Metadata in the UK

Maewyn Cumming, Office of the e-Envoy, UK
maewyn.cumming@cabinet-office.x.gsi.gov.uk

Abstract

The UK Government's Modernising Government programme, begun in 1999, called for the delivery of high quality public services, to be made available seven days a week, 365 days a year. This led to the development of a central portal from which citizens could gain ready access to services and information from many sources.

To achieve this, a programme to develop and implement a standard metadata system was undertaken. The policy is to use simple Dublin Core until an enhanced version, with additional elements, is developed and tested. A system of vocabulary control for the Subject element is being developed. Such a standard is also being considered by other national governments with similar demands.

Keywords; Metadata, government, UK, records management, Dublin Core, elements

Introduction

The government is committed to utilising information and communication technology (ICT) to improve communications between it and citizens. This doesn't just mean having the latest gadgets. It means setting up the right regulatory and legislative environment to support e-commerce, giving people and businesses the confidence to take advantage of the opportunities offered by the Internet, and making sure business and government decisions are informed by the best information available.

The Government isn't simply encouraging businesses to go online, it is taking its own advice by providing quality services electronically to citizens and businesses. The benefits are many, not least that they are available at any time, and through a number of channels, increasing convenience.

In doing this the government has embarked on one of its largest projects ever. The figures speak for themselves: government conducts five billion transactions a year with citizens and businesses, spread over 20 large departments, 480 local authorities and over 200 agencies.

Government and businesses are working towards a common goal: providing services over a new medium to a populace concerned about the security of their information and ease of use. While we have given ourselves the target of having one million small and medium sized enterprises (SMEs) trading online by 2002, we are particularly aware of the problems involved, and need to be in a position where we can drive the technology industry to provide solutions to benefit all.

This paper outlines some of the work the Office of the e-Envoy is doing to meet its own business targets, particularly in the fields of interoperability and metadata. It begins by outlining the underlying programmes and drivers behind the UK metadata policy, describes the policy itself, and then looks in detail at the extensions to Dublin Core we plan to adopt.

Modernising government

The origins of the plan to standardise metadata use across the UK public sector can be traced directly to the 1999 Modernising Government White Paper. [1] This laid out three basic aims; to ensure policy making was more joined up and strategic; to make sure that public service users, not providers, became the focus; and to deliver high quality and efficient public services.

The paper also outlined a long-term programme of improvements, starting with a commitment to ensure that public services are available 24 hours a day, seven days a week, the promotion of 'joined-up government' with a clear commitment for people to be able to notify different parts of government of details such as a change of address simply and electronically in one transaction, a new drive to remove unnecessary regulation, and a new target of all dealings with government being available electronically by 2005.

It was under the Information Age Government (IAG) heading that some of the more visible transformations were to occur. IAG promised to use new technology to meet the needs of citizens and business, and not trail behind technological developments. It would modernise the business of government itself, achieving joined up working between different parts of government and providing new, efficient and convenient ways for citizens and businesses to communicate with government and to receive services.

The Modernising Government paper was followed in April 2000 by the e-Government Strategic Framework. [2]. This strategy focused on better services for citizens and businesses and more effective use of the Government's information resources. Implementing it means creating an environment for the transformation of government activities by the application of e-business methods throughout the public sector. The strategy laid down four guiding principles:

  * building services around citizens' choices;
  * making government and its services more accessible;
  * social inclusion; and
  * using information more efficiently.

The Office of the e-Envoy

Before the Modernising Government strategy started to develop, government IT systems had largely been developed separately by different public service agencies. The Cabinet Office's Central IT Unit (CITU) was charged with encouraging them to converge and inter-connect.

The Office of e-Envoy (OeE), also based in the Cabinet Office, was established in September 1999 following a central recommendation of the Performance and Innovation Unit's report 'e-

commerce@its.best.uk' [3], which set out a detailed strategy for meeting the Government's target for e-commerce set in the 'Our Competitive Future' White Paper [4].

The OeE and CITU merged in July 2000, and the e-Envoy became the high-level champion for Information Age Government, covering both e-commerce and e-government strategy. Today the OeE is leading the drive to get the UK online, ensuring that the country, its citizens and its businesses derive maximum benefit from the knowledge economy.

The Office of the e-Envoy has three core objectives:

  * to make the UK the best environment in the world for e-commerce by 2002;
  * to ensure that everyone who wants it has access to the Internet by 2005; and
  * to make all government services available electronically by 2005.

