследований.

11. Встает вопрос о том, как использовать эту модель при переходе от обычной методики переписи к переписям, основанным на данных регистров. Поскольку это влечет за собой по крайней мере частичное изменение подхода к процессу в целом, в наибольшей степени затронутыми оказываются те его составляющие, которые на диаграмме находятся слева от вертикальной пунктирной линии. Поскольку на смену разработанной специально в статистических целях классической процедуры сбора данных приходит интеграция данных из различных регистров, то модель оценки качества можно представить следующим образом:



12. Одно из важных различий между двумя подходами заключается в том, что релевантность во втором случае превращается в составляющую, характеризующую "процесс". Действительно, когда в основе переписи лежат данные регистров, эта составляющая превращается в базовый фактор, способный влиять на большинство других составляющих качества.

13. Какие же важнейшие моменты следует учитывать при использовании классической модели оценки качества обследования применительно к переписи, основанной на данных регистров?

- а) Классическая процедура сбора данных в ходе обследований разработана специально в статистических целях, в то время как основное назначение данных регистров в большинстве случаев является совсем иным. В этой связи огромное значение приобретает подробное изучение релевантности каждой переменной.

Особое внимание должно уделяться тем переменным, значения которых берутся из разных источников.

- b) Каждый из регистров, используемых для целей переписи, имеет свою историю, свою административную основу, свои механизмы сбора обработки данных и т.д. Поэтому на этапе подготовки важно внимательно изучать и документировать характеристики каждого такого регистра. В отчеты, составляемые по итогам оценки качества, должно включаться, по крайней мере, резюме этих выводов.
- c) Нормативно-правовое и административное регулирование может оказывать серьезное влияние на качество входящих данных. Поэтому отчет о качестве должен, по крайней мере, частично быть посвящен этим вопросам.
- d) При проведении обычных переписей собранные данные обычно относятся к одному и тому же моменту времени (контрольная дата или час). В случае использования в переписи данных регистров такой временной согласованности данных добиться довольно трудно, поскольку данные берутся из разных источников. Поэтому все отклонения от единого контрольного момента времени должны документироваться и указываться в отчете о качестве.

14. Пока несколько в тени оставались такие дополнительные влияющие на качество факторы, как *издержки* и *нагрузка*. Хотя при переходе от обычной переписи к регистровой сами эти составляющие не должны сильно меняться, реальные различия могут быть связаны с их количественной оценкой. Издержки и особенно нагрузка, уже по определению, должны быть значительно меньше в случае проведения переписи на основе данных регистров. Таким образом, оценка издержек и нагрузки в сравнении с издержками и нагрузкой, связанными с проведением классических обследований, также должна занимать свое (важное) место в отчете об итогах оценки качества.

III. ПЛАНЫ ОТНОСИТЕЛЬНО КАЧЕСТВА СЛОВЕНСКОЙ ПЕРЕПИСИ НА ОСНОВЕ РЕГИСТРОВ

A. Краткая история вопроса и имеющиеся административные и статистические источники данных для словенской переписи 2011 года, которая впервые будет опираться на регистры

15. При проведении переписи на основе регистров "данные для переписи населения получают методом регистровой оценки, при котором для определения значений переменных по каждой статистической единице используются одновременно несколько

регистров"¹. В ходе словенской переписи 2011 года, которая будет проводиться на основе регистров, будут использоваться данные из трех основных административных источников: Центрального регистра населения (далее ЦРН), электронной базы данных о домохозяйствах и регистра жилищ. Эти источники данных позволят определить две целевые совокупности: совокупность населения и совокупность жилищ. Помимо этого, будут задействованы и другие административные и статистические источники, которые помогут охватить обязательные и представляющие интерес для страны вопросы. К числу других имеющихся источников относятся: регистр занятости, регистр безработицы, регистр иностранцев, регистр территориальных единиц, налоговая база данных, данные социальной статистики и т.д. В некоторых случаях (главным образом для определения структуры домохозяйств и семей) будут использоваться и данные переписи 2002 года. Данные из других источников будут увязываться с обследуемыми совокупностями на этапе интеграции данных.

16. В настоящий момент словенский НСИ сталкивается со следующими основными проблемами, касающимися источников данных для проведения основанной на регистрах переписи:

- a) создание регистра жилищ еще не завершено (под руководством Управления картографии Республики Словении);
- b) качество данных в электронной базе данных о домохозяйствах (ведется министерством внутренних дел) нельзя назвать удовлетворительным, поскольку создана она была (компьютеризованна) совсем недавно;
- c) данные об отношениях между отдельными лицами в ЦРН являются недостаточно полными, что не позволяет СУРС легко устанавливать все родственные связи в семьях.

