

**КОМИССИЯ ПО СТАТИСТИКЕ и
ЭКОНОМИЧЕСКАЯ КОМИССИЯ ДЛЯ ЕВРОПЫ**

**КОМИССИЯ
ЕВРОПЕЙСКИХ СООБЩЕСТВ**

КОНФЕРЕНЦИЯ СТАТИСТИКОВ ЕВРОПЫ

ЕВРОСТАТ

**Совместная рабочая сессия ЭКЕ и ЕВРОСТАТА
по статистическим метаданным
(Люксембург, 6-8 марта 2002 г.)**

Рабочий доклад № 6

Тема 1: вопросы инфраструктуры применительно к статистическим метаданным

НЕКОТОРЫЕ ВОПРОСЫ ОРГАНИЗАЦИИ БАЗЫ МЕТАДАНЫХ В НСС АРМЕНИИ.

Представлен: Национальной Статистической Службой Республики Армения (НСС РА)¹

Дополнительный доклад

Резюме.

До настоящего времени развитие автоматизации обработки статистической информации базировалось на создании локальных реляционных баз данных.

В 2001 году НСС РА совместно с представителями фирмы "SYNERGY" начинают разработку концепции реформирования информационной системы статистики Армении. Согласно этой концепции была принята система, ориентированная на потребности пользователя. Предполагается создание единого хранилища статистических данных и, как части этого хранилища, базы данных для прямого доступа к информации через INTERNET. Исполнение этой концепции предполагается разделить на 3 этапа. Корпоративное хранилище данных (DataWareHouse) будет являться ядром в архитектуре хранения данных НСС РА.

С 1999 года круг потребителей статистического продукта НСС РА значительно расширился, и возникла необходимость создания системного описаний статистик.

База метаданных будет содержать описание официальных терминов, принятых в статистике Армении и сопоставление с международными понятиями, описание процессов обработки статистических данных и результирующий статистический продукт. Метаданные должны поддерживать 3 языка: армянский, русский и английский.

Управлением информационных ресурсов и технологий НСС РА разработан программный пакет GENER, формирующий базу метаданных. В базу метаданных вносятся сведения о периодичности обработки информации по форме, перечень организаций, представляющих ее, инструкции по заполнению этой формы. Документ сопровождает справка, описывающая особенности вычислений данной формы.

При распространении данных они будут сопровождаться соответствующими метаданными, графиками и картами для большей наглядности.

¹ Авторы: Вануш Давтян, Наталия Погосян

I. Введение.

1. До настоящего времени развитие автоматизации обработки статистической информации базировалось на создании локальных реляционных баз данных. Развитие технического обеспечения и средств телекоммуникации перевернуло прежние представления о методах получения статистической информации. Пользователь стремится получить информацию в кратчайшие сроки и высокого качества. Разрозненные базы данных уже не отвечают современным требованиям.

2. В 2001 году НСС РА совместно с представителями фирмы "SYNERGY" начинают разработку концепции реформирования информационной системы статистики Армении. Согласно этой концепции была принята система, ориентированная на потребности пользователя. Предполагается создание единого хранилища статистических данных и, как части этого хранилища, базы данных для прямого доступа к информации через INTERNET. Исполнение этой концепции предполагается разделить на 3 этапа.

3. С 1999 года круг потребителей статистического продукта НСС РА значительно расширился, и возникла необходимость создания системного описаний статистик.

II. Основные положения.

4. Опыт международных организаций показывает, что самым оптимальным и надежным средством хранения и использования данных является корпоративное хранилище статистических данных. Принципы организации Хранилища Данных:

- * высокая скорость получения информации;
- * взаимосвязь и целостность данных;
- * возможность получить и сравнить разрезы данных;
- * удобство для пользователя работы с данными;
- * качество данных;
- * защита данных.

5. Корпоративное хранилище данных (DataWareHouse) будет являться ядром в архитектуре хранения данных НСС РА. DataWareHouse будет формироваться на основе центрального реестра объектов наблюдений, для которых осуществляется сбор и обработка данных. Оно будет содержать микроданные и макроданные, которые отражают агрегированные статистические продукты официальной статистики. И микро- и макроданные будут сопровождаться метаданными.

6. Макроданные будут храниться в так называемых macrodata-cubes. Так как процесс формирования макроданных из микроданных должен осуществляться автоматически, возникает вопрос создания базы классификаторов, которые будут участвовать в создании статистического продукта. Если же процесс агрегации включает более сложные функции, алгоритм формирования макроданных будет введен в метаданные.

7. База метаданных будет содержать описание официальных терминов, принятых в статистике Армении и сопоставление с международными понятиями, описание процессов обработки статистических данных и результирующий статистический продукт.