Much of the effort to achieve this is based around UK Online and the Government Gateway. The UK Online programme aims to make transactions available on-line in a joined up fashion by the end of 2005. The most visible aspect of this at the moment is the UK Online portal, aimed at both the citizen and businesses.

The UK Online Citizen Portal (www.ukonline.gov.uk) provides a single point of entry to a wide range of government information and services. Its content is organised around the needs of the citizen, making dealing with government as easy and seamless as possible. Building on recommendations in the Modernising Government White Paper, information is focused around 'Life Episodes', which enable the user to access all the information they need about a particular event such as 'Having A Baby' or 'Learning to Drive' without having to understand the workings of government or departmental delivery structures.

The Portal provides a structured way of accessing and linking to existing information available online. However, it is merely the tip of the iceberg. Behind the scenes, systems provide the information, security, search facilities and other structures that form the complex machinery necessary to run seamless and transparent services that require minimum effort and knowledge on behalf of the users.

The Government Gateway is fundamental to implementing this seamless, joined-up service. It acts as the middleware - a self contained and secure infrastructure that sits between Government's back office systems and the applications layer such as departmental websites, portals (including ukonline.gov.uk) and commercial applications. It has been built to provide "always on" availability, scalability and high volume capacity.

The vision is that citizens and businesses will have individually tailored seamless and secure access to services and information from all government organisations at any time.

The Gateway went live for registration and enrolment on 25 January 2001. Services currently available are:

* Electronic VAT Return (for VAT-registered businesses);
* Department for the Environment, Food and Rural Affairs (DEFRA) Integrated Administration and control system (IACS) Area Aid Application (for farmers and agents who complete DEFRA forms on behalf of farmers); and
* PAYE Internet Services (for employers and their agents).

In the future, by using the Portal in conjunction with the Government Gateway, it will be possible to carry out online transactions with multiple government departments, such as registering a birth, from a single point of entry

The use of a central common infrastructure will also provide economies of scale through a 'build-once, use-many' architecture. For example, when the Gateway implements a new piece of functionality, that functionality will be available to all departments connected to the Gateway, removing the need for multiple individual implementations.

Standardisation of the way in which information is formatted is vital. The UK public sector is made up of several thousand different organisations, including central government departments and agencies, non-departmental public bodies, the National Health Service, schools, the police, advisory councils, committees, museums and local authorities. Not only do these organisations provide services externally; they also provide information and services to each other.

e-Government Metadata Framework: the e-GMF

The e-Government Metadata Framework (e-GMF) was published on 3 May 2001 [6]. The result of several months of consultation and planning, it sets out the UK Government's policy for standardising metadata use throughout the public sector. The main drivers behind the e-GMF were

* the need to make government information easily available to businesses, members of the public, and indeed to civil servants themselves, without them having to be familiar with the structure of government or know what each organisation is responsible for
* the need to develop electronic records management systems in line with the guidelines laid down by the UK Public Record Office
* the requirements of the UK Freedom of Information Act and Data Protection Act
* the continuing need to improve efficiency

Traditionally, information resources have been stored in a number of different repositories, and treated differently in each case. It is not unknown for one document to exist on the internet, an intranet, in the organisation's electronic network and in hard copy in an organisation's official record keeping system. In each of these circumstances different information needed for the retrieval and handling of the resource would have been created, often by different staff.

This involves unnecessary repetition of work, and as practically all information being generated now is held in electronic format it is becoming increasingly clear that the boundaries between record keeping systems, internet sites and intranets are breaking down. The information held in each environment overlaps more and more, especially as the Freedom of Information Act encourages the active publication of all information (except that which is exempt). It has become clear that metadata associated with an information resource can be used throughout the resource's life, from its beginnings as a draft document for internal use only, through to when it is put on the department's web site, and remaining with it when it is archived. The metadata will change throughout this life-cycle, new data will be added and existing data amended, but basic details, such as title, subject and date created, will remain.

The e-GMF was therefore designed to provide for both resource discovery and records management needs. It outlines the policy and use of metadata in government, lists the principles of development of the metadata structure and standard and covers management

and some implementation issues. The e-GMF is complemented by the e-Government Metadata Standard (e-GMS), which describes the elements and refinements.

The main thrusts of the policy are;
    *   the mandate of the addition of metadata to information resources in systems that involved information being exchanged with any other public sector system;
    *   the adoption of simple Dublin Core as the e-Government Metadata Standard, to be enhanced as necessary; and
    *   the study and development of a controlled vocabulary.

