Distr.
GENERAL

ECE/CES/SEM.54/21
22 May 2006

ENGLISH ONLY

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

EUROPEAN COMMISSION STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES (EUROSTAT)

ORGANIZATION FOR ECONOMIC COOPERATION AND DEVELOPMENT (OECD) STATISTICS DIRECTORATE

Joint UNECE/Eurostat/OECD Seminar on the Management of Statistical Information Systems (MSIS)
Sofia, Bulgaria, 21-23 June 2006

Topic (iii): Dissemination and client relations

TOWARDS A SERVICE-ORIENTATED ARCHITECTURE (SOA) FOR THE STATISTICAL VALUE CHAIN (SVC)

Supporting Paper prepared by Vic Duoba, Office for National Statistics, United Kingdom

I. INTRODUCTION

1. The concept of generalised statistical processing systems has been implemented by a number of National Statistical Institutes (NSIs) over the past 20 years. It is ironic to some of us who have been observing this generalisation "industry" that generalised systems never seem to be so "general" that they are adopted by multiple agencies, which would be a wonderful proof of generalisability. However, it is a mitigating fact that the individual peculiarities of NSIs are partly explained by national legislative requirements and also by the fact that selecting a consistent mix of technologies along the entire SVC has not yet been formulated as a "hard" optimisation problem. Therefore individual NSIs make methodology choices based upon what they deem best for their local, unique circumstances.

2. Despite the above rationale for explaining the current lack of truly internationally common systems, is it possible that the current ensemble of Object-Oriented software development tools (for the masses), e.g. Java, plus simple, universal and powerful specifications such as XML (plusXSD, XSLT, etc.), SOAP, HTTP, UDDI, WSDL plus the Web are sufficient to implement generalisability in a useful way? It might be possible now or within two years to build robust "backbones" for generalised systems that permit interchangeable software limbs to be attached with relative safety and ease. The backbone would reflect a Web Services-oriented applications architecture and the limbs would be statistical functions as described by a common SVC and implemented as Web Services.

3. The UK Office for National Statistics has been conducting over the past 3 years a comprehensive modernisation programme and has put to the test a number of ambitious ideas, including designing metadata-driven processing systems and highly modularised, extensible and reusable components that support its view of realising in software a statistical value chain. (Note: ONS's modernisation programme encompasses much more than assembling systems of statistical processing software, and includes geographic re-location, managing down-sizing, outsourcing many services, expanding HR self-service, standardising methodology and, more recently, fleshing out a government-mandated concept of statistical independence.) This paper will discuss some of the outcomes and lessons for the future from the survey data management perspective, as well as the more modest ambitions for utilising Web Services "some time" in the future.

II. WEB SERVICES (WS)

4. There is an excellent description of Web Services architecture to be found at http://www.w3.org/TR/ws-arch/. W3C is described at this web site by *"The World Wide Web Consortium (W3C) develops interoperable technologies (specifications, guidelines, software, and tools) to lead the Web to its full potential"*. The essential idea of a web service is very simple:

(a) A user wants a service (in our case often some statistical functionality)
(b) The user locates some potential providers on the WWW by some initial discovery process (web crawling or more formally via special registries)
(c) Having identified some likely providers, the user can find out technical details about the services (e.g. via WSDL)
(d) The consumer can investigate how his/her own system can be made to comply with the connection, invocation and messaging semantics of the best web service (perhaps one or more mediating and anciliary web services will be required)
(e) The user, after obtaining connection permissions from the web service provider(s), utilises the web service(s).

5. An essential feature of this architecture is that it must not matter what are the user and provider platforms (hardware, operating systems applications languages and databases). Of course, the point of interaction of disparate systems must be something that is in common to each, and the evolutionary forces that operate within the computing industry indicate XML (for message encoding), SOAP (for applications invocation and applications (inter)operability) and HTTP over TCP (for transporting messages).

6. But all this sounds fine for someone wanting to web-enable their store for on-line shopping, or to provide electronic survey returns to an NSI, but is it relevant on a big scale to providing scalable and efficient data processing solutions that are needed by a typical NSI?

