

**КОМПОНЕНТЫ СТАТИСТИКИ И
ЭКОНОМИЧЕСКАЯ КОМИССИЯ ДЛЯ ЕВРОПЫ**

**КОМИССИЯ
ЕВРОПЕЙСКИХ СООБЩЕСТВ**

КОНФЕРЕНЦИЯ СТАТИСТИКОВ ЕВРОПЫ

ЕВРОСТАТ

**Совместная рабочая сессия ЭКЕ и ЕВРОСТАТА
по электронному представлению данных
(Женева, Швейцария, 13 - 15 Февраля 2002 г.)**

Рабочий доклад №17

Тема 3: стандарты и концептуальные модели метаданных

ЭЛЕКТРОННОЕ СООБЩЕНИЕ ДАННЫХ (ЭСД), МЕТАДААННЫЕ И ЕВРОПЕЙСКАЯ СТАТИСТИЧЕСКАЯ СИСТЕМА (ЕСС)¹

Заказной доклад

Представлен Юв Кунзлер², Евростат³

Краткое содержание

В настоящем докладе рассматриваются следующие вопросы: что такое ЭСД, что такое метаданные для ЭСД, какие стандарты для ЭСД существуют и какие создаются, насколько причастна к их разработке или участвует в ней ЕСС. Более подробно рассматривается текущая ситуация со стандартизацией ЭОД (электронного обмена данными): переход от ЭДИФАКТА (электронного обмена данными для служб администрации, коммерции и транспорта) к XML (Расширяемому языку описания документов) и принятие к использованию в ЕСС языка ebXML (Расширяемого языка описания деловой информации).

Ключевые слова

ЕСС (Европейская Статистическая система), ЭСД, Метаданные, ЭДИФАКТ, XML, ebXML

I. Введение

В статье 1 Регламента Совета о статистике Сообщества [1] говорится: «На государственном уровне за получение статистической информации Сообщества в соответствии с принципом subsidiarity отвечают национальные органы власти, на уровне Сообщества – орган власти Сообщества. Чтобы гарантировать сопоставимость получаемых результатов, статистические данные Сообщества должны формироваться на основе единых стандартов, а в конкретных требующих того случаях – на основе согласованной методики.»

Сообщество – это Европейский Союз (ЕС). Национальные органы власти – это Национальные статистические институты (НСИ) и другие организации (например, центральные банки, таможня,...), обязанностью которых в государствах-членах ЕС является составление статистических данных Сообщества. Органом власти Сообщества является Евростат, Статистическое бюро Европейских Сообществ в Люксембурге. Все вместе эти учреждения образуют Статистическую систему Сообщества (ССС). Термин Европейская

¹ Взгляды, представленные в настоящем документе, выражают точку зрения автора и не всегда являются отражением политики Европейской Комиссии.

² Юв Кунзлер работает в Евростате, в отделе А-2 по Информационно-коммуникационным технологиям статистической системы Сообщества

³ Евростат является Статистическим бюро Европейских Сообществ. Евростат является Генеральным Директоратом Европейской Комиссии.

Статистическая система употребляется, когда речь идёт ещё и о странах-партнёрах в Европейском экономическом пространстве.

Принцип subsidiarity призван гарантировать принятие решений в максимальной близости от гражданина. Суть этого принципа заключается в том, что Союз принимает какие бы то ни было меры лишь в том случае, если они являются более эффективными, чем те, которые принимаются на национальном, региональном или местном уровнях. Это означает, например, что сбор исходных статистических данных в государствах-членах обычно производят Компетентные национальные власти (КНВ).

Статистический Программный Комитет (СПК) был учреждён в 1989 г. [2] с целью содействия Европейской Комиссии в процессе общей координации многолетних статистических программ для ЕСС [3]. СПК состоит из представителей Национальных статистических институтов (генеральных директоров и президентов) под председательством генерального директора Евростата. Важнейшая часть работы СПК заключается в проведении совместных действий с целью постоянного улучшения деятельности ЕСС. Достижение этой цели во многом зависит от успешной работы систем ИТ по всей ЕСС.

СПК создал Руководящий Комитет по ИТ (ИТРК), который даёт рекомендации по вопросам ИТ. В сентябре 2001 г. ИТРК представил в СПК своё видение ИТ на службе у ЕСС [4]. В представленном документе говорится: "В идеальном мире ЕСС согласовала бы или тесно скоординировала статистические процессы [...] Эти процессы охватывали бы весь жизненный цикл статистических данных, начиная с проектирования вопросника и сбора данных и кончая их производством и распространением."

В документе намечено восемь стратегических целей. Первая из них «Сотрудничество в области стандартизации данных, метаданных и сбора данных». В тексте документа пояснено: «Сбор данных является важной функцией ЕСС, и с расширением ЕС приобретёт ещё большее значение. Большой прогресс в области автоматизации сбора данных позволил повысить его эффективность, однако этот процесс ещё далёк от полного завершения. Важную роль в нём играет стандартизация, без которой автоматизация гораздо более сложна и неэффективна. Важно также удовлетворить растущие требования новых технологий и обновить используемые нами стандарты до версий XML [...]. Действия по стандартизации должны будут охватывать не только техническую стандартизацию (к примеру, разработку стандартных форматов сообщений XML), но ещё и стандартизацию метаданных.»

