1. Introduction (Australia)

1.1 Metadata Strategy

The document entitled *A Brief History of Metadata (in the ABS)* attached to the Case Study (referenced simply as *BHM* hereafter) provides information on the evolution of ABS strategies related to metadata over time.

As described in *BHM*, the ABS Metadata Strategy has evolved over the past three decades. It was formalised in an 18 month process, involving stakeholder consultation across the ABS, which culminated at the end of 2003 with the *Strategy for End-to-End Management of ABS Metadata* being reviewed, and broadly endorsed, by the ABS Executive. More information on the details of that strategy is provided below.

While the formal strategy document from 2003 hasn't yet been updated, the actual strategy employed by the ABS has evolved considerably over the past few years. The 2003 document was a milestone in the evolution of corporate thinking and planning in regard to metadata management, but to some extent it only represents a "snapshot" of thinking at a particular point of time.

As described in *BHM*, the ABS has now embarked on the path towards another milestone in terms of a 2020 Vision. This is expected to be drawn together during 2009 and then provide a platform for strategic planning over coming years.

Ultimately any ABS metadata strategy exists to support the ABS mission and objectives as set out in the organisation's corporate plan.

http://www.abs.gov.au/website dbs/d3310114.nsf/51c9a3d36edfd0dfca256acb00118404/b1042c4ee5af9c71ca256a46008278d9!OpenDocument
In particular, the availability of appropriate metadata and the application of sound metadata management practices are critical to supporting informed use of statistics and the quality of the statistical services we deliver to the nation.

The twelve principles defined as a cornerstone of the 2003 strategy continue to be applied within the ABS:

1. Manage metadata with a life-cycle focus
2. All data is well supported by accessible metadata that is of appropriate quality
3. Ensure that metadata is readily available and useable in the context of client's information need (whether client is internal or external)
4. Single, authoritative source ('registration authority') for each metadata element
5. Registration process (workflow) associated with each metadata element, so that there is a clear identification of ownership, approval status, date of operation etc.
6. Describe metadata flow with the statistical and business processes (alongside the data flow and business logic).
7. Reuse metadata where possible for statistical integration as well as efficiency reasons (no new metadata elements are created until the designer/architect has determined that no appropriate element exists and this fact has been agreed by the relevant 'standards area')
8. Capture at source and enter only once, where possible
9. Capture derivable metadata automatically, where possible
10. Cost/benefit mechanism to ensure that the cost to producers of metadata is justified by the benefit to users of metadata
11. Variations from standards are tightly managed/approved, documented and visible
12. Make metadata active to the greatest extent possible

The strategy proposed the twelve principles be applied when planning and authorising all ABS projects that provide, and/or make use of, metadata management capabilities, even those where metadata management is a secondary rather than primary objective or requirement.

Other key points in the 2003 strategy include:

- There is an agreed conceptual metadata model which is linked to processes that are part of the statistical processing cycle and this linkage is used to determine what metadata should be collected.
- The ABS metadata model takes account of and uses international standards where possible.
- The physical implementation of the metadata model is the Corporate Metadata Repository (CMR) which is used by all ABS projects. It consists of a number of shared physical databases.
- All metadata entities are managed by a 'registration authority'
- Roles and responsibilities are identified
- Data Management and Classifications Branch (DMCB) is responsible for coordination, definition and maintenance of metadata policies, procedure, systems and provides advice and consultancy to developers related to metadata matters.
- DMCB is the 'registration authority' for the CMR and ensures that other organisational units with this role for particular metadata entities understand that role, are trained and have relevant tools.
- Metadata management is part of every project and should be considered alongside resource allocations and accountabilities in the same way as business processes and data flows are considered.
- Governance of metadata management developments and the oversight of outcomes realisation is vested in line management, existing project and program boards with ABS Executive taking an ultimate corporate view.

1.2 Current Situation

_BHM_ describes how the current situation has evolved within the ABS.

The majority of data collection and input processing activities for business and household surveys are moving toward implementation of high level metadata frameworks that are informed by ISO/IEC 11179. These frameworks were developed over the past seven years and postdate the ABS specific metadata framework which was implemented for the corporate output data warehouse which was developed during the 1990s. This creates a looming challenge for end to end metadata management within the ABS. The ABS response to this issue will be thought through further as an outcome of the 2020 Vision process.

