Metadata Standards: Dublin Core + DACS

Dipa Desai and Sara Morrell
HIST 7251: Digital Archives
Prof. Jessica Parr
Spring 2024, February 29th

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

Workshop Agenda

- Metadata use in digital scholarship
- Essentials of metadata
  - Dublin Core Review
  - DACS
  - Class Activity: Northeastern Finding Aids
- Data Mapping

Class materials available at: **https://bit.ly/sp24-parr-hist7251-dacs**

Feel free to ask questions at any point during the presentation!

Metadata Usage in Digital Scholarship

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

Class Discussion: Student experience with metadata

- Where does metadata show up in your work (research, coursework, digital projects, etc.)?
- What formats have you used?
- What standards are you familiar with?

Feel free to ask questions at any point during the presentation!

Review: Key Concepts

Data Interoperability: the ability for data (through formatting and consistency) to be merged, aggregated, or accessed across platforms and data structures.

Relational Databases: databases that store information based on a relational model of data, relying on consistency in structure, constraints, and normalization.

Metadata: is "the information we create, store, and share to describe things" and "allows us to interact with these things to obtain the knowledge we need."

Understanding Metadata by Jenn Riley (2017)

Feel free to ask questions at any point during the presentation!

Review: Key Concepts

Metadata Standards: sets of rules for the creation and implementation of metadata for different purposes, created and published by specific communities, specialized disciplines, or for more broad knowledge domains.

Controlled Vocabularies: standardized and organized arrangements of words or phrases to consistently describe data, typically include subject headings.

- **Taxonomies**: Alphabetical lists of terms arranged within a hierarchical structure for broader and narrower terms.
- **Ontologies**: more specific taxonomies that further describe terms by their relationship to other terms and concepts in a hierarchy.

Feel free to ask questions at any point during the presentation!

Types of Metadata

Dublin Core provides descriptive metadata.

Next workshop we'll talk about **TEI**, an example of a markup language that provides structural metadata.

DACS is a content standard. Content standards are guides to structure textual values in metadata (e.g. abbreviations, capitalization).

Metadata Type	Example Properties	Primary Uses
Descriptive metadata	Title Author Subject Genre Publication date	Discovery Display Interoperability
Technical metadata	File type File size Creation date/time Compression scheme	Interoperability Digital object management Preservation
Preservation metadata	Checksum Preservation event	Interoperability Digital object management Preservation
Rights metadata	Copyright status License terms Rights holder	Interoperability Digital object management
Structural metadata	Sequence Place in hierarchy	Navigation
Markup languages	Paragraph Heading List Name Date	Navigation Interoperability

Source: NISO's Metadata Primer

Feel free to ask questions at any point during the presentation!

Types of Metadata Standards

There are four central kinds of metadata standards based on what they are regularizing and describing for consistency:

- **Structure standards**: sets of metadata elements defined for a structural purpose (known as schemes or schemas).
- **Content standards:** rules for the input data in elements that describe textual data.
- **Value standards**: types of standards that restrain or narrow the possibilities of input for terms to reduce variation (known as controlled vocabularies).
- **Format standards:** technical specifications for encoding metadata for machine readability and processing (known as data formats or encoding standards).

Feel free to ask questions at any point during the presentation!

Review: Dublin Core

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

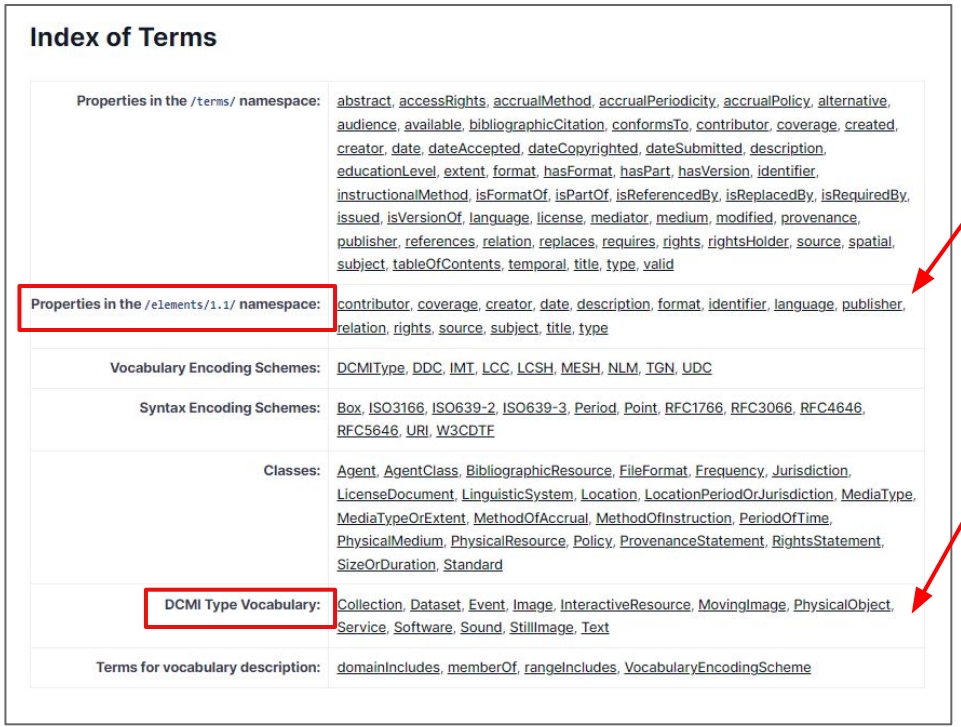
Dublin Core

Dublin Core is an example of a content standard for descriptive metadata for items.

