Sharing Data with D-Scholarship@Pitt

University Library System (ULS), University of Pittsburgh

“Sharing Data with D-Scholarship@Pitt” by the ULS Research Data Management Team is licensed under a Creative Commons Attribution 4.0 International License.
Table of Contents

1. Introduction ...................................................................................................................................... 3

2. File Naming ..................................................................................................................................... 3

3. Data Versioning ................................................................................................................................. 4

4. File Formats ...................................................................................................................................... 6

5. Rights and Permissions ................................................................................................................... 6
   Quantitative Data .................................................................................................................................. 7
   Qualitative Data .................................................................................................................................. 7
   Access Control and Permissions for Sensitive Data .............................................................................. 8
   Licensing Options ............................................................................................................................... 8

6. readme Files ..................................................................................................................................... 9

7. Data Dictionaries .............................................................................................................................. 10

8. D-Scholarship@Pitt ......................................................................................................................... 10
   Depositing your data in D-Scholarship@Pitt .................................................................................. 10
   D-Scholarship@Pitt Metadata ........................................................................................................... 11
     D-Scholarship@Pitt Required Elements for Data Sets ................................................................... 11
   D-Scholarship@Pitt Data Dictionary ............................................................................................... 11
     Details ............................................................................................................................................... 11
     Dataset Details ............................................................................................................................. 15
     Upload ............................................................................................................................................. 16
     Type of Data Vocabulary .............................................................................................................. 17

9. Decision Tree for Depositing Data in D-Scholarship@Pitt .............................................................. 18
1. Introduction

D-Scholarship@Pitt, the University of Pittsburgh’s institutional repository, offers long-term preservation of scholarly materials, including research data. This guide is intended to assist Pitt researchers with the process of data sharing through D-Scholarship@Pitt. We offer guidance on approaches for data organization and documentation to implement during a research project, with the intent of facilitating data sharing.

Good data documentation describes the creation, context, structure, contents, and any manipulations made to the data. Data documentation helps to ensure that your data are interpretable and useful to others. There is a variety of components to data documentation. Relatively simple practices, like using consistent file naming conventions and data versioning will allow users to more easily find the appropriate files. Readme files and data dictionaries help users to navigate your data, by explaining variables, software requirements, structures, and manipulations to the data. Metadata allow potential users to discover and better understand your data.

What is important to document?

- Context of data collection
- Data collection methodology
- Structure and organization of data files
- Data validation and quality assurance
- Data manipulations through data analysis from raw data
- Data confidentiality, access and use conditions

Data-level documentation should include:

- Variable names and descriptions
- Definition of codes and classification schemes
- Codes of, and reasons for, missing values
- Definitions of specialty terminology and acronyms
- Algorithms used to transform data
- File format and software used

2. File Naming

Before you begin your research, decide on a naming convention for your files. File names should allow you to identify a precise experiment from the name. It is a good idea to set up a clear directory structure that includes information like the project title, a date, and some type of unique identifier. Individual directories may be set up by date, researcher, experimental run, or whatever makes sense for you and your research. Choose a format for naming your files and use it consistently. Document the naming convention you choose, and make sure that you and your collaborators follow it. It will save you time and will help others who may use your files in the future.
Best practices include:

- Give files a meaningful name. A file name might include a combination of elements, such as type of equipment used, date, and researcher's surname. Decide on the best order for elements in a file name; it will affect how the files are sorted.
- Keep names a reasonable length; some applications will not work well with long file names. A maximum of 25 characters is a good rule of thumb.
- To separate elements in a file name, consider using underscores (_) or hyphens (-). Avoid using blank spaces in a file name. Use periods only to separate the file name from the file type extension (.txt, .jpg, etc.)
- If including date as part of the file name, use the standard format yyyymmdd to ensure that files sort in chronological order.
- If your file name will include a numerical component, such as a subject number or version number, use leading zeros (001, 002, etc.) so that files sort in sequential order.
- Avoid special characters like ~ ! @ # $ % ^ & ( ) ; < > , [ ] { } ‘ “

More considerations for naming files can be found at these websites:

