



ЭКОНОМИЧЕСКИЙ  
И СОЦИАЛЬНЫЙ СОВЕТ

Distr.  
GENERAL

CES/2003/7  
7 April 2003

RUSSIAN  
Original: ENGLISH

СТАТИСТИЧЕСКАЯ КОМИССИЯ и  
ЕВРОПЕЙСКАЯ ЭКОНОМИЧЕСКАЯ КОМИССИЯ

КОНФЕРЕНЦИЯ ЕВРОПЕЙСКИХ СТАТИСТИКОВ

Пятьдесят первая пленарная сессия  
(Женева, 10-12 июня 2003 года)

**ПРЕДМЕТНЫЙ ОБЗОР ВОПРОСОВ КОНФИДЕНЦИАЛЬНОСТИ  
СТАТИСТИЧЕСКИХ ДАННЫХ И ДОСТУПА К МИКРОДАНЫМ**

Документ, представленный Статистическим управлением Швеции<sup>1</sup>

**I. ВВЕДЕНИЕ**

1. Главной задачей национального статистического института (НСИ) в области конфиденциальности статистических данных и микроданных является отыскание и установление определенного баланса между требованиями защиты конфиденциальности и широкого использования микроданных. Поскольку широкое использование микроданных предполагает улучшенные возможности предоставления более качественных данных для удовлетворения потребностей пользователей, данный баланс приобретает центральное значение в системе официальной статистики, которая должна "*...являться одним из необходимых элементов информационной системы демократического общества, обеспечивающей правительство, экономические круги и общественность данными...*"<sup>2</sup>. Одновременно "*личностные данные, собираемые статистическими ведомствами для разработки статистической информации, независимо от того, относятся ли они к*

<sup>1</sup> Авторы: Матти Нива при содействии Бо Сундгрена и Ингрид Люберг.

<sup>2</sup> Основные принципы официальной статистики ООН, статья 1, Статистическая комиссия ООН, 1994 год.

*физическим или юридическим лицам, должны носить строго конфиденциальный характер и использоваться исключительно для статистических целей*<sup>3</sup>. Поиск данного баланса неизбежно предусматривает комбинирование различных мер и процедур и изучение различных правовых, технических, административных и методологических вопросов.

2. Целью настоящего документа является изложение некоторых из основных вопросов, возникающих в ходе решения проблем, связанных с этими различными аспектами. Сначала в нем излагаются некоторые идеи, касающиеся перспектив использования микроданных. Следующий раздел посвящен вопросам конфиденциальности и состава микроданных. Затем в нем рассматриваются правовые вопросы. В заключение в документе кратко обсуждаются различные организационные подходы, касающиеся доступа к микроданным. Это также примерно соответствует программе работы семинара.

## **II. ИСПОЛЬЗОВАНИЕ МИКРОДАНЫХ**

3. Многие совпадающие во времени события содействовали расширению возможностей использования микроданных в исследовательских целях. К ним относятся технологические достижения в области аппаратного и программного обеспечения, документирования данных и Интернета. Современные ПК способны в настоящее время обрабатывать сложные и большие по объему наборы микроданных. Это предполагает, что НСИ могут относительно легко предоставлять свои крупные наборы данных исследователям. В качестве одной из важных задач НСИ должно рассматриваться обеспечение того, чтобы вся совокупность хранящихся у них микроданных могла в полной степени использоваться исследователями и другими санкционированными пользователями.

4. Традиционно НСИ публиковали те агрегированные статистические данные, которые они считали важными, хотя пользователи также, естественно, оказывали влияние на такие решения. Предоставление агрегированных статистических данных в табличной форме также подразумевало установление четких ограничений на то, каким образом официальные статистические данные могут использоваться в целях социально-экономических исследований.

5. Следующим шагом в направлении повышения полезности хранимых в НСИ данных для пользователей явилось создание статистических баз данных, содержащих матрицы

---

<sup>3</sup> Там же, статья 6.

агрегированных данных, которые открывали пользователю широкие возможности с точки зрения разработки своих собственных статистических данных.

6. Предоставление доступа к микроданным является новым важным шагом вперед, поскольку исследователи и другие пользователи могут самостоятельно отбирать данные, соответствующие потребностям их исследований. Это также имело последствия для теоретических разработок в области социально-экономических исследований. Многие исследователи могут подтвердить важность использования микроданных для анализа возможных последствий мер политики (см., например, Erikson, p. 2). Теоретические пояснения общих условий могут, таким образом, дополняться анализом с использованием статистических данных НСИ механизмов, проводимым на индивидуальном уровне.