- DCMI stands for the Dublin Core Metadata Initiative and is the **namespace** the Dublin Core and associated classes and further specifications.
- DCMI terms are expressed in RDF vocabularies.

Feel free to ask questions at any point during the presentation!

Navigating the DCMI Specifications



Index of Terms

Properties in the `/terms/` namespace:	abstract, accessRights, accrualMethod, accrualPeriodicity, accrualPolicy, alternative, audience, available, bibliographicCitation, conformsTo, contributor, coverage, created, creator, date, dateAccepted, dateCopyrighted, dateSubmitted, description, educationLevel, extent, format, hasFormat, hasPart, hasVersion, identifier, instructionalMethod, isFormatOf, isPartOf, isReferencedBy, isReplacedBy, isRequiredBy, issued, isVersionOf, language, license, mediator, medium, modified, provenance, publisher, references, relation, replaces, requires, rights, rightsHolder, source, spatial, subject, tableOfContents, temporal, title, type, valid
Properties in the `/elements/1.1/` namespace:	contributor, coverage, creator, date, description, format, identifier, language, publisher, relation, rights, source, subject, title, type
Vocabulary Encoding Schemes:	DCMIType, DDC, IMT, LCC, LCSH, MESH, NLM, TGN, UDC
Syntax Encoding Schemes:	Box, ISO3166, ISO639-2, ISO639-3, Period, Point, RFC1766, RFC3066, RFC4646, RFC5646, URI, W3CDTF
Classes:	Agent, AgentClass, BibliographicResource, FileFormat, Frequency, Jurisdiction, LicenseDocument, LinguisticSystem, Location, LocationPeriodOrJurisdiction, MediaType, MediaTypeOrExtent, MethodOfAccrual, MethodOfInstruction, PeriodOfTime, PhysicalMedium, PhysicalResource, Policy, ProvenanceStatement, RightsStatement, SizeOrDuration, Standard
DCMI Type Vocabulary:	Collection, Dataset, Event, Image, InteractiveResource, MovingImage, PhysicalObject, Service, Software, Sound, StillImage, Text
Terms for vocabulary description:	domainIncludes, memberOf, rangeIncludes, VocabularyEncodingScheme

Elements (1.1) Namespace: the fifteen main elements for Dublin Core.

DCMI Type Vocabulary: the specific vocabulary (controlled) for type for Dublin Core. This (as we see in Omeka) is often expanded and customized by projects.

Feel free to ask questions at any point during the presentation!

Anatomy of a DCMI Element Specification

Each term is specified with the following minimal set of attributes:

Name:	A token appended to the URI of a DCMI namespace to create the URI of the term.
Label:	The human-readable label assigned to the term.
URI:	The Uniform Resource Identifier used to uniquely identify a term.
Definition:	A statement that represents the concept and essential nature of the term.
Type of Term:	The type of term: property, class, datatype, or vocabulary encoding scheme.

Feel free to ask questions at any point during the presentation!

Creator

Vocabulary:	DCMI Metadata Terms
URI	http://purl.org/dc/elements/1.1/creator
Label	Creator
Definition	An entity primarily responsible for making the resource.
Comment	Examples of a Creator include a person, an organization, or a service. Typically, the name of a Creator should be used to indicate the entity.
Type of Term	Property
Note	A second property with the same name as this property has been declared in the dcterms: namespace. See the Introduction to the document DCMI Metadata Terms for an explanation.

Feel free to ask questions at any point during the presentation!

Description

Vocabulary:	DCMI Metadata Terms
URI	http://purl.org/dc/elements/1.1/description
Label	Description
Definition	An account of the resource.
Comment	Description may include but is not limited to: an abstract, a table of contents, a graphical representation, or a free-text account of the resource.
Type of Term	Property
Note	A second property with the same name as this property has been declared in the dcterms: namespace. See the Introduction to the document DCMI Metadata Terms for an explanation.

Feel free to ask questions at any point during the presentation!

Type

Vocabulary:	DCMI Metadata Terms
URI	http://purl.org/dc/elements/1.1/type
Label	Type
Definition	The nature or genre of the resource.
Comment	Recommended practice is to use a controlled vocabulary such as the DCMI Type Vocabulary [DCMI-TYPE]. To describe the file format, physical medium, or dimensions of the resource, use the Format element.
Type of Term	Property
Note	A second property with the same name as this property has been declared in the dcterms: namespace. See the Introduction to the document DCMI Metadata Terms for an explanation.

Feel free to ask questions at any point during the presentation!

