WP. 21
ENGLISH ONLY

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

EUROPEAN COMMISSION STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES (EUROSTAT)

ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT (OECD) STATISTICS DIRECTORATE

Joint UNECE/Eurostat/OECD workshop on statistical metadata (METIS)
(Lisbon, 11-13 March 2009)

Session 4 Metadata case studies

CASE STUDY - STATISTICS SOUTH AFRICA[1]

Submitted by Statistics South Africa[2]

0. BACKGROUND

Organization Details

Name: Statistics South Africa

Contact Details:

Physical Address: De Bruyn Park
170 Andries Street
Pretoria
0001
South Africa

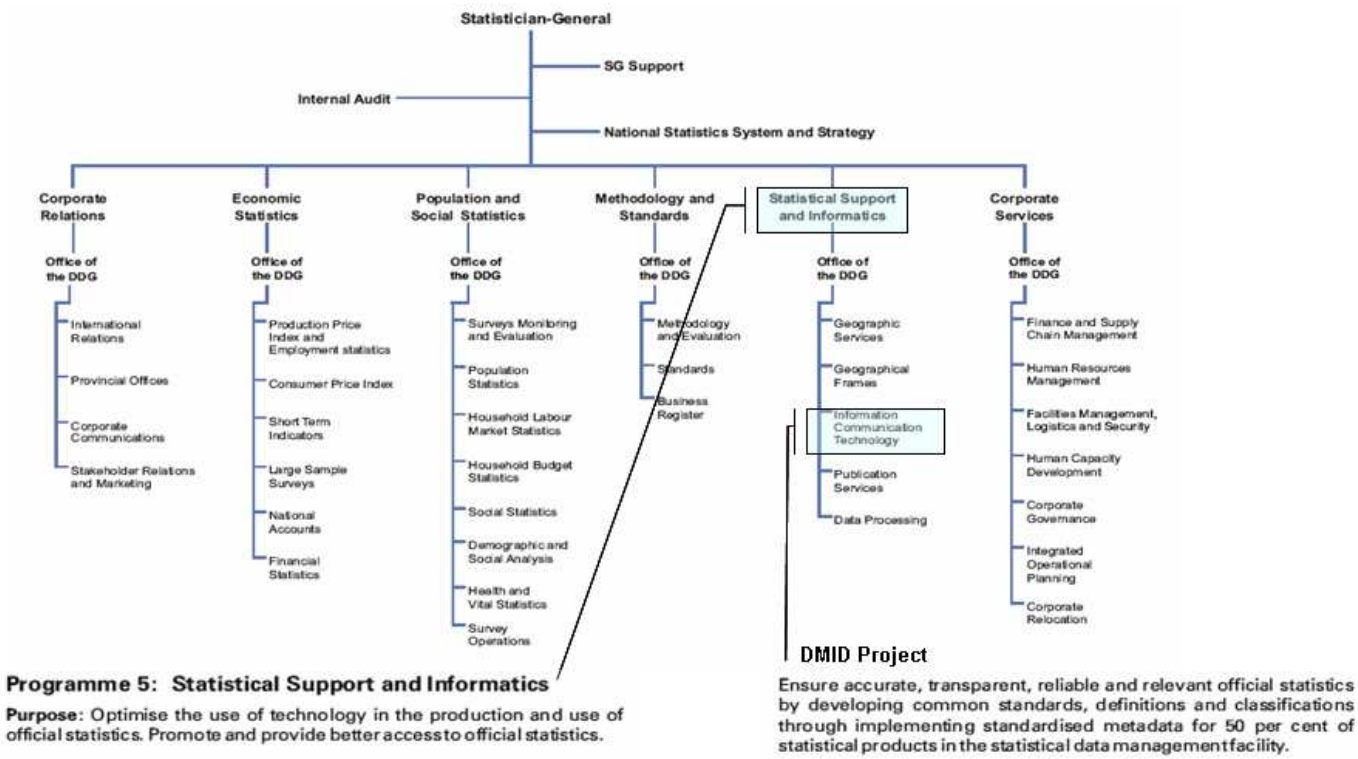
Mailing Address: Private Bag x44
Pretoria
0001
South Africa

Telephone Number: +27-12-310-8911
Facsimile Number: +27-12-321-7381

---

[1] Prepared by, Matile Malimabe (MatileM@statssa.gov.za) and Ashwell Jenneker (AshwellJ@statssa.gov.za) all from Statistics South Africa (www.statssa.gov.za)
[2] Prepared by Sibongile Madonsela (SibongileMA@statssa.gov.za) , Matile Malimabe (MatileM@statssa.gov.za) Bubele Vakalisa (BubeleV@statssa.gov.za) and Ashwell Jenneker (AshwellJ@statssa.gov.za) all from Statistics South Africa (www.statssa.gov.za)

High-Level Organisational Structure



Number of Staff: ± 2,000

Figure 1: Stats SA Organization Chart

The Data Management and Information Delivery (DMID) project (light blue shaded box) is located within the Information Communication Technology (ICT) division within the Statistical Support and Informatics cluster.

The following chart shows how the DMID project is structured:

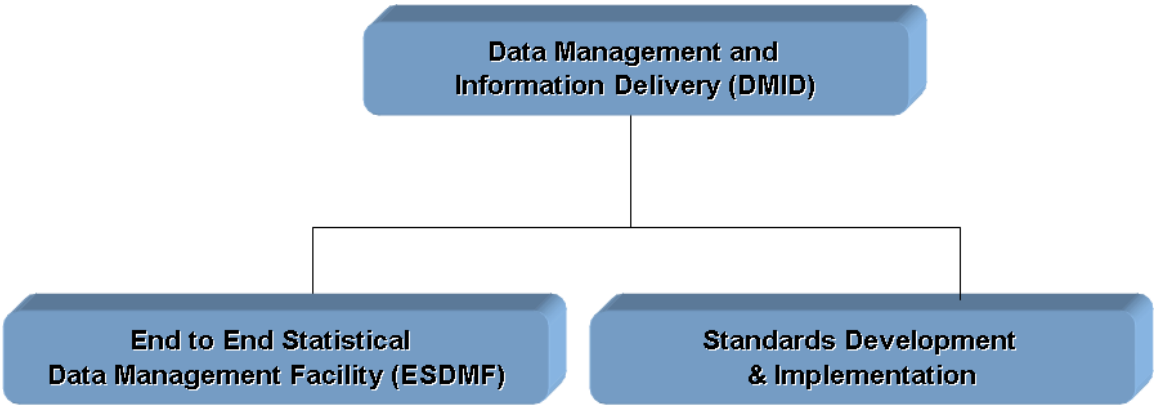


Figure 2: The DMID Project Structure

The following chart shows the organogram of the DMID:

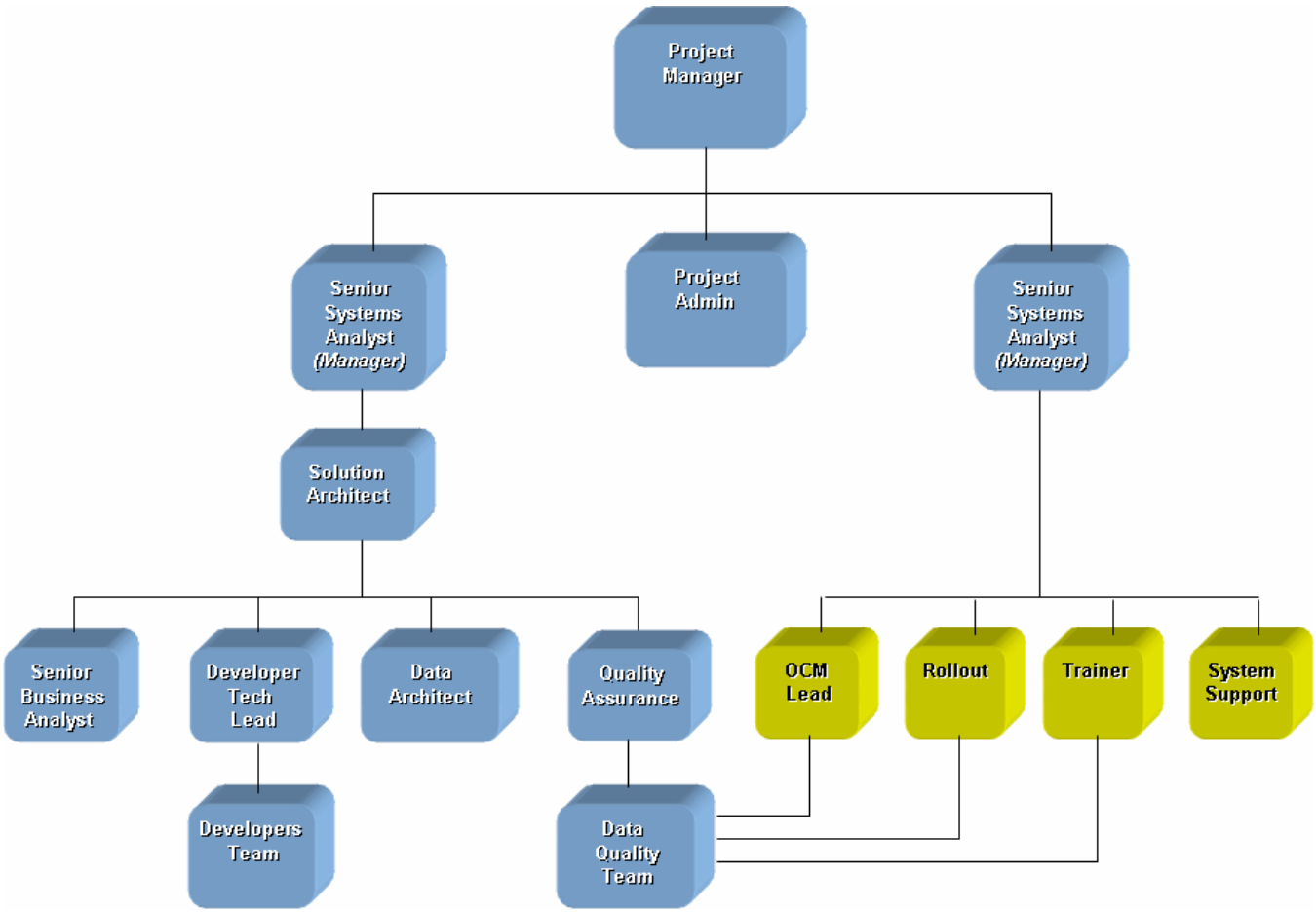


Figure 3: The DMID Project Organogram

Number of staff: ±20

1. INTRODUCTION

Metadata Strategy

Statistics South Africa's development of the metadata management system has its origins in the organisation's requirement to develop a data warehouse. The idea of a data warehouse came about because the organisation wanted to improve the quality of the statistics produced. It was believed that the data warehouse would play a major role in positioning the organisation within its vision of becoming the "preferred supplier of quality statistics". To begin our data warehouse initiative, we paid exploratory visits to various statistical organisations that had embarked on data warehouse developments in order to learn from their experiences. These visits taught us that a number of things about the complexities, difficulties and peculiarities of developing a data warehouse. In particular, our visit to the Australian Bureau of Statistics showed us that for a data warehouse to have any chance of succeeding in a statistical organisation, it needs to have a strong foundation of standards and policies that govern the statistical production processes. Standardisation of concepts and their definitions, as well as classifications of the terms of the actual survey process, were all found to be necessary for the production of quality statistics. For it to be successful, a data warehouse also needs to operate in this environment.

A formal process for standardization was developed through consultation with standards experts standards development and implementation lifecycle was developed to monitor the standardization process. The following is the standards development lifecycle.

