e-government

E-GOVERNMENT METADATA FRAMEWORK

MAY 2001

© Crown copyright 2001

The text in this document may be reproduced free of charge in any format or media without requiring specific permission. This is subject to the material not being used in a derogatory manner or in a misleading context. The source of the material must be acknowledged as Crown copyright and the title of the document must be included when being reproduced as part of another publication or service.

First published 2001

ISBN 0 7115 0408 3

Online copies of this document available at: http://www.govtalk.gov.uk/egif/home.html

Office of the e-Envoy Stockley House 130 Wilton Road London SW1V 1LQ

email: ukgovtalk@e-envoy.gsi.gov.uk

Website: www.e-envoy.gov.uk

Designed and typeset by Format Information Design

Website: www.format-info.co.uk

Foreword by Rt Hon Ian McCartney MP, Minister of State, Cabinet Office



Getting to the right information fast is the key to internal government efficiency and serving the citizen. Metadata helps us to do that.

The Internet revolution has led to an information explosion, but this sheer quantity makes it difficult to find what we need. Tagging our information assets with metadata will help us to identify and exploit our information assets, allowing industry and intermediaries to repackage and add value to them. More importantly, it will make it easier for businesses and citizens to find what they want.

This policy framework has agreed that a metadata standard, based on Dublin Core, should be applied across government. It has also agreed to initiate the Pan-Government Thesaurus project, which will build up a list of keywords to be used in the implementation of search engines and the categorisation of information across government.

The Metadata Framework is another important step in our campaign to deliver e-Government and make life simpler for all.

Muching

Contents

	Executive summary	1
1	Metadata	3
	Introduction	
	What is metadata?	
	Better information faster: the value of metadata	
	Key policy decisions	
	Elements, Application Profiles and the Pan-Government Thesaurus	
	Dublin Core	
	Pan-Government Thesaurus (PGT)	
	Principles of the e-Government Metadata Standard	
2	Management processes and implementation strategy	11
	Implementation support – UK GovTalk	
	Roles and responsibilities	
	Compliance	
	Updating and maintaining the e-GMF	
	Consultation and innovation	
	Future versions of the e-GMF	
	Glossary	17

Executive summary

Information is the key resource. This is as true of government as it is of businesses. Technology and standards that allow efficient exploitation of information are the key to organisational success. Metadata is one such standard.

The e-Government Metadata Framework (e-GMF) sets out the Government's policies for establishing and implementing metadata standards across the public sector. These standards will be used across all of our information systems.

Metadata is data about data, or a summary of the form and content of a resource. Structured metadata is divided into elements, each of which describes a particular aspect of the information resource. Consistency is important for effective information management and retrieval, so clearly defined standards must be set and widely applied.

Two headline decisions are announced in this framework. First is that the UK Government has adopted simple Dublin Core as the UK Government Metadata Standard (e-GMS). However, we recognise that Dublin Core alone is not sufficient to serve our needs. We will therefore add further elements and refinements, according to the principles laid down in this Framework.

The second headline decision is to develop a UK Pan-Government Thesaurus. The thesaurus will contain a structured list of terms and keywords to help us define information accurately and find information faster. The Pan-Government Thesaurus will be made available to government, industry and the citizen.

Adherence to the e-GMF is mandatory, and will be implemented in stages. Simple Dublin Core elements are mandated immediately, other elements and the Pan-Government Thesaurus will be mandated once they are developed and approved.

The implementation of the e-GMF is a substantial and ongoing task, which will require central support and mechanisms for development. The UK GovTalk website will host the e-GMF, the e-GMS and the Pan-Government Thesaurus. The site also hosts the mechanisms for consultation and innovation and will provide best practice guidelines and toolkits.

1 Metadata

Introduction

- 1. In this Information Age, government information resources are of increasing value to the citizen, businesses, government officers and users around the world. Effective information management is essential for the provision of services to the citizen and businesses, and for the efficient running of government.
- 2. Government information resources are not only of value in themselves. They should also be seen as valuable economic assets, the fuel of the knowledge economy.
- 3. It is generally accepted that information generated by the statutory and normal workings of government forms the largest single information resource in any developed economy. By making sure the information we hold can be found and passed between the public and private sectors, taking account of privacy and security obligations, we can help to make the most of this asset, thereby driving and stimulating our economy.
- 4. Simply putting information on a website isn't enough. There is so much information on the Internet that it has become very difficult to find anything specific. Adding metadata involves tagging our information resources to help people navigate through masses of web pages and confidently locate whatever they are looking for. For metadata to be effective across the whole range of government it must be well structured and consistently applied.