Key elements of current metadata infrastructure 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 recent development based on the metamodel found in ISO/IEC 11179 Part 3.

• questions and question modules
  ° This was developed recently for household surveys with an aim to generalise the facility in future.

• collection instruments
  ° This was developed 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 hoped that a common high level metadata registration framework can be implemented across the ABS for all classes of metadata. (This does not imply that all classes of metadata undergo exactly the same registration processes, but that the processes for each class of metadata are 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.

Further information in regard to the NSS is available via


Understandably, metadata is a particular area of focus for the NSS. This includes a simplified and generalised set of principles.


The 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, will shape the upcoming 2020 Vision process.

Next section

2. Statistical metadata systems and the statistical business process

Next section

2.1 Statistical business process

While a few areas insist on their own variations on the following theme, the following diagram is affectionately known as "The Caterpillar" within the ABS.
A strength of The Caterpillar is that it highlights the activities which take place throughout the cycle (eg managing quality and processes, including managing metadata) as well as six main steps within the body of the caterpillar, and the "linking" steps at the beginning and end which open and close the cycle.

The Caterpillar was developed as part of the Business Statistics Innovation Program (BSIP) launched at the dawn of the new century as described in BHM.

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. (Due to legitimate differences there is usually not just one single practice that is best for every survey. Typically a limited set of leading practice models are identified for each step in the cycle from which one can be selected depending on the specific needs and nature of the survey. This is far preferable to 50 different surveys choosing 50 completely different practices for each step.)

In terms of numbering from the GSBPM:

1. Needs emerge from the Statistical Leadership aspect of the caterpillar (possibly having been fed back from Evaluate and Tune). Needs can be clarified in detail during the Design and Tune phase (eg consulting with an external stakeholder group on design issues).
2. Design very much equates to the Design and Tune arrow at the start of the caterpillar.
3. Build is also largely in the Design and Tune arrow. Within the ABS currently, partly due to a lack of metadata driven systems, "design" and "build" often tend to go hand in hand (eg for collection instruments). Also, many business surveys - which informed the original design of the caterpillar - are conducted very regularly so the actual amount of "design and tune" (and build) effort for each cycle of the survey is very limited compared with the amount of work which then occurs in subsequent steps in the caterpillar. Separating out Build from Design in the GSBPM, however, better reflects where the ABS seeks to go in the future.
4. Collect largely equates to Acquire Data.
6. Analyse in the GSBPM includes some aspects of Transform Inputs Into Statistics in the ABS model (eg seasonal analysis, macro editing). It also covers Analysis and Explanation in the ABS model. The difference in the ABS model is that Analysis and Explanation commences once data is "finalised" (unless an anomaly is detected in the phase that causes earlier work to be redone) rather than including finalising some aspects of the data (eg producing seasonally adjusted and trended estimates).
7. Disseminate in the GSBPM largely equates to Assemble and Disseminate and stretches into Decision Support in terms of helping clients make use of the content we have disseminated and answering their questions.
8. Archive in the GSBPM is underdone in the Caterpillar. In reality it is covered in a low key manner as longer term aspects of Assemble and Disseminate and Decision Support as well as part of data management policy under Manage quality and process at the bottom of the diagram. In the context of sharing data on a sustainable long term basis within the National Statistical Service (NSS), however, its relevance to Statistical Leadership is increasingly being recognised. The explicit reference in the GSBPM, therefore, makes sense in the ABS context.
9. Evaluate in the GSBPM corresponds with the Evaluate and Tune arrow in the Caterpillar.

2.2 Current System(s)

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

- Some are fully corporate. The main examples of these are described briefly below.
- Some are "shadow systems" which extend corporate systems to supplement the standard content with attributes of local interest.
° Making the corporate systems more readily "extensible" would help to address this issue, as would an enterprise architecture that makes it easy to marry up "local" low level system/context specific metadata with "corporate" metadata.
° Some of the "shadow systems" have been designed and maintained to ensure they can be easily reintegrated with the corporate system in future while others have not.
  • Some are 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.

Collection Management System (CMS)

This manages high level information about "statistical activities" ("collections") undertaken by the ABS. These "statistical activities" include surveys, censuses, statistical analysis of administrative data sources and statistical "compilation" activities such as preparing the national accounts.

