Distr.
GENERAL

Working Paper
28 March 2014

ENGLISH ONLY

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

EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN UNION (EUROSTAT)

ORGANISATION FOR ECONOMIC COOPERATION
AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE

UNITED NATIONS
ECONOMIC AND SOCIAL COMMISSION
FOR ASIA AND THE PACIFIC (ESCAP)

Meeting on the Management of Statistical Information Systems (MSIS 2014)
(Dublin, Ireland and Manila, Philippines 14-16 April 2014)

Topic (ii): Standards-based modernization

Statistical Metadata Driven eForms

Improved Business Efficiency through Standardised Automated Statistical Processes

Prepared by Oleg Volguine, Australian Bureau of Statistics

I. Introduction

1. The Australian Bureau of Statistics (ABS) as part of its ABS 2017 transformation program is currently changing the way it conducts its statistical business. A major component of this transformation is the development of the next generation of statistical information infrastructure with the goal of modernising and automating statistical processes to achieve business efficiencies and building a more flexible and responsive national statistical service.[1,2,3,4]

2. In 2012, ABS made the strategic decision to transition its web surveys to the Blaise Internet eForm collection platform. The eForm collection platform is expected to host a large number of quarterly, annual and other periodic and new surveys requested by industry and government, with the potential of running and processing hundreds of thousands of survey forms each year. [3,4,5]

3. Currently Blaise Internet instruments are manually developed by programmers based on eForm specifications produced by business teams. This is a time-consuming and resource intensive process. The ABS aims to improve on its current eForm production processes by providing tools and technology to streamline and automate eForm development processes.

4. This paper will discuss the history, progress, key challenges and lessons learned, as well as the vision for the future for statistical metadata driven eForms at ABS. In particular the paper will explore the use of the DDI standard to specify statistical metadata for collection instruments, which are then used to generate eForms for the Blaise Internet eForm collection platform. Moreover, the paper will also explore the use of strategic ABS information infrastructure such as the Metadata Registry and Repository (MRR) and the Statistical Workflow Management system (SWM) to achieve business process automation for eForm development, as well as potential business efficiencies.

II. Statistical Metadata Driven Blaise eForms

A. Metadata Driven Blaise eForms Project

5. In July 2013, the ABS undertook a prototype project to investigate the feasibility of developing a system which enables a conceptual survey instrument specification defined in DDI 3.1 to be tranformed into a fully functional eForm on the Blaise Internet eForm collection platform.

6. The goal of the project was to develop tools and technology to empower subject matter experts such as survey instrument designers, by giving them an instrument design interface to specify collection instruments. This approach would allow instrument designers to specify instruments in business terminology, and automatically capture the specifications in DDI. The survey instrument specification would then be automatically transformed into Blaise code, forming the foundation of the eForm, which Blaise programmers could then polish and tailor to meet more advanced business requirements.

7. In order to achieve this goal, the ABS is developing an Instrument Authoring Tool (IAT), which enables subject matter experts to specify survey instruments, as well as a transformation service, which generates functional Blaise eForm code using a DDI instrument specification.

B. Metadata Driven Blaise eForms Goals

The key goals of the 2013 Metadata Driven Blaise eForms project included:

(a) **2013-14 Land Management Practices Survey (LaMPS) eForm generated from DDI -** The use of DDI metadata to specify the instrument for the 2013-14 Land Management Practices Survey (LaMPS) and the use of the DDI to Blaise eForm transformation service to generate functional Blaise eForm code, as a foundation step in building the LaMPS production instrument.

(b) **Instrument Authoring Tool -** Development of a statistical Instrument Authoring Tool (IAT), which enables subject matter experts to specify a collection instrument in business terms. The instrument specifications developed using the IAT would be automatically written in the DDI standard.

(c) **DDI to Blaise eForm Transformation Service -** Development of an automated service, which uses DDI instrument metadata to generate functional Blaise eForm code.

