

Complete case study (Australia)

Organization name	Australian Bureau of Statistics
Total number of staff	3400 (approx) 47% in Central Office (Canberra) 53% across 8 state/territory offices
Contact person for metadata	Alistair Hamilton
Job title:	Chief Statistical Information Architect Information Management Transformation
Email:	alistair.hamilton@abs.gov.au
Telephone:	+ 61 2 6252 5416

Table of contents

- [1. Introduction \(Australian Bureau of Statistics\)](#)
- [2. Modelling the Information and Processes of a Statistical Organization \(Australian Bureau of Statistics\)](#)
- [3. Statistical Metadata in each phase of the Statistical Business Process \(Australian Bureau of Statistics\)](#)
- [4. Statistical Metadata Systems \(Australian Bureau of Statistics\)](#)
- [5. System and design issues \(Australian Bureau of Statistics\)](#)
- [6. Organizational and workplace culture issues \(Australian Bureau of Statistics\)](#)
- [7. Lessons learned \(Australian Bureau of Statistics\)](#)
- [A Brief History of Metadata \(in the ABS\)](#)
- [ABS IMTP](#)

Create the sections of your case study

- If you have any problems, please [send us an email](#).

Click to add an introduction
Click to create section 2 (Modelling the Information and Processes of a Statistical Organization)
Click to create section 3 (Statistical Metadata in each phase of the Statistical Business Process)
Click to create section 4 (Statistical Metadata Systems)
Click to create section 5 (System and design issues)
Click to create section 6 (Organizational and workplace culture issues)
Click to create section 7 (Lessons learned)
Click to create section 8 (Attachments and links)

1. Introduction

[Next section](#)

Preface to most recent update (2011.1)

The previous major update to this case study occurred in the first half of 2009. As recorded in the document entitled *A Brief History of Metadata (in the ABS)* (referenced simply as *BHM* hereafter), which is attached to this Case Study, the second half of 2009 saw fundamental decisions made by the ABS leading to initiation of the [ABS Information Management Transformation Program \(IMTP\)](#) in February 2010. While IMTP designates a specific program within the ABS, including a specific top level unit within the organisation chart, the aim is for the ABS to achieve IMT (Information Management Transformation). All staff within the ABS have a role in achieving IMT.

IMT will include fundamental reshaping of policies and strategies related to metadata management developed by the ABS over the past two decades. At this stage, however, IMT remains an early "work in progress".

IMT can be seen as focused on the "to be" environment for the ABS, in terms of business architecture, data/information architecture and other elements of enterprise architecture, as well as on the process for achieving the transformation (including business process re-engineering) required to realise the "to be" state. At this time many details contained in the previous version of this case study continue to accurately describe the "as is" environment for metadata management within the ABS.

In the initial update for 2011 it has been decided to focus the main body of the case study on the "to be" state and initial steps toward that state. Many details of the "as is" environment have been moved into supporting documents. Other aspects can be found by referring to the 2009 version of the case study. It is possible within the wiki to view earlier versions of each page. For convenience, also, PDF versions of the 2009 edition of this [case study](#) and 2009 edition of [BHM](#) have been made available.

The result of this approach is that the Case Study document is now shorter than the 2009 edition. Additional details will be added to the documentation as IMT progresses, including its new approach to statistical information management, including metadata management.

It should also be noted that, as with all content in the METIS wiki maintained by ABS practitioners, this is an informal working document shared with colleagues in the field of statistical information management. Unless unambiguously indicated otherwise, no content accessed via these wiki pages should be considered to represent a formal statement on behalf of the Australian Bureau of Statistics.

1.1 Metadata Strategy

Ultimately any strategy exists to support the ABS mission and objectives as set out in the organisation's [corporate plan](#). In particular, the availability of appropriate metadata and the application of sound statistical information management practices are critical to supporting informed use of statistics and the quality of the statistical services we deliver to the nation.

[BHM](#) provides information on the evolution of ABS strategies related to metadata over time, including extensive information in regard to Strategy for End-to-End Management of ABS Metadata established in 2003.

[IMT](#) can be seen as superseding the 2003 strategy, although at this stage there is no "direct replacement" strategy focused specifically on metadata and its management. IMT focuses instead on strategies related to "statistical information" management which spans metadata (in its broadest sense) and data.

Although IMT supersedes the 2003 strategy, most of the fundamental ideas contained in the 2003 strategy remain relevant. For example, none of the twelve cornerstone principles outlined in the 2003 strategy have been disavowed as irrelevant or inappropriate. In this example, however, IMTP seeks principles

- focused on statistical information management rather than simply metadata management
- rationalised where relevant with principles underpinning other frameworks applied within the ABS (eg enterprise architecture, quality management of statistical processes)
- that take account of relevant standards and frameworks associated with the global and national "industry" of producing official statistics, as well as standards and frameworks relevant to data providers and to users of official statistics
- expressed concisely in terms meaningful, and motivating, to business staff
- supported by relevant guidelines and training to assist in them being applied appropriately to specific design processes and decision making

This process of rationalisation is underway currently and it is expected the updated set of principles will be added to the case study once available.

More generally, strategic planning for IMT can be seen as learning from experience with the 2003 strategy (eg much slower progress, and much more mixed success, in putting the strategy into effect than had been anticipated).

Well defined, corporately accepted and supported, governance for information management is much more of a foundation consideration for IMT. This includes clearly established norms/principles/expectations, clearly established authority and accountability and clearly established processes for assessing compliance and actively managing non-compliance. The corporate positioning of IMTP (eg reporting directly to the head of the organisation and independent of any one operational or support division) promotes its ability to address governance requirements successfully compared with the implementation of the 2003 strategy.

The IMT strategy of starting with the Metadata Registry Repository (MRR) as the key enabling infrastructure, including its integration with Statistical Workflow Management capabilities, can be seen as establishing a "central nervous system" to support the new environment – including supporting its relationship with "legacy" applications and repositories – where the 2003 strategy primarily targeted developing new repositories and redeveloping existing repositories without such a well developed strategy for achieving "business integration" in practice.

Compared with the 2003 strategy, IMT work on the Statistical Information Management Framework also includes much greater integration with external frameworks such as GSBPM (Generic Statistical Business Process Model) and GSIM (Generic Statistical Information Model) as well as with other frameworks applied within the ABS (eg Enterprise Architecture).

In October 2009, the ABS Executive formally agreed on Statistical Data and Metadata Exchange (SDMX) and Data Documentation Initiative (DDI) as the standards that will form the core of the ABS's future directions and developments with regard to statistical information management. This means strategic engagement with the two standards communities, including encouraging them to co-ordinate their work in order to support NSIs and others who seek to use both standards, is a high priority.

Participation in the Statistical Network strategy is the primary, but far from only, example of IMT's strategic focus on collaboration (internationally and/or nationally) when it comes to statistical information management.

More detailed formal statements of IMT strategy in regard to statistical information management are still being reviewed within the ABS. Any encapsulation of strategies which is agreed for general release beyond the ABS will be added to this case study once available.

1.2 Current Situation

BHM describes how the current situation has evolved within the ABS. Documentation of IMT outlines the current situation.

The majority of data collection and input processing activities for business and household surveys have moved toward implementation of high level metadata frameworks informed by ISO/IEC 11179. These frameworks were developed over the past decade and postdate the ABS specific metadata framework which was implemented for the corporate output data warehouse which was developed during the 1990s.

Key elements of current metadata infrastructure, which predate initiation of IMTP in 2010, include major repositories related to

- statistical activities
 - Termed "collections" by the ABS, these activities include surveys, censuses, statistical analysis of administrative data sources and statistical "compilation" activities such as preparing the national accounts.
- datasets
 - These are specific structured data files, data cubes and tables associated with statistical activities. Examples include various "unit record files" and aggregate outputs.
- classifications
 - This is a "legacy" system based on an ABS specific data model.
- data elements
 - This is a more recent development based on the metamodel found in ISO/IEC 11179 Part 3.
- questions and question modules
 - This was developed more recently for household surveys with an aim to generalise the facility in future.
- collection instruments
 - This was developed more recently for household surveys with an aim to generalise the facility in future.

