Proceedings of MSP2009
Modernisation of Statistics Production 2009

Uptake of Service Oriented Architecture in Statistical Agencies - Are We Really so Different? [1]

Matjaz Jug

While both SOA theory and implementation practice appear to be in a relatively mature stage it's uptake in statistical agencies still isn't very high. Many organisations are considering SOA approach, building web services interfaces or implementing early versions of Enterprise Service Bus however even early adopters still appear to stay within early levels of SOA maturity model. On the other side there are many benefits offered by implementation of SOA which are definitely very high on a preference lists of statistical information system's architects and designers. The paper will try to uncover the reasons for the slow uptake and to identify main challenges as well as drivers and areas where SOA could offer significant benefits to our organisations.

The implications of SOA implementation within organisations whose core business is production of statistics can definitely be compared to those implied by introduction of some other paradigms like data warehousing or metadata-driven approach. Paper will compare current SOA uptake issues with lessons learnt from the past introductions of some other paradigms and put SOA in the context of statistical business which could help us to understand which of current challenges originate in general SOA implementation issues so we can search for solutions in wider community of SOA practice but also which of them are specific to statistical business process or organisational/cultural issues so we will have to search for solutions within our organisations and statistical community.

Key words: Service Oriented Architecture; Data Warehousing; Metadata-driven Systems, Information Architecture.

1. The Story of Service Oriented Architecture (SOA)

The Open Group (2009) describes Service Oriented Architecture (SOA) as a "style of IT architecture that delivers agility and Boundaryless Information Flow™. It is deployed on an increasing scale in enterprises today." It is however not easy to assess the current level of SOA uptake in enterprises and even more difficult to measure realised benefits of implementation. Even though the SOA buzzword and corresponding philosophy, best-practice and tools have been around for a decade now, there are still many discussions about the benefits of the approach, maturity level of technology, and implementation costs. Enterprises in some industries (for

[1] The views expressed in this paper are those of the author and do not necessarily reflect the policies of Statistics New Zealand.

example travel industry) have adopted SOA in much larger extent than some other (for example government).

1.1. Definition and Benefits of SOA - Why Should Statistical Organisations Be Interested

According to Wikipedia, service-oriented architecture (SOA) provides a set of principles of governing concepts used during phases of systems development and integration. Such an architecture will package functionality as *interoperable services*: software modules provided as a service can be integrated or used by several organizations, even if their respective client systems are substantially different. Benefits organisations should expect from implementation and their potential reflection in statistical organisations are:

- Increased agility. Organisations should be able to more quickly respond to changes in business model and external environment. This benefit looks very attractive for Statistical organisation knowing the shift towards the increased use of administrative data and more automated data processing as well as more user-oriented services for data collection and dissemination,
- Reduction of development cost trough reuse: service-oriented architectures promote the reuse of existing services therefore new IT systems should be able to leverage the most readily available code and services from across organization and externally. One of the main challenges in statistical organisations is how to reduce the high cost of development (and maintenance!) of statistical production systems and how to apply new methods (for example automated editing and imputation methods) across many new and existing statistical processing systems.
- Better possibilities for integration between different systems/platforms including integration of legacy systems into the service-oriented environment. Interfaces used to implement SOA can enable better integration between different organisations. In statistical organisation this means new possibilities to automate data collection and dissemination and possibility to externalise statistical services to other organisations producing official statistics. With the better possibility to integrate old monolithic applications into the corporate technical environment it is possible to protect old investments in legacy systems and prolong their use. Statistical organisations usually have many legacy systems particularly due to the historical stove-pipe development approach.
- Configuration rather than programming could help us to more efficiently respond to volatile data sources and frequent changes in questionnaires, methodology and classifications.

Yet we can't exactly say the SOA uptake in statistical organisation has been high. Statistical community isn't large enough to be measured by market analysts so there is no data about the level of planned or realised investment in SOA. Without the attempt to undertake a survey, it is still possible to get an indicative picture reviewing documentation about IT architectures and SOA implementation projects presented internationally. But before that we should define the overall level of SOA uptake we would like to benchmark with.