What is metadata?

- 5. The most common definition of metadata is 'data about data'. This is a simplification however; metadata is a summary of the form and content of a resource. For example, a catalogue selling household items gives the metadata of those items; the brand, price, colour and capacity of a kettle, or the size, fabric, colour and price of a suit. A library catalogue contains metadata relating to books; their titles, authors, publishers etc.
- 6. In its broadest sense, 'metadata' can be used to describe information structures such as the technical standards and interconnection policies that are the province of the e-Government Interoperability Framework. The e-GMF is concerned with the particular facets of metadata intended to support resource discovery and records management. The e-Government Metadata Standard (e-GMS) covers the core set of 'elements' that contain data needed for the effective retrieval and management of official information. Each element contains information relating to a particular aspect of the information resource, e.g. 'title' or 'creator'.

- 7. It is often more effective to search through metadata than through the resources. Metadata gives information needed by seekers and managers of information, such as subject keywords, date created and intended audience, information that is not always available in the resource itself.
- 8. An important feature of metadata is that it is separable from the information resource itself, and can be made available when the actual resource cannot.
- 9. Many of the different metadata schemes in use internationally have developed independently from each other. As a result, terms used to describe aspects of metadata vary from scheme to scheme. A full glossary is therefore provided at the end of this document.

Better information faster: the value of metadata

- 10. The best place to hide a tree is in a forest. One can equally, if unintentionally, hide information on the Internet or in any large electronic information collection. Although vast quantities of information and data are now readily available through electronic means, it is still not always possible to find what is needed, or to feel confident that the source being used is the best one. This applies to departmental intranets and extranets as well as the Internet.
- 11. The reasons for this are complex, and relate to the way people search for data, the way search engines work, the way information is stored and indexed, and the sheer volume of data available.
- 12. Sound, consistently applied metadata standards cannot guarantee perfect results every time. They can however considerably improve the granularity of a search, essential when navigating around large information resources.
- 13. Metadata is also a valuable tool for those managing records. Specialised local versions have been used for many years to identify, authenticate, describe and manage official records. The change to electronic record keeping provides an opportunity to streamline records management systems while making them more flexible, efficient and joined-up. The Public Record Office's new Electronic Records Management programme will be developed in line with the e-GMF.
- 14. Metadata is also invaluable when making non-text items, such as photo or music collections, forms and services, available electronically. A search engine cannot scan these types of file for keywords, because they don't contain words. Filenames and broad categories are seldom enough to lead users quickly to the right images or service providers.

Example

Hamid, researching a new project, is looking for statistics on egg production in East Anglia for the last three years. He goes to his favourite web search engine and types in 'egg production east anglia'. All search engines work in different ways; this one looks for each of these words, and returns a list of all websites containing at least one of them, arranged so that sites containing all four words are listed first. From his list of 80,000-odd hits, Hamid finds the site for the Egg Marketing Board, the best of the 300 hits he looks through, and uses the official statistics he finds there, with success, throughout the life of the project.

Had he searched on 'poultry', or gone through another 6,000 of his 80,000 hits, he would have found the East Anglian Poultry Farmers Association, an independent body that provides more detailed, current and projected statistics, local knowledge, and some useful, expert opinions on potential market growth and opportunities.

This sort of thing happens all the time.

Had consistent metadata standards been applied, and the search engine programmed to recognise them, Hamid would have been able to limit his search to items of type 'statistics' relating to the geographic region 'East Anglia' only. It would also have increased the probability that the Poultry Farmers' Association would have included the word 'eggs' as a search term. Together these factors would have resulted in a smaller but more accurate search result, and who knows, Hamid might be driving a sports car now instead of that moped.

Key policy decisions

- 15. Four key decisions have shaped the development of the UK Government Metadata Framework:
 - the endorsement of simple Dublin Core and its adoption as the e-Government Metadata Standard.
 - the plan to develop **additional elements** to enhance the e-Government Metadata Standard to meet all of government's information management and retrieval needs.
 - the plan to develop the e-Government Metadata Standard Application Profile to include element refinements and encoding schemes
 - the plan to develop a Pan-Government Thesaurus.

