

Updating Metadata Guidelines to Improve ETD Discoverability and Use

USETDA

September 30, 2015 || Austin, Texas

This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.



Working Group Members

- ★ Santi Thompson (University of Houston)
- ★ Monica Rivero (Rice University)
- ★ Sarah Potvin (Texas A&M University) - Chair
- ★ Kristi Park (Texas Digital Library)
- ★ Colleen Lyon (University of Texas) - Liaison to Vireo User Group
- ★ Kara Long (Baylor University)

What You'll Hear About Today

Some background on the effort: who/what/where/why

Discussion of case studies

Recommendations of changes to Vireo

How these guidelines diverge from 2008 standard

Next steps

Discussion

What's metadata?

Playing soldiers: martial subjects in early modern English drama, 1590-1660	
Show simple record	
dc.contributor.advisor	Whigham, Leslie
dc.creator	Pasupathi, Vimala Cleamona
dc.date.accessioned	2008-08-28T22:30:01Z
dc.date.available	2008-08-28T22:30:01Z
dc.date.created	2005
dc.date.issued	2008-08-28T22:30:01Z
dc.identifier.uri	http://hdl.handle.net/2152/2038
dc.description.abstract	Not available
dc.format.medium	electronic
dc.language.iso	eng
dc.rights	Copyright © is held by the author. Presentation of this material on the Libraries' web site by University Libraries, The University of Texas at Austin was made possible under a limited license grant from the author who has retained all copyrights in the works.
dc.subject.lcsh	SOLDIERS IN THE ARMY
dc.subject.lcsh	ARMY OF ENGLAND--17TH CENTURY
dc.title	Playing soldiers: martial subjects in early modern English drama, 1590-1660
dc.description.department	English
dc.identifier.oclc	68668387

Advisor
Author

Dates

Rights information
Subject headings

Title

PLAYING SOLDIERS:
MARTIAL SUBJECTS IN EARLY MODERN ENGLISH DRAMA, 1590-1660

by

Vimala Cleamona Pasupathi, B.A.; M.A.

Dissertation

Presented to the Faculty of the Graduate School of
the University of Texas at Austin
in Partial Fulfillment
of the requirements
for the Degree of
Doctor of Philosophy

The University of Texas at Austin

August 2005

Why is it important?

Facilitates discovery.

What is this thesis about? Find everything about “soldiers in literature.”

Helps with management and tracking of records.

Who owns the rights to this work? Can we make this ETD publicly accessible?

Clarifies relationships among objects.

What else has this author published? How do I know that this John Smith is the same person as that John Smith?

Why ETDs?

“Electronic Theses and Dissertations (ETDs) represent an important, yet untapped, resource for our State. The collaborating institutions should work together to make ETDs be the first federated digital library collection hosted by the Texas Digital Library.”

-John Leggett, Mark McFarland, and Drew Racine, “The Texas Digital Library: A Business Case,” 2005, rev. 2006.

2008 Guidelines

The 2008 group's charge:
To "support TDL's goal of a single, unified collection of ETDs" by working to "identify the issues and policies involved with ETDs in the member institutions, and make recommendations to the team charged with the development of an ingestion application [Vireo]."



Texas Digital Library Descriptive Metadata Guidelines for Electronic Theses and Dissertations

Version 1.0

June 2008

Prepared by the
TDL Metadata Working Group:
Amy Rushing (The University of Texas at Austin) - Chair
Jay Koenig (Texas A&M University)
Anne Mitchell (University of Houston)
William Moen (University of North Texas)
Tim Strawn (The University of Texas at Austin)
Jason Thomale (Texas Tech University)

Federated Electronic Theses and Dissertations

BROWSE BY

[By Issue Date](#) [Authors](#) [Titles](#) [Subjects](#)

Search within this community and its collections:

The TDL Federated ETD Collection contains records of electronic theses and dissertations (ETDs) submitted by participating member institutions, making these otherwise hard-to-find scholarly works available to a much wider audience.

Collections in this community

[Angelo State University](#)

[Baylor University](#)

[Texas A&M University at College Station](#)

[Texas A&M University at Qatar](#)

[Texas A&M University-Corpus Christi](#)

[Texas A&M University Galveston - Galveston Historic Beach Profiles](#)

[Galveston Historic Beach Profiles](#)

[Texas Tech University](#)

[University of Houston](#)

[University of Texas at Arlington](#)

[University of Texas at Austin](#)

[University of Texas at Brownsville](#)

[University of Texas Medical Branch](#)

[University of Texas Southwestern Medical Center](#)

“A single, unified collection of ETDs...”