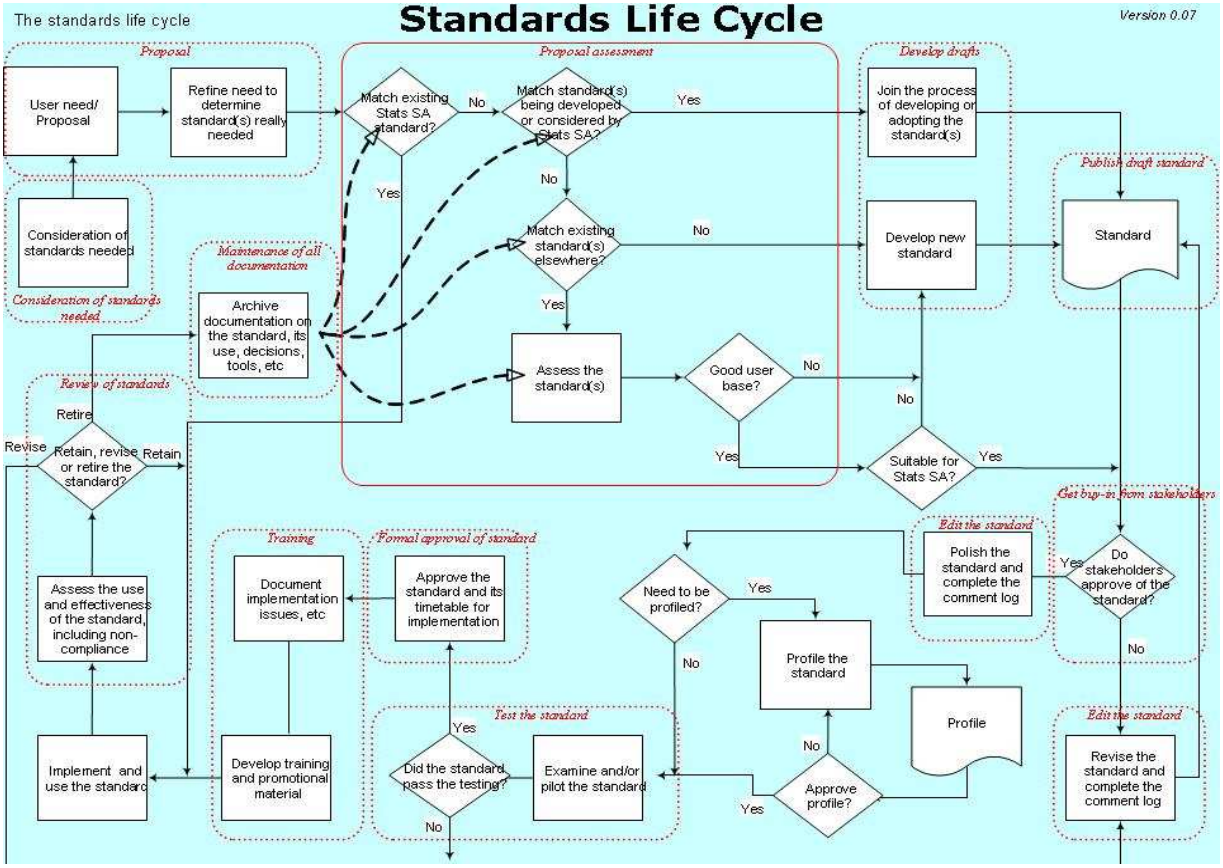


Figure 4: Standards Lifecycle

The next step for us was to investigate the strength of our standards and policy foundation. Upon this investigation, a number of gaps were identified. Chief among these was the lack of standard metadata in the organisation. The need for standardisation of metadata necessitated the development of a metadata management system. However, this had to form a good mix with all the other identified ingredients necessary for the production of quality statistics.

Strategically, our metadata management system forms part of a larger system of applications called the End-to-end Statistical Data Management Facility (ESDMF). As an end-to-end system, the ESDMF will consist of tools and applications to support the whole statistical production process. Within this facility exists a metadata subsystem (refer to figure 5), which plays a central role as the ESDMF was conceived to be metadata driven. In a statistical organisation, a metadata driven system is inevitable because metadata is used and generated at every stage of the statistical production process.

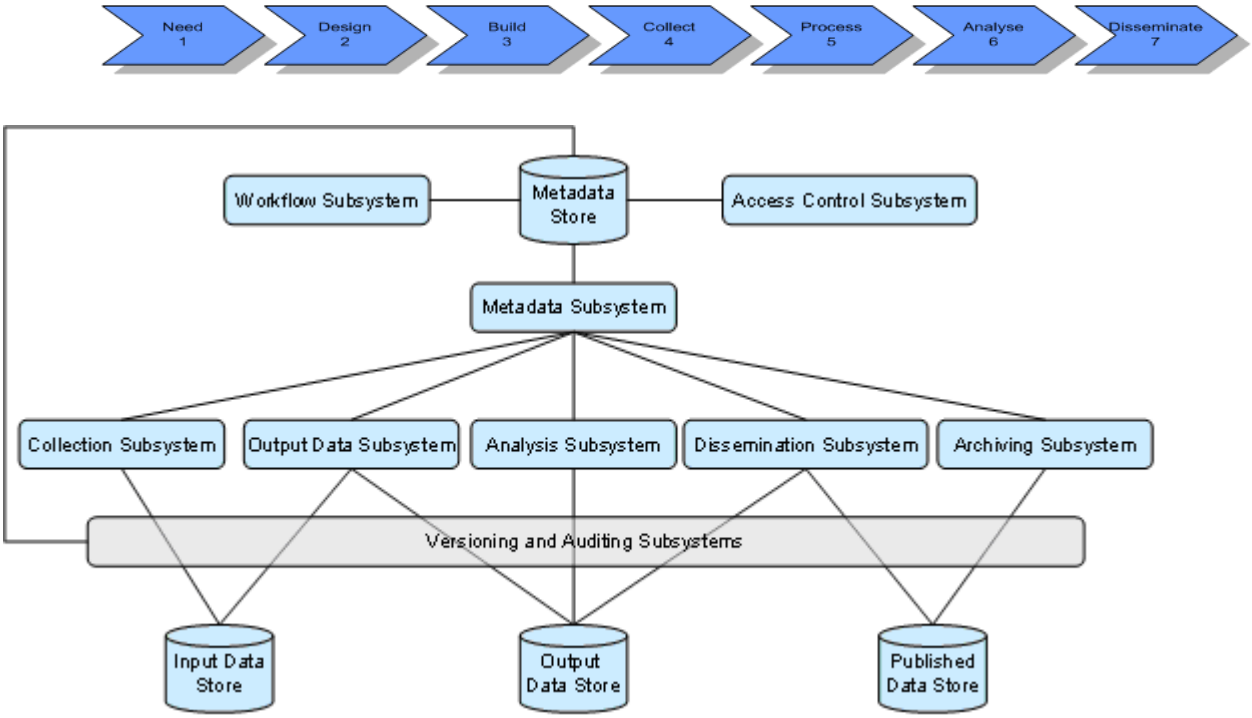


Figure 5: Conceptual components of the ESDMF

As a data factory, a statistical organisation needs to organise and package data in ways that make it useful to the end user. Produced data must also meet certain minimum quality standards. To satisfy both these requirements, use of metadata is invoked. In packaging its data and statistical products, a statistical organisation must ensure that they are attached with metadata for ease of analysis and interpretation by their users. Metadata also play a key role in ensuring that the end products of this data factory are of good quality. Such metadata includes descriptions of concepts used in the organisation, classifications of these concepts, methodologies and business rules. These are all necessary metadata to ensure that products are of good quality.

The development of a metadata management system was informed by the following principles:

- *Maintenance of trust in official statistics:* Descriptions of data collection methods, data processing, and storage needed form part how statistical data are presented to the end user. When presented like this, statistical data and products engender trust to the users.

- *Facilitation of correct interpretation of statistical data:* Metadata accompanying datasets and other statistical products.

- *Quality of statistics:* Standard metadata contributes to the improvement of a number of quality dimensions. Standardisation of concepts and their definitions and classifications are essential ingredients of standardized metadata.

Programme Providing Frame for Stats SA Projects

The work of all Stats SA components is mapped out in the organisation's Work Programme. Organisational units must support the following strategic themes to advance the work of the organisation:

- *Providing Relevant Statistical Information to meet user Needs*

- *Enhancing the Quality of Products and Services*

- *Developing and Promoting Statistical Coordination and Partnerships*

- *Building Human Capacity*

This project is aimed at supporting the strategic theme "Enhancing the Quality of Products and Services". Within the DMID project, the metadata management system, more than any of its components, addresses this strategic theme.

Overall Project Objective

Statistics South Africa's metadata management system therefore forms part of the organisation's broader objective to continuously improve the quality of its products. As the driver of the overall facility, the metadata management system is the first deliverables of the DMID project. The metadata management system is also divided into smaller logical units based on the organisation's classification of its metadata. Survey metadata, consisting of elements for providing the overall description of a statistical survey is the first of these metadata deliverables. The survey metadata component is fashioned along the lines of Statistics Canada's Integrated Metadata Database (IMDB) Metastat.

Following the survey metadata component will be the definitional metadata component. This will incorporate into the metadata management system the standardised organisation-wide concepts and their definitions and classifications as well as other components that form part of definitional metadata.

2. THE STATISTICAL METADATA SYSTEMS AND THE STATISTICAL BUSINESS PROCESS

Overview of the Process Model

Also shown in figure 6 are the main phases of Statistics South Africa's statistical cycle, internally referred to as the statistical value chain (SVC). The development of the SVC forms part of the organisation's standardisation of processes and systems. The SVC will provide broad guidelines for survey areas.

It was decided that the organisation should not re-invent the wheel in its development of the SVC. Therefore in developing the survey cycle, the starting point was to study how national statistics organizations (NSOs) similar to ours had logically structured their survey cycles. Another guiding principle was that Stats SA was not re-engineering the way it conducts surveys, but wanted to map and structure it as much as possible. We had to adapt survey cycles from organisations whose survey operations mimicked those of us. This work was done in 2005. Having made all considerations, we found the Statistics New Zealand's business process model for survey cycles to be the most aligned to our survey operations.

How Stats SA's Statistical Process Maps into Common Metadata Framework (CMF) Lifecycle Model

Resulting form the process stated above, Stats SA's survey cycle consists of the following phases:

1) **Need**

   Although Stats SA already produces statistics that satisfy needs of a large and diverse group of the country's socio-political landscape, requirements for new or supplementary statistical products often arise. Requests for such new products may emanate from a number of sources, including government departments, private business and other stakeholders. When this happens, the first step in statistical production is to understand the need for the required statistics, i.e., what the required statistics are going to be used for in concrete terms by their users. Often for new projects, a lot of detailed information about what is needed is not at first clear and in some cases, not present, even from the perspective of the initiators of the project. It is therefore important to go through a process of refining the understanding the information (statistics) needs to be addressed by the results of the project.

2) **Design**

   The design phase consists of preparing ground for the execution of a statistical production project. This stage is reached when either a new project has been given the go ahead or a frequent on-going project is about to begin. In the case of on-going surveys, the design is usually in place already, but in a few cases it might need to be altered to cater for additional requirements or special circumstances.

3) **Build**

The build phase puts together all the pieces of the infrastructure for a statistical production project. These include the computer system, scanners, printing out of questionnaires, etc. The build phase also includes the procurement of the pieces of the infrastructure as necessary. The amount of work done in this phase also varies for new and on-going surveys.

4) **Collect**

Although the term collection may have different meanings in a statistical organization, it is used in the SVC to refer to both *direct* and *administrative* methods of data collection. The direct collection method refers to data collection in which Stats SA sources data directly from the respondents. In administrative collection, data are drawn from databases of other organizations which in turn source them from their respondents. It is important to note that this phase has a major influence on the activities of the design phase.

5) **Process**

The Process phase includes capturing collected data into databases so that data processing may be done. Data processing is necessitated by a number of issues. Chief among these is a fact that the data collection process is fraught with errors. The process phase is undergone to remove these data validity errors so as to improve data quality, and to package the data for use by analysis tools.

6) **Analyse**

After data have been cleaned during the Process phase, they are now ready for manipulation using analytical tools. This is the analysis done by domain experts to get insight into the meaning of the data. Further data quality enhancements may be done at this phase.

7) **Disseminate**

Stats SA collects data in order to produce statistics to be used by different stakeholders in the country including the general South African public. This means that the organization has to have ways of giving these communities access to the data and the resultant statistics. The Disseminate phase formalizes the steps Stats SA needs to go through in order to distribute information to the different communities as well as give them access to data repositories.

Stats SA uses a number of dissemination methods to ensure that the data produced by the organization is accessible to the widest user community. These include: electronic (e.g. via the internet), printed output and compact disks.

Table 1 below shows the mapping between Stats SA's survey cycle and the METIS cycle:

CMF	Stats SA
Survey planning and design	Need and Design Phases
Survey preparation	Part of Design Phase
Data collection	Collection Phase
Input processing	Processing Phase
Derivation, Estimation, Aggregation	Processing Phase
Analysis	Analysis Phase
Dissemination	Dissemination Phase
Post Survey Evaluation	

Table 1: CMF vs. Stats SA Business Processes

Post Survey Evaluation is currently done outside the statistical cycle. It is performed only for the large surveys such as the population census and the community survey.

Current Systems

The development of Statistics South Africa's metadata management system (Meta-Information system) is part of a larger system, the ESDMF. The central components of the ESDMF will follow the completion of the meta-information system, because the ESDMF is driven by the metadata. Although the ESDMF is a new system, it is merely a means to centralize the organisation's disparate statistical information systems. Figure 6 below shows the conceptual ESDMF subsystems and how they are placed relative to other organizational subsystems. The metadata subsystem supports the entire statistical cycle.