Elements, Application Profiles and the Pan-Government Thesaurus

16. For metadata to be effective it must be structured. The e-Government Metadata Standard sets out the initial structure for all metadata applied to government information systems. At

the moment it comprises the 15 simple Dublin Core elements only; when fully developed, the structure will consist of elements, an application profile, refinements and encoding schemes.

- 17. The **Elements** form the building blocks of the metadata for any given resource. They will be complemented by the **Application Profile**, which will consist of
 - Element refinements, which will break down the elements into smaller divisions, for example the 'Date' element may be broken down into 'date created' 'date published' 'date made available' 'date valid' 'acquisition/accession date'.
 - Encoding schemes, where appropriate, will specify the format in which the value for each element will be recorded, for example it may specify that the value for the 'Date' element will always appear in the format yyyy-mm-dd, which is the ISO 8601:1988 standard.
- 18. Proposed enhancements to the e-Government Metadata Standard will be placed on the UK GovTalk website for consultation. The final set of elements and refinements will be comprehensive and designed to provide for many purposes, including information retrieval, information management and records management. Project managers will select the elements and refinements they consider suitable for each application, using guidelines laid down in the Standard. Different types of resource will thereby be catered for by the same standard, with consistency being maintained throughout.
- 19. As part of the proposed enhancements, each element will be marked with a status of Mandatory, Recommended or Optional. In due course Deprecated may also be needed, to mark elements and refinements that should no longer be used. Any elements or element refinements added under the Extensibility provisions (see section 34) will be marked as 'Optional' or 'Recommended' for a period of time and only marked as 'Mandatory' after due consultation and if circumstances warrant.
- 20. The diagram opposite describes the relationship that will exist between the component parts of the e-Government Metadata Standard; the elements, application profile and the Pan-Government Thesaurus. It shows a partial metadata record describing a fictional document.

Please note that this diagram is illustrative only; the Application Profile is under development and may vary from that shown here.

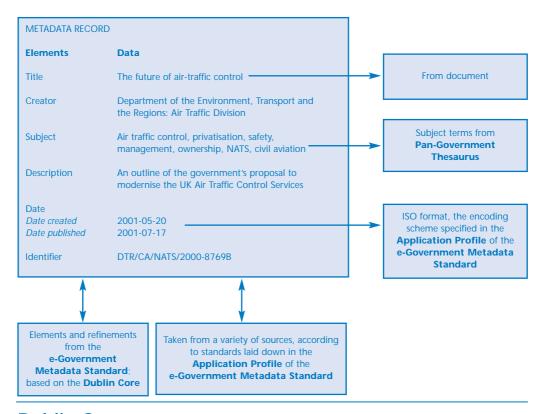


Figure 1 Elements, Application Profiles and the PGT

Dublin Core

- 21. The Dublin Core is a highly developed, flexible and internationally recognised standard, and has therefore been adopted as the core of the e-GMF. See http://dublincore.org for background information on the Dublin Core and details of the standards.
- 22. 'Simple' Dublin Core establishes the basic set of fifteen elements developed and supported by the Dublin Core Metadata Initiative (DCMI). 'Qualified' Dublin Core refers to the same set of elements and recommends additional regulation of each element, usually by an international standard, e.g. it may require the 'language' element to be filled with an item from the ISO list of languages (ISO 639-2). While 'Simple DC' has already been adopted as the e-GMS, an Application Profile in which element refinements and encoding schemes are specified will also be needed. The refinements used in DC will be given first consideration when the Application Profile is being developed.
- 23. Dublin Core was developed to meet resource discovery requirements. It does not provide for other areas, notably records management. To meet the purposes of the e-GMF, further elements will be added to the e-GMS.

Simple Dublin Core Title Format

Author or creator Resource identifier

Subject and keywords

Description

Publisher

Other contributor

Source

Language

Relation

Coverage

Date Rights management

Resource type

Pan-Government Thesaurus (PGT)

24. One of the greatest barriers to effective information retrieval across a number of organisations is the lack of a common vocabulary. A thesaurus acts as a guide to terminology, helping reduce the number of 'false hits' and ensuring all of the correct items are found. It has therefore been decided that a pan-government thesaurus will be developed and maintained, to help ensure semantic consistency of metadata records by providing a structured list of terms covering all aspects of government and its business.

25. The PGT will play a major part in ensuring easier, more accurate information retrieval. It will however be a major undertaking, in some cases requiring considerable resources to develop and maintain. A study is therefore being conducted to determine the

- · development and implementation strategy
- cost
- · ongoing management and control strategy
- · size and depth
- use of existing thesauri
- use of software applications.

