Chapter 1
Introduction

overview

THIS CHAPTER INTRODUCES the idea of metadata and illustrates it with some early examples of the use of metadata concepts before the term 'metadata' was coined. The development of metadata is placed in the context of the history of cataloguing as well as parallel developments in other disciplines. This leads to discussion of the definitions of 'metadata' and a suggested form of words that is appropriate for this book. Some examples of metadata from library catalogues and websites are used to illustrate the concept. The chapter then considers why metadata is important in the information culture that many people work in. This provides a way of assessing the models of metadata in terms of its uses. The chapter concludes with a five-point model for the use (or purposes) of metadata.

Metadata is important in the information society. It is having a profound impact on most aspects of information work and is an enabler of the information systems that underpin the knowledge economy, e-commerce and e-government. Understanding what metadata is and how it works is crucial for those working in the knowledge, information, cultural and learning sectors. Use of information modelling techniques associated with metadata analysis will help the next generation of information services providers and systems designers to deliver more effective services and systems. General users need to understand how metadata works in order to make best use of resources that are available on the internet, for instance, and to have an idea of where things might be going in future.

The historical background is the starting point for an understanding of metadata, what it is, and how it is used.

Some historical background

Library catalogues

Although the term 'metadata' is a recent one, many of the concepts and techniques of metadata creation, management and use originated with the development of library catalogues. Books (and scrolls) are repositories of information and a catalogue contains data about that information and can therefore be regarded as metadata. An understanding of what people tried to do before the term 'metadata' was coined helps to explain the concept of metadata. The historical background also gives a perspective on why metadata has become so important in recent years.

The idea of cataloguing information has been around at least since the Alexandria Library in ancient Egypt. Callimachus of Cyrene (305–235BC), the poet and author, was a librarian at Alexandria. He is widely credited with creating the first catalogue, the 'Pinakes', of the Alexandria Library's 500,000 scrolls. The catalogue was itself a work of 120 scrolls with titles grouped by subject and genre (Ellens, 1997). This was in effect the first recorded compilation of metadata.

In Western Europe library cataloguing developed in the ecclesiastical and, later, academic libraries. In the eighth century the books donated by Gregory the Great to the Church of St Clement in Rome were catalogued in the form of a prayer. During the same era, Alcuin of York (735–804) developed a metrical catalogue for the library at York Cathedral. Cataloguing developed so that by the 14th century the location of books started to appear in catalogue records and by the 16th century the first alphabetical arrangements began to appear. Up until that time catalogues were used as inventories of stock rather than for finding books or for managing collections.

Modern library catalogues date back to the French code of 1791, the first national cataloguing code with author entry, which used catalogue cards and rules of accessioning and guiding. Cataloguing rules (an important aspect of metadata) were developed by Sir Anthony Panizzi for the British Museum Library and these were published in 1841. In the USA Charles A. Cutter prepared *Rules of a Dictionary Catalog*, which was published in 1876. The American Library Association and the Library Association in the UK both developed cataloguing rules around the turn

of the century. This led to an agreement in 1904 to co-operate to produce an international cataloguing code which was published in separate American and British editions in 1908.

Later, the International Conference on Cataloguing Principles in Paris in 1961 established a set of principles on the choice and form of headings in author/title catalogues. These were incorporated into the first edition of the *Anglo-American Cataloguing Rules* (AACR) in 1967, published in two versions by the Library Association and the American Library Association. The Canadian Library Association, the British Library and the Library of Congress were closely involved in the Joint Steering Committee for the Revision of the AACR (Joint Steering Committee for Revision of AACR, 2002a).

The International Standard Bibliographic Descriptions (ISBDs) were developed by IFLA and have been incorporated into the second edition of the *Anglo-American Cataloguing Rules* (AACR2), published in 1988 and revised in 2002; they have been adopted by the Library of Congress, the National Library of Canada, the British Library and the Australian National Library (Joint Steering Committee for the Revision of AACR, 2002b). AACR2 specifies the sources of information used to describe a publication, the order in which the data elements appear and the punctuation used to separate the elements. This was an important development because it made catalogues more interchangeable and allowed for conversion into machine-readable form (Bowman, 2003).

