Lessons learned from implementing an integrated methodological architecture

Marlize Pistorius * and Evelyn Maelane
Statistics South Africa, South Africa
E-mail: MarlizeP@statssa.gov.za
EvelynM@statssa.gov.za

Abstracts

In 2011, Statistics South Africa (Stats SA) documented the methodological architecture for Stats SA outlining how integration of systems, standards development and methodological support during the design and implementation of statistical production processes would enhance public confidence and trust in the official statistics. The Project kicked off using a phased approach, where the focus in terms of the first phase was the development and the implementation of the various components required across the Meta-data layer, Data Layer and Application Layer. The paper focuses on the challenges and achievements during the integration thereof. Some of the challenges that will be discussed in the paper would be over commitment of resources, reprioritisation of key projects, relevant skills set, buy-in from the organisation and budget. As part of the achievements the paper will cover e-collections, enhanced data visualisation and interaction through Roambi flow, data analytics and mobile applications as well as the documentation of best practices, methodologies used, and development of applicable standards. Other achievements include the development of a Central Metadata Repository (CMDR) which stores concepts and definitions, variables, classifications. The next phase within the CMDR development would be to expand the CMDR to include statistical metadata as well as a questionnaire builder. Effort has been invested in ensuring that the different components of the architecture are in place, going forward is the process of integration of the developed components across the layers.

Key Words: Methodological architecture, Meta-data layer, Data Layer, Application Layer, Roambi, Central Metadata Repository.

1. Introduction

Methodology & Standards is a cluster within Statistics South Africa which provides methodological and technical support services to producers of statistics as well as maintaining the Business Sampling Frame. Methodology & Evaluation is a division within the Methodology & Standards (M &S) cluster which provides methodological and system support services. The Methodology & Evaluation division is made up of the ADAPT (Application Development And Processing Techniques) component providing system solutions support, the Methodology component providing methodological support that is in line with international best practices and the Evaluation component which is responsible for the evaluation of the processes and standards used by the survey areas. M & S cluster also comprises of the Survey Standards division which is responsible for ensuring the development of standards and putting measures in place to ensure that the standards are adhered to by all business users in the organisation and also ensuring continuous evaluation and improvement of these standards to make sure that they remain relevant to the business users.

Challenges were identified in the cluster which indicated that although the mentioned divisions are working to achieve the same set objective of enhancing public confidence and trust in official statistics, there is lack of integration amongst the set divisions and components with them. The architecture document was compiled outlining the challenges hampering integration of process and procedures within the M & S cluster as well the improvements which need to be implemented in the M & S cluster in order to optimally achieve the set objective of enhancing public confidence and trust in the official statistics.

Areas of improvement were identified in the methodology, the systems development and the standards processes. In order to achieve these improvements, a phased approach has been adopted. The first phase is the development and the implementation of the tool to store the Statistical Standards and Methodologies (Concepts, Definitions, Classifications, and variables), electronic questionnaire collection development, enhanced data visualisation and interaction through Roambi, data analytics and mobile application, the documentation of best practices, methodologies used as well as the development of applicable standards. Figure 1 illustrates the Architecture Diagram.

