

Distr.
GENERAL

CES/SEM.40/10
4 September 1998

RUSSIAN
Original: ENGLISH

СТАТИСТИЧЕСКАЯ КОМИССИЯ и
ЕВРОПЕЙСКАЯ ЭКОНОМИЧЕСКАЯ КОМИССИЯ

СТАТИСТИЧЕСКОЕ УПРАВЛЕНИЕ
ЕВРОПЕЙСКИХ СООБЩЕСТВ (ЕВРОСТАТ)

КОНФЕРЕНЦИЯ ЕВРОПЕЙСКИХ СТАТИСТИКОВ

Совместная рабочая сессия ЕЭК-ЕВРОСТАТА
по переписям населения и жилищ 1/
(Дублин, Ирландия, 9-11 ноября 1998 года)

Тема 3

ВОПРОСЫ РАСПРОСТРАНЕНИЯ ДАННЫХ В ПРОГРАММЕ ПЕРЕПИСИ НАСЕЛЕНИЯ И ЖИЛИЩ 2001 ГОДА В ВЕНГРИИ

Документ, представленный Центральным статистическим управлением Венгрии 2/

I. Обзорная информация

1. Результаты переписей, как и другая статистическая информация, традиционно публиковались в печатной форме. Это, безусловно, ограничивало возможности пользователей в области получения и использования такой информации в своих конкретных целях. Внедрение методов электронной обработки данных позволило статистическим управленим более гибко реагировать на потребности пользователей. При этом степень такой гибкости еще более возросла в последнее время в связи с новыми разработками в области методов ЭОИ, технических и программных средств и связи.

2. Все более широкое использование статистической информации влечет за собой также диверсификацию пользователей, их потребностей и запросов. В связи с этим Статистическому управлению становится все труднее удовлетворять широкий круг конкретных требований. Поэтому, если мы не хотим постоянно подвергаться критике и намерены в полной мере выполнять свои обязанности, нам необходимо найти гибкие и отвечающие потребностям пользователей средства распространения данных.

1/ Документы, подготовленные для этой рабочей сессии, будут иметь тот же статус, что и документы, подготовленные для семинаров.

2/ Автор: Габриэлла Вукович.

3. В прошлом большинство пользователей вряд ли могли считаться специалистами в области использования статистических данных, и, по общему мнению, статистики более или менее точно определяли потребности пользователей. В последние 10-15 лет эта ситуация коренным образом изменилась. У всех групп пользователей возросла потребность в надежных данных, и в связи с этим роль пользователей в определении методов и рамок анализа также возросла. Среди пользователей много хороших специалистов, и они сами могут определять и определяют собственные конкретные информационные запросы, а также их желательные рамки и форму. Сейчас производство статистических данных гораздо сильнее зависит от спроса, чем в прошлом.

4. Печатные материалы по-прежнему имеют определенное значение, особенно при их использовании в целях информирования широкой общественности по социальным и экономическим вопросам, охватываемым в переписи, однако, помимо этого, пользователи нуждаются в представлении статистической информации в машинно считываемой электронной или оптической форме, частично в форме набора таблиц, но еще в более значительной степени в базах данных, позволяющих свободно отбирать и комбинировать переменные показатели.

5. При анализе проблем, связанных с распространением данных, возникает ряд вопросов.

- Как определить потребности пользователей? Как выбрать наилучшее время для распространения данных? Как обслужить максимальное число пользователей и удовлетворить максимальное число их потребностей?
- Каким образом Статистическое управление может соблюдать требования в области защиты данных, если пользователи имеют доступ к базам данных? Данные переписи имеют специфический характер, в частности с точки зрения географических и личностных данных. В связи с этим свободный доступ к первоначальной базе данных переписи позволит косвенным образом получить личностные данные даже при анонимном характере файлов.
- Какие меры должно принять Статистическое управление для покрытия издержек? Какую информацию следует предоставлять бесплатно, и за какие услуги следует взимать плату? Какие следует установить цены? Следует ли обращаться с пользователями на равной основе, или необходимо определить категории пользователей, которые будут иметь доступ к информации бесплатно или по низкой цене, в то время как в отношении других будут применяться более высокие тарифы?

II. Определение потребностей пользователей

6. Процесс определения потребностей пользователей в данных переписи в целом хорошо известен всем статистическим управлением. В этой связи существуют различные организационные возможности. Обычной практикой при планировании переписи являются консультации в постоянных и специальных комитетах в среде внутренних, внешних и международных пользователей данных переписи. Вместе с тем планирование переписи осуществляется задолго до периода распространения данных, и поэтому основное внимание при ее разработке уделяется определению вопросов, которые должны вноситься в вопросник, их формулировке, а также вопросам кодирования и обработки результатов.

7. На этом этапе фактической форме распространения данных практически не уделяется никакого внимания. В какой-то степени это оправдано, поскольку процесс планирования более или менее завершается за 12-18 месяцев до дня переписи, в то время как распространение ее результатов начинается лишь 9-12 месяцев спустя после проведения переписи. Для наших информационных сообществ 30 месяцев – большой срок, в течение которого последнее звено в коммуникационной цепи может полностью измениться. Мы уже столкнулись с определенными трудностями при попытке максимально позднего определения нашей ИТ без нанесения ущерба самой обработке данных и использования ИТ, которая будет не слишком устаревшей на момент начала обработки данных. Наши клиенты, однако, могут руководствоваться иными подходами, а используемая многими из них информационная технология может быть гораздо более передовой по сравнению с нашей.

