

Distr.
GENERAL

CES/AC.71/2005/23 (Summary)
3 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 iii): XML и вебсерверы

ПРЕДСТАВЛЕНИЕ СТАТИСТИЧЕСКИХ ДАННЫХ В РЕЖИМЕ ОНЛАЙН С ИСПОЛЬЗОВАНИЕМ XML

Вспомогательный документ

Представлено Управлением национальной статистики Соединенного Королевства¹

Резюме

I. ВВЕДЕНИЕ

1. Настоящий документ содержит резюме полной версии документа, посвященного описанию того, каким образом в рамках проекта "Районная статистика" Управление национальной статистики (УНС) Соединенного Королевства широко применяет XML в программном обеспечении, используемом для представления статистических данных в

¹ Автор: Дайянтха Джошуа (dayantha.joshua@ons.gsi.gov.uk).

Сети, а также того, каким образом в ходе проектирования программного обеспечения предусматривается потенциальное повторное использование в качестве вебсервиса.

II. ВЕБПРОЕКТ "РАЙОННАЯ СТАТИСТИКА": СПРАВОЧНАЯ ИНФОРМАЦИЯ И ОХВАТ

2. Осуществление программы "Районная статистика" (NeSS) было начато в 2000 году, и главной ее целью является предоставление более совершенной информации о планах правительства в области реконструкции жилых районов.

3. Одним из результатов NeSS станет создание общедоступного вебсайта, который позволит пользователям просматривать статистические данные, собираемые из широкого круга правительственных источников Соединенного Королевства, в том числе проводящейся раз в десять лет переписи УНС. Данные имеются в наличии по полному набору демографических уровней, начиная с небольших групп домохозяйств и кончая региональным и национальным уровнями.

4. За период с 2001 года было создано три версии вебсайта, который постепенно пополняется новыми функциями и данными. Четвертую версию запланировано опубликовать весной 2005 года. Четвертая версия была написана с чистого листа с использованием технологий Java J2EE и Oracle и обеспечивает всестороннее удовлетворение требований NeSS.

III. ТРЕБОВАНИЯ К ПРЕДСТАВЛЕНИЮ ДАННЫХ

5. Перечень требований NeSS к представлению статистических данных можно кратко охарактеризовать следующим образом: в виде таблиц HTML; в виде графиков - гистограмм и линейных графиков; тематических карт (с использованием масштабируемой векторной графики - SVG²); данных в форме динамических электронных таблиц (в формате CSV).

IV. ПОДХОД К (СТАТИСТИЧЕСКИМ) ВЕБСАЙТАМ КАК НАБОРУ ВЗАИМОДЕЙСТВУЮЩИХ СЕРВИСОВ

6. Вебсайты могут рассматриваться в качестве комбинации взаимодействующих *сервисов*. Речь идет о сервисах в широком смысле этого понятия, а не конкретно о вебсервисах, хотя данная технология может являться частью возможного решения. На

² <http://www.w3.org/TR/SVG/index.html>, <http://www.abode.com/svg>.

высоком уровне набор предлагаемых УНС статистических вебсервисов может выглядеть следующим образом:

- Общая навигация и контент, включая жизненный цикл публикации данного контента. Эти сервисы могли бы предоставляться с помощью одной из коммерческих систем управления контентом (СУК).
- Более развитая система навигации (например, с использованием статистической таксономии) и механизмы поиска. Большинство из сервисов могут быть обеспечены с помощью коммерческого механизма поиска.
- Управление сеансами пользователей и профили пользователей - а) сбор информации о том, кто пользуется сайтами УНС, и б) с тем чтобы пользователь мог избежать необходимости осуществления повторных процессов выбора и формирования запросов в ходе последующих посещений.
- Отображение статистики и в основе этого поиск по онлайн-источникам данных, таких, как реляционные базы данных.

7. Очевидны две вещи. Во-первых, независимо от того, каким образом технологии, а также конкретные производственные потребности будут меняться во времени, необходимо обеспечить реализацию вышеупомянутых решений. Во-вторых, благодаря этим взаимодействующим сервисам их "скрытая" техническая реализация будет иметь меньшее значение, чем обеспечение *интерфейса* между ними.

8. XML является привлекательным инструментом обмена информацией между системами, комбинирующим широкие возможности отображения с простотой текстуальных данных. Этот язык также является основополагающим инструментом представления данных для последнего поколения технологий интеграции - вебсервисов (комбинирующих SOAP, XML и UDDI). УНС сделало вывод о том, что XML является надлежащим языком интерфейса между взаимодействующими сервисами.

9. УНС уже определило технологию корпоративной системы управления контентом (СУК) и механизма поиска на основе коммерческих пакетов. Требовалось найти новое решение для представления данных в Сети (рендеринга данных). Проектная группа сделала вывод о необходимости разработки вышеупомянутого решения на основе XML. По воле случая корпоративная СУК также в значительной степени опирается на XML.