The basic definition of a "collection" suitable to be registered in CMS involves inputs, processing/ transformation and output. Simply collating data from other collections, therefore, results in a new "product" rather than being a new "collection" in its own right.

Each collection may have many instances (cycles) - such as a monthly survey. Information can be recorded at the collection, cycle or an intermediate level called "profile". (One purpose of the "profile" level is to document small to medium "redesigns" and other changes that can occur over time within a collection.)

Many (but not all) end to end processing systems do refer to the Collection ID and Cycle ID based on the registration of the relevant activity to CMS. This provides a good starting point in terms of end to end "metadata glue" and means the corporate registry function of CMS is being used relatively actively.

As a repository for descriptive information about statistical activities undertaken by the ABS, however, it sits to one side of the processes themselves and the content is often of relatively poor quality to start with and then poorly maintained over time. This is despite the fact that managers of these activities are asked to sign off on CMS content. Much of the content visible through CMS, therefore, cannot be relied upon as an accurate, up to date description of activities in the ABS.

A subset of this content is signed off to the ABS website to become visible in the ABS Directory of Statistical Sources.


This disseminated content does tend to be better (but not perfectly) maintained.

CMS also hosts "Quality Declarations" that have started being disseminated alongside ABS data in recent months. For example, see


The basic design of the CMS dates back to the 1990s although it was updated to Version 5 in 2001. It's structure for describing activities doesn't correspond to the ABS Caterpillar that was developed subsequently. Also

    • more of the information entered in CMS should be actively driving actual business processes rather than being "passive" independent documentation, and
    • more of the content visible through CMS should be sourced from other stores of actively used metadata.

A redeveloped CMS might also, for example, be aligned with the top level modules ("Group", "Study Unit", "Data Collection") associated with DDI (Data Documentation Initiative) V3.

Redevelopment of CMS is recognised as a priority, but not imminent.
Dataset Registry

This is a widely, but not universally, used registry for defining "dataset" metadata associated with a specific unit record file, data cube etc. This metadata includes

- the set of individual "data elements" included within the dataset
- where the data is stored and how it is structured (eg field names)
- what business unit owns the dataset, when it was last updated etc
- what statistical activity (collection) produced the data

This catalogues all available "output" datasets within the ABS and assists in their management including long term retention.

Some systems working with data in specific environments have their own dataset registries, which includes structuring "dataset" metadata in somewhat different ways. Extending the corporate registry to integrate with the definition and management of "input" and "intermediate" datasets would be of value in an end to end context including being able to trace metadata usage within the ABS. (Querying the metadata model currently allows us to know, for example, which output datasets make use of a particular classification but not which input or intermediate datasets might do likewise.)

The main corporate register dates back to the 1990s and the characteristics of "data elements" recognised within its model are not fully harmonised with ISO/IEC 11179 although the differences are not monumental. This is another driver for updating the model underpinning the registry, in addition to the need to extend that model to better support definition and management of input and intermediate datasets.

While extending and updating the register is desirable it is not imminent. The issue may be "forced", however, when the ABS starts trying to "join up" the IDW and ISHS based data collection and input processing developed during recent years with output processes operating in an environment that currently dates to the 1990s. (See BHM for more details.)

Classification Management System (ClaMS)

This is another system that largely dates back to the 1990s. It features a "pre Neuchatel" ABS developed model for classifications. As infrastructure it is used relatively widely (although not universally) in end to end statistical processes within the ABS. For example, in addition to being used universally as part of the defining metadata for output datasets, these classifications can be linked into metadata definition for

- the Input Data Warehouse
- processing of Household Surveys
- driving aggregation, estimation and consequential confidentialisation processes
- driving the layout of publication tables
  - eg indenting labels according to the depth of the classification item in the classification hierarchy
- labelling and describing time series
  - eg based on the classification item labels associated with each dimension of the "key" for that particular time series.

