

Distr.
GENERAL

CES/SEM.47/9 (Summary)
29 January 2002

RUSSIAN
Original: ENGLISH

**СТАТИСТИЧЕСКАЯ КОМИССИЯ и
ЕВРОПЕЙСКАЯ ЭКОНОМИЧЕСКАЯ КОМИССИЯ**

**КОМИССИЯ ЕВРОПЕЙСКИХ
СООБЩЕСТВ**

КОНФЕРЕНЦИЯ ЕВРОПЕЙСКИХ СТАТИСТИКОВ

ЕВРОСТАТ

**Совместный семинар ЕЭК ООН/Евростата
по интегрированным статистическим
информационным системам и связанным
с ними вопросам (ИСИС-2000)**

(17-19 апреля 2002 года, Женева, Швейцария)

Тема II: Надежные средства связи и конфиденциальность данных

ПРОЕКТ "ВЫЧИСЛИТЕЛЬНЫЕ АСПЕКТЫ СТАТИСТИЧЕСКОЙ КОНФИДЕНЦИАЛЬНОСТИ"

Специальный документ

Представлен статистическим управлением Нидерландов¹

Резюме

Введение

1. Контроль за предоставлением доступа к статистической информации (КДСИ) является одной из областей статистики, которая привлекает к себе в последние годы повышенное внимание. Директивные органы все в большей степени нуждаются в подробной статистической информации. Исследователи способны проводить сложный статистический анализ на своих мощных ПК и желают получать подробные микроданные. С учетом этого к статистическим управлениям предъявляются все более настойчивые требования о публикации высокодетализированной информации. Однако статистические

¹ Автор: Анко Хюндепол (ahnl@krypton.vb.cbs.nl).

институты должны сбалансированно подходить к решению своей задачи в качестве поставщика данных и выполнению своего обязательства по сохранению анонимности респондентов, которые доверили им индивидуальную информацию о себе.

Статистическая информация не может существовать без респондентов.

2. Проект CASC представляет собой инициативу по координации исследований и разработок в Европе. Он частично субсидируется из средств Пятой рамочной программы Европейского союза. Будучи продолжением проекта КДСИ, он направлен на объединение исследований и разработки практических средств - программы ARGUS. Нашей задачей является одновременно решение проблем КДСИ в отношении микроданных и данных в табличной форме.

Микроданные

3. Существующая программа μ -ARGUS уже предоставляет полезные функции защиты социальных микроданных благодаря глобальному перекодированию и локальному исключению. Однако в случае микроданных о предприятиях защита от несанкционированного доступа является намного более сложной задачей. Наличие специальных (несимметричных) распределений в этих наборах данных делает традиционные методы неадекватными. Исследовательские инициативы касаются:

- добавления помех;
- пострандомизации;
- микроагрегирования.

4. Эти методы будут изучаться и внедряться в новых версиях μ -ARGUS. Это создаст возможность для удобного сопоставления результатов различных методов. Впервые будет решена задача обеспечения безопасности микроданных о предприятиях. Тем не менее мы должны проявлять повышенную осторожность и также учитывать "архитектуру безопасности", которая представляет собой центр контролируемого анализа в рамках НСИ.

5. Кроме того, мы планируем изучить модели рисков, которые могут использоваться для расчета степени безопасности отдельных записей и наборов данных.

Данные в табличной форме

6. Текущая версия τ -ARGUS обеспечивает защиту от несанкционированного доступа к неструктурированным статистическим таблицам на основе методов оптимизации. Однако опыт свидетельствует о том, что многие используемые в реальной жизни таблицы обладают более сложной структурой. Большинство перечней кодов имеют иерархическую структуру, что создает большие проблемы защиты этих таблиц. Многие вспомогательные материалы, как, например, таблицы, могут содействовать восстановлению исключенных клеток.

7. Для решения этой проблемы в настоящее время изучается ряд новых подходов. Специалисты работают над распространением текущих методов оптимизации на иерархические таблицы. Мы планируем включить данное решение в новые версии τ -ARGUS. В то же время с учетом того, что сложность этой проблемы может потребовать весьма длительных вычислений для защиты очень крупных таблиц, мы также занимаемся изучением альтернатив для включения в программу. В нее включены решения, опирающиеся на рациональную декомпозицию таблиц, а также решения, опирающиеся на гиперкубы. Это позволит создать программное средство, которое будет способно предложить широкий набор решений, начиная с оптимального решения и кончая быстрым, но не полностью оптимальным решением. Кроме того, методы, разработанные для иерархических таблиц, также будут применяться для решения проблемы связанных таблиц.

Тестирование

8. Большое внимание будет уделяться тестированию программного обеспечения и методологии. Мы можем внедрить много прекрасных методов, однако в конечном итоге важно, чтобы результирующие данные были полезными. В связи с этим изучаются такие вопросы, как "аналитическая способность" набора данных. Кроме того, результирующее программное обеспечение пройдет проверку в реальных условиях использования.

Выводы

9. Мы надеемся добиться значительного прогресса в области как исследований, так и практического применения методов контроля за предоставлением доступа к статистической информации. Дополнительную информацию о проекте можно найти на вебсайте: <http://neon.vb.cbs.nl/casc/default.htm>.