Electronic catalogues

In the mid-1960s computers started being used for the purpose of cataloguing and a new standard for the data format of catalogue records, MARC (MAchine Readable Cataloguing), was established. MARC covers all kinds of library materials and is usable in automated library management systems. Although MARC was initially used to process and generate catalogue cards more quickly, libraries soon started to use it as a means of exchanging cataloguing data, which helped to reduce the cost of cataloguing original materials. The availability of MARC records stimulated the development of searchable electronic catalogues. The user benefited from wider access to searchable catalogues, and later on to union catalogues, which allowed them to search several library catalogues at once. Different versions of MARC emerged, largely based on national variations, for example USMARC, UKMARC and NORMARC. Although the different MARC versions were designed to reflect the particular needs and

interests of different countries or communities of interest, they inhibited international exchange of records. It is only with the recent widespread adoption of MARC21 by the national bibliographic authorities that a degree of harmonization of national bibliographies is being achieved.

The growth of electronic catalogues and the development of textual databases able to handle summaries of published articles demanded new skills, which in turn contributed to the development of information science as a discipline. Information scientists developed many of the early electronic catalogues and bibliographic databases (Feather and Sturges, 1997). They adapted library cataloguing rules for an electronic environment and did much of the pioneering work on information retrieval theory, including the measures of precision and recall which are discussed in Chapter 5.

Although metadata was first used in library catalogues it is now widely used in records management, the publishing industry, the recording industry, government, the geospatial community and among statisticians. Metadata is widely used because it provides the tools to describe electronic information resources, allowing for more consistent retrieval, better management of data sources and exchange of data records between applications and organizations.

Origin of the term

Vellucci (1998) suggests that the term 'metadata' dates back to the 1960s but became established in the context of Database Management Systems (DBMSs) in the 1970s. Woodley (1999) traces the first reference to 'meta-data' back to a PhD dissertation on 'An infological approach to data bases' which made the distinction between:

- objects (real-world phenomena);
- information about the object; and
- data representing information about the object (i.e. metadata).

The term had begun to be widely used in the database research community by the mid-1970s in Northern Europe.

A parallel development occurred in the geographical information systems (GIS) community and in particular the digital spatial information discipline. In the late 1980s and early 1990s there was considerable activity within the GIS community to develop metadata standards to encourage interoperability between systems. Because government (especially local government) activity often requires data to describe location, there are significant ben-

efits to be gained from a standard to describe location or spatial position across databases and agencies. The metadata associated with location data has allowed organizations to maintain their often considerable internal investments in geospatial data, while still co-operating with other organizations and institutions. Metadata is a way of sharing details of their data in catalogues of geographic information, clearing houses or via vendors of information. Metadata also gives users the information they need to process and interpret a particular set of geospatial data.

Metadata initiatives

In the mid-1990s the idea of a core set of semantics for web-based resources was put forward for categorizing the web and to enhance retrieval. This became known as the Dublin Core Metadata Initiative (DCMI), which has established a standard for describing web content and which is not discipline- or language-specific. The DCMI defines a set of data elements which can be used as containers for metadata. The metadata is embedded in the resources so that it is accessible to users and systems that use the web resource. DCMI is an ongoing initiative which continues to develop and elaborate the metadata standards for web applications (see Chapter 3, page 51 for more details).

This position is questioned by Gorman (2003) who suggests that metadata schemes such as Dublin Core are merely subsets of much more sophisticated frameworks such as MARC. He suggests that without authority control and use of controlled vocabularies, Dublin Core and other metadata schemes cannot achieve their aim of improving the precision and recall from a large database (such as web resources on the internet). His solution is that existing metadata standards should be enriched to bring them up to the standards of cataloguing. However, his arguments depend on a distinction being drawn between 'full cataloguing' and 'metadata'. An alternative view (and one supported in this book) is that catalogues are a form of metadata.

All of these approaches to metadata have begun to come together as the different communities have become aware of the others' activities and have started to work together. The DCMI involved the database and LIS communities from the beginning with the first workshop in 1995 in Dublin, Ohio, and has gradually drawn in other groups that manage and use metadata. Work has also gone into the development of crosswalks of standards so that it is possible to map the data elements of one metadata standard on to those of another. Bodies such as the International Orga-

nization for Standardization (ISO) are currently developing metadata standards and this activity is described in detail in Chapter 3.

Looking at existing trends, therefore, metadata is becoming more widely recognized and it is being included in the specification of IT applications and software products. For example, the UK government's specification for electronic document and records management systems specifies minimum metadata standards. In addition, content management systems are increasingly being called on to handle resources that contain embedded metadata so that the resulting web and intranet pages are retrievable and can be exchanged between different systems. As a result, manufacturers of digital hardware and suppliers of software applications are incorporating metadata standards into their products and this in turn will stimulate further uptake of metadata.

