Metadata

7.1 Databases use Metadata: Metadata is "data about data". The term is ambiguous, as it is used for two fundamentally different concepts (types). Structural metadata is about the design and specification of data structures and is more properly called "data about the containers of data"; descriptive metadata, on the other hand, is about individual instances of application data, the data content.

Metadata is traditionally in the card catalogs of libraries. As information has become increasingly digital, metadata are also used to describe digital data using metadata standards specific to a particular discipline. By describing the contents and context of data files, the usefulness of the original data/files is greatly increased. For example, a webpage may include metadata specifying what language it is written in, what tools were used to create it, and where to go for more on the subject, allowing browsers to automatically improve the experience of users. Wikipedia encourages the use of metadata by asking editors to add category names to articles, and to include information with citations such as title, source and access date.

The main purpose of metadata is to facilitate in the discovery of relevant information, more often classified as resource discovery. Metadata also helps organize electronic resources, provide digital identification, and helps support archiving and preservation of the resource. Metadata assists in resource discovery by "allowing resources to be found by relevant criteria, identifying resources, bringing similar resources together, distinguishing dissimilar resources, and giving location information."

Definition

Metadata (metacontent) is defined as the data providing information about one or more aspects of the data, such as:

- Means of creation of the data
- Purpose of the data
- Time and date of creation
- Creator or author of the data
- Location on a computer network where the data were created
- Standards used

For example, a digital image may include metadata that describe how large the picture is, the color depth, the image resolution, when the image was created, and other data. A text document's metadata may contain information about how long the document is, who the author is, when the document was written, and a short summary of the document.

Metadata is data. As such, metadata can be stored and managed in a database, often called a metadata registry or metadata repository. However, without context and a point of reference, it might be impossible to identify metadata just by looking at them. For example: by itself, a database containing several numbers, all 13 digits long could be the results of calculations or a list of numbers to plug into an equation - without any other context, the numbers themselves can be perceived as the data. But if given the context that this database is a log of a book collection, those 13-digit numbers may now be identified as ISBNs - information that refers to the book, but is not itself the information within the book.

The term "metadata" was coined in 1968 by Philip Bagley, in his book "Extension of programming language concepts" where it is clear that he uses the term in the ISO 11179 "traditional" sense, which is "structural metadata" i.e. "data about the containers of data"; rather than the alternate sense "content about individual instances of data content" or metacontent, the type of data usually found in library catalogues. Since then the fields of information management, information science, information technology, librarianship and GIS have widely adopted the term. In these fields the word *metadata* is defined as "data about data". While this is the generally accepted definition, various disciplines have adopted their own more specific explanation and uses of the term.

Libraries

Metadata has been used in various forms as a means of cataloging archived information. The Dewey Decimal System employed by libraries for the classification of library materials is an early example of metadata usage. Library catalogues used 3x5 inch cards to display a book's title, author, subject matter, and a brief plot synopsis along with an abbreviated alpha-numeric identification system which indicated the physical location of the book within the library's shelves. Such data help classify, aggregate, identify, and locate a particular book. Another form of older metadata collection is the use by US Census Bureau of what is known as the "Long Form." The Long Form asks questions that are used to create demographic data to find patterns of distribution.

Photographs

Metadata may be written into a digital photo file that will identify who owns it, copyright and contact information, what camera created the file, along with exposure information and descriptive information such as keywords about the photo, making the file searchable on the computer and/or the Internet. Some metadata is written by the camera and some is input by the photographer and/or software after downloading to a computer. Most digital cameras write metadata, and some enable you to edit it; this functionality has been available on most Nikon DSLRs since the Nikon D3 and on most new Canon cameras since the Canon EOS 7D.

Photographic Metadata Standards are governed by organizations that develop the following standards. They include, but are not limited to:

- IPTC Information Interchange Model IIM (International Press Telecommunications Council),
- IPTC Core Schema for XMP
- XMP – Extensible Metadata Platform (an ISO standard)
- Exif – Exchangeable image file format, Maintained by CIPA (Camera & Imaging Products Association) and published by JEITA (Japan Electronics and Information Technology Industries Association)
- Dublin Core (Dublin Core Metadata Initiative – DCMI)
- PLUS (Picture Licensing Universal System).

Video

Metadata is particularly useful in video, where information about its contents (such as transcripts of conversations and text descriptions of its scenes) is not directly understandable by a computer, but where efficient search is desirable.

Web pages

Web pages often include metadata in the form of meta tags. Description and keywords meta tags are commonly used to describe the Web page's content. Most search engines use these data when adding pages to their search index.

Creation of metadata

Metadata can be created either by automated information processing or by manual work. Elementary metadata captured by computers can include information about

when an object was created, who created it, when it was last updated, file size and file extension.

For the purposes of this article, an "object" refers to any of the following:

- A physical item such as a book, CD, DVD, map, chair, table, flower pot, etc.
- An electronic file such as a digital image, digital photo, document, program file, database table, etc.

7.2 Metadata types

The metadata application is manyfold covering a large variety of fields of application there are nothing but specialised and well accepted models to specify types of metadata. Bretheron & Singley (1994) distinguish between two distinct classes: structural/control metadata and guide metadata. *Structural metadata* is used to describe the structure of database objects such as tables, columns, keys and indexes. *Guide metadata* is used to help humans find specific items and is usually expressed as a set of keywords in a natural language. According to Ralph Kimball metadata can be divided into 2 similar categories: technical metadata and business metadata. *Technical metadata* correspond to internal metadata, and *business metadata* correspond to external metadata. Kimball adds a third category named *process metadata*. On the other hand, NISO distinguishes among three types of metadata: descriptive, structural and administrative.

