briefing paper

digital preservation europe

The Uses of Metadata in Public Administration

Efficiently managing enormous quantities of government data requires the development and deployment of robust, sustainable and interoperable metadata regimes. Governments need metadata to manage, understand, enable access to, and preserve their vital data assets over time and across domains of use. Different communities of practitioners think of metadata differently because of its different uses. ICT professionals think of metadata as data that describes data and data systems: that is, the structure of databases, their characteristics, location and usage. Information management professionals, on the other hand, regard metadata as structured information that describes and/or enables finding, managing, controlling, understanding or preserving other information over time. In other words, metadata documents the content, context and structure of information resources in order to support the ongoing use of those resources.

Objectives

Governments are now heavily reliant on the use of information and communications technology, and data or digital information is now the currency of public administration. In every country public sector information represents a significant component of the overall information and knowledge economy. Government data provides the evidence base that informs government decision making. It provides the virtual memory of the government's past decisions and activities, thus allowing government to account for itself. Public sector information is an often underutilised national strategic resource. As such, it is vital for public sector information to be efficiently managed in ways that maximise the returns on this enormous national investment.

Uses of metadata in public administration include the following:

- Data management (eg for statistical data sets)
- Resource discovery (eg standardised descriptions of government online resources and e-government services)
- Recordkeeping
- Managing and enabling the use of geospatial data sets
- Privacy protection
- Information rights management (eg for intellectual property management and security management)
- Digital preservation (ie ensuring the longevity and continuity of vital data assets)
- Documenting the levels of compliance of information resources with World Wide Web Consortium accessibility standards.

Inadequate metadata results in poor management and underutilisation of public sector information assets, thus wasting public monies. For example, inadequate metadata may result in:

- The failure to locate and/or share important information when it is needed because of a lack of adequate descriptive metadata;
- The inability to be able to read/use digital information due a lack of technical metadata about the structure and technical properties of the digital information object;
- An inability to attribute meaning or value to information due to a lack of contextual metadata; and
- An inability to verify the authenticity and reliability of information.

Experience

In a networked world where 'joined up Government' is an imperative, metadata usually needs to be able to be shared/exchanged and reused by different entities for different purposes, usually by automated systems. Because machines need predictability to successfully process metadata (ie. to be interoperable), it is important for metadata regimes to comply with accepted industry standards. Metadata standards standardise one or more of three main aspects of metadata:

1. Structure (how the metadata is structured – often into elements of information or 'properties' consistent with an explicit data model or ontology);

2. Semantics (what the metadata elements or properties mean); and

3. Syntax (how the metadata is written/expressed/encoded using common mark-up languages such as HTML and XML and data values consistent with designated controlled vocabularies and encoding schemes).

Some important metadata standards used in public administration include:

• ISO/IEC 11179 – Metadata Registry Standard.
Metadata registries provide a reliable source of information on the meaning of a given metadata element or piece of metadata. ISO-compliant metadata registries consist of a hierarchy of 'concepts' with associated properties for each concept. Each concept and property must have a precisely worded data element definition.
• ISO 15836 - Dublin Core Metadata Element Set (a standard commonly used by governments for online resource discovery. For example, Australian governments have adopted the AGLS Metadata Standard, which is an extension and application profile of Dublin Core and is used to describe government information and services).
• ISO 19115 – Geographic Information – Metadata.
• ISO 23081 – Metadata for Records – Part 1: Principles; and Part 2: Conceptual and Implementation Issues.
• ISO 8601 – a standard for encoding date and time information.
• PREMIS – Preservation Metadata Implementation Strategies (2005).
• METS – Metadata Encoding and Transmission Standard.
• SCORM – Sharable Content Object Reference Model (for e-learning).
• RDF – Resource Description Framework (a family of World Wide Web Consortium (W3C) specifications, originally designed as a metadata data model, which is used as a general method of modeling information through a variety of syntax formats).
• MODS – Metadata Object Description Schema (US Library of Congress).
• SOAP – Simple Object Access Protocol.
• OAI – Open Archives Initiative (metadata harvesting protocol).

It is important to note that, because the different uses of metadata are not mutually exclusive, many of these standards often have some areas of overlap. Also, many of these standards have jurisdiction-specific and sector-specific (eg health, education) manifestations or application profiles which extend the scope or level of granularity of the relevant international standard.

When developing and implementing metadata regimes there are a variety of issues and considerations that should be addressed. These include:

• Human beings generally dislike creating metadata manually, as they often regard it as an onerous imposition on their already busy schedules. It is therefore important for metadata creation and management to be as automated as possible.

