Metadata:
A Love Letter to the Future

MAGGIE DULL, DIRECTOR OF METADATA STRATEGIES
RIVER CAMPUS LIBRARIES
Overview

What is metadata?
What can we do with metadata?
How can we create metadata?
How can we share metadata?
Why does metadata matter?
What is metadata?

DEFINITIONS
The Classic Definition

Data about data.
“Metadata, the information we create, store, and share to describe things, allows us to interact with these things to obtain the knowledge we need.”

"Perhaps a more useful, ‘big picture’ way of thinking about metadata is as the sum total of what one can say at a given moment about any information object at any level of aggregation."

My Definition

Metadata are *structured statements* about resources.

These statements allow us to *discover, access, use, preserve, and share* these resources.

One person’s data may be another person’s metadata (or vice versa).
CAUTION!

STAY CLEAR OF NET
AND BE ALERT FOR
OBJECTS LEAVING
THE PLAYING FIELD
De-identified COVID-19 NYU Langone Database

**Description**
The de-identified COVID-19 NYU Langone Database is an active database comprising data of inpatients and outpatients with or at risk for COVID-19 at NYU Langone Health facilities beginning January 1, 2020 to further understand and characterize illness due to the novel coronavirus (COVID-19 disease). The goal of this database is to gather data on COVID-19 related symptoms, comorbidities, risk factors, diagnoses, clinical findings and outcomes, and thereby facilitate pooling of data to ask and answer numerous COVID-19 clinical and research questions. Identifying information, including names and medical record numbers, have been removed from the dataset.

**Timeframe**
2020 - Present

**Geographic Coverage**
New York (State) - New York City

**Subject Domain**
COVID-19
- Delivery of Health Care
- Electronic Health Records
- Health Status
- Population Characteristics
- Quality of Health Care
- Risk Factors

**Population Ages**
Child (2 years - 12 years)
Adult (19 years - 64 years)
Senior (65 years - 79 years)
Aged (80 years and over)

**Keywords**
Comorbidity
COVID-19
- Delivery of health care
- Electronic health records

**Restrictions**
NYU Langone Health Only

**Instructions**
The data is only available to NYU Langone employees after signing the Data Use Agreement supplied through the Access Link. The link can be accessed by employees on-campus or through the institutional IP. Questions or concerns about the NYU COVID-19 de-identified clinical database can be directed to: covid19_ded_data@nyulangone.org

**Associated Publications**

**Access via NYULH**

**DATA NY GOV**

**Canal System Locks**

The New York State Canal System is a 524-mile inland waterway that includes 57 locks, which are used to transfer vessels from a navigation pool at one elevation to another. Information provided in this data set includes the name of each lock, its phone number and specific location by mileage along the canal and its associated elevation.

**More**

**Updated**
October 31, 2021

**Data Provided by**
New York State Canal Corporation
What can we do with metadata?

FUNCTION AND PURPOSE
Leveraging Metadata

Metadata allows us to discover, access, use, preserve, and share information and knowledge.

We rely on metadata, often without being aware of it, to navigate and understand the world and our work.

We rely on the familiar, or standardized, structure inherent in the metadata as much as we do the data itself.


Miller, Lucasta. *L. E. L: The Lost Life and Scandalous Death of Letitia Elizabeth Landon, the Celebrated "Female Byron"*. Alfred A. Knopf, New York, 2019.
What are the types of metadata?
Types of Metadata

Descriptive

Administrative

Structural
Descriptive Metadata

Helps you identify or describe your information.

- Date, time, duration, geographic location
- File name, file type, who created the file, file path
- Name, address, high school, GPA
- Equipment, methods, related datasets
Administrative Metadata

Helps you manage and administer your information.

- **Technical** metadata accounts for information about the file/the resource itself.
- **Rights** metadata helps users understand the intellectual property status or access options for a resource.
- **Preservation** metadata aids in the management of resources over a period of time.
Structural Metadata

Helps you define the relationship between parts.

- Defines connects between portions of a dataset
- Defines chapters or sections within chapters
- Defines level of a hierarchy
How do we do this?

PAUSE TO CONSIDER YOUR STUFF
Questions to Consider...

What do you have?
What do you want to do with it?
What can you say about it?
What can’t you say about it?
Who is your intended audience?
Who else may be using or leveraging this information?
Do not forget your audience...

Assignment1.pdf
Assignment1(2).pdf
Assignment1(3).pdf
Dull_Assignment1.pdf
2022_Dull_Assignment1.pdf

“I need that book on Matthew Vassar. The cover is red.”

“I need a snapshot of all the relevant data from between now and last month.”

“How many papers did my lab publish between 2012 and 2022?”
You can only leverage metadata that you have or that you can understand.
Remember...

Metadata are *structured statements* about resources.

We rely on the familiar, or standardized, *structure* inherent in the metadata as much as we do the data itself.
Date: 06-08-10
You can only leverage metadata that you have or that you can understand.
How do we create structured statements?

METADATA SCHEMA
Metadata Schema

What is a schema?
A set of metadata definitions focused on the resources you are describing and/or a community of practice.