8. Результаты переписи 1990 года были опубликованы в стандартной печатной форме, в одном томе объемом примерно 550 страниц с таблицами по каждому из 19 медье, а также в основном и тематических томах. В целом эта серия состоит из 88 томов общим объемом примерно в 20 000 печатных страниц. Серия завершается томом с описанием методологии и определений. Перепись проводилась 1 января 1990 года, тома по медье вышли в июне и июле 1992 года, после чего были опубликованы тематические тома, последний из которых вышел в феврале 1995 года. Кроме того, была подготовлена дискета с базой данных, содержащей основную информацию по всем населенным пунктам Венгрии, и картографическую программу. Был выпущен КД-ПЗУ с полным файлом данных и программой поиска в базе данных, а также картографическая программа. На момент подготовки КД-ПЗУ считался очень современным средством – настолько современным, что, несмотря на его популярность, многие заинтересованные пользователи были вынуждены отложить его покупку, не имея возможности купить соответствующее устройство для считывания оптических дисков. В настоящее время таким устройством оборудованы практически все персональные компьютеры, публикация на КД в системе DOS практически изжила себя, и мы вынуждены разрабатывать вариант в системе Windows. (Последний, однако, по-прежнему связан с некоторыми нерешенными вопросами. Он удобен для внутреннего пользования, однако мы не призываем клиентов использовать его.)

9. Многие пользователи не могут найти необходимых данных в публикации 1990 года, и, поскольку им необходимы лишь некоторые таблицы, для них удобнее обращаться в Статистическое управление со специальными запросами о представлении таблиц. К концу 90-х годов мы по-прежнему получаем в среднем 20 запросов в год с просьбой о представлении данных переписи 1990 года, особенно по небольшим районам. Эти запросы тщательно регистрируются и позволяют нам учитывать текущие потребности пользователей в таком виде данных, а также выяснить, в какой форме им необходимы эти данные. Другие индивидуальные запросы в статистических данных, помимо данных переписей, также регистрируются и используются в качестве справочной информации при планировании стратегии распространения данных переписи 2001 года.

10. Кроме того, на этапе планирования переписи 2001 года пользователям предлагается сообщить о своих предпочтениях в области распространения её результатов. Как и предполагалось, их ответы говорят о том, что стандартные таблицы теряют свое значение и для профессиональных пользователей гораздо более важен прямой доступ к файлам данных. Пользователи также хотят получить данные вскоре после проведения переписи (гораздо скорее, чем мы сможем их подготовить), из чего следует, что график подготовки публикаций, использовавшийся в 1990 году, уже не актуален, и в связи с этим следует предусмотреть гораздо более оперативное распространение данных. Такая потребность в своевременном распространении данных также подчеркивает необходимость скорейшего обеспечения доступа к базе данных без задержек, которые неизбежно возникают при использовании традиционных форм публикации.

11. Для тех пользователей, которые заинтересованы в более общих формах представления результатов переписи, наилучшим выходом, возможно, являются стандартные таблицы. Обычная книга на полке по-прежнему удовлетворяет многих пользователей, которые хотят ознакомиться или использовать лишь некоторые данные переписи. Для этих пользователей мы намерены опубликовать региональные (по медье) и тематические выпуски, содержащие стандартные таблицы с данными переписи 2001 года, однако эти таблицы будут гораздо менее подробными, чем во время переписи 1990 года. Эти выпуски будут предназначаться скорее для получения общего представления, а не для сложного анализа.

12. Ряд пользователей нуждается в файлах или базах данных в различных исследовательских или учебных целях. Лучшим выходом для таких пользователей представляется файл микроданных переписи. Вопрос о размерах выборки пока еще не решен. Эту задачу, вероятно, можно будет решить с помощью национальной выборки обезличенных записей в 2%, которую можно будет легко обработать на ПК.

13. Не решен также вопрос о региональных данных. Национальная выборка в 2% является недостаточно подробной для некоторых субнациональных пользователей, однако в связи с относительно недавним расширением полномочий местных властей и разработкой

политики регионального развития перед нами по-прежнему стоит задача обсудить с пользователями вопросы составления и наличия региональных выборок и выборок мелких районов.

III. Защита личностных данных

14. Одним из реальных вопросов, возникающих при публикации данных по мелким географическим единицам, является возможность косвенного выявления респондентов. Для предотвращения такого косвенного выявления статистические службы пользуются различными методами, например, "барнардизация" (произвольное добавление +1 или -1), ограничение переменных показателей, которые могут публиковаться по мелким районам, или установление порогового показателя – минимального числа лиц, домашних хозяйств, или жилищ, которые могут находиться в районе, по которому публикуются данные.

