Who? What? Why? When? A Tale of ETD Metadata

Kara Long, Colleen Lyon, Kristi Park, Sarah Potvin, Monica Rivero, Santi Thompson

TxETDA/USETDA Region 3 Joint Conference, 2015
Waco 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.

Overview

Background

Approach

Use Cases

Next Steps

Discussion

Who?

TDL ETD Metadata Working Group

- Kara Long (Baylor University Libraries)
- Colleen Lyon (University of Texas Libraries)
- Kristi Park (Texas Digital Library)
- Sarah Potvin (Texas A&M University Libraries), Chair
- Monica Rivero (Fondren Library, Rice University)
- Santi Thompson (University of Houston Libraries)

What?

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.

What’s metadata?

Advisor
Author

Dates

Rights information
Subject headings

Title

Playing soldiers: martial subjects in early modern English drama, 1590-1660

Show simple record

dc.contributor.advisor	Whigham, Frank.
dc.creator	Pasupathi, Vimala Claeamona
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 literature.
dc.subject.lcsh	English drama--17th century
dc.title	Playing soldiers: martial subjects in early modern English drama, 1590-1660
dc.description.department	English
dc.identifier.oclc	68668387

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?

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

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:

Go

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

VIREO

"...make recommendations [for an] ingestion application…"

Why update?

Inconsistent application

New use cases

Vireo development

Need for systems & standards to work together

Approach

- Gather data: Review of existing standards
  Feedback from stakeholders
- Engage with Vireo Users Group
- Develop use cases
- Draft recommendations
- Use “Task and Review” Group

Use Case 1:
Name disambiguation

Use Case 1: Name disambiguation

Motivating problem:

Institutions have come up with local approaches to name disambiguation to address problems and create work-around solutions. These issues were not addressed in the original ETD guidelines.

Columbus, Christopher vs Columbus, Christopher

Columbus, Christopher vs Columbus, Christopher, 1958-

Use Case 1: Name disambiguation

Problems:

It is difficult to accurately search by author within a repository - potential problems with harvested metadata causing further ambiguity

Use Case 1: Name disambiguation

Problems:

Exposure of birth years is not a workable option for all institutions

Lyon, Colleen, active 2013

Use Case 1: Name disambiguation

Problems:

Students don’t consistently provide all the necessary information in order to create a unique name heading, or the information provided is inaccurate.

Use Case 1: Name disambiguation

Methodology:

- Conversations with stakeholders
- Reviewing standards around name disambiguation
- Reviewing methods of name disambiguation

Use Case 1: Name disambiguation

Current practices:

- Locally created authority files
- Crosswalking from/to MARC
- ORCiD integration

Use Case 1: Name disambiguation

Recommendation:

Option 1: ORCiD integration

Use Case 1: Name disambiguation

Recommendation:

Option 2: Integration with campus information systems

Use Case 2:
Problems with Dates

Use Case 2: Problems with Dates

Motivating Problems:

Thesis Offices, libraries, and users are confused by the dates found in each record-- where did they come from? what do they mean?

Use Case 2: Problems with Dates

Current (2008) TDL Guidelines on dates:

- Date of Creation [dc.date.created]
  - Graduation date <u>or</u> date degree is conferred
- Date of Publication [dc.date.issued]
  - Date ETD released to public
- Author Birth Date
- Advisor Birth Date
- Committee Member Birth Date
- Record Creation Date
- Record Change Date

Use Case 2: Problems with Dates

Problems:

Growing pains in moving away from simple best practices

Type of Document	Dissertation
Author	Eagle, Kimberly S
Author's Email Address	keagle@vt.edu
URN	etd-04282005-095141
Title	New Public Management in Charlotte, North Carolins: A Case Study of Managed Competition
Degree	PhD
Department	Public Administration and Public Affairs
Advisory Committee	**Advisor Name** **Title** Dudley, Larkin S. Committee Chair Gooden, Susan T. Committee Member Svara, James Committee Member Wamsley, Gary L. Committee Member Wolf, James F. Committee Member
Keywords	• Public Management • Charlotte • managed competition • normative theory
Date of Defense	2005-04-25

Use Case 2: Problems with Dates

Problems:

Unclear what various date fields mean for users and administrators

- Example:
  - dc.date.accessioned
  - dc.date.issued
  - dc.date.available

Use Case 2: Problems with Dates

Problems:

Local uses of dates include customizations that may diverge from existing guidelines.

