ETD Plus: When Non-traditional is the New Normal, What's the Norm for ETD Programs?

Martin Halbert
University of North Carolina at Greensboro, mdhalber@uncg.edu

Follow this and additional works at: https://docs.lib.purdue.edu/etdgiantleaps

Part of the Archival Science Commons, Higher Education Commons, and the Scholarly Publishing Commons

Recommended Citation
https://docs.lib.purdue.edu/etdgiantleaps/13

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
ETD Plus:
When Non-traditional is the New Normal, What's the Norm for ETD Programs?

Martin Halbert (MLS, PhD)
Dean of University Libraries
University of North Carolina at Greensboro
Giant Leaps - Symposium on Electronic Theses and Dissertations
Purdue, Indiana
Thursday, May 23, 2019
Overview

1. Core problems
2. IMLS Funded Inter-Institutional Projects
   a. Lifecycle Management of ETDs
   b. DataRes & ICAMP projects
   c. ETDplus project
3. ETDs and Broader Open Research Data Management Issues – APLU/AAU/NSF Workshop Initiative
4. Future Directions – Where do we go now?
Core Issues

• What made me care about any of this?
• What problems did I and others notice that we thought needed to be addressed through a series of collaborative projects?
Traditional Print Theses and Dissertations Repositories
Continuity with the Past through ETD Programs

• Most ETD repositories have primarily been collections of PDFs
• PDFs as format of choice for long term preservation
But not all dissertations are solely print

Associated Recitals are an example:


http://digital.library.unt.edu/ark:/67531/metadc30516/
Many are primarily non-textual

- POP BAROQUE, by Jennifer Gassiraro, B.F.A
- “Problem in Lieu of Thesis” (?) Prepared for the Degree of MASTER OF FINE ARTS
- “...vivid colors, textures, patterns and designs collected from my environment...”
- Only 9 pages of text, primarily description of a fabric arts exhibit that was the actual work
The Only Surviving Evidence of the Real Work

- Eight thumbnail images of the exhibit as an appendix
...and then there are truly avant-garde works

- **Critical Discussion of Pleroma: A Digital Drama and Its Relevance to Tragic Form in Music**, by Stephen Lucas, B.M. Thesis Prepared for the Degree of MASTER of MUSIC

- “Pleroma is a digital drama: a work composed of digital animation combined with electroacoustic music, presenting an original dramatic narrative. Pleroma's dramatic elements evoke both the classical form of tragedy and the concept of perceptual paradox...”
Complex Digital Media are Complicated to Preserve and Access

• Pleroma is depicted as a floating eyeball with half of a blue, translucent eyelid encircling its form. Abraxas and Gnosis are two sides of a double-headed creature and serve as advisers to Pleroma. Abraxas has a head that resembles some type of bird, with hair made of fire, and Gnosis's head resembles a metallic robot, with hair made of wires...

• [http://digital.library.unt.edu/ark:/67531/metadc33228/](http://digital.library.unt.edu/ark:/67531/metadc33228/)

Fig. 1. Appearance of Pleroma.
Why ETDs are Different from TDs

• ETDs present us with new opportunities and challenges that we never considered with print theses and dissertations
• We can potentially preserve far more sophisticated content types, ancillary materials (datasets, digital media, executables, etc.)
• However, this is a much more complex prospect than traditional TD library storage
• Many of the same issues as the broader Research Data Management challenge
Field Needs Identified by NDLTD in 2011

- Many smaller academic institutions still did not have a basic framework for implementing ETD programs & services
- Further, many institutions were struggling to understand ETD programs in terms of long term lifecycle management concepts
- Widespread need for documents and toolkits to assist in developing ETD programs
Collaborative IMLS-funded Projects 2011-2017

- Lifecycle Management of ETDs
- DataRes & iCAMP projects
- ETDplus project
Some Key Challenges of all ETD Programs

• How will institutions address the entire life cycle of ETDs?
• Can we ensure that ETDs acquired from students today will be available to future researchers? In 10 years? In a century?
• How will libraries identify and institutionalize the best long-term curatorial practices for this important genre of digital content?
IMLS ETD Lifecycle Management Project 2011-2014

• Funded by $268K award from IMLS

• General aim of the project was to analyze, document, and address the lifecycle management challenges presented by ETDs to ensure that colleges and universities have the requisite knowledge to properly curate these new collections permanently