7. Наличие больших массивов продольных микроданных обуславливает появление новых аналитических возможностей. Так, например, увязка различных файлов микроданных за ряд лет открывает новые возможности в области динамического анализа. Данный тип исследований на основе микроданных стал в последние десятилетия широко распространенной практикой. Возможности использования данного подхода также являются очевидными и в области экономических исследований. Одним из типичных примеров экономического анализа рынка труда является увязка данных о работодателях и наемных работниках для анализа одновременно спроса и предложения на рынке труда (см. Westergaard-Nielsen, p. 2).

8. НСИ могут также осуществлять интеграцию различных регистров микроданных и создавать новые базы данных. Однако, как правило, обеспечение приемлемого качества данных требует большого объема статистической работы. Статистическое управление Швеции разработало несколько баз данных такого рода. Примером может служить продольная база данных "Луиза", содержащая обезличенные микроданные об отдельных лицах и семьях, касающиеся их образовательного уровня, доходов и занятости. Следует добавить, что эта база данных содержит годовые данные обо всем взрослом населении Швеции начиная с 1990 года и обновляется каждый год. Такая интегрированная база данных открывает широкие возможности для проведения различных видов анализа. Альтернативой созданию интегрированной базы данных может служить увязка ряда регистров микроданных друг с другом на специальной основе в конкретных целях.

9. Более широкая доступность микроданных в совокупности с разработками в области ИТ также привели к появлению нового подхода, получившего название "добыча данных" или в более широком смысле "извлечение знаний из баз данных". Данный подход представляет особый интерес с учетом возможности "добычи данных" из множественных баз (см. Torra et al).

10. Для НСИ расширенное использование микроданных предполагает прирост полезности за счет более эффективного использования данных, хранящихся в НСИ, что должно также содействовать укреплению их авторитета в глазах респондентов и общественности в целом. Это также означает, что инвестиции в официальную статистику могут обеспечивать высокую отдачу.

### III. КОНФИДЕНЦИАЛЬНОСТЬ

11. Одним из основных вопросов, который неразрывно связан с использованием любых микроданных, является их конфиденциальность. Упрощенно говоря, любое использование микроданных, даже в обезличенном виде, может создавать угрозу конфиденциальности. Хотя нарушение конфиденциальности в связи с использованием микроданных вряд ли может произойти когда-либо в рамках использования в исследовательских проектах данных НСИ, защита конфиденциальности по-прежнему является и должна оставаться одной из главных задач и сохранять свою актуальность. Весьма положительные результаты, достигнутые в этой области в настоящее время, частично объясняются усилиями НСИ. Еще одной возможной причиной может являться тот факт, что исследователи, работающие с микроданными, рискуют в данном случае своей репутацией. Кроме того, предоставление доступа к микроданным традиционно содействует налаживанию контактов между НСИ и исследовательскими кругами.

12. Необходимость защиты и обеспечения целостности статистических данных является одной из давних тем. Одним из первых о необходимости установления баланса между правом на неприкосновенность личной жизни и растущей потребностью в информации заговорил Винсент П. Барабба в 1974 году, когда он, выступая от имени Бюро переписей США, заявил, что "*...существует внутреннее противоречие при сборе данных от граждан. Речь идет о противоречии между правом этих граждан на неприкосновенность личной жизни, с одной стороны, и использованием правительством обязательных процессов получения информации для удовлетворения законных интересов, с другой стороны*" (см. Varabba, p. 34).

13. Вопрос конфиденциальности разрабатывался и нашел свое отражение в документах международного статистического сообщества. В декларации МСИ о профессиональной этике 1985 года подчеркивалось, что "*Статистические данные не имеют никакого отношения к личности граждан*", что предполагает, что *... "личностные характеристики и сведения сотрудничающих (или не сотрудничающих) респондентов должны ... оставаться конфиденциальными, независимо от того, гарантируется ли конфиденциальность явным образом или нет"*. (ISI, 4.5 Maintaining confidentiality of

records). Кроме того, статистики не должны допускать публикации своих данных "*...в форме, позволяющей разглашение или установление личности любого субъекта*" (ISI, 4.4 Inhabiting disclosure of identities).