C. Metadata Driven Blaise eForms Approach

8. The project approach consisted of developing a prototype solution which was designed to use strategic ABS techonolgies, platforms and tools, including the Statistical Workflow Management (SWM) and the Metadata Registry and Repository (MRR), as well standards such as GSIM, GSBPM, and DDI to support automation of statistical business processes for eForm production. These technologies and tools are described in more detail in the next section.

D. Metadata Driven Blaise eForms Solution

9. The prototype solution design consisted of the following components as depicted in *Fig 1.0 Metadata Driven Blaise eForms solution 2013*:

(a) **Instrument Authoring Tool (IAT)** – Application which allows users to specify survey instrument metadata in a user friendly manner, using business concepts aligned with GSIM business objects. The instrument specification is automatically written in DDI by the IAT and the tool uses SWM services to register and store the specification on the MRR;

(b) **Metadata Transformation Service Module (DDI to Blaise)** – Java program to transform DDI to Blaise Internet 4.8.4 eForm program code. This module is self-contained for future web service implementation;

(c) **Blaise Internet eForm Collection Platform** - ABS online survey collection system which is based on the Statistics Netherlands Blaise Internet 4.8.4 product. The system comprises the Blaise Internet product, which is integrated with ABS back-end statistical processing systems using web services;

(d) **Statistical Workflow Management User Interface –** SWM user interface which allows users, such as reviewers, approvers, and certifiers of instruments to interact with SWM and action business processes for instrument validation and approval;

(e) **Statistical Workflow Management system (SWM)** – ABS business process workflow system. SWM hosts core processes to register, search and retrieve statistical metadata on the MRR, as well as GSBPM aligned processes for eForm development; SWM is based on ActiveVos software which is a BPMN2.0/BPEL standard based commercial of the shelf solution.

(f) **Metadata Registry and Repository (MRR)** – Oracle based data store which stores, versions, queries and retrieves statistical metadata;

