

Information Management Advice 6 Information Classification tools

Introduction

The process of classification helps describe, organise and control information. In Information and Records Management practice, classification is undertaken for the purpose of managing records and other business information according to its business context within an organisation. Information can be organised in different ways for different purposes, for example 'security classification' which describes the level of restriction (public, confidential, etc) or sensitivity of information.

With regard to business information, classification builds order into the understanding what an organisation does and how it does it. AS ISO 15489 defines classification as “..the systematic identification and arrangement of business activities and records into categories accordingly to logically structured conventions, methods, and procedural (business) rules represented in a classification system..”. This definition can equally apply to all information content, not just records. The agency's classification framework in turn enables and enhances the capacity of the organisation to share information and knowledge more broadly than the traditional 'silos' established by business units working in isolation.

The management of records and other business information can be greatly facilitated by developing and implementing tools to assist with classification, titling, retrieval, security and access, sentencing and disposal of records. This Advice is intended to assist agencies to gain an awareness of the various classification tools, and provide guidance when considering the introduction of information classification schemes for agency use.

Understanding Classification tools and terms

The purpose of classification is to create a uniform approach or 'common language' that both people and systems can use for saving, organising, sharing and retrieving content. This uniformity provides a number of benefits, including:

- Ability to link classification to access, security and disposal decisions
- Ease of retrieval through consistent application of defined terms
- Linking of records to business context
- Consistency across business units and agencies (not focused on 'who' did it, but what was done)
- Flexibility and stability in response to organisational change

Some of the more common terms and concepts related to information classification are discussed in turn on the following pages.

Thesaurus

A thesaurus is a reference tool that lists words grouped together according to similarity of meaning, in contrast to a dictionary, which provides definitions for words, and generally lists them in alphabetical order. The main purpose of a thesaurus is to help the user “to find the word, or words, by which an idea may be most fitly and aptly expressed”¹.

Functional thesaurus: A functional thesaurus is a controlled list of terms linked together by hierarchical, associative or equivalence relationships. It assists users to find ‘preferred’ terminology by which to classify information, by including terms with similar, related or opposite meanings. Building on an organisation’s business classification scheme, its use can be extended from classifying, titling and retrieving records and business information, to managing, sentencing and disposing of records.

A functional thesaurus covers terminology relating to the agency’s unique or core business functions and activities. It can provide classification to three, four or more levels, allowing comprehensive classification of records and other business information. Terms are allocated according to established rules or conventions. Although a functional thesaurus displays terms in alphabetical order, when the classification terms are selected they are usually presented in the function | activity | topic | subtopic combination.

A functional thesaurus also includes terms that are not authorised for classifying and titling records. These ‘non-preferred terms’ (words or phrases) are similar to terms that can be applied to records. The purpose of adding ‘non-preferred terms’ is to provide links for users to similar ‘preferred’ terms that should be applied to records. As a functional thesaurus includes more navigational paths for users, it incorporates more functionality than a records classification scheme but it is also more complex, requiring more resources to develop, implement, maintain and to train users.

Keyword thesaurus: A thesaurus of general terms based on the keyword classification method is Keyword AAA (developed by State Records NSW). Keyword AAA is designed for use in classifying, titling and indexing many types of records. It is important to note, however, that Keyword AAA was not designed to cover ALL the records of an organisation, as it covers only those administrative functions and activities common to most agencies. Whilst it is not mandatory that agencies use Keyword AAA, it is a recommended starting point for classification tool development. Note that the current Disposal Schedule for Common Administrative Functions DA2157 is based on Keyword AAA.

Keyword for Councils (developed by State Records NSW) is the local government equivalent of Keyword AAA. It is intended to cover all the functions performed by local government authorities, in addition to administrative functions, which means that Councils generally do not need to add their own specific functional terms. However, geographical, municipal and/or jurisdictional differences may apply, with the result Councils may wish to add or amend terms to suit the Tasmanian environment. Note that the current Disposal Schedule for Records of Local Government Councils DA2200 is based on Keyword for Councils.

