Distr. GENERAL
29 April 2013
WP.21
ENGLISH ONLY

UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN UNION (EUROSTAT)

ORGANISATION FOR ECONOMIC COOPERATION
AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE

Work Session on Statistical Metadata
(Geneva, Switzerland 6-8 May 2013)
Topic (iii): Metadata in the statistical business process

METADATA DRIVEN BUSINESS PROCESS IN THE AUSTRALIAN BUREAU OF STATISTICS

Working Paper

Prepared by[1] Aurito Rivera, ABS; Simon Wall, ABS; Michael Glasson, ABS

I. Introduction

1. It is becoming widely accepted that National Statistical Institutes (NSIs) are confronted with fundamental challenges in regard to remaining relevant and sustainable in a fast evolving digital age[2]. This paper briefly revisits the strategy adopted by the High Level Group for the Modernisation of Statistical Production and Services (HLG) - for addressing these challenges; examines the role of metadata driven business process in this strategy; briefly discusses recent examples of what the Australian Bureau of Statistics (ABS) is doing in regard to 'metadata driven business process'; and explores what the metadata driven business process paradigm might mean to metadata management in NSIs.

2. There is a strong affinity between the concept of metadata driven business processes and the concept of a Common Statistical Production Architecture (CSPA)[3] (or 'plug and play' architecture) being pursued by HLG during 2013. While there are strong synergies, metadata driven business processes could be implemented without global agreement on a CSPA. In addition, CSPA requires agreement on more facets of architectural design than simply how metadata will drive business processes.

II. NSIs in the digital age

3. As early as 2009, the Chief Statistician of the Australian Bureau of Statistics (ABS), Brian Pink, together with some members of the ABS Senior Management, articulated a compelling case for statistical innovation and international collaboration[4] – prompted by the changes brought on by the

[1] The authors would like to acknowledge the contributions made by Alistair Hamilton (ABS), Oleg Volguine (ABS) and Kelly Boettcher (ABS)
[2] http://www1.unece.org/stat/platform/display/hlgbas/Strategic+vision+of+the+HLG
[3] http://www1.unece.org/stat/platform/download/attachments/58492100/Plug+and+Play+project+outline.docx
[4] http://www.unescap.org/stat/MSIS/egm-Jun2011/session-opening-ABS.pdf

digital age. This was influential in mobilising the ‘statistics industry’ to form what is now known as the HLG.

4. As direct response to the aforementioned challenges the HLG has articulated a vision and implementation strategy with the following objectives;

    - New and better products that serve a more global purpose – through innovative re-use/re-purposing of existing data.
    - Production processes which are automated and sustainable (to allow reallocation of resources) and agile (e.g. readily configurable) to respond to changing needs and opportunities.
        a. The ability to share new processes, methods and IT components internationally, facilitated through standardisation in regard to the 'industry' of producing official statistics, is seen as a key means of achieving modernised statistical production.

III. The role of metadata driven business process in the industrialisation and standardisation of statistics production

5. The HLG Vision from 2010 outlines 'The road to industrialisation and standardisation'. This includes statistical production processes which are designed using methods and IT components which are configurable, modular, exchangeable between organisations and as independent as possible from subject matter constraints.

6. The Vision strongly shaped the high level directions HLG provided for the development of the Generic Statistical Information Model (GSIM)[5]. HLG asked that GSIM provide a set of standardized information objects, which would be the inputs and outputs in the design and production of statistics.

7. Among other benefits, GSIM would:

    - Create an environment prepared for reuse and sharing of methods, components and processes, and
    - Provide the opportunity to implement rule based process control, and minimize human intervention in the production process.

8. This highlights the essential role HLG sees for standardised statistical information (particularly various forms of metadata) in determining - in practice - how statistical business processes are reused, assembled and configured - to provide an automated flow for transforming data obtained from various sources into statistics in future.

9. The thinking of the ABS in this regard is similar to that of INE Spain. INE summarises the aim as configurable, rule-based and modular ways of producing statistics[6]. They highlight the need for parameterized tools, which allow different behaviours in different surveys and collection channels, in order to achieve integration and reuse of components. INE proposes that each of the properties and parameters that define the production phases will be reflected in a corporate-wide metadata system, standardized and integrated for all statistical operations.