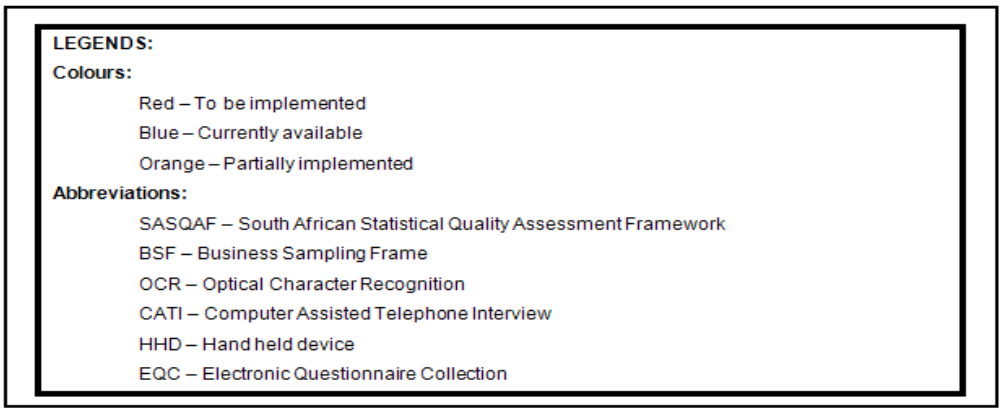

Figure 1 – Architecture Diagram Legends

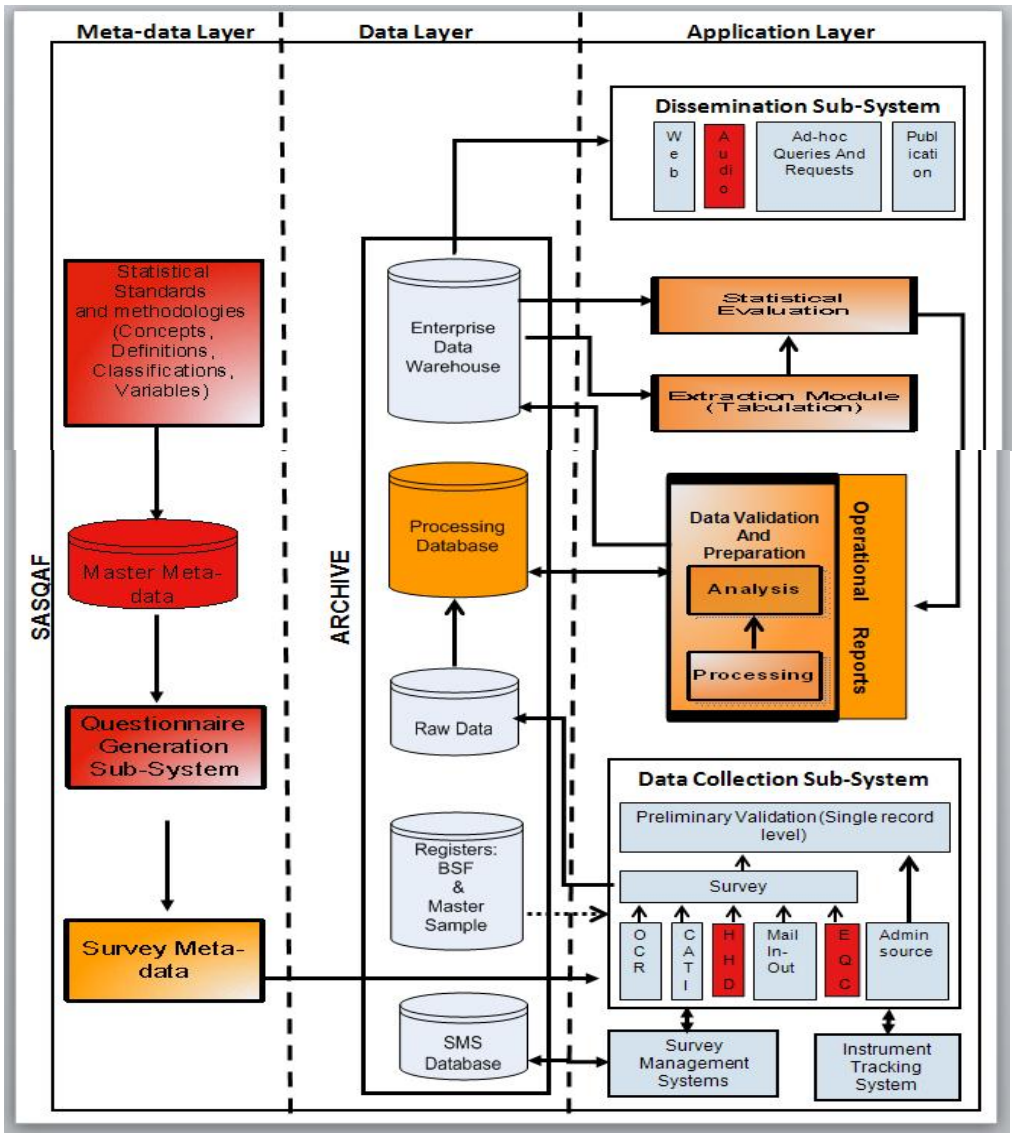

Figure 2 – Architecture Diagram

2. Results

The key process in this phase was the standardisation of Statistical Standards and methodologies (Concepts, Definitions, Classifications, and variables) to be used by survey areas which are part of the meta-data layer according to the architecture document. Although we had approved Statistical Standards and methodologies (Concepts, Definitions, Classifications, and variables) but due to the lack of a tool to be used these were not enforced.