• The project team engaged in spinoff activities not originally foreseen in the project (NDLTD international ETD survey)
IMLS ETD Lifecycle Management Project - Partners

1. Networked Digital Library of Theses and Dissertations (NDLTD)
2. Educopia Institute/MetaArchive Cooperative
3. University of North Texas
4. Virginia Tech
5. Rice University
6. Boston College
7. Indiana State University
8. Pennsylvania State University
9. University of Arizona
IMLS ETD Lifecycle Management Project - Goals

A. Dissemination of Guidance Documents for Lifecycle Management of ETDs
B. Production of ETD Lifecycle Management Tools
C. Creation of Educational Materials and Associated Workshop
2014 Guidance Documents for Lifecycle Management of ETDs

1. Briefing on Access Levels and Embargoes of ETDs
2. Briefing on ETD Copyright Issues and Fair Use
3. Guidelines for Implementing ETD Programs - Roles & Responsibilities
4. Guidelines for Collecting Usage Metrics & Demonstrations of Value for ETD Programs
5. Overview of Formats, Complex Content Objects, and Format Migration Scenarios for ETDs
6. Overview of ETD Metadata & Lifecycle Event Record-Keeping for ETDs
8. Guide to Options for ETD Programs

Other Project Outputs

ETD Lifecycle Management Tools:
- PREMIS Event Service
- ETD Drop
- Lifecycle Management Tools Manual

ETD Lifecycle Management Workshop:
- Instructor Manual, Example Slides, Example Agenda,
  Module 1 Handout, Module 2 Handout, Module 3 Handout,
  Pre-workshop Survey, Post-workshop Survey, Example Syllabus

NDLTD International Survey of ETD Practices
2013 NDLTD International Survey of ETD Practices – Selected Findings

• 161 institutions responded (132 US, 29 International)
• Confirmed predominance of PDFs versus other formats
• Indicated significant increase in preservation planning compared to 2008
• US institutions were much more likely to allow access restrictions at the request of ETD authors
• Average inst. annual growth was in the hundreds of ETDs

http://hdl.handle.net/10919/50978
Comment: How Many Institutions Pictured ETD Intake Workflows

https://media.giphy.com/media/cmx1gjLskcGWytbLm6/giphy.mp4
What It Was More Like In Practice

https://media.giphy.com/media/l3q2ymonjpIgGTmWQ/giphy.mp4
DataRes & iCamp Projects 2011-2013

• The DataRes Project was funded by a 2011 grant of US$ 226,786 from the Institute of Museum and Library Services 21st Century Librarian (21CL) program.

• Goal: investigate the current status of research data management in universities and how the library and information science (LIS) profession can best respond to emerging needs of.

• DataRes is a collaboration between the University of North Texas Libraries, the UNT College of Information, and the Council on Library and Information Resources.

• Paired with the iCAMP curriculum redesign project, another IMLS 21CL 2011 grant of US$ 624,663 to UNT to assess educational needs and develop new shared curricula to train new LIS professionals seeking to fill data management positions.
Methodologies Used

DataRes Project

- Surveys of institutional data management policies, and views of individuals concerning research data management
- Textual analysis of agency requirements and institutional data management policies
- Focus groups of agency officials, university admins, and librarians
- Engaging community experts in producing a peer-reviewed Council on Library and Information Resources (CLIR) monograph discussing the future of research data management

iCAMP Project

- Competency analysis of advertised jobs in data management
- Curriculum redesign based on competency analysis
Notable Findings: Institutional Policy Scan of 230 Universities

Policy Scan Results

- 82% Institutional-Level Policy
- 18% No Institutional-Level Policy
Text Analysis: Institutional Policy Scan

- 38 institutional policies identified and analyzed
- Policy language was often weakly assertive, example:

“The University recognizes the importance of data sharing in the advancement of knowledge and education.”