26. The Category List will complement the thesaurus, providing a high-level categorisation scheme. Seekers of information will simply browse the list of categories and subcategories; they will not have to think of keywords. A category list of policy keywords has already been incorporated into the Knowledge Network, and is being developed further in preparation for incorporation into the Citizen Portal.

27. The Pan-Government Thesaurus (PGT) and Category List will form integral parts of the e-GMF. They will be dynamic, constantly evolving tools, maintained via UK GovTalk for use by all. More about the Citizen Portal, Knowledge Network and other initiatives that will benefit from the e-GMF can also be found on the site.

Principles of the e-Government Metadata Standard

28. The e-Government Metadata Standard will be further developed and maintained

according to the following principles. Some of these are necessarily contradictory, and it will be the task of the Information Age Government Champions' Metadata Working Group to ensure a practical balance is maintained between conflicting requirements.

- 29. It will be **Independent**. It will not be software, application or project based, but flexible enough to meet the information retrieval and records management needs of any information held in any format.
- 30. It will be **Simple** to use. The standard must be readily applicable by those with widely varying experience of preparing resource descriptions.
- 31. It will be **Compliant** with other UK Government standards and policies, such as the e-GIF standards and the Government Data Standards Catalogue.
- 32. It will be **Compliant with international standards**. Information is an international resource, and the UK aims to remain a leader in the global information revolution. To achieve this, the metadata standard must reflect international standards and systems. If an international standard is appropriate and kept up to date it will be incorporated into the e-GMS. Preference will be given to standards with the broadest remit, so appropriate international standards will take preference over EU standards, EU will take preference over UK standards.
- 33.It will be **Stable**. Changes to a standard that will become embedded in all information systems will require considerable effort, time and resources to implement. The e-GMS must therefore be flexible enough to meet future as well as current needs.
- 34. It will be **Extensible**. Additional element refinements can be added where it can be shown that these are essential and the existing set does not make provision for the requirement. A balance will need to be struck between the need for extensibility and the need for stability.
- 35. It will be economical and give value for money.
- 36. It will be **Inclusive**, taking into account the many existing metadata schemes, with the aim of minimising the need to rework existing products. This will be balanced with the need for maximum interoperability, which requires consistency across all information resource descriptions.
- 37. **Above all**, it will meet the information retrieval and management needs of the citizen and of government.

METADATA

2 Management processes and implementation strategy

This section covers the processes by which the e-GMF and the tools needed to implement it will be developed, applied and maintained.

Implementation support – UK GovTalk

- 1. UK GovTalk is a government initiative initially designed to get public and private sector organisations to work together to support and develop the XML schemas in accordance with the e-Government Interoperability Framework.
- 2. The UK GovTalk website http://www.govtalk.gov.uk will be used to support the e-GMF implementation strategy. It provides a forum for discussion and information sharing on all aspects of the Metadata Framework and its implementation. The e-GMF and related information can be found at http://www.govtalk.gov.uk/egif/home.html.

Roles and responsibilities

3. The main roles and responsibilities of the key players are:

Cabinet Office, Office of the e-Envoy

- is the lead authority for implementing and maintaining the e-GMF and associated documentation and tools, with ultimate responsibility for its content and implementation
- provides the resources needed for the development and maintenance of the e-GMF and the Pan-Government Thesaurus
- ensures that the development of the e-GMF is co-ordinated with other related UK Government initiatives and developments
- ensures that, as far as is practical, the e-GMF develops in line with related UK and international standards
- leads the Change Management process
- · chairs the relevant working groups
- promotes and audits the implementation of the e-GMF
- manages the UK GovTalk website

Information Age Government Champions (IAGC)

- ensure that the e-GMF is implemented within departments and other public bodies
- ensure use of the e-GMS is maintained and standards are followed

IAGC Metadata Working Group (MWG)

- provides advice and comments on all aspects of the e-GMF
- prepares and distributes papers relating to the development of the e-GMF
- develops and maintains the e-Government Metadata Standard

IAGC Pan-Government Thesaurus Working Group

- reports to the IAGC Metadata Working Group
- · plans and develops the PGT and the Category List
- develops the implementation and ongoing management processes for the PGT
- prepares and distributes papers relating to the PGT

Committee of Departmental Librarians Metadata Working Group (CDL MWG)

 provides professional input and advice, especially relating to the Pan-Government Thesaurus