«Координация наших собственных разработок в области метаданных требует более активного сотрудничества с НСИ и Департаментами статистики организаций-партнёров. Эту работу надлежит проводить по двум параллельным направлениям: гармонизации содержания статистических метаданных и разработки более прогрессивных систем обмена и взаимного использования информации. ЕСС следует одновременно заниматься обоими видами стандартизации ("технической" и "содержательной").

«Цель этой работы - сформулировать общие электронные стандарты (в частности, стандарты на основе синтаксиса XML) и наметить ряд высокоприоритетных действий по стандартизации в области производства, обмена и распространения метаданных, которые, возможно, позволили бы нам повысить эффективность и избежать двойной работы. К тому же, внедрив общую платформу для создания и обмена статистической информацией, мы также смогли бы уменьшить объём отчётности государств-членов».

В том же документе под заголовком "Возможности: новые технологии, Интернет-технология" сказано: «Самая основная явившаяся сейчас возможность – бурный рост Интернета и технологий, связанных с Интернетом (включая XML, микропрограммное обеспечение и т.д.)» Технологии Интернета в значительной мере содействуют решению проблем сопряжения различных платформ. Согласно результатам анализа, проведённого Группой Гартнера, мы вступаем в третью стадию развития Интернета. На первой стадии стала возможной "реклама" в Интернете, для которой главным было "обеспечение собственного присутствия" и возможность контакта с потенциальными покупателями. Она получила название "брошюрного программного обеспечения". Затем появилась возможность покупать и продавать или электронная торговля. Важным её элементом стала возможность взаимодействия с собственными покупателями. Фаза, в которую мы вступаем сейчас, - это использование Интернета для упрощения взаимодействия со своими коллегами. Появляются прикладные программы, которые способны создавать более глубокие отношения между партнёрами, которые предлагают более совершенные способы работы с системой поставок и методы получения и использования знаний (вне зависимости от того, кто является их источником – информационные базы или люди). Эта система получила название "программного обеспечения для коллективной работы". В качестве других примеров можно назвать видеоконференции с помощью настольных компьютеров, программное обеспечение группового пользования и системы управления базами знаний."

В настоящем докладе рассматриваются возможные пути использования технологий и стандартов Интернета при ЭСД в качестве "программного обеспечения коллективной работы". В следующих главах обсуждаются темы "Электронное сообщение данных", "Метаданные" и "Стандарты". Настоящий документ отчасти основывается на ранее представленных докладах ([5], [6]).

II. ЭСД – Электронное сообщение данных

Что такое ЭСД? Существует множество его определений, одни из которых частично включают в себя другие, но редко определяют строго один и тот же феномен, а иногда даже значительно расходятся (см., например, [7]). Я постараюсь несколько сконкретизировать область определения.

В ЕСС сбор данных охватывает две фазы: сбор исходных данных, который в странах-членах ЕС осуществляется в основном государственными властями, и передачу (обычно накопленных данных) в Евростат. ЭСД подпадает под первую категорию – сбор исходных данных.

Сбор исходных данных включает прямой (или первичный) и непрямой (вторичный) сбор данных. Прямой сбор данных базируется на опросах, которые проводятся с респондентами (или их агентами). При непрямом сборе данных используются уже существующие данные, например, с баз данных налоговых или таможенных служб. ЭСД является прямым сбором данных.

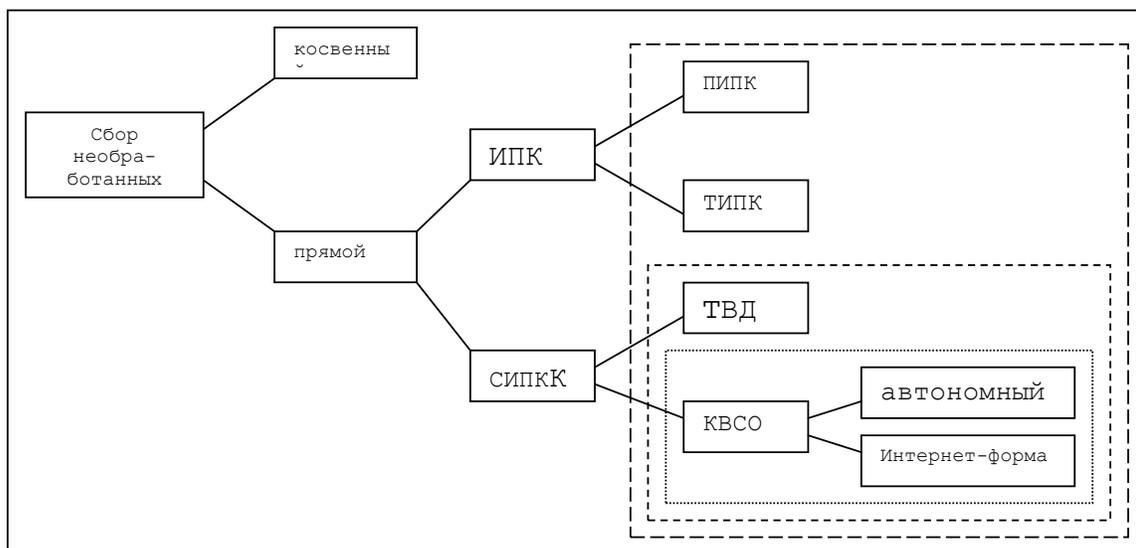


Схема 1: Что такое ЭДС?