1.2. Is SOA still in an early immature stage, in wide use or maybe dead?

There are many maturity models developed by different vendors and interest groups. Most of them are however similar to the Open Group Service Integration Maturity Model (OSIMM) - the industry's first collaborative maturity model for SOA adoption. There are 7 levels in OSIMM model with level 1 being the less take up (silo) and level 7 being the greatest take up (dynamically reconfigurable services). For the purpose of this paper we will use relatively simple Sonic Maturity Model (see picture 1) which represents maturity level in relation to the benefits offered by particular maturity stage. Sonic model has been also used in more comprehensive SOA maturity assessment methodology developed by Inaganti and Sriram which offers multi-point view that encompasses as many aspects of the organization's SOA implementation as possible, to arrive at its true state of SOA maturity (Inaganti & Sriram, 2007). The graphic representation of this model is in the appendix 1.

Writing the abstract for this paper nearly one year ago I thought it should be easy to get information about the SOA uptake in different sectors of industry, including government. Reports developed by analysts and IT gurus blogs however revealed it is quite difficult to understand the SOA's current uptake, level of maturity and benefit realisation. While some researchers back in 2007 still claimed SOA had been in immature stage (Mackie, 2007): "A report from Research 2.0 describes the current use of service-oriented architecture (SOA) as experimental. The research firm's report of May 31, 2007 predicts that SOA will be embraced as mainstream technology by the year 2015." the following year 2008 saw more optimistic forecasts (Gartner, 2008): "Gartner conducted a user survey in 2008 regarding service-oriented architecture and application development trends. Survey results reveal a strong uptake of SOA around the world." Optimism of course didn't last very long: the signifficant cut of IT investments coused by global financial crisis in 2009 had significant impact on many SOA implementation projects. In January 2009 Anne Thomas Manes made ironic announcement (Manes, 2009): "SOA met its demise on January 1, 2009, when it was wiped out by the catastrophic impact of the economic recession. SOA is survived by its offspring: mashups, BPM, SaaS, Cloud Computing, and all other architectural approaches that depend on "services"." Which was perhaps the ultimate proof that SOA isn't silver bullet for all our problems anymore. It is simply technology which can deliver *some* benefits if used appropriately. At the and of the day it isn't very important, what is the level of uptake across different sectors/industries – organisations should use it if it can deliver benefits justifying the investment.