15. В качестве самых мелких географических единиц, по которым публикуются данные, в большинстве стран используются переписные районы. В Венгрии переписной район состоит из примерно 100 домашних хозяйств (или несколько менее 300 человек). В связи с очень строгими правилами защиты данных этот пороговой показатель обычно считается заниженным, и от нас требуют представлять данные по нескольким (обычно 5–10) переписным районам одновременно. В ходе переписи 2001 года нам было необходимо использовать переписные районы именно в тех целях, для чего они первоначально "создавались", т.е. в качестве определения рабочей нагрузки на счетчиков. При распространении данных переписи 2001 года мы намерены отказаться от переписных районов как географических единиц. Вместо этого мы планируем определить специальные единицы на основе регистра адресов (представляющего собой файл адресов всех жилищ, созданный в целях переписи и дополненный важной информацией по каждому жилищу и прилегающей к нему территории в связи с переписью). Такой подход позволяет добиться высокой степени гибкости при удовлетворении потребностей пользователей. Все эти территориальные единицы, безусловно, должны будут отвечать пороговым критериям.

IV. Посредники в распространении данных

16. Еще одним нерешенным вопросом является проблема передачи базы данных переписей или результатов переписи распространителям-посредникам. До настоящего времени за решение этой задачи отвечало Центральное статистическое управление Венгрии. По-прежнему не решены вопросы, связанные с правовыми (защита данных, авторские права), практическими и финансовыми аспектами отношений с распространителями-посредниками.

17. Нам хотелось бы поощрить тех посредников, которые повышают качество наших данных, и тем самым расширить рамки услуг, предоставляемых пользователям. К числу таких продуктов с большой добавленной стоимостью, которые уже пользуются спросом (однако получить которые пока не просто), относятся прикладные программы ГИС, разрабатываемые на основе данных переписи и другой статистической информации. Это

направление деятельности будет представлять собой особый интерес для пользователей и посредников-распространителей, особенно в связи с тем, что Статистическое управление на настоящий момент располагает лишь ограниченными возможностями в области ГИС. Следует признать, что эта работа будет полезна и для самого Статистического управления как при планировании и переведении переписей и обследований, так и при анализе данных.

v. Покрытие издержек

18. Вероятно всем статистическим управлениям приходится решать проблемы нехватки средств, и при этом некоторые из них в соответствии с законодательством должны быть рентабельными. Это относится также и к Центральному статистическому управлению Венгрии. В соответствии с законом о годовом бюджете управление должно обеспечивать определенную сумму поступлений. Этот закон не поощряет государственный сектор в получении прибыли и предусматривает только покрытие издержек. В соответствии с Законом о статистике и внутренними правилами пользователи обязаны покрывать дополнительные расходы, связанные с обработкой данных, составлением таблиц и другими услугами, выходящими за рамки обычной работы Статистического управления. Вместе с тем четкого определения этих рамок работы и услуг не существует. За планирование результатов отвечает само Статистическое управление, и статистики в определенной степени склонны придерживаться традиционных решений. В связи с этим многие пользователи вынуждены оплачивать услуги, которые могли бы быть включены в стандартные рамки официальной информации, вместо тех из них, которые во многом уже утратили свою актуальность.

19. Статистическое управление не планирует покрывать расходы, связанные со сбором или обработкой данных, до этапа создания базы данных по результатам переписи. Начиная с этого этапа, некоторые данные приобретают открытый характер, а другие требуют дополнительных затрат со стороны Статистического управления. В качестве простейшего примера можно привести выпуски печатных таблиц, которые предназначены для продажи, однако цена на них устанавливается на уровне затрат на публикацию, и не предусматривает прибыли для Статистического управления.

20. Возможная схема покрытия расходов на распространение данных будет включать в себя следующие элементы:

- бесплатное предоставление некоторой базовой информации через Интернет и в виде брошюр,
- ряд стандартных таблиц, предоставляемых по низкой цене,
- файл микроданных, предоставляемый по низкой цене,

- региональные выборки и выборки мелких районов, предоставляемые по несколько более высокой цене по сравнению с национальной выборкой в связи с их более важным значением с аналитической и практической точек зрения,
- файлы, предоставляемые распространителям-посредникам на договорных условиях,
- наборы данных, отвечающих конкретным индивидуальным потребностям и предоставляемых за дополнительную плату, связанную с составлением специальных таблиц или разработкой конкретной базы данных,
- в соответствии с практикой Центрального статистического управления Венгрии учебным заведениям и исследовательским институтам данные или файлы данных предоставляются по льготным ставкам или бесплатно, а ограничения в отношении использования таких данных оговариваются в конкретном соглашении между клиентом и ЦСУВ. Этот подход, наиболее вероятно, будет использоваться при распространении данных переписи 2001 года.

21. В отношении доступа пользователей к данным и файлам данных или базам данных сохраняется некоторая неопределенность в связи с быстрым развитием электронных средств. КД-ПЗУ по-прежнему может пользоваться спросом, в то время как все более широкую популярность приобретает Интернет, а на момент распространения данных переписи могут появиться и другие средства. Важно проанализировать предпочитаемую пользователями форму представления данных и попытаться удовлетворить их потребности.