Среди методов прямого сбора данных можно выделить методы интервью и самоинтервью. Интервью проводятся при личной встрече либо по телефону. В обоих случаях к проведению интервью можно привлечь компьютер. Эта методика известна под названием ИПК – интервью посредством компьютера (включает ПИПК и ТИПК – соответственно персональное или телефонное интервьюирование посредством компьютера).

Самоинтервью основаны на вопросниках, которые раздаются респондентам. Традиционно, письменный вопросник в этом случае высылается респонденту по почте; тот его заполняет и возвращает в коллектор данных. Для максимальной автоматизации этого процесса сегодня используется информационно-коммуникационная техника, а сам процесс называется СИПК – самоинтервью посредством компьютера. СИПК включает электронные вопросники для самоинтервью и ТВД – тональный (телефонный) ввод данных. Электронные вопросники СИПК часто называют КВСО – компьютеризированный вопросник для самоопроса. Их взаимодействие видно на 1-ой схеме. Я буду использовать термин ЭДС применительно к прямому сбору первичных данных посредством электронных вопросников для самоинтервьюирования (КВСО). В более широком смысле ЭДС могло бы включать в себя ТВД, а также, возможно, и ИПК.

Существуют две основные технологии вопросников: в первом случае программные средства устанавливаются на автоматизированном рабочем месте респондента, во втором случае требуется лишь наличие Интернет-связи и обычного Интернет-браузера. В настоящем докладе я буду называть первый тип системы «автономной системой КВСО», а второй «Интернет-формой».

Автономная система КВСО может использоваться без подключения к Интернету, к примеру, через магнитные средства передачи данных. Такие системы КВСО существовали уже до свершения Интернет-революции. С их помощью статистические данные можно регистрировать и упаковывать в формат, приемлемый для коллектора данных. В современных автономных системах КВСО Интернет находит всё большее и большее применение: данные пересылаются через Интернет, корректировки программ загружаются с коллекторов данных Интернет-сайта, на Интернет-сервере можно просмотреть длинные списки кодов, и т.д.

Удачным примером автономной КВСО является IDEP/CN8 ([8], [9]), электронный вопросник для декларации Интрастата. Интрастат – статистическая система для данных, связанных с торговлей между государствами-членами ЕС. IDEP/CN8 разработан и используется

Евростатом. В большинстве стран-членов существует его версия на языке страны-пользователя. Его распространением и обеспечением пользователей занимаются Госадминистрации. В настоящее время IDEP/CN8 используется более, чем пятьюдесятью тысячами предприятий. Существование и успех IDEP/CN8 явились результатом тесного и плодотворного сотрудничества заинтересованных партнёров в ЕСС.

Тогда как традиционные автономные системы КВСО, такие как IDEP/CN8, содержат только один статистический опрос, системы КВСО с множественными вопросниками можно использовать для одновременного проведения нескольких параллельных опросов. По мере надобности в них можно включать новые вопросники, которые можно загружать, например, из Интернета. ЭДИСЕНТ, одна из первых систем с множественными вопросниками, была разработана в ходе исследовательского проекта ТЕЛЕР (система оценки медобслуживания по методу Л.Е.Ру) (1996-1999г.г.) [10]. Ещё более поздняя разработка сейчас используется в Австрии (это система множественных электронных вопросников e-Quest, к которой в апреле 2001 г. был открыт общественный доступ) [11]. Австрийское статистическое бюро представит e-Quest на этом рабочем заседании.

Интернет-формы – электронные вопросники, не требующие установки программного обеспечения на компьютере провайдера. Интернет-форму можно найти на коллекторе данных Интернет-сервера; её можно изучить и заполнить с помощью обычного Интернет-браузера. Этот тип решения всё больше предлагают статистические бюро стран-членов ЕС [12]. В качестве примера можно назвать Интернет-форму Интрастата, используемую в Великобритании (существует с 1998 г.) [13], или немецкую систему w3stat (существует с 1999 г.) [14].

Интернет-формы на основе XML также разрабатываются в ходе исследовательского проекта IQML (2000 - 2003) [15], который будет отдельно представлен на этом рабочем заседании.