This brief history of metadata demonstrates that it had several starting points and arose independently in different quarters. In the 1990s, there was growing awareness about metadata, and the work of bodies such as the Dublin Core Metadata Initiative has done a great deal to raise the profile of metadata and its widespread use in different communities. It has become an established part of the information environment today. However, its history does mean that there are distinct differences in the understanding of metadata and it is necessary to develop some universal definitions of the term.

What is metadata?

At this stage it is worth interrogating the concept of metadata more fully. The previous section described how the concept of metadata arose from several different intellectual traditions. This is reflected in the different uses of metadata that are described later in this chapter (pp 12–17). They depend on the priorities of the communities that are using it. This leads to speculation on whether there is a common understanding of what metadata is, and whether there is a definition that is generally applicable.

Metadata was originally referred to as 'meta-data', with an emphasis on the two-word fragments that make up the term. The word fragment 'meta', which comes from the Greek μετα, translates into several distinct meanings in English. In this context it can be taken to mean a higher or ulterior view of the word it prefixes. In other words metadata is data about data, or data that describes data or information.

Many of the available definitions focus on the function that metadata performs. For instance Tozer looks at metadata from the perspective of

database management systems. So in this definition metadata is for control of data and it is seen as contributing to business success:

> A more useful perspective is gained by seeing metadata as the means by which the structure and behavior of data is recorded, controlled, and published across an organization. (Tozer, 1999, xix)

Some in the library and information community have taken a similar approach, also defining metadata in terms of function or purpose. However in this context metadata has more wide-ranging purposes and includes purposes such as retrieval and management of information resources. It also includes rights management and information about the provenance of data. Aspects of access control relate to the information retrieval purpose of metadata:

> any data that aids in the identification, description and location of networked electronic resources. . . . Another important function provided by metadata is control of the electronic resource, whether through ownership and provenance metadata for validating information and tracking use; rights and permissions metadata for controlling access; or content ratings metadata, a key component of some Web filtering applications.
>
> (Hudgins, Agnew and Brown, 1999)

Gilliland-Swetland takes a similarly wide-ranging view of the purpose of metadata in her definition. Metadata is described in terms of the range of uses to which it is put. Interestingly she specifically mentions paper documents as being within the scope of metadata. This is a position that is endorsed in this book in contrast to an earlier definition of metadata which focused on digital objects. The Gilliland-Swetland definition of metadata is:

> there is more to metadata than description; a more inclusive conceptualisation of metadata is needed as information professionals consider the range of their activities that may end up being incorporated into digital information systems. Repositories also create metadata relating to the administration, accessioning, preservation, and use of collections. Acquisition records, exhibition catalogs, and use data are all examples of these, even though they are largely still created in paper form. (Gilliland-Swetland, 1998, 1)

For many people, metadata is seen almost exclusively in the context of the internet and access to web resources. This is evident in the definition

devised by the UK government's Office of the e-Envoy (OeE), which is particularly relevant for web resources, again describing it in terms of its uses. It emphasizes resource discovery and also mentions management, without elaborating on it:

> Metadata can be understood as data about data, a tool enabling users, seekers and owners of information resources to find and manage them.
>
> (UK Office of the e-Envoy, 2003)

These detailed definitions or explanations of metadata are based on specific application areas such as libraries or web resources and are too specific to apply across the range of applications that will be considered in this book. Some of them imply a definition by describing the way in which metadata is applied. Although the meaning of metadata depends on the context and the community that is using it, there is some benefit in adopting a general definition that can apply across the full range of contexts that it may be found in.

Although there is an attractive simplicity in the original definition: 'Metadata is data about data', it is not adequate to describe the complexity of the subject and the range of situations in which it might be used.

A further description is proposed to cover the range of situations in which it might be applied, while still making meaningful distinctions from the wider set of data about objects. If the object (say a packet of cereal on the supermarket shelf) is not an information resource, then data about that object is merely data, not metadata. Our definition of metadata is therefore as follows:

> Metadata is data that describes the content, format or attributes of a data record or information resource. It can be used to describe highly structured resources or unstructured information such as text documents. Metadata can be applied to description of: electronic resources; digital data (including digital images); and to printed documents such as books, journals and reports. Metadata can be embedded within the information resource (as is often the case with web resources) or it can be held separately in a database.