DCMI Type Vocabulary: Examples

Dataset

Vocabulary:	DCMI Metadata Terms
URI	http://purl.org/dc/dcmitype/Dataset
Label	Dataset
Definition	Data encoded in a defined structure.
Comment	Examples include lists, tables, and databases. A dataset may be useful for direct machine processing.
Type of Term	Class
Member Of:	http://purl.org/dc/terms/DCMIType

Image

Vocabulary:	DCMI Metadata Terms
URI	http://purl.org/dc/dcmitype/Image
Label	Image
Definition	A visual representation other than text.
Comment	Examples include images and photographs of physical objects, paintings, prints, drawings, other images and graphics, animations and moving pictures, film, diagrams, maps, musical notation. Note that Image may include both electronic and physical representations.
Type of Term	Class
Member Of:	http://purl.org/dc/terms/DCMIType

Feel free to ask questions at any point during the presentation!

Introduction to DACS

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

Society of American Archivists

Society of American Archivists (SAA): the largest and oldest professional association in the United States dedicated to the needs and interests of archivists and archivists. Founded in 1936, the SAA represents over 6,000 professional archivists today.

- Includes SAA Council, a group that creates, develops, and implements a strategic plan for the organization.
- Maintains and facilitates changes to the content standards for archives in the United States.

Feel free to ask questions at any point during the presentation!

Describing Archives: A Content Standard (DACS)

Describing Archives: A Content Standard (DACS): the official content standard for the archival community in the United States since 2005.

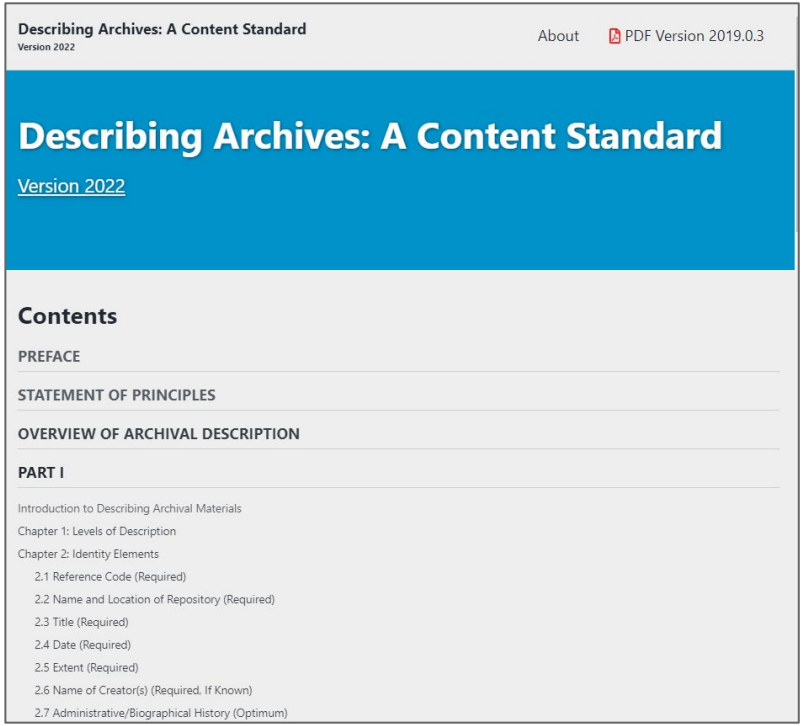
- Designed to be used for a variety of archival descriptions, including finding aids, catalog records, etc.
- Updated in a continuous cycle as needed. Current version and past versions are on the TS-DACS GitHub.

Feel free to ask questions at any point during the presentation!

Accessing DACS

DACS can be accessed on the [TS-DACS GitHub], either in a user-friendly interface or as a collection of files (in markdown) that can be downloaded.

- [DACS (User-friendly) using Jekyll on Github]



Feel free to ask questions at any point during the presentation!

Levels of Description

Archival description exists on different levels, depending on the arrangement and hierarchy of a collection. DACS does *not* define a "proper" level of description. Instead, it contains elements helpful to creating a system as needed including:

- **Single-level Descriptions:** a single record (can vary in content)
- **Multi-level Descriptions:** linked records, detailing an inventory, register, collection, etc. Multi-level records must include at least one sublevel.

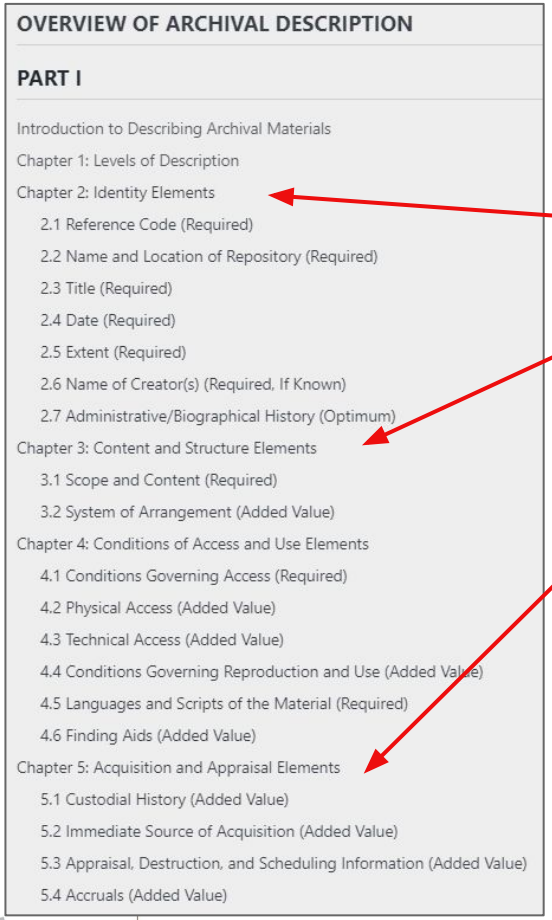
Feel free to ask questions at any point during the presentation!