III. Метаданные

Что такое метаданные? Точнее, что означают метаданные в нашем статистическом контексте? В течении последних 30 лет этот вопрос явился объектом обсуждения множества материалов и конференций. Различные органы и учреждения представляли разные определения статистических моделей метаданных, но международное согласие в отношении него не было достигнуто. Недавно шесть солидных международных организаций – БМР (Банк международных расчётов), ЕЦБ (Европейский центральный банк), Евростат, МВФ (Международный валютный фонд), ОЭСР (Организация международного сотрудничества и развития) и ООН (Организация Объединённых наций) – решили объединить свои силы и заняться «изучением общих электронных стандартов и производимых в настоящее время действий по стандартизации, которые позволили бы нам избежать дублирования усилий в своей собственной работе и, возможно, в работе других организаций в области статистической информации» [16]. 6-7 сентября 2001 г. в Вашингтоне, федеральный округ Колумбия, проводилась встреча группы по ОСМД (Обмен статистическими и метаданными), целью которой был обмен информацией и начало общей работы в области статистических метаданных. Эта работа только что начата, поэтому важно проследить, чтобы эта инициатива ОСМД включила в себя и ЭСД.

Существуют различные типы метаданных. Метаданные можно использовать для поиска данных (метаданные выборки данных); метаданные могут описывать параметры статистических данных (статистические метаданные); метаданные можно использовать в ИТ-системах для организации, обработки и обмена данными (метаданные инфраструктуры ИТ). Существуют и другие типы метаданных, каждый из которых в свою очередь можно

подразделить на более конкретные типы (более подробно об этом можно узнать, к примеру, в [17] и [18]).

В нашем случае мы задаём вопрос: «Что такое метаданные в ЭСД?» Чтобы ответить на этот вопрос, я сначала рассмотрю другой вопрос: «Как функционирует ЭСД?» Как объяснялось выше, в настоящем докладе говорится об электронных вопросниках, так называемых системах КВСО. Существуют автономные системы КВСО и Интернет-формы. Автономные системы КВСО могут содержать один (как IDEP/CN8) или несколько вопросников (как ЭДИСЕНТ или e-Quest). В настоящем докладе будут рассмотрены лишь метаданные, связанные с ИТ; а статистические метаданные, такие как данные об участниках опроса, выборах, методике и т.д., рассматриваться не будут.

Опрос на основе ЭСД начинается с проектирования вопросника. Электронный вопросник, предлагаемый респонденту, строится на ряде различных метаданных вопросника:

- Метаданных, описывающих сам вопросник (их можно назвать структурными метаданными вопросника): вопросы, варианты ответов, тексты заставок, информация о маршрутизации и т.д.
- Метаданные формата вопросника: шрифты, позиции, цвета, логотипы, фоны, и т.д.
- Метаданные, помогающие респонденту заполнять вопросник, среди которых могут быть:
 - тексты подсказок, сообщения об ошибках, предупреждения
 - системы условных обозначений, классификации, списки кодов
 - правила проверки достоверности данных
 - правила самостоятельного заполнения
 - правила получения выходных данных (данные на вывод)
 - заранее внесённые данные (их, однако, можно рассматривать как данные, а не как метаданные)

Этот список не является полным и не претендует на роль эталонного определения; он приведён в качестве примера, дабы очертить рамки нашей дискуссии.

В автономной системе КВСО, содержащей один вопросник, метаданные вопросника обычно встраиваются в программные средства КВСО. Среди этих средств и функциональные средства, которые обеспечивают функцию воспроизведения текста на экране. Во всех остальных случаях респондент просматривает вопросник с помощью программного обеспечения для воспроизведения вопросника (множественные вопросники в автономных КВСО) либо с помощью стандартного браузера (Интернет-форма). В этих случаях программные средства воспроизведения и метаданные вопросника разделены. В последнем случае существуют некоторые преимущества: для множества различных вопросников требуется лишь одно программное обеспечение, цикл обработки данных более краток, следовательно, и затраты ниже. Файлы вопросников гораздо меньше, чем сложный комплект программного обеспечения, а это значит, что его распространение проще и дешевле.

На следующей стадии жизненного цикла электронного вопросника возможны три варианта: (а) респондент заполняет его вручную, (б) файлы генерируются специальными программными средствами и импортируются, (в) вопросник имеет прямой доступ к существующим базам данных, откуда выбираются данные. Во всех этих случаях автоматически применяются правила проверки достоверности данных (метаданные!). Вариант (б) и особенно вариант (в) требуют большего обеспечения метаданных; требуется определить формат импортируемых данных, а в варианте (в) ИТ системы должны знать, как общаться друг с другом, чтобы идентифицировать требуемые данные.

Когда вопросник заполнен, данные передаются в коллектор. Передача данных производится через ППГТ, электронную почту или другим доступным способом передачи файлов. Возможны два варианта формата: можно сдать полное описание первоначального вопросника, включающего информацию, внесённую респондентом, а можно сгенерировать и отослать конкретный (менее объёмный) файл с выходными данными. В последнем случае для конкретизации выходных данных потребуются метаданные.