The more recent developments also incorporate an approach to metadata registration based on ISO/IEC 11179 Part 6. Even if some of the older repositories cannot be completely replaced in the next few years it is anticipated that a common high level metadata registration framework, harnessing the MRR, can be implemented across the ABS for all classes of metadata. This does not imply that all classes of metadata will undergo exactly the same registration workflow, but the workflows for each class of metadata will be consistent with a higher level "metamodel" for registration.

Interoperability of the current ABS metadata models, including the legacy "output" model, with third party software (eg SAS, Blaise, SuperCROSS) continues to be an issue.

The increasing focus of the ABS and other agencies on the National Statistical Service (NSS) requires development of metadata models and capabilities which are usable beyond the ABS. The NSS needs to interoperate with agencies whose data content is more "administrative", "geospatial" or "research oriented" than "statistically" oriented. This provides additional challenges and issues in regard to metadata modelling.

While many of those agencies are at least as passionate about metadata as the ABS - but from a different "school" - the NSS also needs to support content producers and users for whom metadata is much less of an interest and priority. This raises questions about minimum metadata content and quality standards.

Understandably, metadata is a particular area of focus for the NSS. This includes a [simplified and generalised set of principles for managing metadata](#).

Challenges associated with the current situation, such as achieving a coherent "end to end" metadata driven environment(s) within the ABS and better supporting the NSS, underpin [IMT](#).

[Next section](#)

2. Statistical metadata systems and the statistical business process

[Next section](#)

2.1 Statistical business process

Since [GSBPM \(Generic Statistical Business Process Model\)](#) reached full maturity with the release of V4.0 agreed in April 2009 it has been regarded as the preferred reference model for the **statistical business process** within the ABS. This status for the GSBPM was confirmed with the area leading the enterprise architecture initiative within the ABS at that time. The [IMT](#) Program launched in February 2010 also confirmed this status for the GSBPM.

A number of reference models for the statistical business process existed, and were harnessed, within the ABS prior to the development of the GSBPM.

A particularly broadly applied model was affectionately known as "The Caterpillar" within the ABS. The Caterpillar was developed by the ABS as a reference model to support the Business Statistics Innovation Program (BSIP) launched in 2002.. It allowed a disparate range of surveys and other statistical activities whose processes were (especially prior to BSIP) very different in detail to describe what they did, why and how (eg what systems and data stores were used) in terms of a common high level reference point for the statistical life cycle. It later allowed "leading practice" to be identified in different parts of the statistical cycle.

The broad relationship between the Caterpillar and GSBPM, documented previously in this section of the case study, has been moved to a [supporting page](#).

Extensive process documentation, together with categorisations of information and even software interfaces, was developed during the course of initiatives such as BSIP and ISHS (both described in [BHM](#)). This activity was undertaken based on the "pre GSBPM" reference models for the statistical business process associated with those initiatives. It has been agreed existing process documentation within the ABS will not be rewritten for the sole purpose of referring to the GSBPM. Formalising, and making readily available, mappings between the GSBPM and the local reference models has been particularly important.

The paper [Applying the GSBPM within an NSI : Experiences and examples from Australia](#), prepared for the METIS Work Session in March 2010, provides more information in regard to ABS utilisation of GSBPM as well as in regard to statistical business process models that preceded GSBPM. Annex 1 of that paper provides a full description of the Caterpillar.

Key points include

- GSBPM is harnessed, within the business domain of ABS Enterprise Architecture, as the primary reference model for the for the statistical business process.
- Recognition within business architecture allows GSBPM to serve (among many other roles) the purpose for which it was originally commissioned internationally, namely linking metadata management (and statistical information management more generally within the data/information domain of Enterprise Architecture) to the statistical business process.
- GSBPM is seen as a vital enabler of practical, purposeful and collaborative external engagement at the international level, national level and sub-national level.
- GSBPM is harnessed as a reference model for the statistical business process for corporate planning and management purposes. A number of specific ABS applications of the GSBPM in this regard are listed in Section VI of the paper for the METIS work session.
- Experience with BSIP and ISHS highlighted the practical value of common (within each initiative) reference model for the statistical business process prior to GSBPM maturing to fulfil this role across the ABS as a whole, as well as internationally. (See Annex 2 in the paper for the METIS work session)
- While ABS waited until GSBPM achieved the level of maturity, and breadth of international recognition, associated with V4.0 prior to adopting it formally, it has had an active interest in GSBPM since the inception of its development. (See Annex 3 in the paper for the METIS work session)

2.2 Metadata System(s)

There are currently many systems within the ABS that encompass significant metadata definition and management aspects.

The MRR (Metadata Registry/Repository) associated with [IMT](#) is by far the most significant metadata system currently under development. The MRR's Registry capabilities will act as a "central nervous system" for systems across the ABS that define, manage and use metadata. At this stage the MRR is at the Proof of Concept phase.

An early activity associated with IMT was the first phase of a "metadata census". As suggested by selection of the term "census" it was originally hoped that this activity would provide a much clearer and more comprehensive "as is" picture of metadata management at a local level within the ABS as well as at a corporate level. GSBPM was an important point of reference for indicating what phases and sub-processes each system was supporting with which metadata.

An early issue encountered was the ability for those responsible for systems to describe in a clear and consistent manner the "types" of metadata managed within their systems. For example, if one system was said to work with "variables", another with "data items" and another with "data elements" were all three systems talking about the same "type" of metadata, or about different "types" of metadata (that maybe related to each other in some way)? Work associated with GSIM and the MRR should lead to this issue being more tractable in future.

The "as is" picture is also complicated the fact that many local systems currently need to "replicate", possibly in a specialised format with local content additions, metadata held in existing corporate repositories. The issue about consistently typing metadata compounds the issue of being able to establish which systems are managing which metadata simply because they currently can't source it from elsewhere – as opposed to managing metadata for which that system should be considered an authoritative source within the ABS. A significant number of processing systems currently have a secondary role as a "metadata system" only because – for a variety of reasons - they can't source the metadata they need systematically from elsewhere.

The second phase of the "metadata census" focused in more depth on metadata associated with core corporate stores and systems. The outputs have already contributed to the design of the Proof of Concept for the MRR and will contribute more broadly to the development of the MRR in future as well as being used as one practical test of the scope and nature of metadata requirements addressed by GSIM.

Early work on the metadata census confirmed that some current metadata systems are

- fully corporate.
- "shadow systems" which extend corporate systems to supplement the standard content with attributes of local interest.
 - The need for "shadow systems" should be eliminated once, via the MRR, modelling of information objects, attributes and relationships address the needs of the organisation as a whole.
 - Some systems may still not be able to only use "standard" metadata but the metadata actually used by these systems will be able to be described, and registered, together with the relationship of that metadata to "standard" metadata..
 - Some of the "shadow systems" have been designed and maintained to ensure they can be easily reintegrated with corporate systems in future while others have not.
- truly "local" systems
 - These exist for a variety of legitimate and not so legitimate reasons.
 - The best of them source relevant content from the Corporate Metadata Repository (CMR) as a properly maintained snapshot but then reformat that content to meet local needs (eg to support systems that cannot "read" the metadata directly and require it to be translated/packaged in a special way).
 - The worst of these update, evolve and create new metadata for local use independently of the CMR.
 - Others deal with classes of metadata (eg methodological parameters to drive specific processes) which are not currently managed within the CMR.

Information about key existing corporate metadata systems, documented previously in this section of the case study, has been moved to a [supporting page](#).