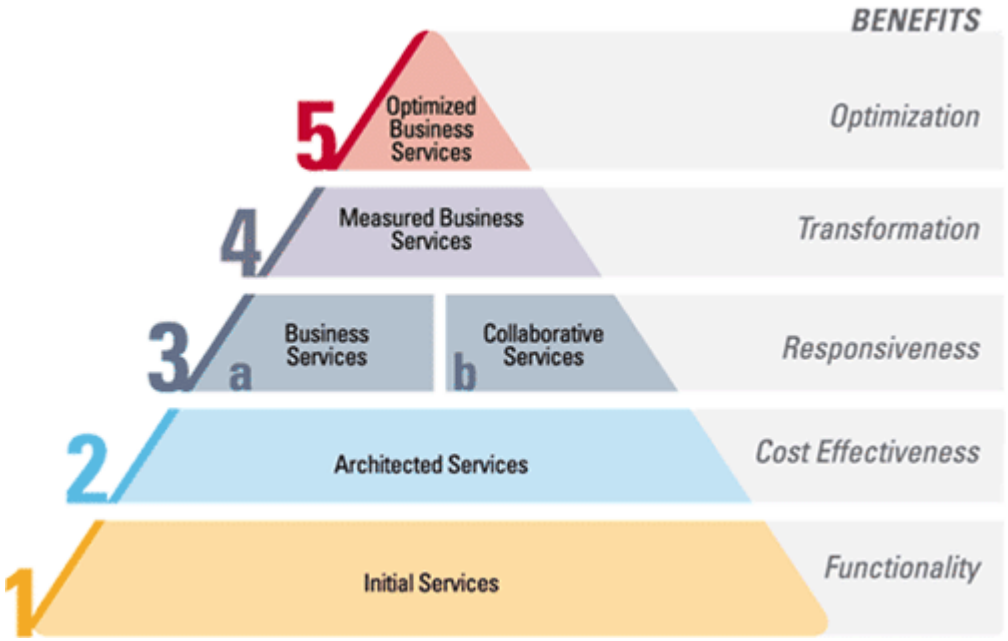

Figure 1: Service-oriented Architecture Maturity Model Levels with Key Business Impact

1.3. Some early implementation projects in Statistical Organisations

SOA implementation in some statistical organisations started back in 2004. There were at least three important projects at that time:

- Statistics Finland has presented their SOA framework, called MetaAPI (Mäkinen et al, 2004) which aimed to develop application programming interface built on top of the existing metadata systems.
- ONS has presented large redevelopment project, important part of which was SOA-based integration between many different statistical tools as well as data and metadata repositories.
- Statistics New Zealand has started implementation of programme called Business model Transformation Strategy (BmTS) with the aim to develop new generic business process model. Two key architectural elements of the BmTS were the Ten Component Information Model and the utilisation of a Service Oriented Architecture.

Three years later some initial lessons learnt and challenges were reported from Statistics Finland (Rouhuvirta, 2007) and Statistics New Zealand (Dunnet, 2007):

- The exploitation of SOA in application development requires a common data model of statistical data. Without it, it will not be possible to utilise all the productivity benefits that come from progress in information technology.
- There were examples of the problems with the web services compliance of third-party software used in statistical organisations
- The adoption and implementation of SOA as a Statistical Information Architecture requires a significant mind shift from data processing to enabling enterprise business processes through the delivery of enterprise services.

- Expecting delivery of generic services from input / output specific projects leads to significant tensions, particularly in relation to added scope elements within fixed resource schedules.
- Delivery of business services at the same time as developing and delivering the underlying architecture services adds significant complexity to implementation.
- Skilled resources, familiar with SOA concepts and application are very difficult to recruit, and equally difficult to grow.
- Without significant governance it is very easy to start with a generic service concept and yet still deliver a silo solution.

Five years after the first implementation projects the situation hasn't changed a lot. Most statistical organisations appears to be in a very early phase of SOA uptake. Good indication where we are at is the METIS Wiki with case studies describing experience and lessons learned concerning the development of statistical metadata systems in national and international statistical organizations. Part 4 – System in Design Issues includes a description of IT architecture. Out of total 12 case studies, 6 of them have mentioned SOA as a guiding or promising approach however only 3 claimed they are actively implementing it. It is actually quite hard to find statistical organisation with maturity level higher than 2. We are apparently "behind" in SOA maturity level comparing to some other sectors like Airline Industry.

2. SOA in Historical Context

2.1. Two other Paradigms affecting Statistical Organisations IT Architecture in the last decade

SOA has been impacting architectures in statistical organisations since 2004. It is probably too early to assess the benefits and risks of the implementation in statistical community however we can compare the SOA with some earlier high profile architectural approaches. The paper which has provided theoretical basis for information systems architecture in many national and international statistical organisations (Sundgren, 1999) lists two key paradigms: Data Warehousing and Metadata-driven systems. "In summary, the statistical information system architecture of a statistical organisation should provide an efficient common framework for individual information systems corresponding to (Sundgren, 1999):

- survey processing systems
- clearing-house systems, "data warehouses"
- registers
- analytical processing systems.

Data warehousing should provide an answer for at least two types of systems listed above: clearing-house systems and analytical systems.

