

An Analysis of Evolving Metadata Influences, Standards, and Practices in Electronic Theses and Dissertations

Sarah Potvin and Santi Thompson

This study uses a mixed methods approach to raise awareness of divergences between and among current practices and metadata standards and guidelines for electronic theses and dissertations (ETDs). Analysis is rooted in literatures on metadata quality, shareable or federated metadata, and interoperability, with attention to the impact of systems, tools, and practices on ETD date metadata. We consider the philosophies that have guided the design of several metadata standards. An examination of semantic interoperability issues serves as an articulation of the need for a more robust ideal moving forward, rooted in lifecycle models of metadata and concerned with the long-term curation and preservation of ETDs.

Sarah Potvin (spotvin@library.tamu.edu) is a Digital Scholarship Librarian at Texas A&M University Libraries. **Santi Thompson** (sathompson3@uh.edu) is the Head of Digital Repository Services, University of Houston.

Submitted August 7, 2015; returned to authors for revision November 3, 2015; revised manuscript submitted December 18, 2015; accepted for publication January 4, 2016.

The authors benefitted tremendously from conversations with our colleagues on the Texas Digital Library Electronic Thesis and Dissertation Metadata Working Group: Kara Long (Baylor University), Colleen Lyon (University of Texas), Monica Rivero (Rice University), and Kristi Park (Texas Digital Library). We are thankful for their expertise and insight; this article would not have taken shape without them. Hannah Tarver (University of North Texas) and Shelley Barba (Texas Tech University) provided invaluable, generous feedback on our draft, as did Amy Rushing (University of Texas at San Antonio) and Brian Surratt, chairs of prior TDL metadata working groups. Many thanks to *Library Resources and Technical Services* reviewers and to Mary Beth Weber for their suggestions and careful editorial guidance.

As theses and dissertations have evolved in format from shelved print resources to electronic files housed in institutional repositories, recordkeeping practices have been developed to account for the description of theses' content and their administration across a lifecycle marked by institutional approval, deposit, publication, and preservation.¹ These practices are based in standards and recommendations issued at institutional, regional, national, and international levels. As Pargman and Palme have argued, "What can and what cannot be expressed when it comes to electronic communication is, in the end, determined by the underlying and in many respects invisible infrastructure of standards that enables (and, at the same time, constrains and restricts) such communication."² This paper attempts to raise the visibility of the standards and infrastructure, philosophies and practices that enable and constrain the expression of electronic theses and dissertations (ETDs) as records.

Development and Application of ETD Metadata Standards

The development and application of ETD metadata standards, and the resulting quality, consistency, and interoperability of the metadata produced and exchanged, incur major implications for the discovery and long-term preservation of these unique student works. As Arms et al. asserted, "The goal of interoperability is to build coherent services for users, from components that are technically different and managed by different organizations." They noted that "This requires agreements to cooperate at three levels: technical, content

and organizational.” Here, we focus primarily on content and organizational aspects of interoperability, aspects that emphasize semantic agreement (content) and “ground rules for access, preservation of collections and services, payment, authentication” (organization).³ We observe, in particular, failures of semantic interoperability, which distort the meaningful, consistent interpretation of metadata values associated with particular elements.

What forces have proliferated inconsistent metadata, further complicating interoperability?² Broadly, we argue that the failures of interoperability, particularly for date-related metadata, are exacerbated by divergent philosophies about the role of metadata, viewed either as primarily descriptive or as a distinct component in the lifecycle management of electronic documents, and are shaped by the constraints enforced by the systems and tools developed to shepherd ETDs. This argument is an attempt to reconcile how philosophies and tools have restrained and expanded metadata practices, and to document the incongruities between reality and ideal. A view of the issue that considers recent history, coupled with close analysis of standards, positions us to identify gaps in the sociotechnical infrastructure, and to understand forces, whether decisions, compromises, or trends, that have separated practice from ideal.

This paper uses a mixed methods approach to illustrate divergent metadata philosophies and the impact of systems, tools, and practices on ETD date metadata. First, we review the historical developments of ETDs and ETD metadata. This analysis of the guiding principles of the ETD movement highlights how practices have changed over time. We then conduct a meta-analysis of various ETD standards and guidelines, designed to show areas of agreement and confusion across these ideals, as well as to indicate the distinctive goals and philosophies underpinning these approaches. Next, we sample data from selected Networked Digital Library of Theses and Dissertations (NDLTD) institutions to better understand the current quality and consistency of ETD date metadata. Finally, we consider how tools affect metadata standards and practice, using DSpace repository software and the coevolution of tools and standards produced by one state consortium as examples. Attention to date metadata is prompted by ETD stakeholders’ confusion over the quantity and meaning of dates provided in ETD metadata, and in the interest of analyzing an aspect of metadata with descriptive, technical, administrative, and preservation implications.

By raising awareness of these shortcomings, and the forces behind them, we hope to begin to move closer to and engage with approaches that consider the long-term curation and preservation of ETDs. Our goal, too, is to promote a deeper understanding of the tradeoffs incurred in emphasizing a union catalog model for the discovery and administration of ETDs. These tradeoffs are suggested by

the union catalog model itself, which privileges metadata over full-text, as well as by the application of this model for ETDs. Generally, in the union catalog approach, metadata are emphasized as the basis for resource discovery: whether in its traditional form (aggregating records contributed by member institutions into a central database) or its more recent incarnation (aggregating records from multiple repositories automatically, via harvesting protocols), union catalogs unify multiples source into a single record set.⁴ In the particular case of ETDs, the union catalog approach has been tremendously successful in enabling search and discovery of ETDs across repositories and countries, providing a low barrier to entry for institutions contributing metadata and users searching across metadata records. The dominant metadata standard for ETD exchange, NDLTD’s ETD-MS, is a relatively lean standard, designed to emphasize ease of inclusion.⁵ It follows that institutional approaches to ETD metadata that reify NDLTD compliance as the ultimate objective, rather than the most basic format of exchange, may forfeit the rich affordances of these digital objects, including discovery and information retrieval enabled by full-text search.⁶

Additionally, because the union catalog model for ETD exchange emphasizes descriptive metadata and largely ignores administrative, technical, and preservation metadata, a lack of awareness of the limits of the model may threaten ETDs’ long-term survival. Administrative, technical, and preservation metadata document the structure of an object and trace its provenance throughout the object’s lifecycle. Popular preservation metadata schema, like Preservation Metadata: Implementation Strategies (PREMIS), frequently contain information on the composition of an object (including file size and formats), chronicle important actions and decisions made over time to extend access to an object (including decisions to migrate file formats), and outline specific rights management issues that can determine an object’s accessibility.⁷ Maintaining this information helps build trust in records by ensuring that they are authentic and reliable. “The Society of American Archivists’ *Glossary of Archival Terminology* defines authenticity as ‘the quality of being genuine, not a counterfeit, and free from tampering’ and reliability as ‘the quality of being consistent and undeviating.’”⁸ Because the union catalog model focuses primarily on descriptive metadata, it might lack the evidence needed (found in administrative, technical, and preservation metadata) to ensure that libraries have maintained authenticity and reliability. The need for metadata beyond descriptive becomes apparent in some real world scenarios: for example, as institutions migrate content from one repository to another, the descriptive metadata frequently privileged by the union catalog model may prove insufficient in capturing the structure of complex objects, in explaining metadata decisions developed to meet specific

system requirements or functionality, or in accounting for an object's administration over time (including, for example, when an embargo expired). Failure to account for technical, administrative, and preservation metadata incurs the risk of limiting functionality in the new system or losing the ability to faithfully render the object. Alternatives to the union catalog model are addressed in greater detail in subsequent sections of the paper.

Metadata Concerns with the Emergence and Growth of ETDs: A Recent History

The roots of the ETD movement extend to experiments in dissertations produced in Standard Generalized Markup Language (SGML) markup in the 1980s, growing out of discussions between UMI (later ProQuest) and the Virginia Polytechnic Institute and State University (Virginia Tech). As Fox described it, meetings in the late 1980s and 1990s brought together the Coalition for Networked Information, the Council of Graduate Schools, UMI, Virginia Tech, SoftQuad, and Adobe (then testing their Portable Document Format), with coordinated efforts and progress with community building and technology.⁹ Virginia Tech initiated a requirement that students submit ETDs, rather than printed documents, in 1997, the first institution to do so.