In regard to systems envisaged for the future, the Statistical Workflow Management (SWM) facility designed to work with the MRR is expected to provide a source for information related to, for example,

- reusable process specifications
- the assembly of processes into specific workflows
- results from executing defined workflows
- reusable business rules for driving and chaining processes and workflows

Earlier conceptual and exploratory work identified seven types of "process metadata" from "configuration" metadata about the IT environment and the user running the process, through to metadata which is a formal "input to", or "output from" the process, through to metadata which describes the process itself and which describes how chains of processes fit together. (None of these seven types of process metadata corresponded to "process metrics" as described below. Given there are already more than enough types of "process metadata", the ABS tends not to favour using the term to also denote "process metrics".)

Achieving a clearer path forward in regard to structuring and managing "process" metadata is seen as an important enabler to having other metadata (eg the structural definition of data elements) actively drive statistical processes.

It is intended that the work related to structural definition and description of processes harness appropriate standards such as BPMN (Business Process Model and Notation) and BPEL (Business Process Execution Language).

It is anticipated that, through SWM working with the MRR, it will become possible to specify and analyse detailed information related to the statistical information used by, and produced from, specific process steps.

A further priority is to better capture and store (for automated and interactive analysis and reporting) "process metrics" related to how statistical processes are performing (eg response rates, imputation rates, edit rates etc). Such data about the outcomes of processes is sometimes referred to as "process metadata", "operational metadata" or (typically in specific circumstances) "paradata" by others. Process metrics can be useful for internal monitoring, management and tuning of processes as well as generating data quality indicators for external dissemination.

2.3 Costs and Benefits

Section 2.2 details infrastructure delivered as the result of diverse projects, some of which first delivered outputs more than a decade ago. Lifecycle costs and benefits are extremely difficult to even estimate meaningfully.

Costs and benefits for new developments and redevelopments were estimated when developing business cases. While much better than a vacuum for planning purposes, past experience suggests these cost benefit analyses were seldom borne out with any precision in practice. Often this was because decisions were made over time to diverge from the original project plan in some way rather than just because the original estimation process was flawed or based on imperfect information.

IMTP is instituting a much more rigorous approach to

- estimation of costs and benefits during the planning stage
- establishing compelling evidence that the planned benefits are achievable in practice, together with establishing well defined outcome realisation plans to ensure the benefits will be achieved
- managing projects to ensure the implications for planned costs and benefits are understood in regard to, and refactored as a result of, any variation from the original plan
- managing related projects as a coherent program to ensure any benefits which rely on successful completion, and co-ordination, of multiple projects are realised (and that dependencies between projects are understood and supported appropriately)

For future developments, therefore, more concrete information should be able to report in this section.

During formulation of the detailed business case for IMT, however, it is not appropriate for the ABS to release to the public domain the details of estimated costs and benefits associated with the program.

2.4 Implementation Strategy

Information related to the implementation strategy can be gleaned from the description of [IMT](#) (including resources linked to the page) and from Section 1.1.

The challenges that provide the drivers for IMT must be addressed in one form or another. In order to achieve the transformation in a timely manner (eg in well under a decade), and realise maximum benefits for users of ABS (and other NSS) statistics, significant resources in addition to those allocated to undertaking and supporting current "business as usual" activities within the ABS will be required. This approach

- achieves greatest efficiency overall (a more protracted approach requires a smaller budget each year but stretches over many more years and ends up costing more in total)
- reduces, through a focused approach, risks to business continuity and sustainability during the transition period

The first generation of the information management framework and other enabling infrastructure such as the MRR, together with generic tool sets, is required before the main transformation (including re-engineering) across statistical production streams can begin in earnest. As has been the case for all elements of IMT, the main transformation period will be planned in detail prior to commencement (eg which re-engineering for which statistical business process will occur at which time during, eg, a four year period).

In terms of metadata management, the swinging of a pendulum can be seen to some extent in the [BHM](#).

Developments in the 1990s tended to be on a "big bang" basis. These were sometimes pejoratively referred to as "Cathedral Projects" for being too grandiose in ambition and design, and for taking much longer and much more money to complete than originally expected. Nevertheless, many of the results of these projects have proved to be of enduring value - so much so that many outputs have lived on long beyond their prime.

The strategy next (eg as formulated in 2003) became "opportunistic" and "incremental". There was notionally a "master plan" of what should exist in the longer term, but individual "construction projects" were much more modest in scale. Progress toward the "master plan" was much slower, less direct and more difficult than anticipated and hoped.

[IMT](#) is establishing a much clearer, more compelling, more widely shared and more actionable "master plan" together with the active corporate mandate and governance to achieve progress. Where the cathedrals of the 1990s tended to be largely designed and built in isolation, the IMT approach focuses on collaborative and sharable solutions underpinned by common standards and frameworks

A consistent learning has been that a well developed and managed implementation strategy (in addition to a development strategy) is essential. New capabilities are being delivered into a complex context of existing processes and infrastructure. Uptake of those new capabilities needs to be managed and promoted appropriately. (The simple "Field of Dreams" approach of "Build it and they will come!" has never yet worked for us.) Often the new capability and/or the implementation and communication strategy for it, needs to be refined based on early uptake experience. Whether it is managed by the development team or some other team, every major project requires a well planned and actively managed "Outcome Realisation" phase after it has finished delivering its major outputs.

[Next section](#)

3. Metadata in each phase of the statistical business process

[Next section](#)

3.1 Metadata Classification

The ABS doesn't have a formal "taxonomy" of metadata. One was proposed early in development of the 2003 metadata strategy but it wasn't included in the final document. It was found that discussions about how to "class" particular instances of metadata (in borderline cases rather than all cases) could become very protracted without that discussion seeming to generate any real value.

In general ABS concurs with the findings of Bo Sundgren in regard to [Classification of Statistical Metadata](#), namely that multiple valid approaches exist, with the optimum depending on why classification is being attempted.

One form of categorisation sometimes used within the ABS relates to purpose/use of metadata. This means a particular "piece" of metadata may (and often should) support more than one type of use. The categories are

1. (Search and) Discovery - Help users find data (or a metadata object in its own right, such as a classification) of relevance to their needs and interests
2. Definition - Help users understand data (or a metadata object in its own right, such as the definition of a data element)
3. Quality - Help users assess the fitness of associated data for their specific purpose
4. Process - Apply metadata to run processes, such as using a classification to drive an aggregation process or to provide a list of valid encoding values for editing purposes. It also includes defining other parameters that drive a process as metadata, such as the choice of which imputation method to use for which data element.
5. Operational - These are metrics on the results of the operation of processes such as edit rates, imputation rates etc. These can feed into internal decisions on managing and improving survey processes and into external "quality" decisions. This metadata is sometimes termed "paradata".
6. System - Low level information about files, servers etc that helps allow the physical IT environment to be updated without end user processes needing to be respecified.

The ABS also recognises "objects" in regard to which metadata can be assembled and registered. These include

- high level end to end statistical activities ("collections")
- individual datasets
- data elements
- classifications
- individual processes
- terms
- questions
- question modules
- collection instruments

These "objects" can be further broken down (eg data elements into properties, object classes, value domains etc).

The main way forward from the ABS perspective at this time is work toward [GSIM \(Generic Statistical Information Model\)](#). This should (among other things) provide a reference classification (or taxonomy) of "information objects" (including "metadata objects") that is shared in common beyond just the ABS.

Work on the Metadata Census within the ABS is also providing a "bottom up" approach to classifying/grouping "information objects" based on the requirements of existing systems and processes (including seeking to harmonise the sets of requirements and align them with constructs described within SDMX and DDI). As described in Section 2.2, this work (together with GSIM) will input to classing the objects supported by the MRR and also provide a use based checklist for testing the GSIM "taxonomy".

3.2 Metadata used/created at each phase

The ABS actively aspires to manage metadata consistently throughout the statistical business process. The MRR to be delivered under [IMT](#) will have a crucial role in realising this objective in the medium and longer term.