7. My answer to this question is not surprisingly "Yes – with reservations". There is no reason technically why a WS cannot be implemented on an NSI's intranet so that data transfer volume is not a dominating issue. Indeed, ONS's mixture of legacy platforms that require modernisation over the next 5-7 years, points to the option of a WS architecture as being the source of highly relevant standards and guidance. It would facilitate first assimilation then

replacement of legacy components. However, it is the enforcement of a set of standards for achieving interoperability and replaceability between and within bundles of naturally disparate tools (which could be legacy, gathered from many vendors or from donor NSIs) that makes WS attractive for ONS. It seems sensible to follow the natural Darwinian forces driving the computer industry, rather than introspectively designing alternative technology mixes heavily based on older technology, e.g. CORBA and RPC.

III. STATISTICAL VALUE CHAIN (SVC)

8. One of the decisions that have to be made in implementing Web Services is what to encapsulate as a Web Service. From the commercial perspective the decision will be based roughly on what is a useful piece of business functionality that can be recognised by clients as being of value to their business. But in practice, this might mean a small credit-checking tool, or it could mean a small business accounting system with the option of another tax-accounting web-service add-in. For NSIs it could be an electronic survey return option, a confidentialising tool, or it could be an entire editing and imputation system with individual methodology options each expressed as a WS.

9. A starting point for ONS is the Statistical Value Chain, which I have edited down in size to the following table:

1	Stakeholder consultation and prioritising
2	Methodological studies
3	Administrative sources assessment
4	Sampling design (technical)
5	Sampling design (implementation)
6	Data collection (field, electronic, mail)
7	Editing, validation, coding, derivations
8	Weighting and estimation
9	Simple analyses and quality assessments
10	Index number construction
11	Time-series analyses
12	Other advanced analyses
13	Confidentiality assurance
14	Information and data dissemination
15	Data management

10. The details in the above table can be debated and some of the statements are “soft” and some are “hard”, e.g., “Stakeholder consultation” as compared to “Time-series analyses”. However, the activities encapsulated in the SVC table above are what NSIs do repeatedly. ONS has been, for the statistical technical items in the table, collecting and testing software from a number of NSIs over the past few years. Even for some of the “softer” topics, it has been investigating CRM software and other tools that support systematised approaches to work-planning and prioritisation.

11. It is not impossible to think of all of the 15 items in the table as being realised by some bundle of tools. I use the term bundle deliberately, as often individual tools are a poor fit with other tools in the same item heading. For example, in ONS we use SuperCROSS, GES and Tau Argus. Each of these tools can perform several statistical functions chosen from "tabulate, calculate sampling error, confidentialise". This surfeit of functionality per tool actually causes more decision-making problems than you can at first imagine. However, that aside, it is possible that NSIs could agree on a statistical value chain that could be at the level of discrete web services. A topic such as "Other Advanced Analyses" could be subdivided usefully into further analysis domains and specific WSs developed for each.

12. If such a categorisation of the SVC into a widely recognised set of WS were agreed upon and the services described in sufficient detail, it might even encourage commercial development of the components. At least it would encourage NSIs to develop the components using WS architecture standards and save beneficiary NSIs the considerable trouble of programming adaptive layers (which is what ONS is doing).