As ETDs moved from theory to practice, the literature emphasized two key areas of promise and innovation in the transition away from print: expression and access.¹⁰ The former considered the possibility that students, now unrestricted by print format requirements, could more fully express their creative and scholarly vision.¹¹ This hope was wedded to the more pragmatic idea that graduate education would be enhanced by students' mastering those digital production tools necessary to author even a basic ETD.¹² In the latter scenario, the format of ETDs is linked to possibilities of access, and to works distributed, aggregated, and made available worldwide, to wider audiences than bound, shelved volumes had permitted.

The new formats and promise of ETDs posed a challenge to libraries. As Virginia Tech librarian Gail McMillan observed, "Theses and dissertations as electronic files transferred from the Graduate School to the Library may well be the first major source of electronic texts that many libraries and their catalogers will regularly encounter," and the "first significant body of electronic materials [that] regularly requiring cataloging."¹³ McMillan identified two goals, based on quality and efficiency, that developed in Virginia Tech's initial efforts to process ETDs: (1) ensuring that "access would be at least as good as it is for a hard copy" and (2) developing workflows and practices to "derive cataloging information from the electronic text and avoid rekeying as much as possible."¹⁴

These concerns about access and avoidance of redundant labor were taken up in an extensive subsequent literature examining efficiencies in creating bibliographic records for ETDs and developing workflows. The literature reflects an anxiety surrounding the shift from bibliographic records created by expert catalogers to metadata records supplied by ETD authors. Particular attention was paid to the enhancement of author-contributed metadata and cost-benefit analyses of expert-assigned subject headings.¹⁵ As full-text electronic documents associated with bibliographic records, ETDs represented a significant shift from a machine-readable record serving as surrogate for a separately located print item. Lubas observed, "ETDs are full-text searchable in DSpace and other repository systems, so the need for a metadata quality control process or application of a controlled vocabulary may not appear paramount."¹⁶ Yet the union catalog model of ETD discovery, promoted by groups such as NDLTD, continues to rely on metadata, not full-text search, in aggregated discovery environments.¹⁷

Part of the challenge of cataloging ETDs was specific to the genre of theses and dissertations, rather than the electronic format. As unique items, theses and dissertations, even before the advent of ETDs, prompted special considerations for catalogers. Repp and Glaviano explained in a 1987 article, "As Library of Congress priorities preclude cataloging of even depository copies of dissertations submitted for copyright, no LC cataloging for dissertations appears on the bibliographic utilities, and full responsibility for bibliographic control falls to the degree-granting institution."¹⁸ Local responsibility for creating records, where abundant information was relegated to local fields, took its toll. As McMillan observed in the mid-1990s, "Even the full MARC record for a dissertation is not very robust and often has a local twist, presenting valuable information in a unique format that can be seen only at the originating institution because it is masked to users of OCLC or other centralized cataloging repository."¹⁹ In the pre-ETD era, scholars interested in viewing graduate works either traveled to the holding institution, requested a print copy via interlibrary loan, or viewed a UMI reproduction. Repp and Glaviano described significant barriers to discovering dissertations, barriers that were lessened for the intramural scholar, who was likely to have access to records tailored for local access, locally maintained indexes, or "special shelving arrangements, amenities lost to the extramural scholar."²⁰

Irregularities, idiosyncrasies, and local conventions for cataloging theses and dissertations have contributed to ongoing metadata interoperability issues for union catalogs and other shared records. These challenges were magnified and significantly altered as graduate works moved into the sphere of digital delivery and non-MARC metadata.

Models for ETD Metadata: Discovery and/or Curation

As the management of theses and dissertations evolved from print to the electronic environment, those responsible for ETDs focused on creating policies, tools, and workflows address the deposit, access, and preservation of these documents. These included the capacity to manage file formats, support categories of metadata (including descriptive, administrative, structural), assert the rights of authors or publishers, and elucidate access policies. Accounting for data that assists in the management of ETDs was a change in practice for libraries accustomed to emphasizing retrieval and access when generating cataloging records. Greenberg argued that strides have been made toward conceptions of “metadata as structured data about an object that supports functions associated with that object” and noted that repositories, with their connection to “archival or recordkeeping practices,” may diverge from goals and metadata types and functions that dominate in libraries.²¹ This shift reflects potentially divergent philosophies of metadata: one was founded in a simplified vision of library cataloging approaches and theories grounded in print, seen as emphasizing the record as descriptive surrogate; the second moved toward managing electronic and networked objects and a pressing need to consider long-term access and curation.

As noted above, McMillan observed in the mid-1990s, “Even the full MARC record for a dissertation is not very robust and often has a local twist, presenting valuable information in a unique format that can be seen only at the originating institution because it is masked to users of OCLC or other centralized cataloging repository.”²² Both cataloging and metadata practices are aimed at resource description to facilitate discovery and access. Approaches to ETD metadata that focus exclusively on adherence to the NDLTD union catalog model are the equivalent of cataloging approaches attentive only to OCLC exchange, stripped of the administrative information related to a work’s acquisition, circulation, preservation, and access requirements. As discussed in an earlier section, cataloging practices provided the foundation for metadata creation for the first ETDs. In this section, we explore the influence of lifecycle records management in relation to the development of ETD standards and guidelines and address the distinctive goals of describing items and curating ETDs.

The record lifecycle model, popularized by researchers examining the collection, description, and preservation of records, recognized that objects are not static, but are born, change and evolve as they age, and eventually die.²³ Building on this metaphor, the lifecycle model traced important events that took place while the document ages. As technology shaped how records were created, shared, and preserved, information professionals adapted the broad lifecycle model

to fit new record keeping challenges. Some frameworks, like the Digital Curation Center’s Lifecycle Model, illustrate the iterative roles that curation and preservation play in the long-term maintenance of digital objects (see figure 1).

Researchers have argued for the explicit application of a lifecycle model to metadata, helping us to both understand metadata and create metadata models that complement and embody the lifecycle approach to digital resource management. As Greenberg explained,

A key reason for using lifecycle concepts for repositories is that digital resources are more mutable and sharable than their physical printed counterparts; and the mutable nature presents a seemingly *organic* object . . . like the digital resource, metadata—in digital form—is more mutable and sharable than traditional cataloging records printed for library card catalogs, or maintained in closed databases.²⁴

The lifecycle model, she argued, “not only [has] appeal, but a proven applicability.”²⁵

Literature on ETD management has also aligned with the lifecycle model. According to the *Guidance Documents for Lifecycle Management of ETDs*, this model has sought to “study and document the progression of digital objects through stages of creation, dissemination, use, update and re-use, storage retention or archiving, and sometimes destruction or disposal, of digital objects.”²⁶ Because of its expansive scope and iterative approach, the lifecycle model approach is well suited to facilitate the processes of acquiring, administering, providing access to, and preserving ETDs. Since the model focuses on an object from creation to either its destruction or disposition in a repository for long-term access and preservation (and further evaluation for retention in the future), it incorporates all of the stakeholders who play a role in the ETD process, including the student/creator, faculty committee, graduate school, university library, and university information technology.²⁷ The model also accommodates a complex workflow that can allow for simultaneous actions from different contributors.

Review of Standards: Treatment of Dates

As ETD management embraces the lifecycle management approach, ETD standards are developing recommendations that better account for key dates in an ETD document’s lifecycle. While we argue that capturing dates in the work’s lifecycle is integral to any robust method for administering these materials, ETD standards have not always supported this approach. The earliest ETD standards, which predate the dominance of the lifecycle management model, focused on a philosophy of metadata that emphasized data exchange

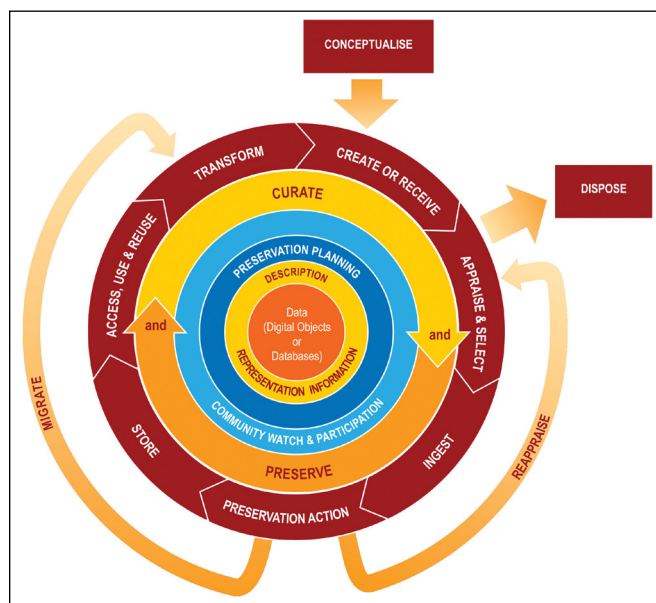


Figure 1. Digital Curation Center Curation Lifecycle Model. Source: Digital Curation Center, University of Edinburgh, "DCC Curation Lifecycle Model," last modified 2014, www.dcc.ac.uk/resources/curation-lifecycle-model.

and discoverability. As such, these standards focused on descriptive metadata elements, such as identifying title, author, and subjects.

