

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.43/7 (Summary)
10 February 2000

КОНФЕРЕНЦИЯ ЕВРОПЕЙСКИХ
СТАТИСТИКОВ

RUSSIAN

Семинар по интегрированным статистическим
информационным системам и связанным с ними
вопросам (ИСИС-2000)

Original: ENGLISH

(Рига, Латвия, 29-31 мая 2000 года)

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

**ОПЫТ СТАТИСТИЧЕСКОГО УПРАВЛЕНИЯ ФИНЛЯНДИИ
В ОБЛАСТИ ПОСТРОЕНИЯ АРХИТЕКТУРЫ ДАННЫХ И
РАСПРОСТРАНЕНИЯ ДАННЫХ ЧЕРЕЗ WWW**

Специальный доклад

Представлен Статистическим управлением Финляндии¹

Резюме

I. Архитектура данных

1. В период 1994-1997 годов основное внимание в рамках проекта по построению архитектуры данных Статистического управления Финляндии уделялось разработке и созданию ориентированной на пользователей модели реализации архитектуры данных в среде ПК/Unix. Этот проект опирался на унифицированную систему файлов и

¹ Авторы: Свен Бьерквист и Пире Тойвонен.

классификационную базу данных, которые были созданы ранее в Статистическом управлении Финляндии в среде больших ЭВМ, а также на соответствующую многомерную базу данных. Одним из приоритетов являлось определение процедур обработки метаданных. Модель реализации состояла из трех компонентов: файлового компонента, компонента статистических продуктов и компонента метаданных. Модель определяла структуру, в соответствии с которой данные статистической системы преобразовывались из первичных данных в статистические продукты, т.е. последовательность преобразования данных статистической системы одной из ступеней трехуровневой иерархии файлов (отдельные файлы, унифицированные файлы и многомерные базы данных) в статистические продукты.

2. Данный проект позволил создать систему унифицированных файлов с описательной базой данных, классификационную базу данных и систему взаимных связей. В рамках этого проекта также была предпринята попытка разработать новую версию многомерной базы выходных данных взамен интерактивной базы данных, которая была создана в 80-е годы в среде больших ЭВМ. Однако данная попытка не оправдала наших ожиданий по причине нехватки ресурсов, в связи с чем в начале 1998 года было принято решение о прекращении деятельности по данному направлению работы.

3. В марте 1998 года Статистическое управление Финляндии приступило к поиску нового решения с целью создания хранилища агрегированных данных для внутреннего использования и распространения информации с использованием технологии WWW и КД-ПЗУ. Было начато осуществление проекта, задачей которого являлась оценка существующих вариантов и разработка предложения по рабочей модели. Была предложена рабочая модель, в рамках которой центральное хранилище данных выполняло роль накопителя агрегированных данных в табличном формате. Осенью 1998 года участники проекта пришли к выводу о том, что средством реализации данной рабочей модели будет являться разработанная Статистическим управлением Нидерландов производственная система StatLine.

4. Также в 1998 году мы приступили к разработке концептуальной базы данных, которая, как и система StatLine, отвечает потребностям определенной нами архитектуры данных. Архитектура данных Статистического управления Финляндии описывает процедуры поступления информации из рабочих баз данных в базы выходных данных. На сегодняшний день модель реализации состоит из файловой системы U-формата с унифицированными файлами и описательной базой данных, классификационной и концептуальной баз данных и базы выходных данных StatLine с соответствующей системой взаимных связей. В Статистическом управлении Финляндии в настоящее время

существует возможность и рекомендуется формировать таблицы StatLine с необходимыми метаданными непосредственно из файлов U-формата.

II. Опыт распространения данных через WWW

5. В начале 1999 года было начато осуществление проекта по внедрению производственной системы StatLine в Статистическом управлении Финляндии. Ранее была проделана определенная подготовительная работа, которая позволила оперативно приступить к осуществлению проекта. Конечный срок реализации проекта установлен на 31 декабря 1999 года. Основные этапы предусматривали создание и начало эксплуатации внутренней служебной системы рабочих данных к январю 1999 года бесплатной службы общественного распространения с большим объемом данных к маю 1999 года и платной службы общественного распространения к началу 2000 года.

