

Metadata

Who needs to know this?

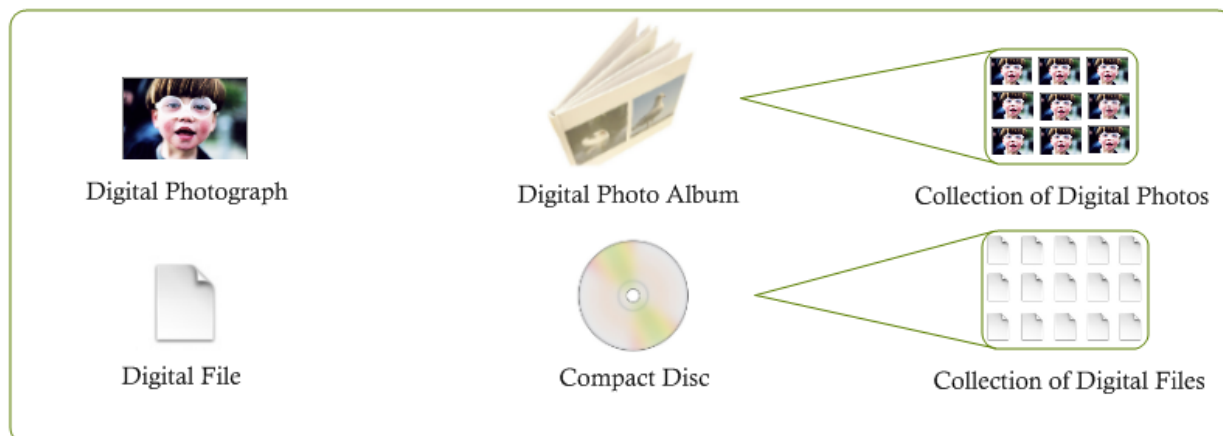
This is a general introduction which is likely to be of interest to researchers, their support staff, data centre and repository staff and research administrators.

Definition

The term *metadata* refers to information used to describe items and groups of items. It is data about data. It can be used to describe physical items as well as digital items (files, documents, images, datasets, etc.). A library catalogue, for example, is made up of metadata describing the books, journals and other items held by the library. The File Properties for a word processing document is a rudimentary (and imperfect) metadata record.

Item level metadata is used to describe a single object such as a photograph: who took the photograph, who is in it, the date it was taken, the place it was taken, the type of camera used to take the photograph, and so on.

Collection level metadata is used to describe an aggregation of objects such as the photo album (or CD-ROM or file folder) that contains a group of photographs: the size of the collection, who took the photographs (there may be more than one person), the time period over which the photographs were taken, and so on. Some of these



attributes, such as 'Title' may be the same as those used to describe an individual photograph.

Metadata adds value to documents or images. For scientific data, metadata is even more important because it provides the context needed to make sense of what would otherwise be a collection of random numbers.

Types of metadata

The metadata elements used to describe either an item or a collection can serve different purposes. Some examples include:

- *Descriptive* metadata, such as the name of the photographer, the subject of the photograph, the date and time that the photograph was taken;
- *Technical* metadata, such as the type of camera used, the file format in which the photograph is stored, the exposure time and dimensions of the photograph, and so on;



- *Access or rights* metadata, defining who is allowed to view to this photograph and under what conditions; and
- *Preservation* metadata, which allows a digital preservation expert to keep track of actions taken to preserve or sustain the photograph for later access and use.

Metadata creation

Metadata can be created *by hand*, or it can be created *automatically*. The camera can tell you the time and date, the type of camera, exposure times, file format, and so on, and can attach this metadata to the image file automatically. The camera cannot tell you who the photographer is, or what the subject of the photograph is. This information must be provided by a human. There is a significant cost associated with assigning metadata by hand and little cost associated with collecting it automatically.

Why we need metadata standards

If you have ever tried to find a photograph on your own computer, you know how useful metadata can be.

If you are creating a database of your own photographs, then as long as you are internally consistent you can use any terms you've chosen to search, sort or rank the items. For example, if you want to search for all photographs of the Taj Mahal, then each image needs to have the subject descriptor 'Taj Mahal'. If you want to search for all photographs of Bendigo in 1858, then the photographs must be described with both the geographic element 'Bendigo', and the date '1858'. The same goes for any searchable database (or repository) of research publications or collections of data.

Standardised vocabularies and *ontologies* describe ways in which terms are standardised and grouped to provide consistency when ascribing metadata. This will help to make sure that preferred terms (such as 'thesis' as opposed to 'dissertation', or 'directory' rather than 'folder') are used.

Once you start combining those indexes, lists, or databases, you need to have some agreed standards in place to allow for the interchange of data. There are many different standards for metadata, some of which are discipline-specific. Computers can then retrieve metadata from different sources (a process known as harvesting) and combine it automatically to create bigger collections of metadata that make for better discovery services. Examples include services such as Picture Australia or the ANDS 'Register My Data' service.

Further information

ANDS Guides and other Resources: <http://ands.org.au/guides>

For examples of metadata in action, see:

The Western Australia Marine Data site's Department of Transport – Navigational Aids

<http://waodn.ivec.org/geonetwork/srv/en/metadata.show?uuid=9e54803a-9cbc-4906-9ae2-e5a319c6c002>

The Indexgeo metadata record for Wastewater artificial wetlands listed as important wetlands in Australia.

<http://www.indexgeo.com.au/ec/pub/crossley/dataset/ANZCW1003100029.html>

For more about metadata, see:

Understanding Metadata, NISO Press, <http://www.niso.org/publications/press/UnderstandingMetadata.pdf>

'Documentation and Metadata', from the MIT Libraries' *Data Management and Publishing* guide,

<http://libraries.mit.edu/guides/subjects/data-management/metadata.html>

An Introduction to Metadata, from the University of Queensland Library web site,

<http://www.library.uq.edu.au/iad/ctmeta4.html>