10. ABS commonly refers to 'configurable, rule-based and modular ways of producing statistics' as 'metadata driven processes'. Similarly to INE, ABS scopes metadata more broadly than just traditional structural and reference metadata to include, for example, 'process metadata'.

[5] http://www1.unece.org/stat/platform/pages/viewpage.action?pageId=59703371
[6] http://www.q2012.gr/articlefiles/sessions/17.1_Revilla_INE%20Spain%20Topic%20II%20Session%2017.pdf

IV. Metadata Driven Business Process in the ABS

11. For the purposes of this paper, key concepts are defined as follows:

    - **Statistical metadata** is taken to mean ‘data about statistical data’ and comprise data and other documentation that describe objects in a formalised way[7].
    - **Statistical process metadata** is taken to mean (a kind of) metadata that:
        a. Describe processes, or
        b. Represent ‘process metrics’. Examples of process metrics are; response rates, imputation rates, edit rates.

12. Whenever the word ‘metadata’ is used without qualification, it should be taken to include both ‘statistical metadata’ and ‘statistical process metadata’.

13. The IT industry and the statistical community offer a range of definitions of what constitutes a 'metadata-driven business process'.

14. In the ABS, it is defined as the systematic and consistent use of metadata to determine the inputs, outputs and behaviour of a statistical business process.

15. 'Metadata-driven business processes' in the ABS should exhibit the following characteristics:

    - Metadata is used systematically
        a. Metadata is used in a planned and managed way across the organisation.
    - Metadata is used consistently
        a. Authoritative ‘single source of truth’ metadata is used for every GSBPM sub-processes where it is relevant.
    - Metadata is used actively
        a. Metadata is used to automate the definition and execution of statistical processes.
        b. Metadata is structured so as to be machine-consumable.

16. When 'metadata-driven business processes' are implemented, a standard metadata framework and information standard are essential elements, because they provide a key underlying system within which business process components interact. The ABS will use the GSIM framework, and the DDI and SDMX Standards.

17. The metadata essential to metadata-driven process include:

    - Metadata to describe the transformable and reference inputs to a business process, e.g. 'Australia and New Zealand Standard Industrial Classification', ‘Output File Location’, ‘Hot Deck File Location’;
    - Metadata parameters that instruct a process on how it should operate, e.g. ‘number of iterations’, ‘quality thresholds which must be met before moving on to the next step', ‘derivation formula’;
    - Metadata that describe the outputs the business process is required to produce;
    - Process logging outputs that report on the success, elapsed time or other aspect of a business process.

18. Of course a metadata framework is not the only enabler of this plug and play vision. The other necessary components for implementation are:

    - Libraries of business processes, methods and IT components. These components can be assembled in a ‘plug and play’ way. Processes are atomic or are complex processes that are assembled from atomic process components.

[7] http://www.unece.org/fileadmin/DAM/stats/publications/CMF_PartA.pdf

- Effective governance framework to ensure the appropriate specification and use of contents. The governance framework dictates what and how metadata is used in what processes.

19. For a number of years now the ABS has been on the path towards metadata-driven business process in one form or another. A brief history of metadata in the ABS[8] shows:

    - In 2003, the ABS developed and formalised its strategy for end-to-end management of metadata
    - Around then, it undertook the Business Statistics Innovation Program (BSIP)[9] – which was aimed at re-engineering the ABS business statistics processes using innovative technologies and methodologies
    - In 2009 the ABS formally adopted the use of DDI and SDMX Standards
    - In 2010 it initiated the Information Management Transformation Program (IMTP) – which sought to re-engineer business processes, develop key conceptual information models, capabilities and information management best practice that are required to put in place metadata-driven statistics production.
    - Today, under the ABS 2017 program, the work towards the metadata driven business process paradigm is continuing and have widened in scope.

20. Following are two recent examples of projects directly aimed at realising metadata-driven business process in the ABS.

21. It should be kept in mind that there are other significant undertakings in the ABS that are also integral to making metadata-driven business processes a reality in the Bureau. For instance, the ABS Enterprise Data Warehouse.