These standards allocated one or several fields for capturing date information. For example, NDLTD's ETD-MS: an Interoperability Metadata Standard for Electronic Theses and Dissertations, first published in 2001, served primarily to promote exchange of metadata and the creation of a union catalog among NDLTD member institutions. Early, ambitious attempts by that organization to build an XML DTD standard for encoding the full text of an ETD had been met with resistance from members. Instead, ETD-MS "emerged as a flexible set of guidelines for encoding and sharing very basic metadata regarding ETDs among institutions."²⁸ ETD-MS identifies one date category that should be recorded, mapped to the DC element date and requiring the user to capture the date "that appears on the title page or equivalent of the work."²⁹ Created in 2009, the Electronic Thesis Online Service (EThOS) metadata standard, used in the United Kingdom as the basis for a national union catalog, outlined two date fields to capture: the date the thesis is awarded and, if applicable, the date that an embargo on the document ends.³⁰

Standards and guidelines evolved to incorporate more than the date of creation or publication; many of these standards embraced another philosophy of metadata that began to emphasize the management and preservation of these objects as records. As such, these standards and guidelines

paid greater attention to administrative dates. The 2014 *Guidance Documents for Lifecycle Management of ETDs* identifies four key areas where dates should be recorded: a general date (ideally, publication or graduation date); a date when an embargo ends; birth and death years of the creator to track copyright issues; and dates to track preservation work on the document.³¹ In 2014, OhioLINK, a consortium of academic libraries in Ohio, which hosts an ETD Center, established a standard for recording ETD metadata in RDA. Like the *Guidance Documents*, the OhioLINK standard identified four key dates to capture, including copyright date, production (or publication) date, the date the degree was awarded, and the date that any restricted access on the document ends.³² In 2015, the Texas Digital Library, a consortium of academic libraries in Texas, which hosts a shared metadata repository for ETDs and the Vireo thesis management tool, issued updated metadata guidelines that included an expansive set of dates to capture and publish, including copyright date, graduation date, date of repository ingest, date made public in the repository, date of embargo lift, and author birth date. These guidelines recommended that date fields "be revised and enhanced with increasing reliance on provenance fields to supply additional context for ambiguous date values. Given the likelihood of fields to change meaning over time, explicit encoding of meaningful lifecycle dates in `dc.description.provenance` fields will help administrators make sense of the myriad dates associated with an item."³³ We consider the coevolution of standards and tools maintained by the Texas Digital Library in a subsequent section of this paper.

While the Thèses Électroniques Françaises (TEF) standard used in France does not explicitly reference the lifecycle model, the standard is exceptional in its articulation of eight areas where dates should be captured. Created to ensure that ETD metadata were both recorded and transferred in the differing contexts and applications used to administer the documents, the TEF guidelines address the holistic approach needed to generate important dates about metadata throughout the workflow. According to TEF, ETDs "reflect three dimensions that characterize the whole theses," including information that documents the "academic work validated by peers," "intellectual work subject to the law of intellectual property," and "an administrative document that governs the grant of a national diploma."³⁴ The dates captured by the standard reflect both descriptive and administrative metadata. The types of dates associated with this standard include: date of defense, date of publication, author birth date, date of record creation, date of record modification, date that embargo ends, and temporal coverage of the thesis.

Analyzing the variety of date fields reflected in ETD standards and guidelines reveals inconsistencies between the types, definitions, and granularity of dates to be captured by

ETD stakeholders, platforms, and tools. These inconsistencies are shaped by the differing philosophical approaches to metadata promoted in ETD standards. Standards such as ETD-MS focus on broad dates that represent the beginning of a document's lifecycle. This approach makes little data available for the long-term management of ETDs. Other standards leave the interpretation of the date being captured to the creator or ETD administrator (for example, reflecting the date shown on the cover page of the document, the date the document was submitted by the student to the Graduate School, or the date of the student's graduation—which may or may not be the same date depending on institutional policies and specific contexts). The lack of semantic clarity may create values that do not correspond between documents and impede interoperability. Still other standards vary in the amount and the detail of dates to be captured. For example, the TEF standard has specific fields for the date of the thesis defense and the date of thesis approval. Divergent standards guide the production of inconsistent metadata, which impede both the management and discoverability of ETDs.

A Snapshot of Metadata Quality and Consistency: NDLTD Member Institutions

The quality of ETD metadata, including fields associated with dates, presents another barrier to interoperability. Regional and national digital library consortia, like Open Access Theses and Dissertations and the Digital Public Library of America, rely heavily on metadata aggregation to bring disparate collections together into one user interface. For content to be discoverable in an aggregated environment, the metadata must be robust enough to include information queried by the user. Furthermore, records must contain similar fields and valid values in those fields. These properties require records creators to have standardized data entry practices and to use common guidelines for describing content.

With metadata driving how objects are discovered and reused, concerns about maintaining quality metadata have increased.³⁵ Information professionals have developed categories for analyzing metadata to evaluate its quality. The literature frequently cites Park's metadata quality measurement criteria as one of the most practical benchmarks in metadata evaluation.³⁶ Park identifies three core categories of metadata quality: completeness, accuracy, and consistency. Completeness of a metadata record "can be measured by full access capacity to individual local objects and connection to the parent local collection(s)."³⁷ Park notes that completeness does not necessarily correlate with populating a high number of elements with values that describe an object. Instead, it "can be measured by full access capacity to individual local objects and connection to the parent local collection(s). This reflects the functional purpose of metadata in resource

discovery and use."³⁸ Accuracy focuses on the "correctness" of an object's descriptive representation and can address spelling, formatting, and intellectual content.³⁹ Consistency, according to Park, accounts for data values at the "conceptual" and the "structural" levels. At the conceptual level, consistency "entails the degree to which the same data values or elements are used for delivering similar concepts in the description of a resource."⁴⁰ At the structural level, it addresses "the extent to which the same structure or format issued for presenting similar data attributes."⁴¹ Date values, expressed in a variety of ways, including natural language (January 1, 2015) and ISO 8601-compliant (2015-01-01), are examples of structural-level consistency. Collectively, these criteria provide information professionals with a framework to assess existing data and descriptive practices.

Drawing on this framework, we analyzed the consistency and accuracy of date elements across institutions to evaluate the quality of ETD metadata. We conducted an environmental scan of metadata records from sixteen NDLTD members. We harvested records from institutions' digital asset management systems, including DSpace, Digital Commons, and homegrown repositories, using OAI-PMH requests. We documented the categories, frequency, and purposes of dates being captured and made accessible by NDLTD member institutions. Our approach relies on sampling to provide insight into the current state of metadata quality related to dates. This approach requires close interpretation to match dates in records with their semantic meaning. Because ETD records are typically produced using tools that assure regularity, the dates included in these random samples are likely to be repeated across collections. However, readers should not assume that the information reflected in figure 2 necessarily reflects the practices of each institution.

The environmental scan revealed a diverse array of dates being captured by NDLTD member institutions. We analyzed one record per institutions. The number of date fields used by institutions varied from as little as one to as many as twelve.

Most dates conform to ETD-MS, including `dc.date`, `dc.date.available`, and `dc.description.provenance`. Complying with this standard promotes consistency among key dates in the ETD lifecycle, allowing high quality metadata (at least in relation to consistency) to be shared among numerous systems and libraries.