Merged thesaurus: A single alphabetical list of terminology for both general administrative and core functional terms, which is used to classify records. For example, an internally developed thesaurus of functional terms may be merged with an ‘off the shelf’ product like Keyword AAA, to provide comprehensive coverage across both the ‘common’ and ‘unique’ functions (and by definition all official records) of an agency.

¹ Peter Mark Roget

Taxonomy

“The classification of entities in an ordered system that indicates natural relationships”².

Taxonomy is a hierarchical classification system that groups objects based on assumed natural relationships or sets of characteristics. In information management, the intended purpose is to enable you to understand the structure of your information environment, and be able to predict where you are likely to locate information about particular things. It is also to enable you to group related things together, which is the ‘classification’ function of a taxonomy, and why records and information management-specific taxonomies tend to be known as ‘classification schemes’ (see Business Classification Scheme and Records Classification Scheme later in this Advice for more information).

Taxonomies can be represented in different forms:

Lists: A collection of items that have some basic relationship to each other. The relationship defined may be similarity of characteristics or purpose, steps in a process, projects that you are involved in, etc. When you look at a list, you should be able to understand the relationship that makes them ‘similar’. The relationship between items on a shopping list, may simply be the fact they are required household staples, and available for purchase from the local supermarket.



The main drawback to lists is they tend to become difficult to scan quickly after about a dozen entries. Further, making lists predictable can be challenging, for example, a list of job roles may be sorted by hierarchy, or activities sorted by sequence in a business process. When working with multiple lists (and associated business rules) this can become difficult for end users, who have to work out the sorting sequence before they can predict where to look. This is why many lists are usually sorted alphabetically, even if this is not the natural or preferred way of sorting, as it makes navigation absolutely consistent.

² <http://naa.gov.au/records-management/publications/glossary.aspx#>

Tree: Once a list gets too long, users typically start breaking it up into clusters, creating a tree structure. The tree is the structure most traditionally associated with taxonomies, categories at the top level, and sub-categories underneath.

In the shopping list example pictured above, you can see how items have been further clustered into sub-groupings, like Bread, Herbs, Dairy, Meat, etc.

Whilst the above example might be straightforward, most individually created tree structures are not intuitive to navigate, as people are often inconsistent when making decisions on groupings and categorisation. For example, you might sort things into some folders by type of document, but for others you might decide to use dates or topics. Whilst it might make sense to you, the lack of consistency makes it hard for others to follow, and subsequently locate information easily. Consider the structure of network share drives within your agency, particularly those where individuals are freely able to create their own systems including folder and/or document naming conventions!

Hierarchy: A hierarchy is a tree structure that follows very strict rules about how it is sub-categorised. The same principles must be consistently applied at every level. Each topic can only exist in one place in the structure. Strict hierarchies may not be very practical, however they do illustrate important principles about limiting the number of ways content can be organised and following consistent rules. The use of polyhierarchies attempts to solve the problem of overly restrictive hierarchies by ‘selectively’ breaking rules, however the problem of creating too many cross linkages is that it starts to impact on the principles of consistency and predictability. Very clear principles for use must therefore be developed and strictly adhered to.

Business Classification Schemes

“A conceptual model of what an organisation does, and how it does it. It involves the identification and documentation of each business function, activity and transaction and the documentation of the flow of business processes, and the transactions which comprise them”.³

A business classification scheme represents the business activity performed by an organisation. It is a by-product of the analysis of business activity undertaken, and is typically presented as a hierarchical model of what an organisation does, structured from the general to the specific (Function, Activity, Transaction/Task):

- First level – the business function
- Second level – activities constituting the function
- Third level – groups of transactions that take place within each activity

Business classification scheme structures typically fall into four categories:

Functional: Based on an analysis of the unique business functions and activities of an organisation, but independent of the organisation’s administrative structure. AS ISO 15489 endorses this approach (Functions → Activities → Transactions) to classifying records, as core functions of an agency change less frequently than organisational structure; and it allows for easier transfer of functions (and related records) between Government agencies and any associated amendments required.