- [Best practices for file naming](http://library.stanford.edu/data-management-services/file-naming-and-versioning.html) (Stanford University Libraries - Data Management Services)
- [File naming and versioning](http://researchdata.wisc.edu/file-naming-and-versioning.html) (University of Wisconsin Research Data Services)
- [File naming conventions](http://library.purdue.edu/data-management-grad-researchers/file-naming-conventions.html) (Purdue University Libraries - Data Management for Graduate Researchers)
- [File management](http://researchdata.cornell.edu/file-management.html) (Cornell University Research Data Management Service Group)

3. Data Versioning

Versioning refers to saving new copies of your files when you make changes so that you can go back and retrieve specific versions of your files later. Saving multiple versions makes it possible to decide later that you prefer an earlier version. You can then immediately revert to that version instead of having to retrace your steps to recreate it.

In its most basic form, versioning relies on a sequential numbering system. Within a given version number category (major, minor), these numbers are generally assigned in increasing order and correspond to changes in the data. The US Geological Survey recommends the following structure:

- DataFileName_1.0 = original document
- DataFileName_1.1 = original document with minor revisions
- DataFileName_2.0 = document with substantial revisions

The ETDplus project, led by the Educopia Institute, offers additional guidance for version control. Versioning should be taken into account when developing the folder and file naming structure. The following guidance is taken from the ETDplus brief on versioning, available on the [project site](http://etdplus.org):
At the beginning of a research project, it is important to create a stable folder structure in which you can organize materials. The specific folders will depend on your own research process. File organization could be based on how you plan to gather materials, which experiment or process generated them, when they were created, or other strategies. The key is to use folders that make sense to you and allow you to easily find your materials. A simple method to designate a revision is to note it at the end of the file name. This way, files can be grouped by their name and sorted by version number. For example:

- image1_v1.jpg
- image1_v2.jpg
- image2_v1.jpg
- image2_v2.jpg
- ...

If you use version numbers, one issue that can arise is that computers will sort files based on the position of the characters. This can lead to strange, unhelpful results. For example:

- image1_v1.jpg
- image1_v10.jpg
- image1_v2.jpg
- ...

A good practice that can help you to avoid these problems is to use dates to designate version numbers. If you choose this strategy, format dates as year-month-day (20150930). Using this order will help avoid confusion when collaborating with other researchers or systems that use a day-month-year or month-day-year, and it will help your computer sort versions in chronological order. For example:

- image1_20151021
- image1_20151214
- image1_20160123
- ...

If the files you are using are created or edited collaboratively, you may want to incorporate names or initials into your file naming conventions so that you know which versions contain updates by each individual on your team. For example:

- dataset1_20160402_KES
- dataset1_20160301_WTC
- dataset1_20160814_GSC
- ...

4. File Formats
When selecting a file format for your data deposit, the format should ideally be non-proprietary, unencrypted, uncompressed, and in common usage in your research community. The format of the electronic data files you work with during your research may be determined by the research equipment and computer hardware and software. However, for long-term preservation and ease of sharing, best practices may dictate that the files be converted to a different format after your project has ended. If conversion to an open data format will result in some data loss from your files, you might consider saving the data in both the proprietary format and an open format. When it is necessary to save files in a proprietary format, consider including a readme.txt file in your directory that documents the name and version of the software used to generate the file, as well as the company who made the software.

Some preferred file formats:
- Containers: TAR, GZIP, ZIP
- Databases: XML, CSV
- Geospatial: SHP, DBF, GeoTIFF, NetCDF
- Moving images: MOV, MPEG, AVI, MXF
- Sounds: WAVE, AIFF, MP3, MXF
- Statistics: ASCII, DTA, POR, SAS, SAV
- Still images: TIFF, JPEG 2000, PDF, PNG, GIF, BMP
- Tabular data: CSV
- Text: XML, PDF/A, HTML, ASCII, UTF-8
- Web archive: WARC