Welcome to the Thesis & Dissertation Submission System

Once you have passed the final defense and satisfied the requirements of the committee, you are ready to submit your manuscript to the Thesis Office for review. This submission process is fully electronic, and is made through an online application developed and maintained by the Texas Digital Library, in conjunction with the Texas A&M, MIT, and UIUC.

To get started with your submission, click the link below. You will be asked to authenticate using your NetID:

[Start your submission](#)



“...make recommendations [for an] ingestion application...”



VIREO



HARVARD
LIBRARY



Why update?

Inconsistent application

New use cases

Vireo development

Need for systems & standards to work together

Our Charge

The TDL ETD Metadata Working Group will provide guidance to TDL member institutions and other ETD practitioners on metadata for electronic theses and dissertations, with a particular focus on works published through the Vireo ETD submission and management application.

(1) Update the “**Descriptive Metadata Guidelines for Electronic Theses and Dissertations (Version 1)**” to reflect current and emergent practices;

(2) Evaluate the recommended metadata guidelines for **divergences with current Vireo functionality** and make recommendations to the Vireo Users Group for correcting those divergences;

(3) Publish and **promote use of the updated guidelines** among TDL members and Vireo users.

All work of the group will be done in close cooperation with the Vireo Users Group and TDL staff.

Challenges of updating a standard; Or: “the art of compromise”

- This effort faced 2 overarching challenges:
 - balancing optimal solutions with ease of adoption;
 - weighing multiple existing standards: current Vireo output (a de facto standard); 2008 ETD Metadata Guidelines; NDLTD requirements
- * **There is no one solution that achieves this balance. Throughout, we have documented our decision-making processes, so that future updates can more easily decipher the rationale(s) behind decisions. ***



“The standards specified in standards document (formal standards) and the standards used in real practice (de facto standards) are two different things.”

■ Daniel Pargman and Jacob Palme, “ASCII Imperialism,” in *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life*, ed. by Martha Lampland and Susan Leigh Star (Ithaca: Cornell University Press, 2009).

Case Studies

- Methodology for identifying and constructing case studies
- What were the problem areas, and how do the Guidelines address them?
 - Access/availability metadata
 - Proliferation of dates
 - Crosswalking issues
 - Name disambiguation
 - Complex objects / supplementary files

Methodology

Stage 1: Data gathering & identification of case studies

Began by gathering information through informal interviews, literature reviews, and reviews of existing standards.

Stage 2: Close investigation of case studies, drafting of recommendations

Once emergent case studies were identified, working group divided into pairs to draft recommendations. We relied on environmental scans, research into existing literature, and consultation with experts.

Stage 3 [actually iterative]: Feedback & finalization

In an attempt to vet and confirm our recommendations and process, we presented at Texas / US Region 3 ETD Association, formed and sought feedback from Task & Review group, and engaged with the Vireo SG.

Access & Availability

Problem:

Inconsistent recording of access, rights, and availability metadata at the item-level.

Objective:

To clearly communicate the rights and re-use status of the ETD, the availability of the ETD, and whether a Creative Commons license applies to the ETD.

Access & Availability

Recommendations:

- Rights and re-use of work:
 - dc.rights and dc.rights.rightsHolder
- Availability of the document:
 - dc.rights.accessRights
- Regarding licensing:
 - dc.rights and dc.rights.uri
- Vireo:
 - Provide Creative Commons option

Dates

Problem(s):

Date information is being exported to many different fields. What does it all mean?!

Objective:

Account for all dates generated by the user, ETD administrators, or the systems that manage ETDs, and determine which are the “important” dates.

Dates

Recommendations:

- Graduation Date and Date Made Public
 - most important!
 - dc.date.issued || dc.date.created
 - dc.date.available
- Explicit encoding of other meaningful dates
 - dc.description.provenance

Crosswalking

Problem:

Inconsistencies in mapping ETD metadata resulting in poor record quality and data loss.

Records harvested *without* author information and degree information.

Objective:

Provide reliable, quality ETD metadata accessible through OAI harvesting.

Crosswalking

Recommendations:

Update the crosswalk, so that OAI-ETD and the NDLTD's standard, ETD-MS v1.1 are in better alignment.

Name Disambiguation

Why is this important?

Name disambiguation is vitally important for search and discovery of ETDs, and for the management of a scholar's academic reputation and body of work.

Columbus, Christopher

Columbus, Christopher, 1958-

Name Disambiguation

The Motivation:

There is diversity of local approaches. None of the challenges of name disambiguation were addressed in the 2008 guidelines.