What does a record that uses DACS look like?

Records that utilize DACS look very different, but contain (at minimum) the following elements for both single-level and multi-level descriptions:

- Reference Code
- Name and Location of Repository
- Title
- Date
- Extent
- Name of Creator(s)
- Scope and Content
- Conditions Governing Access
- Language and Script of the Material
- Rights Statement

Feel free to ask questions at any point during the presentation!



Sections of DACS

- **Identity Elements**
- **Content and Structure Elements**
- Access and Use Elements
- **Acquisition and Appraisal Elements**
- Related Material Elements
- Notes Elements
- Description Control Elements

Feel free to ask questions at any point during the presentation!

Access Points

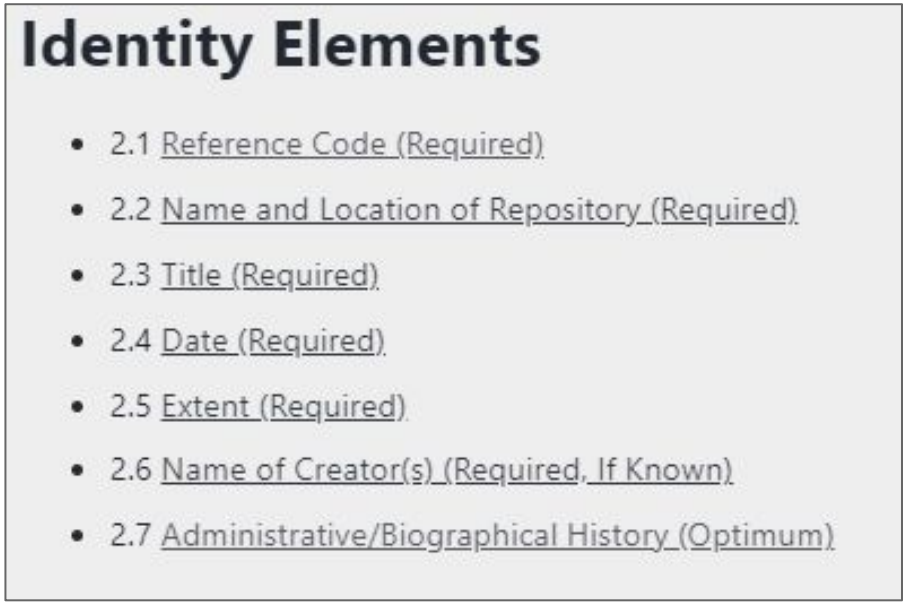
Archival description serves to create *access* to information about materials to make them searchable. This requires ***access points***, or certain forms of coded information in certain fields that tools, regardless of the form of a system, will use for searching.

Access points include:

- Names
- Places
- Subjects
- Documentary Forms
- Occupations
- Functions

Feel free to ask questions at any point during the presentation!

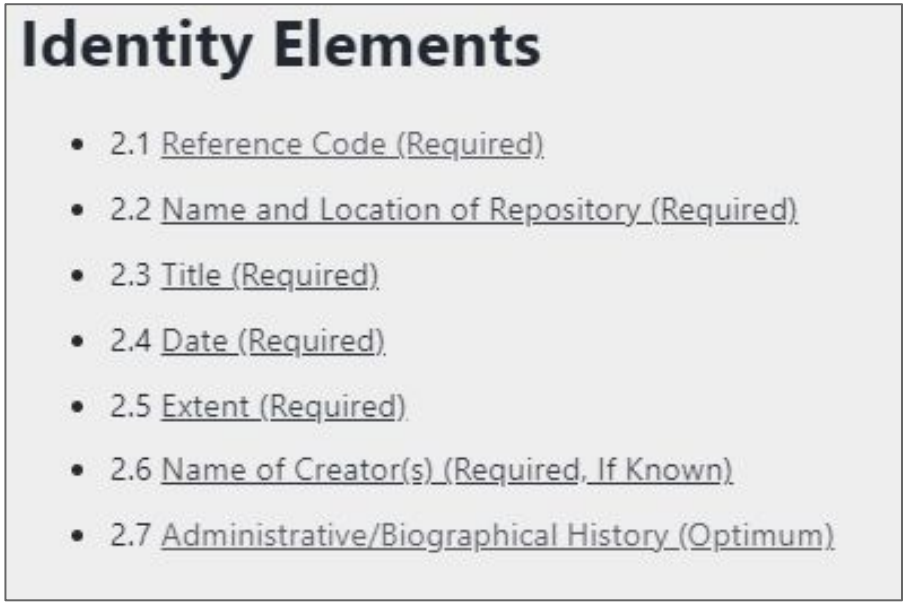
Identity Elements



- **Reference Code**: a unique identifier for the materials described (including unit).
- **Name and Location of Repository**: information about the repository that holds the archival material.
- **Title**: word or phrase for the materials to be identified by.
- **Date:** identifies and records the dates that pertain to the materials described (including creation, assembly, accumulation, etc.)