V. Case Study 1 – Metadata Registry/Repository (MRR), Statistical Workflow Management System (SWM)

22. MRR and SWM are two core ABS applications which are currently being developed under the ABS 2017 Program but which have their origins in the Information Management Transformation Program. Together they are considered the 'central nervous system' of the future metadata management infrastructure in the ABS.

23. The MRR is a store of statistical metadata that is authoritative, centralised, and standards based. It has two components:

    - the Repository – a logically centralised store for standards-based metadata
    - the Registry - a metadata catalogue and data directory - that tells the user (human or machine) what metadata are in the Repository and where statistical data are located.

24. The MRR and SWM aim to guide and quicken the work of Statistics Producers in their various activities throughout the end-to-end statistics production cycle. When authoring metadata, the user is guided by SWM through the step-by-step work that needs to be undertaken. At the same time, SWM also helps the user effectively re-use existing statistical metadata. SWM performs this by orchestrating the dialogue between 'user process steps' and the MRR. There are three key MRR functions - these are depicted in Figure 1.

25. As can be seen in Figure 1, the operations that can be performed by the MRR are:

    - Register metadata
    - Search for existing metadata in the Repository

[8] http://www1.unece.org/stat/platform/display/metis/A+Brief+History+of+Metadata+(in+the+ABS)

[9] http://www.abs.gov.au/ausstats/abs@.nsf/95553f4ed9b60a374a2568030012e707/a34271aed0c116a7ca25719b000498c5!OpenDocument

- Retrieve existing metadata from the Repository

When used in the context of SWM, these will eliminate duplication of metadata as well save time and effort through re-use.

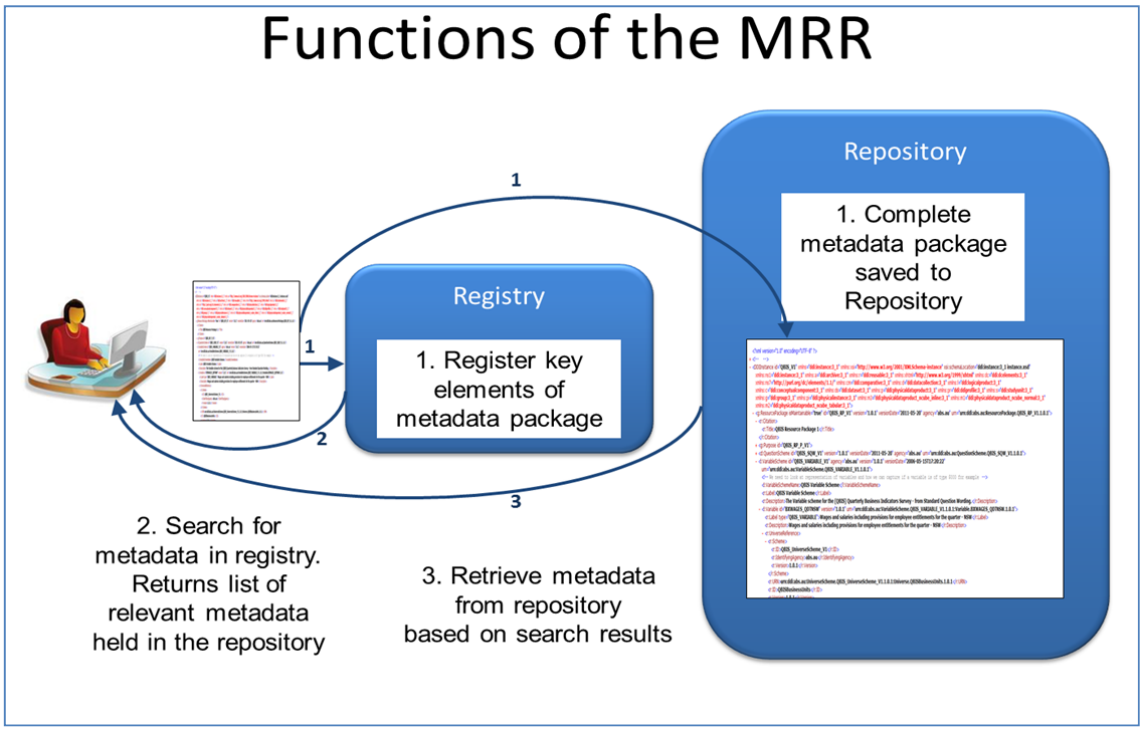

Figure 1: Functions of the MRR

26. The MRR will have a Registry Information Model which is a subset of GSIM V1.0. The Registry Information Model spans attributes and relationships that are relevant when registering information objects (e.g. for the purposes of discovering statistical metadata for potential reuse). This means it will not necessarily record every attribute associated with the detailed definition of each object.

