

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 work session on statistical metadata (METIS)
(Geneva, 3-5 April 2006)

Topic (iii): Metadata and the Statistical Cycle

STATISTICS NEW ZEALAND'S END-TO-END METADATA LIFE-CYCLE

Invited Paper

Submitted by Statistics New Zealand, New Zealand¹

I. INTRODUCTION

1. Statistics New Zealand has embarked on a major change initiative focussed on its business processes and systems, called the Business model Transformation Strategy (BmTS). The BmTS aims to facilitate the development of processes and systems to deliver on specific outcomes and priorities, while informing and driving the design and building of new capability and competency models for the future organisation. The paper covers Statistics New Zealand's view of the end-to-end statistical cycle, describes where metadata fits in, and gives an example of an implementation of the proposed data and metadata life-cycle approach.

II. INTRODUCTION TO THE BUSINESS MODEL TRANSFORMATION STRATEGY.

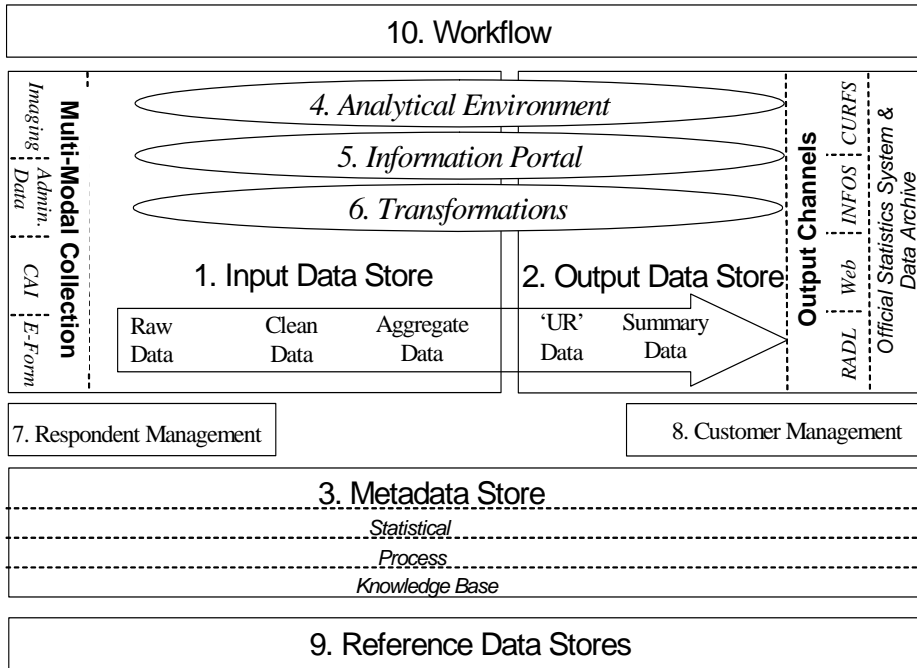
2. The Business model Transformation Strategy is focused on Statistics NZ business *processes and systems* including:

- a. the identification of a need for statistical output;
- b. the design and build of systems for meeting that need;
- c. the collection, processing and analysis of input data; and,
- d. the production and dissemination of output.

3. Its aim is to improve, streamline and standardise in order to optimise the Statistics NZ business model from end to end.

4. In order to deliver on Statistics New Zealand's organisational key strategies the following conceptual view of the ten key business components that are required to operationalise the new business model has been developed.

¹ Prepared by Gary Dunnet, gary.dunnet@stats.govt.nz; and Rebecca Merrington.