• Close attention needs to be given to metadata quality assurance. It is generally not difficult to create good metadata, but it is also very easy to create bad metadata. Bad metadata is worse than no metadata.

References, part1

AGLS Metadata Standard
http://www.naa.gov.au/agls

Davies, John, Semantic Web Technologies: Trends and Research in Ontology-based Systems . Wiley, 2006. ISBN 0470025964

Day, Michael, 'Metadata', in: DCC Digital Curation Manual (2005). This instalment of the DCC Digital Curation Manual is a good introduction to metadata and preservation metadata, especially within the context of digital curation. It provides definitions, discusses the importance of, and uses for, metadata. There are also sections devoted to the OAIS model, preservation metadata, standards and package formats such as METS.
http://www.dcc.ac.uk/resource/curation-manual/chapters/metadata/

Dublin Core Metadata Initiative
http://www.dublincore.org/

Dublin Core Metadata Initiative Glossary
http://www.dublincore.org/documents/usageguide/glossary.shtml

Evans, Joanne, Barbara Reed and Sue McKemmish, 'Interoperable Data: Sustainable Frameworks for Creating and Managing Recordkeeping Metadata', Records Management Bulletin, vol. 18, no. 2, 2008, pp. 115-129.

Introduction to Ontologies and the Semantic Web.
http://obitko.com/tutorials/ontologies-semantic-web/

References, part 2

National Archives of Australia, Australian Government Recordkeeping Metadata Standard version 2.0, 2008. http://www.naa.gov.au/records-management/create-capture-describe/describe/RKMS/index.aspx

PREMIS (PREservation Metadata: Implementation Strategies) Working Group, Data Dictionary for Preservation Metadata : Final Report of the PREMIS Working Group, May 2005 (United States of America) http://www.oclc.org/research/projects/pmwg/premis-final.pdf

Tambouris, Efthimios and Konstantinos Tarabanis, 'An Overview of DC-Based e-Government Metadata Standards and Initiatives', Lecture Notes in Computer Science, Springer, Berlin, 2004, pp. 40-47.

Author: Adrian Cunningham (National Archives of Australia)

• Metadata implementations need to give careful consideration to what metadata needs to be created and maintained. All metadata creation and maintenance comes at a cost, and all such costs need to be justified by business needs. Metadata should not be created and maintained for the sake of it. Organisations should only create and maintain metadata that will deliver a demonstrable return on investment.

• Metadata regimes can be simple or highly complex. Generally speaking, the more simple the metadata the better because complex metadata is expensive to maintain and more difficult to quality assure. However, simple metadata may not always deliver the desired business outcomes – so, greater levels of complexity and precision may be pursued when justified by clearly understood business drivers.

• Because public sector information is dynamic and ever evolving, it is important for metadata to be kept up to date to reflect the changing nature of the information resources to which the metadata relates. Metadata should not be static, but rather should be actively maintained, managed and updated.

• As a general rule the technical infrastructure supporting metadata implementation should rely on flexible rather than hard-wired enterprise architectures. Service Oriented Architecture (SOA), for example, promises to provide an ideal approach to implementing flexible, dynamic, interoperable and reusable metadata.

Metadata and 'The Semantic Web'

The Semantic Web is an evolving extension of the World Wide Web in which the semantics of information and services on the web is defined, making it possible for the web to understand and satisfy the requests of people and machines to use the web content. It derives from World Wide Web Consortium director Sir Tim Berners-Lee's vision of the Web as a universal medium for data, information, and knowledge exchange. At its core, the semantic web comprises a set of design principles, collaborative working groups, and a variety of enabling technologies. Some elements of the semantic web are expressed as prospective future possibilities that are yet to be implemented or realized. Other elements of the semantic web are expressed in formal specifications. Some of these include Resource Description Framework (RDF), a variety of data interchange formats (e.g. RDF/XML, N3, Turtle, N-Triples), and notations such as RDF Schema (RDFS) and the Web Ontology Language (OWL), all of which are intended to provide a formal description of concepts, terms, and relationships within a given knowledge domain.

Public sector metadata promises to be a key enabler of the Semantic Web. Government metadata implementations should therefore consider the potential benefits of compliance with Semantic Web protocols such as RDF and OWL.

Author: Adrian Cunningham, *National Archives of Australia*
adrian.cunningham@naa.gov.au