27. The SWM is targeted at:

    - Providing Process Designers with the capability to assemble, configure, deploy and 'chain' the various processes within the end-to-end statistics production cycle. The aim is - wherever possible – to support automatic execution based on rules-based business logic.
    - Providing Statistics Production staff the opportunity to add higher value because lower value tasks are automated in a quality-assured manner and they have additional information to guide them in making higher level choices.

    The tool is intended to support the creation of high-level workflow processes (e.g. building the process for a particular survey) through to low-level technical processes (e.g. building re-useable lower level processes that can be used as 'building blocks' for high-level processes). Authored statistical process metadata are registered in MRR – thus making it possible to re-use existing statistical process metadata.

28. Together the MRR and SWM are intended to bring about the following changes to activities throughout the end-to-end statistics production cycle:

    - significant increase in the level of automation,

- significant increase in the level of re-use of statistical information, business processes, methods and IT services/components which support production of statistics, and
- the machine-assisted assembly and execution of business processes

29. In turn, it is anticipated that these changes to the way of doing business in the ABS will bring about the following business benefits:

    - Significant reduction in the time and cost of business operation in the future,
    - Significant reduction in the time and cost of developing and delivering new products and services and harnessing new data sources.

30. It is understood in the ABS that realising these benefits is not simply a matter of putting MRR and SWM in place. There are other parts that must be delivered and activities that must be carried out. These other components and activities are portrayed in Figure 2.

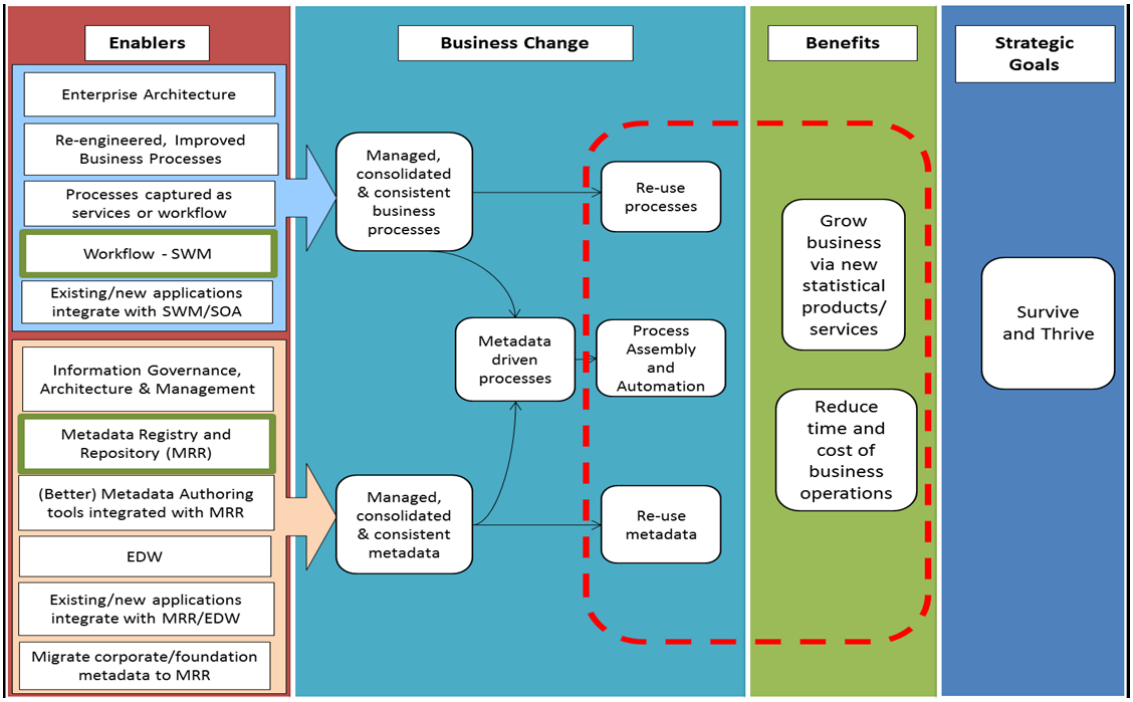

Figure 2: MRR, SWM and other enablers – and how they all contribute to realising the forecasted business benefits.