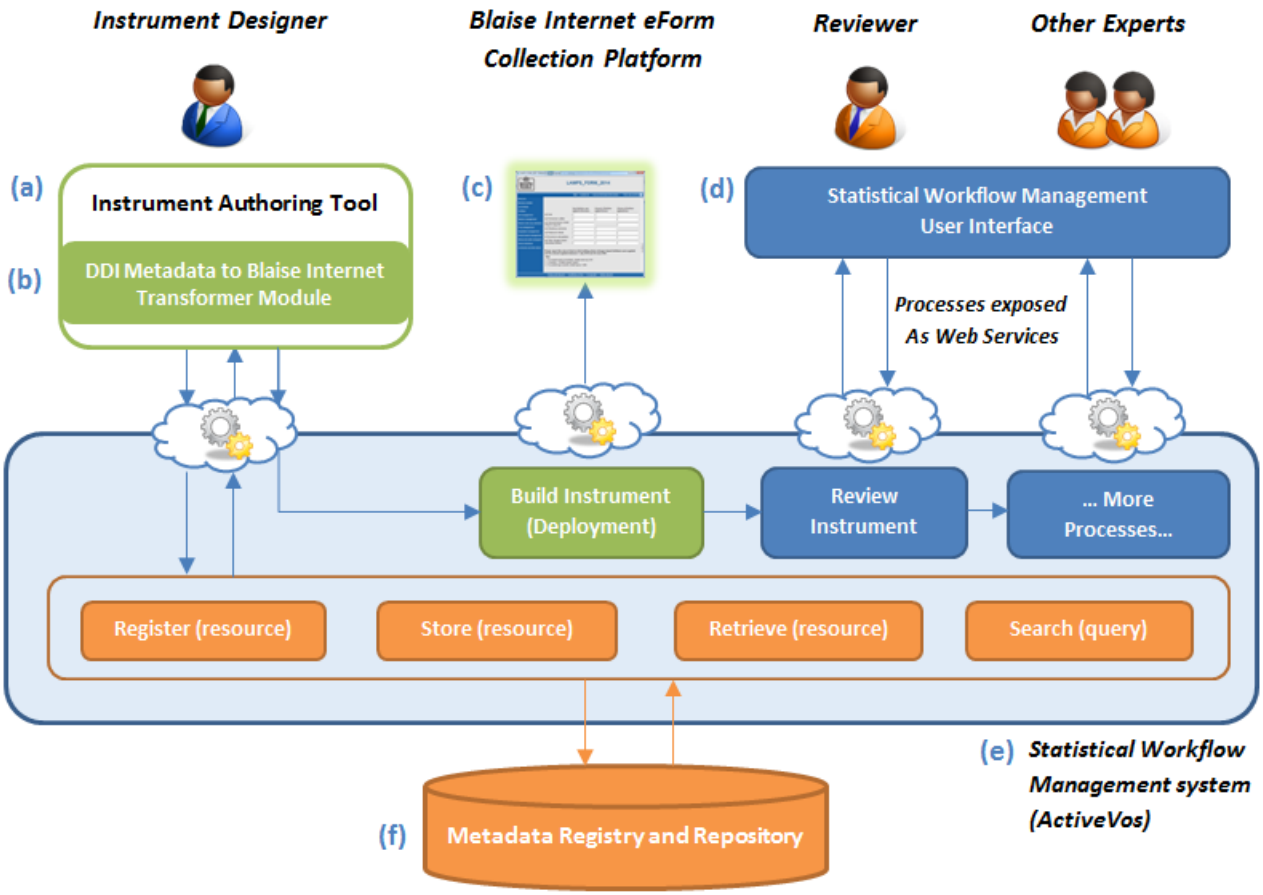


Figure 1.0: Metadata Driven Blaise eForms solution 2013

E. Business Use Case: Creating a Metadata Driven eForm

10. The business use case consisted of a metadata instrument designer using the Instrument Authoring Tool (IAT) to create the survey instrument specification for the LaMPS survey. The IAT allowed the designer to create the instrument specification using a graphical interface to create elements such as questions and question sequences in business terms, which are aligned to GSIM business objects. The specification was then automatically written into the DDI3.1 standard, such that the instrument designer was not required to have any knowledge of DDI. In the full implementation of the business process the DDI specification would then be registered and stored on the MRR using core SWM processes (services) for registration and storage. The instrument designer could then preview the LaMPS instrument by clicking on a "*Preview eForm*" button, which would transform DDI into a Blaise eForm code, as well deploying it to a development environment for a live review by the user.

11. Certain processes such as registration and storage of the DDI to MRR were postponed in terms of implementation as they required the ABS DDI Instrument Profile to be fully defined. These processes and their integration are also generally well understood and were trialled in the previous Metadata Driven ACES eForm prototype project in 2012. [2] For similar reasons the GSBPM processes in SWM for review, approve and certify instrument, which were prototyped in 2012, were also left out of the implementation scope of this project.

F. DDI Instrument Sample for Land Management Practices Survey

12. One of the main focus points of this project was the development of a DDI instrument sample for the Land Management Practices Survey, which would in turn facilitate the development of the ABS DDI Instrument Profile. The purpose of the instrument profile is to clearly specify the allowed and required instrument constructs, such that DDI instrument specifications can be easily and consistently used by existing and future ABS statistical processing systems and applications.

13. The project used version 3.1 of the DDI standard to define the LaMPS instrument sample. However, some minor extensions were needed to the DDI to enable a more complete specification of the instrument as well as a more complete generation of the Blaise eForm. These DDI extensions are noted in Table 1. The LaMPS survey consisted of 56 questions with 345 data items. The DDI instrument specification was approximately 16,000 lines and took under 40 seconds to be processed and transformed into Blaise code. On a technical note, the project team believes that the transformation module could be made to run much more quickly through the optimisation of the XML parser, if required for business reasons, for example to allow more responsive instrument design and more frequent eForm preview/review.

