

Distr.
GENERAL

CES/AC.71/2005/16 (Summary)
2 February 2005

RUSSIAN
Original: ENGLISH

СТАТИСТИЧЕСКАЯ КОМИССИЯ и ЕВРОПЕЙСКАЯ ЭКОНОМИЧЕСКАЯ КОМИССИЯ (ЕЭК) ОРГАНИЗАЦИИ ОБЪЕДИНЕННЫХ НАЦИЙ КОНФЕРЕНЦИЯ ЕВРОПЕЙСКИХ СТАТИСТИКОВ **ЕВРОПЕЙСКАЯ КОМИССИЯ СТАТИСТИЧЕСКОЕ УПРАВЛЕНИЕ ЕВРОПЕЙСКИХ СООБЩЕСТВ (ЕВРОСТАТ)**

ОРГАНИЗАЦИЯ ЭКОНОМИЧЕСКОГО СОТРУДНИЧЕСТВА И РАЗВИТИЯ (ОЭСР) СТАТИСТИЧЕСКИЙ ДИРЕКТОРАТ

Совместное совещание ЕЭК/Евростата/ОЭСР по вопросам управления статистическими информационными системами (УСИС)
(Братислава, Словакия, 18-20 апреля 2005 года)

Тема ii): Стратегия развития статистических информационных систем

СТАТИСТИЧЕСКАЯ МЕТАИНФОРМАЦИОННАЯ СИСТЕМА И ХРАНИЛИЩЕ ДАННЫХ В КАЧЕСТВЕ ОСНОВНЫХ ПОДСИСТЕМ СИС - ТЕКУЩЕЕ ПОЛОЖЕНИЕ И ПЛАНЫ

Вспомогательный документ

Представлен Статистическим управлением Польши¹

1. Статистическая информационная система (СИС) представляет собой глобальную систему, которая будет охватывать практически все системы обработки статистических данных в Центральном статистическом управлении Польши (ЦСУ). Статистическая информационная система и хранилище данных являются двумя главными подсистемами, работа по созданию которых велась на протяжении последних трех лет.
2. Одной из главных целей построения СИС являлось сведение воедино существующих статистических распределенных баз данных для обеспечения того, чтобы услуги, предлагаемые официальной статистикой, могли использоваться в любой момент времени за счет улучшения доступа к уже собранной информации.

¹ Авторы: Кшиштоф Курковски (k.kurkowski@stat.gov.pl) и Станислав Силужицки (s.sieluzycski@stat.gov.pl).

3. Хранилище данных облегчает проведение анализа и вторичных наблюдений на основе распределенных источников данных благодаря распространению результатов различных отчетов из одного места, в котором хранятся все необходимые данные. Данные как микро-, так и макроуровня доступны статистикам в режиме онлайн. Данные защищены от повреждений и несанкционированного доступа.

4. В основе функционирования хранилища данных лежат метаданные. Это означает, что управление хранящимися данными осуществляется с использованием метаданных и любое изменение в структуре данных должно предваряться соответствующими изменениями в метаданных. Построение СИС было начато с интеграции метаданных (так называемой "тематической интеграции") и создания хранилища данных на основе метаинформационной системы (так называемой "физической интеграции").

5. Тематическая интеграция означает единое определение терминов и классификаций, используемых для документирования статистических обследований, в то время как под физической интеграцией понимается внутренняя непротиворечивость хранящихся данных. Хранилище данных организовано таким образом, чтобы одинаковые результаты можно получать различными способами, например на основе индивидуальных (микро-) данных или на основе агрегированных (макро-) данных.

6. Метаинформационная система является первой системой, в которой непосредственно будут задействованы статистики. Она будет обеспечивать ведение глоссария терминов, классификаций и других метаинформационных элементов в режиме онлайн. Ведение хранилища данных будут обеспечивать специалисты в области ИТ.

7. Метаинформационная система может рассматриваться в качестве зонтичной структуры, охватывающей все другие системы, т.е. любая метаинформация, используемая другими системами, должна браться только из одного места, т.е. базы метаданных. В свою очередь хранилище данных представляет собой базовую систему, в которой хранятся данные. Однако существуют также другие подсистемы, которые будут включены в состав СИС. Эти системы могут рассматриваться в качестве пользовательских и служебных систем с точки зрения хранилища данных. К пользовательским системам относятся: система разработки и осуществления обследований (включая систему документирования обследований); системы электронных формуляров для ввода, редактирования и проверки данных; система использования данных из административных источников. К служебным системам относятся: система сбора данных, аналитические системы, системы распространения данных (порталы), системы управления контентом (СУК), общедоступное хранилище данных, включая географические информационные системы (ГИС).

8. Системы формирования выборочных совокупностей и документирования обследований будут опираться на метаинформационную систему. Статистики будут иметь возможность создавать выборки с использованием необходимых терминов, а также всей информации, хранящейся в базе метаданных. Они будут задавать условия документирования обследования, включая условия процедур проверки, которые должны выполняться в ходе ввода данных. Как правило, такие условия являются наиболее трудными элементами системы, поскольку они требуют разработки внутреннего языка, интерпретируемого и исполняемого в автоматическом режиме в ходе следующих стадий обработки данных.

9. Система электронных формуляров опирается на базовую предпосылку, согласно которой в ближайшем будущем интернет-системы будут обеспечивать большинство статистических обследований. Хотя такая тенденция является очевидной, существует много проблем с реализацией, и, как это можно увидеть на примере других стран, внедрение таких систем требует времени. Системы электронных формуляров будут обеспечивать получение информации из систем формирования выборки, а также из метаинформационной системы. Проблемы безопасности также имеют важное значение, поскольку индивидуальные данные передаются непосредственно через Интернет. ЦСУ в настоящее время занимается реализацией первой системы (DG1), которая будет охватывать экономические данные и позволит компаниям передавать свои сведения непосредственно через Интернет.

10. Использование административных источников требует в первую очередь заключения большого числа двусторонних соглашений между ЦСУ и другими агентствами. Эти соглашения касаются правовых аспектов, а также технических решений. Использование стандарта XML является одним из типичных рабочих решений.

11. Что касается служебных систем, то широко распространенным подходом является использование системы сбора и анализа данных. Как правило, такие системы зависят от используемой системы управления реляционной базой данных (СУРБД). Хранилище данных ЦСУ реализовано на базе Oracle, вследствие чего статистики используют инструменты данной компании. Средство Oracle Discoverer позволяет онлайнное формирование отчетов непосредственно самими статистиками. Знание структурированного языка запросов (SQL) не обязательно. Средства Oracle Reports используются для решения более сложных задач, которые требуют некоторого программирования (осуществляемого специалистами в области ИТ). Собранные данные могут быть экспортированы в Excel или формат значений с разделительной запятой (CSV). Под аналитическими средствами мы понимаем прежде всего пакет статистического анализа SAS. Увязка Oracle с пакетом SAS еще не обеспечена, но с технической точки зрения это может быть сделано в короткий срок.

12. Последним, но не менее важным компонентом являются системы распространения данных. Наиболее важным из них является портал, поскольку он обеспечивает доступ не только к документам под управлением СУК, но также к данным и метаданным, которые могут непосредственно получаться из общедоступного хранилища данных и метаинформационной системе. ЦСУ в настоящее время готовит тендер на создание нового портала, который должен быть реализован с использованием новейшей технологии, а также опыта и наилучшей практики других стран.