2.2. Data Warehousing

According to the Wikipedia the data warehousing concept dates back to the late 1980s when IBM researchers Barry Devlin and Paul Murphy developed the "business data warehouse". Most of the work to develop data warehousing architecture and modelling techniques and widely promote the concept was however undertaken in early nineties. "But if we do not strictly limit the concept of

the "data warehouse" to commercial enterprises and their business data, the first manifestations can be identified as early as in the 1970s – in the form of statistical output databases. At Statistics Austria, the output database ISIS ("Integrated Statistical Information System", also known internationally as LASD – "Large Scale Statistical Data System") was developed as early as 1972/73 – many years before the expressions "data warehouse" and "OLAP" were invented." (Zettl, 2000).

Statistics Austria was hardly the only statistical office using the data warehousing principles and techniques before the expression "data warehouse" was even invented. Similar concepts were used in another mainframe-based output database called Axis (predecessor of PC-Axis) developed by Statistics Sweden. This shows high level of need for data management, analysis and dissemination solutions in statistical organisations. During the last decade of the 20 century the multidimensional data concept and data management ETL (Extract-Transform-Load) techniques were finally developed, adopted by many commercial vendors and started to become widely used in commercial sector. Many statistical organisations started using "new" technology for their data analysis and dissemination solutions. Moreover, data warehousing approach suddenly looked as a perfect candidate for the architectural framework for the whole end to end statistical processing: "We could now go one step further and call the entire statistical office a warehouse." (Zettl, 2000).

To see how well data warehousing technology fits end to end statistical process it is important to understand main benefits of the data warehousing approach and put them in the "statistical" context:

- Decreased cost of data access and analysis. With the third-party analytical and reporting tools it is possible to access and analyse data without the need for (expensive) custom IT development resource.
- More cohesive view on data. Common data model regardless of the data's source enables inconsistencies are identified and resolved. It also supports the drive towards the increased use of administrative data due to possibility to integrate data from different sources.
- Faster and more automated data management process. ETL tools and techniques can significantly improve timeliness and reduce operating costs but still leveraging historical investments in the legacy survey processing systems.

There are however also disadvantages to using a data warehouse. Some of them related to the use in statistical organisations are:

- Over their life, data warehouses can have high costs particularly related to implementation of change. The data warehouse is relatively static meaning that change in data sources, date structures and processing methods require involvement of IT experts. Maintenance costs are therefore high if change occurs often.
- Third-party tools are not always delivering functionality required by statistical organisations, for example confidentiality rules, "statistical-specific" processing like imputation etc.
- Data warehouse concept doesn't really cover the whole statistical process – while it is very strong in Analyse and Disseminate phase it offer less in Process phase and hardly anything in Collect phase.

- DW & BI expertise is required.

2.3. Metadata-driven systems

The specific of metadata-driven concept is that it is generally not possible to buy the tools and applications from the commercial vendors. Wikipedia (as at September 2009) doesn't have a record for metadata-driven system. Even for metadata topic there is a statement: "This article is in need of attention from an expert on the subject". There is however a fair approximation to metadata-driven systems in commercial sector: Enterprise Resource Planning, introduced by research and analysis firm Gartner in 1990. ERP systems now attempt to cover all core functions of an enterprise, regardless of the organization's business or charter. These systems can now be found in non-manufacturing businesses, non-profit organizations and governments. (Wikipedia).
Metadata-driven systems are sort of home-made ERP systems for statistical organisations. They share some general features like configurability, end-to-end coverage and complexity. Main benefits metadata-driven systems are offering are as follows:

- Increased quality and reduced cost of metadata management. Metadata systems have been often developed to efficiently manage different types of metadata like classifications, concepts/variables etc.
- Easier to implement change in existing processes due to configurable metadata. Metadata is used to describe statistical data and methods therefore has a great potential for automated statistical production. With the data structure, processing parameters and workflow stored as a metadata it is possible to configure statistical process in a much more flexible way and implement change without the need of IT redevelopment.
- Increased number of surveys reusing developed metadata-driven solutions. Metadata-driven systems act as a platforms which can host many surveys – the prerequisite is to have metadata standardised across all surveys.