(University of New Hampshire “UNH.VII.C.9”)
Notable Findings: Survey of Individuals

Does your institution have a policy governing the retention and sharing of research data? (231 responses)

- No: 72%
- Yes: 9%
- Don't Know: 19%
Textual and Focus Groups Analysis of Agency Requirements

• Analyzed NIH, NEH, and NSF data management requirements
• There was significant variation across the different agencies in terms of emphasis and assertions
• Both disciplinary foci and historical level of emphasis on data sharing in the agencies was apparent
• Focus groups revealed a high degree of skepticism among many research communities that requirements would be enforced
The Denton Declaration: An Open Data Manifesto

http://openaccess.unt.edu/denton_declaration

- Developed by national gathering of university administrators, technologists, librarians, researchers, and other stakeholders gathered to discuss and articulate best practices and emerging trends in research data management
- Declaration is a statement of prescriptive assertions and values concerning research data management
- Bridges the converging interests of these stakeholders and promotes collaboration, transparency, and accountability across organizational and disciplinary boundaries
Synergistic iCAMP Project Findings

• 110 job advertisements were collected for analysis between October 2011 and March 2012
• Used to identify new competencies required for data management jobs
• New core curriculum of 4 courses designed and being taught for the first time in 2013
• Utilizing a “teaching library” model for practical training
Key Findings

• Major disconnect between assertions of the importance of research data management (by both agencies and individuals) and actual practice

• Much more prescriptive guidance and requirements will be necessary to actually encourage disciplines to take RDM requirements seriously in evaluating funding applications

http://www.clir.org/pubs/reports/pub160
Key Findings (cont.)

• New competencies being requested for contemporary data management jobs are indeed significantly different from past.

• While curricula can be redesigned, it is exceedingly difficult to find and recruit qualified instructors for updated curricula.
Key Findings (cont.)

- Institution-level policies are driven by practice, not the other way around
- Research Data Management is not a single department issue, nor is it the purview of a single discipline
- Collaboration, domain knowledge, and infrastructure are all key to the success of any RDM response
Comment: How Librarians Probably Pictured Research Data Management

https://gph.is/1hehotm
What It Was More Like In Practice

https://gph.is/2fEvagf
ETDplus Project

How can institutions best ensure the longevity and availability of ETD research data and complex digital objects (e.g., software, multimedia files) that comprise an integral component of student theses and dissertations?
ETDplus Project 2014-2017

• Built on the momentum of the earlier “Lifecycle Management of ETDs” project to research and build tools to help manage a growing challenge in ETD programs: the creation and submission of materials beyond the PDF of an electronic thesis or dissertation.

• Ranging from research data sets to video installations, from websites to music recitals, these digital objects are pieces of intellectual work that cannot be captured in words alone.

• The project has produced guidance documentation, workshop materials, and software tools for students and staff to use in managing these complex digital objects.
# ETDplus Project Group

1. NDLTD
2. Educopia Institute
3. MetaArchive Cooperative
4. ProQuest
5. Carnegie Mellon University
6. Colorado State University
7. HBCU Library Alliance
8. Indiana State University
9. Oregon State University
10. Penn State University
11. Purdue University
12. University of Louisville
13. UNC School of Library and Information Science
14. University of North Texas
15. University of Tennessee Knoxville
16. Virginia Tech University
Deployed surveys at beginning of project to gather information regarding current stakeholder community needs:

- What research outputs are students creating as part of their thesis/dissertation research process?
- Which of these research outputs do the students consider valuable or essential for understanding and building upon their findings?
- Which of these research outputs are they currently planning to or able to submit as part of their theses/dissertations packages?
- What are some of the common barriers institutions report in accepting complex (non-PDF) submissions of theses and dissertations?
Surveys (cont.)

- Two surveys: 1) Graduate students; 2) Institutional ETD program staff
- 12 universities took part (w/ 12 IRBs...)
- March-May 2015
- Good response: 795 total graduate student responses
Types of Information/Materials Generated during Research Process as Reported by Graduate Students (respondents N=795)
Question to Graduate Students: What is the most valuable part of your thesis or dissertation research for you?

- 62% of respondents found the ETD Text to be the most important content.
- 38% found Other Content to be the most important content.

Responses:
- ETD Text: 62%
- Other Content: 38%
The ETD Preservation Gap

- Although more than a third of students said that the materials beyond their PDF are the most important, only 100 of the 795 surveyed students (13%) reported plans to actually submit those materials.
- An additional 521 (66%) specifically report that they will not submit materials beyond the PDF, and 174 (22%) reported that this question did not apply to their work.
- Students’ perceptions of importance, in other words, seem not to be the key drivers for submitting their research outputs as part of their thesis/dissertation packages.
Survey of 65 administrators and ETD staff at the same 12 institutions asked about submission policies concerning non-pdf objects.

Does your institution accept objects in addition to the PDF of the thesis or dissertation?