5. The BmTS will be judged successful if, at the end of the three year period (2007/08):
 - a. The ten components of the new business model have been identified, solutions designed, and the solutions are accepted by the organisation (and internationally) as consistent with known best practice or best method;
 - b. The ten components of the new business model have been implemented within the six pilot programmes;
 - c. There is a reduction in the operating cost to produce a statistical output (that are operating on a separate subject matter system) by between 10 – 20% after moving to the new end-to-end business model;
 - d. There is a reduction of 50% in the investment (of time and money) required to implement the end to end processes and systems required for a new statistical output;
 - e. Staff throughout Statistics NZ have a good understanding of the business solutions and that business areas have assessed how the business solutions will impact their functions and business processes; and,
 - f. A high level plan to implement the business model across the remainder of Statistics NZ is developed.
6. In order to meet these success criteria, there are four fundamental elements that are seen as critical to the creation of a new business environment for Statistics NZ. These are:
 - a. A corporate approach of Statistics NZ's data resource to prevail over a collection or even subject specific view.
 - b. A robust metadata environment that supports the full lifecycle of Statistics NZ created data (an end to end view of data) and will facilitate the implementation of the OSRDAC metadata and data repository.
 - c. The definition of a set of generic business processes that will be supported by standard technology environments, that provides the flexibility to adapt to new statistical requirements.
 - d. An organisational structure is created that supports the appropriate balance of data management across the Official Statistical System and the specific needs of the various internal business areas and external researchers.
7. The intention is to provide a new business environment that:
 - a. Facilitates the more challenging data collection, integration and analysis necessary to meet the increasingly complex policy and research needs of government and the wider research community

by abstracting the business users and their business processes from the underlying data structures and database systems, moving our statistical staff up the analytical 'value' chain.

- b. Creates the flexibility to respond to changes in users needs and demands, to make use of new data sources or methods and to provide a flexible range of generic information access methods. While also providing the ability to more easily match and confront data in order to increase the quality of Statistics NZ information.
- c. Increases the accuracy, relevance, coherence and interpretability of statistical outputs. It is intended to increase the value from the investment made in official statistics through improving accessibility to statistical information while preserving the confidentiality of data providers and continuing to reduce respondent burden, particularly sub-populations.
- d. Increases the use of administrative data, to make better use of the data that's available, reducing the number of individual collections or the need for new collections to create new statistics.
- e. Simplifies the migration of data and systems as underlying technologies change.
- f. Builds a professional environment that creates a more satisfying working experience.
- g. Provides a standard environment and uniform systems that will allow staff to quickly get up to speed with new subject matter.
- h. Standardises the skills sets and professional development costs of our staff.
- i. Allows Statistics NZ to provide standard information management tools and services for official statistical purposes.

8. In order to develop the new business environment the following principles will be applied across all processes and systems in the End-to-End Model (E2EM):

- a. A standard business process model and E2EM will be implemented for all new Statistics NZ developments, with the aim of removing stove-pipe developments. A process will be put in place to approve changes, exceptions and additions to the E2EM once it is in place.
- b. The systems and processes implemented throughout each of the ten components elements (modules) will be built in a manner that is generic and well documented, utilising externally available modules where possible. The systems will comply with the I&T Strategy. The 80/20 rule will be used when developing generalised systems - core and clerical work will be covered by the 80%, technical complexities will interface with the developed 80%.
- c. Data and metadata must be known, accessible, secure and useable. Data structures and change history will be documented and managed with a life-cycle focus.
- d. There are two key business requirements:
 - i. The production of quality official statistics; and
 - ii. The ability to produce and maintain a quality research and development data archive environment.
- e. A (yet to be determined) consistent efficient approach to methods, processes and systems, will be used when developing, compiling, accessing, analysing and disseminating statistical information. Efficient processes will enable more time for the analysis and dissemination of statistical information or shorter release cycles.

9. The ten key business components that are required to operationalise the new end-to-end business model need to be considered in the context of both the Business Process Model and the statistical data flow through the organisation. Utilising both these contexts could assist in enabling each business area to assess how the business solutions will impact their functions and business processes as the changes will lead to a very different environment for some areas.