Divergences from the most common elements referenced in the previous chart occurred partly because repository systems generated different date fields over time. Two date fields, `dc.date.accessioned` and `dc.date.issued`, were used interchangeably to denote the date that content was deposited in a particular repository. DSpace metadata recommendations note that versions before 4.0 supported `dc.date.issued` for tracking an object's entry into the

Table 1. Comparison of Date Fields in ETD Metadata Standards and/or Guidelines

Source	Date Field	Field Definition
Networked Digital Library of Theses and Dissertations ETD-MS 1.1 (2009)	dc.date	The date “that appears on the title page or equivalent of the work”
EThOS UK ETD (n.d.)	dcterms:issued	The date the thesis was awarded
	uketdterms:embargodate	The date that an embargo on a document ends
Guidance Documents for Lifecycle Management of ETDs (2014)	date	Publication date. Graduation date.
	embargo lift date	“the metadata should include information sufficient to allow a repository system to know the date upon which the embargo is lifted.” ⁱ
	creator’s birth and death years	“Knowing the birth and death dates of the creator and the year in which the ETD was created will help to calculate and determine the copyright status.” ⁱⁱ
	preservation event date/time	The date when objects are altered by administrators
OhioLINK Standard for Cataloging ETDs in RDA (2014) ⁱⁱⁱ	264 #4 \$c © [year]	“Copyright date, if available. (RDA 2.11). Optional if there is a publication date.” ^{iv}
	264 #1 \$A \$c [year]	Publication date
	500 ##\$a [year]	“Quote ‘Year and Degree’ information from OhioLINK ETD Center website.” ^v
	502 ## \$d [year]	Degree granted date (“calendar year in which a granting institution or faculty conferred an academic degree on a candidate”)
	506 ## \$a Full text release delayed at author’s request until [year month day]	Restriction on access—Full date that an embargo on the document ends
Thèses Électroniques Françaises 2.0 (2006)	dcterms:dateAccepted	Date of thesis defense
	dcterms:issued	Date of publication
	tef:dateNaissance	Author birth date
	dcterms:temporal	Temporal coverage
	metsRights:ConstraintDescription	Date that embargo lifts
	mets:metsHdr CREATEDATE mets:dmdSec ID="CREATED"	Date of record creation
	mets:metdsHdr LASTMODDATE	Date of record modification
TDL Descriptive Metadata Guidelines for ETDs 1.0 (2008)	mods:dateCreated	“The date the student graduates or the date the degree is conferred” ^{vi}
	mods:dateIssued	“The date the ETD is released to the public.” ^{vii}
	mods:nameType="personal" mods:namePartType="date"	Birth year of author
	mods:nameType="personal" mods:namePartType="date"	Birth year of advisor
	mods:nameType="personal" mods:namePartType="date"	Birth year of committee member
	mods:recordCreationDate	“month, year, and day of the creation date of the record” ^{viii}
	mods:recordChangeDate	“month, year, and day of the change date [of the record]” ^{ix}

i. Alemneh, et. al., *Guidance Documents*, 6–3.ii. *Ibid.*

iii. While not explicitly stated in the standard, an appendix to the standard, “ETDs in RDA template, as of Oct. 2014; KSU example,” includes the dates (including day, month, and year) that the record was entered and replaced.

iv. OhioLINK, “Standards for Cataloging Electronic Theses and Dissertations.”

v. *Ibid.*, note included in “ETDs in RDA template, as of Oct. 2014; KSU example” appended to standard.

vi. Texas Digital Library, “Descriptive Metadata Guidelines,” 12.

vii. *Ibid.*

viii. Texas Digital Library, “Descriptive Metadata Guidelines,” 17.

ix. *Ibid.*

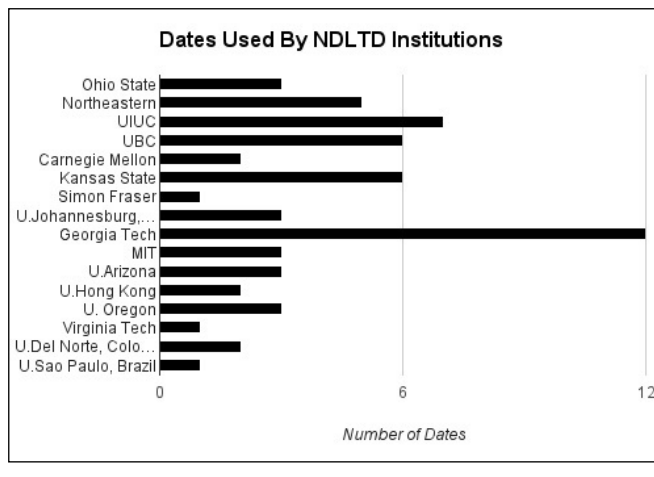


Figure 2. Dates Used by Selected NDLTD Institutions

repository, while DSpace versions at 4.0 or higher supported `dc.date.accessioned`.⁴² This change has direct implications on the quality of ETD date information. If institutions migrate to a newer version of DSpace but fail to transfer values from `dc.date.issued` to `dc.date.accessioned`, they store and disseminate inconsistent date elements throughout their repository. These inconsistencies decrease metadata quality.

Additionally, divergences existed because local metadata practices incorporated the usage of unique date elements to describe ETD temporal content. Popular dates among the ETD lifecycle received unique fields for some institutions. These fields included the following:

- `dc.date.graduation`
- `dc.date.graduationmonth`
- `dc.date.published`
- `dc.date.updated`
- `dc.dateAccepted`
- “Available in [name of repository]”
- “Date Deposited”
- “Embargo Period” portion of metadata record header
- date stamp in metadata record header

Multiple instances of date fields for graduation date, embargo date, and the content approval date suggest that these kinds of dates present unmet needs among metadata creators, digital repositories, and/or metadata guidelines. Because common date elements (like those ones listed in table 2) may not adequately address the rationale for these unique fields, future metadata guidelines should identify ways to accommodate some of the temporal data being captured in these local fields. Until this occurs, the proliferation of local date elements fosters inconsistent and inaccurate uses of temporal fields and compromises the overall quality of ETD metadata.

Understanding the consistency and accuracy of ETD date information becomes more complicated when analyzing the relationships between the types of dates captured by NDLTD institutions and the frequency with which they are used. Table 4 tracks the type of date used by NDLTD institutions, how often the institution used each type of date, and the date element where they recorded the temporal information. The table divides the latter information into two categories: common uses of the elements (used by over half of the sixteen NDLTD member institutions) and “localized” uses (used by fewer than half of the NDLTD institutions surveyed).

Inconsistent practices between even the most frequently used date type (the date an embargo ended, the date an object is published to the digital repository, and the date an object is submitted) suggest that future metadata guidelines should address some of the more specific ETD temporal data to promote more consistent and accurate uses of date elements. The varying ways that institutions convey the copyright date (`dc.rights`, `dc.date.copyright`, `dc.description`) also complicates description and accessibility, as some institutions repurpose the value in this date for other important administrative functions (including determining embargo start and end dates). Finally, the lack of guidance for graduation date continues to lead to the creation of localized fields, which further impede consistency across NDLTD institutions.

How Tools have Shaped de facto Standards

We have alluded to the influence of tools and systems such as repositories in the production of metadata. Our study of metadata standards and resulting practices would be incomplete without an examination of the influence of tools and systems in the development of de facto standards. In this final section, we consider the coevolution of tools and standards, concluding with the case of the Texas Digital Library (TDL).

Access platforms, the digital asset management systems or repositories into which documents and records are ingested, serve as influential factors in the creation and management of ETD metadata. These systems shape the de facto metadata standards for ETDs through automated processes of metadata creation and assignment, even as they are integrated into a wider system of Internet standards and protocols (like OAI-PMH) for discovery, persistence, and aggregation. Given our observation that the lack of definitional clarity in standards may create values that do not correspond between documents and impede interoperability, how do the systems used to ingest, manage, and steward ETDs reinforce, shift, or ameliorate these issues? How do the constraints of tools shape ETD management?

In some cases, it proves impossible to square the ideal of platform-neutral standards with the reality of platform

Table 2. Common Date Elements Used by NDLTD Institutions

Metadata Field	Definition ⁱ
dc.date	A point or period of time associated with an event in the lifecycle of the resource.
dc.date.available	Date (often a range) that the resource became or will become available.
dc.date.copyright	Date of copyright [dateCopyrighted].
dc.date.created	Date of creation of the resource.
dc.date.issued	Date of formal issuance (e.g., publication) of the resource.
dc.description	An account of the resource.
dc.description.provenance	A statement of any changes in ownership and custody of the resource since its creation that are significant for its authenticity, integrity, and interpretation.
dc.identifier.bibliographicCitation	A bibliographic reference for the resource.
dc.rights	A legal document giving official permission to do something with the resource.

i. "Section 2: Properties in the /terms/ namespace," Dublin Core Metadata Initiative, <http://dublincore.org/documents/dcmi-terms/#H2>.