What does metadata look like?

Having considered some definitions of metadata, it is useful to see some examples. What does it look like? Some metadata is never seen by humans, because it is transient and used for exchange of data between

Figure 1.1

```
100    1     $aPedley, Paul
245    10    $aEssential law for information professionals/ $cPaul
Pedley
260          $aLondon: $bFacet, $c2003
300          $axviii, 222p.; $c24cm
```

Example of a MARC21 catalogue record

systems. Visible examples of metadata range from HTML metatags on web pages to MARC records used for exchanging cataloguing data between library management systems. It can be expressed in a structured language such as XML (Extensible Mark-up Language) and may follow guidelines or schema for particular domains of activity.

The two examples here show metadata associated with different objects. Figure 1.1 is an extract taken from a library catalogue and describes a book in a library, using the MARC21 standard ISO 2709:1996. It uses a three-digit tag for labelling fields, so that '100' represents the author field, '245' represents the title field, '260' the publication details and '300' the format of the publication. Immediately after the field tag is a two-digit modifier, generally used to give extra information to the computer, such as how the content of the field is to be sorted. Within each field are subfields separated by '$' delimiters. In field 100, subfield $a is the author name. This use of field tags, modifiers and delimiters ensures that different systems are able to interpret the data in a consistent way and know how to treat all information of a particular sort. It also ensures consistent layout of standard documents such as catalogue cards or screen displays. MARC is an early example of metadata used to allow interoperability.

The second example, Figure 1.2, is of metadata embedded in an HTML web page. The 'DC' label refers to the Dublin Core Element Set, a metadata standard for describing web-based resources. This extract is from an official government web page, and demonstrates how the content of the web page can be enriched with metadata.

Figure 1.2

```
<meta name="DC.Identifier" scheme="URI"
content="http://www.doh.gov.uk/nsf/diabetes"/>
<meta name="DC.Creator" lang="en" content="Department of Health, UK"
/>
<meta name="DC.Title" lang="en" content="NHS National Service
Framework for Diabetes: standards"/>
<meta name="DC.Subject" lang="en"
content="doh.gov.uk/nsf/diabetes; diabetes; department of health; NHS,
England; united kingdom; public health; uk; national health service;
NSF; National Service Framework;"/>
<meta name="DC.Description" lang="en"
content="Full text of National Service Framework for Diabetes:
standards - including additional material, service models and
intervention details"/>
<meta name="DC.Date.created" scheme="ISO8601" content="2000-08-27"/>
```

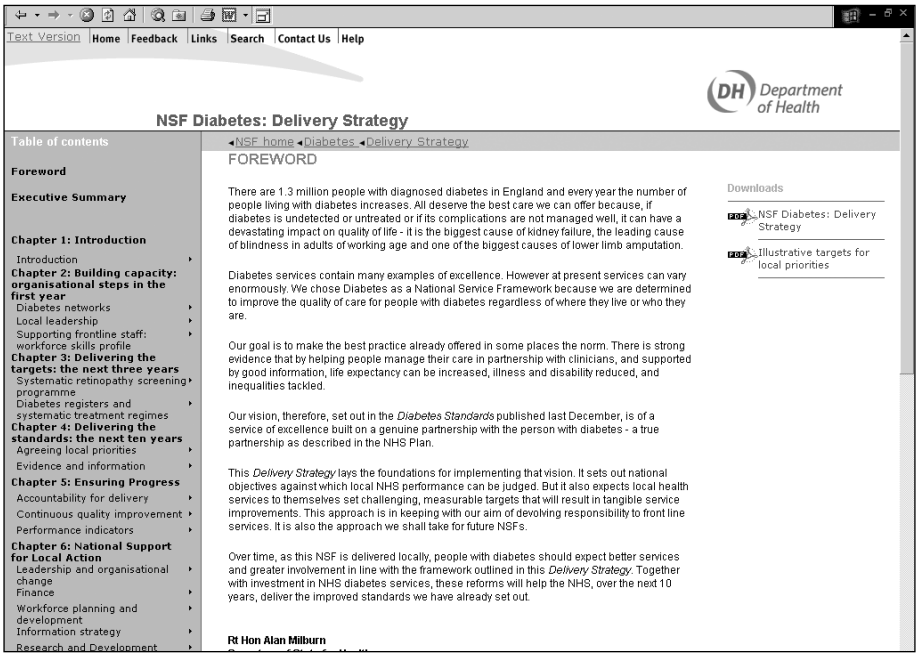
Extract from the source coding of a web page, www.doh.gov.uk/nsf/diabetes/delivery/foreward.htm

