Metadata

Preservation and Curation of ETD Research Data and Complex Digital Objects

INSTITUTE *of* **Museum** and **Library** SERVICES

EDUCOPIA INSTITUTE

Metadata

After years of work, your dissertation is finished and submitted, and you have now entered your chosen profession as a researcher. One of the datasets you used in your dissertation has become an important element in a new research project that you are directing. You share it with your project team, and they begin building upon your work. A month later, as you are reviewing your team's progress, you realize that your colleagues have misinterpreted several variables in the dataset, and as a result, they have now conducted and coded several weeks' worth of research incorrectly. When you begin researching how this error happened, you find that your data dictionary—the last piece you finalized and included in your dissertation submission package—contained several errors that have now impacted your current research. ■

Rationale and Motivations – Why

Metadata describes and documents research, data, and publications. More simply, it is information that is created and stored alongside content (such as a thesis or dissertation) in order to help users find and understand that content. Metadata documentation supports other systems that manage the content in many ways, from determining how it can be shared (e.g., publicly or not), to the ways people are able to search for and find it via Google, library catalogs, or other search methods and platforms. Some forms of metadata—for example, a data dictionary or codebook—are essential components of research, without which a dataset may be impossible to understand at all. The quality of the metadata you provide is key to ensuring the future accessibility and usability of your work.

There are few hard and fast rules about what constitutes "metadata." Choices about what information is important to know are driven by the needs people have and the systems they create to manage and find information. Understanding the value and uses of metadata, and knowing what you can do to check and improve it, will go a long way to ensuring that your work is discoverable in the future.
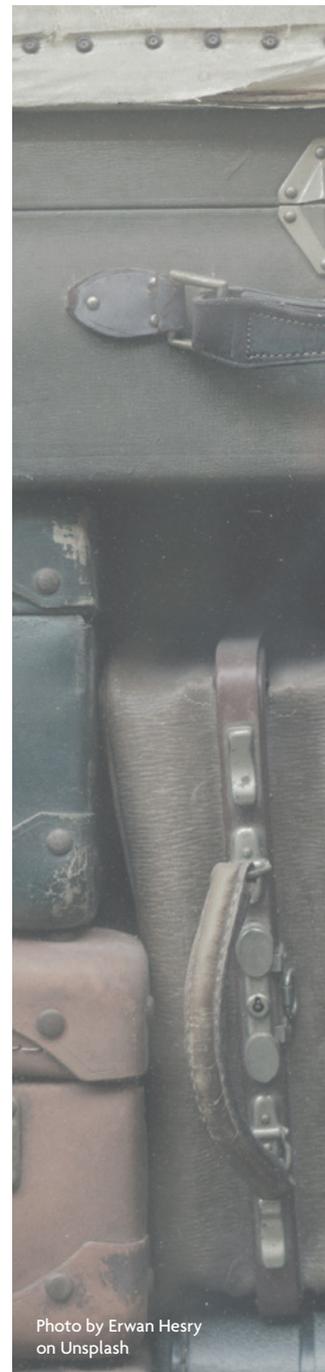
The Basics – How to Do It

As you submit your thesis or dissertation, you will be asked a series of questions (usually via a web form), and your answers will be stored as metadata alongside your text-based submission. If you produce a text-only thesis or dissertation, this metadata may be sufficient for your work.

However, if your research produces non-text components—images, audio, video, software, datasets, etc.—these components may not be encompassed within the metadata collected during the submission process. In order to make sure these files are discoverable and usable in the future, you will need to actively create metadata and store this information with the files. This process is described in more detail below.

As you are finishing your dissertation, the following tips will help you to avoid a few of the more common metadata pitfalls.

1. The abstract you supply will play a role in how others can find your work, so it is important that it includes a clear description and keywords relevant to your work.

2. Be careful with over-reliance on spell-check functions. For example, Microsoft Word does not spell-check capital letters, which can impact titles.

3. Create keywords that are not in your title. This will increase the discoverability of your work.

4. Define any acronyms you use (repeating them in both letters and in natural language).

5. Proofread all of your metadata, including department name and advisor name, prior to submission.

Tools – What to Use

Each document you work with, including the narrative of your dissertation and any accompanying supplementary materials or data you produce or rely upon, can (and should) be identified and described so that you and your future readers know what they are, who created them, and how they can be used. This process is extremely easy, and yet most people tend to overlook this step.

Simply adhering to your university and departmental guidelines for submitting the dissertation will go a long way in making sure that several very important metadata fields are accurately created, including author name, title, and details concerning your institution and degree granted.