- Example:
  - Graduation date stored in:
    - dc.date.created
    - dc.date.graduation
    - dc.date.graduationMonth

Use Case 2: Problems with Dates

Methodology:

- Analyze standards around dates in ETD guidelines
- Analyze how dates are generated in and packaged by Vireo
- Benchmark sample ETDs from NDLTD institutions

Use Case 2: Problems with Dates

Results:

Identified few dates that were purely descriptive

- coverage dates
  - dc.coverage.temporal
- dates associated with subjects
  - dc.subject
- dates included in other descriptive fields
  - dc.title
    - *Innovation in the Electronics Industry (1955-2010)*
  - dc.description.abstract
    - “Using a broad panel of US and Japanese publicly listed IT firms in the period 1983-2004, it shows this change in the nature of IT…”

Use Case 2: Problems with Dates

Results:

Identified dates that were both descriptive/administrative

- author or advisor birth dates
- date of defense
- date of graduation
- date of publication
- period of embargo

- Some institutions are specifying copyright dates
  - dc.date.copyright
  - dc.date.description

Use Case 2: Problems with Dates

Results:

Awareness that systems [such as Vireo and DSpace] are generating or packaging the dates we are capturing and exposing to the user.

Use Case 2: Problems with Dates

Results:

Variations in how ‘readable’ institutional metadata is.

Example, simple (9) and full (19) item view:

http://hdl.handle.net/2429/52252

dc.language.iso	eng	en_US
dc.publisher	University of British Columbia	en
dc.rights	Attribution-NonCommercial-NoDerivs 2.5 Canada	
dc.rights.uri	http://creativecommons.org/licenses/by-nc-nd/2.5/ca/	
dc.title	Effect of water matrix on Vacuum UV process for the removal of organic micropollutants in surface water	en_US
dc.type	Electronic Thesis or Dissertation	en
dc.degree.name	Doctor of Philosophy - PhD	en_US
dc.degree.discipline	Chemical and Biological Engineering	en_US
dc.degree.grantor	University of British Columbia	en
dc.date.graduation	2015-05	en_US
dc.degree.campus	UBCV	en_US
dc.description.scholarlevel	Graduate	en

Files in this item

Files	Size	Format	Description	View
ubc_2015_may_duca_clara.pdf	4.053Mb	Adobe Portable Document Format		View/ Open

This item appears in the following Collection(s)

Electronic Theses and Dissertations (ETDs) 2008+

Show simple item record

All items in cIRcle are protected by copyright, with all rights reserved.

creative commons Except where otherwise noted, this item's license is described as Attribution-NonCommercial-NoDerivs 2.5 Canada

Show Statistical Information

Use Case 2: Problems with Dates

Results:

Variations in how ‘readable’ institutional metadata is.

dc.date.accessioned	2015-02-24T21:38:29Z
dc.date.available	2015-02-24T21:38:29Z
dc.date.copyright	2015
dc.date.issued	2015-02-24
dc.date.graduation	2015-05

Copyright Date:	2015
Publicly Available in clRcle	2015-02-24

Use Case 2: Problems with Dates

dc.date.accessioned	2015-01-12T21:11:09Z
dc.date.available	2015-01-12T21:11:09Z
dc.date.created	2014-12
dc.date.issued	2014-08-21
dc.date.submitted	December 2014