User Organisations

- inform the Office of the e-Envoy about related initiatives within their own organisations
- contribute to the Pan-Government Thesaurus (PGT)
- describe how e-GMF has been used in their organisations and what effect it has had, thereby contributing to its development
- · comment on proposals for changes to the e-GMF

External metadata specialists

• provide expert advice for the development of the e-GMF

Compliance

- 4. The e-Government Metadata Standard applies to all intradepartmental systems and the interactions between:
 - UK Government department and other UK Government departments
 - UK Government and wider public sector
 - UK Government and foreign governments (UK/EU, UK/US etc)
 - UK Government and businesses (world wide)
 - UK Government and citizens.
- **5**. The compliance rules are:
 - the e-GMF is mandated on all new systems
 - legacy systems which link to the GSI, Government Portal (Gateway and UK Online), the Knowledge Network, or other systems which are part of electronic service delivery or electronic records management, will need to comply with the e-GMF over time.

- 6. Mandation will take place in four stages:
 - Use of e-Government Metadata Standard (simple Dublin Core) in all specified systems. Effective immediately
 - Use of enhanced e-Government Metadata Standard, including additional elements and application profile. Effective from date of completion.
 - Use of the Category List. Effective immediately for use in the Knowledge Network Project, effective for use in other systems from date of completion.
 - Use of Pan-Government Thesaurus. Effective from date of completion.
- 7. The e-Government Metadata Standard will undergo further development. Following wide consultation and testing in the field, in a variety of information systems, and receiving approval from the MWG, mandation of the advanced e-GMS will come into force.
- 8. The Category List is already in use in the Knowledge Network Project. It will require further development before it is suitable for use by UK Online and other areas. Mandation will take place in stages as the Category List gains approval for specific projects or information systems.
- 9. Implementation of the e-GMF is also recommended for all system procurement and major upgrades that fall outside the mandate. In practice, information sharing between organisations will become increasingly difficult if they do not comply with the e-GMF.
- 10. Each department will include compliance with these standards in its own project approval process. Whilst compliance is by self-regulation, it is expected that the policies and standards set out in the e-GMF will be used by organisations involved in project audits, reviews and approvals by the National Audit Office and the Audit Commission.
- 11. The e-GMF standards will also be part of the conformance check for the release of funding by the e-Envoy and HM Treasury.
- 12. However, the Government's approach to compliance is to help stakeholders as much as possible. Organisations will also be able get help with compliance problems through the website.

Updating and maintaining

- 13. Changes to the e-Government Metadata Standard or Application Profile may involve considerable effort on the part of all user organisations. The e-GMS must change, however, to keep up with developments in technology and the Dublin Core and other international standards being used.
- 14. The Office of the e-Envoy, as lead authority, is responsible for managing the change process, and for ensuring that the e-GMF is at the forefront of development.

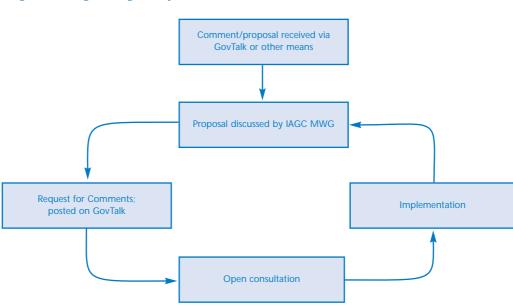


Figure 2 Change management process

Consultation and innovation

- 15. All users and parties with a vested interest in the e-GMF must have the opportunity to provide input to the change process, comment on all aspects of it, and describe the requirements of their individual projects. This will be done through UK GovTalk. The consultation process will actively target organisations and individuals who have expressed an interest in the e-GMF. Unsolicited comments and suggestions will also be encouraged.
- 16. The Metadata Working Group of the Information Age Government Champions will initially discuss comments and suggestions received. Those that meet the policies and principles laid down in this document, and that potentially offer positive benefits to user organisations, will be posted on the UK GovTalk site with Requests for Comments. They will also be sent to active user organisations, IAGCs and other interested parties. Comments will be invited from all.
- 17. When proposed developments require innovative solutions, the government will issue a Request for Proposals outlining the requirement. The aim is to attract innovation and the most cost-effective solution using the world-wide industry and population base. The MWG will also request expert opinions from information specialists if it is considered valuable to do so. Updates and additional information will appear on the website, along with comments received.
- 18. Following open consultation, the final decision rests with the Office of the e-Envoy.