31. From Figure 2, in addition to MRR/SWM, activities such as re-engineering business processes and instilling a new metadata management culture, etc. - are all required to realise the envisaged business benefits. Thus MRR and SWM are only parts of a wider business transformation program.

32. From the ABS statistical metadata management perspective specifically, having MRR and SWM to facilitate metadata driven statistics production will entail the following changes to the way the ABS manages statistical metadata:

    - SWM/MRR will be the trusted and authoritative source of statistical and process metadata.
    - Applications that define and/or use metadata must link to and interoperate with MRR/ SWM.
    - GSIM and GSBPM frameworks will be used to conceptualise metadata in a standard way. In turn, the DDI and SDMX Standards are used to implement metadata in a standard way. These enable new and existing applications to communicate with MRR/SWM.
    - Governance will be required to ensure appropriate definition, registration and re-use of statistical and process metadata.

- Statistical data and metadata production processes in the ABS will need re-engineering to take advantage of the automation brought on by the SWM/MRR.

33. An illustration of the possible change is that when ABS implemented a new version of our standard Industry Classification (ANZSIC 2006 replaced ANZSIC 1993) managing the transition across every statistical collection which refers to ANZSIC took approximately 5 years, cost around 12 million Australian dollars (approximately 9.5 million Euros) and required a dedicated program management and analysis team of up to 5 people (not mentioning the contributions of the various Business and Support Area teams) just to coordinate the many activities . Examples of issues included

    - Software which had hard coded logic related to use of ANZSIC 1993 and therefore required rewriting of software to work on the basis of ANZSIC 2006
    - Use of undocumented and inconsistent variations of ANZSIC 1993 with unrecognised sub totals, 'dump codes', 'exception codes', labels etc.

    While definition of correspondences, back casting of data and so on would still be required for migration to the next version of ANZSIC beyond 2006, it is considered that appropriate practices related to the use of MRR, SWM, governance of metadata definition and reuse - might more than halve the cost and time, as well as decrease the business continuity risk - associated with the migration.

34. For the ABS Standards and Classifications area (which expertly designs and negotiates new version of ANZSIC) - the change is anticipated to result in increased confidence that their work more directly drives what happens when defining and driving business processes and their outputs across the ABS. Another likely impact to the team is that - it increases the need for the design of, and supporting material associated with, the authoritative version of ANZSIC to fully address (by one means or another) practical business requirements from all the collections which collect or use data that is classified/qualified by Industry. There needs to be stronger governance, with binding decisions on whether a particular requirement from a particular business area should be met (or granted some form of 'exemption') or whether the variation needs to be ceased because that variation does not add sufficient value to warrant the deviation from standards.

35. The first production versions of the SWM and MRR are scheduled for release by December 2013. A Proof of Concept project that was undertaken in 2011/2012 demonstrated that the fundamental concepts and technologies behind the MRR and SWM are sound (for example, it was demonstrated that the ABS eForms system (a web-based data collection tool) can successfully integrate with SWM/MRR and re-use its services (‘Register metadata’, ‘Find metadata’ and ‘Retrieve metadata’). It was also demonstrated that the systematic use of statistical and business metadata can automate business processes. For example, electronic forms can be automatically generated using registered metadata in the MRR and the ‘approval process’ workflow for the newly constructed electronic form can be automated. The following brief case study discusses these in greater detail.

VI. Case Study 2 – ABS Online Forms Prototype Project (eForms)

36. The eForms Project was intended to build a prototype that would serve as the basis for developing future ABS capability for automatically building and deploying web-based collection tools. The idea behind eForms is highly consistent with what Jeremy Iverson describes in his Metadata-Driven Survey Design paper[10].The prototype was primarily targeted at building electronic forms (eForms) but it can be extended to be able to cater for paper-based or interviewer-facilitated methods of collection.

