Content Standard Principles

LIBR 506
Week 4 Feb 1 2022

1 Lecture

Key questions we're focusing on in this lecture:

1. **What is a content standard?**
2. **What should we describe about a resource?**
    - Which aspects of the resource do we want to represent?
    - What constraints are placed on those representations? (e.g., how detailed can it be, what sources of information are "trustworthy"
    - How much time / expertise does the person doing the description have? (can you trust the person doing the judgements)
3. **What are the minimal agreed-upon attributes to include?**

Central concepts we typically represent in a library: title, creator (person, group, event "responsible" for the resource), and version
Some other concepts that could be employed: what the resource looks like, what type of resource it is (e.g., text, sound recording, etc.), who published it and where, how to identify the resource (unambiguous, unique identifiers—e.g., ISBN, ISSN, DOI), what the resource is about (thesaurus terms, class number)

NB: schemas and standards are not interchangeable; **schemas** might become **standards** when they are regulated, managed, and shared among institutions

1.1 (Authority) Control

When describing items, we are trying to create accurate, reliable versions so they can be used by many institutions, provide accurate depictions of intellectual content

- **Language and consistency:** There can be issues around language and consistency: minor differences in language necessitate creating a repeatable system of representing something in a character set (e.g., an author's name might have changed their name over time, or name appears differently depending on character set; need to be able to identify which version of someone's name to use in the catalogue, but also connect all those forms of the name)
- **Bibliographic relationships:** Also, works created in a library has relationships to other items; info professionals trying to create a network that people can navigate through (some new things:

connecting series information (prequels or sequels of each other, critiques of each other))—how might people be interested in connections between items?

1.2 Abstraction and Specificity

How important is it to make similar resources appear the same? (sometimes important to remove some of the specificity and move toward abstraction—to engage with intellectual content of the resource
But specificity can be important in some areas: e.g., distinguish similar resources from each other (e.g., different edition of a textbook)

1.3 Access

When we create a full summary (surrogate) of an item in the catalogue:

- which elements need to be searchable?
- Filterable? (types might become relevant—e.g., only want text, don't want sound recordings, only want peer reviewed articles)
- What will users need to find, sort, identify, and evaluate a record? How is that summary facilitating all these other tasks someone has in the catalogue? "Do I think this thing is going to serve my information need?"

1.4 International Federation of Library Associations (IFLA) General Principles

Principles: convenience of user, common usage, representation, accuracy, sufficiency and necessity, significance, economy, consistency and standardization, integration, interoperability, openness, accessibility, rationality

- Some of these are in tension with each other (e.g., convenience of user and "economy")

1.5 Example of a content schema: the Dublin Core

Devised in Dublin Ohio in 1995 by people in different domains trying to come up with an agreed upon "base" that will probably be available, and worth describing about an item, regardless of what it is (e.g., biological specimen, dataset, published work)

Goal: easier search and retrieval mechanisms

15 main elements:

General Information	Physical Description	Intellectual Content
Title	Identifier (e.g., ISBN, URI)	Description
Date	Type (e.g., moving image)	Subject (e.g., LCSH, DDC)
Creator	Format (e.g., mpeg)	Coverage (place, time)
Contributor	Language	Relation
Publisher		Source
Rights		

Description of each Dublin Core elements is very vague—applicable to wide array of contexts

Examples of Dublin Core as applied:

- UBC: https://circle.ubc.ca/handle/2429/46007?show=full
- Museum example: http://dlib.indiana.edu/omeka/mathers/items/show/221

Dublin Core is meant to allow specific collections to surface specific information their users might be interested in; does through **qualifiers**

- E.g., “Date” is very vague; might need date digitized, date modified, date available, etc.

Dublin Core Metadata Initiative (DCMI) built up large array

1.6 Metadata and Cataloguing

Both are information about a potentially informative resource, many parallels between these approaches Arguably, catalogue records are a kind of metadata

- Most **metadata** is created specifically to describe digital content
    - Challenges that come with digital content: changeability / mutability, need for version control; more dynamic properties than for a print edition; metadata often employed to create control on access, e.g., create copyright restrictions

1.7 Choosing a Content Standard / Schema

- Balance between functionality and simplicity (how many elements?)
- Support both human and machine use (readable by a human, but interpretable by a machine—these are often at tension, finding balance)
- Supporting **interoperability** (metadata from one institution readable by another; translations, crosswalks) and **extensibility** (expectation that we won’t know future information needs that this will have to support; schema should be flexible enough that we can adapt over time to suit needs of users—can be simplified, or made more specific)

1.8 Summary

Each schema has

- A set of values
- Some instructions on which elements are necessary
- Some instructions on how to modify elements
- Some instructions on how to fill out values
    - Possibly a reference to controlled vocabularies for particular fields