While quite useful for many systematic purposes, the current system is very weak in terms of enforcing rational reuse of classifications across the ABS. For example, while a business area might define their own version of a classification and use that version more or less on an end to end basis, they are unlikely to reuse a classification defined by another area. This is because

- it is relatively hard to find existing classifications that would be structurally suitable to be reused for the area's purpose(s)
- it is relatively easy for areas to define new classifications that meet their required specifications
- areas like to exercise full control over "their" classifications rather than being dependent on other management processes

In addition, ClaMS does not properly support the following

- detailed definitions (as opposed to labels) for individual classification items
- item by item mappings from one version of a classification to another version of the same classification
- item by item mappings from one classification to another
"special" concepts such as "cut off values" used to translate continuous variables to categorical codes

At the same time, however, the levels of sophistication and complexity of classifications which can be supported within ClaMS can make it "overpowering" for users who have very simple and basic requirements.

It should be noted, also, that ClaMS is sometimes used for defining lists (eg of valid values) rather than only "proper" classifications.

Redevelopment of ClaMS is recognised as a priority, but not imminent.

Data Element Registry (DER)

This is a newly developed ISO/IEC 11179 based facility which replaces a number of older "Data Item" systems.

It has been developed using a "services architecture". At the core is a repository of data elements and their building blocks (eg object classes, properties, value domains etc). There are then low level Create, Read, Update, Delete services which are in turn called by a higher level "business based" service layer. A generic user interface is supplied for the DER but it is expected that most users will be interacting with the DER as part of more general "business workflow level" metadata assembly (including reuse) tools that will work with data elements in combination with questions, question modules, collection instruments etc rather than in isolation.

The first main "take up" of the DER will be via the Questionnaire Development Tool (QDT) developed as part of the ISHS project. (See BHM for more information). The second main "take up" is expected to relate to the Input Data Warehouse associated with business statistics. This means that the first uses of DER will be at the "input" end of the statistical cycle, but full end to end utilisation, including support for dissemination requirements, is expected in future.

In addition to the "data element" repository component based on the ISO/IEC 11179 Part 3 metamodel, the DER includes a more general "metadata registration" component based on ISO/IEC 11179 Part 6. The latter has been designed to be able to be separated out as a register and set of services in its own right which could support registration and management of metadata "objects" that are outside the Part 3 metamodel (eg questions, "collections", "collection instruments", datasets). This separation is likely to occur (at a logical level, if not a physical level) to support the rolling out of a common high level framework for metadata registration across the ABS.

Questions, Question Modules, Collection Instruments

The ISHS project for household surveys has developed new metadata repositories and associated services related to the above, as well as making use of the new corporate Data Element Registry and the existing Collection Management System.

While the actual development work on these repositories and services to date has concentrated on household survey requirements, the high level design and IT architecture has been selected with an expectation that these repositories will be generalised and "corporatised" in future even if the higher level business services and workflow interfaces developed as part of ISHS, which currently interact with these repositories, remain specific to household survey processes.

Analysis to date suggests that some extensions to the repositories and services will be required to support business statistics and other corporate uses but this should not impact existing use by household surveys.

The infrastructure developed by ISHS is only now in the process of being "commissioned" for actual use by household surveys so it is possible there will be some further refinement to the repositories and services for that purpose prior to any thought of wider "corporatisation".

The initial use of these repositories and services will focus on survey development and input processing but full end to end utilisation, including support for dissemination requirements, is expected in future - first by household survey processes and then more generally.

Quality Infrastructure System (QIS) and Business Activity Monitoring (BAM)

Both of these systems, recently released to production, store metrics on how statistical processes are performing (eg response rates, imputation rates, edit rates etc) and support reporting and analysis
related to these metrics. This data about the outcomes of processes can be termed "operational metadata" or "paradata" within the ABS. It can be useful for internal monitoring, management and tuning of processes as well as generating data quality indicators for external dissemination.

These systems rely on individual processes being "instrumented" to write relevant metrics to the QIS or BAM store. (QIS is informed more by the IDW "business statistics" data model and BAM by the household surveys approach.) This allows for progressive uptake.

At the moment the metrics recorded in QIS and BAM tend to relate to early stages in the statistical cycle but both are designed to be able to accept metrics from later in the cycle.

**Process Metadata**

Some early conceptual and exploratory work has been done in this area but, as yet, no major design work as occurred. Seven types of "process metadata" were identified in this early work, 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.

The simple SDMX package related to process definition was also reviewed during 2008. It was considered to provide a fundamentally sound "core" structure, although one which would most likely need to be extended through use of a specialised language/standard for defining and managing business processes (eg BPEL, XPDL) if metadata were to fully drive process and workflow execution in practice.

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.