Feel free to ask questions at any point during the presentation!

Identity Elements



- **Extent:** the physical nature (or extent in linear feet) of materials.
- **Name of Creator(s):** element identifying the corporate bodies, persons, and families associated with the creation, assembling or use of materials.
- **Administrative/Biographical History**: element to describe biographical or historical information about the creators or people described in materials.

Feel free to ask questions at any point during the presentation!

Content and Structure Elements


Chapter 3

Content and Structure Elements

- 3.1 Scope and Content (Required)
- 3.2 System of Arrangement (Added Value)

- **Scope and Content:** element to provide information about the nature of materials and activities reflected in units being described. This includes: functions, documentary forms, content dates, geographic areas, subject matter, etc.
- **System of Arrangement**: information about the organization of an collection, including the arrangement of file, subsections, etc.

Feel free to ask questions at any point during the presentation!

Acquisition and Appraisal Elements

- **Custodial History:** information on changes of ownership or custody of materials.
- **Immediate Source of Acquisition**: element identifying the source from which a repository acquired materials (as well as methods, date, etc.)
- **Appraisal, Destruction, and Scheduling Information**: information about appraisal decisions, any information about material destruction, or retention.
- **Accruals:** information about any additions to materials being described.

Feel free to ask questions at any point during the presentation!

Archival Authority Records

Archival authority records are records (maintained and located in different repositories) that act as a standardizing access point to bibliographical and topical materials. This includes information, most commonly, for names and subjects.

- Library of Congress Name Authority File (LCNAF)
- Library of Congress Subject Headings (LCSH)

Feel free to ask questions at any point during the presentation!

Group Activity: NEU Finding Aids

Now that we have covered the main elements for archival description using DACS, we will look at example records in the Northeastern Archives and Special Collections finding aids.

Access the Northeastern Archive Collections for examples!

Try using keywords, names, or subject headings related to your collection or items that you described in Dublin Core.

- Are there any collections that come up?
- What do you notice about the finding aids?
- Does each finding aid or collection description have the same information?
- What are some observations you have about certain elements?
- Is there anything that confuses you?

Feel free to ask questions at any point during the presentation!

Overview

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

Key Metadata Vocabulary

- **Namespace**: a descriptor for a **container** of information, used in reference with metadata to indicate the specific **vocabulary** a file structure or data system is using and can stand in for a **URI** or **IRI** (uniform resource identifier or internationalized resource identifier).
- **Schema:** the organization or structure for data, a dataset, or database that acts as a model or representation of information.
- **Metadata Languages:** different systems of notation for data structure and the syntax related to these structures and systems (i.e. XML).

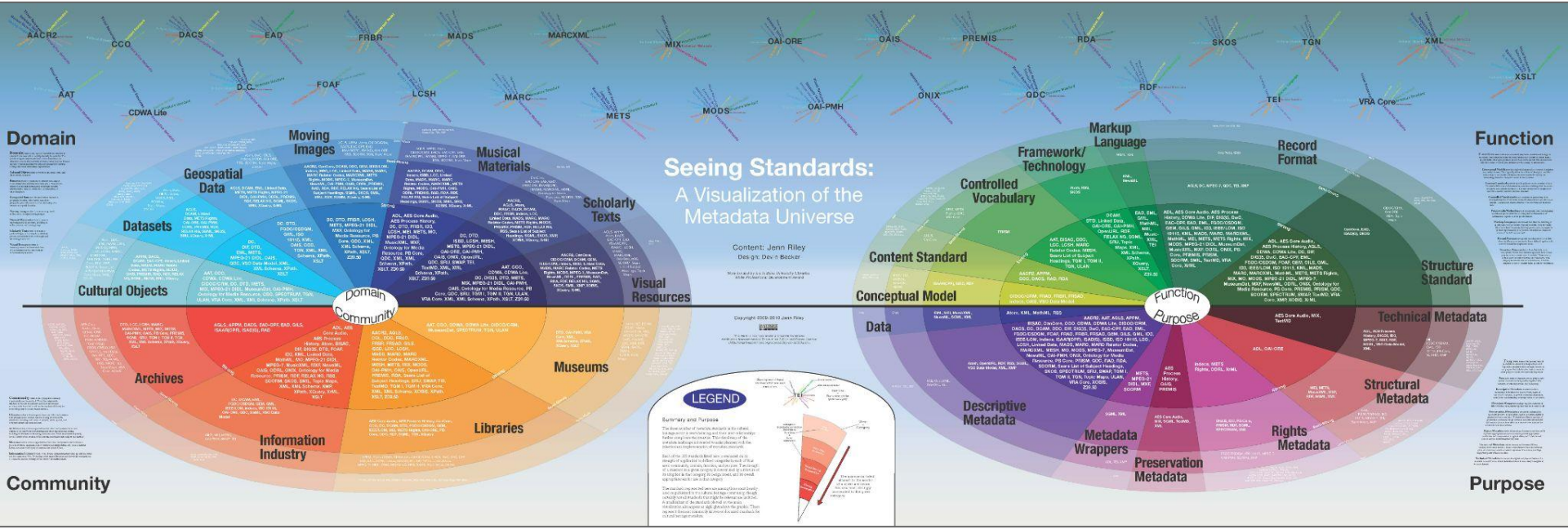
Feel free to ask questions at any point during the presentation!