Additional helpful guidelines for selecting file formats can be found at these websites:
- [Best Practices for File Formats](https://www.stanford.edu/group/DMS/bestpractices.html) (Stanford University Libraries - Data Management Services)
- [Choosing formats](https://www.lib.cam.ac.uk/services/datamanagement/chooseformat.html) (Cambridge University Libraries - Data Management)
- [File formats](https://www.library.cornell.edu/researchdata/fileformats.html) (Cornell University - Research Data Management Service Group)
- [ResearchWorks Archive List of Preferred File Formats](https://www.library.washington.edu/researchworks/preferredfileformats.html) (University of Washington - University Libraries)

5. Rights and Permissions
As a data creator, you have certain rights over the work and an opportunity to license your data appropriately to facilitate sharing and re-use. The application of copyright and licensing depends on several factors - whether your data set contains quantitative data, qualitative data, or sensitive information. Copyright and licensing options vary depending on the type of data and its sensitivity.

Best practices include:
- Understanding the nature of your data set and whether your data are subject to copyright.
• Making your data as open and reusable as possible, ideally by dedicating it to the Public Domain.
• Identifying any restrictions of sharing data, e.g. from Terms of Use.
• Asserting your rights under the Doctrine of Fair Use if necessary.
• Considering carefully any ethical questions involved in sharing your openly and choosing licensing and access options to match.

The information presented here is a brief overview of a very complicated topic. Please get in touch with the Research Data Management team for help with any of the rights and permissions considerations below.

Quantitative Data
By quantitative data, we mean data that are numerical values or measurements of facts about the universe. Because facts are not subject to copyright, most quantitative data are not copyrightable in the United States and copyright laws usually do not apply or are not enforceable.

However, the arrangement, selection, and coordination of the data set as a whole may be subject to copyright. This depends on the creativity involved with arranging and displaying the data.

- [Copyright Basics](#) from the U.S. Copyright Office (see page 3, "What is not Protected by Copyright?")
- [U.S. Copyright Office Report on Legal Protection for Databases](#)

Many researchers believe in the importance of sharing quantitative data openly to facilitate the greatest possible reuse of the data. For example, [Dryad](#) and the [Panton Principles](#) for Open Data strongly recommend that data be contributed to the public domain. When a data set is dedicated to the Public Domain, then the creator declares that others may use the data set in its current form (and, therefore, the potential copyright of the arrangement, selection, and coordination of the data set are dedicated to the Public Domain). Below are two examples of licenses that a data set creator can apply to a quantitative data set to dedicate it to the Public Domain.

- [Creative Commons CC0 License](#)
- [Open Data Commons Public Domain Dedication and License (PDDL)](#)

Qualitative Data
By qualitative data, we mean data that contain observations, texts, conversations, artistic or creative works, which are usually collected in the humanities and social science fields. Some examples of qualitative data include text corpora, interviews, photographs, and social media output. Because these are often creative expressions made by individuals that are fixed in a tangible form, many of these data sets are subject to copyright and permission may need to be obtained for their use. For those compiling qualitative data sets, privacy, ethics, and licenses are of key concern.

For those collecting interviews or other recordings and documentation made by research subjects, clear guidelines for the usage and ownership of these materials should be set out
in a Consent Form and cleared with the IRB. This is also the case when research work is conducted via the Internet.

- **Considerations and Recommendations Concerning Internet Research and Human Subjects Research Regulations** from the US Department of Health and Human Services
- **Guidelines for Ethical Conduct in Participant Observation** (University of Toronto) - contains advice on what to consider when writing a consent form and protocol.
- **Communicating Qualitative Research Study Designs to Research Ethics Review Boards** (2011) by Carolyn Ells - a discussion on ethical issues in collecting data for qualitative research studies and how to construct a protocol that reflects these considerations.

Researchers must identify whether the data are in the public domain, subject to licensing terms, or may qualify as Fair Use. Because these data sets often include substantial transformative use, a Fair Use argument may be particularly powerful for qualitative data sets.