The extract shows metadata arranged into different elements corresponding to fields. The name of the metadata standard is embedded in the coding – in this case Dublin Core (see pp 51–4). For instance in the line:

```
<meta name="DC.Creator" lang="en" content="Department of Health, UK"/>
```

the content of the Dublin Core metadata element 'Creator' is expressed in English and the content of that data element (i.e. the creator) is the Department of Health, UK. This kind of element can be generated automatically. The actual web page from which this is extracted is shown in Figure 1.3.

Figure 1.3



Web page containing metadata (UK Department of Health, 2003)

Why is metadata important?

A more comprehensive understanding of metadata can be developed from studying the above examples. The development of cataloguing over two millennia has provided a set of tools for describing published information. This has been drawn on by the web community. Correspondingly the growth of the

internet has focused public attention on the importance of information retrieval and management and has stimulated the development of tools to improve retrieval performance. Having a clear understanding of what metadata is and how it works also provides a means of managing of information resources more effectively. In answering the question 'Why is metadata important?' several arguments emerge:

- *Metadata enhances retrieval performance* – Metadata can improve retrieval by establishing a context for individual descriptors. For instance the word 'Green' in the Creator or Author field indicates the name of an individual, whereas 'green' in the title of a document may be a subject retrieval term. Appropriate metadata tags around the different data elements allow search engines to seek information in a more discriminating way. The presence of a subject field (metadata element) can be used as a prompt for entering keywords, or for use of controlled indexing terms to describe the document. Knowing how metadata works provides information managers with a mechanism for indexing documents more precisely.
- *Metadata provides a way of managing electronic digital objects* – Many software packages use metadata as a way of managing electronic resources, whether it is for records retention schedules or for digital preservation. Content management systems, for instance, use metadata to track when a digital object was last updated or verified, who was responsible for its creation and whether any special access conditions apply. Unlike paper records or printed publications there is not a long tradition of managing digital objects and metadata provides a focus for the establishment of standard practices. It is the metadata associated with digital objects that provides a common format for management and manipulation of resources.
- *Metadata can help to determine the authenticity of data* – Metadata provides an audit trail to establish ownership and authenticity of a digital object such as an electronic document or image. The history of what has happened to a document or record in its life becomes an important part of this. Metadata provides evidence about the provenance of a resource and this underpins good governance, transparency and accountability. This is increasingly important for the many organizations that depend on electronic records rather than paper files. It becomes necessary to demonstrate that the electronic document has been kept securely, it is a complete record, and it has not been tampered with. Metadata provides evidence for the integrity of an electronic document.

This is particularly important in a legal context where electronic documents or physical records may be used, for example as evidence in legal proceedings.

- *Metadata is the key to interoperability* – Interoperability depends on the exchange of metadata between systems to establish the nature of the data being transferred and how it should be handled. E-commerce is one example of interoperating, where several different proprietary systems may need to exchange data. Access to metadata helps to establish the protocols of exchange of data and ways in which it might be exploited. Another example is in the development of e-government initiatives around the world, where there is a move towards common standards for data held by the different government departments or ministries, to allow greater exchange of data and delivery of 'joined-up' government services. Metadata fulfils an important role in enabling this to work, by establishing standards for data elements and by providing information about the data on one system so that it can be processed and used by other systems or departments.
- *Metadata is the future* – An increasing number of software and systems suppliers are working to metadata standards or are creating their own proprietary standards for metadata. The growth of e-commerce depends on metadata for exchange of data between applications. Many industries are developing their own e-commerce infrastructure to allow software from different suppliers to work together and exchange data. As we saw above, the e-government initiatives around the world are underpinned by the concept of interoperability and this in turn often depends on metadata standards. For instance the UK government has developed its own metadata standard, the e-Government Metadata Standard (eGMS) for use on government websites, based on the Dublin Core Metadata Element Set (DCMES). Metadata generated by content management systems is seeing a renaissance on the internet after its initial use for subject description. Metadata standards are being used by portal software and to provide access to the information content of websites.

These five arguments for the importance of metadata provide a way of assessing the purposes to which metadata is put.