Key Metadata Vocabulary

- **Resource Description Framework (RDF):** a data model and framework used for data exchange on the web. It includes features that facilitate data merging despite schema differences.
- **eXtensible Markup Language (XML):** a markup language and file format for storing, transmitting, and reconstruing data to be human and machine readable.

Feel free to ask questions at any point during the presentation!

Seeing Standards by Jenn Riley and Devin Becker



https://jennriley.com/metadatamap/

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

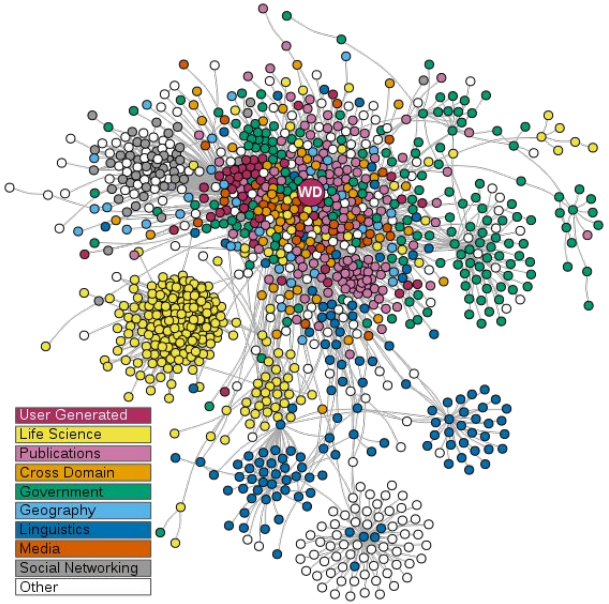
Linked Data

Linked Data is structured data that is interlinked with other data so a person or machine can explore relations in the semantic web of data.

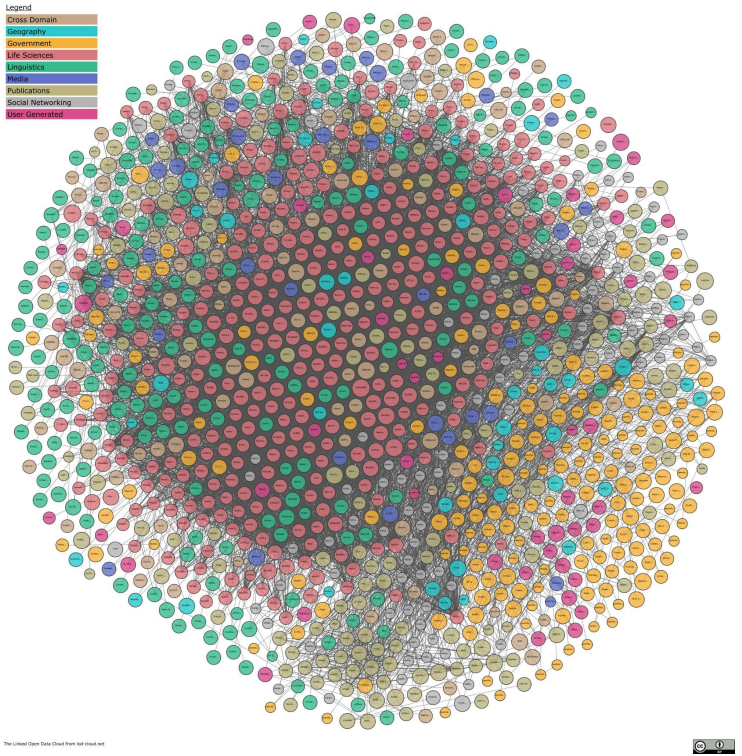
- **Semantic Web** is an extension of the World Wide Web set by the World Wide Web Consortium (W3C) to make internet data machine-readable by encoding **semantics** (meaning and reference with language) in data.

Feel free to ask questions at any point during the presentation!

Linked Data Projects



Wikidata in the [Linked Open Data Cloud.]



Datasets published in Linked Data format from the [Linked Open Data Cloud.]