37. The goals of the project included:

[10] http://www.iassistdata.org/downloads/iqvol3312iverson.pdf

- Investigation of the feasibility of developing a system which uses the DDI Version 3.1 Instrument specification metadata (for EEH - Employee Earnings and Hours survey) to facilitate the creation of a functional web form on the ABS Collections Electronic System (ACES)
- Investigation of the use of strategic ABS technologies, platforms and tools in developing and running the above system. These technologies, tools and platforms include the Statistical Workflow Management system, Metadata Registry and Repository and wbe services, to create, store and process metadata to drive industrialised statistical processes.
- Implementation of relevant GSBPM aligned processes. In particular the implementation included the 'Build Instrument', 'Review Instrument', 'Approve Instrument' and 'Certify Instrument' processes.

38. The metadata driven business process in this context utilises:

    - The appropriate metadata consistently in every activity associated with designing, building, reviewing and approving a collection instrument.
    - Metadata expressed in accordance with the DDI 3.1 standard. This spans the full instrument specification including collection title, reference period, questions, variables and data types, as well as questionnaire structure and question flow.

39. As illustrated in Figure 3, at a high level, the system consists of the following components:

    - Web Instrument Production (WIPA) Web Application,
    - SWM (Statistical Workflow Management) system
    - MRR (Metadata Registry and Repository)
    - DDI to ACES Transformer Web Service
    - ABS Collections Electronic System (ACES)

40. The metadata, references to metadata or information extracted from metadata are used to action the business processes.

41. ABS staff fulfilling the role of Metadata Designer defines the relevant metadata that will be required for all of the subsequent activities. This will contain the instrument specification metadata. This metadata is then registered and stored in the MRR.

42. ABS staff fulfilling the role of Instrument Designer choose an instrument to be built based on the list of surveys that have 'ready' metadata specifications (this step can be automated).

43. During the instrument build process, the build service component transforms and generates a concrete functional web form to be hosted on the ACES production grade eCollection system. In this step, conceptual metadata specification for a given survey which contains metadata information such the full instrument specification in DDI is converted to the programming language which will be understood by the ACES system