?

```
▼<dim:field mdschema="dc" element="description" qualif
   Submission original under the a zero month embargo l
 </dim:field>
▼<dim:field mdschema="dc" element="description" qualifier="provenance">
   The student, Derya Aksaray, accepted the attached license on 2014-08-21 at 00:44.
 </dim:field>
▼<dim:field mdschema="dc" element="description" qualifier="provenance">
   The student, Derya Aksaray, submitted this Dissertation for approval on 2014-08-21 at 00:56.
 </dim:field>
▼<dim:field mdschema="dc" element="description" qualifier="provenance">
   This Dissertation was approved for publication on 2014-08-21 at 11:53.
 </dim:field>
▼<dim:field mdschema="dc" element="description" qualifier="provenance">
   DSpace METS Submission Ingestion Package generated from Vireo submission #1378 on 2015-01-12 at 16:11:08
 </dim:field>
▼<dim:field mdschema="dc" element="description" qualifier="provenance" lang="en">
   Submitted by Vireo Sword-App (vireo@library.gatech.edu) on 2015-01-12T21:11:09Z No. of bitstreams: 3 AKSARAY-DISSERTATIO
   (MD5) LICENSE_1.txt: 3958 bytes, checksum: 590b85873d26898a20b86f43a098eaf7 (MD5) LICENSE.txt: 3958 bytes, checksum: 410
 </dim:field>
▼<dim:field mdschema="dc" element="description" qualifier="provenance" lang="en">
   Made available in DSpace on 2015-01-12T21:11:09Z (GMT). No. of bitstreams: 3 AKSARAY-DISSERTATION-2014.pdf: 12078669 byt
   bytes, checksum: 590b85873d26898a20b86f43a098eaf7 (MD5) LICENSE.txt: 3958 bytes, checksum: 410aba714d1d9c76920c35615aa74
 </dim:field>
```

Use Case 2: Problems with Dates

Results:

Most frequently used dates correspond to lifecycle management / administrative data, which has not necessarily been addressed by descriptive metadata standards.

Examples:

- date accessioned
- date available
- date approved by committee
- date submitted
- date of license approval

Use Case 2: Problems with Dates

Recommendations:

Future standards should:
(1) identify dates to be collected by Vireo, stored with ETD, and exposed in a repository
(2) define how these dates are coded/populated

Example: “dc.date.approved should be filled as YYYY-MM with the date that the graduate school approved the ETD”

Use Case 2: Problems with Dates

Recommendations:

Dichotomy of descriptive/administrative does not hold up when it comes to ETDs.

Use Case 2: Problems with Dates

Out of scope / longterm goal:

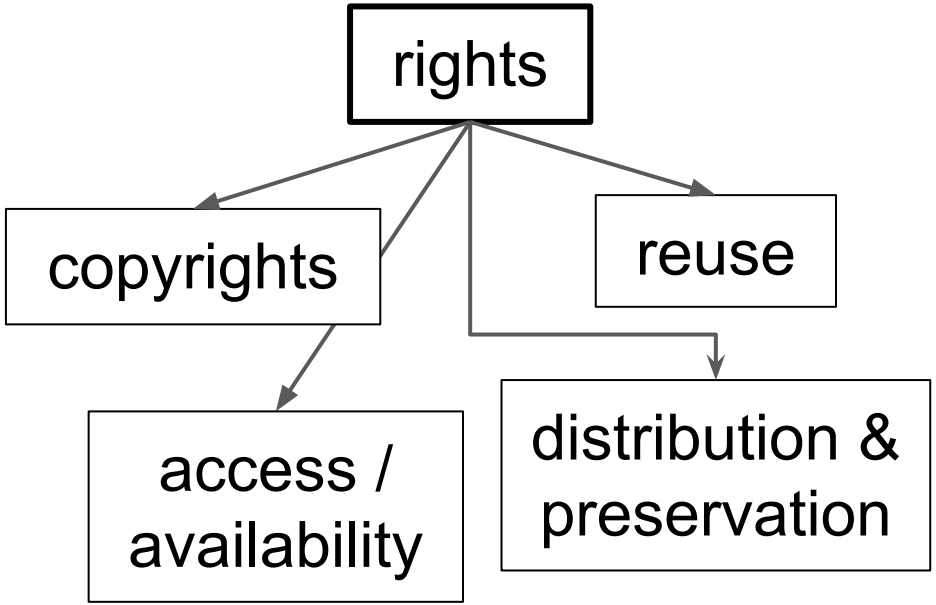
We need more robust systems for generating and including administrative metadata as object persists in a published state.