Descriptive metadata is typically used for discovery and identification, as information used to search and locate an object such as title, author, subjects, keywords, publisher. *Structural metadata* give a description of how the components of an object are organized. An example of structural metadata would be how pages are ordered to form chapters of a book. Finally, *administrative metadata* give information to help manage the source. They refer to the technical information including file type or when and how the file was created. Two sub-types of administrative metadata are rights management metadata and preservation metadata. *Rights management metadata* explain intellectual property rights, while *preservation metadata* contain information that is needed to preserve and save a resource.

Metadata structures

Metadata (metacontent), or more correctly, the vocabularies used to assemble metadata (metacontent) statements, are typically structured according to a standardized concept using a well-defined metadata scheme, including: metadata

standards and metadata models. Tools such as controlled vocabularies, taxonomies, thesauri, data dictionaries and metadata registries can be used to apply further standardization to the metadata. Structural metadata commonality is also of paramount importance in data model development and in database design.

Metadata syntax

Metadata (metacontent) syntax refers to the rules created to structure the fields or elements of metadata (metacontent). A single metadata scheme may be expressed in a number of different markup or programming languages, each of which requires a different syntax. For example, Dublin Core may be expressed in plain text, HTML, XML and RDF.

A common example of (guide) metacontent is the bibliographic classification, the subject, the Dewey Decimal class number. There is always an implied statement in any "classification" of some object. This is a subject-predicate-object triple, or more importantly, a class-attribute-value triple. The first two elements of the triple (class, attribute) are pieces of some structural metadata having a defined semantic. The third element is a value, preferably from some controlled vocabulary, some reference (master) data. The combination of the metadata and master data elements results in a statement which is a metacontent statement i.e. "metacontent = metadata + master data". All these elements can be thought of as "vocabulary". Both metadata and master data are vocabularies which can be assembled into metacontent statements. There are many sources of these vocabularies, both meta and master data: UML, EDIFACT, XSD, Dewey/UDC/LoC, SKOS, ISO-25964, Pantone, Linnaean Binomial Nomenclature etc. Using controlled vocabularies for the components of metacontent statements, whether for indexing or finding, is endorsed by ISO-25964: "If both the indexer and the searcher are guided to choose the same term for the same concept, then relevant documents will be retrieved." This is particularly relevant when considering the behemoth of the internet, Google. It simply indexes pages then matches text strings using its complex algorithm, there is no intelligence or "inferencing" occurring. Just the illusion thereof.

7.3 Hierarchical, linear and planar schemata

Metadata schema can be hierarchical in nature where relationships exist between metadata elements and elements are nested so that parent-child relationships exist between the elements. An example of a hierarchical metadata schema is the IEEE LOM schema where metadata elements may belong to a parent metadata element.

Metadata schema can also be one-dimensional, or linear, where each element is completely discrete from other elements and classified according to one dimension only. An example of a linear metadata schema is Dublin Core schema which is one dimensional. Metadata schema are often two dimensional, or planar, where each element is completely discrete from other elements but classified according to two orthogonal dimensions.

Metadata hypermapping

In all cases where the metadata schemata exceed the planar depiction, some type of hypermapping is required to enable display and view of metadata according to chosen aspect and to serve special views. Hypermapping frequently applies to layering of geographical and geological information overlays.

Granularity

The degree to which the data or metadata is structured is referred to as their granularity. Metadata with a high granularity allow for deeper structured information and enable greater levels of technical manipulation. A lower level of granularity means that metadata can be created for considerably lower costs but will not provide as detailed information. The major impact of granularity is not only on creation and capture, but moreover on maintenance. As soon as the metadata structures get outdated, the access to the referred data will get outdated. Hence granularity shall take into account the effort to create as well as the effort to maintain.

Metadata standards

International standards apply to metadata. Much work is being accomplished in the national and international standards communities, especially ANSI (American National Standards Institute) and ISO (International Organization for Standardization) to reach consensus on standardizing metadata and registries.

The core standard is ISO/IEC 11179-1:2004 and subsequent standards (see ISO/IEC 11179). All yet published registrations according to this standard cover just the definition of metadata and do not serve the structuring of metadata storage or retrieval neither any administrative standardisation. It is important to note that this standard refers to metadata as the data about containers of the data and not to metadata (metacontent) as the data about the data contents. It should also be noted that this standard describes itself originally as a "data element" registry, describing disembodied data elements, and explicitly disavows the capability of containing

complex structures. Thus the original term "data element" is more applicable than the later applied buzzword "metadata".

The Dublin Core metadata terms are a set of vocabulary terms which can be used to describe resources for the purposes of discovery. The original set of 15 classic metadata terms, known as the Dublin Core Metadata Element Set are endorsed in the following standards documents:

- IETF RFC 5013
- ISO Standard 15836-2009
- NISO Standard Z39.85.

Although not a standard, Microformat (also mentioned in the section metadata on the internet below) is a web-based approach to semantic markup which seeks to re-use existing HTML/XHTML tags to convey metadata. Microformat follows XHTML and HTML standards but is not a standard in itself. One advocate of microformats, Tantek Çelik, characterized a problem with alternative approaches:

> “ Here's a new language we want you to learn, and now you need to output these additional files on your server. It's a hassle. (Microformats) lower the barrier to entry. ”