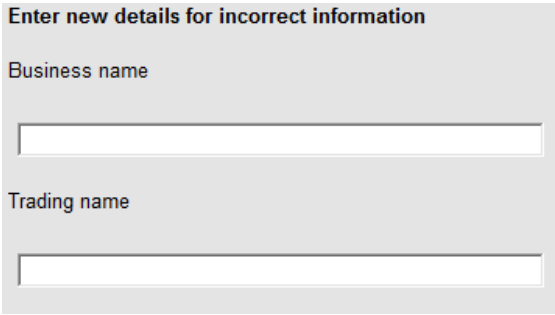
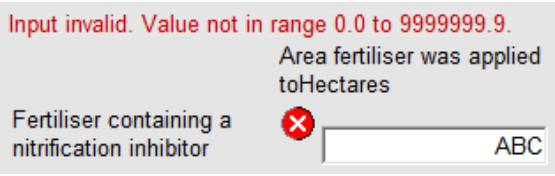
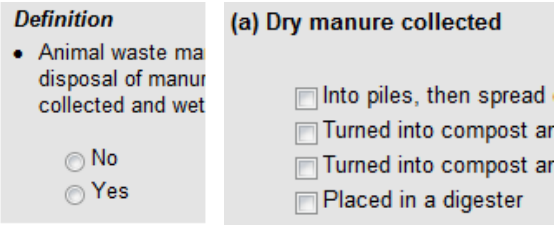
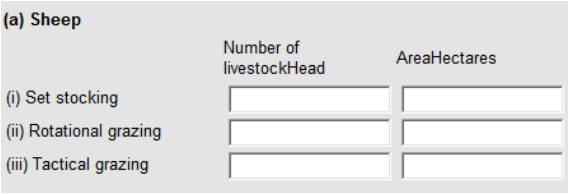
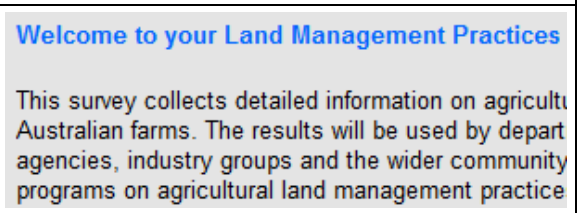
G. Scope of process automation and Blaise eForm generation

14. In terms of Blaise eForm generation from DDI, a comprehensive number of the Blaise code constructs could be automatically generated based on the DDI instrument specification for the LaMPS survey. The project team estimated that approximately 80% of the Blaise eForm code was generated. As result of this, the ABS Blaise programming team was able to complete the first development round of the LaMPS instrument in 2 days rather than the 10 days that was initially estimated. The generated Blaise program code included all the questions, question text and interviewer instructions, including text markup and data types, as well basic sequences via Blaise "RULES" code.

15. Some aspects of the code such as dynamic text (word substitutions), and expressions for If-Then-Else logical conditions were not generated at this time, and were left for manual completion by programmers. This allowed the project team to focus on other deliverables such as Grid/Matrix type questions which were surprisingly complex to specify and produce. Other aspects of the Blaise eForm which were not fully generated for this project included layout instructions, including page breaks and precise positioning and alignment of eForm elements on the screen. A more comprehensive discussion of these challenges is provided in part *H. Key Challenges and Learnings.*

16. The following table (Table 1) provides a concise summary of the DDI constructs and corresponding eForm elements that were defined in DDI and generated into a Blaise eForm. DDI extensions that were needed to achieve automatic generation of Blaise code are also indicated in the table. The intention is not to provide a full or technical specification here, but rather to provide a summarised overview of the type of constructs and transformation that was achieved.

Table 1: **DDI constructs and their generated Blaise code and eForm elements**

DDI Construct/s	Blaise code generated	Blaise eForm elements generated
MutilpleQuestionItems and QuestionItems	Blaise FIELDs and FIELD Text	
Response type: Text, Numeric (length, min, max)	Blaise FIELD Type: Range 0..100, String[length]	
Code Domains and Category Domains *(Note: DDI3.1 extension was used to indicate multiple choice or single choice)	Field type: Enumeration Type and SET of Enumeration Type (Checkbox or RadioButton)	
Structured Mixed Response Domains	Multiple FIELDs, with Other Specify layout	
Grid/Matrix Questions *(Note: DDI3.1 extension was used)	BLOCK with BLOCK and Matrix Layout in Blaise	
Statements and Interviewer Instructions	AUX Fields (data not stored for these fields), or FIELD Text	