17. Если сравнить положение в Словении с положением в странах Северной Европы, то главное различие между ними заключается в том, что, хотя в ЦРН и регистрируются некоторые родственные связи между членами семьи, это делается лишь применительно к таким важнейшим показателям движения населения, как рождения, браки и т.д., и то лишь с апреля 1988 года. Информация о родственных связях является недостаточно полной для того, чтобы легко определять состав семей. Это может делаться на основе данных о

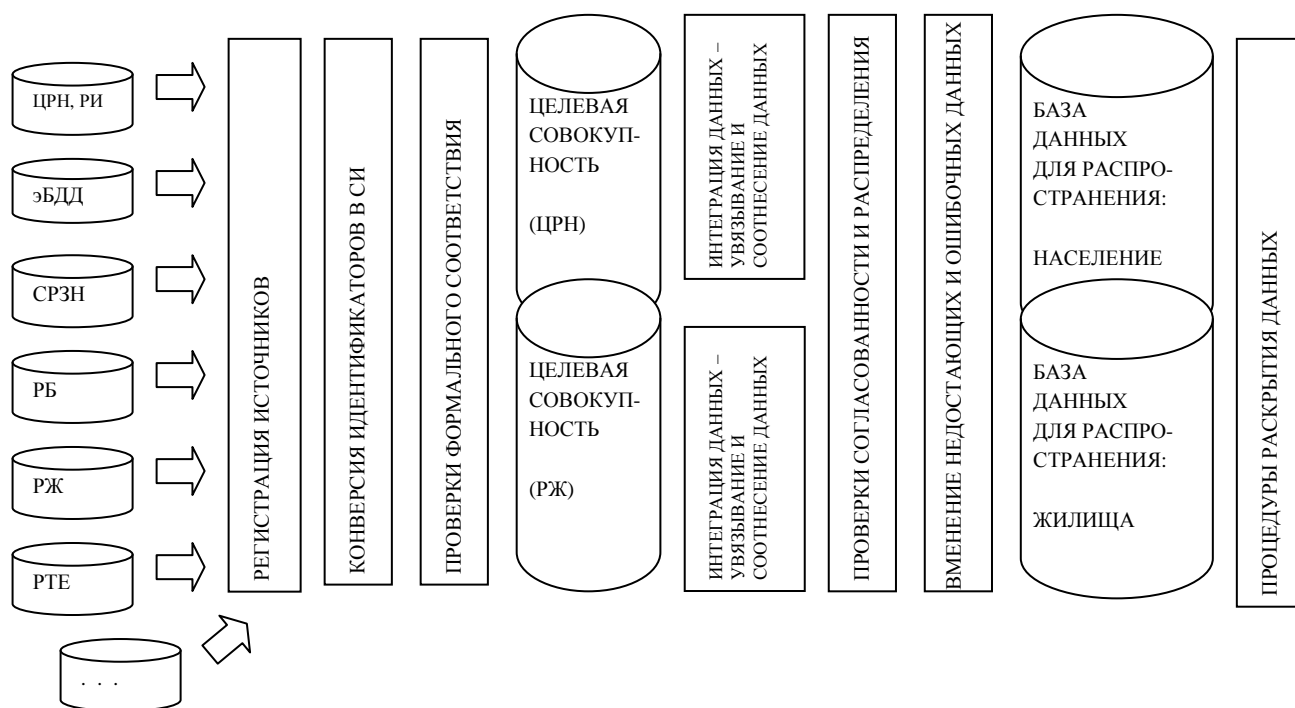
¹ Рекомендации КЕС по проведению переписей населения и жилищного фонда 2010 года, Организация Объединенных Наций, Нью-Йорк и Женева, 2006 год, стр. 271.

возрасте, семейном положении и поле лиц, проживающих вместе в одной жилищной единице, при условии, что в ЦРН указан номер соответствующей жилищной единицы. Кроме того, для некоторых групп населения могут использоваться и данные о родственных связях из переписи 2002 года (например, для лиц, которые не переезжали со времени предыдущей переписи).

18. Конечно же, существует возможность включить в одно из основных обследований домохозяйств специальный модуль, который позволит собирать информацию о семейных связях. Однако эти данные можно будет использовать лишь для оценки полноты охвата и качества данных, полученных из административных и статистических источников.