As documented in the supporting page for Section 2.2, consistent metadata related to the identification of collections and cycles and the definition of classifications is already used widely (although not universally) throughout the statistical business process.

As the ABS does not yet have a definitive taxonomy of "statistical information objects" (including "metadata objects") we do not yet have a definitive mapping of metadata used/created at each phase of the statistical business process. Nevertheless, indicative representations were created in 2006 and 2010.

The 2006 representation predates the ABS adoption of the GSBPM. The [most legible version of the diagram](#) predates the [final version](#) that was ultimately included in a briefing paper for senior management. Unfortunately the source file for the final version of the diagram has not yet been located and its reproduction from the briefing paper itself is hard to read.

The 2010 working document has been loaded to the [METIS wiki](#).

3.3 Metadata relevant to other business processes

IMT has a focus on statistical information management rather than, eg, financial and human resource information within the ABS.

That said, information about costs, information about organisational structures, people and their roles are examples of information that can be relevant to managing and performing statistical business processes. More generally Statistical Information, Corporate Information and Business Information can be visualised as intersecting circles, and the intersections with Statistical Information are certainly in scope for IMT.

Also, as long as it does not distract from the focus on the statistical information used and/or produced in the course of various sub processes within a **statistical business process**, there is no reason why elements of the Statistical Information Management Framework associated with IMT could not be applied to these other information domains.

A further connection is that a range of high level key performance indicators in relation to the core business of the ABS are expected to be able to be sourced via the MRR in future. These will assist in high level corporate monitoring and reporting, on a more consistent and informed basis, of efficiency, productivity, return on investments etc.

One reason IMT does not have a primary focus on corporate information and business information is that well recognised standards and frameworks (not specific to producers of official statistics) already exist for these domains of information. A more natural alignment in this case might be with Australian Government Architecture rather than the common reference architecture for producers of official statistics internationally.

These other domains of information are recognised within ABS Enterprise Architecture (eg in terms of data/information architecture). The redevelopment of information systems related to human resource management is an example of an "architecturally significant" project in this regard.

In this regard, the IMT approach parallels the 2003 metadata strategy which defined its scope as relating to "statistical" metadata (rather than all the metadata potentially relevant to any aspect of ABS operations).

[Next section](#)

4. System and design issues

[Next section](#)

4.1 IT Architecture

ABS Enterprise Architecture harnesses The Open Group Architecture Framework (TOGAF) which recognises domains of business, data, applications and technology architecture. In describing "IT Architecture" below, reference is primarily made to applications and technology architecture. Connections with data architecture are also explored.

Unless otherwise noted, descriptions in this section refer back to the main metadata systems as described in Section 2.2.

The newer metadata facilities are based on a Service Oriented Architecture. The older facilities tend to have monolithic coupling of the repository, the business logic and business rules (which are built into the application rather than embedded in services) and the User Interface.

Nevertheless, selected information about the collections defined in CMS is "projected" from CMS into an Oracle database. While only a small subset of the total information held in CMS, this comprises all of the core "structural" registration details about collections, cycles and profiles. Basic (read only) "collection metadata services" based on this content on Oracle are then provided for statistical processing applications to access.

A similar approach applies in the case of classifications except a much greater percentage of the total information held in regard to classifications is both "structural" and available on Oracle.

Apart from CMS and ClaMS (which include some descriptive content held only in IBM's Lotus Notes product) the other metadata holdings are all based in Oracle. There is extensive use of Oracle Stored Procedures for reusable services/functions and some use of true web services.

In summary, more recently developed facilities based on recent architectural standards within the ABS, tend to consist of

- a store (typically Oracle based)
- that is wrapped with low level Create, Read, Update, Delete (CRUD) services
 - "D" often actually refers to "Deprecate" (eg marking metadata as no longer the version recommended for use) rather than physically deleting metadata.
- that, in turn, are used as building blocks for higher level "business services" related to the store
 - these business services which consistently enforce business logic/rules - including resolving on an authenticated roles basis who is permitted to do what in terms of CRUD operations on specific elements of content within that store of metadata

- there is typically also a generic GUI associated with the store, for general browsing, management and administration purposes
 - typically, however, most business applications (eg for statistical processing and dissemination) simply access and apply the business services in the manner they require to interact with the metadata content rather than making use of the generic GUI
 - external applications are not able to use SQL or other means to interact with the metadata content store except via the CRUD layer

While SOA offers a lot of opportunities and potential, it also comes with a lot of new complexities compared with earlier approaches. It requires new understandings and a new mindset from those developers who are being asked to take up, and interact with, the available services as well as requiring the same from the business analysts and programmers within the team responsible for providing the metadata repositories and services. It can make the overall environment more complicated in some ways (eg services are calling services that call services etc and then somewhere at a low level a service is updated and everything needs to be configured appropriately to allow proper testing of that change). Implementing SOA in environments that include a lot of "legacy" processing systems that are not enabled for the new architectural directions is particularly challenging

During 2008 it became clearer that a significant aspect of the work on establishing an updated and coherent metadata framework for the ABS amounts to defining Enterprise Information Architecture (EIA) in the context of a statistical organisation. Without a clear and coherent EIA, there is a risk each service, or each bundle of services, is delivered with its own explicit or implicit information model. The ABS could have gone from having a dozen or so environments with subtle and not so subtle differences in their underpinning information concepts and structures to having an array of services based on a plethora of different, and unreconciled, information models. On the positive side, SOA can help make EIA practical and consistent. Rather than having the same objects and relationships specified in the EIA implemented, and extended, differently across a number of different environments, a single consistent but flexible bundle of services could be used within each environment. SOA and EIA are complementary rather than alternative directions.

The *IMT* strategy addresses the requirement for SOA and EIA to work together. It enables common information constructs, defined according to schemas aligned with relevant standards such as SDMX and DDI, to be used consistently via service layers. These service layers enforce core business rules. They also mean application developers can work with information objects at a business level without needing to understand, and code based on, the full details of the SDMX and DDI information models. The integration with Statistical Workflow Management is also an important element of the "to be" IT Architecture.

4.2 Metadata Management Tools

Statistical processing applications interact with metadata via services where possible. As described in *BHM*, however, many ABS processing applications and third party vendor products are not yet amenable to this approach. Where this approach is used currently it most often involves the application "reading" relevant content from the metadata repository rather than writing back new or updated records.

The *IMT* strategy seeks to fully, and consistently, realise this approach. Some existing key applications (and repositories) may need to be "wrapped" so they can interact with the MRR on a CRUDS basis. ("S" refers to harnessing the MRR Search capabilities to support discovery, selection of relevant content to Read etc.). Other legacy applications may need to be decommissioned, through delivery of services and interfaces that take their place, and content from a number of legacy repositories will need to be migrated to the (logically) centralised repositories associated with the MRR.

In the meantime, as described in the introduction to 2.2, there are cases where metadata from the Corporate Metadata Repository needs to be restructured and/or repackaged relatively manually to make it suitable for use in particular processing systems.

4.3 Standards and formats

Standards and formats currently in use for major metadata repositories are described in Section 2.2.

Under *IMT*, the primary standards are SDMX and DDI, interoperating with other "purpose specific" standards such as

- ISO 11179 for concepts.
- ISO 19115 (and related standards such as ISO 19139) for geospatial metadata, together with relevant OGC (Open Geospatial Consortium) standards for geospatial data and registries.
- Dublin Core and related standards for discovery metadata.
- BPMN for process modelling
- BPEL for process execution

Regardless of which standard's information model is being harnessed, content for interchange (eg to be read by applications) is typically represented in XML. In order to reduce the need to exchange large XML structures, where only a small proportion of the total information may be needed for a particular application, the XML used to describe an object can refer to sub components and related objects "by reference" rather than including all this information "in line". The calling application can then resolve the specific references (if any) which are relevant to its particular needs – once again typically resulting in smaller packages of XML than would be the case if a comprehensive set of information related to the component was included "in line".