The ABS is currently exploring commercial business process management software offerings, including the extent to which they can be metadata driven. If the business case is proven it is expected the ABS will progress to acquiring and implementing a business process management system. If a decision is made to proceed then the practical focus on process metadata is expected to increase greatly.

### 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 are usually estimated when developing business cases. While much better than a vacuum for planning purposes, past experience suggests these cost benefit analyses are usually not borne out in practice. Often this is because decisions are 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.

None of the major developments are currently at the "business case" stage - they are either not yet at that stage or long past it - so current "business case" estimates are not available.

### 2.4 Implementation Strategy

This question can be viewed from several perspectives. At least 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 became "opportunistic" and "incremental". There was a broad "master plan" of what should exist in the longer term, but individual "construction projects" were much more modest in scale.

The 2020 vision process on which the ABS has now embarked may move the balance back toward the centre.

At another level, 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.

A number of reviews in 2008 highlighted the extent to which - notwithstanding sound outcome realisation plans - the complex jumble of existing processes and infrastructure, which lacks transparency and includes many inconsistencies in its underlying information models, is a major barrier both to major change and to progressive evolution in statistical business processes and in supporting infrastructure. Different processing steps and environments are often joined together through locally designed and maintained "glue steps" which reformat and/or restructure content (data and metadata) to make it suitable for the next step. (For example, if ten different statistical collections perform Process H after Process G there might be ten different customised "glue steps" between G and H - one for each of the collections - rather than a single logically designed "bridge" between the two processes.) "Glue steps" are typically "hardcoded" and often become "black boxes" after the local programmer who built them moves on. However if, for example, Process G changes then the glue steps also need to change (eg F->G and G->H) and there can be further downstream and upstream ripples as well. This adds very significantly to the levels of cost and risk associated with change, creating inertia. This is a major reason why the SESAME Framework described in BHM, or something similar, is seen as a crucial enabler for sustainable change, including implementation of change, in the future.

3. Metadata in each phase of the statistical business process

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.

The primary categorisation in use now 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 uses 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). While the ABS could establish a list of all the high level metadata objects we currently recognise, we wouldn't necessarily recognise a particular list as containing all of, and only, the high level objects that ever should be recognised by any statistical agency.

3.2 Metadata used/created at each phase

The ABS is an agency that aspires to achieve end to end definition, management and reuse of metadata. Section 2.2 records the extent to which we have achieved this so far in regard to our major corporate metadata systems.

While indicative rather than exhaustive, the following diagram sets out ABS aspirations in this regard as captured in a briefing paper from 2006.

![Diagram showing metadata aspirations]

3.3 Metadata relevant to other business processes

The 2003 metadata strategy defined its scope as relating to "statistical" metadata (rather than all the metadata potentially relevant to any aspect of ABS operations). The scope was still broad, however, because some of the metadata required in order to perform core statistical operations may not be thought of as "statistical" in nature.

Briefly exploring some of the borderline cases, the operational metadata (paradata) about statistical processes can be (and is) used for making financial planning and prioritisation decisions. For example, the financial implications of increasing sample size, increasing the length of questionnaires, accepting reduced response rates, raising the threshold for "significance" editing etc can all be gauged better, together with the likely statistical benefits/costs. This can help set priorities for expenditure, or for areas where savings can be reaped.

On the other side, "administrative" information sourced from the ABS "Corporate Directory" about individual staff members, individual positions (which might be temporarily occupied by one person while another is absent), business units, corporately defined "roles" etc is used extensively by statistical systems - including metadata systems. This may be used, for example, to determine who is currently in the set of people who have the right to edit, approve or otherwise manage a particular piece of metadata.

An intersection is early work on a proposed Statistical Content Ownership Framework (ie ownership of data and metadata). This recognises that organisational units change over time, so assigning ownership of individual content to a particular business unit can create maintenance headaches and/or responsibility headaches over time. The idea is to assign ownership/custody for particular data and metadata holdings to particular subject matter based "domains", possibly together with some additional specialised "methodological" domains related to particular concepts, methods and other artefacts. We anticipate these domains should be more stable and enduring. We would then map these domains to the current organisational structure.