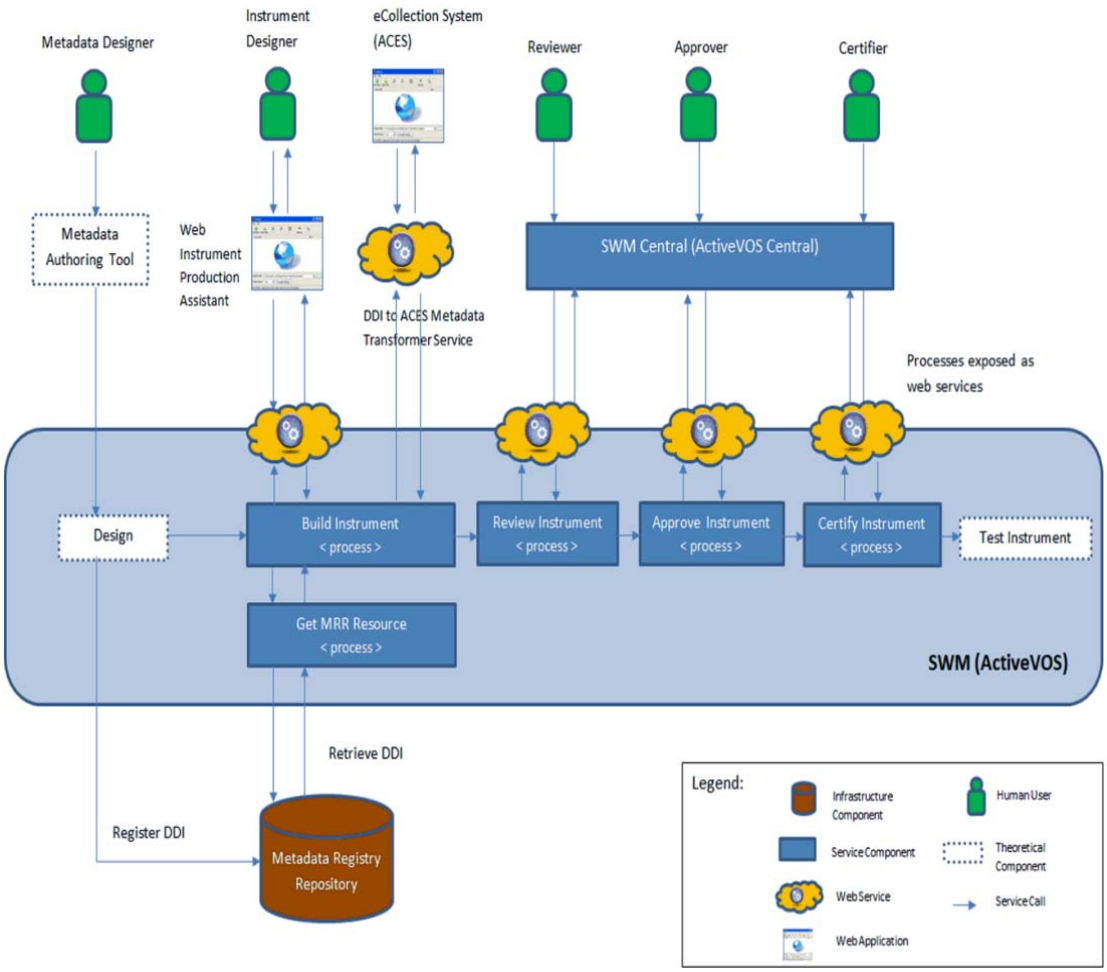

Figure 3: The Web Form Creation System: Show how SWM/MRR is leveraged by metadata systems to automate Web Forms creation processes.

44. Following the creation of the web form, metadata was then used to notify the relevant reviewer, approver and certifier stakeholders about which form is ready for review, approval or certification including any other relevant instrument details that will assist in the quality control process. This included the instrument reference period, despatch or due date as well as other metadata that was relevant to these processes.

45. All three steps; review, approve and certify required human interaction and these processes are actioned through SWM User Interface.

46. The project succeeded in defining the DDI metadata for the EEH survey - which was registered, stored, retrieved and processed using the MRR/SWM to automatically build a functioning web instrument (eForm). This demonstrates that metadata driven business processes are feasible and present the business with improved efficiency and productivity from automation of business processes and re-use.

47. From the metadata management perspective, some of the lessons learnt from this prototyping project include:

    - The need for defining an ABS (or a wider-NSI) DDI Profile[11], with specific guidelines and examples of its use as applies to ABS statistical business. A local DDI Profile will streamline

[11] http://dx.doi.org/10.3886/DDIBestPractices06

and optimise the creation of applications, systems and business processes by providing specific structure and guidelines which will guide business users and technology application delivery teams to create metadata driven solutions.

- The need for adoption or creation of a Metadata Creation and Management Toolbox. The various tools in the Toolbox should be driven by the DDI Profile and allow for the creation of metadata across the entire statistical life cycle, otherwise known as the end-to-end (E2E) metadata. The Metadata Creation and Management Toolbox should naturally integrate with the MRR for storage, versioning and retrieval of metadata resources.
- Centrally registered, accessible and strongly governed metadata is key to a successful implementation.

VII. Impacts of the 'metadata driven business process' paradigm on statistics production and metadata management in the ABS

48. As can be seen from the work already done in the ABS MRR/SWM and eForms activities, the 'metadata driven business process' paradigm is feasible. It is also poised to bring the following changes to the business:

    - significant increase in the level of automation,
    - significant increase in the level of re-use of statistical and process metadata, and
    - the machine-assisted assembly and execution of business processes

    In turn, these changes are expected to bring about the following business benefits:

    - significant reduction in the time and cost of business operation,
    - improvement in quality and coherence of statistical products , and
    - business growth through new statistical products and services

49. The two case studies also point to the ABS needing to put into effect the following changes to the way it manages metadata:

    - Move away from siloed systems to an integrated metadata infrastructure – where SMW/MRR act as the 'central nervous system' and other statistical systems integrate with it.
    - Conceptualise statistical metadata and statistical process metadata in a consistent way (based on GSIM). Then metadata must be implemented for machine consumption in a standard way (e.g. using SDMX and DDI) across various statistical metadata systems.
    - Put in place effective metadata management governance to ensure metadata is defined according to best practice, registered in MRR and re-used appropriately.