- **Copyright and Intellectual Property Toolkit: Public Domain** - contains information on how to determine if an item is in the public domain.
- **Copyright and Intellectual Property Toolkit: Fair Use** - contains information about the doctrine of Fair Use and tools for making a Fair Use argument.
- **Understanding Fair Use: Transformative Use** - read more about the "Fifth Factor" of Fair Use, Transformation.

When obtaining data from the Internet via scraping tools, the restrictions in Terms of Service and Developer Policies apply, especially from social media websites.

- **Fair Use in the Age of Social Media** - an article covering the basics of Fair Use in social media contexts.
- **Challenges of Using Twitter as a Data Source** - covers some of the issues with using and sharing qualitative social media data sets, including licensing issues. See also Twitter’s **Developer Policy**, which applies to those creating data sets by scraping Twitter.

**Access Control and Permissions for Sensitive Data**
For data sets that contain sensitive research, e.g. human subject research, access control may be an option. Mixed levels of access control may be put in place for some data, combining controlled access to confidential data with standard access to non-confidential data.

- **Sharing Sensitive Data** from the Australian National Data Services, including a decision tree on whether to share data or not.
- **Ethical Issues in Research Data Management** from University College Dublin, including what may make data "sensitive" and how to anonymize data sets for open sharing.

**Licensing Options**
Beyond the Public Domain licensing options above, there are some other licensing options that can apply to data sets. Creative Commons licenses allow creators to specify the rights
for reuse - typically with attribution to the creator, but potentially also including bans on commercial use and derivatives. It is not recommended to prohibit derivative works on a data set, as this will compromise the usability of that data set.

- How to License Research Data by the Digital Curation Centre (UK)
- Copyright and Intellectual Property Toolkit: Creative Commons, Copyleft, and Other Licenses
- Open Data Commons - licenses specifically created for data reuse, including a Public Domain dedication as well as an Attribution Required license.

Licenses can work in tandem with access control, Fair Use, and ethical considerations detailed above. For complex situations, contact us for guidance.

6. readme Files

In the context of research data, a readme file is a plain text file (.txt) that helps others understand your data and interconnections among data files. By titling the file "readme," the date creator signals to other users that this file should be looked at first. For researchers depositing data in D-Scholarship@Pitt, the information in the readme file may mirror and augment information included in the metadata form and, if the deposit includes multiple files, may explain the file naming structure, relationship among the files, and abbreviations used.

Cornell University's Research Data Management Service Group has made a useful readme file template available for download. At a minimum, the Cornell group recommends completing the following sections in the readme file template:

General information
- Data set title
- Name and contact information for investigators
- Date (or data range) of data collection
- Geographic location of data collection

Data and file overview
- A short description of each file
- Date that the file was created

Methodological information
- Description of methods for data collection
- Description of methods for data processing

Data specific information
- Variable list, with full names and definitions of column headings if tabular data
- Units of measurement
- Definitions for codes or symbols used to record missing information (see Cornell University, Guide to writing "readme" style metadata)
7. Data Dictionaries

A data dictionary describes all the data stored in a data set or used by a database, including their types, attributes, structure, relationships, and usage in the database or software program. A good data dictionary can be a valuable part of the metadata describing a data set, enabling a user to get a clear understanding of the content and organization of the data and how it could be modified, if necessary. In the context of a database or software package, the data dictionary may be an essential piece of software that programmers and the database management system require to access and use the data properly. The user view of a data dictionary is usually presented as a table or spreadsheet. Dictionaries may also be incorporated into XML files or other mark-up languages. A data dictionary does not contain the data, but only describes it.

A data dictionary typically contains a list of all files in the database, names for each file, the type of data included, a list of all field names and variable names, a description of the information contained in each field, and the various attributes of each field. These may include type (text, date, numeric, etc.), standard formats, units, field length, description, unique identifiers, default values, whether a value is required or not, and more, depending on the specific data.

For some examples of data dictionaries, check the following sites:

- [Data Dictionary Examples](https://datacommons.org) – Ag Data Commons – National Agricultural Library - USDA
- [Sample Dataset 2014](https://ckan.org) - Statistical Consulting at University Libraries, Kent State University. Click on the link to “Data definitions (*pdf)” in the Sample Data Files section.
- [Protein Data Bank Exchange Data Dictionary](https://www.rcsb.org) (PDBx/mmCIF V4.0) – Worldwide Protein Data Bank. There are separate tabs for Category Groups, Data Categories, and Data Items.