DDI Construct/s	Blaise code generated	Blaise eForm elements generated
Sequences for question flow/paths	Blaise RULES sequence for question order and question flow/paths	**Based on the responses you provide in this questionnai questions that do not apply to you. These questions will are only presented with the questions that you need to a** **Are the business name, address details and ABN that ap you correct?** Yes, details are correct No, details need to be updated (including address cha
If-Then-Else branching	Blaise RULES section with If-Then-Else code	Yes - owned and/or operated an agricultural property No - services to agriculture only (e.g. aerial sprayers, No - invest only No - did not operate an agricultural property
Mandatory/optional response	EMPTY response attribute in Blaise	**Enter new details for incorrect information** Business name Please provide a response to this question.
Text Markup: Bold, italicised, underlined text Implemented via inline tags in text For example **Bold Text** (Note: these could also be achieved via XHTML in DDI standard, but are more complex to profile)	Blaise custom tags *@b*, *@i*, @u for HTML markup **, ** etc.	The Australian Government Statistical Clearing House perf Commonwealth Government business surveys as a means necessary, well designed, and place minimum load on bus please click 'Australian Government Statistical Clearing H **Based on the responses you provide in this questionr that do not apply to you. These questions will not be presented with the questions that you need to answe** (e) **Total area of holding** (sum of (a) to (d)) Please make sure that the sum of Questions 4(a) to 4(d)
Text Structure: Columns and Lists for InterviewerInstruction text *(Note: DDI3.1 extension was used to indicate Column or List type)	Question text which included HTML tables and lists (Note: intended to be replaced with CSS for accessibility in the future)	***Definition*** • Soil enhancer - a material added to soil to improve plant g Enhancers are used to correct the soil's deficiencies in st e.g. compost, biochar, gypsum, lime, dolomite. ***Including*** • All in-ground soil management ***Excluding*** • Fertilisers
QuestionItem, StatementItem and InterviewerInstruction constructs	Basic layout instructions for questions, statements and instructions which were based on the type of DDI construct encountered during generation	**Part 3 - Fertiliser** ***Note*** • If the same fertiliser was applied to an area more than on only but the total fertiliser applied to that area. • If two or more types of fertiliser were applied to the same occasion. • Please report in tonnes and/or litres.

H. Key Challenges and Learnings

17. This section outlines the key challenges that the project encountered as well as the key learnings that were achieved.

18. **Layouts for eForm Instruments**

One of the key challenges encountered was the layout of eForm elements such as question text, response boxes, radio buttons and checkboxes, as well as other elements on the actual eForm page. The approach to overcome this challenge was to use standardised corporate templates that specify eForm layouts (including typical positioning of elements), colours and visual styles for those elements. Moreover, a key advantage of having corporate layout templates and styles is to provide a consistent, professional and familiar look and feel for survey respondents.

19. These corporate layout and style standards were applied via mode libraries in Blaise which included standard error message and styles for web pages specified in Cascading Style Sheets (CSS). This approach proved quiet effective, even though, precise alignment of some eForm elements on the page could not be fully achieved, and some manual work was still required by the ABS Blaise programming team. The project team also explored the possibility of further automation to achieve precise positioning of eForm elements, which the team believes may be possible via more advanced techniques that use Blaise application programming interfaces and Blaise metadata layout files to specify more detailed settings.

20. **Grid/Matrix/Table Questions**

Another key challenge, which proved surprisingly complex, was the specification and generation of "*Grid/Matrix/Table"* type questions. These questions constituted the large majority of the questions in the 2013-2014 LaMPS instrument and were characterised by having a table or grid of rows and columns, with different rows of question items having the same column data types. For example, the columns could be *"Age"* and "*Weight"* and the rows could be *"Calves", "Bulls" and "Cows"*. The challenge with these questions was that DDI3.1 does not naturally support the structure or layout of these types of questions, although DDI3.2 does include grid type questions in the standard specification. Moreover, a further complication with these question types was with the development of more in-depth programming code in the *"DDI to Blaise eForm Transformation Module"* in order to handle the grid question structures and produce the correct Blaise code.