8. Хранилище будет являться и основой для базы данных, выставленной в Internet для внешних пользователей. В связи с этим метаданные должны поддерживать 3 языка: армянский, русский и английский. В хранилище включаются только откорректированные данные.

9. На данных и метаданных Хранилища будут основываться и публикации, которые представляет НСС РА ежемесячно. Концепция структуры Хранилища должна охватывать все показатели официальной статистики, обеспечивать совместимость различных платформ обработки данных, совместимость и целостность статистических данных, а также гарантировать возможность добавления информации с течением времени.

10. Агрегация данных будет производиться в соответствии с классификациями, версии которых соответствуют времени получения статистического продукта. Описание процесса сбора информации обязательно будет присутствовать в метаданных еще до получения окончательных микроданных, особенно в области кодировки данных. Будут содержаться и описания для облегчения поиска и доступа к информации.

11. При распространении данных они будут сопровождаться соответствующими метаданными, графиками и картами для большей наглядности. Стандартные периодические таблицы будут получаться автоматически. DataWareHouse позволит прямой доступ пользователям к базе данных неконфиденциальной статистической информации по запросам. Доступ к микроданным будет осуществляться согласно официально принятым актам.

12. Таким образом, мы получим единый продукт вместо разрозненных множественных баз данных, каждый шаг получения которого будет сопровождаться пояснениями и описаниями.. Это обеспечит повышение качества информации, ускорение получения конечного продукта, облегчение доступа к нужной информации.

III. Потоки информации.

13. В настоящее время НСС РА обрабатывает 239 форм статистической отчетности, а также проводит периодически выборочные обследования. В настоящее время, в основном, бумажные формы отчетности посылаются в районы, где они заполняются, затем передаются в центральный офис. В центральном офисе они контролируются, кодируются, записываются на магнитные носители. Предлагается перенести эти функции на региональный уровень.

14. На первом этапе ввод будет производиться в марзовых офисах, а затем можно будет записывать данные и на региональном уровне. Передача информации будет осуществляться по каналам связи. Сначала это будет модемная связь, после создания корпоративной сети данные будут передаваться по WAN NSS RA. Для ввода данных должно быть подготовлено стандартное программное обеспечение, которое тоже будет распространяться посредством телекоммуникаций. Программное обеспечение должно обязательно сопровождаться классификациями и набором единиц выборки, для которых производятся наблюдения, а также описаниями форм отчетности (инструкциями) и описаниями статистик.

IV. Процесс создания базы метаданных НСС РА.

15. Система метаданных разрабатывается в тесной связи с разработкой Хранилища данных. Будут использованы 2 типа метаданных:

- ссылки для удобного поиска требуемой информации;
- информация, интерпретирующая статистические данные.

16. Управлением информационных ресурсов и технологий НСС РА разработан программный пакет GENER, формирующий базу метаданных. На первом этапе был создан репозиторий всех показателей официальной статистики Армении. В его состав входят:

- * код и наименование показателя;
- * описание показателя;
- * код формы отчетности;
- * ключевое слово для поиска;
- * ссылка на классификатор;
- * ссылка на таблицу агрегации;
- * ссылка на таблицу IMF;
- * дата заполнения;
- * код оператора;

Связанная с показателями, предварительно создана таблица форм отчетности, содержащая:

- * код и наименование формы;

- * периодичность представления;
- * организации, представляющие отчетность;
- * ссылка на инструкции по заполнению;
- * ссылка на справку;
- * дата заполнения;
- * код оператора.

17. Таким образом, в базу метаданных вносятся сведения о периодичности обработки информации по этой форме, перечень организаций, представляющих ее, инструкции по заполнению этой формы. Документ сопровождается справкой, описывающая особенности вычислений данной формы. Следовательно, при работе с некоторой формой сотрудник уже будет обладать достаточной информацией, которую почерпнет из базы метаданных.

18. Введенная метаинформация идентичными ключами связана с базами данных, которые производят статистический продукт. При изменениях в форме отчетности должны быть автоматически произведены изменения в базах данных.

19. Используя базу метаданных, представленный пакет GENER формирует и новые формы отчетности, а также с его помощью возможно произвести изменения в старых формах статистики.

20. Предполагается, что в дальнейшем с развитием вышеуказанной системы выходная информация тоже будет сопровождаться метаданными.