6. Поскольку система StatLine была разработана в Нидерландах, было необходимо обеспечить ее адаптацию к рабочей среде Статистического управления Финляндии. Данный проект преследовал следующие основные цели: разработку локальной версии программы для использования в Статистическом управлении Финляндии (определение концепций, перевод программ и справочников); планирование структуры внутреннего хранилища данных и службы WWW; оказание помощи и руководство работой статистических департаментов по разработке содержания системы. Основные цели также включали в себя адаптацию других рабочих инструментов, используемых в Статистическом управлении Финляндии, для обеспечения их успешной интеграции с системой StatLine, определение наилучших методов использования хранилища данных и информационной службы в различных случаях и подготовку необходимых инструкций и справочников. Для достижения этих целей были намечены три основных направления работы, предусматривавших создание технической инфраструктуры, разработку концепций служебной системы и сопровождение служебной системы и системы связи.

7. Проект получил полную поддержку со стороны высшего руководства Управления, и на его годовую реализацию были ассигнованы средства в размере 1,5 млн. финских марок (примерно 254 000 евро). Поскольку данный проект предусматривал преобразование в масштабах всего Управления, его осуществление повлекло за собой расходы в других организационных единицах, вследствие чего ресурсы, задействованные рабочей группой по проекту, составляют лишь небольшую часть от общего объема расходов. В течение первых четырех месяцев в состав рабочей группы входили шесть человек (некоторые из которых работали в ней по совместительству); позднее к ним присоединились еще два человека.

8. Переход от централизованно координируемой интерактивной системы производства распределенной, хотя и по-прежнему централизованно управляемой, интерактивной службе на основе WWW был непростой задачей, в особенности с учетом того, что она должна была быть решена в течение одного года. В связи с этим чрезвычайно важное значение имела поддержка со стороны высшего руководства. Специальная группа, в состав которой вошли 40 сотрудников из различных статистических отделов, сыграла ключевую роль в пропаганде системы в рамках Управления. Без ее содействия проект, скорее всего, потерпел бы неудачу. Кроме того, профессиональная подготовка необходимого минимума поставщиков информации для базы данных облегчила наполнение хранилища необходимыми сведениями. Важное значение также имела увязка новой системы распространения информации с уже существующими в рамках Управления инструментами, что позволило снизить трудозатраты по наполнению системы данными.

9. Интерактивное распространение информации среди широкой общественности требует тщательных пояснений и описаний данных во избежание путаницы, а также адаптации данных к требованиям профессионального использования. Источником пояснений и описаний должны являться централизованные системы метаданных, такие, как наша система файлов U-формата, поскольку в ином случае (повторное) описание данных для опубликования создало бы громадную нагрузку для статистических департаментов.

10. Цели проекта были достигнуты, благодаря чему в настоящее время Статистическое управление Финляндии располагает внутренним хранилищем справочных данных, общедоступной статистической службой WWW и технической инфраструктурой для оказания платных услуг с использованием технологии StatLine. Мы также организовали обучение более 100 сотрудников по вопросам использования данной системы с целью подготовки данных с метаданными для баз данных. Служба общественного распространения содержит около 100 таблиц, насчитывающих в общей сложности 30 млн. ячеек (по состоянию на 15 ноября 1999 года), количество которых быстро растет. Обратная связь с пользователями приносит весьма положительные результаты. Кроме того, в настоящее время другие разработчики официальной статистики в Финляндии проявляют значительный интерес к данной системе. Создание широкого набора данных, распространяемых на бесплатной основе, также в значительной степени содействовало созданию положительного образа Управления в качестве поставщика данных в информационном обществе.

11. В настоящее время система StatLine используется для ежедневного производства статистической информации. Однако для обеспечения успешной интеграции системы

StatLine с нашими производственными процессами необходимо продолжить работу по проектированию, модернизации и стандартизации интерфейсов и потоков данных и метаданных. Решение этих задач и учет непрерывно поступающих пожеланий пользователей будут являться основными направлениями нашей деятельности в предстоящие годы. Распределенная координация не во всех случаях является благом, в связи с чем для координации содержания и структуры нашей службы необходимо по меньшей мере создать своего рода редакционный совет. Полностью централизованная координация также не является идеальным решением, поскольку она ведет к снижению уровня ответственности и позволяет рассматривать обязанности по наполнению и обновлению содержания базы данных в качестве одной из задач этого центрального координационного органа. В связи с этим координация содержания является одной из задач, решение которой требует еще значительных усилий.