Even though there was a lot of work on metadata theory and best practice in international statistical community there are still not a lot of implementations of fully integrated enterprise-wide metadata-driven statistical production systems. Some difficulties with the implementation:

- Very large initial investment – it is not easy to build metadata-driven statistical production system without comprehensive metadata solutions. It is usually not possible to buy metadata management system or metadata-driven components which require significant amount of (expensive) custom development.
- Metadata is providing significant level of configurability of metadata-driven processes however not so high level of flexibility in terms of use and integration of third-party tools and legacy applications.
- Metadata-driven systems are usually complex environments with the multiple databases and applications so redevelopment costs can be significant.
- Expert metadata knowledge is required

Successful implementations usually have data warehouse as a component of wider metadata-driven platform.

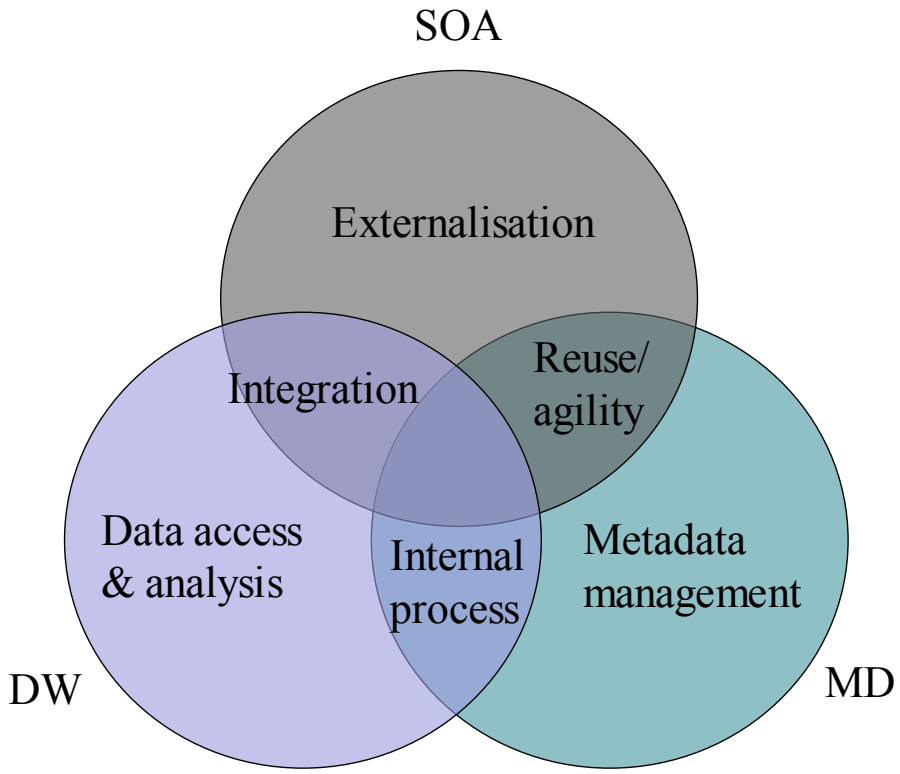

Figure 2: Benefits delivered by three architectural approaches

2.4. Implementation of new architecture is an investment decision

Decision for the implementation of new architectural approach of any kind is in large extent an investment decision. The lessons we have learnt in many projects implementing IT architecture using data warehousing and metadata-driven approach reveal it is a large and expensive organisational change particularly if we aim to do it at the enterprise level. The Victorian government has developed a powerful yet simple approach for investment management: Investment Logic Mapping (ILM) aimed at allowing investment decision-makers within government to 'clearly define the reason for an investment, shape the solution that will best respond to the need and track the delivery of benefits throughout the investment lifecycle'. Description of all three architectural approaches using ILM method (drivers, objectives, benefits, changes and enabling assets) is in the Appendix in Table 1. Please note only elements of ILM have been used here not the whole procedure with facilitated workshops and further work on business case.