Yes
No
Don't know
Prefer not to answer
What resources students would find most helpful in five curation tasks

• When we asked what resources students would find most helpful in five curation tasks (Versioning, Storage, Metadata, IP, and File Formats), students typically preferred **software** and **documentation** over webinars and workshops. Even “no help” was preferable to **webinars** and **workshops**(!).

• We will have to meet students where they are—it’s a safe assumption that students prefer resources that work on their own schedules. The deliverables of the project that are aimed towards students needed to be workable without a lot of in-person guidance—something the students are likely to avoid.
Project Outputs

- ETDplus Guidance Briefs, published 28 October 2017
- ETDplus Virtual Workshops, published 28 November 2017
- Introduction to ETDplus Toolkit, published 12 June 2017
- ETDplus Toolkit, published 03 April 2017
- ETDplus Workbench, published 25 July 2016

Project Outputs - Foundation

Questions we asked ourselves:

• What do students need to know about digital content management as pertains to their research outputs?

• How might the ETD—as a common rite of passage in research careers—be used to help students learn how to structure, share, and manage their digital content appropriately?

• What kinds of adaptable content could we provide institutions to use for student self-preparation?
ETDplus Guidance Briefs

Preserving & Curating ETD Research Data & Complex Digital Objects

The Guidance Briefs have been authored by the ETDplus project team as short (3-4 page) "how-to" oriented briefs that will help ETD/IR programs build and nurture supportive relationships with student researchers.

These briefs will be released as open documents that colleges and universities can adopt and adapt to assist their own student researchers in understanding how their approaches to data and content management impact credibility, replicable research, and general long-term accessibility: knowledge and skills that will impact the health of their careers for years to come.

Interested ETD stakeholders can download and evaluate copies of the Guidance Briefs using the links on this page. We invite your feedback! Please draw our project team's attention to any components that need to be edited, revised, broadened, or narrowed.

We will gladly accept comments between May 3 and June 30, 2016. Please send an email with your suggestions to Courtney Vukasinovic and/or track your changes within the documents and email those back to us. We will integrate the community's feedback before formally issuing these Briefs later this summer.

The project is generously funded by the Institute of Museum and Library Services (IMLS) and led by the Educoopia Institute, in collaboration with the NDLTD, HBCU Alliance, bepress, ProQuest, and the libraries of Carnegie Mellon, Indiana State, Morehouse, Oregon State, Penn State, Purdue, University of Louisville, University of North Texas, and Virginia Tech.

Preserving & Curating ETD Research Data & Complex Digital Objects by Educoopia Institute and the ETDplus Project Team is licensed under a Creative Commons Attribution 4.0 International License.
File Formats in brief

There is no perfect file format. Each will have advantages and disadvantages depending on your research uses. Select a file format, or set of file formats, that helps you complete your research now, and that you can access again in the future. This is true both for your research outputs (what you create) and your research inputs (materials you use in the research process).

Common file types include:
- Images: jpg, gif, tiff, png, ai, svg, ...
- Video: mpeg, m2ts, flv, dv, ...
- GIS: kml, dxf, shp, tiff, ...
- CAD: dxf, dwg, pdf, …
- Data: csv, mdf, fp, spv, xlx, tsv, ...
- Text: txt, rtf, tvi, doc, pdf...

How to select file formats:
- Use software that imports and exports data in common and non-proprietary formats
- Consult with advisors and colleagues
- Convert files from proprietary to non-proprietary formats (e.g., .doc to .txt and/or .pdf)
- Choose a format with functions that support your research needs
- Save final versions of your content in multiple formats in order to spread your risk across multiple software platforms (e.g., docx, pdf, and txt; or mp4, avi, and mpg)

Many ETD programs favor pdf files. If you export your research outputs to pdf, make sure that you:
1. Embed your fonts
2. Embed (and test!) hyperlinks
3. Archive web-based resources and citations (using a tool like Robust Links, Archive-It, or PermaCC)
4. Store supplementary materials as separate files

Before you undertake any conversion, you need to identify what characteristics of your data are important to maintain during the conversion. For example, are the colors in a document or image important? Is the pagination essential? What about references? You will want to test these after your conversion is complete to ensure that you have a conversion that will meet your needs.

Additional Resources:
- List of File Formats (Wikipedia)
- Sustainability of Digital Formats (Library of Congress)
- Evaluating Your File Formats (UK National Archives)
- Reformatting Guides (US National Archives)

What will you do if you no longer can use the software you create your research files in – either because you no longer can afford the software, or the publisher goes out of business, or the latest version is not backwards compatible? Plan for these possibilities by saving your final research files in multiple formats – including a non-proprietary format.