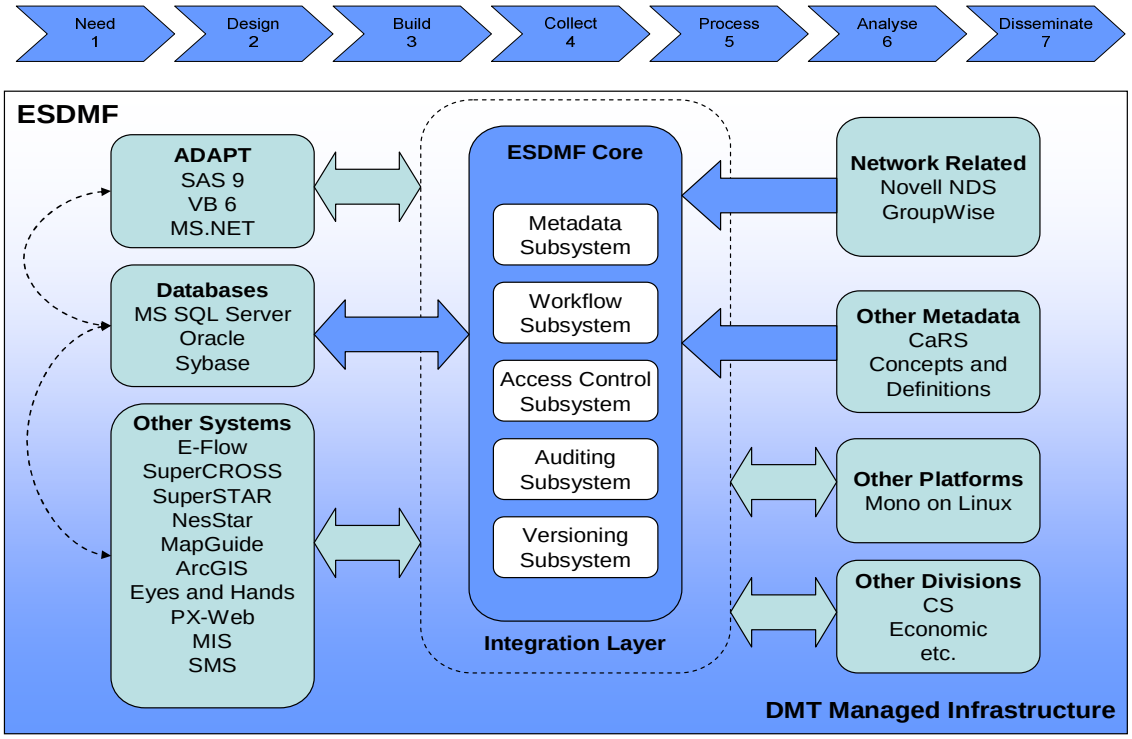


Figure 6: Conceptual components for the ESDMF in relation to other subsystems

Benefits and Costs of the Project

A fully-functional EDMF is defined as one that contains both administrative and survey data, and that is used for 'real time' transaction processing and post-processing analysis and dissemination. There are seven identified business outcomes that a DW containing both administrative and survey data will enable (the EDMF itself is not considered a business outcome). The business outcomes arise from the use of a EDMF in production processes and dissemination and the subsequent refinement of those processes. The main recipient of that business outcome is shown in brackets.

1. Substantial survey processing cost savings that are expected through lower technological support costs for development and maintenance of a smaller number of survey-specific data stores and associated applications. Small to moderate reductions in survey processing costs are expected to eventuate through the use of prior or existing knowledge about selected units in existing business processes, say for imputation, or confrontation during editing [economics, social and censuses].

2. Reduced provider load is expected to eventuate through better use of existing knowledge about a unit. This will also reduce the number of times a provider may be approached unnecessarily (economics).

3. Improved data quality is expected to eventuate through improved and automated confrontation procedures, in real time (national accounts/demographic analysis/ integrative analysis/quality and methodology).

4. Ongoing improvements to data integration beyond statistical units, and integration to data definitions, concepts and items at the data element level. (This is not to suggest that systems alone will resolve current problems. Greater discipline supported by better systems will be required).

5. Improved maintenance and refinement of data security through the use of technological solutions supporting business needs. Improvements in data management practices, policies and archiving are also expected. Note that better data matching and greater accessibility increase risk. There are policy/ culture/awareness issues that will drive the outcome of this. Technology alone is not the solution.

6. Improved ability to support new analytical outputs/products, through the availability of data in a single secure repository

7. Support for National Statistics System objectives through the availability of data from both administrative and survey sources especially around establishing a management system.

Implementation Strategy

Because of the complexity of the project, the project was planned to take place in phases with specific deliverables in each phase. The deliverables were facilities (application tools) to realize the ESDMF.

As it became evident that the approach was not effective, as the project was only in phase 2 out of 12 after a two thirds of the time allocated an alternative approach was then put in place for progress to be based on tool delivery versus phase approach.

The idea of delivering by tool was attractive in the sense that we could get quick wins by deploying each tool as it got completed (to the users). The challenge is to ensure that we choose the tools to deliver in such a way that future tools build up from earlier tools.

Currently, we are in the process of reviewing the project. Consequently, we have not progress much towards the goal of delivering by tool. On the otherhand, the last phase (Phase 2) was somehow a step in the direction as can be seen by the tools described in section 4.2

3. STATISTICAL METADATA IN EACH PHASE OF THE STATISTICAL BUSINESS PROCESS

The essence of Stats SA's meta-information system is captured by how the organisation uses the metadata. Metadata is used internal to the organisation to enable statistical production processes. This means that metadata is used during various stages of statistical production as essential input to production processes. However, the production processes in turn, produce metadata. This metadata is also important in documenting the trail of activities during the statistical production process. The documentation of production activities informs related metadata issues such as the assessment of data quality and its interpretation.

3.1 Metadata Classification (Categories of Metadata)

Because of this diversity of metadata usage, it was decided that contents of the meta-information system should be aligned with these usage activities. The natural progression of this decision was to undertake a project to classify all of the organisation's metadata. The following is a list of the categories of metadata adopted by Stats SA:

- **Survey Metadata**

  Often referred to as dataset metadata, Survey metadata is used to describe, access and update dataset, data structures. Stats SA chose to call this type of metadata *survey* rather than *dataset* because some of the metadata such as information about "the population which the data describe" refer to the broader aspects of the survey, and not only the dataset.

- **Definitional Metadata**

  This is metadata describing the concepts used in producing statistical data. These concepts are often encapsulated into measurement variables used to collect statistical data. Descriptive text is used to define individual concepts, however the concepts are further grouped into logical topics. These main topics are effectively classifications of data. Hence, included in Stats SA's package of definitional metadata classifications drawn from different study domains.

- **Methodological Metadata**

  These metadata relate to the procedures by which data are collected and processed. These may include Sampling, Collection methods, Editing processes, etc

- **System Metadata**

  System metadata refers to active metadata used to drive automated operations. Some of the examples of system metadata are:

  - Publication or dataset identifiers date of last update
  - File size
  - Mapping between logical names and physical names of files
  - Dataset input flows
  - Access methods to databases
  - Coordinates as kept in metadata store
  - Table and column definitions schema and mappings of data

- **Operational Metadata**

  This is metadata arising from and summarising the results of implementing the procedures. Examples include Respondent burden, Response rates, Edit failure rates, Costs and other quality and performance indicators, etc

The different components of Stats SA's meta-information system are logically grouped according to these categories of metadata. This means that the database for the meta-information system has different data structures corresponding to these metadata categories. We have recently (June 2007) finished developing the first metadata component, the survey metadata capturing tool, which is the subject of this case study.

3.2 Metadata Used/Created at Each Phase

The tables below show the phase of the business process (need, design, build, etc.) where the metadata created as well as where it is consumed. This is done for each key stage of the statistical production process.

Need

Metadata	Description	Detail /Doc / Generated / Linked	Phase created	Phase used
Survey request	User needs	Doc	Need Dissemination	Need Design Dissemination
Project requirements	E.g. Usage of data, quality requirements	Doc	Need	Need Design

Planning

Metadata	Description	Detail / Doc / Generated / Linked	Phase created	Phase used
Strategic plan		Doc	Design	All
Operational plan		Doc	Design	All
Business plan		Doc	Design	All
Financial plan		Doc	Design	All
Risk management plan		Doc	Design	All
Quality plan		Doc	Design	All
Communication & change management plan		Doc	Design	All
Publicity plan		Doc	Design	All
Evaluation plan		Doc	Design	All
Capacity development plan		Doc	Design	All
Fieldwork plan		Doc	Design	Collect
Marketing plan		Doc	Design	All

Design

Metadata	Description	Detail / Doc / Generated / Linked	Phase created	Phase used
Scope	Scope and coverage definition	Doc	Design	Design
Objectives	What is the study wishing to achieve	Doc	Design	Design Dissemination
Project Constraints	The limitations of the project	Doc	Design	Design
Tabulation plan	How the data should be disseminated	Detail/doc	Design	Design Analysis Dissemination
Dissemination plan	How the data should be disseminated	Doc	Design	Design Analysis Dissemination
Project methodologies	This will be a collection of other methodologies e.g. collection, sampling, processing etc	Doc	Design	All
Quality measures/metrics		Doc	Design	Design Dissemination

Workflow Processes

Metadata	Description	Detail / Doc / Generated / Linked	Phase created	Phase used
Planned process	Planned process as designed in Workflow	Generated from Workflow	Design	All
Process audit trail	Audit trail of by whom when and how the planned process was executed	Generated from Workflow	All	All
Task specific metadata	Specific metadata for general tasks e.g. approval, meetings, workshops (E.g. Minutes of meeting, Workshop reports, Approval / rejection comments)	Doc	All	All

Questionnaire/Collection Instrument

Metadata	Description	Detail / Doc / Generated / linked	Phase created	Phase used
Project Concept definitions	Definitions of concepts addressed by one or more questions.		Design	All
Project question modules	Questions selected from organisational list to be used in questionnaire. This will include the sequence of questions as well as the question detail (Name, definition, question text, answer format, value sets, answer validations, classification version)	Detail Link to CaRS	Design	Design Build Collect Dissemination
Questionnaire	Questionnaire / collection instrument	Doc	Design	Design Build Collect Dissemination
Observational study report	After questionnaire testing	Doc	Design	Design
Interviewer instructions		Doc	Design	Collection

Sampling & Basket

Metadata	Description	Detail / Doc / Generated / Linked	Phase created	Phase used
Sample requirement / specification	Information about how sample must be drawn e.g. strata, number in each stratum etc	Doc	Design	Design
Sampling method	Description of sampling methodology to follow	Doc	Design	Design Dissemination
Sampling report	Information about how sample was drawn	Doc	Design	Design
Sample evaluation report		Doc	Design	Design
Sampling maintenance documents		Doc	Design	Design
Sample frame information	Description of Sample frame, version of sample frame etc.	Doc	Design	Design Build
Sample Statistical units	Info on the units (organisations/households) in the sample	Generated from IDW	Design	Design Collect Process Analysis
Response rates		Generated from IDW	Collect	Collect Process Analysis Dissemination
Sample detail	Detail on sample including sample size, distribution etc)	Detail or Doc	Design	Design Collect Process Analysis
Sample weights	Weighting sets	Detail	Design	Design Process Analysis
Weighting methodology		Doc	Design	Design Process Analysis Dissemination
Basket description	A description of the basket used	Doc	Design	Design Process Analysis Dissemination

Tool Building

Metadata	Description	Detail / Doc / Generated / Linked	Phase created	Phase used
Capturing / editing tool design	Design spec for capturing / editing tool	Doc	Design	Design Build
Capturing / editing tool description	Description of capturing / editing tool (Collection instrument)	Doc	Build	Build Collect Process Analysis
Batch loading process design	Design spec for batch loading process	Doc	Design	Design Build
Batch loading process description	Description of batch loading process (Collection instrument)	Doc	Build	Build Collect Process Analysis
Project data elements	Data elements included in the capturing/editing tool or batch loading process. Object classes, properties, data elements concepts, value domains (conceptual domains and representations))	Detail (Generated from project questions)	Design	Build Collect Process Analysis
Project data element validations	Data elements included in the capturing/editing tool or batch loading process. Range validations, dependency validations	Detail (Generated from project questions)	Design	Build Collect Process Analysis
Project data element classifications	Data elements included in the capturing/editing tool or batch loading process. Classification used for question including version	Detail (Generated from project questions)	Design	Build Collect Process Analysis
IDW Dataset design	Data fields in IDW dataset	Detail (Generated from project questions/project data elements)	Design	Design Build Collect Process Analysis Dissemination
IDW Data store creation script			Design	Design Build
ODW Dataset design	Data fields in IDW dataset	Detail (Generated from project questions/project data elements)	Design	Design Build Analysis Dissemination
ODW Data store creation script			Design	Design Build Analysis

Collection

Metadata	Description	Detail / Doc / Generated / Linked	Phase created	Phase used
Project Collection methodology	Collection technique to be used in project	Doc	Design	Design Build Collect Process Analysis Dissemination
Collection training manuals		Doc	Design	Design Collection
Fieldwork manuals		Doc	Design	Design Collection
Collection manuals		Doc	Design	Design Collection
Collection progress reports		Doc	Collection	Collection
Fieldwork / Collection survey reports	Fieldwork / collection report	Doc	Collection	Collection Process Analysis Dissemination
Fieldwork/ collection evaluation report		Doc	Collection	Collection Process Analysis Dissemination

Processing

Metadata	Description	Detail / Doc / Generated / Linked	Phase created	Phase used
Processing manual	Describing processing to be performed in project		Design	Design Process Analysis Dissemination
Processing report	Describing processing performed in project	Doc / Generated from workflow	Process	Process Analysis Dissemination
Processing evaluation report	Evaluation report done on processing performed in project	Doc	Process	Process Analysis Dissemination
Imputation methods /specification	Description of imputation to be performed in project	Doc	Design	Process Analysis Dissemination
Imputation rules	To be used in data transformation process creation	Detail	Design	Process Analysis Dissemination
Estimation methods	Description of estimation to be performed in project	Doc	Design	Process Analysis

Metadata	Description	Detail / Doc / Generated / Linked	Phase created	Phase used
				Dissemination
Estimation rules	To be used in data transformation process creation	Detail	Design	Process Analysis Dissemination
Derivation methods	Description of methods to derive a values or datasets in project	Doc / Generated from workflow	Design	Process Analysis Dissemination
Derivation rules	To be used in data transformation process creation	Detail	Design	Process Analysis Dissemination
Coding quality information	Description of coding methods to be used in project	Doc	Design	Process Analysis Dissemination
Coding rules	To be used in data transformation process creation	Detail	Design	Process Analysis Dissemination
Editing instructions	Description of editing rules to be followed in project	Doc	Design	Process Analysis Dissemination
Edit rules	Default values Editing rules Ranges Classification	Detail (Generated from project data element detail)	Design	Process Analysis Dissemination

Analysis

Metadata	Description	Detail /Doc / Generated / linked	Phase created	Phase used
Detail analytical report	Description of analysis done on project	Doc	Analysis	Analysis Dissemination
Quality audit report	Report on error identified in release	Doc	Analysis	Analysis Dissemination
Data confrontation report		Doc	Analysis	Analysis Dissemination
Technical notes	Notes on values in a dataset for release.	Detail	Analysis	Analysis Dissemination
Explanatory notes	Notes on dataset for release (e.g. Reasons for discrepancies)	Detail	Analysis	Analysis Dissemination
Clearance notes	Notes to assist with approval sign-off for dissemination	Doc	Analysis	Analysis Dissemination

Dissemination

Metadata	Description	Detail / Doc / Generated / Linked	Phase created	Phase used
Production schedule		Doc	Design	Design Dissemination
Release and various approval sign-offs	Describing sign-off process for release	Generated from workflow	Process Analysis Dissemination	Process Analysis Dissemination
Statistical release information	Description of release (including info to enable a person to find the release; description of possible use) (Release can be a publication, a dataset or an index value)	Doc	Design	Dissemination
Publication design	Design of publication to be disseminated (based on tabulation plan & standard product templates)	Detail	Design Dissemination	Dissemination
Publication dataset design	Design of dataset to be disseminated (including format e.g. Excel, coma delimited etc; field descriptions; structures; footnotes; titles etc)	Detail	Design Dissemination	Dissemination
Press release	The press release to accompany a specific product	Doc	Dissemination	Dissemination
Automatic dissemination audit trail	A description of customers a specific product was disseminated to automatically	Generated	Dissemination	Dissemination

4. SYSTEMS AND DESIGN ISSUES

The design of the system will conform to Stats SA's Enterprise architecture. One of the main components of this enterprise architecture is the IT architecture. The software architecture of all the applications developed in the ESDMF is dictated by the IT architecture as document in the Stats SA ICT strategy..