IV. ONS's STATISTICAL APPLICATIONS MODERNISATION EXPERIENCE

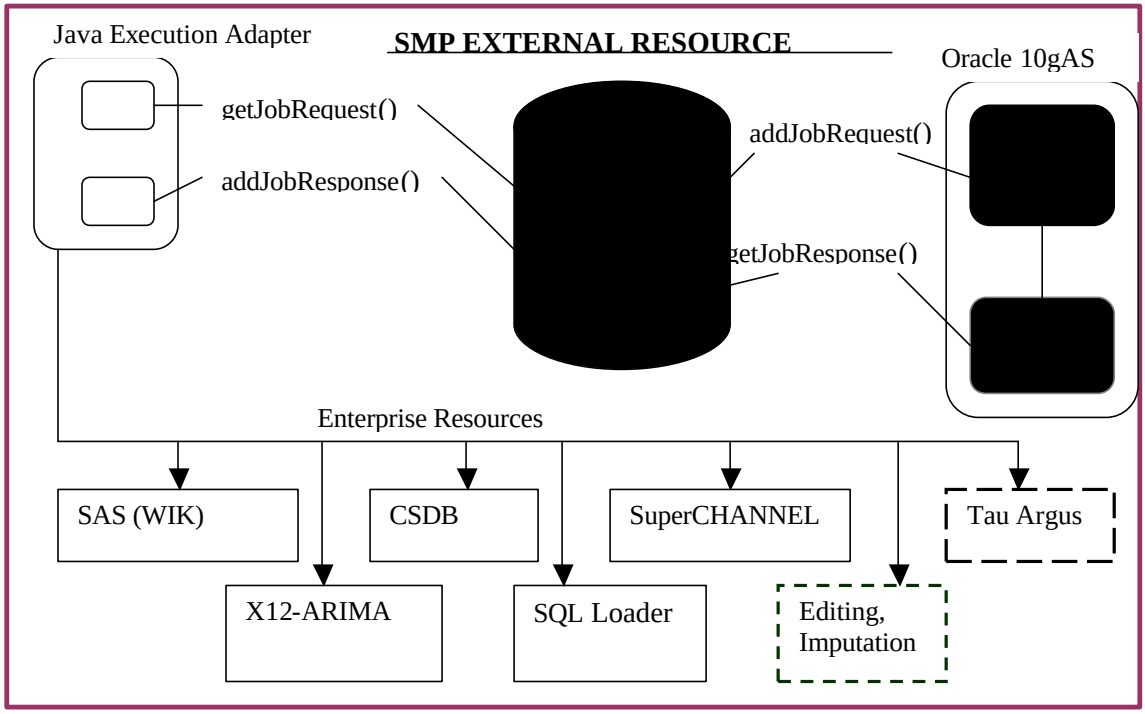
13. The ONS modernisation programme has been operational for about 3 years, and the initial vision was very bold. Metadata-driven processing on a modern platform and architecture would enable businesses to configure end-to-end processing systems against secured data and would also produce quality and performance measures. In other words ONS was to be given a luxury control and information console for statistical managers and a stress-free work environment for the common worker. This Polyanna-ish vision was subsequently dealt a rude blow by the reality of the time and cost taken to up-skill and re-equip personnel, to accurately analyse and agree upon logical and physical modernisation details, and by the costs of upgrading the technology base. In addition, ONS was faced with a downsizing and relocation requirement from government. In short, after two years of intensive effort, the modernisation programme had fallen short of its goals for greatly enhanced statistical capability from the business users' perspectives and funds were parlous. This required a major re-think of the mechanics of accomplishing modernisation and ultimately a long time horizon has been agreed.

14. However, the technology effort during those two years did not by any means represent a waste of resources, as there was much more to modernisation than writing new applications. ONS re-skilled a large number of applications developers, introduced an integrated applications development environment and associated methodologies, established an outsourced server farm with a SAN, and gained a great deal of practical experience in designing and implementing a very sophisticated ONS logical data model. Perhaps most importantly, management as a whole across ONS became fully educated in the modernisation concept and its implications. The fact that the modernisation vision became accepted and understood by ONS junior and senior management rather than being confined to a small band of enthusiasts was one of the important capability improvement outcomes of Phase I. In hindsight, SMP Phase I should have been called "Preparing ONS for Modernisation".

15. In 2005 the SMP was largely replaced by a strategy of trying to get proportionate returns to investment, so that the emphasis changed to delivering highly pragmatic outputs. Generality

and reuseability were component design features that have been reduced to a more subservient role due to cost pressures, but the bigger vision, of course, has had to remain in some form. It is invitable that when decisons on investment are made project-by-projects by different business managers that there is increasing certainty that the resultant whole will not be optimal. Hence, the Information Management Strategies group has been strengthened in order that the scalability and reuseability features of a modern architecture are not omitted from the future developments. Happily, with the help of ONS's Xansa partners, there has been established a generalised processing system architecture within which new business systems will be embedded. This architecture is not classical WS, but there is the capability of making it so.

16. The diagram below shows the execution stack architecture that ONS has created in order to supply a common execution style for its modernised architecture. The disparate resources are called using a collection of utilities in the Job Execution Adapter (comprising adapters for many individual tools). On the business side, the tools are not visible, just the business functionality from the user perspective and incorporated in a business service. Business service requests can be added to an Oracle queue that services the Job Execution Adapter. The latter breaks the request into calls for any combination of tools required by the job and handles any required synchronisation until the job is finished. To be honest, at the time of writing the paper this architecture is in the process of early testing using nascent business applications. It is expected that ONS will learn a lot from this exercise.



17. It can be seen from the above architecture that Web Services could be introduced at the Enterprise Resource level (but in a way that still preserves the above architecture if desirable) or at the Business Services level. Ideally, Web Services will pervade both areas. The corporate difficulty at the moment is that what is technically feasible and interesting is not immediately obvious as being necessary or desirable for any individual application. The situation could have been different if Web Serviced statistical components had been available from the start of

modernisation, but right now there is a reluctance by businesses to bear the costs of extensive generalisation that will, on the surface, benefit late-comers to applications modernisation.