When using website-based materials as evidence or references, take precautions to ensure that if the content moves, changes, or disappears, you still have evidence of its existence. Current tools to help you ensure the longevity of these materials include Robust Links, PermaCC, and Archive-It. You can also take screenshots of important digital content in order to preserve the look and feel of an object.

Source - Guidance Briefs: Managing Your ETD Research Files
Metadata in brief

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. It can be especially useful in providing describing context for the research files that may accompany your dissertation.

For every research file you create, you should also produce metadata describing:
- **Who** created the content
- **What** is the content
- **When** was the content created
- **Where** is it geographically
- **How** was it developed
- **Why** was it developed

**What is a Metadata Standard?**

Metadata standards provide a structure for consistent (predictable) information. They define the structure and categories of information (e.g., “title,” “author,” “date”) and provide controlled vocabulary to enable interpretation across a discipline. Metadata standards foster uniformity, which permits search/retrieval systems to identify and share the content metadata describes.

**ETD metadata tips:**

1. Your abstract needs to include a clear description and keywords relevant to your work, including any research files that accompany your dissertation.
2. Be careful with over-reliance on spell-check functions. For example, Microsoft Office does not spell-check capital letters, which can impact chart or graph 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.

**Typical metadata requested about a pdf during the ETD submission process:**
- Title
- Author/Creator
- Advisor
- Resource Type
- Date
- Language
- Abstract
- Subject
- Identifier
- Degree Information
- Rights information

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

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 programs, etc.). 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. Submit this inventory spreadsheet as part of your ETD package.

Source - Guidance Briefs: Managing Your ETD Research Files
Data Structures in brief

Structuring your data well enables you to:
- Reproduce results
- Reuse it in the future
- Share it with others
- Gain and retain credibility
- Comply with IRB/funder requirements

The decisions you make about how you organize and structure your data today will have implications for how you and others can access and make use (or sense!) of that data in the future.

Data Organization Principles:
1. Use one variable per column
2. Make one observation per row
3. Include one kind of data per column
4. Use human-readable column name
5. Use an ID or key to indicate the relationship between multiple tables (If you apply this principle, you should be using a Relational Database)
6. Include a readme text file detailing why the data has been collected, and what files comprise your data package.

Do:
- Consider what your NULL values are and how they are represented
- Use standard data representation (e.g., YYYYMMDD for dates)
- Use consistent capitalization

Do Not:
- Use formatting to convey information
- Include units in cells along with the data value
- Place comments in cells
- Use special characters in field names
- Use blank spaces or symbols in column names

Questions to consider for any data project:
1. What are your field’s (or funding agency’s) data structure standards and requirements?
2. What are your university’s policies relating to your data
3. What are your data export options?
4. What forms of the data will be needed for future access?

As a first step in your research, create a "Data Management Plan" that documents your practices for collecting, organizing, backing up, and storing any data you generate. This will help you think through ways of structuring your data that increase its long-term accessibility and use.

Discipline-based data repository examples:
- Social Sciences: ICPSR
- Genomics: GenBank
- Earth Sciences: NASA’s Earthdata
- Archaeology: IDAR
- Oceanography: NODC
- BioSciences: Dryad

Source - Guidance Briefs: Managing Your ETD Research Files
Storage in brief

**Back-up:** A copy of your digital content, ideally stored in a different location from the original, usually made to prevent data loss.

**Preservation:** The “series of managed activities necessary to ensure continued access to digital materials for as long as necessary”. –*Digital Preservation Coalition*

Where and how you choose to store your research materials and writings will determine how long they survive. To mitigate against loss, make your own back-ups on a regular, formalized schedule (e.g. daily or weekly).

**Threats to storage environments:**
- Natural disaster
- Human error
- Human malice
- Drive failure
- Format obsolescence
- Media obsolescence
- Bit rot
- Business failure
- Software or hardware error

**Basic recommendations:**
1. Maintain at least one local (i.e., non-cloud-based) copy of your content
2. Maintain at least three separate complete copies of your research content
3. Maintain at least one copy in a different geographic location
4. Maintain a history of changes in at least one location (e.g., using a “Time Capsule” software package to automatically back up your content without deleting older copies)
5. Document in a text file how, when, and where you store and back up your materials
6. Systematize your folder- and file-name conventions using human-identifiable information
7. Use naming conventions to mark versions of files, e.g., using consecutive numbers to track a file through all edits and revisions that take place to it. (e.g., filename-v12.txt)
8. Make sure your filenames are followed by the correct file extension (e.g., .txt, .csv)
9. Avoid using special characters in all file and folder names (e.g., \?.*<>}{][&$,;!)
10. Document the formats you are managing and the potential sustainability issues
11. Save a copy of your research files in non-proprietary formats, so that you don’t need a software license to render and use them.