Developing the e-GMS

Dublin Core (DC), though seen as the ideal starting point, is not sufficient for the varied and specialised needs of government. It doesn't cater for records management, data security, or the requirement of the Data Protection Act or the Freedom of Information Act. Working closely with the UK Public Records Office (PRO) and others, the OeE is extending DC to create an element set comprehensive enough to cope with the job in hand.

The first step in creating the standard was to lay down the principles guiding the decisions regarding the choice of elements, refinements, and future changes. These were listed in the Framework, and are:

1. 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.

2. It will be **Simple** to use. The standard must be readily applicable by those with widely varying experience of preparing resource descriptions.

3. It will be **Compliant with other UK Government standards** and policies, such as the e-GIF standards and the Government Data Standards Catalogue.

4. 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.

5. 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.

6. 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.

7. It will be economical and give **Value for Money**.

8. 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.

9. **Above all**, it will meet the information retrieval and management needs of the citizen and of government.

In addition to a set of elements and refinements, the e-GMS will include guidance for information managers and other background material. Because it covers a broad range of applications, including records management systems, internet sites and intranets, not all of the elements will be needed in all situations, so guidance is given on selecting elements appropriate for different information systems.

At the time of writing the draft e-Government Metadata Standard (e-GMS) is undergoing thorough testing in a variety of 'live' situations, government websites, intranets and document management systems. Testing is essential for three main reasons,
    * to ensure the elements are adequate for the many purposes we have for them;
    * to ensure that the many public servants who will be adding metadata to their documents find the refinements and guidelines to be meaningful and usable; and
    * to get documentary evidence that metadata improves information retrieval and

management enough to justify the time and effort need to add it to documents and files.

Once the testing is completed the draft element set will be edited as necessary. It will then officially become the UK Government standard, to be formally launched in the latter part of 2001.

More elements

The new draft e-GMS is based on the fifteen simple Dublin Core elements. The refinements have been based on DC but adjusted to take into account of our specialist requirements, and four extra elements have been added; Audience, Disposal, Location and Preservation.

Audience is an element being adopted by many organisations, and appears routinely in draft proposals for amendments or variations to the main DC element set. The best known example is probably AGLS, the Australian Government Locator Service. 'Audience' should help in refining searches, and in an ideal world users would be able to set the search engine to default to their preferred audience level. The main difficulty presented by this element will be the encoding scheme, we still do not have a definitive list of audiences, and some of the benefits of this element will be lost of a multitude of schemes is used.

The **Disposal** element holds data relating to the time a record should be reviewed and a decision taken as to its further retention or disposal. Disposal could be seen mainly as a records management and archiving requirement, but webmasters should find it invaluable to be able to flag web pages with review dates, to help keep sites up to date. Current web management practices, for large organisations in particular, often hold the owners or authors of information responsible for ensuring it is updated or removed, but staff move on, or get involved in new projects, and rarely remember to check.

Refinements for the Disposal element have been taken from PRO guidelines. They are:
  * review (the date on which the resource should be reviewed to determine the need to retain it)
  * Conditions (a specific period of time following a specific event determining the period for which the resource must be kept, e.g. 'Five years after completion of programme')
  * Action (the action to be taken when the condition is reached)
  * Review details (details of reviewers and decisions taken)

An example of how this could be used to help manage a web resource involves a web page providing information on the organisation's current work plan. The page contains the statement 'Work plan January 2002 – June 2003'. The metadata will include

Disposal.Review	2003-05-25
Disposal.Action	Replace or move to archive

This can then be used, either automatically or manually, to alert the webmaster or content owner of the need to replace the page with an updated version before it becomes out of date.

Preservation is another information management element, and the long term storage of electronic records presents enough difficulties for it to be an essential element for the foreseeable future. If files are to be readable in the long term, details about file type and version, software requirements, file sizes etc are essential. This is particularly important for records that consist of more than just text, such as images, databases, and snapshots of entire web sites.

Refinements for this element are being developed by the PRO. They will be very extensive, covering many file and software application types.

The final element we have added is **Location**. Although originally meant for items held in a physical format, such as paper files, books and photographs, it will also be useful for describing electronic information held on tapes or floppy disks. Even for electronic media space is limited, older records and documents are often saved to disc and placed in storage centres, while the metadata is kept separately to allow retrieval. In this event it is essential to know not only which disc the document was stored on, but where the actual disc is kept. No refinements have been set for Location.