Table 3. System-Generated Date Elements

Metadata Field	Definition
dc.date.accessioned	Date the repository took possession of the item. ⁱ
dc.date.issued	Date of formal issuance (e.g., publication) of the resource. ⁱⁱ

i. See DuraSpace, "Metadata and Bitstream Format Registries," DSpace 4.x Documentation, 2014, accessed February 8, 2016, <https://wiki.duraspace.org/display/DSDOC4x/Metadata+and+Bitstream+Format+Registries>.

ii. See "Section 2: Properties in the /terms/ namespace," in DCMI Usage Board, "DCMI Metadata Terms," Dublin Core Metadata Initiative website, accessed August 7, 2015, <http://dublincore.org/documents/dcmi-terms/#H2>.

constraints. Metadata manuals specify that, when developing a metadata application profile, one must consider the repository or content management.⁴³ Institutions make design decisions and select standards based on repository functionality. Yet researchers have argued for "the importance to reliable digital preservation management of . . . the practice of packaging digital objects in a repository-independent manner."⁴⁴ These decisions are particularly problematic when the repository-based access copy is the basis for the digital preservation copy. The adjustments made in metadata creation to conform to repository functionality belie the promise of repository-independent digital packages.

Inevitably, the dates associated with ETDs are shaped by the tools used to manage them, as the Vireo ETD submission system and DSpace demonstrate. DSpace, in its function as a core component of ETD management and publication, has contributed to the development of de facto standards that rely on DC and the ETD-MS Thesis schema.⁴⁵ As of 2015, DSpace only supports flat, non-hierarchical metadata schema. As the TDL case study will illustrate, this constrained functionality hastened the abandonment of MODS as the TDL ETD schema, particularly as TDL moved to a reliance on OAI-PMH for harvesting metadata into a portal of TDL ETD metadata, and sought compliance with the ETD-MS metadata standard.

But what are the broader implications of DSpace's emphasis on flat metadata, which has, since its 2004 launch, centered around a DC-dominated library application profile? In considering what level of description was adequate to enable discovery or administration, experts have continually expressed doubt about DC, but ease of use and functionality have hastened adoption. In a generalized critique from 2003, Martin Dillon described "three majors causes that can be adduced for the less than enthusiastic adoption of the library world of the Dublin Core": its "incompleteness," the lack of documentation or agreed-on standards for filling the fields, and "slow adoption."⁴⁶ Yet the use of unqualified DC for ETDs, Lubas argued, proliferated because of institutional repositories and OAI-PMH.⁴⁷ She observed,

While during the early days the use of a simplified metadata element set such as Dublin Core may have seemed limiting, over the course of a decade of experience with electronic theses and dissertations metadata reveals that blending the use of qualified Dublin Core with harvesting and crosswalks, plus creating tools to encourage better results from author-generated metadata have proved useful.⁴⁸

On a more granular level, DSpace has affected the dates that are included in object-level metadata. DSpace versions

Table 4. Type, Frequency, and Uses of Date Metadata at Selected NDLTD Institutions

Type of Date	Instances (N = 16)	Common Metadata Uses (> half of instances):	Local Metadata Uses (< half of instances):
Date embargo ended	12	dc.date.available	dc.date “Embargo Period” portion of metadata record header “Available in [name of repository]”
Date object published in the digital repository	12	dc.date.issued dc.date.available dc.description.provenance	dc.date dc.date.accessioned dc.date.published date stamp in metadata record header
Date object submitted (including to the digital repository)	10	dc.date.submitted dc.date.accessioned dc.description.provenance	“Date Deposited”
Date of degree or graduation	6		dc.date.created dc.date.graduation dc.date.graduationmonth dc.date.published dc.identifier. bibliographicCitation
Date of copyright	5	dc.rights dc.date.copyright dc.description	
Date of approval	2	dc.description.provenance dc.description	
Date of metadata record creation	2	dc.date.created	dc.date.submitted
Date object accepted by academic department	1	dc.dateAccepted	
Date of license agreement	1	dc.description.provenance	
Date of metadata record modification	1	dc.date.updated	
Date object withdrawn	1	dc.description.provenance	

before 4.0 automatically assigned values to dc.date.issued if items lacked values for that element, indicating prior publication. Confusion over this automatic assignment was not limited to ETDs. In 2013, Google and GoogleScholar alerted DuraSpace, the organization that oversees DSpace community development, that the automatic assignment of dc.date.issued (intended as the formal date of publication) as the value of the date of ingest was causing their web crawlers to inaccurately index publication dates. Google reported seeing “repositories, where 30–50 percent of their items all have the same ‘dc.date.issued,’ as those items were all imported on the same date.”⁴⁹ Rodgers noted that the automatic assignment of ingest dates for dc.date.issued was built into the system with a rationale in mind: “The bedrock use-case for DSpace was not published articles, but ‘grey lit’ (born digital content from an institution that was **not** in the official scholarly record): for this sort of content, appearance in the IR essentially **is** the equivalent of publication.”⁵⁰ With DSpace 4.0, the software stopped automatically assigning the date of accession as dc.date.issued, and the documentation specified that dc.date.issued, defined as date of publication, should be supplied by metadata creators.⁵¹ At the time

of publication, the DSpace community had yet to resolve the thorny issue of consistency in dc.date.issued for collections of items ingested before and after 4.0. As an outstanding card indicates: “we still need a way to help individual DSpace sites to locate any existing, possible inaccurate ‘dc.date.issued’ values.”⁵² For some ETDs, the date of ingest into DSpace would constitute a date of publication; for others, ingested into DSpace under embargoes, the date of publication is distinct from ingest.⁵³ By relying on an automated feature of DSpace, those standards that equated ingest with publication failed to anticipate and account for these use cases.⁵⁴

The Texas Digital Library, which develops and hosts the Vireo ETD thesis submission tool and an ETD metadata standard, furnishes an interesting case study in coevolution. The TDL ETD descriptive metadata standard was first developed as the basis of a union catalog of ETDs from TDL member institutions, introduced in the form of a shared ETD metadata repository. In 2005, TDL tasked a working group with “developing a common [descriptive] metadata standard that would allow members to share metadata in the TDL repository.”⁵⁵ The working group, rejecting the

Dublin Core expression of ETD-MS as flawed, issued an ETD-specific application profile for the Metadata Object Description Schema (MODS), which brought ETD-MS elements into MODS.⁵⁶ Rationalizing this decision, authors of the recommendation specifically referenced the limitations inherent to DSpace repositories and OPACs; convinced that the tools to manage MODS would soon be developed and adopted, the working group emphasized that more robust, structured schema.⁵⁷

TDL used the MODS application profile as the basis for Vireo, which, in addition to supporting ingestion, verification, and publication of ETDs, would channel institutions into the production of more consistent metadata.⁵⁸ Vireo was—and is—typically used to generate deposits into the popular open-source repository platform DSpace, where materials are published, embargoed, stored, and integrated into preservation systems. But Vireo's application was limited by its reliance on MODS, DSpace did not natively support; in DSpace, MODS files were inactionable, stored as bitstreams. TDL had developed both the standard and the tool needed to support the standard, but the shared ETD metadata repository, based in DSpace, operated primarily on flat Dublin Core. In practice, then, active metadata did not align with either the TDL or NDLTD standards.

In 2008, TDL organized a new working group to address the gap between the MODS application profile and members' increasing reliance on DSpace and to bring TDL metadata into compliance with NDLTD. The resulting guidelines focused primarily on the bibliographic elements needed to support aggregation of ETD metadata among the various member institutions.⁵⁹ While continuing to position MODS as the canonical schema, the 2008 standard introduced mappings to ETD-MS, expressed as Dublin Core and a flat "thesis" extension. The guidelines dictated that the DC schema mappings "are provided only to assist participants in meeting DSpace requirements, and are not a recommendation to provide qualified Dublin Core as the primary descriptive metadata schema."⁶⁰ New, too, in the 2008 guidelines were explicit references to the metadata that would be generated by or necessitated by DSpace.⁶¹

Over the intervening six years, Vireo was further developed for the management of complex submission and approval workflows. Interestingly, metadata related to dates began to proliferate, in violation of TDL guidelines. When, for example, a student clicked through to approve a license, that action and date were stored as metadata and included as part of the item record upon publication. The tool increased the number of dates generated and retained during the student's submission of the ETD, its approval by committee members, graduate offices, and other required stakeholders, and its ingest into the institution's digital repository. Some of these fields provide supplemental information that can aid the ETD curation process, including the

student's submission date, the student's license agreement date, the approval date from the student's committee chair, and embargo beginning and end dates. Vireo generates this metadata, which is largely administrative in nature. Additionally, several institutions that used Vireo observed that the tool had, at some point, stopped generating MODS files and had changed the way that date fields were populated.⁶² Vireo's metadata output, consistent as it was, constituted a de facto standard.