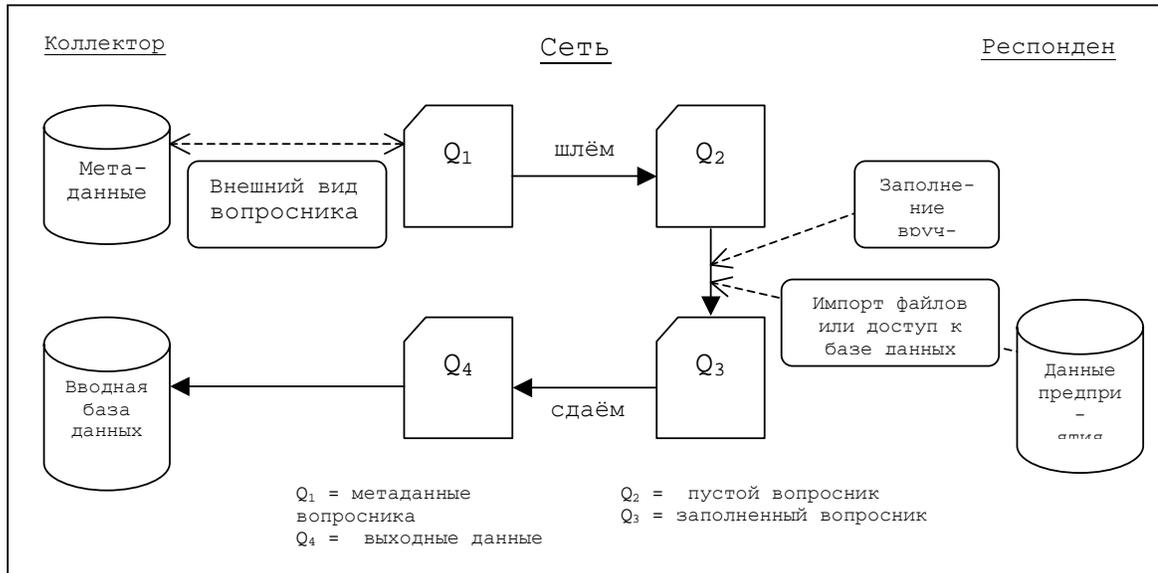


Схема 2: Как функционирует ЭСД?

Этот механизм показан на схеме 2. Он включает различные типы метаданных ЭДС, в том числе:

- Метаданные вопросника, которые выслаются респонденту, включая структурные метаданные, метаданные формата и вспомогательные метаданные, такие как правила проверки, классификации и т.д.;
- Метаданные для выборки данных, с помощью которых осуществляется автоматический доступ к базам данных;
- Обмен данными или метаданные ЭОД.

Метаданные ЭОД являются наиболее известным типом метаданных в контексте ЭСД. ЭДИФАКТ и ANSI X.12 используются уже давно. XML стремительно выходит на сцену. В следующем разделе мы больше узнаем о стандартах.

IV. Стандарты

Мне бы хотелось начать с рассмотрения стандартов для метаданных ЭОД. В общем смысле ЭОД (Электронный обмен данными) означает упорядоченный обмен информацией между двумя прикладными компьютерными программами без человеческого вмешательства. Он применяется в основном для автоматизации сообщений от предприятия к предприятию (B2B), например, для автоматической пересылки заказов на поставку товара, счетов и т.д. ЭОД находит применение и в статистической сфере. Обмен накопленными данными может происходить между двумя статистическими учреждениями, либо лицо (учреждение), предоставляющее исходные данные, может переслать их прямо в коллектор с помощью технологии ЭОД.

Традиционным стандартом для ЭОД является ЭДИФАКТ (а в Америке ANSI X.12). Но XML занимает всё более уверенные позиции. Как это сказывается на сборе статистических данных в ЕСС?

ЭДИФАКТ был доинтернетовским стандартом для ЭОД. Для него типичными средствами сообщения были закреплённые линии или сети с дополнительными услугами (СДУ). Бурный рост Интернета совсем не обязательно означает конец ЭДИФАКТа, поскольку сообщения ЭДИФАКТа можно с равным успехом посылать и через Интернет, например, в качестве файлов-приложений в электронной почте или через передачу файлов.

ЭДИФАКТ стал стандартом ООН в 1987 г. С самого начала ЕСС принимала активное участие в разработке определения сообщений ЭДИФАКТ. ЭДИФАКТОм занимается ЕУСЭ, Европейское управление по стандартизации ЭОД. ЕУСЭ состоит из нескольких экспертных групп согласно различным сферам, которые определяют сообщения ЭДИФАКТ. Вопросами статистики в ЕУСЭ занимается Экспертная группа 6 (ЕЭГ6).

Для конкретных областей статистики были приняты подмножества соответствующих сообщений ЭДИФАКТ, разработанные другими компетентными органами. Например, в статистике внешней торговли используются подгруппы сообщений для таможи, разработанные Экспертной группой ЕУСЭ по вопросам таможи (ЕЭГ3). Однако статистическое сообщество, т.е. ЕЭГ6, также разработало свои собственные универсальные сообщения ЭДИФАКТ: для многомерных данных или хронологических последовательностей – ГЕСМЕС, для обмена классификациями – КЛАСЕТ, а для сбора первичных данных, т.е. ЭОД, – РДРМЕС. Над каждым из этих видов сообщений работает конкретная рабочая группа внутри ЕЭГ6.