How can Vireo help support this very important work?

Name Disambiguation

More Names More Problems:

- Ambiguous search results
- Exposure of birth years - not for everyone
- Not enough information from students

Name Disambiguation

Recommendations:

Use an authorized form of a personal name whenever possible.

- Local name files & campus directories
- Unique information
- ORCiDs or other unambiguous identifiers

Complex objects and supplemental files

Why is this important?

Our students are going beyond the pdf, but this work is not always being captured.

- data sets
- visualizations
- multimedia objects

Complex objects and supplemental files

The Motivation:

There is greater diversity in the kinds of digital objects being created in fulfillment of graduate degrees.

How can Vireo help identify and describe this work?

Complex objects and supplemental files

Recommendations:

- Identify ETDs with supplementary or complex files
 - Assign file types - use a file characterization to analyze uploaded files
- Provide context
 - Support inclusion of readme files to interpret supplementary files and publish to repository
- Allow authors to reference supplementary and complex files in the metadata that cannot be a part of the ETD submission
 - Allow for links to material deposited elsewhere
 - Figshare, Dataverse or other data repositories

Vireo recommendations

Name disambiguation

- Integrate a locally controlled authority file
- Export ORCID data using `dc.identifier.orcid`

Availability & access metadata

- Automatically populate `dc.rights` field - customize at institution level
- Automatically populate `dc.rights.accessRights` to indicate availability
- Allow students to select CC licenses
- Embargo lift date mapped to `dc.rights.accessRights`

Vireo recommendations

Complex objects & supplemental files

- Automatically analyze uploaded files and assign either `dc.type.genre` value or `dc.type.dcmi` value
- Prompt and enable the inclusion of a readme file as a supplement to ETD
- Automatically generate text that identifies supplemental files
- Allow for the inclusion of links to data in data repositories

Vireo recommendations

Dates

- Discontinue mapping of approval date to dc.date.issued.
- Discontinue use of dc.date.submitted
- Graduation date should be mapped to dc.date.issued in YYYY-MM format. May also be included in natural language in dc.date.created
- Graduation date should map to dc.description.provenance with standard language (e.g. FirstName LastName graduated on YYYY-MM)
- Continue and extend practice of including information in dc.description.provenance

Vireo recommendations

Collection and export of degree information

- Program, department, college, and major
- Option of enabling authority control for these fields

Over the longer term

- Take steps towards enabling linked data by building out Vireo to dynamically point to externally-hosted vocabularies

Vireo recommendations

Type

- “Text” should be mapped to `dc.type.dcmi` instead of `dc.type.material`
- “Thesis” should be mapped to `dc.type.genre` per 2008 standard

What's next with Vireo?

Vireo Steering Group

Development cycles

Divergence from 2008 standard

- Schema based on qualified Dublin Core rather than MODS
- MODS -> DC switch means some element names changed
 - Origin Information -> Date
 - Physical Description -> Format
 - Location (URL) -> Identifier
 - Record Information -> removed because info is captured automatically by repository system
- New elements
 - Rights metadata
 - Author identifiers (ORCID)
 - Description (non-abstract)
- Mappings to OAI-ETD and Google Scholar's Highwire Press tags

Read all about it!!

TDL Repository: <http://hdl.handle.net/2249.1/68437>



Next Steps

- Tips for aligning legacy data with updated standard
- Emerging case studies:
 - Linked Data
 - Fedora
 - Preservation
 - Beyond the Union Catalog
 - Beyond the PDF

Task & Review Group Members

- Daniel Alemneh (University of North Texas)
- Shelley Barba (Texas Tech University)
- Christine Brown (Texas A&M University)
- Melanie Cofield (University of Texas)
- James Creel (Texas A&M University)
- Thomas Dowling (Wake Forest University)
- Jeremy Huff (Texas A&M University)
- Sevim McCutcheon (Kent State University)
- Billie Peterson-Lugo (Baylor University)
- Amy Rushing (University of Texas at San Antonio)
- Ben Steinberg (Harvard University)

Additional Thanks To:

- Informal interview participants
- Attendees at the 2015 Texas / US Region 3 ETD Association
- Ryan Steans (Texas Digital Library)
- Vireo Steering Group members who reviewed the recommendations
 - Stephanie Larrison (Texas State University)
 - Joy Perrin (Texas Tech University)
 - David Reynolds (Johns Hopkins University)
 - Laura Spradlin (University of Illinois at Urbana-Champaign)

Questions?