In 2014, in recognition of a growing divergence between its tool and standard, TDL commissioned a new working group to analyze and update the standard. Guidelines issued in 2015, while continuing to emphasize descriptive metadata, were increasingly attentive to lifecycle concerns and advocated for more robust technical, administrative, and preservation metadata.⁶³ As with the 2008 iteration of the standard, the 2015 guidelines advocated for repository-neutrality while tailoring recommendations to DSpace.⁶⁴ In a departure from the 2005 and 2008 guidelines, these guidelines did not include a MODS application profile.⁶⁵ Continued work is underway to align Vireo development with these new standards.

The tools applied over the course of an ETD's lifecycle are not neutral: they were developed by particular groups, with specific use cases, stakeholders, and goals in mind. These tools, which may be influenced by divergent metadata or stewardship philosophies or reflect design decisions made by those who commissioned, built, or guided their development, constrain and shape ETD metadata. In instances where formal standards proved an awkward fit with available tools, we have observed the development of displacing de facto standards, which complicate existing concerns around interoperability.

Conclusion

In this paper, grounded in literatures on metadata quality, interoperability, and standards, we have coupled research into the history of ETDs and the recent evolution of the Texas Digital Library's ETD standard and tool with close readings of institutional metadata records and a meta-analysis of ETD standards. In so doing, we have sought to initiate a conversation around the generation, maintenance, and evolution of ETD metadata. Our findings highlight distinctions between ETD metadata standards—and the philosophies and goals that underpin these standards—and provide insight into the ETD metadata produced at NDLTD institutions. This exercise has identified a proliferation of fields, without standard definitions, whose interpretation requires close human intervention. Given the erosion of meaningfulness that accompanies diverse and sometimes dissonant metadata standards and practices, we need ways

for dates to “speak” and relay their meaning. Possibilities include (1) implementing clearer field and display labels in repository user interfaces; (2) adding clarifying comments in OAI exports; (3) making institutional application profiles more clearly accessible; (4) developing narratives around dates and placing them in description elements; (5) integrating meaningful local fields that are crosswalked into DC, ETD-MS, or other namespaces; and (6) adjusting existing schema and standards to incorporate commonly used or needed date fields.

We have emphasized that our examination of metadata practices serves as a snapshot. Larger-scale or longitudinal investigations are needed to establish statistical significance, which could inform data-driven decisions around the variety, meaningfulness, and interoperability of dates we capture.

Our analysis has shown that ETD metadata has been shaped by forces related to differing philosophies of metadata and the tools and systems that frequently assist in the process of acquiring, managing, and disseminating ETDs. Dominant standards have emphasized a union catalog model, with descriptive metadata as the basis for federated search. ETD-MS is a lean exchange standard that serves as the basis for the NDLTD union catalog; the standard was formulated as “a flexible set of guidelines for encoding and sharing very basic metadata regarding ETDs among institutions.”⁶⁶ Institutions seeking to optimize the management and description of ETDs must look to more robust standards and models, from which ETD-MS metadata can be derived. We hope, here, to have provided an argument toward a lifecycle metadata model—a model more attuned to the long-term curation of these unique digital objects.

References and Notes

1. The authors refer to both theses and dissertations as “theses” or ETDs (electronic theses and dissertations).
2. Daniel Pargman and Jacob Palme, “ASCII Imperialism,” in *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life*, edited by Martha Lampland and Susan Leigh Star (Ithaca: Cornell University Press, 2009), 191.
3. See William Y. Arms et al., “A Spectrum of Interoperability: The Site for Science Prototype for the NSDL,” *D-Lib Magazine* 8, no. 1 (January 2002), accessed August 7, 2015, www.dlib.org/dlib/january02/arms/01arms.html; As Witten, Bainbridge, and Nichols note, interoperability is both created and thwarted by sociotechnical forces, networked systems “did not come about by accident; they required the development of common understandings about the nature of data formats. It is these communication protocols that allow the reach of digital libraries to extend across our networked world and to interoperate”; Ian H. Witten, David Bainbridge, and David M. Nichols, *How to Build a Digital Library* (Burlington, MA: Morgan Kaufmann, 2010), 343. The authors observe: “Creating and sharing quality metadata is not a straightforward task. . . . Although the technologies may be relatively simple, this is only a *necessary* condition for success, and without the associated human support it will not be sufficient” (350).
4. Mary S. Woodley, “Crosswalks, Metadata Harvesting, Federated Searching, Metasearching: Using Metadata to Connect Users and Information,” in *Introduction to Metadata*, 2nd edition, ed. Murtha Baca (Los Angeles: Getty Research Institute, 2008): 38–62; Caplan described union catalogs as enabling search interoperability between multiple sets of records, whether through the ongoing compilation of a stable “central searchable catalog,” the maintenance of a unified union catalog from which local records are copied, or the generation of a “pseudo-union catalog” that builds a virtual index over multiple sets of records. See Priscilla Caplan, *Metadata Fundamentals for All Librarians* (Chicago: American Library Association, 2003), 34.
5. As a press release explains, the NDLTD Global ETD Search, launched in July 2015, “allows researchers to find ETDs based on keyword, date, institution, language and subject.” “NDLTD Announces Global Electronic Thesis and Dissertation Search,” (July 6, 2015), accessed August 7, 2015, www.ndltd.org/events/news/ndltdannouncesglobalelectronicthesisanddissertationsearch.
6. Woodley indicated several concerns with union catalog implementation, including interoperability issues, concerns that aggregated records were presented with inadequate local context, and variability in how service and data providers expose or add value to aggregated metadata. See Woodley, “Crosswalks, Metadata Harvesting, Federated Searching, Metasearching,” 46–48. In the ETD context, institutions that host ETDs are data providers, while the NDLTD, as an aggregator of these documents’ metadata, functions as a service provider. As Arms noted more than 15 years ago, “Full text and fielded searching are both powerful tools, and modern methods of information retrieval often use the techniques in combination.” William Y. Arms, *Digital Libraries*, digital edition (Cambridge: MIT Press, 2000), www.cs.cornell.edu/way/diglib/MS1999/index.html.
7. PREMIS Editorial Committee, *PREMIS Data Dictionary for Preservation Metadata, version 3* (Library of Congress, November 2015), www.loc.gov/standards/premis/v3/premis-3-0-final.pdf.
8. Richard Pearce-Moses, “Authenticity,” in *Glossary of Archival and Records Terminology* (Society of American Archivists, 2005), accessed December 16, 2015, www2.archivists.org/glossary/terms/a/authenticity; Richard Pearce-Moses, “Reliability,” in *Glossary of Archival and Records Terminology*, accessed December 16, 2015, www2.archivists.org/glossary/terms/r/reliability.
9. Edward A. Fox, “Preface,” in *Electronic Theses and Dissertations: A Sourcebook for Educators, Students, and Librarians*,