4.1 Overview of Stats SA's IT Architecture

The Stats SA's IT environment, within which the ESDMF is developed, requires systems to adhere to the following architectural principles:

- **Integration**

The system must integrate with other organizational systems. API's will be built for various applications that need to connect to the ESDMF. However, most of the connection is expected to be at a data level. With the exception of SAS, the odularized uses relational databases. Integration at this level is attained using ODBC connection. SAS supports ODBC and in addition to that, has native support for various databases.

- **Interoperability**

To ensure interoperability, the ESDMF uses Java as a development standard because of its platform independence. The development of the system as a web application also means that only a web browser is needed to access the application.

- **Modularity**

The development of all the components of the ESDMF is based on the odularized al requirement for building modular systems that allow ease of management and flexibility. The metadata management system is odularized according to the different categories of metadata.

- **Scalability**

Stats SA's computer applications have to be built such that they can scale up to accommodate the inevitability of growth of an organization. Both the database designs and storage hardware for all the components of the ESDMF are developed to cater for such growth.

- **Flexibility**

Applications must meet the diverse needs of Stats SA. These needs change with time, and new ones are also discovered. Development of flexible applications that may be easily changed or added to is vital. Part of the insistence on the use of object oriented programming was informed by the need for flexibility. This will minimize "spaghetti programming" associated with large software projects.

4.1.1 IT Infrastructure Specification

The metadata management system is deployed in an IT infrastructure with a set of minimum specifications. These minimum specifications list the hardware items needed to run the system without going into details of the hardware items themselves.

- **Operating System(s)**

Desktops are in Microsoft Windows. The application is deployed in an Open Source operating system (Novell SuSe Linux).

- **Computer Network**

The network architecture is based on open protocols and industry standards. It allows remote access to some employees. This supports both local area (LAN) and wide area (WAN) networks.

- **Computer Servers**

The system is developed as a client-server application. This means that there is a need for powerful computer servers capable of handling intensive processing.

- **Storage**

Because of the vastness of data to be generated and/or captured in the system, there is need for a well-managed storage system. The Storage Area Network (SAN) is the technology used at Stats SA to provide storage management.

A. Development Environment			
Function	**Make/Model**	**Operating System/ Database Engine**	**Comment**
Application Server	HP BL45p Quad processor 4 GB RAM 2 x 72 GB HDD	SuSe Linux Ver. 10	Make/Model exceeds recommendation
Database Server	HP BL45p Quad processor 16 GB RAM 2 x 72 GB HDD	Oracle 10g *or* Sybase ASE and Sybase IQ Unix/Linux/Windows	Make/Model exceeds recommendation
Build Server	HP DL 320 Dual processor 2 GB RAM 2 x 72 GB HDD	SuSe Linux Ver. 10	Make/Model exceeds recommendation
B. User Acceptance Test (UAT) Environment			
Application Servers	2 x HP BL45p Quad processor 8 GB Ram 2 x 72 GB HDD	SuSe Linux Ver. 10	Make and model exceeds recommendation
Database Servers	2 x HP BL45p Quad processor 32 GB Ram 2 x 72 GB HDD	Oracle 10g *or* Sybase ASE and Sybase IQ Linux	
C. Production Environment			
Application Servers	2 x HP BL45p Quad processor 8 GB Ram 2 x 72 GB HDD	SuSe Linux Ver. 10	Make and model exceeds recommendation
Database Servers	2 x HP BL45p Quad processor 32 GB Ram 2 x 72 GB HDD	Oracle 10g *or* Sybase ASE and Sybase IQ Linux	

Table 2: Hardware and software specifications for the ESDMF infrastructure

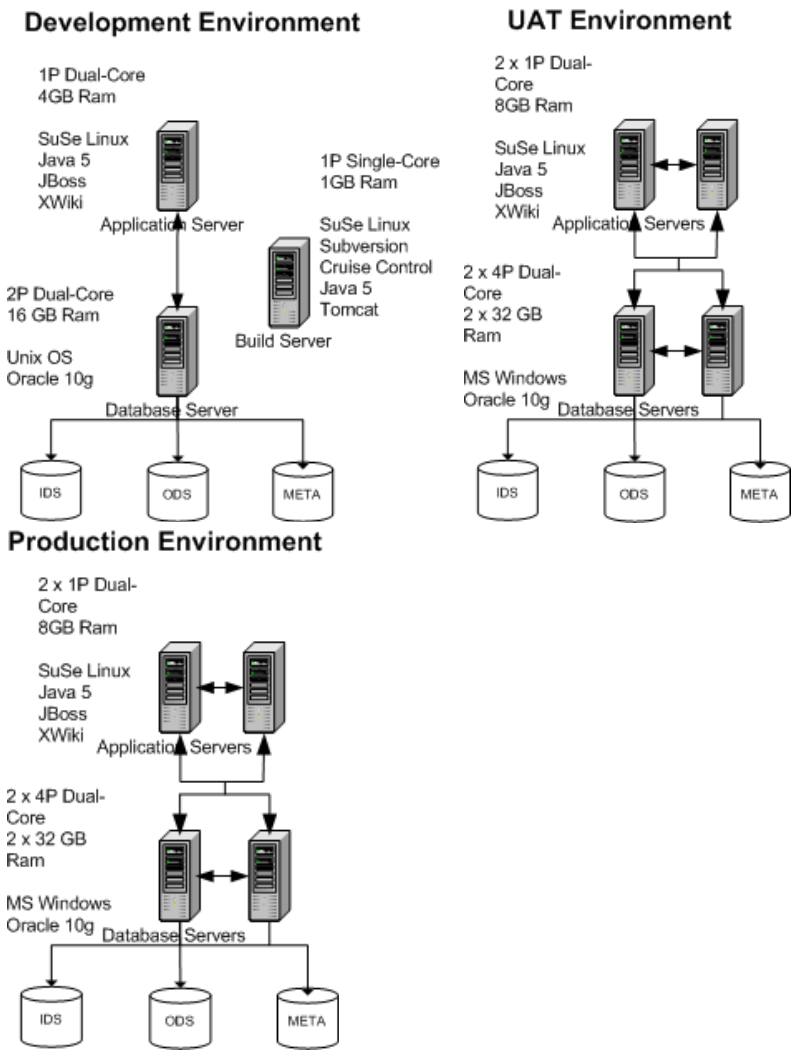


Figure 7: Hardware and software specifications for the ESDMF infrastructure

4.1.2 Components of Metadata Management Application

The application is web-based and developed in Java. Tomcat is used to implement Java Servlet API and HTTP functionality. The following are physical divisions of the application:

- **User Interface (UI)**

The user interfaces for all the metadata management system applications is web-based. This allows us to quickly deploy the tool to users in the organization. Client workstations only need to have a web-browser to access server based applications. The main supported web-browsers are Microsoft Internet Explorer and Firefox.

- **Database**

The application is supported by a relational database management system (RDBMS). Stats SA uses a variety of RDBMS engines. The RDBMS engine of choice for this project is Sybase 12.5.x. The project is currently using the open source RDBMS, MySQL.

- **Business logic**

The business logic controlling the interaction between the UI and the underlying database is coded using Java server side scripting. There is also business logic coded using stored procedures. This mostly performs housekeeping within the database.

- **Application/Web Server**

The application is served to the client via Tomcat, which processes Java code. Tomcat also handles HTTP calls from the web browser.

4.2 Metadata Management Tools

The developed metadata management application allows Stats SA staff members to perform a number of tasks in the metadata management process. Currently, the End to End Statistical Data Management Facility (ESDMF) consists of the following **tools**:

1. The **Access Control Tool** provides a central point for creating and managing access to all ESDMF tools.

2. The **Metadata Registration Tool** provides a central registry for the registration, revision and relating of administered items.

3. The **Metadata Browser** enables users to browse and search for administered items in the central registry.

4. The **Quality Tool** enables users to setup the quality framework that is based on the SASQAF document, and forms the foundation for the assignment of Quality Indicators to Surveys, the capturing of Quality Metadata for a Survey and conduct Quality Assessment of a Survey.

5. The **ESDMF Reporting Tool** allows users to generate static and **ad hoc reports** based on the data contained in the various ESDMF tools.

6. The **Survey Metadata Capture Tool** manages the metadata for a survey on series and instance level.

The following diagram shows a high level view of the above-mentioned tools and their inter-relationships

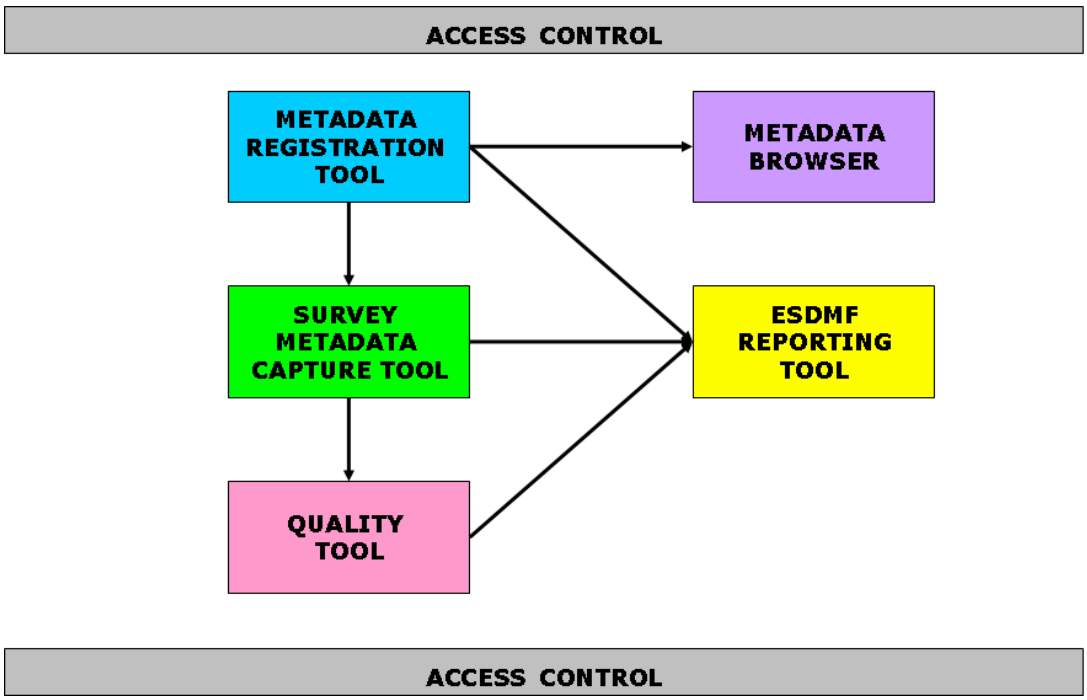


Figure 8: High level view of the relationship between the various ESDMF tools and applications