**Advanced recommendations:**
1. Produce and maintain an inventory of all of your content, documenting file names, sizes, locations, and types
2. Create and regularly check “checksums” or digital signatures for your most important research files. Checksums can be generated by several open source tools and utilities and they can be stored in your inventory.
3. Monitor your content to ensure missing, moved, and renamed files are automatically brought to your attention. A tool like “Fixity” can scan specified folders or directories on a regular basis and report changes to you via email.

**Resources**
- For “back-up” advice, see Jesus Vigo, *Best Practices to Back up Your Data*
- For more on cloud-based backups, please see Charles Beagrie Ltd., *How Cloud Storage can address the need of public archives in the UK*
- For general information, see also *Personal Digital Archiving*

Source - Guidance Briefs: Managing Your ETD Research Files
Version Control

**Version Control:** The process of managing changes to your files over time (aka, revision control or source control)

**Manual Version Control**
A simple method to store the current revision is at the end of the file name. This way, files can be grouped by their names and sorted by version number:
- `filename-v01.jpg`
- `filename-v02.jpg`
- ...

You can also use dates to designate version numbers, using year-month-day (20150930) to help your computer sort versions in chronological order:
- `filename-20160402.jpg`
- `filename-20160407.jpg`
- ...

If the files you are using are created or edited collaboratively, incorporate names or initials so you know who updated which version:
- `filename-20160402-KES.jpg`
- `filename-20160407-WTC.jpg`
- ...

**Software-Assisted Version Control**
There are also software tools that can help you version your content. These tools store your content in such a way that they can remember its state from revision to revision. Usually, they also allow you to “check in” and “check out” your content, ensuring that revisions never happen simultaneously in two different locations (e.g., if collaborating researchers both attempt to revise the same file at the same time, or a researcher unwittingly tries to revise the same file on two different machines). Key differences between these software-assisted methods and the manual methods include:

1. You can only view and edit the working version of a file
2. When you change a file, you can save a revision and attach a short summary of your changes.

Research is active and iterative. You will edit and re-edit your research materials many times before finishing your thesis or dissertation. How will you know that you are working with the most current revision of your materials?

**Resources (For more information)**
- The digital humanities center MATRIX (Michigan State University) provides advice on how to structure file names based on oral history projects that is broadly applicable: [http://ohda.matrix.msu.edu/2012/08/file-naming-in-the-digital-age](http://ohda.matrix.msu.edu/2012/08/file-naming-in-the-digital-age)
- Udacity offers a free online course on how to use Git and GitHub with interactive exercises to familiarize you with using the tools: [https://www.udacity.com/course/how-to-use-git-and-github--ud775](https://www.udacity.com/course/how-to-use-git-and-github--ud775)
- Another helpful GitHub guide is available from Hello World: [https://guides.github.com/activities/hello-world/](https://guides.github.com/activities/hello-world/)
- The Subversion community provides free access to the book Version Control with Subversion: [http://svnbook.red-bean.com/](http://svnbook.red-bean.com/)
Copyright in brief

US Copyright: “that body of exclusive rights granted by law to copyright owners for protection of their work.”
http://www.copyright.gov/help/faq/definitions.html

Copyright vs. Patents: “Copyright protects original works of authorship, while a patent protects inventions or discoveries.
http://www.copyright.gov/help/faq/faq-general.html

If you are using a work that is within copyright, but meets certain “fair use” criteria, courts have found that no formal permission is needed. The criteria that are taken into account include the purpose (e.g., educational and research uses favor fair use while commercial uses do not); the type (e.g., factual or nonfiction-based information may favor fair use; highly creative work likely will not); the amount (e.g., small quantities vs. a significant portion of the original work); and the effect (e.g., not having a negative impact on the copyright holder).