The purposes of metadata

The recent emphasis on building consensus between different communities of interest has led to collaborative efforts to develop international

standards for metadata. One of the main drivers for the evolution of metadata standards within each community is the use to which the metadata is put – its purpose. Even within the library and information profession, a range of metadata purposes have been identified. Two of the most useful models have been adopted as a basis for the general model of metadata used throughout this book.

In the first model Day (2001) suggests that metadata has seven distinct purposes. He starts with resource description – identifying and describing the entity that the metadata is about. The second purpose is focused on information retrieval, which in the context of web resources is called 'resource discovery'. This is one of the primary focuses of the Dublin Core metadata initiative. He recognizes that metadata is used for administering and managing resources (Purpose 3) – for instance flagging items for update after set periods of time have elapsed. The fourth purpose, intellectual property rights, is very important in the context of e-commerce, although e-commerce has not been listed as a purpose in its own right. Again this is possibly a reflection of the fact that the model is oriented towards web resources. Documenting software and hardware environments, the fifth purpose, provides contextual information about a resource, but will not apply to every resource. This could be seen as one aspect of resource description. His sixth purpose, preservation management, is a specialized form of administrative metadata and could be incorporated into Purpose 3. Finally providing information on context and authenticity is very important in a number of areas, particularly records management, where being able to demonstrate the authenticity of a record is a part of good governance, and for collection management where the provenance of individual items may affect their value. The seven purposes of metadata identified by Day can be summarized as follows:

1 Resource description
2 Resource discovery
3 Administration and management of resources
4 Record of intellectual property rights
5 Documenting software and hardware environments
6 Preservation management of digital resources
7 Providing information on context and authenticity.

Gilliland-Swetland (1998) and later Caplan (2003) take a slightly different approach, classifying metadata into categories according to its purpose. The use of metadata is categorized into more specific sub-categories. This

means that a metadata scheme as well as individual metadata elements could fall into several different categories simultaneously. Gilliland-Swetland provides some useful examples of the metadata that falls under each of type (see Table 1.1). There is some common ground with Day, in that they both identify administration (equivalent to management and administration), description (encompassing information retrieval or resource discovery) and preservation as key purposes or types of metadata. The technical metadata in Gilliland-Swetland corresponds to 'Documenting hardware and software environments' in Day. The 'Use' metadata could include transactional data as would be seen in an e-commerce system or could provide an audit trail for documents in a records management system.

Table 1.1 Different types of metadata and their functions, extracted from Gilliland-Swetland

Type	Definition	Examples
Administrative	Metadata used in managing and administering information resources	• Acquisition information • Rights and reproduction tracking • Documentation of legal access requirements • Location information • Selection criteria for digitization • Version control
Descriptive	Metadata used to describe or identify information resources	• Cataloguing records • Finding aids • Specialized indexes • Hyperlinked relationships between resources • Annotations by users
Preservation	Metadata related to the information resources	• Documentation of physical condition of preservation management of resources • Documentation of actions taken to preserve physical and digital versions of resources, e.g. data refreshing and migration
Technical	Metadata related to how a system functions or metadata behaves	• Hardware and software documentation • Digitization information, e.g. formats, compression ratios, scaling routines • Tracking of system response times • Authentication and security data, e.g. encryption keys, passwords
Use	Metadata related to the level and type of use of information resources	• Exhibition records • Use and user tracking • Content re-use and multi-versioning information.

There is a lot of common ground between these two models, and although neither of them specifically mentions 'interoperability' as a purpose, it is alluded to. For instance, Day's Purpose 5 – 'Documenting software and hardware environments' – touches on one aspect of interoperability and

the Gilliland-Swetland model refers to technical metadata 'related to how a system functions or metadata behaves'. There is some scope for simplifying Day's model so that 'Preservation management of digital resources' (Purpose 6) becomes part of 'Administration and management of resources' (Purpose 3), a connection that he previously acknowledged (Day, 1999). Likewise 'Providing information on context and authenticity' (Purpose 7) could be grouped with 'Record of intellectual property rights' (Purpose 4) to become 'Record of context, intellectual property rights and authenticity'. Gilliland-Swetland's model could be extended by separating out the description and the information retrieval purposes, for instance.

The five-point model

In this book I propose a new, five-point model to describe the purposes of metadata. It is a closer reflection of current development in metadata and in particular the growing importance of e-commerce. It also separates description from retrieval as a separate, distinct purpose. Some areas have been consolidated such as management of resources and preservation management (which is presented as a sub-set of management) and rights management which is tied in with provenance and authenticity. This model also makes a distinction between the purposes of metadata (i.e. the ways in which it is used) and the intrinsic properties of metadata elements. In doing this it becomes clear that each data element can be used in a variety of ways and fulfils more than one purpose.