While XML is used for interchange, current repositories tend to store content using RDBMS (relational database) technology. XML stores and graph databases are technologies being considered for future to augment RDBMS approaches.

Expression in RDF format (which builds on simple XML representation) is seen as an important additional capability in future. This is seen as one advantage of harnessing standards – in many cases the community for a standard has already developed a recommended expression in RDF.

4.4 Version control and revisions

The approach to versioning has been a major point of debate within the ABS previously. As the systems have grown up at different times, their approach to version control tends to differ.

In general, where there was not seen to be a compelling case for supporting formal versioning past developments tended to avoid that "complexity". Collections, for example, are not currently versioned. Many aspects of change over time for a collection, however, can be handled through descriptions of the "cycle" or the "profile" rather than edits to the main collection document itself.

Under IMT, however, versioning is seen as a prerequisite for active use and reuse of metadata. The structural definition of a metadata object at the time it was referenced must remain accessible even if a new version of that object is defined subsequently. This is consistent with the approach taken in standards such as SDMX and DDI. Both of these standards have a concept of objects being able to be in "draft" mode in which case they should not be referenced for production purposes. The standards do not require versioning of drafts but it is likely that the MRR will support versioning of drafts.

Past debates over when a change is so fundamental that it should result in definition of a new object, rather than a new version of an existing object, remain to be addressed in the IMT context.

Past debates about changes that are so "trivial" (eg fixing a spelling mistake) that they shouldn't result in version change also remain to be finalised in the IMT context.

An example of problems from lack of appropriate support for versioning in current infrastructure is classification system. It could benefit, for example, from the [Neuchatel](#) approach to modelling classifications, versions and variants as well as the [IMT](#) approach to not overwriting previous content.

Within the current system each registered object is essentially an independent entity (ie a "new classification"). It is possible to designate one classification as being "based on" another but this can mean many different things

- The new classification is a new version of the earlier classification and is in some sense expected to supersede it (although possibly not immediately).
- The new classification is a "variant" of the earlier classification defined for a specific purpose. The earlier classification may "live on" indefinitely for the original purpose.
- Classifications are being "grouped" into a "family" without necessarily being formal variants or versions of each other.

Where revisions are to be made (or new versions created) as much impact analysis as possible is undertaken. This includes, for example, understanding what other metadata objects and processes refer to the object that is about to be revised (or versioned) and whether the revision will have any inappropriate impact (whether the new version should be referenced instead). The lack of fully "joined up" registries (including knowing exactly what metadata is referred to in each processing system) makes impact assessments difficult and only partially reliable in some cases.

The MRR and Statistical Workflow Management working together in future should greatly assist in this regard. While existing metadata objects and business processes will be able to continue referencing the present version of an object that is proposed to be updated/versioned, understanding these existing uses and the requirements associated with them

- may assist in designing the new version of the metadata object to best address "whole of business" needs
- will allow the full set of users of the existing version to consider whether they should now use the new version or continue using the present version

The preceding example illustrates the flow on impacts that versioning can have within a complex and actively used metadata registration system. If the existing metadata objects that refer to the object that just got "versioned" now need to refer to the newer version of that object, all those existing metadata objects themselves now need to get "versioned" (because they're pointing to a different version of the first object). All the objects that refer to the objects that referred to the original object now need to get assessed and potentially versioned themselves, and so on with a ripple effect potentially sweeping across the whole registry originating from just one object being versioned. (While standards such as DDI-L support the option of "late binding", they recommend against it for many purposes. Under "late binding" a reference to another object is always deemed to refer to the most recent version of that object – rather than, eg, to the specific version of the object that was current at the time the reference to it was made. "Late binding" reduces precision and leaves open the possibility that the object referred to will subsequently "evolve" in ways that contradict the initial basis for referring to it.)

The [IMT](#) approach supports user decision points (which may be manual or automated) in regard to the "ripple effect" of versioning. It also provides the greatest systematic support for managing initial and "consequential" versioning processes.

4.5 Outsourcing versus in-house development

While external expert consultants were engaged from time to time, the existing metadata systems described in Section 2.2 were all designed and developed "in-house". Open source and other starting points for the Data Element Registry were seriously considered.

ABS (and the Australian Government) ICT Policy and Strategy is placing a greater emphasis on COTS (Commercial Off The Shelf) & GOTS (Government Off The Shelf) based. "Bespoke" software developments (whether through in house development or commissioning of external developers) to deliver all, or part, of a solution is seen as a last resort if other options are demonstrated not to be viable.

From an ABS perspective, however, it remains typically the case that in house staff

- assemble the functional and non requirements that need to be met by a new solution
- identify solution options and assess them against these requirements
- manage the integration and commissioning of the new solution within the existing IT and business environment

Ensuring solutions are consistent with Enterprise Architecture, including Service Oriented Architecture and support for relevant open standards, promotes effective integration (with minimum need to re-engineer other systems), reduces risks of "vendor lock" and facilitates end of life decommissioning (and possible replacement).

The approach to [IMT](#) aligns with these ICT strategies and policies. This includes

- the emphasis on sharing solutions (eg across NSIs and other agencies)
- the preference, where no existing solution is suitable, for collaborative development
 - this includes harnessing, where applicable, existing common frameworks (such as those developed within the European Statistical System) and utilities as "building blocks" for the solution
 - development of new shared solutions is seen as preferable to undertaking purely local "bespoke" developments

While not all developments related to IMT will necessarily deliver, or harness, open source components, open source is recognised as one important paradigm for sharing solutions and sustaining their evolution over time.

In addition to seeking to collaborate with other agencies, ABS is drawing on input from expert consultants to assist developers understand and apply information standards such as SDMX and DDI and to assist in designing key infrastructure such as the MRR.

Development of [REEM \(Remote Execution Environment for Microdata\)](#) is an example of the ABS working with a vendor that shares our interest in harnessing standards such as SDMX and DDI-L. Elements of the REEM solution include

- the ABS developing services to extract metadata from existing systems to populate a DDI description of unit record data
- the vendor "wrapping" existing COTS components to be able to work with DDI and SDMX, and
- components developed specifically by the ABS for purposes such as user registration

ABS implementation, as [ABS.Stat](#), of the OECD.Stat platform is an example of harnessing an existing standards aligned shared solution and entering into a collaborative partnership (with OECD, IMF, Statistics New Zealand and Istat) to maintain and evolve that solution in future.

4.6 Sharing software components of tools

At present, many systems (as described in section 2.2) used by the ABS are built in a "monolithic" fashion (combining the repository, the business logic and the user interface) and are highly customised for the ABS environment (eg they rely on both IBM Lotus Notes and Oracle databases which are configured in a particular way). CMS, ClAMS and the Dataset Registry are all in this category. While there is no in principle objection to sharing these components with other agencies, doing so in practice would be very complex both for the ABS and for the other agency. In any case, as these facilities were developed more than a decade ago and predate relevant application architecture and metadata standards, it is not anticipated any other agency would be interested in making use of these facilities in their current form.

Newer facilities such as the Data Element Registry (DER) and Questionnaire Development Tool (QDT) are architected in a manner that would make it easier to share them. Both of these facilities are designed so that a user interface interacts with the Oracle database via a "Business Services Layer" (BSL). In addition to full sharing, partial sharing could be supported (eg the ABS providing the repository and BSL, with the other agency choosing to develop its own user interface.)

Sharing could be envisaged in at least two forms. One would be the ABS packaging either the full facility or some layers from the facility in a form which allowed another agency to establish a "stand alone" instance. A second form would be extending the BSL (and probably repositioning the repository) so that authorised and authenticated interactions from outside the ABS became possible in regard to the current instance of the facility. One or more external agencies might then act as registration authorities in their own right. This could have many benefits in terms of sharing, and shared development of, metadata content but would be likely to require more thought in terms of ongoing governance and support arrangements.