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

Data Mapping

Northeastern University
NULab for Texts, Maps, and Networks

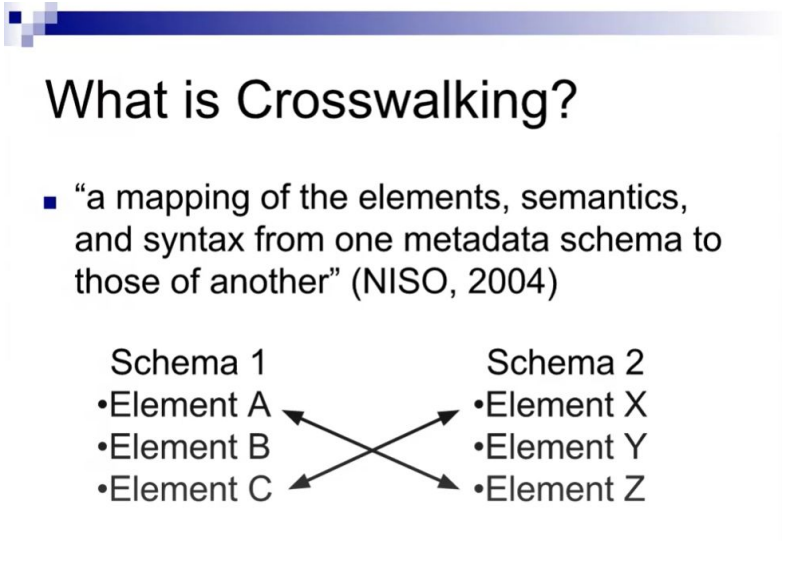
Feel free to ask questions at any point during the presentation!

Metadata Crosswalks

Metadata crosswalks increase data interoperability and discoverability.

They provide data descriptions that map to other, external metadata schema.

Take a look at the DACS Crosswalk: [Crosswalks in DACS (and other archival formats)]



Source: Society of American Archivists' DACS Primer on Youtube

Feel free to ask questions at any point during the presentation!

Example of a Crosswalk

Table 1. Example of Metadata Crosswalk Mapping

	Dublin Core	EAD	MARC 21
Title Element	Title	<titleproper>	245 00$a (Title Statement/Title proper)
Author Element	Creator	<author>	700 1#$a (Added Entry--Personal Name) (with $e=author) 720$a (Added Entry–Uncontrolled Name/Name) (with $e=author)
Date Created Element	Date.Created	<unitdate>	260 ##$c (Date of publication, distribution, etc.)

Source: NISO's Understanding Metadata

Feel free to ask questions at any point during the presentation!

Mapping Metadata in a Database

When mapping metadata from sources into a digital archive or database, there are several processes you can use:

- One-time import: You can compile mapped metadata into a CSV or Excel file and import metadata at once. This is usually used when transferring contents from one digital system to another.
- Importing by file name: You can use a script in your database management system to automatically add metadata based file naming conventions. Ex. Files named using a Title-Creator-AccessionID convention.

Feel free to ask questions at any point during the presentation!

Mapping Metadata (continued)

- Automatic imports: If you're adding content that has existing metadata, like image files with photo date or location, you can use a script in your database management system to automatically import file-specific metadata and map it to your database.
- API: You can use APIs to periodically collect and import metadata from content management systems like WordPress, and sync it with your database management system.

Automated methods may produce mapping errors. It's good practice to combine automated methods with human verification.

Feel free to ask questions at any point during the presentation!

Metadata Application Profile: DTA

Overview of Metadata Requirements

Field	Requirement	Occurrence	Schema
Institution	Required	Non-repeatable	None
Collection	Required	Non-repeatable	None
Title	Required	Non-repeatable	None
Alternative Title	Required if available	Repeatable	None
Creator	Required if available	Repeatable	None
Contributor	Required if available	Repeatable	None
Date Created	Required if available	Repeatable	Extended Date/Time Format
Date Issued	Required if available	Repeatable	Extended Date/Time Format
Temporal Coverage	Required if available	Repeatable	Extended Date/Time Format
Spatial Coverage	Required if available	Repeatable	GeoNames
Subject Coverage: Homosaurus v2	Required if available	Repeatable	Homosaurus
Subject Coverage: LC Subject Headings	Optional	Repeatable	Library of Congress Subject Headings
Subject Coverage: Other	Optional	Repeatable	None
Explicit Content	Required	Non-repeatable	None
Type	Required	Repeatable	LC Resource Types
Genre	Required	Repeatable	DTA Genre Vocabulary

The Digital Transgender Archive's Metadata Application Profile shows which fields are required and can have more than one occurrence or entry, as well as the data schema used to describe each field.

The Metadata application profile also shows where fields map to Dublin Core (DC) terms, and provides examples of field descriptions.

Feel free to ask questions at any point during the presentation!

Metadata Application Profile: DTA

Subject Coverage: Homosaurus v2

Definition	The topic of the resource. Typically, the subject will be represented using keywords, key phrases, or classification codes.
Requirement	Required if available
Occurrence	Repeatable
Schema	Homosaurus v2
DC Terms	subject
Input Guidelines	• The Homosaurus v2 is required for all input values. • Care should be taken to use the terminology already associated with the resource, when available. • To input a value, select "Lookup" and enter a search term. Select the appropriate bold term from the results. • Do not use the Homosaurus field, which queries an earlier version and is only used for editing existing records. • This element can be repeated multiple times for each resource, with each instance containing separate subject terms. • Use the narrowest terms possible. Broader terms may also be used if equally relevant to the resource.
Examples	• HIV-negative people • Transgender studies • LGBTQ zines

The DTA requires use of the Homosaurus controlled vocabulary to describe the Subject Coverage field.