III. THE ROLE OF METADATA IN THE BMTS

10. Statistics New Zealand has always recognised the benefits of robust metadata management, including its role in data integration and managing business processes. In fact, during the last large capability enhancement exercise, the IT Strategy, in the mid 1990's, Statistics New Zealand commissioned Prof Bo Sundgren to provide a strategic direction paper on metadata management. From this work and within the IT Strategy project, Statistics New Zealand created metadata infrastructure like CARS and SIM. This infrastructure has performed exceptionally well in some areas (classifications management) and adequately (at best) in others. However, where there are issues with the existing infrastructure, the issue is not about the

design of the infrastructure, but that this infrastructure is in part, poorly maintained, unstructured, difficult to access, unintegrated, passive and not carefully managed.

11. Essentially, the Business Model Transformation Strategy (BmTS) is designing a metadata management strategy that ensures metadata:

- a. fits into a metadata framework that can adequately describe all of Statistics New Zealand's data, and under the Official Statistics Strategy (OSS) the data of other agencies
- b. documents all the stages of the statistical life cycle from conception to archiving and destruction
- c. is centrally accessible
- d. is automatically populated during the business process, where ever possible
- e. is used to drive the business process
- f. is easily accessible by all potential users
- g. is populated and maintained by data creators
- h. is managed centrally

12. From this list, it is obvious that one of the most exciting aspects is that in this environment metadata will be created, used and stored throughout the end-to-end business process. Metadata becomes the heart of "How we do things" as professional statisticians.

13. The use of metadata to drive business processes is not a new idea, but is yet to be fully realised in other statistical agencies. This is because the ideal requires an extensive understanding of the way an organisation operates, how it should operate, and the relationship between these processes and the enterprise's architecture. Under the BmTS, Statistics New Zealand is currently addressing both its business processes and the infrastructure needed to support these processes. As a result, end-to-end data/metadata management will become a reality.

14. Even though this work is pioneering, Statistics New Zealand is convinced that end-to-end metadata management is the only way to fully realise Statistics New Zealand's leadership of the official statistics system and need for standardisation, simplification and integration of statistical information.

15. From investigation of international best practice, and an understanding of the Statistics NZ's existing metadata infrastructure, it is likely that the following will be developed:

- a. metadata requirements inventory (collected from metadata users)
- b. roles and responsibilities charter establishing metadata owners, populators, maintainers and regulators
- c. a policy framework that provides the right incentives
- d. comprehensive metadata model (to define the content of metadata)
- e. standard data item definition framework/template
- f. thesaurus of statistical terminologies which define and organise terms into a structure
- g. taxonomy of relationships between the terminology structure and other metadata and data
- h. metadata storage plan (this may require the establishment of separate repository/ies)
- i. Extract Transfer and Load (ETL) plan establishing methods of metadata population
- j. metadata registry (to organise the metadata defined in the model)
- k. metadata 'viewer' that ensures that data users have easy access to relevant metadata in a timely manner
- l. classifications repository (CARS already exists)
- m. questions Library
- n. metadata store
- o. training and implementation plan

16. Metadata is the cornerstone of the other initiatives within the BmTS, including:

- a. establishing robust metadata management is essential to the structure of Input and Output Data Environments;

- b. understanding what metadata is required to be used, created and potentially stored throughout the end-to-end business process model and then ensuring systems deliver the required outcome (e.g. Respondent Management);
- c. as the driver of the statistical and operational Transformations and processes;
- d. providing access to information (Information Portal, Analytical Environment and Dissemination channels); and
- e. ensuring metadata is reused and updated from different systems/stages in the business process life cycle.

17. While developing an integrated and dynamic metadata management system has many benefits, some of the risks and issues to be managed include, the size of the initiative and that from initial investigation there does not seem to be a one stop shop solution to metadata management.

IV. THE SHAPE OF METADATA AT STATISTICS NEW ZEALAND

18. Statistics New Zealand has defined its metadata in terms of the "Developing a Common Understanding of Standard Metadata Components - A Statistical Glossary" (Joint UNECE/EuroStat Work Session on Statistical Metadata (METIS), March 2002), along with the Australian Bureau of Statistics model. Metadata is defined in terms of being 'definitional', 'procedural', 'operational', 'systems' or 'dataset' (examples given below).

