Transcript
Metadata basics for the digital humanities
October 2022

Welcome to this video entitled 3Ws and an H: Metadate basics for the humanities. My name is Martha Attridge Bufton, and I am the Interdisciplinary Studies Librarian at the Carleton University Library.

Being able to find things–shoes in our closet, Tupperware in our kitchen cupboards, Christmas ornaments we've put away–is made infinitely easier if we are organized. This can mean putting all our shoes in neat rows or labelling boxes so that we can quickly scan our basement shelves and locate the container of ornaments that we want.

The tidier we are, the easier it is to find things.

This same principle–being organized–also applies to ensuring that digital data and information are properly stored and retrievable.

Metadata is one "knowledge organization system" that is a useful tool for researchers who undertake projects that involve the curation, description, sharing, and preservation of digital objects such as images, audio files, videos, documents, and web pages.

When such projects include the systematic use of metadata, the outcomes can be easily found, retrieved, shared across time and space, and saved for years to come.

The purpose of this video is to introduce humanities researchers to the "3 Ws and 1 H" of metadata: what is metadata, why do we create it, who creates it, and how is it created?

Metadata is a powerful tool for organizing projects and can contribute significantly to knowledge mobilization.

So, the first W: What is metadata?

Definitions of metadata are surprisingly consistent. Here is one definition that captures the key elements.

"Metadata is structured information that describes the attributes of information resources for the purposes of identification, discovery, selection, use, access, and management." (Taylor and Joudrey, 2012, p. 4.)

Martha Attridge Bufton
Licensed under a Creative Commons Attribution-NonCommercial 4.0 International license.

In other words, metadata is structured data about data, where any object can be considered a form of data.

There are different types of metadata. For example, objects can be organized and labelled with:

administrative metadata: this is data that identifies how a resource is managed, such as the name of the organization responsible for creating metadata.

descriptive metadata: this is data that ensures that a resource is discoverable, such as a book title and author's name.

and technical metadata: this is the data that identifies details of a digital object and its creation such as file size and type.

Most of us are familiar with descriptive metadata. For example, descriptive metadata includes the data that helps us find a book in the library or peer-reviewed sources in a database.

And there are different formats, or levels of granularity, to metadata.

Simple metadata is extracted from a resource.

Structured metadata is formal metadata element sets created for the general user.

And rich metadata includes sets of comprehensive, meticulously detailed descriptions.

So, metadata–the data about data–is a tool for organizing objects, including digital objects.

Now we move on to the second W is why: why create metadata?

Metadata is useful when we want to ensure that, at the very least, data and information in a project can be discovered, retrieved, shared, compared, curated, and preserved.

For example, let's say we want to borrow a copy of *The vertical mosaic* by John Porter from the Carleton University Library. Because library cataloguers are experts at properly assigning metadata to records, we can use the Library search engine Omni, to search for the book by title, and then easily retrieve the information needed to find the book in the library.

Martha Attridge Bufton
Licensed under a Creative Commons Attribution-NonCommercial 4.0 International license.

We'll start on the library home page and use the Advanced search option. Then we will change the filter at the top to limit our results to those materials found in our collection.

We will use the drop-down menu next to Any field and choose Title. Then we will put the title of the book in the search field. And then click on Search in the bottom right-hand corner of the field.

Because the metadata has been assigned properly, the book we are looking for is the second result, and the metadata available is the title of the book, the author and the date of publication, along with the floor on which the book is shelved and the call number which is needed to find the book on the shelf.

Likewise, in this Omeka project on documentary filmmaking, the systematic use of metadata allows for the curation of digital objects into collections that can be browsed.

Here we are seeing the metadata for one digital object, an oral history interview. Omeka is an open-source platform for creating digital repositories.

It is important to note that metadata creation is not culturally neutral–it reflects the ways of knowing and being of culture in which it is created.

As a result, cultural bias is embedded in the logic of any given metadata system and as Allison-Cassin and Seeman point out, "although rarely an object of notice or scrutiny by its users, metadata governs the circulation of information and has the power to name, broadcast, normalize, oppress, and exclude"

As a result, it is important to create metadata ethically.