While it is not currently the case, it is possible this could in turn feed into non "statistical" activities such as cost recovering the space used to store data in a particular system related to a particular domain.
4. System and design issues

4.1 IT Architecture

Unless otherwise noted, this section refers 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
- 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 much 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. A highly successful example of implementing an SOA based metadata management environment would be of very high value as a case study for the ABS.

During 2008 it has become 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 go 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.

This is being reflected in an update to ABS Enterprise Architecture, which currently recognises the “standard” business, data, applications and technology architectural perspectives.

4.2 Metadata Management Tools

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

As noted in 1.2, it is hoped a coherent high level corporate reference model and a set of supporting facilities (eg SDMX based data and metadata structure definitions and structure mappings between preferred and "legacy" structures and concepts) might assist in this regard in future.

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

The standards and formats currently in use for the major metadata repositories, together with those we hope to use in future, are described in Section 2.2.

4.4 Version control and revisions

This tends to be a major point of debate within the ABS. As the systems have grown up at different times, their approach to version control tends to differ. The most recent major debate has been in regard to the new Data Element Registry.

In general we are now favouring the general approach to versioning set out in ISO/IEC 11179 Part 6. That standard, however, still leaves a lot of flexibility available to the relevant Registration Authority for a particular registry in terms of how versioning will be applied.

In general, where there is not a compelling case for supporting formal versioning then that complexity is avoided. 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.

The current classification system doesn't handle versioning well and could benefit from the Neuchatel approach. Currently 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 versioning does need to be supported, careful attention needs to be given to defining cases that don't result in new versions ("trivial changes") and cases that must result in whole new objects (ie the change is so fundamental the new object is no longer a "version" of the old object).

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 preceding example of impact assessment in the case of versioning 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, then all those existing metadata objects themselves now potentially need to get “versioned” (because they’re pointing to a new version of the first object). All the objects that refer to the objects that referred to the original object now need to get impact 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. The ABS hasn’t yet resolved this issue.

4.5 Outsourcing versus in-house development

While external expert consultants have been engaged from time to time, the 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. It is expected open source and other collaborative options will increasingly be selected in future, although that is different to complete outsourcing. At a minimum, interoperability between new repositories deployed within the ABS and other relevant "external" repositories, standards and vendor software solutions will be an increasingly important consideration.

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.

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.

Given many ABS metadata system are more than a decade old, and the organisation is now moving toward a more modern, open, standards aligned, information and application architecture, 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.

4.7 Additional Materials

None are supplied at this stage but it is likely that additional information can be made available on request.

5. Organizational and workplace culture issues

5.1 Overview of roles and responsibilities

Realisation of the objectives of the 2003 metadata management strategy, and upholding and advancing the principles set out in it, remains a responsibility shared across the ABS.

As upholding and advancing the principles was seen particularly as the responsibility of every new project within the ABS, the project planners, project managers, business analysts and IT staff associated with these projects had a particularly important role. Data Management Section (DMS) developed guidelines to assist such key people in understanding the practical meaning and intentions of the principles and how they might apply in the context of a specific project. DMS also provides direct interactive advice to planners, analysts and IT staff.

DMS was also assigned the lead role in terms of co-ordinating the development of specific metadata management infrastructure and ensuring this infrastructure fits together as part of a logically integrated Corporate Metadata Repository. It has the lead role in monitoring overall progress in regard to the strategy and identifying areas where refinement to the strategy, updates to policy and practice or other measures might be required.

Statistical subject matter areas are required to make appropriate use of the available facilities, adhere to the policies and follow the relevant guidelines. In particular, these areas remain responsible for the extent, accuracy and other aspects of the fitness for purpose of the metadata content related to their particular collection, including classifications, data elements etc which are specific to their collection. While DMS ensures the necessary "repository infrastructure" is provided, and that the infrastructure remains "fit for purpose" in a changing organisational and technical environment, DMS does not become responsible for the quality of the content held within each repository.

In addition to documenting their metadata initially, senior subject matter staff became responsible for "signing off" that the documented content was both accurate and sufficient. Subject matter areas also became responsible for ongoing custodianship of that metadata, including ensuring it remains up to date and answering any enquiries its definition might generate from others.