V. ПРИЧИНЫ, ЛЕЖАЩИЕ В ОСНОВЕ РАЗРАБОТКИ ВЫШЕУПОМЯНУТОГО РЕШЕНИЯ

10. Эти причины можно кратко изложить следующим образом:

- Группа не нашла коммерческого пакета, который бы удовлетворял всему набору требований NeSS. Полный охват и глубина данных и географический охват не поддерживались в соответствии с требованиями такого рода вебсайта;
- требовалось найти гибкое решение, которое могло бы постоянно дополняться новыми функциями;
- некоторые наиболее вероятные кандидаты из числа коммерческих пакетов не поддерживались в стратегической технической среде, определенной УНС;
- данный компонент должен был беспрепятственно встраиваться в принимающий сайт, а не требовать корректировки конфигурации самого сайта для встраивания этого компонента.

VI. СЕРВИС РЕНДЕРИНГА ДАННЫХ НА ОСНОВЕ XML

11. Что касается вышеупомянутых требований к представлению данных, то каждый стиль выходного материала может рассматриваться в качестве *преобразования* данных. Это является главным аргументом в пользу выбора XML в качестве ключевой технологии для разработки решения рендеринга данных:

- Результаты взаимодействия со статистической базой данных могут подвергаться сортировке в формате XML.
- Представление этих результатов может быть обеспечено путем преобразования формата XML через XSLT.

12. Разработка искомым материалов в формате HTML и CSV является относительно простой, поскольку оба эти формата опираются на текст. Графики и карты должны разрабатываться из источника XML. Что касается графиков, то для этого использовался коммерческий пакет фирмы Corda под названием PopChart³, который может

³ <http://www.corda.com/products/popchart>.

преобразовывать данные XML в различные графические форматы, включая JPEG и SVG. В предыдущих версиях NeSS уже использовался онлайн-инструмент географического картирования из ESRI под названием ArcIMS⁴. В нем также в качестве базовой технологии используется XML. SVG является еще одной опирающейся на XML технологией. Проектная группа предложила создавать тематические карты в формате SVG путем "объединения" материалов XML из статистической базы данных с материалами XML из ArcIMS.

13. Результатом стала проектная концепция сервиса рендеринга данных и использование XML для его разработки. Это затем послужило основой для разработки компонента для NeSS с дополнительным преимуществом, заключающимся в возможности повторного использования данной конфигурации для других проектов.

VII. СЛУЧАИ, КОГДА XML НЕ ИСПОЛЬЗУЕТСЯ

14. XML не подходит для каждого случая, когда требуется осуществить преобразование статистических данных. XML по своей природе является громоздким, содержит много повторяющейся информации и требует большого объема памяти. К счастью, в случае проекта NeSS пользователь является единственным, кто просматривает малый поднабор из всей совокупности имеющихся данных, в связи с чем преобразования являются быстрыми и использование памяти остается в допустимых пределах. Однако использование XML для передачи крупных наборов статистических данных из одной системы в другую вряд ли целесообразно.

VIII. ВЕДЕНИЕ И ПОВТОРНОЕ ИСПОЛЬЗОВАНИЕ, ОБЩАЯ СХЕМА XML

15. Благодаря использованию вышеупомянутой концепции "сервиса" интерфейс (API) удалось свести к максимально простому компоненту, с тем чтобы его можно было использовать на раннем этапе разработки вебсайта. В свою очередь это позволило группе продолжить совершенствование "скрытого" компонента рендеринга. Это наряду с неизменчивостью состояния компонента⁵ делает данный сервис более подходящим для внешнего размещения в качестве вебсервиса в будущем, если возникнет такая необходимость.

⁴ <http://www.esri.com/software/arcgis/arcims/index.html>.

⁵ Цель заключается в том, чтобы компонент не зависел от любых объектов, используемых вебсайтом, запрашивающим его, и не оказывал влияния на информацию, касающуюся запроса к памяти после предоставления ответа. Не меняющие своего состояния вебсервисы являются более масштабируемыми по сравнению с вебсервисами, меняющими свое состояние.

16. Одной из важнейших проблем являлась необходимость обеспечить удовлетворение требований NeSS наряду с обеспечением возможностей будущего повторного использования в других проектах. Данный компонент до сих пор еще не использовался вне проекта NeSS. Однако одним из ключевых проектных элементов данного решения является эволюция общей схемы XML, что теоретически способно обеспечить поддержку любого поднабора многомерных статистических данных.

17. Данная общая схема XML служит основой для любого будущего повторного использования компонента. Опыт многолетней работы свидетельствует о существовании многочисленных различных подходов к хранению статистических данных в реляционной базе данных, выбор которых зависит от уникальных требований каждого проекта. Однако возможности повторного использования нового сервиса рендеринга данных предстоит еще обеспечить путем разработки конкретных адаптеров для баз данных, которые будут способны картировать индивидуальные отображения данных в XML в рамках искомой схемы.

18. Такой же подход используется и в отношении других аспектов проектирования, поскольку мы целенаправленно позволили проектируемой системе поддерживать конкретные конфигурации и расширения (в некоторых случаях путем внесения изменений в простые, опирающиеся на XML конфигурационные файлы).

IX. ВЫВОДЫ

19. В заключение в полной версии документа подробно рассматривается схема компонента рендеринга данных, а также вопросы, которые нам предстоит решить в ходе его разработки.