The purposes described in the Day and Gilliland-Swetland models can be consolidated into a new, five-point model. The new model encompasses the purposes identified above and includes the additional purposes of interoperability and e-commerce. The five purposes of metadata proposed in this book are described below and provide the basis for later chapters.

1 *Resource description* – This is particularly important in organizations that need to describe their information assets. For example, under the Freedom of Information Act in the UK, public authorities have to produce publication schemes which identify all their publications and intended publications. In the USA, federal agencies have to make information available via the Government Information Locator Service (GILS). These both depend on adequate descriptions of the data. Information asset registers compiled by public authorities and increasingly by the corporate sector also require descriptions of information

repositories and resources.

2 *Information retrieval* – In the academic sector a lot of effort has been put into resource discovery on the internet. Some institutions and agencies have devised subject-based gateways or portals that in effect catalogue relevant high-quality web resources in a particular subject area. This provides users with a route to authoritative sources of information. The cataloguing data usually includes a description of the resource, controlled indexing terms and classification headings. This is a metadata resource and may also 'mine' or 'extract' metadata directly from target websites or electronic resources.

3 *Management of information resources* – The growth of electronic document and records management (EDRM) systems has resulted from the emerging requirements of larger organizations to manage both paper and electronic documentation effectively. EDRM systems need access to 'cataloguing' information about individual documents in order to manage record lifecycles. Examples include authorship, ownership (not necessarily the same thing), provenance of the document (for legal purposes) and date of creation and modification. These and other data elements provide a basis for managing the documentation cost-effectively and consistently. Content management systems (CMSs) are also used to manage data resources including material published on intranets and websites. Chapter 6 describes how metadata is used to manage the retention and disposal of records and the publication of web content in CMS applications.

4 *Documenting ownership and authenticity of digital resources* – Metadata provides a way of declaring the ownership of the intellectual content and layout of a document. It also provides a record of the authenticity of the document by providing an audit trail so that, for instance, an electronic document or a digital image will stand up in court as legally admissible evidence. One of the preconditions for widespread acceptance of electronic documents as original evidence is that electronic systems are becoming the preferred medium for long-term storage of documents.

5 *Interoperability* – Metadata acts as an enabler of information and data transfer between systems, and as such is a key component in interoperability. In order to allow software applications that have been designed independently to pass data between them, a common framework for describing the data being transferred is needed so that each 'knows' how to handle that data in the most appropriate manner. This may be at the level of distinguishing between different languages, or

understanding different data formats.

Interoperability is one of the enablers for e-commerce. When a piece of data is passed from one system to another the accompanying (or embedded) metadata allows the new application to make sense of the data and to use it in the appropriate fashion. This can be seen in the book trade, for instance, where many suppliers using different software packages need to be able to exchange data reliably. The widely adopted ONIX standard allows different participants in the chain from author to reader to exchange data without the need to integrate their systems.

This new five-point model of metadata can be tested by revisiting the reasons for the importance of metadata described earlier. The five functions in the model can be mapped on to the issues as shown in Table 1.2.

Table 1.2 The five-point model and why metadata is important

Why is metadata important?	Purposes of metadata
Retrieval performance	Purpose 1: Resource description Purpose 2: Information retrieval
Management of electronic resources	Purpose 3: Resource management
Authenticity	Purpose 4: Ownership and authenticity
Interoperability	Purpose 5: Interoperability
The future	All five purposes

Metadata can be used within one application for several different purposes. The model developed in this chapter helps in the analysis of metadata applications and the understanding of its characteristics in different situations. The five-part model provides the structure for Chapters 4–8 which deal with each purpose of metadata in turn. The next two chapters deal with concepts behind metadata and metadata standards.

summary

Metadata has been around since the first library catalogues were established over 2000 years ago. The term metadata probably first appeared in the 1960s and was adopted by the GIS specialists, database developers, statisticians and latterly, in the 1990s, by the web community.

The term 'metadata' first appeared in the 1960s but became established in the database community in the 1970s. A useful definition sees metadata 'as the means by which the structure and behaviour of data is recorded, controlled, and published across an organization' (Tozer, 1999). This is a useful definition because it does not specify electronic resources and it deals with the key aspects of metadata's purposes.

