



Metadata

Preservation and Curation of ETD Research Data
and Complex Digital Objects



EDUCOPIA
INSTITUTE

Workshop Background

Purpose

- Provide you with resources and tools to help you to create, store, and use metadata to help extend the use of your research outputs.

Context

- Workshop Series: Preservation and Curation of ETD Research Data and Complex Digital Objects
- Other topics: Copyright, Data Organization, File Formats, Storage, Version Control
- <https://educopia.org/research/etdplus>



Learning Objectives: Students

- Understand what metadata is and how it can impact the future usefulness of your research content.
- Learn what types of metadata you can create (and what metadata is auto-created, and how to check it for correctness).
- Learn how to use a basic spreadsheet to build an inventory for your ETD package.



Metadata

- Metadata 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 describes:
 - Who created it?
 - What is it?
 - When, where, how, and why was it created?



Metadata and your PDF submission

Common Metadata Fields

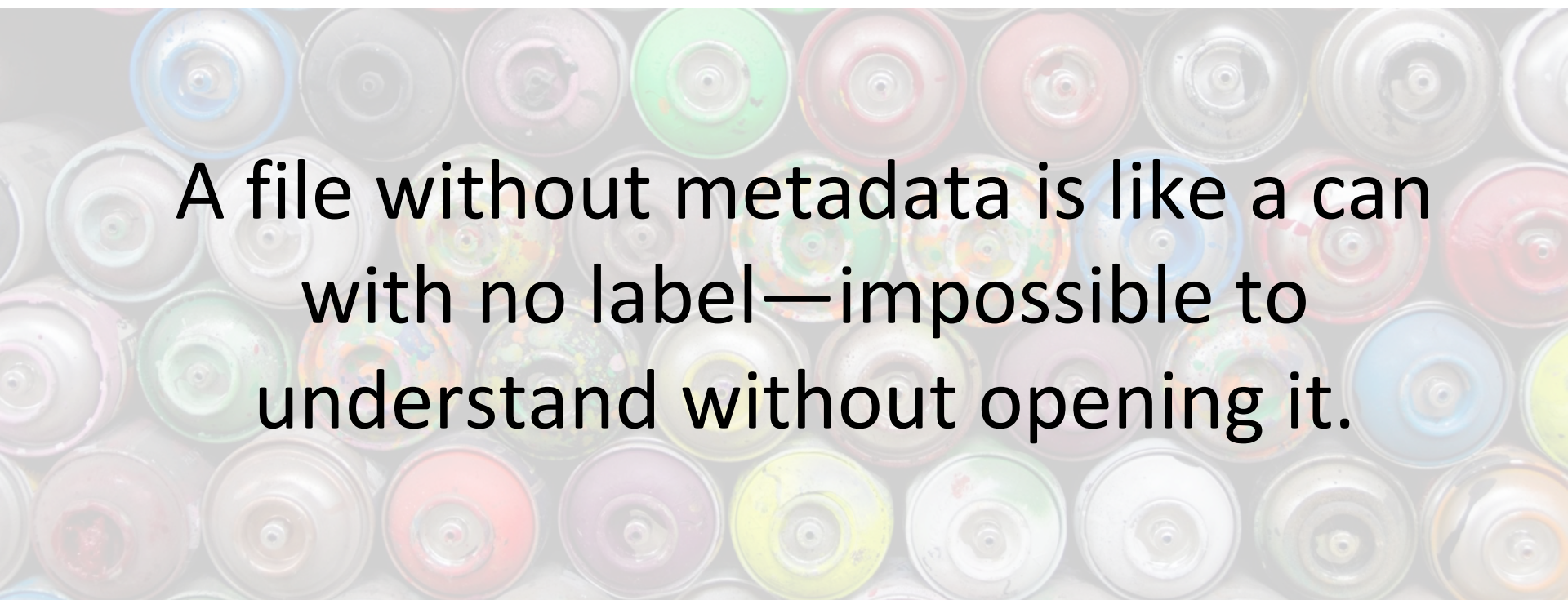
- Title
- Author/Creator
- Advisor
- Resource Type
- Date
- Language
- Description/Abstract
- Subject
- Identifier
- Degree Information
- Rights management information

Tips and “pitfalls” to avoid

- 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.
- Use keywords not in your title. This will increase the discoverability of your work.
- Define any acronyms you use (repeating them in both letters and in natural language).

Metadata and ETD research files

- Text-based PDF file(s)
 - Metadata is generated via webform during submission.
- Non-text, non-PDF components
 - Metadata is not actively collected about “supplementary” research files, even when those actually constitute the core of your submission.



A file without metadata is like a can
with no label—impossible to
understand without opening it.