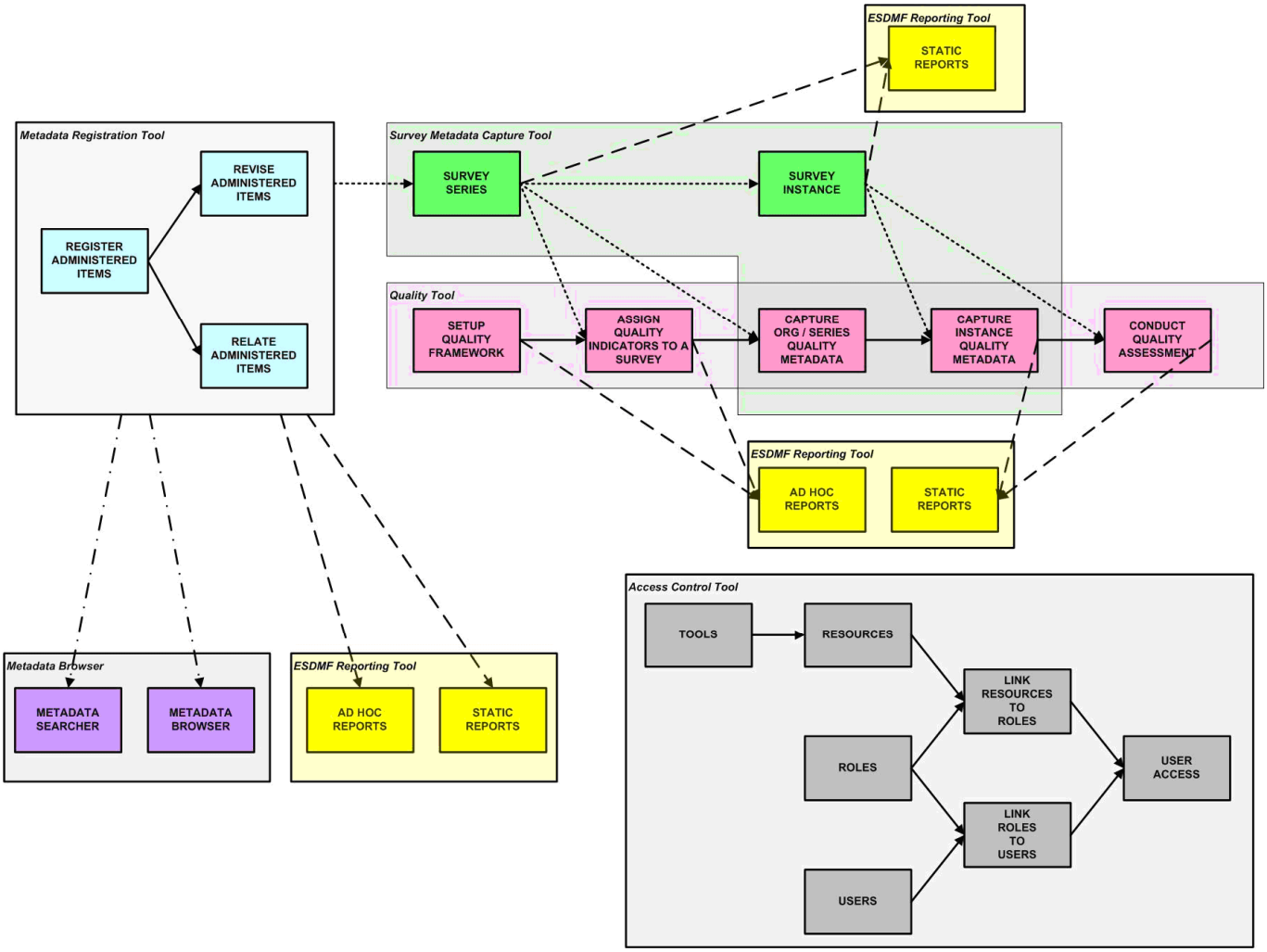


Figure 9: Detailed view of the relationship between the various ESDMF tools and applications

4.2.1 Access Control Tool

The **Access Control Tool** provides a central point for creating and managing access to all ESDMF tools. Tools are linked to resources that can be linked to roles and roles can be linked to users. This association establishes highly customisable user access privileges that are encapsulated in a single sign-on process. This allows a user to sign on once and then access any tool the user has privilege to.

Term	Definition
Tool	A piece of software designed to fulfill a specific purpose in the ESDMF.
Resource	The url for any page in a tool to which access should be granted.
Role	A designated function within a specific tool that requires the user assigned this role to have access to specific functionality.
User	Any individual who exists within the security context of the ESDMF.

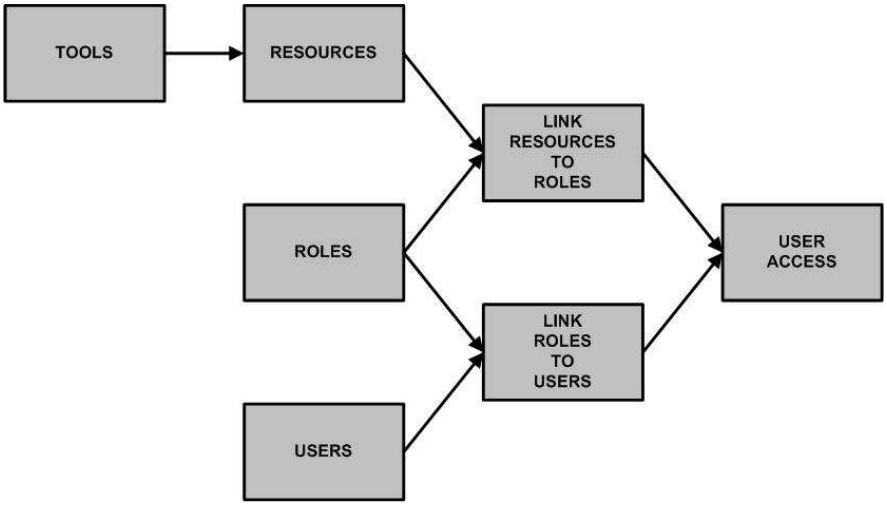


Figure 10: Access Control Tool modules

The following is a screen capture of login within the Access Control Tool

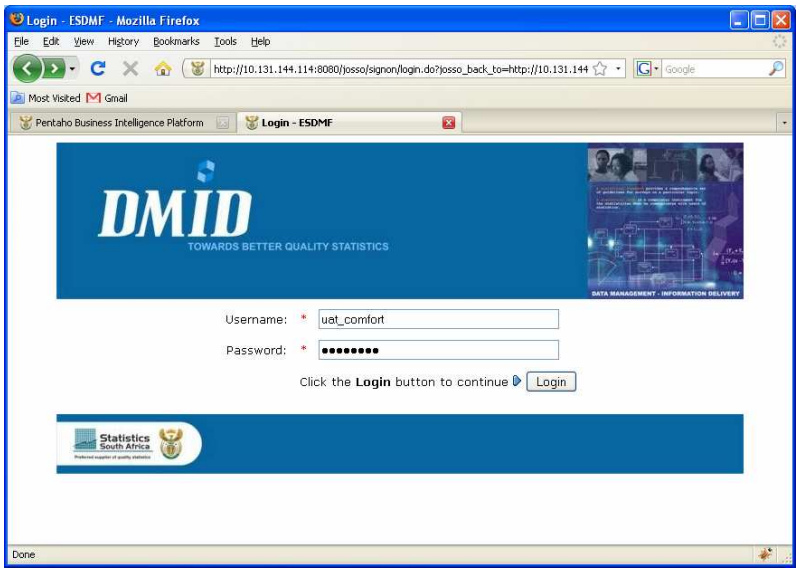


Figure 11: Access Control login screen

4.2.2 Metadata Registration Tool

A modified ISO 11179 registration process is used for registering administered items. Administered items currently include data element concepts, data elements and classifications. The registration process addresses the improvement and progression of administered items towards levels of perfection of the quality of the metadata and the preferences of usage. The progression status include incomplete, candidate, recorded, qualified, standard, preferred standard, superseded and retired. Statuses are grouped together according to usage and visibility.

The relationships between these administered items are as follows: A Data Element Concept has to be registered and must have reached the status of qualified before a relating Data Element can be registered. A Classification is related to a Data Element via the Value Domain. This result of this process is the unified view of administered items.

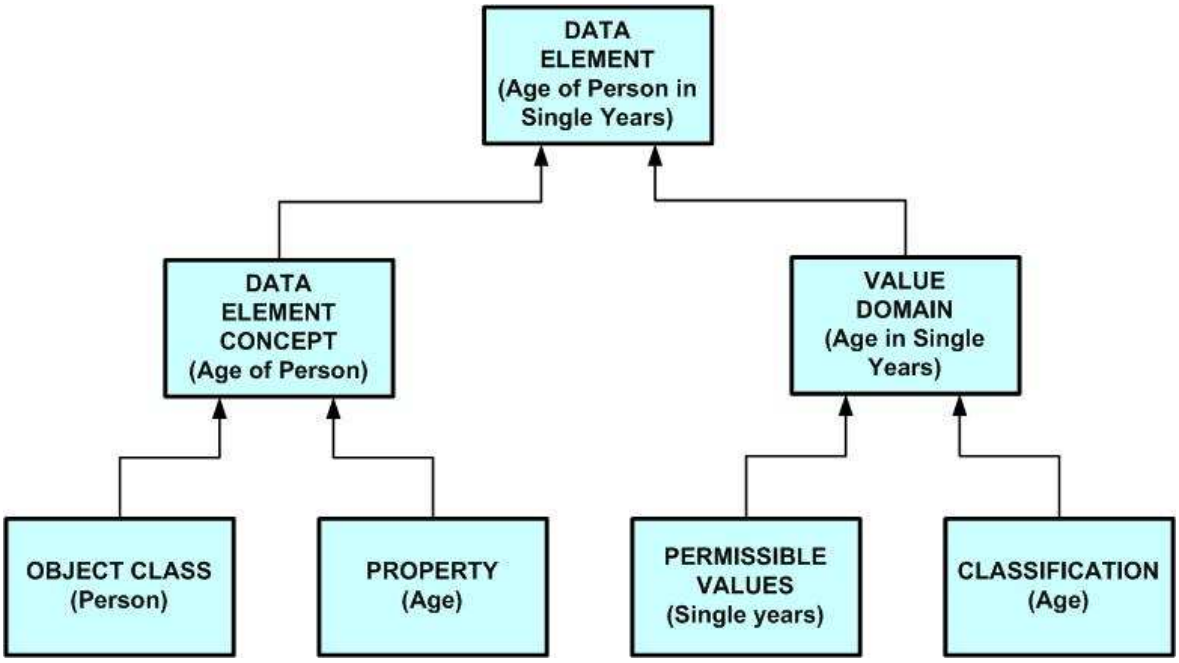


Figure 12: Relationship between Administered Items

Administered Item	Definition	ISO 11179 Definition	Example
Data Element	Basic unit of identifiable and definable information	Unit of data for which the definition, identification, representation and permissible values are specified by means of a set of attributes	Age of Person in Single Years
Data Element Concept	Represents the concept for which data is sought	Concept that can be represented in the form of a data element, described independently of any particular representation	Age of Person
Value Domain	Set of valid values for a data element	A set of attributes describing representational characteristics of instance data with or without enumerated permissible values.	Age in Single Years
Property	Describes an Object Class	A characteristic common to all members of an Object Class	Age
Object Class	Represents a person, organisation, structure or event that is of interest and needs to be describe in other words the thing we collect data about	A set of ideas, abstractions, or things in the real world that are identified with explicit boundaries and meaning and whose properties and behaviour follow the same rules	Person
Classification		A set of discrete, exhaustive and mutually exclusive observations, which can be assigned to one or more variables to be measured in the collation and/or presentation of data.	Age
Conceptual Domain	Description of valid values	Set of possible value meanings of a data element expressed without representation	Ages
Permissible Values	Valid values in a specific value domain	An expression of a value meaning allowed in a specific value domain	Real Numbers 1…120
Administered Item		Registry item for which administrative information is recorded in an administration record	Data Element Concept; Data Element; Classification; Property; Object Class; Conceptual Domain; Value Domain; Permissible Values

Role	Definition
Submitter	Submits an administered i**tem** and provides the minimum information required for registration.
Steward	Responsible for the accuracy, reliability, and currency of descriptive metadata of the Administered Items.
Registrar	Expert in the registration process, responsible for facilitating the registration of Administered Items and making those Administered Items widely accessible and available to the community.

Status	Definition	Visible	Usable
Incomplete	Submitter wishes to make Stats SA community aware of the existence of an administered item	No	No
Candidate	Administered Item is proposed for progression up the registry registration levels	No	No
Recorded	All mandatory metadata attributes for the administered item have been completed	No	No
Qualified	Mandatory metadata attributes for the administered item are complete and conform to applicable quality requirements	Yes	Yes
Standard	Administered item is of sufficient quality and of broad interest for use in Stats SA	Yes	Yes
Preferred Standard	Administered item is preferred for use in Stats SA	Yes	Yes
Superseded	Administered item is no longer recommended for use but can be used under certain predefined circumstances	Yes	No
Retired	Administered item is no longer recommended for use in the Stats SA community and should no longer be used	Yes	No

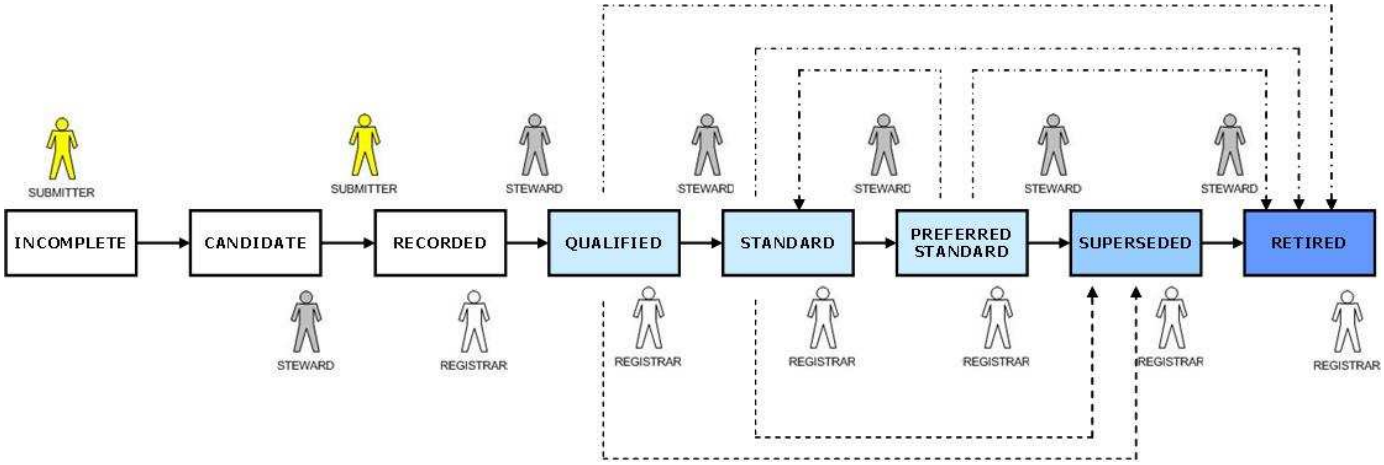


Figure 13 : Metadata Registration Process

4.2.3 Metadata Browser Tool

The **Metadata Browser** enables users to **browse** and **search** the **metadata repository** for visible administered items.

