Distr.
GENERAL

CES/AC.71/2005/23 (Summary)
3 February 2005

Original: ENGLISH

UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE (ECE) CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES (EUROSTAT)

ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT (OECD) STATISTICS DIRECTORATE

Joint ECE/Eurostat/OECD Meeting on the Management of Statistical Information Systems (MSIS)
(Bratislava, Slovakia, 18-20 April 2005)

Topic (iii): XML and web services

ON-LINE STATISTICAL DATA PRESENTATION USING XML

Supporting Paper

Submitted by the Office for National Statistics, United Kingdom[1]

Summary

I. INTRODUCTION

1. This is an overview for a larger paper on how the United Kingdom's Office for National Statistics (ONS) "Neighbourhood Statistics" project has made extensive use of XML in the software used to present statistical data on the web, and how the design of the software will allow for potential reuse as a web-service.

II. NEIGHBOURHOOD STATISTICS WEB PROJECT: BACKGROUND AND SCOPE

2. The Neighbourhood Statistics (NeSS) programme was initiated in 2000, and the core aim is to provide better information for the government's neighbourhood renewal objectives.

3. NeSS will deliver a public website which allows the user to interrogate statistical data collected from a wide range of UK government sources, including the ONS's own decennial Census. The data are available at the full range of geographic detail – from a small number of households through to regional and national levels.

4. There have been three versions of the website, progressively richer in features and data, delivered since 2001. The fourth is due for release in the spring of 2005. This latter delivery has been re-written from scratch using Java J2EE and Oracle technologies, and most comprehensively addresses the requirements of the NeSS programme.

III. DATA PRESENTATION REQUIREMENTS

5. The list of NeSS requirements for presentation of statistical data can be summarised as: tables – i.e. HTML; charts – bar and line charts; thematic maps (using Scalable Vector Graphics – SVG[2]); dynamically-created spreadsheet files (in CSV format).

---

[1] Prepared by Simon Field and Rob Hawkins.
[2] http://www.w3.org/TR/SVG/index.html , http://www.adobe.com/svg

IV. CONSIDERING (STATISTICAL) WEBSITES AS A SET OF CO-OPERATING SERVICES

6. Websites can be considered to be a combination of cooperating *services*. That is, services in a broad sense rather than Web Services specifically, though that technology may provide part of the eventual solution. At a high-level, most statistical web offerings from ONS can be thought as consisting of:

- General navigation and content, including the publishing lifecycle of that content. This can be handled by a commercial Content Management System (CMS).
- More advanced navigation (e.g. statistical taxonomies), and search facilities. Most of this can be handled by a commercial Search Engine.
- User session management and user profiles – a) building a picture of who uses the ONS sites, and b) so the user can avoid having to repeat selection and query processes on subsequent visits.
- The display of statistics, and behind this, the interrogation of on-line data sources such as relational databases.

7. Two things are clear. Firstly, no matter how technologies may change over time, and specific business requirements may fluctuate, solutions to the above are needed. Secondly, with these co-operating services, the technical implementation behind-the-scenes is less important than the *interface* between them.

8. XML is an attractive means of exchanging information between systems, combining richness of expression with the simplicity of textual data. It is also the underlying data representation for the latest generation of integration technologies, Web Services (combining SOAP, XML and UDDI). ONS concluded that XML was an appropriate language for the interface between co-operating services.

9. ONS had already selected a corporate Content Management System (CMS) and Search Engine technology, based on commercial packages. What was required was a new solution for web presentation of data (Data Rendering), and the project team concluded that a bespoke solution based on XML must be developed. Incidentally, the corporate CMS also makes heavy use of XML.

V. WHY DEVELOP A BESPOKE SOLUTION?

10. In brief, this was because:

- The team did not encounter a commercial package that covered the full range of NeSS requirements;
- The sheer range and depth of data and geographic coverage was not supported in a way suitable for the kind of website required;
- A flexible solution was required that could be continually enhanced with new features;
- Some of the better candidate commercial solutions were not supported in the strategic technical environment chosen by ONS;
- The component should fit within the container website seamlessly, rather than having to adjust the design of the website to accommodate the component.

VI. XML-BASED DATA RENDERING SERVICE

11. Referring to the "Data Presentation requirements" above, each output style can be thought of as a *transformation* of the data. This is the crux of why XML was selected as the key technology in the Data Rendering solution:

- The results of the interaction with the statistical database can be retrieved in XML format.
- The presentation of those results can be achieved by transformation of that XML, via XSLT.

12. Production of the required HTML and CSV outputs are straightforward, since both are text-based. Chart and map outputs have to be derived from the source XML. For charts, a commercial package from Corda called PopChart[3] was used, which can turn XML data into a variety of graphic formats including JPEG and SVG. The previous versions of NeSS was already using an on-line geographic mapping tool from ESRI,

[3] http://www.corda.com/products/popchart

called ArcIMS[4]. It too uses XML as a core technology. SVG is another XML-based technology. The project team proposed to create SVG-format thematic maps by 'fusing' together the XML from the statistical database and the XML from ArcIMS.

13. The result was the design concept of a Data Rendering "service", and the use of XML to drive it. This then was the basis of the development of the component for NeSS, with the additional remit of making the design re-usable for other projects.

VII. WHEN NOT TO USE XML

14. XML is not appropriate for every case where transformation of statistical data is needed. XML is by its nature bulky, contains a lot of repeated information, and is memory-hungry. Fortunately, for the NeSS project the user is only ever looking at a small subset of the total data available, so the transformations are quick and the memory usage is kept manageable. However, using XML to transfer large statistical datasets from one system to another would not be recommended.

VIII. MAINTENANCE AND RE-USABILITY, A GENERIC XML SCHEMA

15. By using the "service" concept above, it was possible to keep the interface (API) to the component as simple as possible, so that it could be used early in the development of the website. In turn this allowed the team to continue to enhance the rendering component "behind the scenes". This, together with keeping the component stateless[5], makes the service well-suited to be exposed externally as a Web Service in the future, if required.

16. One of the biggest issues has been the need to deliver the NeSS requirements while balancing that with future re-use on other projects. As yet the component has not been used outside of the NeSS project. However, one of the core design elements in the solution has been the evolution of a generic XML schema that, in theory, supports any sub-set of multi-dimensional statistical data.

17. This generic XML schema is the foundation of any future re-use of the component. Experience has shown over the years that there are many different solutions for how to store the statistical data in a relational database – there are unique aspects to each project's requirements for this. However, reuse of the new Data Rendering Service can still be achieved with the development of database specific adaptors that can map individual data representations to XML in the required schema.

18. This is similar to other aspects of the design – we have deliberately allowed the design cope with specific configurations and extensions (in some cases done by changes to simple XML-based configuration files).

IX. CONCLUSION

14. In conclusion, the full paper will examine in further detail the design of the Data Rendering component, and the issues we have had to overcome in its development.

---

[4] http://www.esri.com/software/arcgis/arcims/index.html

[5] The goal being that component is not dependent on any objects used by the website that is calling it, and does not hold onto information relating to the request in memory after it has produced the response. Stateless web services are more scalable than stateful web services.