One other way in which we have varied the DC Elements set is by listing the elements in alphabetical order in our documentation. Although only a small change, it is the subject of more comment than perhaps any other aspect of the e-GMS. The thinking behind it is simple; the documentation is designed to be used not only by information professionals but civil servants across the UK public sector. They will quickly become familiar with the most common elements, Creator, Date, Title, and will probably only need to refer to the documentation for help with an unusual or unfamiliar situation. Having the elements listed alphabetically makes it easier for them to find the information they need.

e-GMS refinements

Dublin Core was the first port of call when refinements to the elements were being examined, though we also needed to take into account some specialised requirements. In addition to the records management requirements, these included:

* the need to describe collections, databases and services as well as documents and web pages
* official classification systems and other access requirements
* the needs of statisticians, geographers and other specialists
* the need to map to existing systems

It is expected to be several years before all of the refinements can be regarded as completed.

Words

One of the most difficult aspects of ensuring information can be found involves language. The problem is a painfully familiar one; even within a single sector of an essentially monolingual nation. Different government departments and agencies, each with their own area of specialisation, give different meanings to the same words and phrases, then complicate matters further by giving different words and phrases to the same concept. Technical terms and acronyms abound, providing simplicity and brevity to those familiar with them, and total confusion to everybody else.

This leads to the familiar problem of not being able to find services or information through not knowing what word to type into the search box, or of finding thousands of irrelevant pages because the machinery doesn't know the difference between greenhouse (gardening) and greenhouse (gases). The lack of a common vocabulary amongst those who speak the same language is one of the greatest barriers to finding what is needed without being crushed in an information avalanche.

The OeE has carefully examined the practicalities of creating a comprehensive thesaurus for use across government. In doing this we took into account the research being carried out by HILT, (hilt.cdlr.strath.ac.uk) the High Level Thesaurus Project which reports on the problems of cross-searching and browsing by subject across a range of communities. We also looked at what other countries, especially Australia and Canada, were doing, and held a one-day workshop with representatives from all UK government departments to discuss their requirements and decide the preferred solution and a plan of action.

The study into the optimum size and scope of the Pan-Government Thesaurus (available at www.GovTalk.gov.uk/interoperability/metadata.asp) concluded that developing a single, traditional ISO-standard thesaurus for use across government was not a practical solution. Instead we will develop a 'virtual thesaurus' consisting of terms from existing indexing and controlled vocabulary tools. These will be mapped via a central point. This solution allows us to take advantage of the considerable effort that has already gone into vocabulary control within departments and allow them to use their specialist terms, while still allowing the citizen seamless access to government information.

Part of this mechanism will be a set of broad subject terms, known as the Government Category List (GCL). This will be used to provide easy subject browsing, from a broader subject such as 'Health and social care' to a narrower one such as 'Dental health'. It will also act as the backbone of the thesaurus mapping system. When completed it will be made available to government, industry and the citizen. An early draft of the GCL is available at appendix 1.

Knowledge Network

The Government Category List will be based on the Policy Category List, a set of about 250 terms currently being used to aid information retrieval on the Knowledge Network (KN). This is a government-wide, 24-hour electronic communications tool that provides access to briefings, facts and figures on agreed policy. The need for cross departmental information sharing is a key factor in its development, and the KN team is working with key government departments to ensure their individual departmental briefing systems are compatible, using the e-Government Metadata Standard as a linking tool. The departmental systems will then be linked to the Knowledge Network and made available across the whole of the government.

Whilst the project is at an early stage, it is already a prime example of the benefits of implementing metadata standards across a range of organisations.

Europe and beyond

The OeE is always conscious that information is a global resource, and that we are increasingly being called upon to provide services with a European flavour. Contact with our European and other international partners has revealed that many

governments are in a similar position to ours, trying to find solutions to the same problems.

In June this year we were instrumental in organising a two day conference looking at metadata aspects of organising government information. Hosted by the European Commission's Interchange of Data between Administrations (IDA) (europa.eu.int/ISPO/ida/) programme, the European Standards Organisation (CEN) (www.cenorm.be/default.htm), and the DCMI, this was the first in a series of events under the title Managing Information Resources for e-Government (MIReG). Reports from delegates from eleven countries demonstrated that many governments are in a similar position to ours, trying to find solutions to the same problems. Most of them have been looking at an enhanced version of Dublin Core to meet their needs, and all are struggling to find solutions to the vocabulary problem, often in a multilingual environment.

It was clearly agreed that standardisation of metadata structures across Europe and the rest of the world could not only save each government from having to develop its own system, but improve the flow of information internationally. A project to achieve this has been set up under the auspices of IDA, with strong assistance and support from CEN and DCMI.