8. D-Scholarship@Pitt

**Depositing your data in D-Scholarship@Pitt**

The University of Pittsburgh’s institutional repository, D-Scholarship@Pitt, offers long-term storage for scholarly output. Pitt researchers can upload their published or unpublished work to D-Scholarship@Pitt, including data sets. When depositing data, you can select the type of resource you want to add to the repository.

- **D-Scholarship@Pitt**
  - Accepts nearly any format of file including tar.gz and zip files
  - Assigns your data deposits a Digital Object Identifier (DOI), a permanent and unique identifier for a digital object that is used in citations that will help others to find and cite your data
- Allows you to add information that provides important context for your data so that others can discover, understand, and trust the data files
- Is best suited for data sets that are in an inactive state (i.e., after the completion of a research project)
- Tracks your work using alternative metrics to help demonstrate your impact and see how others are using your data
- Can be used to add a catalog-only entry for data sets that you've deposited in another data repository
- Allows you to make data fully public, private, or available only to the Pitt community.

**D-Scholarship@Pitt Metadata**

Metadata are structured information that provide context for information objects of all kinds, including research data. Metadata enables discovery, use, exchange, and preservation of those objects. There are a variety of metadata standards for describing data sets, based on discipline, international standards, and many other characteristics of the data. Common characteristics of data set metadata include information about the researchers involved with the data creation, a name or title of the data set, dates associated with the creation of the data, a brief description or abstract, and terms and conditions associated with the data set.

**D-Scholarship@Pitt Required Elements for Data Sets**

- **Details**
  - Title
  - Creators/Authors
  - Publication Details
    - Status
    - Date
  - Language
    - Resource Language
    - Metadata Language
- **Data set Details**
  - Type of Data
  - Time Period
  - Collection Period
  - Copyright Holders
- **Upload**
  - Visible to
  - File Type
  - License

**D-Scholarship@Pitt Data Dictionary**

The following tables document all of the metadata fields contained in D-Scholarship@Pitt.