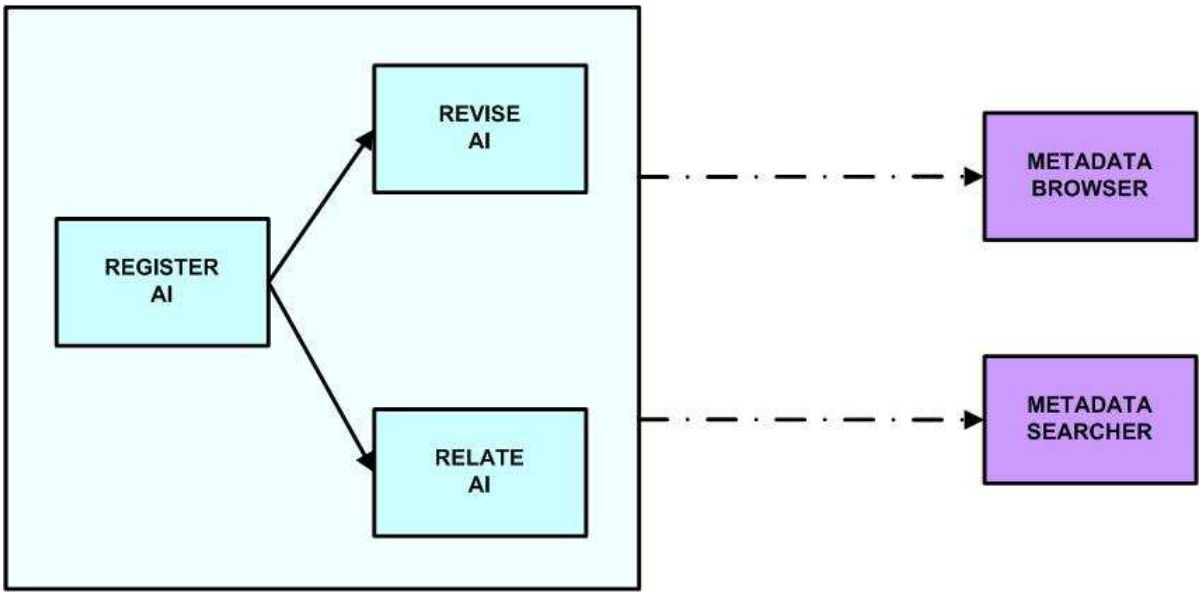


Figure 14: Metadata Browser modules

Figure 15: Metadata Browser main screen

4.2.4 Quality Tool

The SASQAF document forms the foundation for the Quality Tool. Therefore any changes made to the quality framework setup will immediately reflect in the rest of the tool.

A survey area in conjunction with Methodology & Evaluation will decide on which quality indicators a survey will be assessed on. These Quality indicators are then assigned to a survey based on this agreement. Quality indicators can only be assigned to surveys that have been created on the Survey Metadata Capture Tool.

Indicators can be grouped as follows:

- Exempted
  - Provisionally exempted indicators are temporary exempted indicators with compliance criteria and date specified in the statement of intent.
  - Permanently exempted indicators are indicators on which no reporting will be done.
- Not Exempted
  - Series / Organisational Indicators are applicable to the organisation or series level with generic quality metadata that will only be entered once.
  - Instance indicators are only applicable to an instance and quality metadata must be created for each instance created in the Survey Metadata Capture Tool.

The combination of the series / organisational and instance quality metadata statements will produce the quality declaration for the survey instance. The quality declaration then forms the basis on which the quality assessment is done.

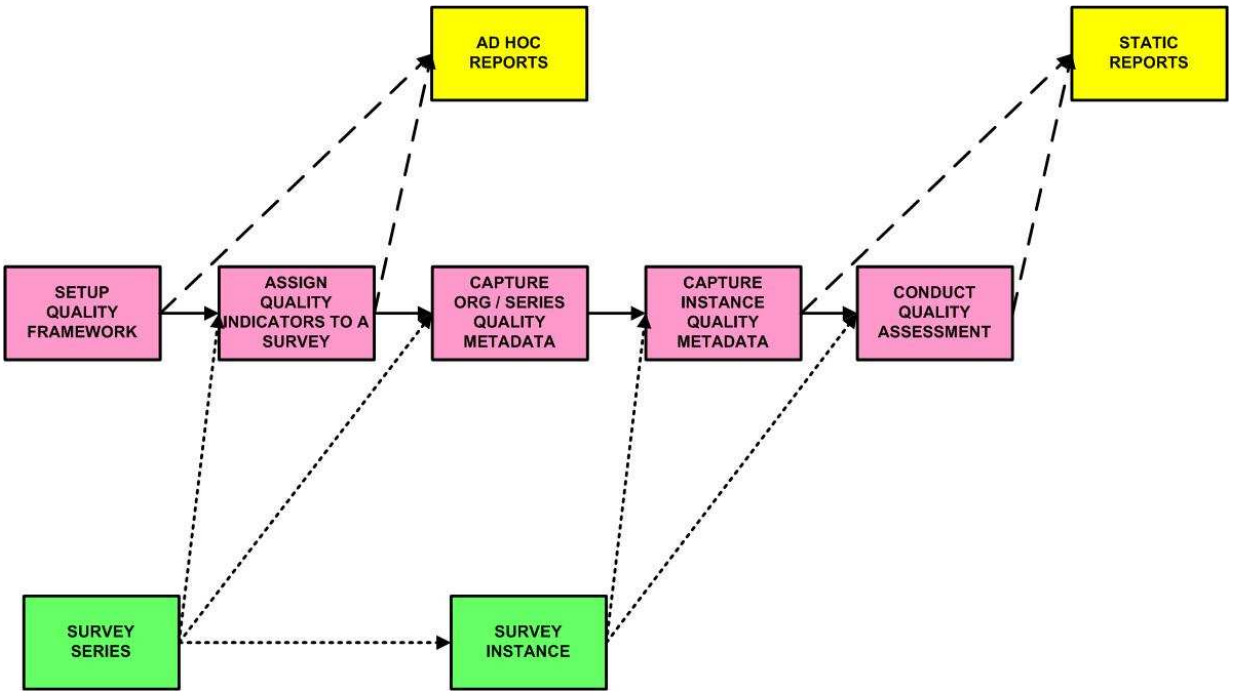


Figure 16: Quality Tool modules and interface to other tools

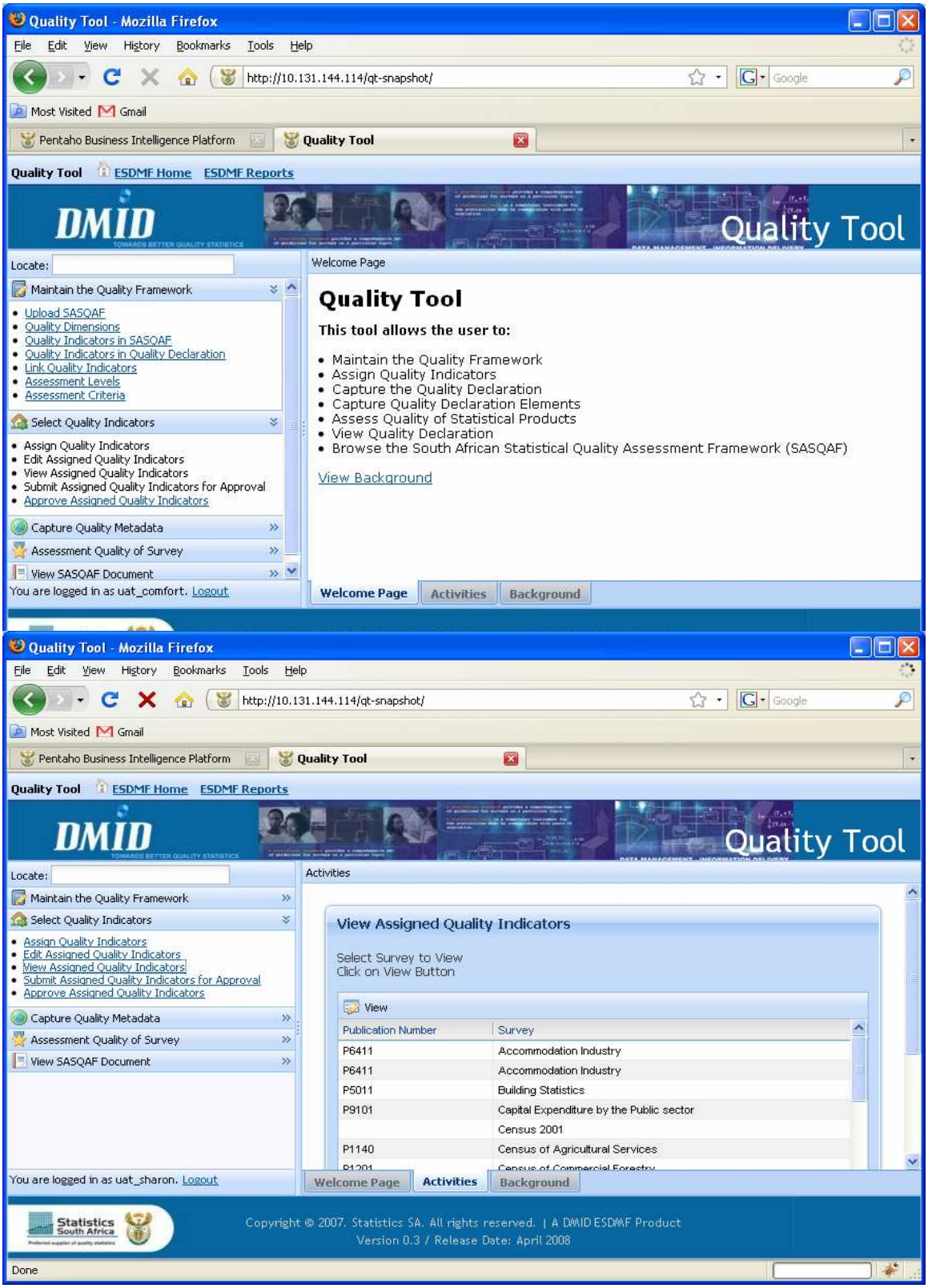


Figure 18: Surveys with assigned quality indicators

4.2.5 ESDMF Reporting Tool

The ESDMF reporting tool allows a user to generate ad hoc or static reports based on data from an ESDMF tool. Static reports have predefined report layouts and ad hoc reports have user defined report layouts.