- ed. Edward A. Fox et al. (New York: Marcel Dekker, 2004), iii–viii.
10. Communication of drafts and training in preparing electronic documents were also cited as advantages. See Gail McMillan, “Electronic Theses and Dissertations: Merging Perspectives,” co-published simultaneously in *Cataloging & Classification Quarterly* 22, no. 3–4 (1996): 105–25 and in *Electronic Resources: Selection and Bibliographic Control*, edited by W. Pattie Ling-yuh and Bonnie Jean Cox (Philadelphia: Haworth, 1996), 105–25.
 11. McMillan recalled that when the Associate Dean of the Graduate School at Virginia Tech approached the library about access to ETDs, they “presented several reasons for providing students with the opportunity to prepare electronic dissertations and over time several more reasons became clear.” First on the list: “Greater freedom for authors to demonstrate creatively the result of their independent research” (106). McMillan, “Electronic Theses and Dissertations.”
 12. Early in the ETD movement, scholars of the movement assessed this expressive potential and predicted the rise of enhanced ETDs as a new genre, with the potential, for authors of truly technologically advanced works, to secure a hiring advantage. See John L. Eaton, “Enhancing Graduate Education Through Electronic Theses and Dissertations” and Seth Katz, “Innovative Hypermedia ETDs and Employment in the Humanities,” both in *Electronic Theses and Dissertations: A Sourcebook for Educators, Students, and Librarians*, edited by Edward A. Fox et al. (New York: Marcel Dekker, 2004), 1–7. Matthew Kirschenbaum advocated for expression by differentiating between ETDs that presented no expressive advantage over print (the “plain vanilla” ETD—which “need avail itself of no method or presentation or organization that could not be duplicated on paper”) and a “multigraphic” thesis and dissertation—which, “*self-conscious of its medium* . . . uses the electronic environment to support scholarship that could not be undertaken in print.” See Matthew G. Kirschenbaum, “From Monograph to Multigraph: Next Generation Electronic Theses and Dissertations,” in *Electronic Theses and Dissertations: A Sourcebook for Educators, Students, and Librarians*, edited by Edward A. Fox et al. (New York: Marcel Dekker, 2004), 19–32 (italics original).
 13. McMillan, “Electronic Theses and Dissertations,” 124.
 14. McMillan related the influences of humanities computing and electronic text efforts in the humanities and pointed to work done at the University of Virginia’s Electronic Text Center, the Text Encoding Initiative, and Annelies Hoogcarpsel’s *Guidelines for Cataloging Monographic Electronic Texts at the Center for Electronic Texts in the Humanities*, Technical Report No. 1, Center for Electronic Texts in the Humanities, Rutgers and Princeton Universities, 1994, as sources and influences. McMillan, “Electronic Theses and Dissertations,” 108.
 15. For a comprehensive overview of this literature, see Rebecca L. Lubas, “Defining Best Practices in Electronic Thesis and Dissertation Metadata,” *Journal of Library Metadata* 9, no. 3–4 (2009): 252–63, <http://dx.doi.org/10.1080/19386380903405165>; “Chapter 6 Theses/Dissertations and ETD Cataloging: An Annotated Bibliography,” *Technical Services Quarterly* (2008): 95–135, <http://dx.doi.org/10.1080/07317130802127934>. See also Sevim McCutcheon, “Basic, Fuller, Fullest: Treatment Options for Electronic Theses and Dissertations,” *Library Collections, Acquisitions, & Technical Services* 35, no. 2–3 (2011): 64–68, <http://dx.doi.org/10.1016/j.lcats.2011.03.019>; Margaret Beecher Maurer, Sevim McCutcheon, and Theda Schwing, “Who’s Doing What? Findability and Author-Supplied ETD Metadata in the Library Catalog,” *Cataloging & Classification Quarterly* 49, no. 4 (2011): 277–310, <http://dx.doi.org/10.1080/01639374.2011.573440>; Michael Boock and Sue Kunda, “Electronic Thesis and Dissertation Metadata Workflow at Oregon State University Libraries,” *Cataloging & Classification Quarterly* 47, no. 3–4 (2009): 297–308; McMillan, describing pioneering work done at Virginia Tech to develop bibliographic standards for ETDs, notes that the library “[pressed] the Graduate School to require authors of theses (in all formats) to provide keywords for use in the bibliographic record. Cataloging had been impressed for years with how labor-intensive was the task of assigning LC subject headings, so have the authors assign (even uncontrolled) subject headings was an appropriate alternative” (110).
 16. Lubas, “Defining Best Practices in Electronic Thesis and Dissertation Metadata,” 255.
 17. As a press release explained, the NDLTD Global ETD Search, launched in July 2015, “allows researchers to find ETDs based on keyword, date, institution, language and subject.” Networked Digital Library of Theses and Dissertations, “NDLTD Announces Global Electronic Thesis and Dissertation Search,” July 6, 2015, www.ndltd.org/events/news/ndltd-announces-globalelectronicthesisanddissertationsearch.
 18. Joan M. Repp and Cliff Glaviano, “Dissertations: A Study of the Scholar’s Approach,” *College & Research Libraries* 48, no. 2 (March 1987): 148–59.
 19. McMillan, “Electronic Theses and Dissertations,” 110.
 20. Repp and Glaviano, “Dissertations,” 149. Indexes such as the *Dissertation Abstracts International* program, American Doctoral Dissertations, University Microfilm International played a significant role in attempting to collate and provide intellectual access to dissertations for these “extramural” users.
 21. Jane Greenberg, “Metadata and the World Wide Web,” in *Encyclopedia of Library and Information Science*, 2nd ed., edited by Miriam A. Drake. See also Jane Greenberg, “Theoretical Considerations of Lifecycle Modeling: An Analysis of the Dryad Repository Demonstrating Automatic Metadata Propagation, Inheritance, and Value System Adoption,” *Cataloging & Classification Quarterly* 47, no. 3–4 (2009): 380–402.
 22. McMillan, “Electronic Theses and Dissertations,” 110.

23. Elizabeth Shepherd and Geoffrey Yeo, *Managing Records: A Handbook of Principles and Practice* (London: Facet, 2002), 5.
24. Greenberg, "Theoretical Considerations of Lifecycle Modeling," 385.
25. *Ibid.*, 398.
26. While the Digital Curation Center's Curation Lifecycle Model is one of the more popular frameworks referenced in the professional literature, it is not the only data lifecycle model available to researchers. For more information on other types and examples of data lifecycle models, see: CEOS Working Group on Information Systems and Services, Data Stewardship Interest Group, "Data Life Cycle Models and Concepts, Version 1.0," 2011, accessed August 7, 2015, <http://wgiss.ceos.org/dsig/whitepapers/Data%20Lifecycle%20Models%20and%20Concepts%20v8.docx>; Bill LeFurgy, "Life Cycle Models for Digital Stewardship," *The Signal: Digital Preservation* (Washington, DC: Library of Congress, 2012), accessed August 7, 2015, <http://blogs.loc.gov/digitalpreservation/2012/02/lifecycle-models-for-digital-stewardship/>. Daniel Alemneh et al., *Guidance Documents for Lifecycle Management of ETDs*, edited by Matt Schultz, Nick Krabbenhoft, and Katherine Skinner (Atlanta: Educopia Institute, 2014), vii. Completed in 2014, the *Guidance Documents* was an IMLS-funded project to frame the roles that various stakeholders play, including students, faculty, administrators, technologists, commercial vendors, and librarians, in confronting the "administrative, legal, and technical challenges presented by ETDs—from submission to long-term preservation" (i). As such, the document extends beyond metadata standards.
27. Alemneh et. al., *Guidance Documents*, viii.
28. Hussein Suleman et al., "Networked Digital Library of Theses and Dissertations," in *Electronic Theses and Dissertations: A Sourcebook for Educators, Students, and Librarians*, edited by Edward A. Fox et al. (New York: Marcel Dekker, 2004), 59.
29. Networked Digital Library of Theses and Dissertations, "Metadata, ETD-MS v1.1: an Interoperability Metadata Standard for Electronic Theses and Dissertations," edited by Thom Hickey, Ana Pavani, and Hussein Suleman, accessed August 7, 2015, www.ndltd.org/standards/metadata.
30. British Library, "The EThOS UKETD_DC application profile," accessed August 5, 2015, http://ethos toolkit.cranfield.ac.uk/tiki-index.php?page=The%20EThOS%20UKETD_DC%20application%20profile. EThOS documentation acknowledges the complexities associated with capturing dates: "Now your repositories might record a submission date, an award date, a digitisation date and/or a publication date and everyone may use different data and fields to record the information. EThOS currently records just one date: the date the thesis was awarded. In future there may be a case for introducing a second date field, to distinguish award date from publication or submission date for example" ("EThOS").
31. Alemneh et al., *Guidance Documents*, 122–23, 130, 133.
32. Database Management and Standards Committee, OhioLINK, "Standards for Cataloging Electronic Theses and Dissertations—Remote Electronic Version (non-Reproduction)," 2014, accessed August 7, 2015, <https://platinum.ohiolink.edu/dms/catstandards/ETD-RDA.pdf>.
33. Texas Digital Library, "Report for Texas Digital Library Description Metadata for Electronic Theses and Dissertations," v. 2, in *Texas Digital Library Descriptive Metadata Guidelines for Electronic Theses and Dissertations*, version 2.0 (September 2015), <http://hdl.handle.net/2249.1/68437>.
34. Thèses Électroniques Françaises, "Les métadonnées des thèses électroniques françaises," second ed., March 2006, www.abes.fr/abes/documents/tef/recommandation/index.html. The authors used Google Translate to translate TEF documentation into English.
35. The literature on metadata quality intersects with a diverse array of topics, including quality control and assessment (see Daniel Gelaw Alemneh, "Metadata Quality Assessment: A Phased Approach to Ensuring Long-term Access to Digital Resources," *Proceedings of the American Society for Information Science and Technology* 46, no. 1 (2009): 1–8, <http://dx.doi.org/10.1002/meet.2009.1450460380>; David Bade, "The Perfect Bibliographic Record: Platonic Ideal, Rhetorical Strategy or Nonsense?," *Cataloging & Classification Quarterly* 46, no. 1 (2008): 109–33; Yen Bui and Jung-ran Park, "An Assessment of Metadata Quality: A Case Study of the National Science Digital Library Metadata Repository," *iDEA: Drexel Libraries E-Repository and Archives*, 2006, <https://idea.library.drexel.edu/islandora/object/idea%3A1600>; Diane Hillmann, "Metadata Quality: From Evaluation to Augmentation," *Cataloging & Classification Quarterly* 46, no. 1 (2008): 65–80), continuing education (see Jung-ran Park, Yuji Tosaka, Susan Maszaros, and Caimei Lu, "From Metadata Creation to Metadata Quality Control: Continuing Education Needs Among Cataloging and Metadata Professionals," *Journal of Education for Library and Information Science* 51, no. 3 (2010): 158–76), revising legacy metadata (see R. Niccole Westbrook et al., "Metadata Clean Sweep: A Digital Library Audit Project," *D-Lib Magazine* 18, no. 5–6, <http://dx.doi.org/10.1045/may2012-westbrook>; Santi Thompson and Annie Wu, "Metadata Overhaul: Upgrading Metadata in the University of Houston Digital Library," *Journal of Digital Media Management* 2, no. 2 (2013): 137–47), and using automation to evaluate metadata quality (see Dongwon Lee, "Practical Maintenance of Evolving Metadata for Digital Preservation: Algorithmic Solution and System Support," *International Journal on Digital Libraries* 6, no. 4 (2007): 313–26, <http://dx.doi.org/10.1007/s00799-007-0014-9>; Mark Phillips, "Metadata Analysis at the Command-Line," *code4lib Journal* 19 (2013), <http://journal.code4lib.org/articles/7818>).
36. Jung-Ran Park, "Metadata Quality in Digital Repositories: A Survey of the Current State of the Art," *Cataloging &*

