Metadata

Digital Preservation Topical Note 5

dpc DigitalPreservationCoalition

What is metadata and why is it needed?

The broad definition of metadata is 'data about data'. For digital preservation, metadata is any secondary information about a digital record that makes it easier to find and use that record. Metadata exists in several forms: some forms are machine-readable, others are created for humans to read, and some are a combination of both. Using a machine (e.g. a computer programme), to search digital records is especially important at scale. Machine-readable metadata can provide access to high volumes of digital records that would be impossible for a person or even team of people to search manually.

When should metadata be created?

As a rule, metadata is much cheaper and simpler to produce at the time of creation or very soon after. The creator of a record will have better, fresher knowledge of the record at the time of creation. Therefore, the more time that passes, the more difficult, time-consuming, and expensive it becomes to reconstruct the required information to create useful metadata.

What types of metadata are there?

Different categories of metadata fulfil different functions. For the purposes of digital preservation, a few different types of metadata are important:

- **Descriptive metadata**: summarises or gives details about a digital record and its content to make it easier to find in a search

- **Structural metadata**: provides information about the internal structure of a digital file, includes information like page, section, or index to help devices with navigation and display

- **Administrative metadata**: refers to the information about the management of a digital record, such as who created it, who is allowed to access it, or what rights issues are associated with it

- **Preservation metadata**: helps archivists and other digital preservation specialists to open and use digital records as far into the future as necessary; might include information about what software or hardware is needed to open and use a digital file

- **Technical metadata**: rather than being created for the purposes of archiving, technical metadata is often captured automatically through the software or hardware used to create a digital record. For example, photos created by a digital camera automatically capture information about the image and embed this information in the file itself.

Key Term: **Interoperability**

Formatted so that it can be read by many different systems; interoperable metadata can operate across different hardware and software environments and across different departments and organisations.

Why is metadata important for digital preservation?

For long-term preservation, the person preserving the digital record is usually someone different from the original creator of the record. Therefore, it is important that information about the record – information the creator would know – is captured in the metadata. Archivists require metadata for several reasons, including:

Decision-making

Information associated with an object, such as the history of changes made to it, software required to open it, or how long it needs to be retained can help archivists make decisions about how or why to preserve it. Metadata also often includes details about rights and ownership so that users of the record know what can and cannot be copied, shared, or modified. If records are openly available, metadata will enhance findability so that more people, such as researchers, can access and use them.

Context for meaning

Metadata, especially descriptive and administrative metadata, can provide the types of context information needed for future users – such as administrators, decision-makers, or even researchers – to better understand the meaning of a record's *content*. This data might include information about why a record was created, who created it, and why it has been preserved.

Usability

Metadata ensures that future users will be able to render and interpret a record, for instance with the correct software. Without the right type of information, a digital record might lose its meaning or structure. For example, a record creator might produce a Word document that links to an Excel spreadsheet. If the spreadsheet becomes lost or deleted, the content in the document might not make any sense. Metadata about this relationship will help ensure that users can find both files. The separation or loss of a dependent file is called *dissociation*. If dissociation occurs and no metadata exists to help re-unite the two files, the Word document and Excel spreadsheet may become useless and essentially lost.

Key Term: **Dissociation**

If two or more digital records depend on one another for meaning or structure and one becomes separated, the files become dissociated. If metadata does not exist to help users find the missing file, the dependent records are essentially lost.

Key Term: **Schema**

A logical plan for structuring metadata so that it can be processed by many different systems

Why are standards important?

Metadata standards help to make metadata as useful as possible by providing guidelines for uniform formatting. *Schemas* are guidelines for uniform metadata formats. Both standards and schemas ensure that metadata for digital records can be interpreted by many different systems, or in other words, so that they will be *interoperable*. Dublin Core, for example, is a standard for discovery metadata. For digital preservation, the PREMIS Data Dictionary provides guidelines for implementing metadata helpful for long-term management.

For more information on Digital Preservation visit the DPC Website: https://www.dpconline.org