

Distr.
GENERAL

CES/AC.71/2003/5
10 February 2003

ENGLISH ONLY

**UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE
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
(Geneva, 17-19 February 2003)

Topic I: Measures for the improvement of quality at the IT management level

**WEB SERVICES – AN INTEGRAL ELEMENT OF OECD'S dot.STAT TECHNOLOGY
FRAMEWORK FOR STATISTICAL APPLICATIONS**

Invited paper

Submitted by OECD¹

Abstract: The combination of evolving ICT standards that increase the interoperability between applications and the ever-increasing need for a more seamless access to and exchange of information is a major driver in the OECD ICT strategy. This paper describes the role and use of web services in the context of recent developments at the OECD to improve the accessibility of statistical information. It follows up on a paper presented at the 2002 ISIS meeting about “Improving Access to Statistical Information at OECD in Response to Users’ Requirements” and describes a technology framework – called “dot.STAT” – that has been devised for the implementation of applications that enable easy access to certain OECD reference data. The concepts outlined in this paper have been developed in close collaboration with the Statistics Directorate, specialised OECD-internal groups and task forces², as well as members of the SDMX Consortium³.

Keywords: statistical information management, statistical data-/metadata repositories, data-warehouse, data dissemination, Registry services, web services, IT development frameworks.

I. INTRODUCTION

1. The dot.STAT development framework is designed to help implement the basic infrastructure for the new OECD Statistical Information System. It incorporates a central data warehouse and is part of a broader statistical strategy, which depends on a glossary of statistical terms and a corporate meta-data repository.

¹ Prepared by Peter Lübker (peter.lubkert@oecd.org) and Cengiz Tarhan (cengiz.tarhan@oecd.org).

² The OECD Statistical Policy Group and the Analytical Statistical Task Force (ASTF).

³ SDMX – Statistical Data and Meta Data eXchange; SDMX sponsors include the BIS, ECB, Eurostat, IMF, OECD, and the UN – www.sdmx.org.

2. The objectives of the dot.STAT framework are threefold. The first aim is to offer a single platform/repository for multiple data/meta-data sets with different characteristics. Secondly it comprises a complete set of information broker tools to make replication from working/production databases as transparent and simple as possible. Thirdly, the architecture aims to offer basic building blocks for the (re)development of statistical production systems as part of an OECD-wide Quality Framework initiative.

3. The application of the “web services” paradigm has its roots in the combination of three factors. First of all, both internal and external users of OECD statistical information typically have their preferred (or standard) data manipulation tools. Secondly, the development of a user interface, which meets “all” potential requirements is virtually impossible and would likely be very costly. And thirdly, there are emerging standards, in particular XML, SOAP, and WSDL which, in combination with already existing standards (HTTP), have great potential for improving the access to and interchanging of data in the Internet space. Another key aspect in the development of this approach is the OECD participation in the SDMX Case Study project⁴, which strives to demonstrate the concept of Registry based, networked (virtual) databases.

4. From the more general business perspective this approach is expected to increase the coherence of ICT solutions for statistical applications in line with recommendation of the OECD Statistics Quality Framework. It should also allow further streamlining application support and maintenance activities.

II. BACKGROUND

5. The OECD has been providing online access to selected data collections via the Internet since the mid-90s. Related developments dovetailed to a large extent with the migration of production systems from a mainframe-dominated environment to client-server architectures with browser-based interfaces for internal and external use. In addition, efforts were undertaken to develop statistical publications with easy-to-use interfaces providing yet a complete range of functionality (including data search, selection, navigation, and export). Though these developments helped to significantly increase the accessibility of OECD statistical information they also illustrated that there is no “one-fits-all” solution in this field. The design of statistical systems is typically driven by production needs and often results in very effective applications that, however, do not really reflect end-users or “consumer” requirements. In this context and taking into account past experience gathered with Internet based statistical applications led to reviewing alternatives to classical web interface developments.

6. The web service based approach appeared to be particularly promising for the reasons mentioned at the outset, but also because of the speed with which the development work could be carried out for both, internal and external usage. It also seemed very compatible with the rest of the dot.STAT development framework, as it allows developing functional components that can be reused in different applications. Another important aspect is the openness of the solution from a technical perspective due to the availability of standards that seem to be rather widely adhered to by the ICT market, thus allowing some degree of vendor independence.