Implementation of the decision will be the responsibility of the Office of the e-Envoy and the Information Age Government Champions, departments and agencies.

Future versions of the e-GMF

- 19. The e-Government Metadata Framework will be formally updated every six months, subject to guidance from the Information Age Government Champions and the Metadata Working Group.
 - Issue 1 was published in April 2001
 - Issue 2 will be published in Autumn 2001.
- 20. Published issues will be available on the Internet at http://www.govtalk.gov.uk/egif/home.html. The site will also contain working drafts of the document giving advance information on government direction and the possible content of the next issue to show how the framework is developing and to invite comments on it.

Glossary

Element

Application Profile Rules, element refinements, encoding schemes and other factors

regulating the content and format of each metadata element.

Category list The simplest type of controlled vocabulary is a high-level categorisation

(or classification) scheme. At the time of input, one or more categories must be selected from the scheme and added to the document metadata. At the time of seeking information, the user does not have to think of keywords, but simply browses the list of categories and subcategories.

One of the items that collectively form a metadata structure. Common

elements are 'title', 'creator', 'date', and 'publisher'. Dividing data into elements allows users to carry out more accurate searches by searching on one element only. For instance, when looking for documents by Jennifer Green, searching the 'creator' field only will retrieve items by Jennifer Green only. It avoids items where the word 'green' appears in other

contexts, as a subject, location etc.

Element refinement A sub-set of an element, to make the meaning narrower or more specific,

e.g. 'Date created', 'Date destroyed' as refinements of 'Date'. A refined element shares the meaning of the unrefined element, but with a more restricted scope. A user who does not understand a specific element refinement term should be able to ignore the refinement and treat the metadata value as if it were the broader element, although this will lose some precision. The definitions of element refinement terms must be

freely available.

Encoding scheme A scheme that controls the content, or 'value' of an element or element

refinement, in order to clarify the meaning or improve resource discovery. These schemes include controlled vocabularies and formal notations or parsing rules. A value expressed using an encoding scheme will thus be a token selected from a controlled vocabulary (e.g. a term from a classification system or set of subject headings) or a string formatted in

accordance with a formal notation (e.g., '2000-01-01' as the standard expression of a date). Encoding schemes are designed to be interpreted

by machines or by human readers.

The definitive description of an encoding scheme must be clearly identified and available for use by those attempting to find information as

well as those creating the metadata records.

Field Commonly used in database applications to describe a space in which

data of the same type is entered (e.g. 'title' or 'price'), 'field' is a similar

concept to 'element'.

Information retrieval Finding the right information. Good information retrieval methods help

ensure users find everything they are looking for, and only what they are

looking for.

Metadata

A summary of information about the form and content of a resource. The term 'metadata' has been used only in the past 15 years, but has become particularly common with the popularity of the World Wide Web. The underlying concepts have been in use for as long as collections of information have been organised. Of particular interest to this Framework are the facets of metadata intended to support resource discovery and records management.

'Metadata' can also be used to describe more technical aspects of information resources; the type of information needed to transfer information from one type of computer or software application to another. 'Metadata' of this type is covered in the e-GIF.

A full set of structured relevant metadata, comprising all relevant elements, describing one information resource. A metadata record can take many forms;

- as part of the main information resource itself, e.g. the metadata of an XML file
- a completely separate record held apart from the information resource itself and even in a different format e.g. an automated library catalogue
- an electronic file held as an extension of the main resource e.g. the 'format' files of a Word document

Term used to refer to both 'Element refinement' and 'Encoding schemes'. Use of this term tends to cause confusion, so it is avoided in this document

Refinement Resource discovery Sub-element Taxonomy

Metadata record

See 'element refinement'.

Finding the right stuff. See 'information retrieval'.

Term sometimes used to refer to an element refinement.

The science of classification, traditionally used to describe a hierarchical scheme for classifying plants and animals. More recently it has been borrowed to describe a classification scheme for organising networked resources and supporting user-friendly navigation among them. Some taxonomies incorporate thesaurus features to augment the hierarchical structure.

Thesaurus

A controlled vocabulary designed to support information retrieval by guiding both the person assigning metadata and the searcher to choose the same terms for the same concept. A thesaurus conforming to ISO 2788 (=BS 5723) supports navigation and term selection by showing relationships between terms that are close in meaning.

A thesaurus can help to ensure:

- · concepts are described in a consistent manner
- experienced users are easily able to refine their searches to locate information easily
- · users do not need to be familiar with technical or local terminology.

Qualifier