A third possibility, which physically "cloned" the repository (ie the first option) but supported a unified logical perspective across the original repository and the clone(s) (ie elements of the second option) would also require significant additional work.

While these facilities are deliberately more compartmentalised and self contained in design, they were not developed from the ground up with the intent of sharing beyond the ABS. Some generalisation of ABS specific aspects (eg linkages of both the DER and QDT to collection information from the CMS) would still be required.

The software the ABS has available should be able to be made available to other statistical agencies free of charge in its current form. If the ABS needed to modify the software and/or provide consultancy support in order for that software to be made operational outside the ABS then that work may need to be cost recovered. Alternatively, and preferably, it may be possible to agree a collaborative arrangement such that the existing facility is extended and generalised in a manner that benefits both the ABS and the other agency.

The ABS seeks to avoid becoming a "software house". Any sharing arrangements would be in the context of either one off provision or, preferably, some form of partnership. A relationship along the lines of the ABS acting as a provider to one or more "customers" does not fit with current ABS aspirations and directions.

A number of other ABS applications (eg ABS Autocoder and REEM) are also listed in the Sharing Advisory Board's [inventory of software available for sharing](#).

Short of sharing software itself, the ABS is very happy to exchange details of data models, application architectures, user experiences etc with other statistical agencies.

New developments such as the MRR are being designed to be more readily sharable, in whole or part.

While the ABS has relatively few components currently that other agencies may be interested in sharing, the ABS is placing a very high priority on establishing collaborative partnerships with other agencies to develop new components, or to extend existing modern standards aligned components that already exist outside the ABS.

[Next section](#)

5. Organizational and workplace culture issues

[Next section](#)

5.1 Overview of roles and responsibilities

Initiation of IMTP in February 2010 led to significant adjustment of roles and responsibilities within the ABS.

The 2003 metadata management strategy had stated that, in terms of governance

Metadata management becomes part of every project and each project ensures that they consider and budget for resources to handle metadata development and maintenance.

It is sometimes suggested that by making something "everyone's business" it becomes nobody's business.

The Data Management Section (DMS) within the ABS was to be "consulted" and had a co-ordinating and advisory role. The aim was that the Corporate Metadata Repository (CMR) and its services would be progressively extended to meet the needs of new application developments. DMS developed guidelines to assist project planners, project managers, business analysts and IT staff in understanding the practical meaning and intentions of the principles and how they might apply in the context of a specific project. DMS also provided direct interactive advice to planners, analysts and IT staff.

In practice, however, the design and development of new metadata repositories and services were driven by the initiatives that required them, and paid for them, such as BSIP and ISHS (see [BHM](#)). Given input from DMS, architectural design panels and other sources the designs were notionally left open for use by other projects, and for integration within the CMR, but these outcomes were given relatively low priority in practice.

DMS also continued to fulfil roles it had prior to the 2003 strategy. DMS has been responsible for Data Management Policy within the ABS and maintaining the ABSDB and selected other infrastructure such as CMS and CiAMS described in the supporting page for Section 2.2 of this case study. The maintenance role includes

- managing IT maintenance and minor new developments
- acting as database administrators
- supporting end users including providing content design advice and training

While DMS ensured necessary "repository infrastructure" was provided, and that the infrastructure remained "fit for purpose" in a changing organisational and technical environment, it is not responsible for the quality of the content held within each repository. That responsibility rests with the subject matter areas and others who provide the content and have an ongoing custodianship responsibility, including ensuring the content remains up to date and answering any enquiries its definition might generate from others.

Data Management Policy mandates use of the corporate facilities for various purposes and subject matter areas are responsible for making use of the facilities in accordance with those policies.

The Standards and Classification Section (SCS) has a number of leadership roles in regard to metadata content within the ABS. SCS develop and support "standard" classifications and variables which are cross domain in nature (eg industry, occupation, language). Many of these are recognised standards for Australia as a whole, not just the ABS. SCS also provide guidelines and advice to help subject matter areas ensure their

"collection specific" metadata is well defined and curated.

DMS and SCS form the Data Management and Classifications Branch (DMCB) within the Methodology and Data Management Division (MDMD). DMCB brings together specialists in metadata modelling and systems with specialists in metadata content, in order to reinforce each other's work and to provide strong integrated support to the ABS and the broader National Statistical Service.

With announcement of the IMTP, an early matter to be clarified was the nature of the new program's relationship with MDMD - and DMCB in particular. The conclusion was that IMTP would assume leadership at the strategic level in regard to (Statistical) Information Management. The Program Board for IMT, for example, consists of the head of the ABS and his four deputies. This Board is therefore able to address organisational governance and alignment issues, including in regard to Statistical Information Management, that the approach to implementing the 2003 strategy had been unable to address in practice.

Naturally the IMTP leadership role entails working closely with MDMD. It is recognised, also, that IMTP is leading a transformation process (which, from July 2011, is expected to take at least six years to complete). At the conclusion of that transformation process IMTP is not expected to continue as an organisational unit in its current form. Strategic leadership therefore needs to transition to a sustainable arrangement within the "post IMT" organisational structure.

As described in *IMT*, this leadership role is reflected in activities such as development of the Statistical Information Management Framework, design work associated with the MRR and leadership of the international OCMIMF collaboration. A team of information management specialists, and business analysts specialising in information management systems, exists within IMTP. At the current time (July 2011) this team comprises half a dozen staff. It is expected to approximately double in size during the coming year..

DMS staff have been seconded to IMTP on a rotating basis to assist with its IM work program.

In addition, DMS has notionally divided its work program between maintenance of existing infrastructure (as described above) and supporting IMTP through

- detailed assistance to "pathfinder" projects in applying standards such as SDMX and DDI as well as sound data management practices more generally
- providing input to the IM related projects and initiatives that IMTP is undertaking
- undertaking practical research projects on behalf of IMTP

There are around a dozen staff within DMS currently, with their duties split fairly evenly between maintenance of existing infrastructure and supporting IMTP.

A third key area (beside IMTP and DMCB) is SISD (Statistical Infrastructure and Solutions Design) unit within the technology oriented division of ABS. One role of SISD is to provide technical leadership and support in regard to Enterprise Architecture, including data/information architecture. SISD also leads and supports the "solutions design" process within the ABS, ensuring that new developments (particularly those that are classed as "Architecturally Significant") are designed with due regard to agreed architectural principles and practices. Alignment with the "to be" business and data/information architectures, whose definition is emerging from IMT, is a key consideration in this regard. The "Metadata Building Code" developed by DMS during 2010 currently provides guidance in this regard.

The solution design process culminates with a formal Design Review that comprises senior executives from Technical Services Division, IMTP and relevant business stakeholders. Where an appropriately consultative solution design process has been followed prior to the formal Design Review, however, key "architectural concerns" from various perspectives should already have been identified by stakeholders and addressed in the design proposal. The Design Review should serve as a formal gate to confirm the solution design process has been conducted appropriately, and confirm high level support for the solution proposed, rather than result in fundamentally new concerns being identified.

In addition to these key organisational units (IMTP, DMCB and SISD) there are a range of governance, reference and advisory groups that include participants from across the ABS. These groups assist in steering and informing ABS priorities and directions related to Statistical Information Management.

Phase 5 of *IMT* will focus on extending facilities to support the discovery of and access to data within the NSS. In the meantime, however, the Data Leadership Initiative (DLI) is being sponsored by NSSLB (National Statistical Service Leadership Branch). DLI aims to promote within the NSS best practice standards to help ensure data is 'fit for purpose' for statistical use. This includes best practice in application of exchange standards such as SDMX and DDI. NSSLB is working closely with IMTP and DMS in regard to these aspects of DLI.