2.5. Summary

Comparing *drivers* for investment it becomes clear that there is a large difference between data warehousing, metadata-driven paradigm and SOA. Data Warehousing investment should primarily try to address problems with data management, access and analysis (which as we seen are very important areas in statistical organisations) without too much attention to address lack of corporate agility and poor reuse. While it looks like both metadata-driven systems and SOA primarily address corporate agility and poor reuse (see figure 2) they actually target different aspects of those two problems: SOA is particularly focused on effective integration and reconfiguration of diverse (internal and external) solutions. Metadata-driven systems provide flexibility on the lower - data structures and parameters level. *Objectives* (impediments) of investment are quite similar: in all three cases standardisation and business process reengineering is required as part of implementation project.

All three approaches seem to be very relevant for statistical organisations and experience with the implementation of data warehousing and metadata-driven systems can now disclose some challenges which might help us to understand why SOA uptake is still so low:

1. Experience with data warehousing and metadata-driven approach shows high degree of organisational change required which is usually a slow process.
2. Main problem is lack of standardisation – it appears every new paradigm requires more of it so it is very important to understand the existing level of standardisation in organisation before we commence new implementation project.
3. It is difficult to establish new governance required for ongoing standardisation and decision making. SOA is dependent on strong governance, particularly to progress to more mature stages.
4. New architecture usually requires complete replacement of existing portfolio of assets. Due to historical silo approach statistical organisations now have a large number of old application assets. Unless there is a lot of money, time and internal capability to undertake transformational project the progress will be very slow. It is interesting that SOA isn't necessary better positioned to address this problem even though it allows evolutionary approach with the integration of legacy systems into the new architecture: the problem is that most of existing legacy systems aren't particularly good candidate for such integration – examples are uniquely developed end-to-end statistical systems with hard-coded metadata, solutions developed by end-users etc.
5. Custom application development is not a core business of statistical organisations. Commercial solutions on the other side usually lacks in some specific functionality required for statistical processing. Tools developed in statistical organisations still don't offer appropriate interfaces required for an easy integration.
6. Common challenge organisations often face involves managing service metadata. Metadata is absolutely crucial for SOA.

Many challenges mentioned above are often listed in a general findings about the SOA implementation challenges. Does it mean we are not so different to other organisations? We are not however we have to take into consideration some challenges specific to the statistical organisations – devil is always in the "details":

- Many semantically diverse data structures. Data models underpinning other sectors/industries are usually less diverse and complicated than ours. At the end of the day we are here to measure the whole environment: economic, social, environmental and more.

- Frequent changes in data sources, data structure, questionnaires. Change is common in many sectors/industries however we are less likely in the position to control it because it is often generated externally.
- Specific requirements like data confidentiality.
- Many stove-piped legacy assets. Nowadays many organisations face challenges how to overcome risks related to the old and complex application assets however statistical processing systems have been historically often implemented as a silo systems. We do have a difficult challenge how to consolidate our applications on standardized application platforms.
- Mainly non-transactional processing. Statistical processing has traditionally been very "batch oriented". Messaging infrastructure underpinning SOA architecture is very "transaction-oriented". The challenge is not only how to adopt transaction-oriented infrastructure for the use in batch environment but also to identify processes where we could benefit from the transition from batch to transaction processing.
- End-user processing environments. Statistical organisations often perform quite a lot of data processing and analysis in end-user environment.

3. Looking forward

3.1. Some areas where SOA can deliver significant benefits

As with any other large organisational change – implementation of SOA in statistical organisation should start in area which can quickly deliver results and provide confidence that approach is feasible and can and will deliver benefits. There are two interesting areas which are a good candidate to start implementation in line with SOA principles:

- Metadata services. The first obvious candidate are metadata management systems, for example classification management system, variable management system etc. Even if they are currently implemented as a legacy systems it is worth developing service interface because metadata services are a good candidate for reuse in many stovepipe and corporate applications.
- Statistical tools. Specialised tools developed by statistical organisations, for example imputation, seasonal adjustment, confidentiality, dissemination, CATI/CAPI are currently the only successful examples of shared development and reuse across statistical community. Making them more interoperable using service interface would significantly improve the possibilities to integrate them in different IT environments and therefore increase their usage.