В. Процедура проведения и оценка качества переписи на основе регистров, запланированной на 2011 год

19. СУРС планирует использовать следующую процедуру переписи, которая будет проводиться с использованием регистров:



ЦРН – Центральный регистр населения

РИ – регистр иностранцев

ЭБДД – электронная база данных домохозяйств

СРЗН – статистический регистр занятого населения

РБ – регистр безработных

РЖ – регистр жилищ

РТЕ – регистр территориальных единиц

... – прочие источники

20. Первым шагом по созданию основанной на регистрах системы должно стать определение подходящих источников административных и статистических данных. Эта задача отчасти была решена при проведении переписи населения и жилищ в 2002 году, однако с тех пор многие данные были компьютеризированы, и появились новые административные источники данных – главным образом в министерстве внутренних дел и министерстве государственной администрации. Что касается оценки качества, то фактор релевантности учитывается уже на этом этапе.

21. Должны документироваться следующие данные (на базе шаблона), полученные из каждого из используемых в проводимой на основе регистра переписи источника:

- a) название источника;
- b) описание источника;
 - i) дата создания,
 - ii) правовой статус,
 - iii) используемые идентификаторы,
 - iv) система обновления,
 - v) структура;
- c) метод и частота использования источника в СУРС;
- d) редактирование данных в СУРС.

22. В каждом источнике будут определены релевантные для переписи 2011 года переменные. По отобранным релевантным переменным будут изучены и задокументированы временные рамки и возможные отклонения от контрольной даты (и часа). В случае существования нескольких источников одной и той же релевантной переменной должна быть определена их приоритетность.

23. По итогам такого документирования каждый источник пройдет процедуру регистрации, конверсии идентификаторов в статистические идентификаторы (так называемые СИ) и апробирования.

24. Регистрацию должен пройти каждый источник, однако на данном этапе единственным качественным критерием может служить ответ на вопрос, удалось ли зарегистрировать источник или нет. На этапе конверсии идентификаторов в статистические идентификаторы следует оценивать число непригодных для конверсии регистровых данных. В результате можно будет оценить, какой объем регистровых данных (в рамках одного источника и всех источников, используемых в проводимой на основе регистров переписи) можно будет использовать на следующих этапах. Апробирование позволит проверить, насколько данные согласуются с заранее отобранным для того или иного источника диапазоном переменных. Данный этап позволит также определить некоторые очевидно выпадающие данные.

25. После апробирования и с учетом установленных правил (в соответствии с согласованными на международном уровне определениями) будут сформированы две основные обследуемые целевые совокупности – населения и жилья. Поскольку считается, что данные ЦРН, по крайней мере, с точки зрения охвата населения, проживающего на территории Словении, имеют высокое качество, определение лиц, постоянно проживающих на территории страны, не должно создать особых проблем. Несоответствия могут возникнуть лишь на этапе увязки данных ЦРН с данными регистра иностранцев. Определенное дублирование возможно из-за различных целей двух регистров. На сегодняшний день немного известно о том, насколько полно целевая совокупность отражена в регистре жилья, который еще только создается. В то же время, поскольку в его основу лягут данные переписи недвижимости, проведенной в 2006 году, недостаточно полно могут быть учтены лишь новые здания и жилища.

26. Этап интеграции данных можно, безусловно, считать важнейшим этапом статистического процесса, когда речь идет о качестве. В отличие от первой части отчета о качестве, посвященной релевантности регистров и переменных, которая носит более или менее описательный и текстовый характер, та часть отчета, в которой речь идет об интеграции данных, должна в гораздо большей степени опираться на количественную информацию. Поскольку этот этап процедуры не описывается "классическими" составляющими качества, ему, безусловно, не хватает стандартных показателей качества. Ожидается, что эти проблемы будут широко обсуждаться в ближайшем будущем; в настоящем документе высказываются два предложения, касающихся возможных показателей.

27. В ходе процесса интеграции для получения информации о целевой совокупности будут объединяться данные из нескольких источников. Такое объединение планируется провести в два последовательных этапа. На первом этапе источники данных будут частично объединены путем прямого сопоставления с единым идентификатором.

К остальной части данных, которые не удастся объединить посредством прямого сопоставления, будет применен статистический метод увязки. Для мониторинга этого процесса могут использоваться три степени соответствия. Общий показатель соответствия указывает на то, какую часть наблюдаемых единиц, характеризуемых данными из административных источников (независимо от целевой совокупности), удалось объединить (напрямую или статистически). Прямой показатель соответствия характеризует единицы наблюдения, объединенные прямым способом, а статистический показатель соответствия – единицы, объединенные статистическими методами.