Some metadata is intended for human readers and some is intended for use by

computer applications to help them process or exchange data. Examples of human-readable metadata include web pages with metatags describing the format and content of the resource and a MARC record from a library catalogue. Metadata plays a number of roles including:

- **improving retrieval performance**
- **providing a way of managing electronic resources**
- **helping to determine the authenticity of data**
- **enabling interoperability.**

Metadata is important because it points to the future of managing information with wider adoption of metadata standards and extensive use in e-commerce.

Different models of metadata have been proposed and a new five-point model is constructed here which describes metadata in terms of the following purposes:

1 Describing information resources
2 Enhancing information retrieval
3 Managing of information resources
4 Documenting ownership and authenticity
5 Exchanging data between systems.

This model reflects current usage of metadata and provides a basis for considering the use of metadata in greater depth. An analysis of the five purposes helps to answer the question 'Why is metadata important?' and gives a framework for the discussion of metadata applications in the later chapters of this book.

References and further sources of information

Bowman, J. H. (2003) *Essential Cataloguing*, London, Facet Publishing.

Caplan, P. (2003) *Metadata Fundamentals for All Librarians*, Chicago, IL, American Library Association.

Day, M. (1999) *Issues and Approaches to Preservation Metadata*. Paper at Joint RLG and NPO Preservation Conference on Guidelines for Digital Imaging, www.rlg.org/preserv/joint/day.html [accessed 25/2/2004I).

Day, M. (2001) Metadata in a Nutshell, *Information Europe*, **6** (2), 11.

Ellens, J. H. (1997) *The Ancient Library of Alexandria: the West's most important repository of learning*, Biblical Review Archives, www.biblereview.org/bswb_AO/brf97library.html [viewed as a cache on Google, 21/1/2004].

Feather, J. and Sturges, P. (eds) (1997) *International Encyclopaedia of Information and Library Science*, London, Routledge.

Gilliland-Swetland, A. (1998) Defining Metadata. In Baca, M. (ed.), *Introduction to Metadata: pathways to digital information*, Los Angeles, CA, Getty Information Institute.

Gorman, M. (2003) *Authority Control in the Context of Bibliographic Control in the Electronic Environment*. Paper at International Conference on Authority Control held in Florence Italy 10–12 February 2003, http://mg.csufresno.edu/papers/Authority_Control.pdf [accessed 25/2/2004].

Hudgins, J., Agnew, G. and Brown, E. (1999) *Getting Mileage out of Metadata: applications for the library*, LITA Guides 5, Chicago, IL, American Library Association.

Hunter, E. J. and Bakewell, K. G. B. (1991) *Cataloguing*, 3rd edn (revised and expanded by E. J. Hunter), London, Library Association Publishing.

ISO 2709:1996. *Information and Documentation – Format for Information Exchange*, Geneva, International Organization for Standardization.

Joint Steering Committee for Revision of AACR (2002a) *A Brief History of AACR*, www.nlc-bnc.ca/jsc/history.html [accessed 24/2/2004].

Joint Steering Committee for the Revision of AACR (2002), American Library Association, Canadian Library Association and CILIP (2000b) *Anglo-American Cataloguing Rules*, 2nd edn, 2002 Revision, Chicago, IL, American Library Association, Ottawa, Canadian Library Association and London, CILIP. Amendments published annually.

Tozer, G. (1999) *Metadata Management for Information Control and Business Success*, Boston, MA, Artech House.

UK Department of Health (n.d.) NSF Diabetes: Delivery Strategy website, www.doh.gov.uk/nsf/diabetes/delivery/foreward.htm [accessed 27/2/2003].

UK Office of the e-Envoy (2001) *E-government Metadata Framework*, London, Office of the e-Envoy.

UK Office of the e-Envoy (2003) *E-government Metadata Standard*, London, Office of the e-Envoy, www.e-envoy.gov.uk/assetRoot/04/00/24/55/04002455.pdf [accessed 25/2/2004].

Vellucci, S. L. (1998) Metadata. In Williams, M. E. (ed.), *Annual Review of Information Science and Technology*, Vol. 33, Medford, NJ, Information Today Inc., 187–222.

Woodley, M. (1999) *Re: History of the term 'metadata'*, Metamarda-l Listserv Archive, 30 March 1999, http://orc.dev.oclc.org:5103/metamarda-l/msg00097.html [accessed 24/5/2004].