**Details**

General metadata about the data set
<table>
<thead>
<tr>
<th>Field Name</th>
<th>Required?</th>
<th>Definition</th>
<th>Allowed Values &amp; Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Yes</td>
<td>A name or title by which a resource is known.</td>
<td>Free text. Titles should be concise and descriptive of the object.</td>
</tr>
<tr>
<td>Abstract</td>
<td>No</td>
<td>A summary of the items content.</td>
<td>Free text. All additional information that does not fit in any of the other categories.</td>
</tr>
<tr>
<td>Creators/Authors</td>
<td>Yes</td>
<td>The main researchers involved in producing the data, or the authors of the publication, in priority order.</td>
<td>Free text. Enter the authors or creators of the work in the order and format given on the title page of the work. If there are more than four authors, click on the [More input rows] button. Pitt Username should be the University of Pittsburgh computer account name you use to access your email. (Non-Pitt users can leave this field blank.)</td>
</tr>
<tr>
<td>Corporate Creators</td>
<td>No</td>
<td>Companies that are considered authors of this item.</td>
<td>Free text.</td>
</tr>
<tr>
<td>Contributors</td>
<td>No</td>
<td>The institution or person responsible for collecting, managing, distributing, or otherwise contributing to the development of the resource.</td>
<td>Free text. If there are more than four contributors, click on the [More input rows] button. Pitt Username should be the University of Pittsburgh computer account name you use to access your email. (Non-Pitt users can leave this field blank.)</td>
</tr>
<tr>
<td>Schools and Programs</td>
<td>No</td>
<td>The Schools and Programs with which this item should be associated.</td>
<td>Select one or more appropriate options from the list. To select more than one school, hold the &quot;Ctrl&quot; key (Windows) or &quot;Command&quot; key (Mac) while selecting.</td>
</tr>
<tr>
<td>Centers</td>
<td>No</td>
<td>The Centers and Institutes with which this item should be associated.</td>
<td>Select one or more appropriate options from the list. To select more than one school, hold the &quot;Ctrl&quot; key (Windows) or &quot;Command&quot; key (Mac) while selecting.</td>
</tr>
<tr>
<td>Publication Details</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>-----</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-- Status</td>
<td>Yes</td>
<td>The current status of the data set.</td>
<td>Select the most appropriate status: Published, In Press, Submitted, Unpublished.</td>
</tr>
<tr>
<td>-- Media of Output</td>
<td>No</td>
<td>Media of output</td>
<td>If there are any physical media associated with the resource, they should be listed here.</td>
</tr>
<tr>
<td>-- Publisher</td>
<td>No</td>
<td>The name of the entity that holds, archives, publishes prints, distributes, releases, issues, or produces the resource</td>
<td>Free text. Do not use abbreviations. Enter full name of the publisher.</td>
</tr>
<tr>
<td>-- Date</td>
<td>Yes</td>
<td>The date when the data was completed, submitted to a publisher, or published.</td>
<td>Four (4) digit year. Select month and day if known.</td>
</tr>
<tr>
<td>-- Date Type</td>
<td>No</td>
<td>The event to which the date applies.</td>
<td>Select most appropriate option.</td>
</tr>
<tr>
<td>-- DOI or Unique Handle</td>
<td>No</td>
<td>The Digital Object Identifier Number (DOI) or unique identification number that works with a Handle System like Handle.net.</td>
<td>Enter the globally unique DOI or Handle.</td>
</tr>
<tr>
<td>-- Official URL</td>
<td>No</td>
<td>The canonical URL for this item.</td>
<td>Some journals require depositors to provide a link to the publisher's version of the object.</td>
</tr>
<tr>
<td>-- Related URLs</td>
<td>No</td>
<td>Links to resources related to the deposited item.</td>
<td>Both the URL and the URL description should be provided. The description will assist others to understand the relevance of the links provided.</td>
</tr>
<tr>
<td>Language</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Field</td>
<td>Required</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>----------------------------</td>
<td>----------</td>
<td>-----------------------------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Resource Language</td>
<td>Yes</td>
<td>The primary language of the resource.</td>
<td>Select one from the list.</td>
</tr>
<tr>
<td>Metadata Language</td>
<td>Yes</td>
<td>The language of the metadata record.</td>
<td>Select one from the list.</td>
</tr>
<tr>
<td>Other ID</td>
<td>No</td>
<td>Any identifiers that should be associated with the document from sources such as arXiv, non-DOI handles, etc.</td>