В некоторых областях ЭДИФАКТ нашёл широкое распространение. Европейская система центральных банков (ЕСЦБ) приняла ГЕСМЕС (а точнее, конкретный диалект ГЕСМЕС/СВ) в качестве собственного формата передачи данных. Обмен данными между Евростатом и его партнёрами (НСИ, ЕЦБ и другими) также основывается на ГЕСМЕС. Более 50000 предприятий используют сообщение ЭДИФАКТ для отчётов Интрастату (КУСДЕК/ИНСТАТ); ежегодно рассылается несколько миллионов таких сообщений.

В других областях ЭДИФАКТ прижился хуже. В особенности эта ситуация характерна для малых и средних предприятий (МСП). Они не спешат внедрять решения ЭДИФАКТ по нескольким причинам. Высокая сложность, высокие затраты, малая гибкость – вот лишь некоторые из них. Эти трудности можно преодолеть с помощью нового подхода, специально созданного в ответ на потребности Интернета, – с помощью XML (Расширяемого языка описания документов).

Чаще всего путают три языка – это SGML, HTML и XML. XML – самый молодой в этом ряду. Для чего он был разработан? Почему в качестве языка Интернета не был принят SGML? Почему было не остановиться на XHTML?

SGML (Стандартный обобщённый язык описания документов) – это общий язык описания документов. Он зародился до появления Интернета (принят МОС в 1986 г.). SGML позволяет давать определение различных типов документов, или иными словами, различных языков, описывающих конкретные типы документов. В этом смысле SGML является метаязыком. К тому же SGML довольно сложен: он слишком сложен для Интернета, что делает его применение дорогим, а прикладные программы медленными.

HTML (Язык описания гипертекстовых документов) был разработан в качестве языка описания документов для Интернета (около 1990 г.). HTML является конкретной прикладной программой SGML, т.е. определяет один фиксированный тип документа. У HTML есть один недостаток: он не поддается расширению, поэтому всегда найдутся прикладные программы, которые нельзя базировать на HTML.

XML был создан чтобы преодолеть проблемы, связанные с SGML (его чрезвычайную сложность) и HTML (негибкость). XML - это подмножество SGML, созданное с целью применения функциональных возможностей SGML в Интернете. Как и SGML (чем отличается от HTML), он является метаязыком, который позволяет для достижения каждой конкретной цели задавать собственный язык.

XML 1.0 был принят Консорциумом Глобальной гипертекстовой системы Интернет (W3C) в 1998г. XML отлично подходит для сетевого ЭОД, т.е. для ЭОД в Интернете. XML поддерживают крупнейшие компании по производству средств программного обеспечения (Микрософт, АйБиЭм, Сан, Оракл,...) и международными административными органами (ОЭСР, МВФ, ЕЦБ, ООН,...). Другое преимущество XML заключается в том, что он обеспечивает связь между предприятием и потребителем (B2C), что было не под силу ЭДИФАКТу. Какой человек или какая семья согласилась бы установить у себя дома дорогой и сложный ЭДИФАКТ (который пока что никто бесплатно не предлагает)? А XML доступен через обычный браузер. Формами XML, которые предприятие представляет на своём сайте в Интернете, может воспользоваться кто угодно – потребителю не надо платить за них дополнительно.

Так что XML – идеальный кандидат в перспективные средства ОЭД. Но XML - это только основа. Для сетевых ОЭД требуется большее. XML – всего лишь синтаксис, семантика определяется отдельно. Здесь и начинается ebXML.

EbXML (XML электронной торговли) – это совместное предприятие ОАСИС (Организация по разработке стандартов структурной информации), производителя программного обеспечения. Оно включает Микрософт, АйБиЭм, САН, САП,...) и ООН/СЕФАКТ (Центр ООН по Упрощению торговли и вопросам электронной коммерции), организации по разработке стандартов международной торговли, ответственной за ЭДИФАКТ. ebXML определил основные рамки XML (базовые компоненты, деловые процессы, услуги регистра,...) применительно к электронной торговле. 14 мая 2001 г. в прессу поступило сообщение о том, что характеристики ebXML получили одобрение [19]:

"Женева, Швейцария, и Бостон, штат Массачусетс, США; 14 мая 2001г. – сегодня ООН/СЕФАКТ и ОАСИС объявили, что 11 мая 2001г. участники встречи в Вене (Австрия) со всего мира, одобрили характеристики ebXML. ebXML родился как инициатива сроком на 18 месяцев, которую финансировали ООН/СЕФАКТ и ОАСИС. ebXML - это модульный комплект характеристик, с помощью которого предприятия любого размера в любой географической точке могут заниматься торговой деятельностью посредством Интернета. В настоящий момент для компаний использование ebXML - это стандартный метод обмена деловыми сообщениями, поддержание деловых связей, передача данных в одинаковых выражениях, а также определение и регистрация коммерческих процессов."