50. In exploring what might be the further impacts of the 'metadata driven business process' paradigm on statistical metadata management, the ABS anticipates the following – based on local experience;

    - GSIM and GSBPM need to be supplemented with an agreed on 'template' that depicts what GSIM objects are typically involved as inputs and outputs in each and every GSBPM sub-processes. This will guide the development and use of metadata-driven applications that are required in aiding/automating statistics production processes.
        a. The work product from the METIS Informal Task Force on Metadata Flows across the GSBPM would provide a good starting point for the above-mentioned template.
    - A versioning strategy for statistical metadata must be in place to effect appropriate use and re-use of statistical metadata.
        a. Where business processes are actively driven by metadata, the process owners need to be confident the metadata they are using cannot be changed/overwritten by others. Nevertheless, metadata definitions evolve over time so new versions can be defined, the owner of the process that uses the older version can be

notified automatically that a new version exists, the owner can assess the implications of the new version for their business process and can use the new version of the metadata when they are ready (and readily 'rollback' to using the older version if there is a problem).

- Practical and effective governance for authoring, registering, re-using and ensuring the quality of statistical metadata is required. However, this particular metadata governance arrangement must be seen as a coherent and integral part of an 'eco-system' of governance arrangements. Where each individual governance arrangement addresses the requirements of a facet of the business. That is, metadata governance must be congruent with:
    a. IT systems development governance,
    b. Business process (re) engineering governance,
    c. Project/Program Management governance,
    d. etc.
- Statistical metadata systems that are intended for use in aiding/automating statistics production should be designed to be easily sharable. This is important in realising the HLG vision of standardising and industrialising statistics production.
- Metadata experts or metadata teams must willingly and skilfully engage with business and IT staff and provide expert advice during the design of statistical metadata systems and new statistical business process. This meaningful participation by metadata people is critical in ensuring the fitness-for-purpose of the products.
- 'Fitness for Purpose' - in this instance - also encompasses the realisation of the HLG's ambition of standardising and industrialising statistics production. It would serve this purpose if in the process of advising development projects - metadata practitioners champion 'global optimisation' rather 'local optimisation'. This may broaden the role of metadata practitioners and deepen their engagement with the business.
- In order to provide xpert advice to business and IT staff, statistical metadata practitioners will need to be trained, knowledgeable and constantly up-to-date with aspects of frameworks (e.g. GSIM, GSBPM) and standards (e.g. , SDMX, DDI) which are relevant to their work. For example, ABS metadata practitioners do not require a detailed technical knowledge of standards such as SDMX but they do need to know broadly how it is being used by the ABS to define and manage statistical information, etc.

VIII. Conclusion

51. Early ABS case studies show that the 'metadata driven business process' paradigm is feasible, although the approach is not without challenges. There are technical challenges, but the more significant challenges are achieving new understandings, cultures and work practices among staff, together with defining changed roles at the organisational level.

52. The case studies also provide credible evidence that it will reduce time and cost, improve statistical product quality and provide agility in producing new statistical products and services.

53. Metadata driven business process directly supports the HLG's approach to industrialising and standardising statistics production by automating or minimising human intervention in statistical processes.

54. Realising and leveraging metadata driven business process will entail potentially significant changes to the way metadata is managed in an NSI.

References

http://www1.unece.org/stat/platform/display/hlgbas/Strategic+vision+of+the+HLG

http://www1.unece.org/stat/platform/download/attachments/58492100/Plug+and+Play+project+outline.docx

http://www.unescap.org/stat/MSIS/egm-Jun2011/session-opening-ABS.pdf

http://www1.unece.org/stat/platform/pages/viewpage.action?pageId=59703371

http://www.q2012.gr/articlefiles/sessions/17.1_Revilla_INE%20Spain%20Topic%20II%20Session%2017.pdf

http://www.unece.org/fileadmin/DAM/stats/publications/CMF_PartA.pdf

http://www1.unece.org/stat/platform/display/metis/A+Brief+History+of+Metadata+(in+the+ABS)

http://www.abs.gov.au/ausstats/abs@.nsf/95553f4ed9b60a374a2568030012e707/a34271aed0c116a7ca25719b000498c5!OpenDocument

http://www.iassistdata.org/downloads/iqvol3312iverson.pdf

http://dx.doi.org/10.3886/DDIBestPractices06