<td>Globally unique identifiers preferred. DOIs and handles should be collected in the &quot;DOI or Unique Handle&quot; field.</td>
</tr>
<tr>
<td>Funders</td>
<td>No</td>
<td>Name of the organization that provided financial support for the development of the resource.</td>
<td>Free text.</td>
</tr>
<tr>
<td>Projects</td>
<td>No</td>
<td>The names or codes of the projects that created this item.</td>
<td>Free text.</td>
</tr>
<tr>
<td>Grant Information</td>
<td>No</td>
<td>Information about the grant that funded the creation of this item.</td>
<td>Free text. Fields should be filled out completely as possible.</td>
</tr>
<tr>
<td>Contact Email Address</td>
<td>No</td>
<td>The contact email address for this item.</td>
<td>The email address will not be made public. If the full-text is not available to the public, then requests to view the full-text will be sent to this email. Please enter a single email address only. This is field is automatically populated by your login information.</td>
</tr>
<tr>
<td>Uncontrolled Keyword</td>
<td>No</td>
<td>Natural language terms to describe the content of the item.</td>
<td>Free text. Keywords should be separated with a comma (,) or semicolon (;).</td>
</tr>
<tr>
<td>Additional Information</td>
<td>No</td>
<td>Any information about your deposit that cannot be entered elsewhere.</td>
<td>Free text. This information will appear on the public summary page for this item.</td>
</tr>
<tr>
<td>Field Name</td>
<td>Required?</td>
<td>Definition</td>
<td>Allowed Values &amp; Recommendations</td>
</tr>
<tr>
<td>----------------------------</td>
<td>-----------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Comments and Suggestions</td>
<td>No</td>
<td>Comments to the D-Scholarship@Pitt repository managers.</td>
<td>Free text. This information will not be displayed to the public.</td>
</tr>
<tr>
<td>Dataset Details</td>
<td></td>
<td>Metadata specific to data sets</td>
<td></td>
</tr>
<tr>
<td><strong>Type of Data</strong></td>
<td>Yes</td>
<td>The primary category of data that is represented in the deposit.</td>
<td>Select the most appropriate entry. See “Type of Data Vocabulary” for definitions of the terms.</td>
</tr>
<tr>
<td><strong>Time Period</strong></td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Collection Period</strong></td>
<td>Yes</td>
<td>Start and end dates for the collection of the data.</td>
<td>Enter year, month, and day when the data was created</td>
</tr>
<tr>
<td><strong>Temporal Coverage</strong></td>
<td>No</td>
<td>Start and end date for this data, if it is different from the data collection period.</td>
<td>Dates for data representing a historical period. Only enter if the dates differ from those in the Collection Period element. Year, month, and day.</td>
</tr>
<tr>
<td>Geographic Location</td>
<td>No</td>
<td>Longitude and Latitude values of a theoretical geographic bounding rectangle that would cover the region in which your data was created</td>
<td>Enter the north and south latitudes and east and west longitudes.</td>
</tr>
<tr>
<td>Field Name</td>
<td>Required?</td>
<td>Definition</td>
<td>Allowed Values &amp; Recommendations</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>-----------</td>
<td>-----------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Geographic Coverage</td>
<td>No</td>
<td>The geographic region(s) in which your data collection was carried out (e.g. name of country, region, location) or the spatial reference system used.</td>
<td>Recommend using an authoritative source for geographic names, such as the Getty Thesaurus of Geographic Names or Geographic Names Information System (GNIS).</td>
</tr>
<tr>
<td>Data Collection Method</td>
<td>No</td>
<td>The methodology used to create this data collection.</td>
<td>Include information on the studied population, studied entities and sampling procedures.</td>
</tr>
<tr>
<td>Data Processing and Preparation Activities</td>
<td>No</td>
<td>Information about the source of any data or data element not derived from previous work, and any changes made to this data prior to deposit.</td>
<td>Include information on data cleaning, transformation, and software/hardware used for data preparation and processing.</td>
</tr>
<tr>
<td>Statement on legal, ethical and access issues</td>
<td>No</td>
<td>A description of any legal, ethical issues encountered in generating the data.</td>
<td>Explain how legal and ethical issues have been dealt with in order to allow sharing, including access restrictions in place.</td>
</tr>
<tr>
<td>Copyright Holders</td>
<td>Yes</td>
<td>A person, firm or corporate body, which holds the copyright for the item.</td>
<td>For questions regarding copyright, consult with staff from the Office of Scholarly Communications and Publishing.</td>
</tr>
</tbody>
</table>