III. CURRENT STATUS

7. The project that spawned the establishment of the dot.STAT development framework – i.e. making OECD Reference Series accessible – has progressed well. A structured design and development environment using dot.NET technology has been put into place. Specific focus has been put on ensuring increased development and maintenance efficiency. In particular, a tool allowing the establishment of unit and functional tests at development time has been integrated into the environment for this purpose. The design and installation of a central data-warehouse based on SQL/OLAP technology has been completed. Last, but not least XML based data loading routines are available for different types of production systems (SQL, Fame, ORACLE Express).

⁴One of the projects sponsored by the SDMX Consortium and lead by the BIS

8. On the end user application side, a first simple web service has been developed under VB and C# and prototyped with Excel, Access, and VB front-ends. First user/client feedback has been very encouraging particularly with regard to an Excel based interactive wizard which offers an environment with which many users are familiar. This development was also used as proof of concept for integrating a web service into the IMF Intranet portal. The same web service example is also being assessed as a potential delivery mechanism in the context of the SDMX initiative. The rollout of a web service/Excel based system for access to reference data and metadata is currently taking place within the OECD Secretariat.

IV. THE CONCEPT

9. Figure 1 illustrates the information flow in a web service scenario. The Universal Discovery and Description Integration service allows finding and locating a service following the concept of yellow pages (see Figure 2). A service provider will publish pertinent information about the content and structure of the service and offer information about the way how to access it.

Figure 1 Web Service Flow Diagram

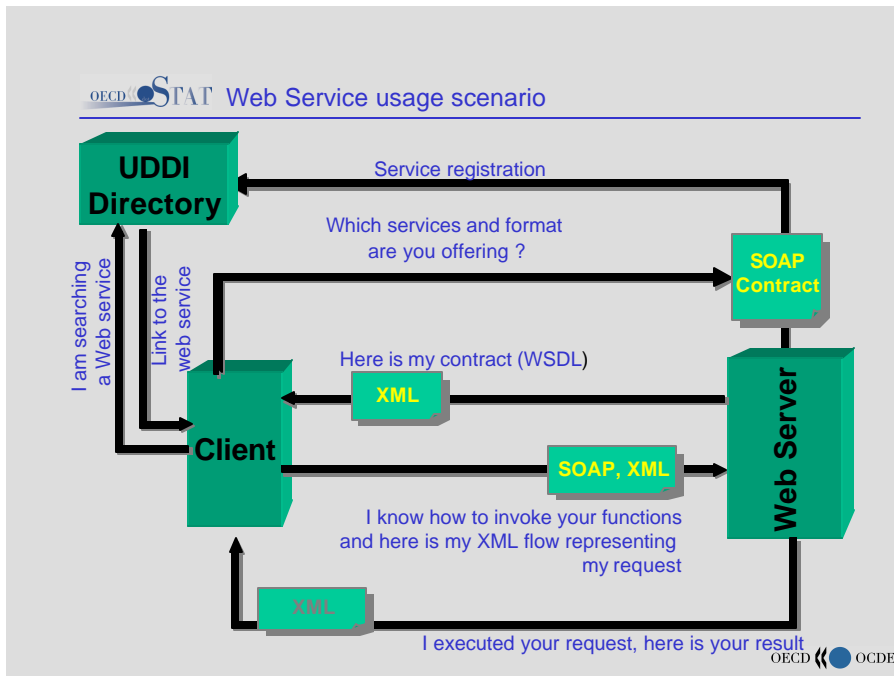
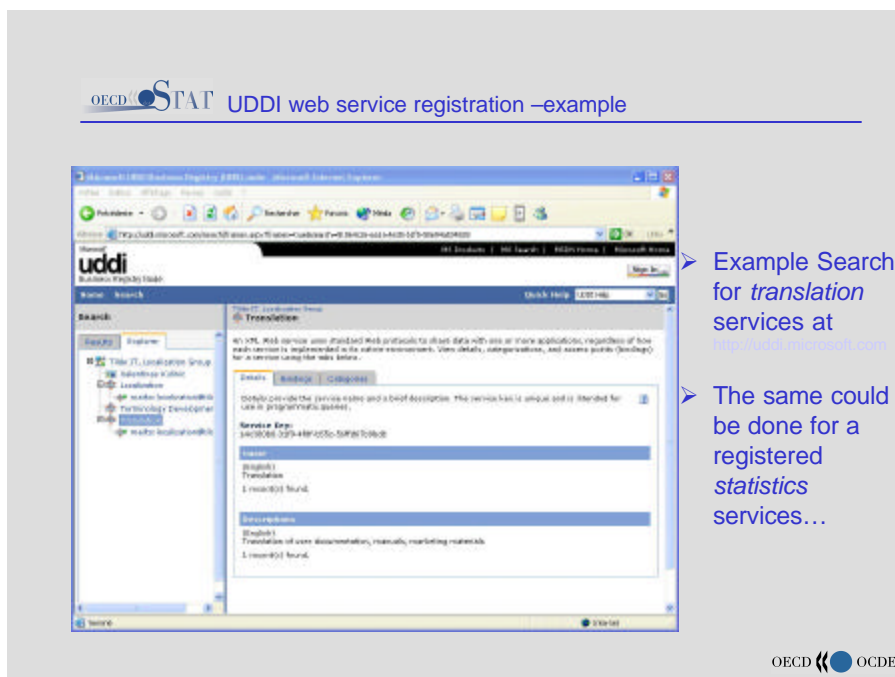