Two "standards" sections, Population Statistics Standards (PSS) and Economic Standards and Classifications (ESC), have leadership roles in regard to metadata content within the ABS. They develop and support "standard" classifications and data elements with are cross domain in nature (eg industry, occupation, language). Many of these are recognised standards for Australia as a whole, not just the ABS. The standards areas also provide guidelines and advice to help subject matter areas ensure their "collection specific" metadata is well defined and curated.
In July 2007 DMS, PSS and ESC were brought together into a single Branch (Data Management and Classifications) within the Methodology and Data Management Division. (PSS and ESC had previously been sections within subject matter divisions.) This brought together specialists in metadata models and systems and 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.

At a higher level the Metadata Strategy Group (MSG) comprising "Branch Heads" drawn from across the ABS exists to elaborate upon, drive forward and "champion" the metadata strategy. This group has direct access to top level management with the ABS and has regularly brought critical issues and proposals before them for input and funding approval.

During 2009 a Metadata Community of Practice (CoP) will be established, with membership at around "Section Head" level, to complement MSG. The Metadata CoP will focus on co-ordination of work programs and activities, sharing of leading practice and information exchange.

Neither MSG or the Metadata CoP have direct control over the activities of DMS or other business units - this remains with line management. By bringing together the relevant line managers, however, and by MSG providing recommendations to forums responsible for corporate decision making, including setting priorities for funding and formally signing off new and changed ABS policies, the two groups have an important practical role in shaping metadata management directions and outcomes within the ABS.

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

<table>
<thead>
<tr>
<th>Name</th>
<th>Role/Position in ABS</th>
<th>Phone Number</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alistair Hamilton</td>
<td>Director - Data Management Section</td>
<td>+61 2 62525416</td>
<td><a href="mailto:alistair.hamilton@abs.gov.au">alistair.hamilton@abs.gov.au</a></td>
</tr>
<tr>
<td>Simon Wall</td>
<td>Leader - DMS Metadata Management Team</td>
<td>+61 2 62526300</td>
<td><a href="mailto:simon.wall@abs.gov.au">simon.wall@abs.gov.au</a></td>
</tr>
<tr>
<td>Helen Toole</td>
<td>Project Officer DMS Metadata Management Team</td>
<td>+61 2 62527762</td>
<td><a href="mailto:helen.toole@abs.gov.au">helen.toole@abs.gov.au</a></td>
</tr>
<tr>
<td>Joan Burnside</td>
<td>Director Population Statistics Standards Section</td>
<td>+61 2 62527074</td>
<td><a href="mailto:joan.burnside@abs.gov.au">joan.burnside@abs.gov.au</a></td>
</tr>
<tr>
<td>Michael Robertson</td>
<td>Team Leader Economic Standards &amp; Classifications Section</td>
<td>+61 2 62525826</td>
<td><a href="mailto:michael.robertson@abs.gov.au">michael.robertson@abs.gov.au</a></td>
</tr>
<tr>
<td>Adrian Paviour</td>
<td>Director Statistical Process Development &amp; Support Section (responsible for ISHS &amp; other business process improvement for household surveys)</td>
<td>+61 2 62525213</td>
<td><a href="mailto:adrian.paviour@abs.gov.au">adrian.paviour@abs.gov.au</a></td>
</tr>
<tr>
<td>Natasha Bartlett</td>
<td>Director - Economic Statistics E2E Project (responsible for IDW &amp; other business process improvement for business surveys)</td>
<td>+61 2 62525314</td>
<td><a href="mailto:natasha.bartlett@abs.gov.au">natasha.bartlett@abs.gov.au</a></td>
</tr>
<tr>
<td>Eric Kerrison</td>
<td>Team Leader - Customer Insights &amp; Strategies Section (strategic directions for future dissemination of ABS data &amp; metadata)</td>
<td>+61 2 62526240</td>
<td><a href="mailto:eric.kerrison@abs.gov.au">eric.kerrison@abs.gov.au</a></td>
</tr>
<tr>
<td>Craig Watson</td>
<td>Team Leader Technology Applications/ IT Architecture (Programmer)</td>
<td>+61 2 62527208</td>
<td><a href="mailto:craig.watson@abs.gov.au">craig.watson@abs.gov.au</a></td>
</tr>
</tbody>
</table>

**5.2 Metadata management team**
Data Management Section (DMS) resides within the Data Management and Classifications Branch (DMCB) of the Methodology and Data Management Division (MDMD) of the ABS. DMS consists of around a dozen staff supported by around half a dozen programmers (application developers) from the ABS Technology Services Division. In addition to looking after

- policy and strategy related to metadata
- the work program related to the Corporate Metadata Repository (CMR)
- user support and training related to the CMR

DMS also look after the work program, user support and training for the output data warehouse and other aspects of data management policy and practice within the ABS. As part of supporting the National Statistical Service, DMS also increasingly engages with other agencies in regard to data and metadata management issues. In some cases this takes the form of providing guidance and advice, in other cases the focus is more on bilateral sharing of perspectives, experiences and plans.