Photo by Aleksander Naug on Unsplash

Metadata for “supplementary” files

- Develop a simple spreadsheet-based inventory of these items:
 - How many
 - What are they
 - Who created them
 - What rights/licensing information are they governed by
- Submit this inventory spreadsheet as one of your supplementary files in your ETD package

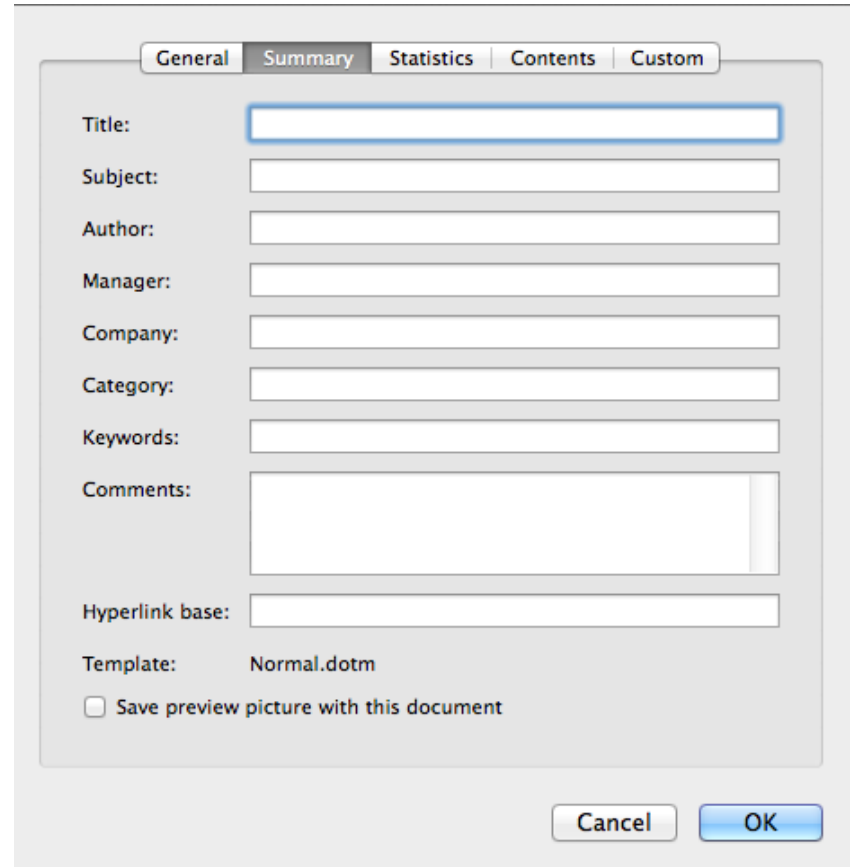


Inventory Example

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

Autogenerated Metadata

Metadata is not always created intentionally; some is auto-generated through software programs (e.g., Microsoft Office, Adobe Photoshop, SPSS) and devices (e.g., scanners, digital cameras, digital recorders). Always check these metadata fields under “Properties” in the navigation window for accuracy.



The image shows a screenshot of the Microsoft Office Properties dialog box, specifically the Summary tab. The dialog has five tabs: General, Summary (selected), Statistics, Contents, and Custom. The Summary tab contains several text input fields: Title, Subject, Author, Manager, Company, Category, Keywords, and Comments. Below these fields are Hyperlink base and Template (set to Normal.dotm). At the bottom, there is a checkbox labeled "Save preview picture with this document" which is currently unchecked. The dialog box has "Cancel" and "OK" buttons at the bottom right.

Example: Properties window, Microsoft Office (access via the “File” tab of the main navigation menu in Office by selecting “File – Properties”.)

Metadata

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. The quality of the metadata you provide is key to ensuring the future accessibility and usability of your work.

For every research file you create, you should also produce metadata describing:

- **Who** created the content?
- **What** is it?
- **When and Where** was it created?
- **How and Why** was it developed?

Why might you need metadata?

if your research produces images, audio, video, software, datasets, or other non-text components, these 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 it with the files.

ETD metadata tips:

1. Your abstract needs a clear description and keywords relevant to your work.
2. Be careful with over-reliance on spell-check functions. For example, Microsoft Office 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 (repeat them in both letters and in natural language).
5. Proofread all of your metadata, including department name and advisor name, prior to submission.



A file without metadata is like a can with no label - impossible to understand without opening it (and perhaps even then!)

Typical metadata requested about a pdf during the ETD submission process:

- Title
- Author/Creator
- Advisor
- Resource Type
- Date
- Language
- Description/Abstract
- Subject
- Identifier
- Degree Information
- Rights management information

Most ETD submission processes **do not** collect metadata about the additional files you may submit (e.g., datasets, audio or video files, image files, GIS files, CAD files, software code, etc.). To help make sure 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. Submit this inventory spreadsheet as part of your ETD package.

Activity

- Choose three (3) documents you are currently working on, preferably in different formats (Word, Excel, etc.).
 - 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!
 - Review the fields and update them as needed, making sure to insert the most current information concerning the project title, category, and keywords.