- Classification Quarterly* 47 (2009): 213-228, <http://dx.doi.org/10.1080/01639370902737240>.
37. *Ibid.*, 220.
 38. *Ibid.*, 219. Park notes that several factors impact an object's completeness. For example, local metadata rules and requirements that dictate whether elements are required or optional frequently determine the extent to which record creators utilize certain fields to describe objects.
 39. *Ibid.*, 220.
 40. *Ibid.*, 221.
 41. *Ibid.*
 42. DuraSpace, "Metadata Recommendations," *DSpace 4.X Documentation*, 2014, accessed August 7, 2015, <https://wiki.duraspace.org/display/DSDOC4x/Metadata+Recommendations>.
 43. Timothy W. Cole and Myung-Ja K. Han, *XML for Catalogers and Metadata Librarians* (Santa Barbara, CA: ABC-CLIO, 2013).
 44. Kyle Rinkus and Thomas Having, "Medusa at the University of Illinois at Urbana-Champaign: A Digital Preservation Service Based on PREMIS," (paper, Joint Conference on Digital Libraries, July 22–26, 2013, Indianapolis, Indiana), www.ideals.illinois.edu/bitstream/handle/2142/45232/p49.pdf?sequence=2.
 45. DSpace is not the only system for ETD management. Other popular alternatives include ETD-MS, EPrints, Fedora, and homegrown platforms.
 46. Martin Dillon, "Prefatory Commentary from the Editor," in Rebecca Guenther, "MODS: The Metadata Object Description Schema," *portal: Libraries and the Academy* 3, no. 1 (2003): 137–50.
 47. Lubas, "Defining Best Practices in Electronic Thesis and Dissertation Metadata," 255–57.
 48. *Ibid.*, 257.
 49. DSpace JIRA card, "'dc.date.issued' is often incorrectly set (reported from Google)," created February 8, 2013, last modified May 8, 2015, DSpace website, <https://jira.duraspace.org/browse/DS-1481>.
 50. Richard Rodgers, comment on DSpace JIRA card, "'dc.date.issued' is often incorrectly set (reported from Google)," February 8, 2013, accessed August 7, 2015, <https://jira.duraspace.org/browse/DS-1481>.
 51. DuraSpace, "Metadata Recommendations."
 52. "DS-1822: Find a way to report on existing, possibly inaccurate 'dc.date.issued' values," Duraspace website, accessed August 16, 2015, <https://jira.duraspace.org/browse/DS-1822>.
 53. While this discussion focuses on DSpace, Vireo offered a compounding factor in the confusion over the date field. Vireo had been altered to include the "date of approval" as dc.date.issued, shifting the meaning of the values in the element without clear explanation.
 54. The 2008 TDL guidelines specify that mods:dateIssued / dc.date.issued should be filled with the publication date, defined as "the date the ETD is released to the public." The guidelines note: "This date is automatically generated by DSpace upon ingest and does not need to be encoded prior to ingest." Texas Digital Library, "Descriptive Metadata Guidelines for Electronic Theses and Dissertations."
 55. Brian E. Surratt, "MODS Meets Manakin: Innovations in the Texas Digital Library's Thesis and Dissertation Collection," (paper, 9th International Symposium on Electronic Theses and Dissertations, Quebec City, Canada, June 7–10, 2006), accessed April 22, 2015, http://docs.ndltd.org/dspace/bitstream/2340/668/1/SP6_Brian_SURRATT.pdf.
 56. Texas Digital Library Metadata Working Group, "MODS Application Profile for Electronic Theses and Dissertations," Version 1 (December 2005), accessed November 4, 2015, www.tdl.org/wp-content/uploads/2009/04/etd_mods_profile.pdf; Alisha Little (University of Texas at Austin), Anne Mitchell (University of Houston), and Jason Thomale (Texas Tech University) were members of the 2005 working group. Brian E. Surratt (Texas A&M University) chaired the working group.
 57. Surratt, "MODS Meets Manakin," 1–3.
 58. Adam Mikeal et al., "Developing a Common Submission System for ETDs in the Texas Digital Library," July 2007, <http://hdl.handle.net/1969.1/5679>.
 59. Texas Digital Library, "Descriptive Metadata Guidelines for Electronic Theses and Dissertations," Version 1.0, June 2008, www.tdl.org/wp-content/uploads/2009/04/tld-descriptive-metadata-guidelines-for-etc-v1.pdf; Jay Koenig (Texas A&M University), Anne Mitchell (University of Houston), William Moen (University of North Texas), Tim Strawn (University of Texas at Austin), and Jason Thomale (Texas Tech University) were members of the 2008 working group. Amy Rushing (University of Texas at Austin) chaired the working group.
 60. *Ibid.*
 61. Included in the 2005 and 2008 guidelines were dates associated with the creation of the document, derived from both the "date of creation" and the "date of publication," and encoded in MODS as origin information. The 2008 standard mapped both of these dates to ETD-MS as dc.date. Additional date fields included birth dates of authors and thesis committee members. The guidelines accounted for ETD lifecycle events by capturing date information associated with the creation and revision of ETD bibliographic records. "MODS Application Profile for Electronic Theses and Dissertations"; "Descriptive Metadata Guidelines for Electronic Theses and Dissertations."
 62. Texas Digital Library, "Dictionary of Texas Digital Library Description Metadata for Electronic Theses and Dissertations," v. 2, in *Texas Digital Library Descriptive Metadata Guidelines for Electronic Theses and Dissertations*, Version 2.0, September 2015, <http://hdl.handle.net/2249.1/68437>.
 63. Texas Digital Library, *Texas Digital Library Descriptive Metadata Guidelines for Electronic Theses and Dissertations*, version 2.0, September 2015, <http://hdl.handle.net/2249.1/68437>.

.net/2249.1/68437; Santi Thompson (University of Houston), Monica Rivero (Rice University), Kara Long (Baylor University), Colleen Lyon (University of Texas), and Kristi Park (Texas Digital Library) were members of the 2014–15 working group. Sarah Potvin (Texas A&M University) chaired the working group.

64. The *Guidelines* noted, “The authors of these recommendations . . . worked towards an ideal of repository-neutral guidelines. But . . . the constraints of DSpace, and its dominance in the TDL and Vireo User communities, provided an argument for tailoring some recommendations to the known constraints and behavior of DSpace repositories. As Vireo and TDL diversify to incorporate Fedora repositories, greater awareness should be paid to the aspects of the guidelines that are not repository-neutral, and to considering the need to tailor recommendations to Fedora and other repository systems.” Texas Digital Library, “Report for Texas Digital Library Description Metadata for Electronic Theses and Dissertations,” 18.
65. See “A Note on MODS, “Dictionary of Texas Digital Library Description Metadata for Electronic Theses and Dissertations,” 4–5.
66. Suleman et al., “Networked Digital Library of Theses and Dissertations,” 59.