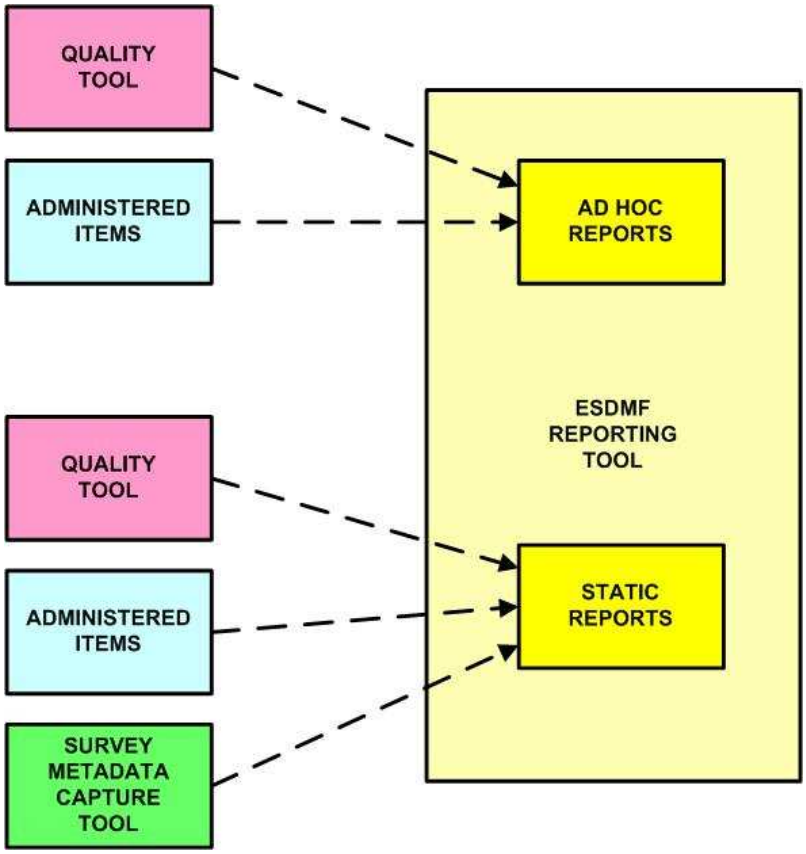


Figure 19: Reporting Tool modules and interfaces to other tools

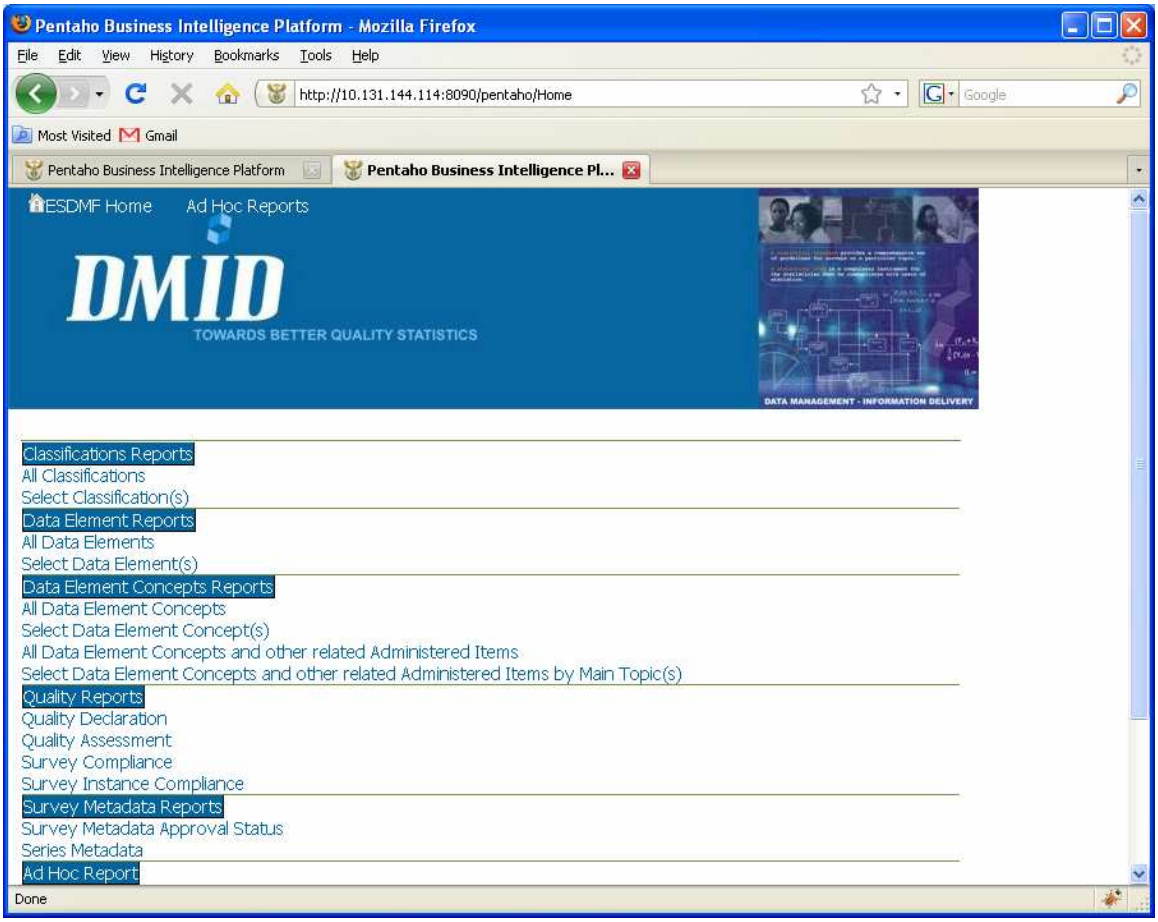


Figure 20: Reporting Tool main screen

4.2.6 Survey Metadata Capture Tool

The Survey Metadata Capture Tool manages the metadata for a survey on series and instance level.

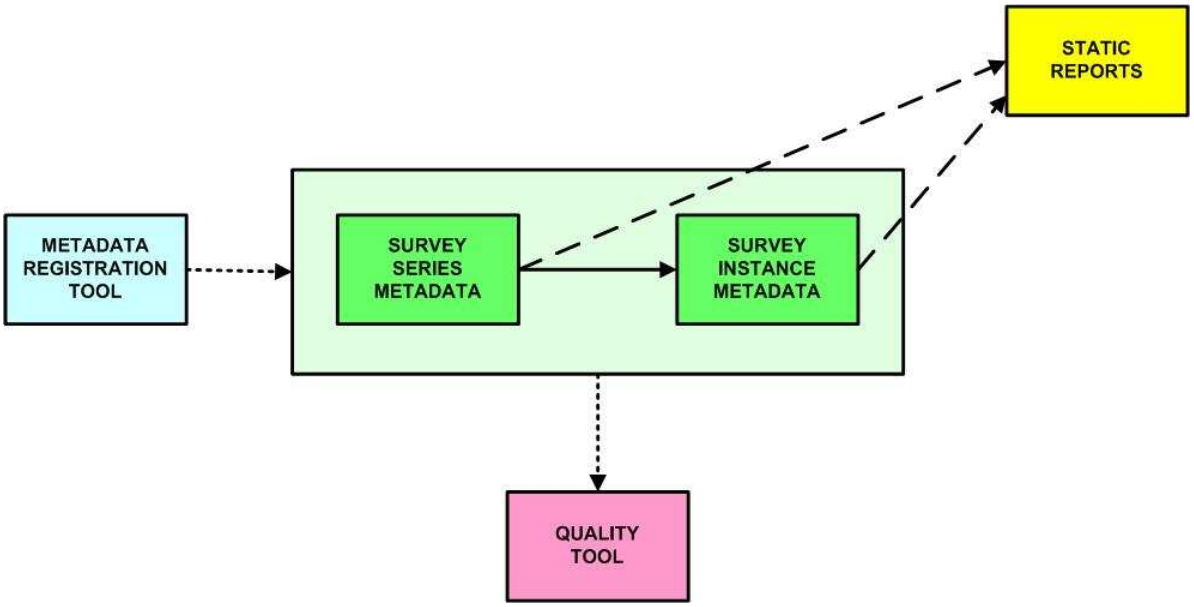


Figure 21: Survey Metadata Capture Tool and interfaces to other tools

Figure 22: Survey Metadata Capture Tool main screen

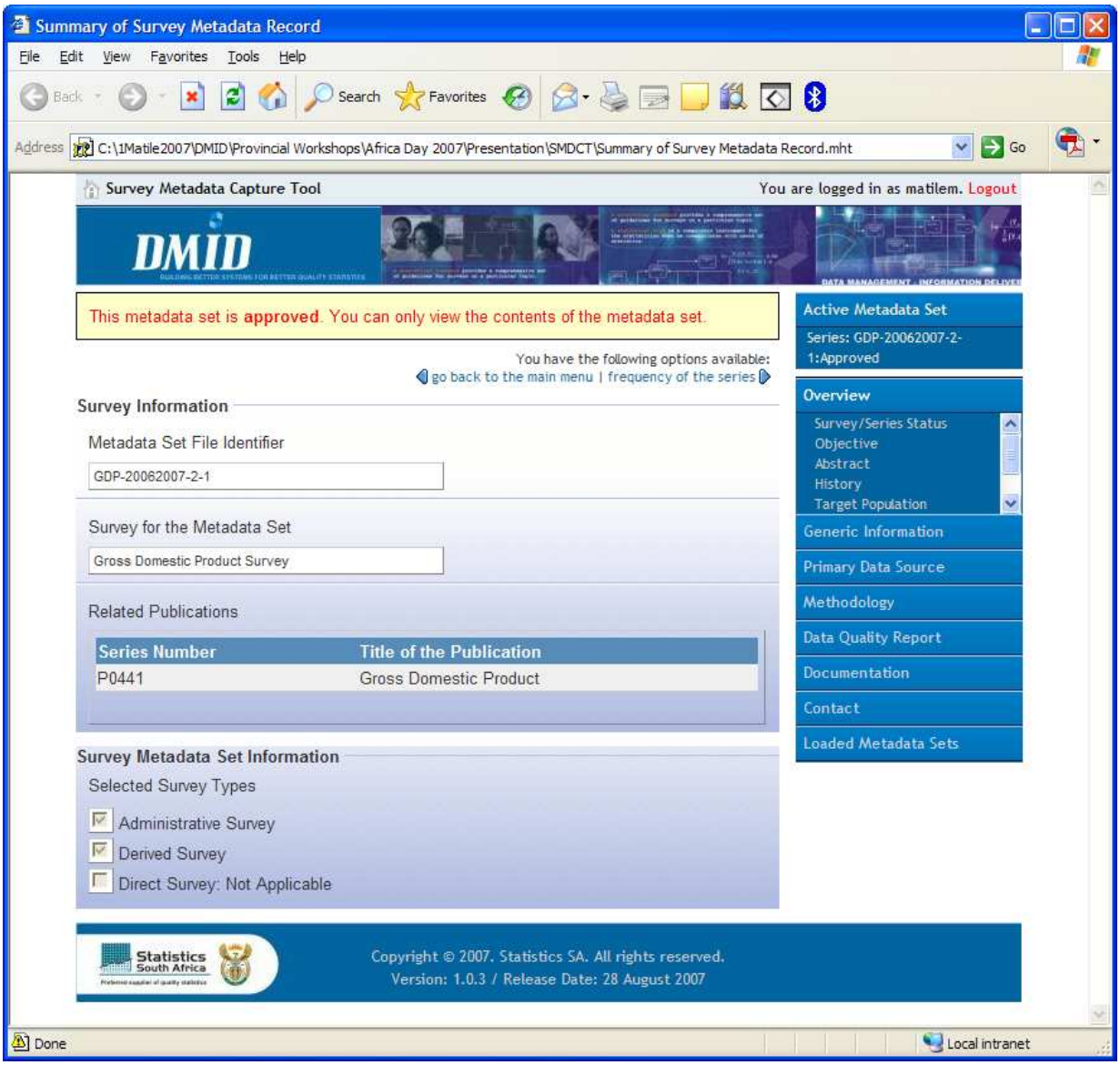


Figure 23: Survey Metadata Capture Tool Survey information page with navigation on the right hand side

4.3 Standards and Formats

- **Metadata Registration Tool**

The Metadata Registration Tool is based on the ISO 11179 Part 6 standard, with some customization for Stats SA purposes.

- **Quality Tool**

The Quality Tool is based on the South African Quality Assessment Framework (SASQAF) standard, which is based on the IMF Data Quality Framework.

4.4 Metadata Revisions and Version Control

Metadata is expected to change due to revisions of concepts and their definitions, changes to classifications, business rules and user requirements. Sometimes more than one version of certain metadata used for the same purpose may exist at the same time.

In the current Survey Metadata tool the "Edit" functionality of the application allows for the revision of captured Survey metadata. These revisions may only be performed by users with requisite permissions. For changes to be effected, revised/edited metadata must be approved by an assigned *Approver*. Survey metadata can only have a single version. This means that the *Edit* process serves to update the metadata repository.

Version control will be introduced when metadata categories with metadata that can have more than one version are incrementally built into the system.

It is important to note that version control will be built into every aspect of the ESDMF.

4.5 Outsourced vs. In-house Development

The development of all of the ESDMF, including the metadata management system, is outsourced. Two issues influenced the decision to outsource. These were: the fact that Stats SA does not have enough skilled resources and the need to have views which would not be obscured by prior opinions of a statistical environment. This scenario requires that the outsourced resources invest a lot of time in understanding the organisation and analysing the requirements.

It is important to note that we conducted two stages of outsourcing. In the first stage we outsourced the task of gathering the requirements for the whole of the ESDMF. These requirements contain details of each of the components of the ESDMF, including the metadata management system. The second stage is the development of the system. The two tasks were done by two different organisations. This separation of tasks was done in order to maintain the focus on requirements gathering. In this development model, the development team mainly verifies existing requirements.

5. ORGANIZATIONAL AND CULTURAL ISSUES

5.1 Roles in metadata/statistical lifecycle management

In order to understand the user requirements, we engaged the survey divisions as pilot groups. We involved them in verifying our understanding of the requirements, which was used to design and implement the system. These pilot groups were also involved during User Acceptance Testing (UAT).

The Survey Metadata Capture Tool can be used by different users depending on the roles that they were assigned. For example, a Capturer could capture metadata but this must be approved by an Approver, who is usually the supervisor or manager. There is

also a role of viewer, whereby metadata could be viewed but the rights are restricted. For example, a viewer cannot edit, change or approve metadata.

The network infrastructure for both development and user environments is supported by the IT department. This includes configuring the environments as well as housing the different servers in the data centre of the organization. The databases are also managed by the IT department. The ESDMF is based on the Linux open source operating system. Because the IT department does not have the skills to service and maintain this environment, we have outsourced these services from a private company. However, this is done in conjunction with the IT department, who are in the process of raising their skill level in order to be able to support the ESDMF in the Linux environment.

During User Acceptance Testing (UAT) any identified defects were logged on the CA Unicentre system, which is used for IT help desk support. With the help of the IT help desk technicians, we were able to customise the system so that the unique categories of defects for the ESDMF system could be recorded.