The Central Metadata Repository tool has been developed to enable the user to do the following:

- Register administered items - Ability to register a new administered item,

registered items refer to Concepts, Definitions, Classifications, and variables

- Modify administered items - Ability to modify all administered items, allow version control of modified administered items, ability to only have the latest version of the administered item available for selection and older version will be viewed as history of item.
- Link administered items - Ability to link an administered item to its associated administered items.
- Select administered items - Ability to select technical and statistical metadata to be used for a particular survey, creating questionnaire metadata and only the latest version of the administered item should be available for selection.
- Locking Questionnaire metadata - Ability to lock questionnaire metadata.
- Copy locked questionnaire metadata - Ability to copy locked questionnaire metadata from previous survey period.
- Search/View administered items - Ability to search and view all administered items regardless of the version using a key word, and the ability to view all available survey methodology documents.
- The ability to export administered items and reports to Excel, Word and PDF. The user to will be able to generate and view the following reports: Unified view report, Report by administered item, Report by questionnaire metadata, Report by technical metadata, Report by statistical metadata and Administered item history report.

The above mentioned information will be stored in the SQL Master meta-data database called the CMDR. These conclude the functions in the Meta data layer, which is one of the deliverables for the first phase.

Another deliverable in this phase was the development of the **PDF Electronic questionnaires** to enable the user to receive the questionnaire electronically, save the questionnaire on their local machines, complete the questionnaire without requiring internet connection during completion and only require internet connection during submission of the questionnaire. This method of data collection subsystem forms part of the Application layer. The electronic questionnaires were developed using ADOBE Life Cycle.

Electronic questionnaires have been developed for the following surveys: Annual Financial Statistics (AFS), Quarterly Financial Statistics (QFS), Motor Trade Survey, Wholesale Trade Survey, Retail Trade Survey, Food and Beverage Survey, Tourism Accommodation Survey, Producer Price Index, Electricity Survey and the Utilisation Survey.

Two approaches of the electronic questionnaires were followed. The dynamic approaches allows for one electronic questionnaire form to be used to generate several surveys. This approach further demonstrated the benefits of adopting a metadata driven approach for systems development. Through metadata more than one survey questionnaire is generated from one electronic questionnaire designed form resulting in significant reduction in development and maintenance time. The second approach was the static design were only one survey is generated from an electronic questionnaire form. This method of design is aimed at surveys with unique questionnaire designs which are not similar to other surveys, therefore requiring a different approach to the metadata design.

The solution developed will enable Stats SA to render and distribute interactive, electronic surveys to the various respondents, as identified in the Survey Management System. Distribution will generally take place in batch mode, with all respondents of a

particular survey receiving their surveys at the same time. Each respondent will then complete the survey and submit the results to a secure web service for further processing. The diagram below provides a graphical overview of the solution.