28. Что касается переменных, значения которых будут браться из разных источников, можно измерить степень их согласованности. Если первый, основной источник значений переменной обозначить Y_R , а альтернативный источник - Y_A , то показатель согласованности можно определить как отношение числа обследуемых единиц, для которых верно условие $|Y_R - Y_A|/Y_R < p$ (например, $p = 0,01$), к числу всех обследуемых единиц, по которым имеются данные из обоих источников.

29. Помимо той части отчета о качестве, который будет посвящен непосредственно мониторингу статистического процесса, внимание, безусловно, будет уделяться и остальным составляющим качества: сопоставимости и последовательности. Согласно нынешним планам, задача обеспечения сопоставимости будет главным образом решаться путем сравнения результатов основанной на данных регистра переписи 2011 года с результатами классической переписи 2002 года. Особое внимание будет уделяться изучению структуры населения по районам регистрации. Поскольку известно, что административный адрес проживания лиц может отличаться от фактического (обычного) места жительства, в этих данных могут быть обнаружены существенные расхождения.

30. В связи с требованием обеспечения последовательности можно задать следующий вопрос: "Согласуются ли итоги основанной на данных регистра переписи с результатами других, "классических" статистических обследований, таких, как ОДУЖ, ОРС и ОБДХ?" Поэтому та часть отчета о качестве, в которой речь пойдет о последовательности, должна быть в первую очередь посвящена представлению и описанию основных расхождений между этими результатами. Цель заключается в том, чтобы дать пользователям четкое представление о тех последствиях, которыми чреваты различия между двумя подходами.

IV. ВЫВОДЫ

31. В рамках европейской статистической системы большинство национальных статистических институтов уже приступили к активной подготовке следующего раунда переписей населения и жилищ 2010 года. По примеру стран Северной Европы большинство стран на этапе сбора данных для переписи планируют как можно шире пользоваться уже имеющейся административной информацией. Словения принадлежит к числу тех стран, которые планируют провести перепись исключительно на основе данных регистров.

32. К числу важных вопросов, которые, безусловно, должны быть учтены на этапе подготовки переписи, относятся планы в отношении оценки качества процедуры и результатов переписи. Поскольку нынешние концепции качества, появившиеся в европейской статистической системе в последнее десятилетие, были в основном адаптированы к потребностям классических статистических обследований, требуется изучить возможность использования новых "подходов к качеству". В настоящем документе высказывались некоторые соображения по данному вопросу. В первой части речь в основном шла о теоретических постулатах, а во второй - о конкретном положении дел в Словении. Некоторые из основных соображений перечислены ниже:

- a) Переход от обычной процедуры к статистическим обследованиям, основанным на регистрах, требует несколько иного подхода к оценке качества данных. Главные отличия связаны с оценкой процедур подготовки обследований и сбора данных.
- b) Релевантность административных и статистических источников, равно как и полученных на их базе статистических переменных, является важнейшей составляющей нового подхода.
- c) Важной составной частью отчета о качестве должна стать исчерпывающая документация, описывающая различные характеристики источников используемых данных и их административный статус.
- d) Кроме того, для основанных на регистрах обследований должен быть разработан набор стандартных показателей качества. Большинство новых показателей должны описывать качество этапа интеграции данных.

- е) Важнейшая цель обеспечения сопоставимости и последовательности должна заключаться в том, чтобы дать пользователям полное представление о тех отличиях, которые связаны с использованием данных регистров.

Справочная литература

Рекомендации КЕС по проведению переписей населения и жилищного фонда 2010 года, Организация Объединенных Наций, Нью-Йорк и Женева, 2006 год.

CES Recommendations for the 2010 Censuses of Population and Housing. United Nations, New York and Geneva, 2006.

Biemer Paul B., Lyberg Lars E.: Introduction to Survey Quality: John Wiley & sons, 2003.

Methodological documents, Handbook “How to make quality report: Working Group “Assessment of quality in statistics”, Sixth meeting, Luxembourg 2-3 October 2003.

Methodological documents, Standard Report: Working Group “Assessment of quality in statistics”, Sixth meeting, Luxembourg 2-3 October 2003.

Oblak Flander, A., 2007. Opportunities and Challenges of a Register-Based Census of Population and Housing – the Case of Slovenia. (Unpublished. Document prepared for the Seminar on Registers in Statistics - methodology and quality, 21–23 May, 2007, Helsinki).

Standard Quality Indicators - Producer oriented: Working Group “Assessment of quality in statistics”, Sixth meeting, Luxembourg 2-3 October 2003.

Wallgren A., Wallgren B.: “Register-based Statistics; Administrative Data for Statistical Purposes”: John Wiley & sons, 2007.
