

Metadata

Summary

Metadata, sometimes referred to as “data about data,” is the additional descriptive information about digital content that makes records useful, meaningful, and findable. Maintaining and updating metadata for an agency’s electronic resources provides benefits such as easier and more efficient discovery of relevant information. Beyond resource discovery, metadata is used to describe the character or content of a digital object such as page number order, technical aspects like image resolution, and is also used to ensure continued access to a digital object.

Not only does metadata facilitate resource discovery within an agency but it can help organize and manage electronic resources when the inevitable system backup, restoration, or migration takes place at an agency.

Metadata Categories and Functions

There are several types of metadata to keep in mind. They will all prove useful in maintaining the content, context, and structure of your records and in keeping them useful and available for the long-term.

- *Descriptive Metadata* – Describes a resource for the purpose of indexing, discovery, and identification. Common descriptive metadata fields include creator, title, and subject.
- *Administrative Metadata* – Helps manage a resource by describing management information such as ownership and rights management.
- *Structural Metadata* – Used to display and navigate digital resources and describes relationships between multiple digital files, such as page order in a digitized book.
- *Technical Metadata* – Describes the features of the digital file, such as resolution, pixel dimension, and creating hardware. The information is critical for migration and long-term sustainability of the digital resource.
- *Preservation Metadata* – Contains the information needed to preserve a digital object and protect the object from harm, deterioration, or destruction. Preservation metadata may encompass the aforementioned forms of metadata.

One of the most pressing reasons an agency will want to create descriptive metadata is to ensure the discovery of information. If an agency receives a request by the public for information concerning the business of the state, high quality descriptive metadata will make the search and retrieval of an electronic resource relating to the request much more manageable. An electronic resource with high quality metadata allows the user to identify resources, distinguish relationships with other objects, bring similar resources together, and determine location information.

There are many compelling reasons for recording metadata but, in terms of support for government agencies, metadata is useful for:

- Legal and statutory reasons¹
- Advancements in technology (upgrading servers)
- Providing service to citizens and other agencies (identifying , locating, and sharing information)
- Optimal workflow (easily finding documents and understanding their context)
- Operational, administrative and preservation reasons (Decision making documentation)

Where Does Metadata Live?

To better understand metadata it is important to know where the information or metadata is stored. Metadata can be embedded within a file. For example, when an image is scanned, the associated metadata such as file type, date scanned, file name, and image resolution, lives with the file or is embedded in the file.

A file's metadata does not necessarily have to be embedded within the file. An external catalog of metadata for an agency's files can provide an efficient avenue for managing and discovering files at a later date. Utilizing a data spreadsheet such as Microsoft Excel or Access, allows a user to tailor metadata entries that best suit the needs of the agency. When using an external catalog, two spreadsheets should be maintained when practicing the catalog method for recording metadata: a master copy with permissions granted to select users and a copy that allows access to all individuals who may need to use the data to perform work duties. A metadata catalog is an efficient way to manage electronic resources and boasts advantages such as:

- Offline searching
- Collection-wide searching
- Provides a record of an agency's electronic records.

Metadata Schemas and Element Sets

Many different metadata standards, or schemas, exist for a variety of users and disciplines. One of the most common metadata schemas, Dublin Core, is versatile and easily applied to an assortment of objects from many different professions and disciplines². The Dublin Core schema consists of fifteen elements: Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, and Rights. Dublin Core was meant to be simple and concise and able to accommodate the increasing presence of electronic resources.

When using a metadata schema it is also important to exercise a controlled vocabulary. A controlled vocabulary consists of an approved set of terms for the content of the elements. For example, there is more than one way to write a date so it is important to set forth an approved method (i.e. 02_02_2005

¹ California Public Records Act, Government Code Sections 6250-6276.48 Government Code 6252 (e); California State Records Management Act, Government Code Sections 12270-12279 Government Code 12275 (a)

² The Dublin Core Metadata Element Set is a set of guidelines for cross-domain resource description. ISO 15836:2009

or February 2, 2015). The format for proper nouns should also be agreed upon (e.g. Smith, Joan or Joan Smith). It is also important to be mindful of abbreviations (i.e. Cal, CA, Calif.)

Ultimately an agency must decide on what metadata schema will work best for them, but employing a high quality standard will ensure consistency across an agency's files and allow objects to be found and compared more easily.

File Naming Conventions

A file name is the key identifier of a digital object and provides metadata for the record. Consistent and descriptive file names will provide a more organized and easily understood collection of records. How a file is named will have a large impact on finding the files at a later date and understanding their contents. The following information might be considered when creating a file naming policy; ultimately file names should reflect the purpose and need of an agency:

- Project name
- Date or date range of creation
- Version number
- Name of intended audience
- Description of content
- Department
- Publication date
- Release date
- Record series
- Name of creator

When creating a file naming policy the following should also be considered:

- Create unique file names.
- File names should be easy to understand and not overly complex.
- Do not use spaces but rather () or (-) to represent a space.
- Avoid using special characters such as \$ # @ & ^ % * ! and use only alpha-numeric characters.
- Limit file names to 25 characters or less.
- Use the three character file extension with a period (e.g. .tif not .tiff) at the end of the name.
- Don't rely on the system to differentiate between upper and lower case and be consistent in what is used.
- If digits are included in the file name, include the appropriate amount of leading zeros and be generous so your project can be scalable.
- It's helpful to include metadata in the file name but this can be cumbersome if you have huge numbers of files. That said, consider using shortened versions of 1) a standardized date, 2) version number (only if this can/will vary), 3) creator's name, and 4) description/type of document/subject in the file name and in a logical order.
- Document whatever naming convention is settled on. This really is the key. Include the naming convention document any time records are transferred elsewhere.