21. The approach that was used to overcome the challenge of question grids in DDI3.1 was to use DDI *MultipleQuestionItems* as column headers nested within other *MultipleQuestionItems* for rows. The DDI extension mechanism was also used to indicate whether the question was intended to be presented as a *"Grid"* question. The project team also experimented with a more advanced extension for Grid type questions where sub-questions residing in the Grid question were tagged with Cartesian coordinates *<row, column>* for each of the grid cells. This extension was explored to allow for more complex and precise Grid question layouts.

22. **End-2-End System Metadata**

End-2-End (E2E) system metadata is one of key requirements that must be considered in an auto-generated eForm solution. This is metadata that is needed for statistical processing across the statistical lifecycle for a survey collection from beginning to end. The reason for this is that, in order to achieve effective automation of setting up and processing a statistical collection, the metadata which drives and populates statistical processing systems must align and match across systems. For example, in the case of metadata driven eForms and ABS processing systems, the project team needed to ensure that question identifiers used to generate eForm fields that store collected data must match the identifiers in the ABS collection storage and processing system such as the ABS Input Data Warehouse (IDW). Moreover, depending on how the collected data is captured and stored by the eForm collection system and existing E2E processing solutions it may or may not be possible to easily ensure alignment of data item identifiers, as collected data items may not have a one-to-one relationship between systems.

23. The current approach which is used in the ABS to ensure alignment between collection, storage and processing systems is to use an automated service that runs as part of eForm data extraction and uses machine actionable rules to map and convert data items between systems for compatibility. However, the creation of the mapping rules between the eForm and data warehouse system is a manual process which is both time and resource consuming. For example, for the LaMPS instrument which had 345 data items, 345 rules specifying the mappings between systems were required.

24. The solution that was explored in this space was the use of the Instrument Authoring Tool to create the DDI instrument specification which would then be used to generate both the eForm field names and populate the system names for the back-end data warehouse system, as well as automatically generate the mapping rules for the eForm data extraction service. However, due to project time constraints the full implementation of this solution was postponed.

25. **Instrument Authoring Tool**

The ABS Instrument Authoring Tool (IAT) for instrument specification creation posed another key challenge. The main reason behind the challenge was that the Instrument Authoring Tool was built around an existing ABS legacy application called the Questionnaire Design Tool (QDT) which did not use the DDI standard to specify instruments. The process to align the QDT to use the DDI standard required a lot of effort in terms of re-design and implementation and is still ongoing. However, some very positive results and outcomes have already been achieved from the work on the Instrument Authoring Tool. The results to date included the project team being able to use the IAT graphical user interface to specify the entire LaMPS instrument in 1 day. The same task using a manual approach to create the LaMPS DDI took up to 2-3 weeks. The time to create manual DDI specifications is also comparable to the time it currently takes business teams to create eForm specifications for the Blaise programming team, which was approximately 2 weeks for LaMPS. In the end, as the IAT DDI was not yet fully integrated with the Blaise eForm generation service, the project team relied on using a large portion of the IAT generated DDI combined with the hand-crafted DDI sample for the survey to create the complete DDI specification. The survey specification was then successfully transformed into a functioning Blaise eForm.

I. Metadata Driven Blaise eForms Results

26. The Metadata Driven Blaise eForms Project concluded in February 2014 and succeeded in producing an automatically generated Blaise eForm from a DDI instrument specification. In addition to this, approximately 80% of Blaise eForm code was automatically generated and was used as the foundation for the 2013-14 Land Management Practices Survey. The use of an automatatically generated Blaise code had also cut down the first round of the LaMPS survey development from 10 days to 2 days. Furthermore, the project also laid the foundation for developing the ABS DDI Instrument Profile based on the DDI sample that was produced for the LaMPS instrument. Finally, the project demostrated the use of the Instrument Authoring Tool to define instrument specficaiton in DDI in a efficient manner, which will be used to empower subject matter experts such as survey instrument designers to streamline and automate eForm development processes in the future.