19. In order to ensure that metadata is relevant and useful, it is critical that Statistics New Zealand understands its audience. In 2002, Gareth McGuinness undertook an Statistics New Zealand audience analysis project ("Full Audience Analysis Paper") and identified several broad user groups (Public, Professional, Technical). It is considered necessary to include a fourth group (System) to facilitate a move to a more 'active' metadata environment. So our metadata initiatives recognise the following metadata audiences:

- a. Public - the 'person on the street'
- b. Professional - analysts and researchers
- c. Technical - academics and statisticians
- d. System - computers driving statistical or business processes

20. As a result of technology and prevalent thinking in the mid-1990's, Statistics New Zealand faces the following issues with its existing metadata infrastructure:

- a. metadata is not kept up to date
- b. metadata maintenance is considered a low priority
- c. metadata is not held in a consistent way
- d. relevant information is unavailable
- e. there is confusion about what metadata needs to be stored
- f. the existing metadata infrastructure is being under utilised
- g. there is a failure to meet the metadata needs of advanced data users
- h. it is difficult to find information unless you have some expertise or know it exists
- i. there is inconsistent use of classifications/terminology
- j. in some instances there is little information about data, where it came from, processes it has been under or even the question to which it relates

21. These issues have resulted in the corporate realisation that the significance of metadata to the effective performance of a National Statistics Office has been overlooked with the result that the current metadata infrastructure is not dynamic enough nor suitably integrated.

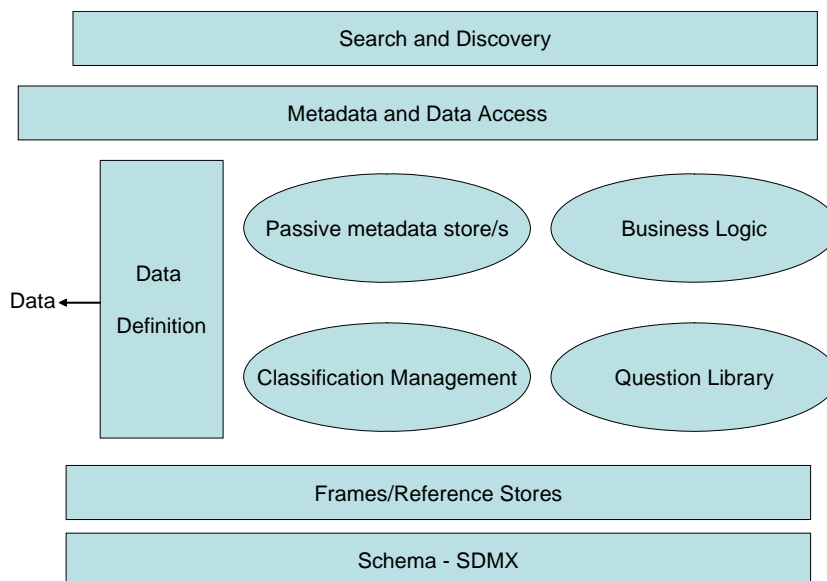
22. However, although some of the existing metadata infrastructure remains passive and unintegrated, the value of the existing metadata infrastructure should not be discounted. Therefore, the metadata management strategy, while addressing the issues highlighted, wherever appropriate re-use either the existing infrastructure or the principles/structure behind it. This is particularly true of Statistics New Zealand's classification management tool - Classifications and Related Standards (CARS).

23. From an assessment of internal and external best practice. The following principles of metadata management have been established. These principles include:

- a. metadata is centrally accessible
- b. metadata structure should be strongly linked to data
- c. metadata is shared between data sets
- d. content structure conforms to standards
- e. metadata is managed from end-to-end in the data life cycle.
- f. there is a registration process (workflow) associated with each metadata element
- g. capture metadata at source, automatically (where possible)
- h. establish a cost/benefit mechanism to ensure that the cost to producers of metadata is justified by the benefit to users of metadata
- i. metadata is considered active
- j. metadata is managed at as a high a level as is possible - managing at the lowest level is prohibitive
- k. metadata is readily available and useable in the context of client's information needs (internal or external)
- l. tracking the use of some types of metadata (eg. classifications)

24. Over the last six months a comprehensive assessment has been undertaken of metadata requirements across both Statistics New Zealand and the broader Official Statistical System (OSS). These requirements have been assess against a number of international models (incl. Dublin Core, ISO 17369 (SDMX), ISO / IEC 11179, ISO 19115 etc.) and we are starting a prototype based around SDMX (Statistical Data and Metadata eXchange). We expect to use SDMX (along with the associated Common Vocabulary Library) as the base for our metadata storage and management.