Subject: Based on what the item or object is about, rather than the purpose or activity that the record was created to document (the intent). The context of the business activity can become imprecise and unclear, and

³ <http://www.informationstrategy.tas.gov.au/Publications/Pages/Glossary.aspx>

will make implementing disposal actions more difficult as they will contain records with differing retention periods. This in turn heightens the agency risk of unauthorised records destruction. Additionally, this practice unnecessarily extends storage requirements of temporary records that would otherwise be routinely destroyed. The file must be kept according to the longest retention period for records on the file, unless labour-intensive routine culling programs are introduced to manage the files.

Subject or theme-based classification was traditionally popular as it enabled the end-user to see all the transactions and activities that occurred under a single subject, such as a specific property or client. However, it is usually more important to establish whether a record is a policy, a complaint, or an application (the 'what') rather than the entity it relates to (the 'who') in order to best serve business, legislative and regulatory requirements.

Organisational: Based on an agency's organisational structure, typically from historic siloed paper environment recordkeeping practices (or network share structures developed to manage security access). Whilst a familiar structure to end users, this approach requires

- significant administrative support and is subject to frequent change;
- is difficult to apply retention and disposal;
- does not manage continuity or information context over time;
- encourages silos of information and information hoarding;
- can result in significant duplication and therefore redundancy of information, with associated storage implications and business inefficiencies.

Particular issues include difficulty with version control, currency and completeness of information available to the agency for decision-making.

Hybrid: This approach enables a compromise between a strict purist approach, and operational flexibility. Similar to a merged thesaurus, it may incorporate the functional-based (Function-Activity) model at the top levels, but permit a mixture of topic, transaction and subject-based elements at the third level to suit business needs, and offer a more intuitive solution to end users who do not naturally tend to think in transactional terms. A hybrid approach also provides flexibility to accommodate 'case files' and project files for example, which are not particularly well suited to a functional approach.

Records Classification Schemes and File plans

A records classification scheme (also known as a file plan or record plan) is a classification tool presented in a hierarchical structure. Founded on an organisation's business classification scheme, it is a tool for classifying records and other business information: the outputs of business activities that generate records. When applied to a business information system it can facilitate the capture, titling, retrieval, maintenance and disposal of records and other corporate information. Agencies should note the term Business Classification Scheme is often used interchangeably here when referring to the classification structure used specifically for the practice of records management.

Often a records classification scheme is depicted as a directory or folder (tree) structure. It can provide classification to two, three, and less frequently, four levels. Terms are allocated according to established rules or conventions. The components of the records classification scheme are displayed hierarchically in the order, or ranking, of Function | activity | topic | (subtopic). The third level topic is typically a mix of subject, transaction and topic titles, usually free text. Titling is typically a collaboration between

- a) officers who have responsibility for records and/or folder creation, and
- b) end users, based on documented agency naming conventions and business requirements.

When a records classification scheme is accessed via a desktop application (for example, in a web browser) users may initially see only the top level of classification terms, i.e. function terms. The terms at this level of the hierarchical ranking are usually listed alphabetically. Depending on the functionality of the classification tool, there may be several ways to navigate to the correct classification terms. Ideally 'scope notes' will be available to steer end users through the structure. Records classification schemes always provide some navigational paths, such as links between related terms, but these navigational paths tend to be limited. For example, records classification schemes do not link terms that a user may think are appropriate (but which are not used for classification, i.e. non-preferred terms) to terms that are used for classification (i.e. preferred terms) as a thesaurus would.

Folksonomy & 'tagging'

A folksonomy is a system of classification derived from the practice of information consumers 'tagging' content with their own personally defined terms in order to facilitate retrieval. Folksonomy, is a hybrid of folk and taxonomy. A broad folksonomy is the one in which multiple users tag particular content with a variety of terms from a variety of vocabularies, thus creating a greater amount of metadata for that content. A narrow folksonomy, on the other hand, occurs when a few users, primarily the content creator, tag an object with a limited number of terms. While both broad and narrow folksonomies enable content to be searched by adding textual description (or access points) to an object, a narrow folksonomy does not have the same benefits as a broad folksonomy, which allows for the tracking of emerging trends in tag usage and developing vocabularies.