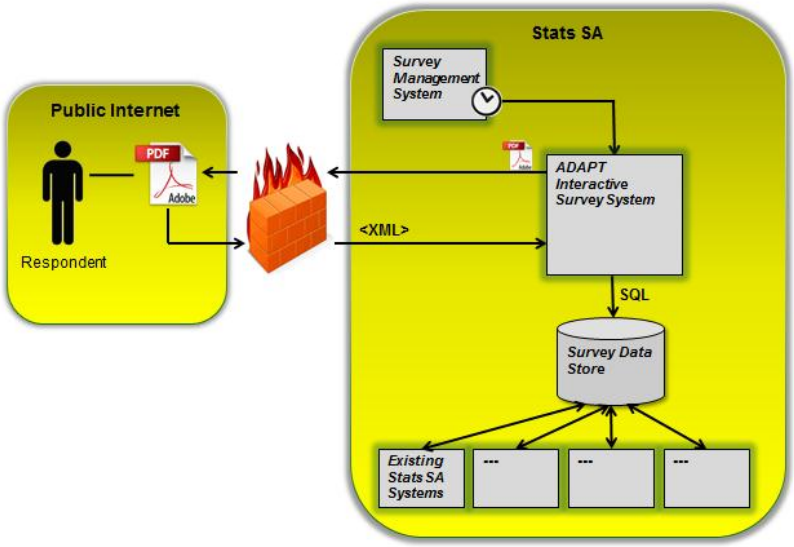

Figure 3 – PDF Solution Overview

For the application layer, **enhancement of data visualisation** was achieved through the dissemination of the Census 2011 results using Roambi as an interactive data dissemination tool and the mobile application which were introduced for the tablet and smart cellular phone users respectively.

To achieve this, Stats SA deployed SAS Business Analytics Platform (BAF), also called the Enterprise Business Intelligence (EBI) platform. The BAF Platform is a comprehensive platform that offers an array of solutions. The solutions range from lite-weight reporting tools such as Web based reports to intricate analytics offered by Enterprise Miner. The BAF platform is used for cleaning, transforming, analysing, processing and preparing publications. The SAS EBI manages all statistical publications that are released by Stats SA from across multiple databases (multiple surveys).

The publications that have been produced will then be used for generating Roambiôs (interactive and analytical graphs and reports). The publication from the Roambi is then downloaded and stored on the mobile devices for users to interact with. Mobile Device users will connect to the Stats SA Roambi portal and download publications to their devices and this will allow them to interact with the data without connecting to the internet.

As part of providing **methodological support** to the economic statistics cluster, the Methodology and Evaluation has successfully standardised the following aspects of the survey value chain for the Application metadata: Sampling methodology, Sample maintenance, Imputation, Estimation, Calculation of response rate, Questionnaire design (done through Questionnaire Clearance Committee) and Seasonal adjustment (the rolling out of the X-12 methodology is in progress).

The Standard division has also concluded the standards documents for the above mentioned aspects excluding estimation.

According to the planned architecture document the following processes are **still outstanding:**

- The next phase within the CMDR development which is expanding the CMDR to include statistical metadata as well as a metadata driven questionnaire builder.
- Part of the Data layer is the Processing database which allows the users to store an extract from the ñRaw dataò database of the required variables and survey periods on which data validations and preparation in the Application layer will be done to prepare data for dissemination. The partitioning and role based control of the processing database into three environments namely the development, preproduction and production still needs to be done. Further changes in this layer will entail the modification of the current SAS programs to use stored processes to facilitate the role based and environment based control.
- Weekly replication of databases in the Data Layer to a disaster recovery site.
- Hand held device data collection method which forms part of the Application layer.
- Although the Meta data driven data validation and preparation, extraction module (tabulation) and statistical evaluation is been done, the integration thereof with other layers has not been achieved yet.
- Investigations into various data quality tools are also underway to determine the suitable to be used for the environment.
- The audio dissemination sub-system which is part of the application layer is still outstanding.
- A PDF portal to reduce the burden on the Stats SA email infrastructure and also add additional possibilities, such as allowing respondents to view the surveys that they need to complete, along with deadline information, download their own survey PDFs, view history of completed surveys and request resubmission of specific surveys, were allowed as per the business rules of the survey.

Major challenges were encountered in the process with the reprioritisation of projects in the organisation in the efforts to ensure the successful completion of the Census 2011 project. This resulted in key skilled resources in the M & S cluster been seconded to the Census 2011 project. The resources seconded included methodologists, database administrators and SAS developers. Significant budgets cuts were also implemented during the financial year which further hampered the progress on the architecture implementation.

3. Conclusions

A lot of effort has been invested in ensuring that the different components of the architecture are in place and they are metadata driven. Going forward is the process of integration of the developed components across the layers. The technology is available to improve the way business operates. It is up to technology houses to take the initiate, investigate and advice business on how these improvements can be made.

4. References

- EN Maelane. (2012) *Architecture Write Up*.
- TN Ntlaba. (2012) *PDF Solution Architecture Stats SA*.
- TN Ntlaba. (2012) *ADAPT - Roambi Solution Overview*.
- TN Ntlaba. (2012) *Census 2011 Mobi Application Solution Review*.