

NATIONS UNIES

COMMISSION ECONOMIQUE
POUR L'EUROPE

ОБЪЕДИНЕННЫЕ НАЦИИ

ЕВРОПЕЙСКАЯ ЭКОНОМИЧЕСКАЯ
КОМИССИЯ

UNITED NATIONS

ECONOMIC COMMISSION
FOR EUROPE

SEMINAIRE

СЕМИНАР

SEMINAR

СТАТИСТИЧЕСКАЯ КОМИССИЯ И
ЕВРОПЕЙСКАЯ ЭКОНОМИЧЕСКАЯ
КОМИССИЯ



Distr.
GENERAL

CES/SEM.43/4 (Summary)
14 January 2000

КОНФЕРЕНЦИЯ ЕВРОПЕЙСКИХ
СТАТИСТИКОВ

RUSSIAN
Original: ENGLISH

Семинар по интегрированным статистическим
и информационным системам и связанным
с ними вопросам (ИСИС-2000)
(Рига, Латвия, 29-31 мая 2000 года)

Тема I: Создание хранилищ данных и разработка
и использование статистических баз данных в сетевой среде

**ОЦЕНКА СОЗДАНИЯ ХРАНИЛИЩ ДАННЫХ ДЛЯ КОРПОРАТИВНОГО
УПРАВЛЕНИЯ ДАННЫМИ В НАЦИОНАЛЬНОМ СТАТИСТИЧЕСКОМ
УПРАВЛЕНИИ**

Специальный доклад

Представлен Федеральным статистическим управлением Швейцарии¹

РЕЗЮМЕ

I. ПРИНЦИПЫ

1. Согласно Сундгрону, корпоративное хранилище призвано стать опорным компонентом архитектуры национального статистического управления. Оно является центральной базой всех итоговых микроданных (регистры итоговых результатов наблюдений) и всех достоверных макро данных (включая показатели) в совокупности с соответствующими метаданными. Хранилище данных должно предоставлять пользователям инструменты для проведения (в качестве неотъемлемой части своих

¹ Автор: Жорж Флети.

функций) большинства процессов по преобразованию микроданных в макроданные. Что касается метаданных, то хранилище должно содержать и обеспечивать поддержку официальных определений универсальных множеств/генеральных совокупностей, объектов/переменных из всех главных классификаций, описания процессов преобразования, а также перечня источников, видов деятельности и статистических продуктов. Хранилище данных должно служить источником информации для интерактивной базы данных с режимом доступа (в виде платной витрины данных) для внешних пользователей (одной из форм доступа является Интернет), обеспечивающим адекватные возможности поиска и просмотра, и в том числе экспорта данных.

II. ИСКЛЮЧЕНИЯ

2. Хранилище данных не обеспечивает функции обработки данных, предшествующие включению сведений в регистры итоговых результатов наблюдений, за исключением интерфейсов для обмена данными и метаданными между центральной базой и рабочей средой. Кроме того, из нее исключены функции управления регистрами статистических единиц (такими, как коммерческие регистры) и его геокодированными микроданными или привязанными к координатной сетке агрегатами. Кроме того, мы не занимаемся изучением различных путей создания конечных материалов для распространения из хранилища.

III. ФУНКЦИИ

3. Будучи одним из опорных компонентов, хранилище должно обеспечивать поддержку основных функций, необходимых в статистическом процессе с целью сведения до минимума всех рутинных операций. Управление большинством из этих функций должно осуществляться с помощью метаданных. Загрузка первичных данных из различных источников в реляционную базу микроданных должна носить наиболее по возможности простой характер и сводить к минимуму все рутинные операции. Эти микроданные должны затем преобразовываться методом автоматического агрегирования в кубы макроданных в рамках многомерной базы данных. Кубы индикаторов, комбинирующие элементы из различных кубов макроданных, являются третьим уровнем хранилища. Одной из основных функций является доступ ко всем (микро- и макро-) данным по специальным запросам. Микроданные должны быть доступны в интерактивном режиме внутренним пользователям для целей формирования наборов или построения специальных макроданных.

IV. ТРЕБОВАНИЯ

4. С точки зрения языков система должна обеспечивать поддержку различных языков как для с точки зрения руководства работой пользователей, так и метаданных. Обязательным условием является наличие высокоэффективных и удобных для пользователей интерфейсов. Необходимо предусмотреть возможность реализации системы на различных платформах ИТ, а также обеспечить эффективное выполнение всех соответствующих функций.

V. ПОЛИТИКА

5. Необходимым условием успешного и эффективного функционирования хранилища данных в рамках статистического управления является разработка и осуществление целенаправленной политики, направленной на то, чтобы ни один из конечных результатов, имеющих статус официальной статистики и распространяемых в любой форме, не содержал цифры или концепции этих цифр, не присутствующие в макрочасти хранилища. Методика структурирования и определения макроданных, включаемых в хранилище, должна опираться на определенные правила и критерии и являться результатом обсуждения с заинтересованными партнерами; она должна включать в себя определение соответствующих метаданных. Доступ к хранилищу должен быть строго ограничен внутренними пользователями.

VI. ЭТАЛОННОЕ ТЕСТИРОВАНИЕ

6. Эталонное тестирование является одним из полезных средств проверки реального существования предлагаемых функций и эффективности работы различных систем баз данных на всех соответствующих этапах в масштабе всего хранилища с использованием определенной платформы.

7. Ключевые элементы для эталонного тестирования можно подразделить на следующие пять групп:

- архитектура,
- система баз данных (микроданные, макроданные, метаданные),
- функции отбора микроданных и макроданных,
- одни и те же метаданные для микроданных и макроданных,
- преобразование микроданных в макроданные путем агрегирования.

8. Эталонное тестирование, проведенное Федеральным статистическим управлением Швейцарии в 1999 году в рамках проекта создания хранилища данных, дало удивительные результаты. В качестве единой платформы для проведения тестов Центр эталонного тестирования фирмы Compaq в Вальбонне (Франция) предоставил нам четырехпроцессорный 64-разрядный Альфа-Сервер с объемом дискового пространства в 1 терабайт, реализованного по технологии RAID-5. В качестве операционной системы использовалась система Tru64-UNIX. Принять участие в испытаниях дали согласие три партнера – ORACLE, SAS и МЦИ, которым было предоставлено четыре полных дня для осуществления оговоренного процесса.

9. Исходные данные были представлены в виде двумерных файлов ASCII, которые были взяты из результатов переписи населения 1990 года, статистики внешней торговли и выборочного обследования квартирной платы, с единой структурой муниципалитетов. Эти данные были продублированы с помощью специального алгоритма на 50-летний период, в результате чего было получено более 2,2 млрд. записей, которые заняли 106 Гбайт дискового пространства.

10. Сопоставление результатов осуществлялось по критериям необходимого дискового пространства, времени выполнения пакетных задач и скорости ответов на специальные запросы. Различия между тремя системами были значительными: системы ORACLE и SAS были в несколько раз (до 10) медленнее системы WIDAS МЦИ. Данные результаты относятся к контексту использования хранилища данных в рамках статистического управления, который ориентирован не на деловые операции, а на обработку запросов, возможности стандартизации которых являются ограниченными.