Figure 2 UDDI Registration



10. This information could be complemented by sample code or even re-usable applications. For more information about UDDI (or UDDI Business Registry) please refer to one of the four existing UDDI servers - IBM, Microsoft, SAP and NTT – which, with version 2 of the standard, also exchange information amongst each other. Thus a Web Service named “dotStat” under a Business named “OECD” can be found at:

- Microsoft: <http://uddi.microsoft.com/search>
- IBM: <https://uddi.ibm.com/ubr/find>
- SAP: https://websmp205.sap-ag.de/~form/uddi_discover/prod
- NTT: <https://www.uddi.ne.jp/ubr-e/find>

11. Once registered, a service can be located by the “customer/client” who can connect to it directly, based on the information provided about content and usage. The application of a number of standards established through the W3C⁵ (see list below) distinguishes the web service approach from traditional application programming interfaces and allows the seamless integration of the web service into existing information processing environments. All information flows between the client and the server will use XML. On the server side, the web service will communicate with the relevant data source. In the case of dot.STAT the source data warehouse is a MS SQL-Server database with OLAP extensions, but could be another type of data store (see Figure 3).

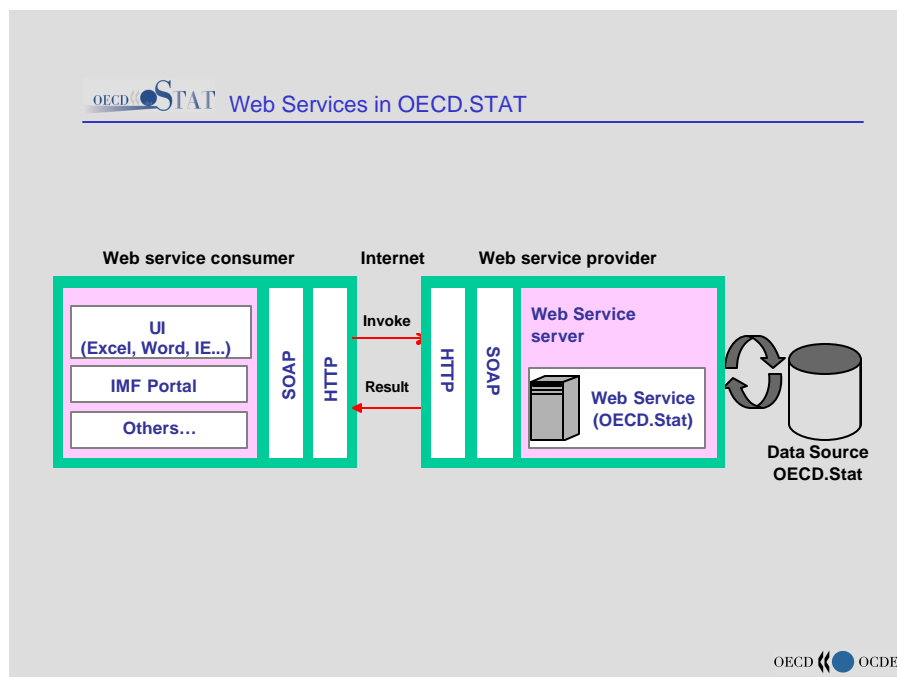
12. Web Services are based on (or utilise) a number of standards, including:

- HTTP (HyperText Transfer Protocol)
- XML (eXtensible Markup Language)
- SOAP (Simple Object Access Protocol)
- UDDI (Universal Description Discovery Integration)