According to Roxanne Lafleur, the Digital Humanities Support Specialist at the University of Ottawa, an ethical process includes considering what is included and what is excluded (and why); the vocabulary chosen to describe a resource; how the resource is interpreted; the cultural context of a resource and if it is acknowledged and respected without judgement or bias.

The third W is who: who creates metadata?

Who creates metadata depends on the given project or activity. A photo on a family Facebook page can be tagged (a form of metadata) by anyone in the group or community.

Metadata entries for an academic digital repository could be created by faculty and/or students. And digital objects in a provincial museum's online collection could be created by trained professionals.

Martha Attridge Bufton
Licensed under a Creative Commons Attribution-NonCommercial 4.0 International license.

So the compensation of the metadata team is project-driven and metadata depends on the scope, the purpose, and the timeframe of the project as well as the required knowledge and skill. Those who do take on this task will need to understand the approach being taken (for example, is a scheme being used) and the purpose for the project overall.

And finally we arrive at H: How do we create metadata and do so ethically?

We know from our everyday lives that there are many ways to organize objects and information. We can organize alphabetically, by colour, by title or name or size. These might be considered simple forms of metadata that do not necessarily require expertise or extra training to use.

However, there are contexts in which it is necessary to create formal or even rich metadata sets.

Take the development and maintenance of institutional library collections, such as those in a public or university library. These are collections (projects if you will) that require expertise and the systematic creation of formal and rich metadata sets to ensure that the collections function well for those who use them.

So, how metadata is created depends on project needs, which drive the metadata format, the use of schema, as well as the use of standards and control.

According to Taylor and Joudrey, metadata must be standardized and controlled if it is to be used to its full potential. This is because standards and control can create clarity rather than confusion for the user. In other words, shared semantics helps to avoid confusion, ambiguity, and missed opportunities to find needed information.

One way to standardize metadata creation is to use a scheme. Metadata schema can be defined as “sets of metadata elements designed for a specific purpose, such as to describe objects.”

Examples of metadata schema include Dublin Core, a set of fifteen elements used widely to describe digital objects. For example, Dublin Core is the metadata scheme used on the Omeka platform.

A metadata scheme can include other forms of standards, such as controlled vocabularies or curated lists of words and phrases to use. The Library of Congress subject headings are an example of a controlled vocabulary.

Martha Attridge Bufton
Licensed under a Creative Commons Attribution-NonCommercial 4.0 International license.

Academic library catalogues or search engines use standardized approaches to creating metadata to ensure that users can easily find and retrieve materials and so that item records can be shared across institutions.

For example, Omni is a “discovery layer” that has been developed by 14 Ontario universities and one of the goals of the project has been to enable users from these institutions to search across collections–this is why the default search option is to search Carleton + Omni libraries.

This cross-institutional searching is made possible by standardized metadata.

Metadata creation in this context includes doing authority work (that is, standardizing the names of individuals attached to a digital object, such as an author and illustrator), doing a subject analysis and assigning Library of Congress subject headings, and then classifying items with a unique identifier (for example, a call number).

However, standards and control over metadata creation must be used ethically. Again, as Allison-Cassin and Seeman point out, metadata has “the power to name, broadcast, normalize, oppress, and exclude” and this power is embedded in standards and the processes used to create metadata.

For example, the Library of Congress subject headings include highly problematic vocabulary such as the use of the term illegal aliens rather than a more neutral term such as undocumented workers, and the term Indians of North America rather than First Nations, Métis, and Inuit.

As such, in order to develop ethical and appropriate metadata, those responsible for its creation must consider not only project needs but also the broader societal need for inclusive descriptors. In short, they must challenge individual and collective biases.

Summary

So, to summarize, metadata is data about data. It is created to organize data so that it can be found and used for the purposes for which it is intended.

Metadata is created by individuals and groups who use ethical approaches to assign metadata to objects and Well-designed metadata is standardized and controlled to avoid confusion, ambiguity, and missed opportunities for users.

This is the end of this video.

Martha Attridge Bufton

Licensed under a Creative Commons Attribution-NonCommercial 4.0 International license.