V. THE WEB SERVICES CHALLENGE FOR ONS

18. I have borrowed the following list of best practices from the book "Service-Oriented Architecture – A Field Guide to Integrating XML and Web Services" (Thomas Erl, 5th printing, ISBN 0-13-142898-5, Prentice Hall). I thought it would be interesting to see how ONS would be evaluated against the list:

1. **Know when to use Web** *services (customise and phase the process)*
   ONS is complying with this advice to a large extent, and is not now going for broke but is committed to doing much more than just re-writing code
2. **Know how to use Web** services (limit to the scope of knowledge)
   The task of re-skilling personnel and adopting a "safe" applications development environment has been very expensive and prolonged. With private sector partners ONS has bought in the missing expertise and experience but is still on a steep learning curve.
3. **Know when to avoid Web services** (keep an eye on adding value)
   Because the WS technology is evolving rapidly, ONS might claw back some modernisation time by it current slow and steady strategy, especially if other NSIs also decide that WS architecture will begin to dictate some in-house applications development strategies.
4. **Moving forward with a transition** *architecture (introduce WS concepts without the technology)*
   This is what ONS is actually doing, partly forced by cost considerations, but it was a strategy mentioned by a consultant at the beginning of modernisation. But Java, the XML family and the Web are being increasing utilised throughout ONS's systems, so we are catching the WS wave.
5. **Leverage the** *legacy (build on what you have and consider re-use)*
   This is what ONS will be forced to do simply because of the size of its legacy estate. However, various wrapper and adapter mechanisms have been tested to embed the legacy code in a modern framework that will improve its interoperability and possibly enable some costly sub-components to be replaced in the shorter term.
6. **Sorry, no refunds (Web service and your bottom line***) (budget for the additional costs associated with WS)*
   It is apparent that creating a fully modernised environment is expensive, especially in regard to the very highly skilled personnel (analysts, designers, architects) that are required to build sophisticated interoperable systems so that security and scalability are realised. ONS underestimated the difficulty of large-scale Information Management modernisation and is now having to be realistic about timelines and ensuring that development monies are allocated wisely and sparingly.
7. **Align ROIs with migration** *strategies (ROI exercises can provide useful anciliary information)*
   Selecting modernising areas from many candidates has been a process that followed some value-for-money or other prioritising mechanism. But ROI to my knowledge was

not carried out to evaluate WS as an option – it was not on the "radar screen" when modernisation began.

8. **Build towards a future state** *(take into account a probable migration path)*
   By investing in an expanding IM Strategies group, ONS is trying to ensure that all systems being built or refreshed are evaluated for conformance to common standards which include emphasis on common subsystems (tools) and common usage patterns. This is partly driven by the desire to cut costs of supporting too high a diversity of technology alternatives that are functionally unjustifiable at the macro level.
9. **Incorporate** *standards (consider WS standards as infrastructure standards)*
   This is one of the possible objectives that the IM Strategies group will continue to evaluate and discuss with its technology advisors and with the funding businesses.
10. **Label the infrastructure** *(consider naming conventions)*
    This have been considered by ONS but not pursued rigorously. However, effectively the essence of the idea has been adopted and a special team within IM Strategies has been established in order to keep track of ONS's OM infrastructure and Strategies itself is a mechanism for formalising and stabilising the new infrastructure.
11. **Design against an interface (not vice versa***) (design service interfaces consistently)*.
    Happily, ONS is managing the design of application interfaces for new or wrapped legacy (or "foreign") systems very seriously, using private sector expertise. This should significantly ease the transition to a fully-fledged WS enterprise infrastructure at some stage in the future.
12. **Use SOA's to streamline business** *models (rethink business models)*
    This has proven more difficult than expected, because the SVC is a long, long chain that has many arcane links to it.

VI. SUMMARY

19. WS architecture is an option that was not clearly visible to ONS at the beginning of its modernisation programme. However, the technology that ONS has adopted for modernisation is highly suited to incorporating the WS paradigm. While large-scale modernisation for an individual NSI is a costly, arduous and prolonged exercise, WS does provide a model for a modern, componentised and interoperable set of statistical services, which if adopted by NSIs could lead to long-term cost reduction. The best-of-breed statistical functions could become shareable amongst NSIs, and this could be accomplished to a useful extent simply by adopting WS standards for future applications development or applications refurbishment – with or without formal cooperation.