The Homosaurus project links their vocabulary to the Library of Congress Subject Headings to enhance the discoverability of LGBTQ+ digital resources. It has also been used to improve ontologies about gender and sex.

Feel free to ask questions at any point during the presentation!

DTA Ontology Example

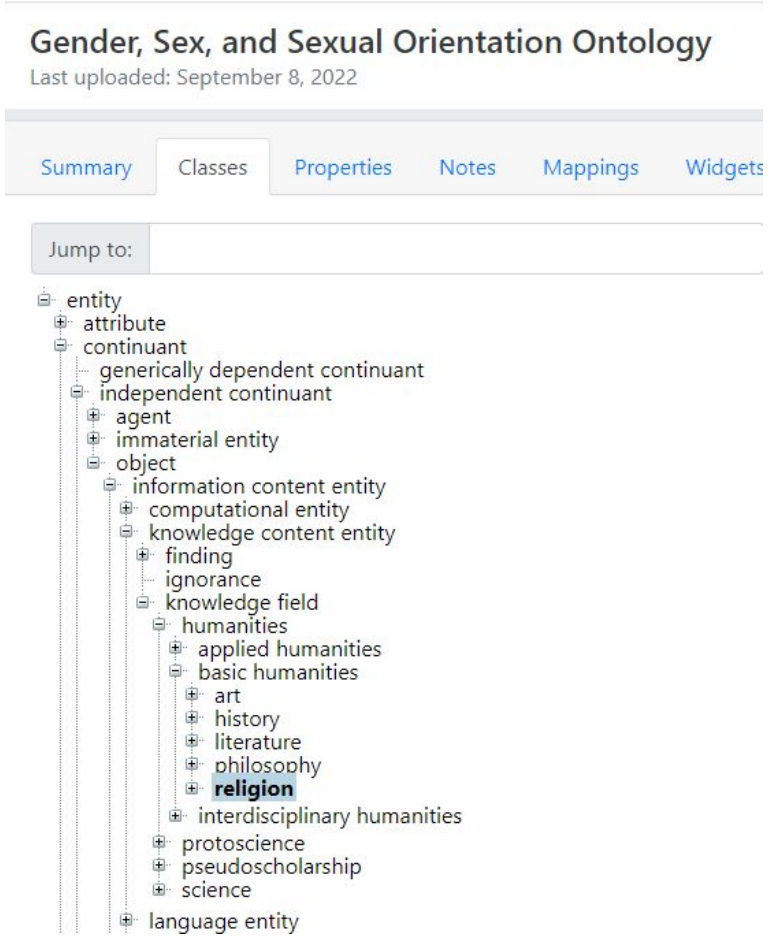
The DTA uses the Gender, Sex, and Sexual Orientation ontology to characterize relationships among terms and concepts in transgender history.

The Gender, Sex, and Sexual Orientation ontology can be mapped to ontologies used in other disciplines, to enhance data discoverability and interoperability.

Feel free to ask questions at any point during the presentation!

DTA Ontology Example

Go to the Gender, Sex, and Sexual Orientation ontology and click on the “Mappings” tab. Choose another ontology (ex. Adverse Childhood Experiences ontology) and click on GSSO terms to see where they overlap with terms from the other ontology hierarchy.



Feel free to ask questions at any point during the presentation!

Mapping Resources

- Society of American Archivists' Crosswalking primer on YouTube
- Crosswalks in DACS (and other archival formats) on Github
- Example of an automated Metadata Mapping model from MidPoint Identity Management
- Example of mapping metadata to Dublin Core for a digital collection from University of Wisconsin-Milwaukee Digital Library Teaching Resource

Feel free to ask questions at any point during the presentation!

Class Activity:
Dublin Core to DACS

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

Your Turn!

Discuss the Dublin Core metadata you used for your Omeka site, or that was used in the "Colored Conventions Project"or "Texts of Taste" sites, with the students near you.

Identify 3–5 fields where your, or the project's, Dublin Core metadata overlaps with the DACS metadata schema for files, items, exhibits, and/or collections you added.

- Are there any Dublin Core metadata fields you used that do not match with DACS metadata?
- Which metadata fields were easy to map to DACS? Which were difficult?
- Which levels of description would you need to map a collection in DC to DACS?

Feel free to ask questions at any point during the presentation!

Questions?

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

Metadata Content Standard Resources

- NISO Metadata Primer
- Seeing Standards Visual Map
- Dublin Core Metadata Terms
- DACS current version and DACS primer videos

Feel free to ask questions at any point during the presentation!

Thank you!

If you have any questions, contact DITI at nulab.info@gmail.com
DITI Office Hours: https://calendly.com/diti-nu/

Developed by: Juniper Johnson, Dipa Desai, Sara Morrell, and Kasya O'Connor Grant

We'd love your feedback! Please fill out a short survey here:
https://bit.ly/diti-feedback

Slides, handouts, and data available at:
https://bit.ly/sp24-parr-hist7251-dacs

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!