14. Кроме того, Основные принципы официальной статистики являются весьма ясными по этому вопросу: "*Личностные данные, собираемые статистическими ведомствами для подготовки статистической информации, независимо от того, относятся ли они к физическим или юридическим лицам, должны носить строго конфиденциальный характер и использоваться исключительно для статистических целей*" (статья 6). Во многих странах для НСИ разработаны национальные кодексы поведения. Как правило, защита конфиденциальности является одним из краеугольных камней таких кодексов. Примером может также служить недавно принятый Кодекс поведения Управления национальной статистики Соединенного Королевства, в котором принцип защиты конфиденциальности является одним из главных обязательств Управления национальной статистики (Кодекс поведения Управления национальной статистики).

15. Вполне очевидно, что защита конфиденциальности является одной из главных обязанностей НСИ. По этическим причинам НСИ, естественно, должны заботиться о защите неприкосновенности личной жизни граждан, а также тайны информации коммерческих заведений, однако приверженность принципу защиты конфиденциальности имеет также и конкретное объяснение. НСИ должны пользоваться в этом отношении полным доверием для того, чтобы иметь возможность собирать данные от респондентов.

16. Однако учитывая, что использование микроданных всегда может быть сопряжено с риском нарушения конфиденциальности (поскольку полностью безопасных микроданных не существует), реально осуществимой задачей является отыскание баланса между требованиями расширения доступа к информации и защита конфиденциальности. НСИ обычно оценивают риск идентификации в различных областях и на различных этапах статистического процесса применительно к различным видам использования и пытаются свести такие риски к минимуму. Для этого НСИ используют целую совокупность различных методологических решений.

17. Некоторые из других вопросов, связанных с защитой конфиденциальности, касаются особых процедур предоставления доступа к микроданным о предприятиях по сравнению с микроданными о частных лицах. Еще один вопрос касается временных рамок, т.е. должна ли защита конфиденциальности обеспечиваться без ограничений по времени (т.е. вечно) или же ограничиваться сроком существования субъекта информации или даже меньшим периодом.

18. В странах центральной и восточной Европы институциональные и правовые рамки статистики претерпели значительные изменения в процессе перехода. Положение с соблюдением конфиденциальности данных в этих стран может иметь некоторые особенности (например, законодательное оформление, применение принципа одностороннего потока<sup>4</sup>, обеспечение защиты конфиденциальности в масштабах всей статистической системы). В исследовании, проведенном секретариатом ЕЭК, был сделан вывод о желательности разработки типовых руководящих принципов в области статистической конфиденциальности (семинар КЕС...).

19. Последние события, подчеркивающие необходимость укрепления национальной безопасности, способны поставить под угрозу принцип статистической конфиденциальности исходя из соображений безопасности. Если это будет предполагать снятие ограничений с доступа к микроданным, которые ранее могли использоваться только в статистических целях, то это может спровоцировать кризис доверия ко всему статистическому сообществу.

20. До настоящего времени защита конфиденциальности являлась в значительной степени проблемой национального масштаба, однако в контексте ЕС она стала проблемой для учреждений ЕС, таких, как, в частности, Евростат. В этой области открываются новые возможности наряду с возникновением новых озабоченностей.

#### **IV. ПРАВОВЫЕ ВОПРОСЫ**

21. Механизмы защиты конфиденциальности могут опираться на правовые акты и/или правила и регламентирующие нормы, применяемые НСИ. Правовая ситуация является различной в зависимости от страны и региона. Если закон о статистике был относительно недавно принят в стране, то он, как правило, содержит нормы, регулирующие доступ к статистической информации. Как правило, одним из центральных принципов является то, что сведения, собираемые в статистических целях, независимо от того, собираются ли они в рамках обязательной отчетности или же предоставляются на добровольной основе, могут в принципе использоваться только для разработки статистических данных. Однако эти данные могут также использоваться в исследовательских целях при условии соблюдения установленных требований.

---

<sup>4</sup> Принцип одностороннего потока предполагает, что НСИ имеют доступ к административным данным, ведущими министерствами или другими правительственными ведомствами, на уровне микроданных, в то время как возможность обратного потока микроданных, на которые распространяются требования статистической конфиденциальности, независимо от того, являются они результатами статистических обследований или взяты из других источников, строго исключается.

22. Как правило вопросы разработки статистических данных и публикации микроданных регулируются также и другими законодательными актами, такими, как Закон о личностных данных. В контексте ЕС важное значение имеет так называемая Директива Совета и Европейского парламента о защите данных (№ 95/46 от 24 октября 1995 года), поскольку она усиливает правовую защиту граждан от автоматической передачи касающейся их личностной информации (речь идет о компьютеризованных личностных данных и данных, хранящихся в структурированных ручных картотеках). Все государства-члены должны принять соответствующий национальный закон о личностных данных, опирающийся на данную директиву ЕС. Это может также предполагать, что другие ведомства, такие, как органы надзора по вопросам информации, могут обладать правом окончательного решения относительно использования микроданных о частных лицах.