Tagging, which is one of the defining characteristics of Web 2.0 services, allows users to collectively classify and find information. Some websites include tag clouds as a way to visualize tags in a folksonomy. However, tag clouds visualize only the vocabulary and not the structure of folksonomies. SharePoint 2010 is one example of an electronic content management tool that allows for the development of folksonomies as an additional way for end users to classify information using words and phrases that are meaningful to them.

Controlled vocabularies

An alphabetical list containing terms which are authorised (or otherwise controlled) in order to ensure that only one term is allowed to represent a concept or name. The alternative to a controlled vocabulary is free text⁴.

Controlled vocabularies provide a way to organize knowledge for subsequent retrieval. Controlled vocabulary schemes mandate the use of predefined, authorised terms that have been pre-selected by the designer of the vocabulary, in contrast to natural language vocabularies where there is no restriction on the vocabulary. The point of controlling a vocabulary is to ensure there is no ambiguity, overlapping or duplication of meaning. Controlled vocabularies can be used in subject indexing schemes, subject headings, thesauri, taxonomies and other forms of knowledge organisation systems.

A controlled vocabulary can be a simple list of terms which is used in your databases, for example, in many agency Customer Relationship Management (CRM) systems the names of customers are strictly controlled to

⁴ <http://www.informationstrategy.tas.gov.au/Publications/Pages/Glossary.aspx>

avoid variations being entered, and thereby scattering information across several entries. This is a form of controlled vocabulary.

Comparison of Classification tools

	Business Classification Scheme	Records Classification Scheme	Functional Thesaurus
Purpose	Conceptual model of the business activity of an organisation; intellectual basis for developing recordkeeping tools, e.g. a records classification scheme or a functional thesaurus	Tool for classifying, titling, accessing and retrieving records	Tool for classifying, titling, accessing and retrieving records
How the tool aids classification and retrieval	Scope notes to define functions and activities	<p>Hierarchical presentation of classification terms</p> <p>Function-level terms are linked to activity-level terms</p> <p>Activity level terms are linked to topic-level terms (where available)</p> <p>Terms in each hierarchical level are listed alphabetically</p> <p>Related terms are linked</p> <p>May include scope notes to define functions and activities</p> <p>Some applications will list classification strings</p>	<p>Alphabetical presentation of classification terms</p> <p>Access can be via activity, function or topic level</p> <p>All relationships between terms are displayed and defined</p> <p>Navigation from non-preferred to preferred terms through synonyms and other variations in description</p> <p>Includes scope notes to define functions and activities</p> <p>History notes connect current terminology with outdated terms</p> <p>Related concepts are linked</p> <p>Displays one-to-many relationships</p>
Limitations of the tool with regard to classification and retrieval	<p>No navigation to preferred terms</p> <p>Displays only one-to-one relationships</p>	<p>No navigation from non-preferred terms to preferred terms</p> <p>Displays only one-to-one relationships</p> <p>Scope notes may not be included</p>	
Levels of classification	<p>Three:</p> <p>Function</p> <p>Activity</p> <p>Transaction</p>	<p>Two or more:</p> <p>Function</p> <p>Activity</p> <p>Topic</p> <p>Subtopic</p>	<p>Two or more:</p> <p>Function</p> <p>Activity</p> <p>Topic</p> <p>Subtopic</p>
Differences in third level of classification	Transactions – the smallest unit of a business activity (the tasks	Topics and subtopics may be a mix of transactions, subjects or record types	Topics and subtopics may be a mix of transactions, subjects or record types

	Business Classification Scheme	Records Classification Scheme	Functional Thesaurus
	undertaken to perform activities)	based on the most appropriate way to title the record for searching and retrieval	based on the most appropriate way to title the record for searching and retrieval
Example of third level of classification	<p>Transactions, eg COMMUNITY RELATIONS Research –</p> <p>Receive research request Determine scope of project Consult internal and external experts Conduct literature search and gather documentation Analyse material Formulate report</p>	<p>Topics, eg COMMUNITY RELATIONS Research –</p> <p>Brochures Business case Equal employment opportunity Ethnic affairs Hazardous substances Literary contributions Oral History Projects Publications Questionnaires</p> <p>(Source: Keyword AAA)</p>	<p>Topics, eg COMMUNITY RELATIONS Research –</p> <p>Brochures Business case Equal employment opportunity Ethnic affairs Hazardous substances Literary contributions Oral History Projects Publications Questionnaires</p> <p>(Source: Keyword AAA)</p>

Practical application in the workplace

As we have seen, agencies already have numerous ‘classification schemes’ in place supporting day-to-day business. Existing classification schemes in your agency may include organisational charts, site maps (web/intranet), asset registers, file plans, etc.