This two areas alone could deliver important benefits and increase collaboration between statistical organisation, for example shared development.

3.2. The next great IT paradigm

While SOA is disappearing from the headlines a new great paradigm which claims it will change IT delivery models has arrived. From the viewpoint of efficiency (value for money), Software as a Service (SaaS) delivered via internet (cloud computing) seems very promising although it will require integration between internal and external systems. It is however not very surprising that the success of this integration and therefore uptake of SaaS will rely on services – similarly to success of SOA which has relied on metadata.

References

Gartner Group (2008). User Survey Analysis: SOA and Application Development and Integration Trends, Worldwide, 2008, 3 September 2008.
http://www.gartner.com/DisplayDocument?ref=g_search&id=750742

Mackie, K. (2007). SOA Uptake Still 'Early,' Analysts Say.
http://adtmag.com/articles/2007/06/18/soa-uptake-still-early-analysts-say.aspx.

Mäkinen, M.; Räikkönen, R; Syvänperä, R.; Ylönen, S. (2004). MetaAPI - improving the usability of existing metadata systems by application programming interface. Work session on statistical metadata, Geneva, 9-11 February 2004.
http://www.unece.org/stats/documents/2004/02/metis/wp.6.e.pdf

Manes, A. T. (2009). SOA is Dead; Long Live Services.
http://apsblog.burtongroup.com/2009/01/soa-is-dead-long-live-services.html.

An examination of the value of the Victorian Government's investment logic map as a tool for front-end evaluation of investment proposals, Evaluation Journal of Australasia, Vol. 8, No. 1, 2008, pp. 26 – 39

Open Group, OSIMM:
http://www.theopengroup.org/projects/osimm/doc.tpl?CALLER=documents.tpl&dcat=14&gdid=17990

OpenGroup (2009) http://www.opengroup.org/projects/soa/ (accessed 4/9/2009)

Rouhuvirta, H. (2007). Developing Statistical Information Systems and XML Information Technologies - Possibilities and Practicable Solutions. Meeting on the Management of Statistical Information Systems (MSIS 2007) (Geneva, 8-10 May 2007)
http://www.unece.org/stats/documents/ece/ces/ge.50/2007/mtg1/wp.9.e.pdf

SOA BluePrint SOA Maturity models. http://www.soablueprint.com/maturity_models.
Sonic Software, Sonic SOA Maturity Pyramid
http://www.sonicsoftware.com/solutions/service_oriented_architecture/soa_maturity_model/index.ssp

Srikanth Inaganti, Sriram Aravamudan: SOA Maturity Model, BPTrends, April 2007

Sundgren, B (1999). An Information Systems Architecture for National and International Statistical Organisation. Meeting on the Management of Statistical Information Technology (Geneva, Switzerland, 15-17 February 1999).
http://www.unece.org/stats/documents/ces/ac.71/1999/4.e.pdf

UNECE, Case Studies. http://www1.unece.org/stat/platform/display/metis/METIS-wiki

Wikipedia, Data warehouse. http://en.wikipedia.org/wiki/Data_warehouse

Wikipedia, Enterprise Resource Planning.
http://en.wikipedia.org/wiki/Enterprise_resource_planning

Wikipedia, Metadata. http://en.wikipedia.org/wiki/Metadata

Zettl, G. (2000). Data Warehouse in a Statistical Office, Seminar on integrated statistical information systems and related matters (ISIS 2000). Riga, Latvia, 29-31 May 2000. http://www.unece.org/stats/documents/ces/sem.43/9.e.pdf