Most colleges and universities have a web form-based submission process that will ask you a series of questions, prompting you to enter basic information that will roughly correspond to the following categories about your main (usually PDF-based) submission:

- Title
- Author/Creator
- Name of Advisor (committee member, contributor)
- Type of Resource
- Date
- Language
- Description/Abstract
- Subject
- Identifier
- Degree Information (name, level, department)
- Rights Management (access, license, rights holder)

Photo by Sanwal Deen on Unsplash

You may also submit additional files – spreadsheets, a collection of images, clips of interviews, or other materials – as parts of your thesis or dissertation. Most ETD submission processes do not collect additional metadata about these files. To help ensure that you and your readers will be able to understand what these additional files are and how they may be referenced, used, or built upon, you can develop a simple spreadsheet-based inventory of these items. This inventory should clearly identify how many additional files you are including, what they are, who created them, and what rights and licensing information they are governed by.

Take, for example, an archaeology thesis that includes original photographs of fieldwork. The narrative will probably include and identify the illustrations or images with captions, but if the images are substantively important to the thesis, the author may also submit the original high-resolution images alongside the dissertation. In so doing, the author could create a listing, or inventory, of them, where the file names correspond to items in the list.

Table 1: Inventory Example

Filename	description	creator	rights
Fieldfile1.tif	HImage of worksite prior to excavation, 2015/04/24	Johanna Brewster	CC By
Fieldfile2.tif	Image of worksite at excavation start, 2015/04/25	Gabriel Foster	CC By
Fieldfile3.tif	Image of worksite after discoveryobject 1, 2015/08/15	Johanna Brewster	CC By
Fieldfile4.tif	Image of discoveryobject 1, 2015/08/15	Gabriel Foster	CC By

Photo by Sergio Rola on Unsplash

6   ETDPlus |

Metadata is not just the information you intentionally create about your work. Some metadata is auto-generated through software programs (e.g., Microsoft Office, Adobe Photoshop, SPSS, etc.) and devices (e.g., scanners, digital cameras, digital recorders, etc.).

For example, in programs that are part of the Microsoft Office suite (Word, Excel, PowerPoint), there is an option called "Properties" under the "File" tab of the main navigation menu. The "Summary" tab then allows you to enter the title, subject, and, perhaps most importantly, the author of the document. Be aware that when you are working with colleagues or are using borrowed hardware or software, it is easy for these fields to become outdated and inaccurate unless they are changed manually. Reviewing the "Properties" of each file you produce and submit will help you to make sure that the correct author, title, and other elements are represented within the file itself.

Image 1: Properties Window, Microsoft Office
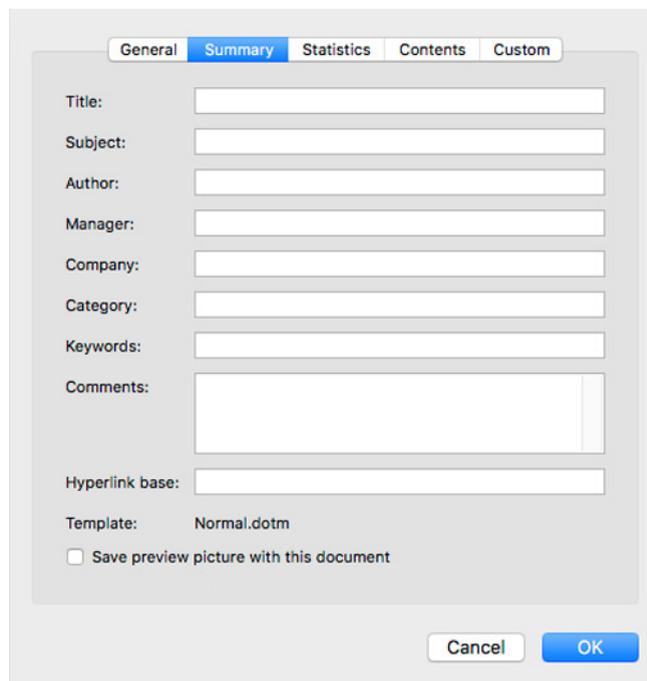
Activities

1. Choose three (3) documents you are currently working on, preferably in different formats (Word, Excel, etc.)

   a. Open the "Properties" tab in the "File" menu and examine the data you see under "Summary." Are you listed as the document's author? If not, this is a good time to update it!

   b. Review the fields and update them as needed, making sure to insert the most current information concerning the project title, category, and keywords.

All links provided were last checked 10/3/2017, and the content we reference here was saved on that date in the Internet Archive's Wayback Machine. For links that no longer work, please visit the Wayback Machine and enter the url to surface the resource.