⁵W3C World Wide Web Consortium - <http://www.w3.org/>

- WSDL (Web Services Description Language)
- Others
 - GXA (Global XML Web Services Architecture) announced for Q1/2003
 - WS-S (Security)
 - WS-I (Interoperability)
 - WS-T (Transaction)
 - Document attachments

Figure 3 Web Services Architecture Overview

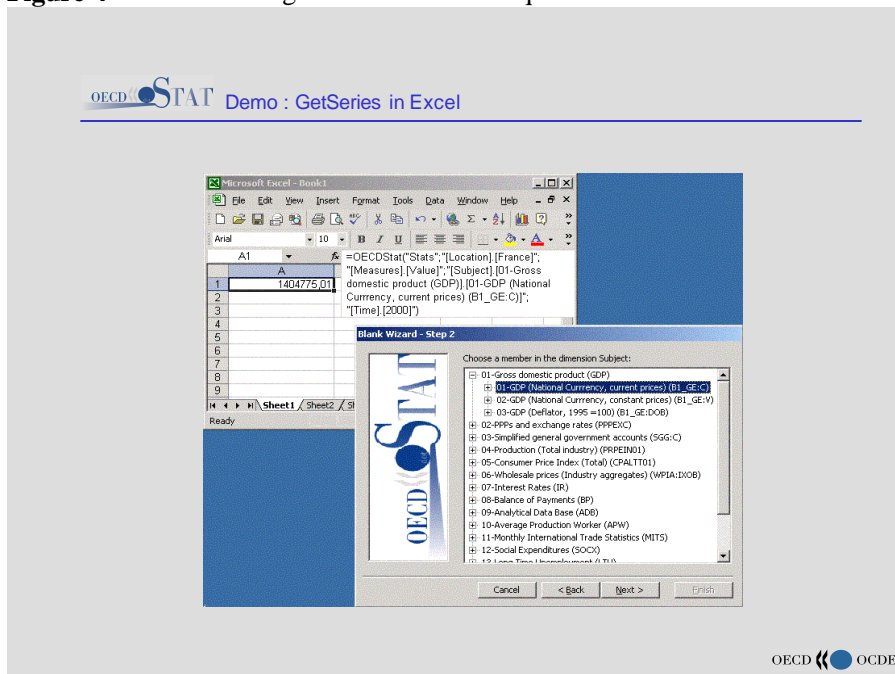


13. The consumer has the choice of exploiting the service interactively or programmatically. Office productivity tools such as Excel can be used for this, as can portal software tools or more classical application programming languages. This quite powerful concept is also illustrated in Figures 3 and was demonstrated through a successful integration test with an IMF⁶ portal application. The service provider further has the possibility to offer shrink-wrapped “applications” that work with commonly used tools. The use of a “wizard” (see Figure 4) to allow the creation of dynamic or static links from a spreadsheet into the web service is a good example of this kind of application. Experience with OECD internal clients and “wizards” has been very positive⁷. In addition to the intuitiveness of the approach in itself, it allows the client to manipulate/adapt the formulas created by the wizard directly later on and apply them in combination with the standard functionality of the spreadsheet product used. A spreadsheet based “briefing book” is another way to offer a web service to clients who require a set of regularly updated sets of information possibly in a graphical format for analysis and reuse.

⁶ The International Monetary Fund has a portal development project using Plumtree© software products

⁷ Example: A wizards to access FAME© from a spreadsheet.