J. Potential for business efficiency through Standardisation and Automation

27. Based on the experience from the metadata driven eForm project as well as ABS prior experience with eForms development across several online collections systems it was noted that there are significant business efficiencies to be gained from automation of statistical business processes and standardised production of eForms. The following table shows an example of the key eForm development steps across systems with manual and automated business processes.

Table 1: **Example comparison of manual and automated eForm development processes**

	eForm System with manual business processes **(Manual processes for metadata definition, limited metadata driven processes)**	**Metadata Driven eForm Prototype** **(DDI instrument specification, SWM/MRR process automation and metadata management)**
Instrument specification	2-4 weeks	2-3 weeks (No Instrument Authoring Tool) 1-3 days (Instrument Authoring Tool)
Production deployment	1-2 weeks (elapsed)	1-5 minutes
Metadata corrections across E2E systems	1 week	Centrally registered and managed metadata *Minutes (*anticipated as more systems use centrally managed metadata)

28. From this example it can be seen that a potential exists for significant business efficiencies from standards based modernisation of statistical business processes which are driven by centrally managed metadata infrastructure and executed through automated and integrated statistical business processes. However, it should also be noted that effective and efficient tools must also be used to create and maintain the metadata, in order to realise maximum business benefits.

III. Vision for the Future and Next Steps

A. Vision for the Future

29. The vision for the future includes statistical subject matter experts working on survey instrument development and using the next generation of statistical information infrastructure to achieve results quickly and efficiently. The subject matter experts will be using such tools as the Instrument Authoring Tool to search for and discover existing survey metadata to quickly update existing surveys for new production cycles and to assemble entirely new survey instruments. In the future, the Instrument Authoring Tool is seamlessly integrated with key parts of the ABS information infrastructure such as SWM and MRR for automated execution of statistical business processes and storage and versioning of statistical metadata assets. With a click of a button the subject matter experts are able to preview and build the completed survey instrument specification to a functional online eForm, with full integration across E2E statistical processing systems. This vision enables the organisation to quickly and efficiently produce collection instruments to support the strategic goals of a more flexible and responsive national statistical service.

B. Next Steps

30. The following is the set of possible next steps for consideration in terms of continuing the development of metadata driven eForms and automation of business processes for eForm production:

(a) Continuation of the development of the Instrument Authoring Tool and the ABS DDI Instrument Profile to support survey instrument creation;

(b) Continuation of the development of the DDI to Blaise eForm transformation service which will support the automated creation of eForms beyond the capabilities developed for the LaMPS survey, for use by other ABS online collections;

(c) Analysis and re-engineering of existing business processes for eForm development, for implementation and automation in SWM;

IV. References

1 Brian Pink, Jenine Borowik & Geoff Lee, *The case for an international statistical innovation program – Transforming national and international statistics systems*, Statistical Journal of the IAOS: Journal of the International Association for Official Statistics, IOS Press, http://iospress.metapress.com/content/2h5764574t6318r4/

2 Stuart Girvan, Chris Conran, Simon Wall, *Applying Information Infrastructure to Business Transformation*, http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.50/2013/Topic_3_Australia_Girvan.pdf

3 ABS Corporate Plan, Jul 2012 (Cat no. 1005.0), 1005.0 - ABS Corporate Plan, Jul 2012

4 ABS Forward Work Program 2012-13 to 2015-16 (Cat no. 1006.0), 1006.0 - Forward Work Program, 2012-13 to 2015-16

5 Adrian Bugg, Kathy Buck, Anthony Davies, Annette Hants, Monica Kempster, *IBUC 2013 Conference Paper: Challenges of Migrating ABS Business and Household Surveys to Blaise Web on a Large Scale and Short Timeframe*, July 2013.