23. Что касается законодательства ЕС, которое непосредственно регулирует вопросы статистической конфиденциальности, то существует Постановление Совета ЕС о статистике Сообщества (№ 322/97), согласно которому как национальные органы, так и Евростат должны обеспечивать защиту конфиденциальных данных. Основной принцип заключается в том, что конфиденциальные данные, полученные исключительно в целях разработки статистики Сообщества, должны использоваться только в целях статистического производства, если только респонденты не дали четкого своего согласия на использование данных в иных целях. Недавно принятое Постановление о порядке применения Постановления о статистике Сообщества касается предоставления доступа к конфиденциальным данным в научных целях (Постановление Комиссии 83/2002). Данное Постановление содержит правила, регулирующие условия предоставления доступа к конфиденциальным данным, получаемым в рамках четырех различных обследований ЕС. Согласно этому Постановлению, исследователям может предоставляться такой доступ на основе подписания с Евростатом договора, регулирующего условия доступа. Однако предстоит еще выяснить, каким образом государства-члены будут применять этот правовой механизм на практике.

24. Что касается информации государственных органов, то здесь в качестве отправного пункта может использоваться принцип, полностью противоречащий защите конфиденциальности. Может утверждаться, что во имя демократии вся информация, издаваемая в государственном секторе, должна являться общедоступной. Согласно этому принципу, все решения, принимаемые государственными органами, а также справочные документы, лежащие в основе этих решений, в том числе переписка государственных органов, должны являться общедоступными. Этот принцип гласного государственного управления применяется в Швеции. Однако все возможные случаи отступления от этого

принципа гласности, обусловленные мотивированной необходимостью сохранения тайны и т.д., должны быть прописаны в конкретных законах, таких, как закон о статистике.

## **V. ДОСТУП К МИКРОДАНЫМ - РАЗЛИЧНЫЕ ПОДХОДЫ**

25. Одной из главных проблем, стоящих перед НСИ в области использования микроданных, является предоставление доступа пользователям различными способами. Такой доступ может в принципе быть организован по различным каналам и, как правило, сам НСИ должен определять подходящие и практически осуществимые решения с учетом существующих институциональных и организационных условий. В то же время в последние годы в этой области проводится большое число сравнительных исследований.

26. Одним из первых способов предоставления контролируемого доступа к микроданным являлось составление обезличенных общедоступных файлов микроданных. Данное решение было принято на вооружение Статистическим управлением Канады в начале 70-х годов прошлого века (см. *Boyko & Watkins*, p. 3). Перед публикацией эти файлы подвергались тщательной обработке с целью снижения вероятности идентификации. С момента начала этой программы было создано более 350 таких файлов, и целый ряд других стран приняли решение об использовании аналогичных решений. Общедоступные файлы микроданных являлись ценными материалами для исследователей из университетов и государственных учреждений. В числе проблем, связанных с их использованием, можно назвать относительно высокие издержки по их подготовке, особенно в период 80-х и начала 90-х годов прошлого века (использование больших ЭВМ, ценовая политика), и тот факт, что процесс обезличивания приводит к снижению полезности данных (там же, p. 5).

27. Во многих странах предоставление обезличенных микроданных исследователям и другим законным пользователям, внешним по отношению к НСИ, по-прежнему является основным способом публикации микроданных. В Статистическом управлении Швеции в 2002 году было зарегистрировано около 200 таких случаев, причем число публикаций быстро растет. Если публикуемые данные являются подробными данными регистров, то они обычно не являются на деле обезличенными. Вполне очевидно, что НСИ в таких случаях должны принимать решение о публикации микроданных на основе действующего законодательства и других правил, регулирующих защиту конфиденциальности. В Швеции попытки идентификации данных преследуются по закону.

28. С учетом чувствительного характера микроданных и возможности идентификации НСИ во многих странах не разрешают использовать микроданные вне своих стен. Для решения этой проблемы они создают поисковые лаборатории для исследователей в

рамках своих учреждений. Примером может служить Дания (см. Access to, p. 14). Данный вариант также предусматривает возможность создания и обеспечения функционирования исследовательского центра данных НСИ, например в одном из университетов. В обоих случаях НСИ имеет больше возможностей контролировать соблюдение правил конфиденциальности.