The project has two aspects. The first is the creation of a set of DC extensions, to be know as DC-gov. This will be presented to the DCMI Usage Board at DC-2001 in Tokyo. The second is the development of an EC Metadata Framework, based on the UK model and supported by CEN and IDA. This is due for completion early in 2002.

It is hoped that encouraging cooperative working and developing a common standard will ease the management of information, and simplify interaction with government for both the citizen and businesses. It should also give software application developers confidence in the size of their market and encourage them to develop the solutions we need to implement a metadata standard.

Finally

New technology offers fantastic opportunities to transform the way governments interact with their citizens, businesses and each other. The opportunities are not without their challenges, and a good marriage of technology and information management techniques is still some way ahead. Increasingly national governments and others are realising the benefits of standardising their information systems, not only internally but in conjunction with other nations, standards organisations and organisations from different sectors.

Standards can help the international flow of information, and break down barriers between nations. The Office of the e-Envoy in the UK is taking the lead in developing these, a lead which will benefit other governments as well as UK citizens and businesses. The OeE is working closely with industry, as well as government departments and agencies, other governments and international bodies, to help ensure the diverse needs of different sectors are catered for.

References

[1] The Cabinet Office. Modernising Government., March 1999. Cm 4310. The Stationery Office, Norwich, 1999. ISBN 0101431023. www.cabinet-office.gov.uk/moderngov/whtpaper/index.htm

[2] Central IT Unit. e-government: a strategic framework for public services in the information age. Central IT Unit, London, April 2000. ISBN 0 7115 0394.X

[3] Cabinet Office, Performance and Innovation Unit. e-commerce @ its best. Cabinet Office, Performance and Innovation Unit, September 1999. www.cabinet-office.gov.uk/innovation/1999/ecomm.shtml

[4] Department of Trade and Industry. Our competitive future: Building the Knowledge Driven Economy, Cm 4176. The Stationery Office, Norwich December 1998. ISBN 0101417624 www.dti.gov.uk/comp/competitive

[5] Cabinet Office, Office of the e-Envoy. e-Government Interoperability Framework First revision. Office of the e-Envoy, April 2001. www.govtalk.gov.uk/egif/home.html

[6] Cabinet Office, Office of the e-Envoy. e-Government Metadata Framework. Office of the e-Envoy, May 2001 ISBN 0711504083 . www.govtalk.gov.uk/egif/home.html

Appendix 1
First draft GCL – Top Two Levels only

Agriculture, environment and natural resources

Energy and fuel
Environmental protection
Farming
Fisheries
Forestry
Horticulture
Land
Plants and animals
Waste management
Water resources

Business and industry

Business practice and regulation
Business sectors
Consumer affairs
Energy and fuel
International trade
Small and medium enterprises
Transport

Crime, law, justice and rights

Animal rights and welfare
Civil and human rights
Crime
Emergencies
Ethical issues
Extradition
Firearms
Justice system
Law
Security

Economics and finance

Capital and financial markets
Economic development
Euro and EMU
Investment
Monopolies and mergers
Nationalisation/privatisation
Personal finance
Public finance
Tax
UK economy

Education and employment

Adult learning and skills
Employment relations
Equal opportunities
Further and higher education
Occupational health & safety
Pre-school
Schools
Unemployment and jobseeking
Working hours, terms and conditions

Government, politics and public administration

Central government
Civil Service
Constitution
Devolved administrations
Electoral system
Honours system
Local government
Political parties
Public administration
Regional policy

Health, care and nutrition

Animal health
Benefits
Care
Carers and health professionals
Disabled people
Drugs and alcohol
Family planning
Food and drink
Health
National Health Service (NHS)
Private health care
Safety

Information and communication

Communications industries
Data security
Freedom of information
Intellectual property
Intelligence
Internet
Libraries and archives
Media and the press
Privacy and data protection

International affairs

Defence
Embassies and consulates
European Union
Extradition
Foreign policy
Immigration and asylum
International conflict
International development & aid
International organisations
International relations

Leisure, recreation and travel

Arts and culture
Gambling and lotteries
Parks and gardens
Passports and visas
Sports and recreation
Tourism
Transport

People, communities and living
      Benefits
      Care
      Charities
      Consumer affairs
      Equal opportunities
      Housing
      Marriage and divorce
      People
      Population and migration
      Race relations
      Religion
      Rural communities
      Social exclusion
      Urban communities

Science, technology and innovation
      Biotechnology
      Information technology
      Research & Development