25. Through the user requirements, the following key metadata infrastructure model has been developed. This model is being continually tested against the enterprise wide metadata requirements. The model is also being used to guide the scope of some small 'metadata related' projects; for example, metadata capture to feed the data definition around the Input Data Environment.



V. THE FIT OF METADATA INTO THE BUSINESS PROCESS MODEL.

26. Statistics New Zealand has created a generic 'as is' Business Process Model (BPM) consists of seven key stages - from the identification of a need at one end, through to the dissemination of statistical outputs to meet that need at the other end. The target end to end model that the BmTS aims to operationalise will focus our resources on the Need, Analyse and Disseminate stages, and reduce the resources spent on the intervening stages (Design, Build, Collect, and Process) once these have been rationalised, streamlined and standardised.

27. Metadata management and the Metadata Store supports all phases of Statistics New Zealand's agreed 'As is' business process model from end-to-end. A more specific discussion about the metadata produced and used at each stage of the business process is outlined below.

VI. PROPOSED METADATA FRAMEWORK

28. The following is the proposed metadata framework. This framework takes the definition of metadata established above and then discusses different attributes of the definition.

Attributes of Metadata	Categories	Types of Metadata				
		Definitional	Procedural	Operational	Systems	Dataset
		<ul style="list-style-type: none"> statistical units/population scope and coverage definition concepts classifications data items - names and definitions, value sets statistical terminology, glossary, thesaurus 	<ul style="list-style-type: none"> objectives collection instruments , instructions, CATI scripts, IFU scripts infrastructure systems used methodologies sample method interviewer instructions imputation and estimation methods question modules 	<ul style="list-style-type: none"> response rates, sample size and distribution sample frame information edit failures coding quality information quality metrics release and various approval sign-offs notes to assist with approval sign-off for dissemination statistical release information 	<ul style="list-style-type: none"> edit rules derivation rules coding rules physical field/database descriptions imputation and estimation rules 	<ul style="list-style-type: none"> datasets - structure, footnotes, titles etc metadata to describe micro and macro output datasets dissemination products annotations technical notes
What is it used for?	<ol style="list-style-type: none"> Assess data quality (all the quality dimensions). Understand how data is created. Identify and locate data Remember production steps. Train new staff. Tune and improve processes. Run software tools. Understand software tools. 	1,3	1,2, 4,5	1,6	1,7	8

Where in the business process is this used?	1. Need 2. Design 3. Build 4. Collect 5. Process 6. Analyse 7. Disseminate (incl. archive)	1,2,3,5,6,7	2,7	2,4,5,7	2,4,5,6,7	2,3,6,7
Who defines it?	1. Statistical Designers (IT, Methodologists, Questionnaire) 2. Input data providers 3. Production statisticians 4. Software tools	1	1,3	2,3	1	4
Who uses it? (in general, not related to access rights)	1. Public 2. Professional 3. System 4. Technical - External 5. Technical - Operational	1,2,3,4,5	3,4,5	3,5	3,5	3,5
How is it used?	1. Active 2. Passive	1,2	1,2	1,2	1	1
How is it populated?	1. Automatic 2. Manual	1,2	2	1	1,2	1
What format is used?	1. Structured 2. Unstructured	1,2	1	1	1	1

VII. CONCLUSION

29. Robust metadata management is the key to the success of the BmTS. Collecting comprehensive requirements and learning from past and international experience will ensure that Statistics New Zealand develops a robust metadata management system.