**Upload**

Metadata that are specific to data files. These fields should be completed for each document that makes up this deposit.

<table>
<thead>
<tr>
<th>Field Name</th>
<th>Required?</th>
<th>Definition</th>
<th>Allowed Values &amp; Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visible to</td>
<td>Yes</td>
<td>The security level of the document.</td>
<td>Select either &quot;Anyone (Open Access)&quot; or &quot;University of Pittsburgh users only&quot;. If you limit to University of Pittsburgh users, consider adding an email address</td>
</tr>
</tbody>
</table>
on the "Details" tab. This will allow users to use a web form to request a copy of the document. Depositors can decide on a case-by-case basis to provide access.

<table>
<thead>
<tr>
<th>Content</th>
<th>No</th>
<th>The content of this document.</th>
<th>Select the most appropriate entry.</th>
</tr>
</thead>
<tbody>
<tr>
<td>File Type</td>
<td>Yes</td>
<td>The storage format of the document</td>
<td>Select the most appropriate entry.</td>
</tr>
<tr>
<td>Description</td>
<td>No</td>
<td>Additional brief description of the format.</td>
<td>Include any other information that a user may need to access the specific file.</td>
</tr>
<tr>
<td>Embargo expiry date</td>
<td>No</td>
<td>Date that a publisher-imposed or sponsor-imposed embargo expires.</td>
<td>On and after this date, the document will be made publicly accessible.</td>
</tr>
<tr>
<td>License</td>
<td>Yes</td>
<td>Creative Commons license applied to the document.</td>
<td>Specify an explicit license for the document. Specifying a license does not alter the access rights your grant at the end of the deposit process. You may specify a license for items you deposit only if your own the copyright for that item.</td>
</tr>
</tbody>
</table>

**Type of Data Vocabulary**

<table>
<thead>
<tr>
<th><strong>Option</strong></th>
<th><strong>Description</strong></th>
<th><strong>Examples</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td>A resource consisting primarily of plain text.</td>
<td>Tab-delimited (TSV) files, comma-delimited (CSV) files, and spreadsheets. May include other text-based documents, such as grey literature, lab notes, etc.</td>
</tr>
<tr>
<td>Audio</td>
<td>A resource primarily intended to be heard.</td>
<td>Sound recordings, including interviews, music files, oral histories, etc.</td>
</tr>
<tr>
<td>Video</td>
<td>A series of visual representations imparting an impression of motion when</td>
<td>Films, video, etc.</td>
</tr>
</tbody>
</table>
shown in succession. May or may not include sound.

<table>
<thead>
<tr>
<th>Image</th>
<th>A visual representation other than text.</th>
<th>Digitized or born digital images, drawings or photographs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>An abstract, conceptual, graphical, mathematical or visualization model that represents empirical objects, phenomena, or physical processes.</td>
<td>Modelled descriptions of, for example, different aspects of languages or a molecular biology reaction chain.</td>
</tr>
<tr>
<td>Software</td>
<td>A computer program in compiled form.</td>
<td>Software supporting research</td>
</tr>
<tr>
<td>Code</td>
<td>A computer program in source code (text).</td>
<td>Code supporting research</td>
</tr>
<tr>
<td>Machine/Instrument Log</td>
<td>Raw data output from a tool such as a computer or other instrument.</td>
<td>Log files generated from instruments.</td>
</tr>
<tr>
<td>Database</td>
<td>Data encoded in one or more related tables.</td>
<td>Relational datasets, Microsoft Access, Oracle, MySQL, etc.</td>
</tr>
<tr>
<td>Mixed</td>
<td>Data consisting of significant materials in two or more forms.</td>
<td>Mixed forms of materials, such as text, photographs, and sound recordings, contained in a dataset.</td>
</tr>
<tr>
<td>Other</td>
<td>Used when no other “Type of Data” option is appropriate.</td>
<td>If a suitable category is not listed, select 'Other'</td>
</tr>
</tbody>
</table>

**9. Decision Tree for Depositing Data in D-Scholarship@Pitt**

This decision tree is designed to assist researchers completing an ETD as part of a degree program as they decide whether to submit the data as its own record in D-Scholarship@Pitt or as a supplemental file. The same guidance can be used for data that supports another type of scholarly output.
Did you collect data as part of your thesis or dissertation research?

Yes

Would you like your data to be discoverable and citable separate from your ETD?

If Yes to either

Create a separate record for your data in D-Scholarship@Pitt.

Before creating a separate record for your data, ask yourself

Are you finished updating your data?
Are you the owner of the data?
For research with human subjects, have you anonymized your data?

If Yes to all

Deposit your data as its own record in D-Scholarship@Pitt. Read our guide on data sharing and contact the library for help.

If No to any

Deposit your data as a supplemental file or files with your ETD.

Contact the library for assistance navigating these issues!

If No to all

If you'd like ETD formatting assistance, the library has classes and drop-in hours.

Visit us at Library.pitt.edu

Sharing your data with D-Scholarship@Pitt