Use Case 3:
Rights, Availability and Access Metadata

Use Case 3: Rights, Availability and Access

Motivating Problem

- New element to the TDL metadata standard



Use Case 3: Rights, Availability and Access

Methodology

Section 3: Metadata Elements

3.1 free_to_read

3.2 license_reference

3.2.1 Description

3.2.2 License URI Expression

Section 4: Recommended Mechanisms for Distributing Metadat

Section 5: Use Case Review

5.1 Use Case: End User Seeks to Discover, Identify, and Access Free-to-Read Items

5.2 Use Case: End User Seeks to Know the Readability Status of an Item

5.3 Use Case: End User Seeks to Know Re-Use Permissions of an Item

5.4 Use Case: End User Seeks to Know Re-Use Permissions of a Sub-Component of an Item

5.5 Use Case: Repositories Seek to Expose Free-to-Read Items

5.6 Use Case: End User Seeks to Text Mine Content

5.7 Use Case: Ensure Author/Publisher Rights Assertions Align with License Statements

5.8 Use Case: Funding Agency Seeks to Track Compliance of Research Outputs to Open Access Mandates

5.9 Use Case: Institution Seeks to Report on Open Access Compliance of Research Outputs

NISO

The Recommended Practice Access and License Indicators (RP-22-2015)

NDLTD

PREMIS

```
targetNamespace="http://www.loc.gov/standards/RightsMD/" elementFormDefault="qualified" attributeFormDefault
RightsMD: Rights Metadata Extension Schema
<xsd:element name="RightsMD" type="RightsMDType"/>
<xsd:annotation>
<xsd:element name="registration" type="registrationType"/>
<xsd:element name="restriction" type="restrictionType"/>
<xsd:complexType name="RightsMDType">...</xsd:complexType>
<xsd:complexType name="credit_lineType">...</xsd:complexType>
<xsd:complexType name="registrationType">...</xsd:complexType>
<xsd:complexType name="rights_holderType">...</xsd:complexType>
<xsd:complexType name="restrictionType">
<xsd:annotation>...</xsd:annotation>
<xsd:sequence>
<xsd:element name="restriction_LC">
<xsd:simpleType>
<xsd:restriction base="xsd:string">
<xsd:enumeration value="unrestricted_copyright_expired"/>
<xsd:enumeration value="unrestricted_deed_of_gift"/>
<xsd:enumeration value="unrestricted_US_government"/>
<xsd:enumeration value="unrestricted_LC_review_unrestricted"/>
<xsd:enumeration value="restricted_LC_review_low risk"/>
<xsd:enumeration value="restricted_LC_review_high risk"/>
<xsd:enumeration value="restricted_copyrighted_permission_not_sought"/>
<xsd:enumeration value="restricted_copyrighted_permission_granted"/>
<xsd:enumeration value="restricted_copyrighted_permission_denied"/>
umeration value="restricted_donor_deed_of_gift"/>
umeration value="restricted_classified"/>
umeration value="restricted_non_copyright_issue"/>
umeration value="unrestricted_copyright_expired"/>
umeration value="restricted_recsound_permission_not_sought"/>
umeration value="restricted_folklife_permission_not_sought"/>
umeration value="restricted_folklife_permission_denied"/>
umeration value="restricted_recsound_permission_granted"/>
umeration value="restricted_co_in_process"/>
umeration value="restricted_co_not_released"/>
umeration value="restricted_co_released_no_access"/>
umeration value="restricted_co_released_selectable"/>
triction>
eType>
name="restriction_expiration_date" type="xsd:date"/>
<xsd:element name="restriction_inception_date" type="xsd:date"/>
```