2 The Discipline of Organizing, Section 5.1, 5.3

Resource descriptions and metadata provide meaning—but what is meaningful can depend on context, and may change over time. (e.g., artwork titles)

Some important questions when it comes to describing resources:

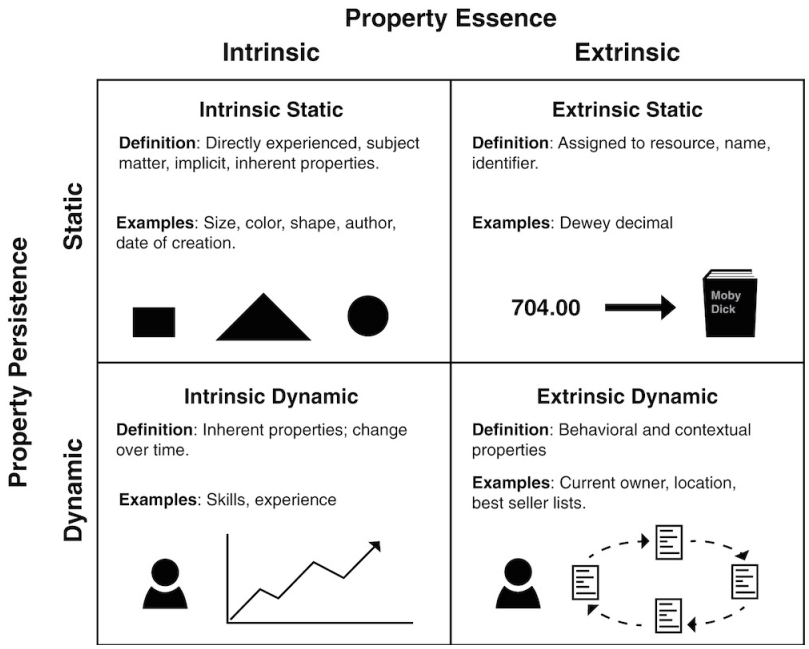
- What is the purpose of resource description?
- What properties should be described?
- How to create good resource descriptions?
- What makes a good resource description?

7 interdependent and iterative steps to describing resources:

1. **Determine scope and focus** (and think about how **granular**[1] and how **abstract**[2] descriptions should be)[3]
    - "One person's metadata is another person's data"
    - Granularity: what should be treated as a resource versus a collection of resources? Granularity might differ for different users or purposes
    - Larger collections need more complex descriptions to differentiate resources from each other; as collection grows, descriptions might need to be revised
        - **Dublin Core:** metadata element set within only 15 elements; simpler description vocabulary for non-professionals: Contributor, coverage, creator, date, description, format, identifier, language, publisher, relation, rights, source, subject, title, type
        - Tension between complexity and loss of precision (e.g., "creator" vs "artist, composer, author, etc.")
        - Also tension between investing work / effort in an extensible system vs only doing what is necessary now
2. Think about **purpose** of description (what interactions to support)
    - Selecting: e.g., product marketing (prunes / plums = dietary supplement / snack food)
    - Organizing
    - Interacting: e.g., finding, identifying, selecting, obtaining, navigation / exploring
    - Maintaining: e.g., version numbering, useful life information, equipment maintenance schedule
    - Resource description for sensemaking
3. **Identify properties** to encode in descriptions (think about robustness and reliability)

    Can think about 4 types of properties: **essence** (intrinsic or extrinsic) and **persistence** (static or dynamic)

[1] Describe resources, or collections of resources?
[2] Describe resource instances, parts of them, or resource types?
[3] Focus: which resource is the focus of our attention? (e.g., the primary resource? The description resource?)



4. **Design vocabulary** of description (what words or values will represent the properties?). Svenonius (2000) proposes principles of good description: **user convenience** (commonly used among target audience), **representation** (should somehow reflect resource themselves), **sufficiency and necessity** (enough info to serve purposes, but none that isn't necessary), **standardization**, **integration** (consistency across resources)
    - Keeping the users front of mind can help when tensions or conflicts arise in these best practices
    - **Controlled vocabulary**
    - **Content rules:** e.g., specific data type, can be limited by logical expressions
5. Design **form** and **implementation** of the description
6. **Create** resource description (could be group / individual, formal / informal, person / computer)
    - Each scenario of who is creating description (e.g., author, user, professional, computer) has strengths and tradeoffs
    - **Semantic gap:** difference between automated and human description
7. **Evaluate** description: does it support intended purposes? Might need to iterate through steps
    - Many different criteria could be used; e.g., accuracy, completeness, consistency, timeliness, interoperability, usability (might conflict with each other)