Figure 4 Wizard used to generate web service queries



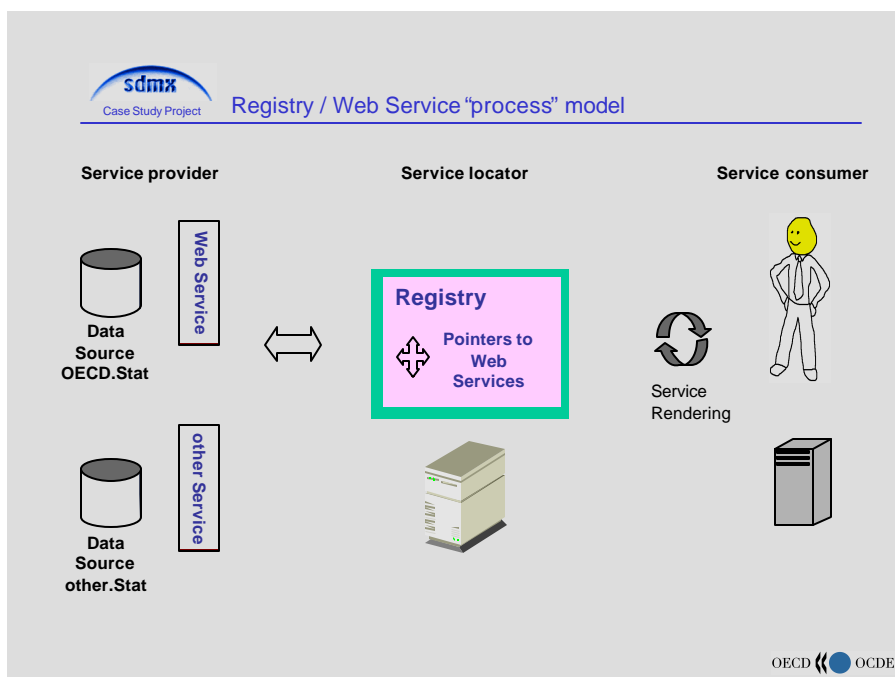
V. THE VISION

14. The main drivers for adopting the dot.STAT web services approach are outlined in the introduction. In terms of target audiences, the system was primarily devised for use with well-defined constituencies within the OECD or externally. Whereas the adoption and management of the web services concept internally is expected to be rather simple, there are both opportunities and challenges with external clients. By default, web services are not geared for general interactive use on the Internet. At the same time they offer the possibility to establish powerful applications, which eventually could change the common practice in statistical data exchange. Ultimately it could be envisaged to offer distributed web services that allow to significantly reducing the need for replicating data between different locations. This, in turn, could lighten the overall burden of reporting and decrease the multiplication of efforts in making information available more seamlessly.

15. An essential element in this distributed (or virtual) database approach is the "Registry" paradigm. Following on from the above discussion of UDDI, it could be envisaged to substitute the UDDI Directory by a Registry that does not only contain information about a service, but would actually integrate more closely with the actual service. The Registry would contain information about services and mechanisms to connect through to the actual information service. Figure 5 shows the role of the Registry as a central hub. It is an abstract illustration of the process model behind a combined Registry/web services approach. The rendering of content can be an interactive or programmatic process as with the provision of UDDI based web services. For the programmatic access it might be appropriate to offer the possibility of channelling calls directly to the web service provider, bypassing the Registry once the initial contact has been established. This, in turn, would require the management of certain metadata in both the Registry and the web service.

16. The integration of web services with a Registry, which is an essential part of the SDMX initiative, heavily relies on standard setting in both the statistical and the ICT domains. It also opens up a series of questions about access and information security, performance and bandwidth, as well as feature evolution. These are, however, not issues which are specific to the Registry concept, as they also exist with UDDI based web services. Another challenge lies in the necessity to harmonise nomenclature and terminology, particularly if a web service is made available to anonymous clients.

Figure 5 Registry - a central hub in a distributed database environment



17. As part of the SDMX initiative, current efforts in the OECD focus on a feasibility study to integrate an existing web service with a Registry prototype. The project is not only meant to address some of the data exchange and transmission issues outlined above, but is also supposed to demonstrate the feasibility of mixing and matching applications developed under different environments⁸ using currently available standards.

VI. CONCLUSIONS

18. The concept of web services is not new per se, but offers interesting opportunities in modern applications development due to emerging standards in the Internet space. These standards, assuming that they will continue to evolve, can significantly alter the way of exchanging information, at least amongst peer organisations. The web services paradigm could help materialise ideas about more effective and less resource intensive information exchange. Major advantages of this approach/concept include:

- Less information to be replicated and reduced transmission efforts
- Opportunities to reporting burden and duplication of data repositories
- Possible convergence of data dissemination and data exchange efforts

19. The focus of the Registry/web services idea is on the process by which well defined/structured content is made available. The dot.STAT development framework attempts to embrace this model for both access to information repositories internally and in the context of advanced international data exchange initiatives in which the OECD is involved.

⁸e.g., a web service based on proprietary technology with an Open Source Software based Registry

Acknowledgements

This paper was elaborated with a lot of invaluable input from colleagues within the OECD on an ongoing basis. A special word of thanks goes to Stéphane Varin for his contribution in the establishment of the dot.STAT development framework. Particular thanks also go to Stuart Feder from the BIS, as well as Christian Lallemand and Didier Simon from MCS for numerous extremely fruitful and stimulating strategic discussions and practical expertise.