EbXML получает поддержку со стороны ЕУСЭ. Чтобы отразить этот факт, а также то, что в будущем центром его интереса станет XML, ЕУСЭ решил сменить своё название. Теперь он называется Управление электронной коммерции по вопросам евростандартизации (УЭКЕС). Подобные изменения происходят и в других органах стандартизации, также на структурном уровне. Бывшая ЭРГ (Рабочая группа ООН/СЕФАКТ по ЭДИФАКТу) станет этРГ (Рабочей

группой по электронной торговле) и будет отвечать за нетехнические аспекты eXML: базовые компоненты, коммерческие процессы и разработку сообщений.

Развитие и внедрение версий статистических сообщений, используемых в ЕСС, и совпадающих с eXML, объявлено стратегией Евростата. В основу разработок лягут модели данных УМІ (Унифицированного языка моделирования). Инструментальные средства передачи и распространения данных, применяемые Евростатом, такие как Стадиум или НьюКронос, будут использоваться в качестве обеспечения XML. В то же время существующие решения ЭДИФАКТа будут продолжать своё существование так долго, как это будет необходимо.

eXML не является единственным семантическим стандартом на основе XML. В принципе, каждый вид деятельности может иметь свой конкретный набор стандартов, построенный на основе XML. Некоторые из этих стандартов можно было бы использовать в статистических целях: например, статистические вопросники могли бы автоматически производить опрос соответствующих им баз данных.

Расширяемый язык деловой отчётности (XBRL) является спецификацией на основе XML и служит для подготовки и обмена финансовыми отчётами и данными [20]. Развитие XBRL идёт вперёд благодаря бухгалтерам, в особенности Американскому Институту дипломированных аудиторов. XBRL использует принятые стандарты и практику финансовой отчётности для обмена финансовыми отчётами посредством любых средств программного обеспечения и технологий, включая Интернет. В 2001 г. Евростат создал Рабочую группу по вопросам бухгалтерии. Она будет помогать проведению действий ЕСС в различных областях, включая сбор и интерпретацию полученных от компаний коммерческих данных. Ввиду этого рабочей группе непременно придётся изучить достижения XBRL.

Проект по IQML (комплект ПО и XML-стандарт для интеллектуальных вопросников) развивает Метамоделю общего хранилища данных (ОХМ)[21], которая проявилась в результате объявления о принятии заявок, сделанного Рабочей группой по управлению объектами в 1998 г. Среди участвующих в проекте партнёров – АйБиЭм, Унисис, Оракл и другие. "Целью Инициативы по общему хранилищу метаданных (ИОХМ) является упрощение обмена метаданными между инструментальными средствами доступа к данным и хранилищами данных в распределённых неоднородных эксплуатационных условиях."

Реализация проекта по Xформам была начата Консорциумом Глобальной гипертекстовой системы Интернет (W3C) [22]. Он утверждает, что Xформы - это язык XML для нового поколения Интернет-форм. "Xформой W3C называет спецификацию Интернет-форм, которую можно использовать с различнейшими видами платформ, включая настольные компьютеры, портативные приборы, информационные устройства и даже бумагу."

Существует ещё великое множество прикладных программ и производственных инициатив по XML (см., например, [23]). Поскольку eXML избран в качестве семантического стандарта XML для ЕСС, важно не терять из виду и остальные стандарты. В определённых обстоятельствах они могут стать более выигрышным решением, так что необходимо хотя бы знать, где эти решения смогут выгоднее проявить себя.

Как мы видим, в области метаданных ЭОД проводится широкая деятельность по стандартизации. XФормы являются одной из инициатив по стандартизации метаданных вопросников, включая вопросы, типы ответов, правила проверки достоверности данных, классификации и т.д. Внутри ЕСС существуют также проекты, связанные с метаданными вопросников.

На стадии проектирования вопросника метаданные вопросника можно сохранять в хранилище метаданных до следующего их использования или извлекать из этого хранилища для составления вопросника. Такой вариант решения рассматривается в проекте IQML [24] (иной подход представлен в [25]). В ходе исследовательского проекта ТАДЕК была рассмотрена другая проблема: проблема автоматической генерации документов электронных вопросников [2].

Такие инициативы, как XBRL или OXM продвигаются в направлении автоматизированной выборки данных на основе стандартизированных метаданных. Автоматическая выборка данных также является основой «Семантической глобальной "паутины"». Это выражение придумал Тим Бернерс-Ли, изобретший Глобальную гипертекстовую систему Интернет. «Семантическая глобальная "паутина"» не есть отдельная сеть, она является продолжением существующей Глобальной "паутины". В ней информация получает чётко определённое значение, предоставляя больше возможностей компьютерам и людям работать сообща. Уже делаются первые шаги на пути вплетения Глобальной семантической "паутины" в структуру уже существующей Глобальной "паутины.» Движущей силой Глобальной семантической "паутины" являются метаданные, выраженные через XML, COP (Систему описания ресурса) и антологии [27]. Возможно, в недалёком будущем эти идеи и действия заслужат более пристальное внимание со стороны ЕСС.

V. Выводы

Интернет-революция изменяет наш способ общения, в том числе и в области статистического ЭСД. В качестве стандарта обмена документами и данными в Интернете на первый план выступает XML. Основные стандарты, базирующиеся на XML, такие как eXML и XBRL, в будущем окажут огромное влияние на автоматизацию сбора данных.