The IT procurement group was used to procure all the hardware and software used in the development and deployment of the system.

The development of the ESDMF was not done in isolation of the existing projects within Stats SA. For example, the following projects were ongoing and in parallel with the development of the ESDMF:

- SAS 9 migration
- Re-engineering of other surveys
- Community Survey 2007
- Census 2011

Some members of these other projects were also involved in the development of the requirements and review of the architecture of the ESDMF. The goal is to ensure that we do not do things in isolation so that we can share our knowledge and ease the integration of the new system into existing systems.

Staff from the Methodology and Standards division was seconded to the ESDMF project. Their role was to develop policies, procedures and standards for the system. Our development process is that policies are developed and approved. Thereafter, the procedures and standards are developed. So, for each phase, the policies are used to develop and implement the system deliverables for that phase.

For example, for the first phase, we developed a policy for Data Quality and a policy for Metadata. As a result, Phase One was focused on capturing metadata (Metadata policy) in order to ensure quality of the output product (Data Quality policy). For the Second Phase, we already have approved policies for Concepts and Definitions as well as for Classifications.

5.2 Description Of The Team/Individuals Involved In Development And Maintenance Of Metainformation Systems.

5.2.1. System Developers

The deliverables expected from the supplier include a Skills Transfer Plan and Strategy. The goal is that the supplier will train Stats SA system developers in how the system is designed and implemented. At the end of the contract, these Stats SA developers should be knowledgeable to maintain, upgrade and/or enhance the system. Thus, we should not be dependent on the supplier for any development beyond the expiry of the contract.

5.2.2. Data Quality Officers and Specialists

The Data Quality Officers and Specialists are trained on how to use the system. They are also trained to be trainers ("train the trainer"). Once again, the supplier's deliverables includes training Stats SA Data Quality Officers and Specialists in how to train users on how to train other users to be trainers themselves.

5.2.3. Methodology and Standards Professionals

The Methodology and Standards staff members provide support by developing Policies and Standards. They are subject matter experts in survey operations. They are also involved during the design phase in order to help explain and clarify the requirements.

5.2.4. Project Managers

The Stats SA project manager works closely with the supplier's project manager. They bridge the gaps between the two organizations and make sure that the deliverables are managed properly and on time.

5.2.5. Training approaches and knowledge management

Users are required to spend at least a day in a training session, taking them through the functionality of the system as well as how to use it.

The Training Manual is used during the training sessions. The Training Manual contains complete descriptions of the system. The users can also use this document for reference purposes.

The system is designed such that tool tips (online help) are available to the user when hovering over certain areas of the user interface. These tool tips explain the features over which the mouse may be hovering. This allows the user to have information directly at a point of need without having to go through the Training Manual.

5.3 Training and Knowledge Management

The training strategy was to train the ESDMF project staff who would train other members of the Stats SA users, the so-called, train the trainer approach. This way knowledge about the ESDMF and its tools would be transferred to the eventual users. The ESDMF staff members received training on how to be trainers. Additionally, we engaged the organisations training department in this endeavour. By so doing, we were aligning the training programme of the ESDMF system of tools to that of the organisation as a whole.

5.4 Partnerships And Cooperation Between Agencies

In Latvia, we learned that during the development of their system, their outsourced supplier took a while to understand the business of the statistical organization. It came as no surprise when we ran into similar problems with our supplier, as much as we were not happy about it.

Their Integrated Statistical Data Management System (ISDMS) uses Bo Sundgren's model of metadata system, which they used as a firm foundation for the theoretical definition of metadata. We learned the importance of having a solid foundation in the definition of metadata

In Ireland, we learned about the cultural issues regarding communication between the customer and the supplier. Additionally, they had the same problem as in Latvia in that the development of their system also took longer than originally planned. This happened even after Ireland provided very detailed documentation on most of the major aspects of the system. Once again, when we ran into similar problems, we were not surprised, as much as we did not like it.

In Slovenia, their metadata model is also based on Bo Sundgren's model, with some modifications in areas where they believe that their components are adequate to meet Bo Sundgren's requirements for a metadata system.

Their development model is to build the system in-house and outsource when they get to maintenance phase. They continuously re-skill and train their staff as they bring in new technologies aboard.

From New Zealand, we adopted a few of their practises. For example, we brought in the Statistical Value Chain into Stats SA. This is how we view the business of statistical production processes within Stats SA. We also adopted the way they broke down metadata into five categories, namely, definitional, operational, system, dataset and procedural/methodological metadata. One of their experts helped us to evaluate the respondents to the tender for the development of the ESDMF.

In our trip to Australia, we learned that in order to have a successful data warehouse project, there is a need to develop policies and standards which will define how the system should be designed. When we returned to South Africa from that trip, we

restructured the team into two groups, the Policies and Standards team and the Technology team. The Standards and Policies team developed policies and standards which were used by the Technology team in the development and implementation of the ESDMF.

Experts from Sweden occasionally came to Stats SA to advise us on various aspects of metadata and statistical production processes. For example, a few years ago, Bo Sundgren, a well known expert on metadata, came to Stats SA to advise us on how to proceed in the development of a metadata system. Recently, another expert from Stats Sweden came to conduct a workshop on SCBDOK, the Stats Sweden metadata template. He also conducted training on quality definition and quality declaration of official statistics. This gave us a better idea on how to develop a data quality template, as well as how data quality should be reported on.

In 2006, we met Alice Born (from Stats Canada) when we attended the METIS conference. We engaged her regarding their development efforts of their metadata system, Integrated Metadata Data Base (IMDB). We applied that knowledge during the development of our Survey Metadata Capturing Tool.

Consultants from Canada help us in other projects within Stats SA. During their tenure we engage them for advice and other consultation.

We used the Corporate Metadata Repository (CMR) model by Dan Gillman, from the US Bureau of Statistics in our understanding the metadata model, especially with regard to the ISO 11179 Specification. We also sent our metadata model to him and other metadata experts for review and critique.

5.5 Other Issues - Organizational Change Management

5.5.1. Climate and Culture Assessment

Preliminary Organisational Change Management (OCM) initiatives necessitated a review of the operating culture at Stats SA in order to understand the 'lie of the land' in which the system will be introduced. The information contained in the Culture & Climate Assessment was obtained through a number of OCM diagnostic interventions, targeted specifically at internal stakeholders. This was done by holding focus groups as well as running an online survey via Stats SA intranet website.

A key challenge to Stats SA is to focus the organisation on the strategic importance of the DMID project, not only in as far as it assists an individual in their immediate job function, but even more importantly how it contributes to the overall wellbeing of the South African society at large and the contribution it makes to strategic decision making at government level. DMID communication messages need to create a sense of higher purpose to help individuals with long term strategic thinking.

5.5.2. Change Readiness Assessment

A Change Readiness Assessment was conducted to determine the current capacity of Stats SA to change, and to identify areas of resistance towards DMID requiring Organisation Change Management (OCM) interventions.

The Change Readiness Assessment was conducted via a survey and series of focus groups.

The following 'change readiness dimensions' are integral to enable commitment towards DMID and formed the basis of the Change Readiness Assessment:

- Clear vision
- Effective leadership
- Positive experience with past change initiatives
- Motivation to do the project
- Effective communication
- Adequate project team resources

5.5.3. What is Change Readiness?

OCM is a critical, although often bypassed element in organisations. It focuses on the 'human response to change', helping people understand, accept and commit to a new way of working. One of the key upfront steps in the change process is the Change Readiness Assessment.

The Change Readiness Assessment is a process used to determine the levels of understanding, acceptance and commitment likely to affect the success of the planned change. Change readiness is gauged along an axis known as the Change Commitment Curve, which is depicted below:

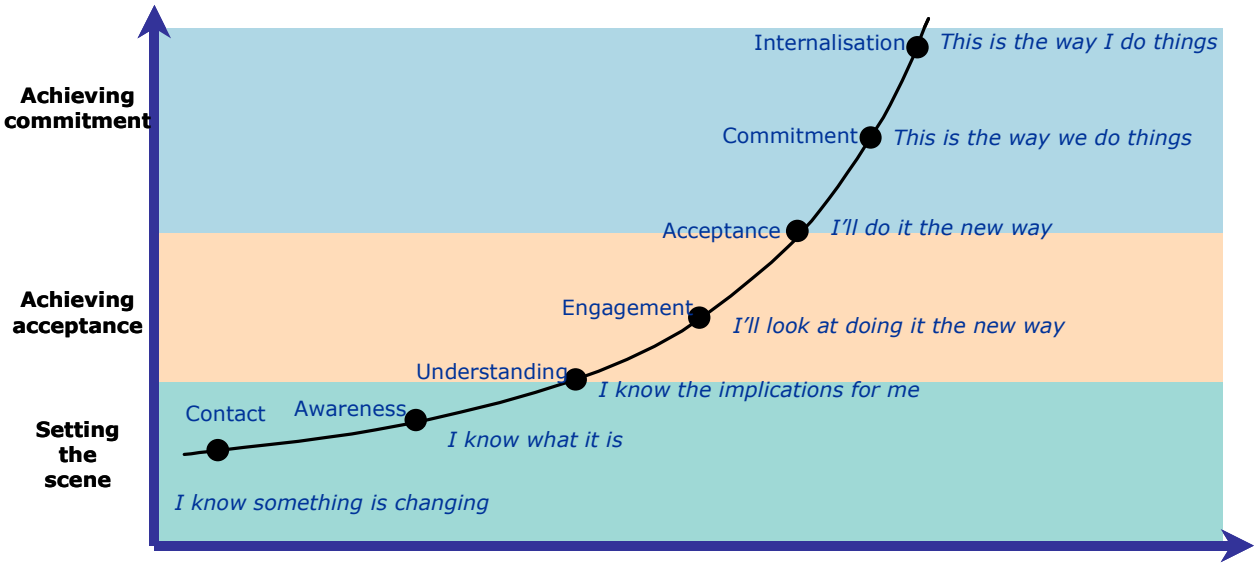


Figure 24: Change Commitment Curve

As the DMID project phases roll out, different stakeholders will need to be at specific levels of commitment. The level of commitment required will be dependent on the role they play in the DMID project and their ability to influence the program. The Change Commitment Curve will provide a framework for understanding and tracking the requisite levels of commitment that stakeholders need to be facilitated through so that OCM interventions can be developed accordingly.

A Change Readiness Assessment will become an obligatory OCM intervention prior to the rollout of a new phase on the DMID project.

5.5.4. Findings

The following were the finding from the assessments:

- Executive Management does not have the same understanding of the DMID project.
- Lack of communication between management and sub-ordinates; this makes it difficult for sub-ordinates to understand the purpose of the project and the impact it has on their working lives.
- Lack of support from Executive management will result to resistance and difficult success of the project
- If management does not communicate, does not understand, and does not promote the project, it will result in difficulty to deliver the message and get buy-in from staff in the organisation.

5.5.5. Next Steps from the Findings

The findings of the assessments resulted in identifying where some of the key staff members belonged on the Change Commitment Curve. In general, most were in the “Setting the Scene” and “Achieving Acceptance” area bounded by in time by “Contact” (“I know something is changing”) and “Understanding” (“I know the implications for me”). Obviously, a lot of effort is needed in order to move from that area to “Achieving Commitment” demonstrated by “Internalisation” wherein staff can claim that “This is the way I do things”

Another outcome of these assessments was to organize a Leadership Alignment workshop. In this workshop, the Executive Committee was given a presentation of the findings and the path forward. The path forward is to ensure that the leadership understands the goals of the project and how they line up with the vision of Stats SA. The leadership was also instructed on how to communicate the same message about the project.

6. LESSONS LEARNED

6.1 **Lessons** Learned

The supplier had a difficult time understanding the business of Stats SA, which is statistical production processes. Additionally, the goal of the project is to improve quality, which will help support the vision of Stats SA "to be the preferred supplier of quality statistics". Even in the face of this vision, the supplier failed to recognize that quality was a primary business objective.

Under pressure of meeting the deliverables, the supplier ignored the Skills Transfer Plan, with the result that the Stats SA developers were not involved in the final design and development of the system.

For a project of this magnitude (three years) and complexity, we decided to break down the deliverables into twelve phases. Each phase was planned to be three months long in duration. Also, each phase was planned to be a complete deliverable in its own right, even though the next phase was planned to build on the previous phases. The first phase was delivered late mainly due to the lack of understanding that the supplier demonstrated. The key is that clear understanding of the requirements is very important in meeting the deliverables as well as milestones for those deliverables.

Phase 1 took very long, considering the time allocated for the project, due to challenges uncovered in the capabilities of the service provider as this was a "new" undertaking for them and the capacity due to staff turnover (for both Stats SA and the service provider).

It became evident that the approach was not effective, as the project was only in phase 2 out of 12 phases after two thirds of the time allocated, an alternative approach was then put in place for progress to be based on tool delivery versus phase approach.

It was during this process that the service provider instituted some claims against Stats SA that has led to a disassociation, pending legal conclusion, resulting in the threat to the DMID project. At this stage only 7 out of 52 tools have been delivered.

6. ATTACHMENTS & LINKS

Documents to be attached:

1. Survey Metadata Standard Template (*Survey Metadata Capturing Tool_v0.10.doc)*
2. Web page of the Metadata Capture Tool in MHT format *(Summary of Survey Metadata Record.mht)*
3. SASQAF *(South_African_Statistical_Quality_Assessment_Framework_V05.doc)*

*** END ***