V. Программное и техническое обеспечение.

21. На данный момент технический парк НСС РА позволяет приступить к построению хранилища данных (WareHouse), где предполагается хранить как микроданные, так и макроданные, сопровождаемые метаданными. Для решения этой задачи был выбран программный продукт Microsoft SQL Server 7.0. Программное обеспечение может быть разработано с использованием MS Visual C++ и MS Visual Basic. Для связи с базами данных MS Access и MS SQL будет использоваться ADO(Active Data Object). В архитектуре клиент-сервер модуль серверной программы, преобразующий пользовательский протокол в запрос SQL.RDS (Remote Data Service), является простым и эффективным средством, поддерживающим three-tier систему. Такая система позволяет осуществлять связь клиента и сервера на уровне запросов и наборов данных с использованием мер безопасности.

22. Централизованное Хранилище Данных будет размещаться на SQL сервере в центральном офисе НСС РА. Основой микроданных являются базы Бизнес регистра, Административных регистров и Домашних хозяйств. Данные будут храниться 5 лет. Объем базы - 1000000 записей. Макроданные являются агрегированной информацией, полученной на основании классификаций или дополнительных описаний. В связи с тем, что информация постоянно претерпевает изменения, предполагается хранить версии классификаций, и сохранять в базе данных всю информацию об этих изменениях, а также версии классификаторов. Классификаторы сопровождаются таблицами переходных ключей. С 1 июля 2001г. для агрегации данных по видам деятельности должен использоваться национальный классификатор на базе NACE; по территориям - территориальный классификатор, который был разработан в НСС РА.

23. База метаданных, также расположенная на сервере, будет содержать наименования и определения всех показателей, а также описание методологии сбора данных, и их обработки.

VI. Участие в международных проектах по развитию системы метаданных.

24. Учитывая то, что использование общей системы распространения данных (ОСРД) Международного Валютного Фонда (МВФ) расширит возможности развития сферы статистики, в частности, в предоставлении своевременной, достоверной статистической информации

пользователям, в начале ноября 1999 года НСС РА проинформировала статистическое управление МВФ о желании Армении присоединиться к этой системе.

25. Уже начиная с апреля 1999 года статистические публикации НСС составляются и производятся согласно общей системе распространения данных МВФ и выставлена в INTERNET-е на сайте www.armstat.am.

26. С июля 2000 года начаты работы по сбору необходимой информации для составления таблиц метаданных для основных статистических показателей по формату представленному статистической комиссией Международного Валютного Фонда. Эти таблицы, составленные статистическим управлением (МВФ) специально для основных данных входящих в ОСРД, предназначены для специального сайта ОСРД, в которой планируется иметь метаданные всех стран подписчиков ОСРД. Информация, содержащаяся в этих таблицах (в общей сложности 22 таблицы), будет обновляться каждый год.

27. Основой составления метаданных для Армении служат образцы пилотных метаданных для некоторых стран, в частности Казахстана, которые составлены с помощью специалистов МВФ специально с целью предоставления их в качестве руководства для других стран. Вместе с отделом методологии статистических работ вышеупомянутые работы координируются также внутриведомственной рабочей группой по системе распространения данных, которая в частности координирует работы связанные с получением необходимой информации для составления метаданных из других ведомств (Центральный Банк Республики Армения и Министерство финансов и экономики). Составление таблиц метаданных ведется на английском и русском языках.

28. Осознавая важность таких данных для пользователей статистической информации внутри страны в дальнейшем планируется перевести содержание таблиц на армянский язык для предоставления их пользователям статистической информации внутри страны. Вместе с этим предполагается расширить круг статистических показателей для которых должны составляться метаданные. Предполагается, что они могут предоставляться широкой общественности как по запросу, так и в статистических публикациях.

VII. Распространение информации.

29. Хранилище данных должно поддерживать процесс принятия решений, оно содержит стабильные данные, изменение производится только добавлением данных к существующим. Данные в Хранилище должны быть организованы таким образом, чтобы отвечать на запросы пользователей. Используя OLAP-технологии часть информации будет выставлена в INTERNET-е на сайте www.armstat.am.

30. Модель Хранилища, которое обеспечивает систему принятия решений, основанная на CUBES-технологиях, содержит описания категорий и их значения. Такая модель множественного определения упрощает получение пользователем комплексных запросов, упорядочить данные для отчета, от суммарных данных обратиться к деталям, получать различные срезы данных. Основой для такого решения будет база метаданных.

VIII. Заключение.

31. Учитывая опыт международных статистических агентств в разработке концепции интегрированного статистического пространства, НСС РА считает развитие информационной системы метаданных одной из первоочередных задач. К сожалению опыт сотрудников НСС РА в этой области невелик, и нам бы хотелось изучить более подробно опыт стран, где система метаданных действует уже многие годы. Мы строим систему метаданных НСС РА предполагая ее развитие, как в части добавления и расширения существующей информации, так развития новых положений. Так, перспективы существуют в сфере развития стандартов метаинформации. Внедрив их мы сможем войти в международное информационное статистическое пространство.