Use Case 3: Rights, Availability and Access

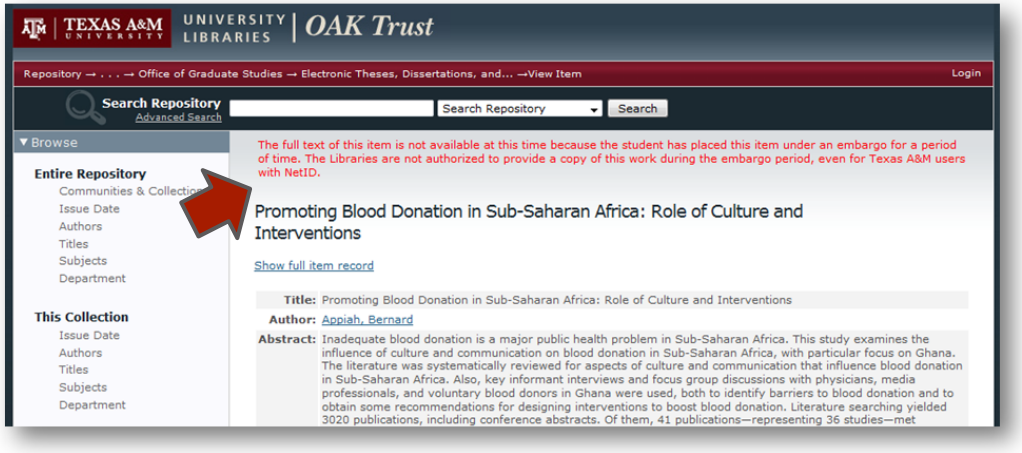
Practices

dc.date.copyright dc.description dc.description.embargo

local.embargo.lift dc.description.embargopolicy local.embargo.terms

dc.rights.cc dc.description.release dc.description.

statementofresponsibility dc.rights dc.rights.uri dc.rights.availability



Use Case 3: Rights, Availability and Access

Access levels examples

NDLTD ETD-MS (rights levels)	**OATD**/Dowling	**Open Data Project**
● Not publicly accessible ● Limited public access ● Publicly accessible	● restricted ● unrestricted	● public ● restricted public ● non-public

Use Case 3: Rights, Availability and Access

Recommendations

- dc.rights.accessRights

Proposed Levels :

- Publicly accessible
- Access restricted to campus
- Access restricted due to embargo. Release date xxxx-xx-xx.

Use Case 3: Rights, Availability and Access

Recommendations

- dc.rights.rightsHolder
  e.g. Copyright 2013 © John Smith
- dc.rights
  - e.g. The authors of the theses and dissertations are the copyright owners. The Digital Library and Archives has their permission to store and provide access to these works.
  - e.g. Copyright © is held by the author. Digital access to this material is made possible by the University Libraries. Further transmission, reproduction or presentation (such as public display or performance) of protected items is prohibited except with permission of the author.

Use Case 3: Rights, Availability and Access

Recommendations

- dc.rights.uri
  - Creative Commons: Attribution 3.0 Hong Kong License
  - http://creativecommons.org/licenses/by-nc-nd/3.0/us/

Next steps

- Draft report expected April, 2015
- Task & review group
- Vireo User Group approval
- Final report will be made available

Acknowledgements

Thank you to all who have contributed and provided feedback, including (but not limited to):

Renee Babcock, Shelley Barba, Jane Costanza, Jon Crossno, Thomas Dowling, Susan Elkins, Laura Hammons, Sandra Harman, Jeanne Hazzard, Cameron Kainerstorfer, Stephanie Larrison, Ann Marchock, Billie Peterson-Lugo, Amy Rushing

Discussion

- How are you disambiguating names? Who handles this in your university?
- Do you understand what the dates in ETD records mean?
- In the work you do to curate ETDs, what dates do you think are most important?
- What dates in the provenance field (hidden from immediate view and, in some cases, restricted from public access entirely) should be exposed, and how should they be exposed?
- Have you encountered cases of complex rights-- for example, where students have previously published their T or D as an article or articles, with copyright transferred to the publisher?