The following table contains a list of specialists in metadata management in the ABS:

Name	Role/Position in ABS	Phone Number	Email
Alistair Hamilton	Chief Statistical Information Architect - IMTP	+61 2 62525416	alistair.hamilton@abs.gov.au
Simon Wall	Director - DMS	+61 2 62526300	simon.wall@abs.gov.au
Graeme Brown	Director - SCS	+61 2 62525920	graeme.brown@abs.gov.au
Ric Clarke	Chief Architect - SISD	+61 2 62526736	ric.clarke@abs.gov.au

Marie Apostolou	Director, Statistical Coordination, NSSLB	+61 3 96157500	marie.apostolou@abs.gov.au
-----------------	--	----------------	----------------------------

5.2 Training and Knowledge Management

General training in regard to IMT, including the future for statistical information management, is only starting to be developed and remains at a general level.

Capability building for information management specialists and IT developers in regard to SDMX and DDI has been a priority since the decision of the ABS Executive in October 2009 that these standards will form the core of the ABS's future directions and developments with regard to statistical information management. To date this has primarily been achieved through engaging international experts to present structured courses and workshops. On line learning packages and other training materials developed overseas have also been researched and evaluated, and then utilised where appropriate. It is planned to "train trainers" within the ABS to be able to deliver basic and intermediate (but not necessarily advanced) training in regard to these standards (and their application within the ABS) in future.

Several of the deliverables from the current activities being undertaken by IMTP (eg the Metadata Registry/Repository, the Statistical Information Management Framework) create training needs in order for these outputs to be harnessed appropriately by business and technical users.

In regard to existing infrastructure, DMS provides a range of training.

In addition, a Corporate Metadata Repository (CMR) Assistant is available from the home page of the ABS intranet. This provides a portal to overview and detailed information about the available facilities as well as related policies, guidelines and training courses. It also provides direct access to the facilities themselves by allowing users to click on the component of interest as represented in a high level diagram showing how the various facilities fit together.

As the CMR is "part of the way the ABS does business", the generic training offered by DMS is only one strand. The training about dissemination processes in the ABS, for example, includes information about how content defined in the CMR can be drawn into the various dissemination channels and made available outside the ABS. DMS provides development assistance and input on the components of these training courses that relate to the CMR.

Similarly the corporate "Assistants" related to Business Statistics, to Household Surveys and to Publishing cross reference relevant content from the CMR Assistant where appropriate.

The strategy of presenting information about the CMR in the context of a particular wider business process, rather than trying to present everything about it exhaustively in a major CMR specific training program, appears to work well.

5.3 Partnerships and cooperation

The major partnership specifically related to metadata management, as described under [IMT](#), is the ABS work on the OCMIMF Collaboration with five other NSIs.

Given the ABS Executive decision in regard to SDMX and DDI in October 2009, the ABS has a strong interest in how effectively and efficiently these standards work together, both currently and into the future. The ABS is therefore a very active participant in the [SDMX/DDI dialogue process](#) which also engages the two standards bodies together with a number of other NSIs and international agencies.

While developments such as the MRR and the Statistical Information Management Framework are not formally structured as collaborative projects, plans and experiences in regard to them are shared at an informal working level. Informal interchange is particularly common with agencies that are undertaking similar developments which harness SDMX and DDI working together.

More generally, the ABS is very keen to share information and experiences and to collaborate within METIS generally as well as on a narrower (eg bilateral or "working group") basis.

At a national level, ABS is undertaking a number of metadata related projects in conjunction with [ANDS \(Australian National Data Service\)](#). The primary focus for ANDS is infrastructure to better support the data management and access requirements of researchers. Public Sector Information, including statistical information from the ABS, is a key information resource of interest, and value, to the research community.

The National Statistical Service (NSS) provides many opportunities for other collaborations. These include working with State and Territory Government agencies that are undertaking major data related initiatives as well as working with sector specific initiatives (eg the Australian Transport Data Action Network) that span agencies at the State and Territory as well as the Australian level. NSS initiatives take the ABS beyond simply collaborating with other statistical agencies and into collaborating with other metadata communities, such as the geospatial community, the research community, and others.

One collaborative project, for example, with a state government agency and the university sector involved developing "injectors" for technical metadata about usage rights under the Creative Commons framework. The software allowed information on usage conditions to be "injected" into spreadsheets and other products so this information remained associated with the content even after it had been downloaded from the web. The Creative Commons organisation itself has now expressed interest in assuming responsibility for ongoing custodianship and development of the software.

5.4 Other issues

Over the past 15 years the term "metadata" has become common parlance within the ABS. The value and importance of metadata is widely recognised.

There is also a degree of disappointment, frustration or scepticism expressed in some quarters because more progress hasn't been made more quickly and we haven't yet made metadata simple to manage and maintain as well as "all powerful" in driving and describing all processes and outputs. The vision expressed in the 2003 metadata strategy to some extent fostered expectations that were unable to be met during the subsequent years of implementation. Questions have been raised in regard to what is different about IMTP which will allow larger scale success this time.

As illustrated elsewhere in this case study, however, the corporate positioning of, and support for, IMTP is incomparably stronger than the positioning for implementation of the 2003 strategy. The profile of IMTP has led to much more active business (including senior executive) engagement from across the ABS in shaping the IMT strategy and its expression. It is a corporate initiative, driven by business strategy and requirements, rather than an initiative driven (in reality or in terms of common perception) by IT and/or IM specialists. In addition there are enablers (eg mature standards and technologies) capable of supporting IMTP that did not exist eight years ago. Partly because of these enablers, great strides have been made within the wider community of producers of official statistics which mean IMT is able to harness collaboration, and shared solutions, in a manner that was not possible in 2003. In addition, IMT learns from past ABS experiences in this field - and the experiences of other agencies.

The fact the term "metadata" is so widely used, in a variety of valid but different contexts, is emerging as an issue. The focus on IMT on "statistical information" rather than specifically "data" or "metadata" is seen as an advantage in this regard. At a minimum most references to "metadata" in discussions within the ABS, or outside the ABS, require clarification of which type(s) of metadata is being referenced.

Being primarily aware of low level technical examples, some managers are unsure why metadata should be considered a strategic business challenge and enabler within the ABS rather than a purely technical matter. Once again, a focus on "statistical information" (which is the core business of the ABS as an NSI) can be helpful.

Similarly there is frequent confusion between "metadata concepts, models, systems etc" and "metadata content". It is challenging to promote a message that investment in well designed and integrated metadata infrastructure is a necessary, but not sufficient, condition for achieving consistently high quality of metadata content. Senior managers have tended to have unrealistically high expectations of what will be delivered - which would lead to disillusionment if not addressed in advance - or else their expectations are so low that they are unwilling to commit resources to the effort.

Very significant challenges arise from the fact staff often enjoy the challenge, and receive satisfaction, from developing definitions, structures, frameworks etc from first principles. They often also find it hard to resist the temptation to "tweak" the wording of a definition, the details of a structure etc that they already recognise as basically fit for purpose but which they believe could be improved upon slightly for their specific purpose. This can be seen as part of a culture of "local optimisation" rather than "global optimisation". A series of poorly integrated local optimisations, however, may result in an inefficient, sub-optimal end to end business process. In addition, a diversity of "locally optimised" processes/systems across the organisation typically proves very hard to sustain over time.

Seeking of "local optimisation" by employees can be linked to a sense of professionalism and pride in their work. It is vital not to undermine the latter two when seeking to address the former. Aiming for "local optimisation" also tends to be simpler than seeking global optimisation.

It is also the case that simple reuse isn't always the answer. Sometimes local divergences are appropriate even when viewed from a wider perspective. The trick becomes identifying when this is the case. Such cases typically require "designing the divergence" such that re-use of existing concepts and content is maximised, with the divergence being only to the extent required. This becomes a difficult balancing act. There is a temptation to revert to "starting with a blank slate" as soon as it becomes apparent re-use will not be simple.

As illustrated in the preceding two paragraphs, there is scope for the aim "think globally, act locally" to create even more challenging and satisfying roles for staff, but the extent of the cultural change required to reach that point appears daunting.