Giving credit is no substitute for asking permission!
Creative Commons (recommended)

Fair Use and Public Domain Resources
- Cornell University, Copyright Term/Public Domain in the United States
- CMSI Code of Best Practices in Fair Use for Scholarly Research in Communication
- CAA Code of Best Practices in Fair Use for the Visual Arts
- Columbia University Fair Use Checklist

What can copyright protect?
1. literary works
2. musical works, including any accompanying words
3. dramatic works, including any accompanying music
4. pantomimes and choreographic works
5. pictorial, graphic, and sculptural works
6. motion pictures and other audiovisual works
7. sound recordings
8. architectural works

Do I need a patent?
Universities often have designated offices to deal with questions arising about new inventions or innovations. These questions involve the policies of the university around ownership and IP, and understanding your own institutions’ policies is a must. Examples include
- Stanford University’s Office of Technology Licensing
- Columbia University Tech Ventures

More: https://creativecommons.org/

Source - Guidance Briefs: Managing Your ETD Research Files
Electronic Theses and Dissertations (ETD) plus Toolkit

• The ETD+ Toolkit is an approach to improving student and faculty research output management.

• Focusing on the Electronic Thesis and Dissertation (ETD) as a milestone in a student’s research trajectory, it provides practical advice to students and faculty about avoiding common digital loss scenarios for the ETD and all of its affiliated files.

• The ETD+ Toolkit provides free introductory self-paced training modules on crucial data curation and digital longevity techniques.
Electronic Theses and Dissertations (ETD) plus Virtual Workshops

• The ETD+ Virtual Workshop Series, taught by Dr. Katherine Skinner, is a set of free introductory training resources on crucial data curation and digital longevity techniques.

• Focusing on the Electronic Thesis and Dissertation (ETD) as a mile-marker in a student’s research trajectory, it provides in-time advice to students and faculty about avoiding common digital loss scenarios for the ETD and all of its affiliated files.
ETDplus Workbench

• A web based tool designed to assist students in preparing and packaging ETD supplementary materials for long-term preservation and access

• Includes configurable functions that integrate basic preservation actions such as virus scans, integrity checks, file format identification and validation, personally identifiable information scans, and metadata and versioning support into a simple data upload and review workflow

• The tool also packages uploaded data and metadata as Bags that users can download and ingest into an array of repository and storage environments

• Designed as a reference implementation of reusable Hydra-Sufia code for system geeks to tinker with in their own local environments
ETDs and Broader Research Data Management Issues

APLU/AAU/NSF Workshop Initiative
As much as we’d like to train researchers to properly manage their data...

http://gph.is/1EBMp3m
They often just do what makes sense to them

http://gph.is/1c7xeFt
It’s Hard to Get Scholars Organized

• They have their own agendas, which are largely about moving on to their next research projects in order to continue accumulating academic renown / credit

• They don’t get incentivized to preserve and organize their outputs in the same way librarians are motivated

http://gph.is/1Lil1Qp
We can’t herd them, but we know they can be motivated to take action

http://gph.is/XMgQDx
Purpose was to provide a venue for learning, sharing, and planning to support research universities as they implement systems for public access to research data

Foster cross-institutional collaboration that yields alternative models to publicly sharing data, reduces total effort of developing public access to data across the system of research universities, and builds consensus on key system elements that foster effective storage and sharing of data in ways that are findable, accessible, interoperable and reusable (FAIR)

Federal Agency Officials Encouraged Universities to Make Research Data Publicly Accessible

• Presentations by officials from NSF, DOE, NIH, and other funding agencies all emphasized the importance of this point
• This urgency was impressed on university administrators

https://gph.is/2x9gahj
UNCG Workshop Team Got Along Well!

- Comprised of senior campus officials from Office of Research, IT services, Library (me), Interdisciplinary Center for Community Engagement, and County Data Manager
- Generated a remarkable (unprecedented?) synergy and interest in collaboration
- Initiated six months of planning for a cross-campus open data initiative, together with the county

https://gph.is/1xJoMWM
Future Directions

Where do we go from here?
Maybe we can get alignment on these issues after all, if we reorient our efforts on what gets the attention of academics

http://gph.is/PfT3JV
What got the attention of the UNCG Administration

• Possibilities for attracting more research dollars and fostering better collaborative research efforts on campus and externally
• Opportunity to build new kinds of partnerships with the community
• Prospect of realizing greater impacts of university research activities
• Potential to increase visibility of research and how it benefits the public
If we work with their motivations, we may be able to align and synergize with researchers and administrators after all.

http://gph.is/1szaw3Y