Records and information can therefore be classified in a number of different ways, depending on business objectives, including:

- Security level eg Confidential
- Retention eg PERMANENT
- Subject/topic eg Annual reports
- Location eg Australia > Tasmania > Hobart
- Function eg Financial Management
- Document type eg customer request

Given the volume of information in an agency, and the often widely different methods employed for classifying it, the value of developing and implementing an agency-wide classification scheme for use with relevant information repositories, cannot be understated. Significant analysis should be undertaken to ensure appropriate application in individual environments, however, potential benefits may include:

- Consistent application of classification and terminology across business units, departments and business systems in turn, leading to improved search functionality, and improved scope for business intelligence/analytics and reporting

- Improved access/retrieval as end users become familiar with 'authorised' structure (less ambiguity or subjectivity)
- Improved efficiencies and compliance in terms of improved management of unstructured information, and potentially reduced 'dark data'. (Dark data is generally defined as information assets collected, processed and stored, but not effectively re-used or leveraged to support business). Enterprise-wide classification structures and processes may be adapted and applied to traditionally unstructured information, and/or help to identify information of low business value, or time expired records that may be appropriately destroyed
- Improved implementation of e-discovery strategies
- Consistent application of business rules, security and access rights

Whilst the work to develop functional Business Classification Schemes for the agency is significant, additional records management tools may be developed from the functional analysis undertaken, thereby adding value to the undertaking. These include:

- File naming conventions - hard copy or electronic thesaurus
- Controlled vocabularies for use with collaboration tools eg SharePoint 2010
- File plans for EDRMS systems
- Development of functional retention and disposal schedules
- Network drive structuring and data mapping in preparation for data migration to EDRMS systems

Additional Resources

Developing a Thesaurus, State Records of South Australia ⁵

Developing a Functions Thesaurus, National Archives of Australia ⁶

International Standard ISO 2788-1986 Documentation - *Guidelines for the establishment and development of monolingual thesauri*

⁵

http://government.archives.sa.gov.au/sites/default/files/20120124%20Developing%20a%20Thesaurus%20Final%20V1.3_Copy.pdf

⁶ http://www.naa.gov.au/Images/developing-a-thesaurus_tcm16-47228.pdf

Further Advice

For more detailed advice please contact:

Government Information Strategy Unit
Tasmanian Archive and Heritage Office
91 Murray Street
HOBART TASMANIA 7000
Telephone: 03 6165 5581
Email: gisu@education.tas.gov.au

Acknowledgements

This document is largely based on:

- *Overview of Classification Tools for Records Management*, National Archives of Australia.

Additional information has been drawn from:

- *Business Classification Scheme Design*, National Archives UK
- *Public Records Brief: Functional vs Subject-based Classification*, Queensland State Archives
- *Public Records Brief: An Introduction to Business Classification Schemes and Thesauri*, Queensland State Archives
- *AIIM Training Brief: How to Develop Taxonomies to Support Navigation, Information Discovery, and Findability* - Carl Weise AIIM 2012⁷,
- Information Security Classification

Information Security Classification

This document has been security classified using the Tasmanian Government Information Security classification standard as PUBLIC and will be managed according to the requirements of the Tasmanian Government Information Security Policy.

Document Development History

Build Status

Version	Date	Author	Reason	Sections
3.0	February 2015	Christine Woods	Template	All
2.0	August 2014	Sam Foster-Davies	Rewrite	All
1.0	13 July 2005	D Benjamin	Initial Release	All

⁷ www.aiim.org

Amendments in this Release

Section Title	Section Number	Amendment Summary
All	All	Document imported into new template

Issued: September 2014

Ross Latham
State Archivist