Exactly the same "local optimisation" issues described above in regard to subject matter staff reusing metadata structures and content have been observed in terms of programmers re-using existing services as part of Service Oriented Architecture.

[Next section](#)

6. Lessons learned

6.1 Lessons Learned

1. While technology is a vital enabler, metadata management should be driven, governed and presented as primarily a business issue rather than a technical issue.
 - This requires proponents of metadata management focus first on business outcomes and benefits (eg improved productivity, increased utility of statistical outputs) rather than on metadata management itself.
 - Metadata management as a topic in its own right is of interest to very few, and is typically viewed as a technical specialisation. Its potential as one (of a number of) means to achieve business process improvement is generally acknowledged. A key challenge is to demonstrate, in practical terms meaningful to business areas, that it should be one of the preferred means - and one that is supported through investment and through business practices and culture.

Within reason, the less the term "metadata" and the names of various metadata standards are used in discussions with senior management and business areas the better - the focus should be on what will be different, and what outcomes will be achieved, from a business perspective.

- "Metadata projects" should not be designed and promoted as "IT system developments" but rather focus on the development and deployment of new and improved capabilities, business processes etc. Such projects will often include new or extended IT systems but they should not be "about" IT systems. Among other drawbacks, narrowing the focus to IT systems will mean business areas - at best - see themselves as relatively remote stakeholders with some interest in the results of the project rather than feeling they are active participants with direct roles to play in ensuring the success of the project.
2. All high level organisational units need to be engaged by the metadata management program and have defined responsibilities in relation to it.
 - Some units' primary responsibilities may simply be to contribute to corporate sign off on the objectives, strategies, policies and high level design of deliverables (systems and processes) and then to acceptance test, take up and apply the outputs in an agreed manner to contribute to the achievement of the corporate outcomes sought from the project.
 - Other units will have a much more extensive role in terms of leadership, co-ordination, business analysis, design, development, implementation and ongoing management of systems and processes.
 - If only a few specific organisational units are seen to have a direct stake in the project then it's much less likely to achieve overall success.
 3. It's become more and more apparent over time that applying externally recognised and supported standards, in regard to design of data models for example, has a lot of benefits - including as a means of building upon a wealth of intellectual efforts and experiences from others.
 - At the same time, application of standards must be driven, and moderated, by the organisation's particular context and needs. The underlying effectiveness of the infrastructure should not be sacrificed in favour of complying "to the letter" with a standard, although the business case and the management arrangements for any divergence need to be defined and agreed.
 4. In addition to developing and deploying infrastructure, a metadata management project should be understood, and managed, as a "cultural change" initiative for an organisation. Metadata management aims to make information explicit, visible and reusable (in whole or in part) - with these aims requiring a somewhat standardised and structured approach. This can be a "culture shock" for some business areas who are used to operating in a more autonomous and self contained and often a less structured manner.
 - It needs to be acknowledged there were sometimes benefits from the former approach and there will be some overheads with the changed approach. If at all possible, however, it needs to be demonstrated - or at least plausibly posited - there will be net positive benefits in practice (not just notionally) from the changed approach even if, eg, many of those benefits accrue over time - rather than being immediate - and/or accrue to users of the content rather than producers of the content.
 - Sharing and re-use can lead to concerns about loss of absolute "control" over metadata. It is important to ensure practical processes and governance around content use and change management (eg stakeholder consultation, ease of resolution/divergence if a required content change isn't tenable for one of the users of the existing content) address legitimate concerns in this regard
 - Note also the cultural change aspects discussed in Section 5.5 in terms of moving from "local optimisation" to a paradigm of "global optimisation".
 5. Sufficient attention needs to be focused, by the project team and by other areas, on ensuring the metadata management infrastructure (systems and processes) is fully integrated with other business processes and IT infrastructure rather than being a "stand alone" development.
 - This needs to be factored in from high level design onwards. This, in turn, requires that as part of the initial requirements gathering, analysis and sign off phase there is detailed attention focused on practical matters related to implementation, including uptake and ongoing use as part of end to end business processes.
 - This is also a reason why acceptance testing of deliverables by a fully representative selection of the business areas expected to use them is essential. The aim of this testing is not so much to confirm the specifications have been implemented faithfully (detailed system testing should already have been completed) but that the results meet practical business needs, including integrating with other workflows and systems and meeting performance and other usability requirements.
 - "Acceptance" testing should mean just that. If (for whatever reason) what has been delivered is not yet at a stage where it is fundamentally fit for purpose from a business perspective then it should not be deployed in its current form. (On the other hand, if the deliverable is imperfect but basically "fit for purpose" then the remaining issues may be held over to be addressed in a later release.) The phase ends either with business agreement the deliverable is fit for use within the broader production environment - possibly with some caveats - or else no release occurs.
 - Sound project management and engagement with business stakeholders in earlier phases should minimise the risk of failure at the Acceptance Testing stage. That said, it is counterproductive for all concerned if software that is not fit for purpose is forced on business areas.
 6. In addition to allowing sufficient time and resources for the business analysis, design and development process it is crucial there is sufficient resourcing focused on
 - implementation of the new infrastructure
 - includes training, best practice advice and technical troubleshooting support for business users
 - maintaining and upgrading the infrastructure as business requirements, and as other elements of the IT environment, evolve over time
 - co-ordinating and promoting "outcome realisation" from the investment
 7. Business areas must be able to engage with implementation processes.
 - In many cases there may need to be scope for business areas to negotiate and agree (not decide unilaterally) short term or longer term exemptions from, or variations on, the standard implementation process.
 - Exemptions and variations should be actively managed and reviewed with the aim of achieving convergence over time wherever practicable
 - Metadata systems should clearly identify preferred and non preferred definitions and structures, so that - wherever practicable - areas with a need to diverge from standard practices and definitions remain "within the system" while at the same time those practices/definitions are clearly identified as non preferred.
 - Feedback from business areas needs to be able to influence the details of the implementation process. For example, if it appears too many exemptions and variations will be required it may be that the design of the implementation process doesn't properly

reflect business needs and realities.

- If business areas are not provided with a genuine opportunity to "work with" a change process they are more likely to covertly "work around" that process in a manner which undermines the business objectives of the change.

8. Metadata management is largely about connections of various forms, such as

- between documentation of agreed processes, methodologies, definitions and structures and what happens systematically
- between producer and consumer perspectives on statistics
- similarities and differences between different sets of data, different structures and definitions etc

9. Due to the wide variety of roles it must perform, and perspectives it must support, there is not one particular structure/format for metadata that is, in itself, ideal for all purposes.

- The key appears to be modelling and managing metadata in way that can support the different views, and preserve the integrity of the connections underlying these different views.
 - The ideal appears to be a relatively simple, robust, standards aligned but highly extensible core model, together with well defined and managed means to map and transform locally required metadata into and out of that core model and, where necessary, to define, manage and integrate local specialised extensions to that common core.
 - A single central metadata model that aims to span all content for all purposes is likely to be too complex, too unwieldy and too static.

10. "Statistical metadata management" is increasingly expected to interoperate with metadata management as practised in other communities (eg geospatial, academic/research) and sectors (eg use of XBRL by businesses and by regulatory agencies). This provides a huge opportunity (as well as a challenge) in being able to efficiently and effectively open up and harness (from statistical and other perspectives) a vastly increased suite of information resources. It also provides a practical affirmation that other communities and sectors recognise the value of metadata and standards, although because their primary purposes vary the details of their schemas and standards also vary.

- This reinforces the previous point. It appears both impractical and undesirable to establish a single approach that supports the primary purpose of each different community and sector. On the other hand, statistical agencies are strategically placed to provide a simple core that might be used to bring together information across communities/sectors and to exchange it across them.
- It also reinforces the value of international standards and collaborations. Most of the community and sector specific standards are internationally based. Rather than each NSO needing to work out mappings "from scratch" there is a lot of opportunity to share a core of analysis and mapping between NSOs.