В то же время организацию метаданных всё больше начинают воспринимать в качестве критического фактора успеха, в том числе и в отношении работы статистических бюро. Многообещающими являются инициативы типа SDMX.

Совместив Интернет-стандарты с гармонизацией метаданных, можно было бы создать «Глобальную семантическую статистическую "паутину"», охватывающую весь жизненный цикл статистических данных – от их сбора к созданию и затем к распространению. ЕСС уже начала работать в этом направлении.

Список использованной литературы

- [1] Регламент Совета (ЕС) о статистике Сообщества № 322/97 от 17 февраля 1997г.. ОЖ 52, 22.2.1997 г.
- [2] Решение Совета (89/382/ЕЕС, Евратом) от 19 июня 1989 г., которым учреждён Комитет по статистическим программам Европейских Сообществ. ОЖ 181, 28.6.1989 г.
- [3] Решение Совета (ЕС) № 1999/126 от 22 декабря 1998 г. о статистической программе Сообщества на 1998 - 2002 г.г. ОЖ 42, 16.2.1999 г.
- [4] CPS 2001/42/14/EN: стратегия сотрудничества в области ИТ в ЕСС. 42-ое заседание СПК, Люксембург, 19 - 20 сентября 2001 г.
- [5] Стратегия Электронного сбора данных (ЭСД) в Европейской статистической системе (ЕСС). Юв Кунзлер, Евростат. Конференция NTT/ЕТК, Крит, Греция, июнь 2001 г.
- [6] Влияние Интернета на сбор и распространение информации в Европейской статистической системе (ЕСС). Вольфганг Кньюппель и Юв Кунзлер, Евростат. Дополнительное заседание IAOS по вопросам статистики в Информационном обществе, Токио, Япония, август 2001 г.

- [7] Методика проведения Интернет-опросов и других телеопросов. Гад Натан, Древнееврейский Университет, Ерусалим. Конференция NTTS/ЕТК, Крит, Греция, июнь 2001 г.
- [8] IDEP/CN8 – Европейский пакет средств программного обеспечения. Евростат, август 1998 г.
- [9] Электронный сбор исходных данных (eCoRD) – Европейская перспектива. Юв Кунзлер, Евростат. Семинар по ЭСД, Гулль, Канада, сентябрь 2000 г.
- [10] Заключительный отчёт ТЕЛЕР. Филипп Кай (Сезия), Луи- Эме де Фукьер (Сезия) и Ганс Штоль (Статистическое бюро Нидерландов). Май 1999 г.
- [11] Электронный опрос: использование средств программного обеспечения на основе метаданных для электронного сбора данных в Статистическом бюро Австрии. Вольфганг Коллер и Гюнтер Цеттль, Статистическое бюро Австрии. Конференция NTTS/ЕТК, Крит, Греция, июнь 2001 г.
- [12] Интернет-формы для Интрастата. Евростат, Март 2000 г.
- [13] <http://www.hmce.gov.uk>
- [14] <http://www.w3stat.de>
- [15] <http://www.epros.ed.ac.uk/iqml/>
- [16] Рабочая группа по обмену статистическими данными и метаданными (ОСДМ) – Совместное заявление учреждений-участников. Люксембург, май 2001 г.
- [17] Обмен оптимальными методами и ноу-хау с целью совершенствования создания и использования метаданных. Джоанна Лэмб, Эдинбургский Университет. Конференция NTTS/ЕТК, Крит, Греция, июнь 2001 г.
- [18] Выработка единого понимания стандартных компонентов метаданных: статистический глоссарий. Денис Уорд (ОЭСР) и Марко Пеллегрини (Евростат). Семинар по вопросам обмена статистическими данными и метаданными (ОСДМ). Вашингтон, федеральный округ Колумбия, сентябрь 2001 г.
- [19] <http://www.ebxml.org>
- [20] <http://www.xbrl.org>
- [21] <http://www.cwmforum.org>
- [22] <http://www.w3.org/MarkUp/Forms>
- [23] <http://xml.coverpages.org>
- [24] Интеллектуальное использование метаданных в проектировании вопросника. Карен Браннен, Эдинбургский Университет. Конференция NTTS/ЕТК, Крит, Греция, июнь 2001 г.
- [25] Составные элементы связи: знаки контекста в проектировании вопросников. Стивен Дженкинс (Меркатор Компьютер Системз Лимитед) и Тони Соломонидес (Университет Западной Англии). Конференция NTTS/ЕТК, Крит, Греция, июнь 2001 г.
- [26] ТАДЕК, Инструментальное средство анализа и документирования электронных вопросников. Елке Бентлем и Анко Хандепул, Статистическое бюро Нидерландов. Конференция NTTS/ЕТК, Крит, Греция, июнь 2001 г.
- [27] Глобальная семантическая "паутина". Тим Бернерс-Ли, Джеймс Хендлер и Ора Лассила. Сайнтифик Америкэн, май 2001 г.