29. Еще одним, более осторожным решением является разрешение использования микроданных только сотрудниками НСИ. В некоторых случаях исследователь может становиться сотрудником НСИ на период времени, необходимый ему для проведения исследований на основе микроданных. Кроме того, НСИ могут иметь программы найма исследовательского персонала для работы с имеющимися микроданными в целях удовлетворения нужд внешних клиентов.

30. Однако более широкое распространение приобретает практика выдачи разрешений определенным исследовательским учреждениям на получение интерактивного доступа к обезличенным микроданным НСИ. Выбор в пользу данного решения сделали Австралия, Дания и Португалия. Наборы данных, предоставляемые в интерактивном режиме, могут подвергаться ограничениям с учетом их чувствительности, а также модифицироваться с учетом конкретных потребностей и специализации деятельности соответствующих исследовательских учреждений. Вполне очевидно, что системы удаленного доступа, такие, как интерактивный доступ через Интернет, высоко ценятся исследователями. Они также могут становиться привлекательными для все большего числа НСИ, поскольку данное решение позволяет осуществлять определенный контроль за использованием микроданных.

31. При обсуждении доступа к микроданным также необходимо учитывать вопрос установления цен. Довольно распространенной практикой является финансирование этой деятельности НСИ за счет ассигнований, выделяемых на проведение значительной части работ по составлению и ведению регистров микроданных. В данном случае представляется вполне разумным, если назначаемая цена соответствует дополнительным издержкам, связанным с публикацией микроданных. Такие издержки могут, конечно же, определяться различными способами в зависимости от применяемых принципов расчета. Однако ценовая политика в области публикации микроданных может также опираться и на другие принципы, такие, как рыночные цены или, даже в определенных условиях, принцип бесплатности.

## VI. ЗАКЛЮЧИТЕЛЬНЫЕ ЗАМЕЧАНИЯ

32. В настоящем документе подчеркивается необходимость установления определенного баланса между требованиями защиты конфиденциальности статистических данных и улучшения доступа к микроданным. В нем также поясняется, что решение этой задачи требует использования определенной комбинации правовых, административных и технических мер. Это также означает, что данный баланс может обеспечиваться различными способами. Статистическое сообщество могло бы выработать общую политику и согласовать базовые принципы в области доступа к микроданным. Выражается надежда, что данный семинар КЕС будет содействовать более глубокому пониманию существующих вариантов.

## СПРАВОЧНЫЕ МАТЕРИАЛЫ

*Access to Microdata in the Nordic Countries* (2003). Statistics Sweden.

Barabba Vincent P. (1974): *The Right of Privacy and the Need to Know*. Proceedings of the Social Statistics Section, American Statistical Association 33.

Boyko Ernie & Watkins Wendy (2002) *Safe Data, Safe Places: No Either/Or Solutions*. Paper to the 19th CEIES seminar 'Innovative Solutions in Providing Access to Microdata'

*CES Seminar on Data Confidentiality* (2003). Note by the ECE secretariat.  
CES/BUR.2003/27/Add.1

*Declaration on professional ethics* (1985). International Statistical Institute.

Erikson Robert (2002): *The right to privacy and the right to information*. Paper to the 19th CEIES seminar 'Innovative Solutions in Providing Access to Microdata'

*National Statistics Code of Practice. Statement of Principles* (2002). National Statistics, UK.

Perpétuo Fernanda (2002) *Statistical Information System for Researchers*. Paper to the 19th CEIES seminar 'Innovative Solutions in Providing Access to Microdata'

Sundgren Bo (2001): *Statistical Microdata – Confidentiality Protection vs. Freedom of Information*. Joint ECE/Eurostat Work Session on Statistical Data Confidentiality. Skopje, FYROM.

Torra Vincenç & Domingo-Ferrer Josep & Torres Àngel (2003): *Data Mining Methods for Linking Data from Several Sources*. Paper to the joint ECE/Eurostat work session on statistical data confidentiality. Luxembourg 7-9 April 2003.

*UN Fundamental Principles of Official Statistics* (1994). Adopted by the UN Statistical Commission.

*Основные принципы официальной статистики ООН (1994 год), утвержденные Статистической комиссией ООН.*

Westergaard-Nielsen Niels (2002): *Linking employer-employee data*. Paper to the 19th CEIES seminar 'Innovative Solutions in Providing Access to Microdata'

-----