The two other sections within DMCB are the standards areas for economic and population statistics, looking after the development, definition and promotion of key content related statistical frameworks, concepts and classifications.

Each of the three subject matter Groups within the ABS (each Group, loosely, consists of two Divisions) includes a "co-ordination section" that assists with

- requirements gathering and prioritisation for new metadata facilities and for improvements to existing ones
- targeting and co-ordinating "user acceptance testing" of, and feedback on, new/changed facilities
- co-ordinating definition of implementation programs for new processes and systems, monitoring progress of implementation and escalating the most common and most serious implementation issues to ensure they are addressed
- aiding communication between DMS and end users (translating terminology and impacts between the two)
- meshing the CMR work program with other work programs relevant to that Group

DMS also works closely with the publishing area to facilitate appropriate content from the CMR flowing through appropriately into publications or directly onto ABS web pages. The level of content flowing through in this way is increasing in terms of its scope, volume and interlinkage. Making metadata available to the public via the web raises a range of additional content, process and management issues for subject matter areas, DMS and Publishing that need to understood and addressed in an appropriate and sustainable manner.

5.3 Training and Knowledge Management

DMS provides a range of training. This includes an overview of concepts and systems related to metadata management. Such training is regularly made available to new starters within the ABS and other staff.

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 economic and social statistics areas provide training that includes explanations of how the CMR facilities fit within, and are used within, their business processes. The training about dissemination processes in the ABS likewise 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 be working very well.

5.4 Partnerships and cooperation
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.

A second major international opportunity for partnership and cooperation is seen to be around SDMX. ABS has provided extensive feedback on previous proposals and outputs from the consortium, and volunteered to take part in case studies. The consortium seems committed to providing National Statistical Offices with even greater opportunities to shape, rather than just respond to, the initiative in future. The ABS looks forward to that.

The ABS also contributes actively to international committees associated with other metadata standards of relevance to it, such as ISO 11179.

The ABS interest in collaborating on relevant software development, whether on an open source or alternative basis, has been noted earlier.

As described in Section 1.2, the National Statistical Service (NSS) provides many opportunities for various collaborations. Many of these collaborations are within Australia but they also include international collaborations, such as with the US Bureau of the Census. The NSS initiative takes 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.

A recent 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.

5.5 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.

Some of the practical complexities of managing and actively reusing metadata throughout the statistical cycle are not yet widely and well understood. This means there is a degree of disappointment and frustration expressed in some quarters that more progress hasn't been made more quickly and that we haven't yet made metadata simple to manage and maintain as well as "all powerful" in driving and describing all processes and outputs.

There is also still a tendency for projects to want to structure metadata in an exactly optimal manner for their processes and manage it directly in that form. The services layer "plumbing" to allow such an approach to be implemented efficiently, while drawing content from corporate metadata repositories that are not structured in the same manner, is far from being largely - let alone fully - in place. The fact the term "metadata" is so widely used, in a variety of valid but different contexts, is also emerging as an issue. 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. 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 either tend 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. There is some thought, echoed in other Australian agencies beyond the ABS, that promoting improved focus and outcomes related to Information Management (possibly Statistical Information Management in the case of the ABS) would be more effective than badging our work as data and metadata management.

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 see as already fundamentally fit for purpose but 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", where a series of poorly integrated local optimisations may result in an inefficient, sub-optimal end to end business process and/or a diversity of "locally optimised" processes/systems across the organisation may be very hard to sustain.
Staff's seeking of "local optimisation" can be linked to a sense of professionalism and pride in their work, so it is vital not to undermine the latter when seeking to address the former. Aiming for "local optimisation" also tends to be simpler than seeking global optimisation, so there may be an increased risk of missing the mark if the aim shifts to the latter.

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 vast.

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.

---

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 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.