What does it do?
Schema create common vocabulary and semantics that allows for a shared understanding of your metadata and its structure.
Making Statements with Metadata Schemas

“Metadata schemes (also called schema) are sets of metadata elements designed for a specific purpose, such as describing a particular type of information resource. The definition or meaning of the elements themselves is known as the semantics of the scheme. The values given to metadata elements are the content.”

(Hodge’s Understanding Metadata)
Resource → title → Pickwick Papers
Properties and Values

Metadata Schemas, such as Dublin Core, provide the properties we can use to make our statements.

Also makes suggestions to the kinds or types of values that go best with each property. These include controlled vocabularies (e.g., a list of values) or data formats (e.g., date).
**Term Name: date**

<table>
<thead>
<tr>
<th>URI</th>
<th><a href="http://purl.org/dc/terms/date">http://purl.org/dc/terms/date</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Label</td>
<td>Date</td>
</tr>
<tr>
<td>Definition</td>
<td>A point or period of time associated with an event in the lifecycle of the resource.</td>
</tr>
<tr>
<td>Comment</td>
<td>Date may be used to express temporal information at any level of granularity. Recommended practice is to express the date, date/time, or period of time according to ISO 8601-1 [ISO 8601-1] or a published profile of the ISO standard, such as the W3C Note on Date and Time Formats [W3C3DTF] or the Extended Date/Time Format Specification [EDTF]. If the full date is unknown, month and year (YYYY-MM) or just year (YYYY) may be used. Date ranges may be specified using ISO 8601 period of time specification in which start and end dates are separated by a / (slash) character. Either the start or end date may be missing.</td>
</tr>
<tr>
<td>Type of Term</td>
<td>Property</td>
</tr>
<tr>
<td>Has Range</td>
<td><a href="http://www.w3.org/2000/01/rdf-schema#Literal">http://www.w3.org/2000/01/rdf-schema#Literal</a></td>
</tr>
<tr>
<td>Subproperty of</td>
<td>- Date (<a href="http://purl.org/dc/elements/1.1/date">http://purl.org/dc/elements/1.1/date</a>)</td>
</tr>
</tbody>
</table>

**Term Name: title**

<table>
<thead>
<tr>
<th>URI</th>
<th><a href="http://purl.org/dc/terms/title">http://purl.org/dc/terms/title</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Label</td>
<td>Title</td>
</tr>
<tr>
<td>Definition</td>
<td>A name given to the resource.</td>
</tr>
<tr>
<td>Type of Term</td>
<td>Property</td>
</tr>
<tr>
<td>Has Range</td>
<td><a href="http://www.w3.org/2000/01/rdf-schema#Literal">http://www.w3.org/2000/01/rdf-schema#Literal</a></td>
</tr>
<tr>
<td>Subproperty of</td>
<td>- Title (<a href="http://purl.org/dc/elements/1.1/title">http://purl.org/dc/elements/1.1/title</a>)</td>
</tr>
</tbody>
</table>
How do we share metadata?

REDUCE, REUSE, RECYCLE
Reduce, Reuse, Recycle

**Reduce:** do not re-invent the wheel or re-invent a standard.

**Reuse:** borrowing and use statements created by others.

**Recycle:** allow others to borrow and use your statements.
Metadata Standards

Metadata created with standards are consistent, understandable, and shareable.

**Controlled vocabularies:** list of countries (ISO 3166)

**Encoding standards:** YYYYMMDD (ISO 8601)

**Defined Elements/Schema:** Creator or Contributor
Discipline Driven Metadata

Many communities of practice or disciplines established metadata schema and controlled vocabularies that align with or best represent their resources, audiences, and tools.

Other disciplines rely on adaptations, sometimes known as “extensions”, of existing schemas. This approach allows a shared understanding of the basics of the schema as well as discipline-specific properties.

Check out: https://guides.lib.unc.edu/metadata/
Which standards should I use?

What are you colleagues and peers using?

What standards are inherent in your tools or workflows?

What standards align with your stuff?

What standards help you create statements that are useful and actionable?
Documenting your metadata decisions allows you and your colleagues to create *consistent* metadata.

Documentation also allows anyone interested in your metadata to understand *how and why* your metadata looks and works.

This is especially important if you’re diverging from a standard or have created a standard of your own.
<table>
<thead>
<tr>
<th>Project Element</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mapping</td>
<td>Title</td>
</tr>
<tr>
<td>Required/Optional</td>
<td>Required — ALL items in the collection must have a title.</td>
</tr>
<tr>
<td>Controlled Vocabulary</td>
<td>N</td>
</tr>
<tr>
<td>Definition</td>
<td>A name given to the resource.</td>
</tr>
<tr>
<td>Usage</td>
<td>The title for the photograph can be found on a white sticker on the back of the photograph or text written on the photocopy of the photo located in 3-ring binder in the stacks by the collection. If a title cannot be located by these means, one should be created by the person creating the metadata. The title should be short and descriptive. See examples for models.</td>
</tr>
<tr>
<td></td>
<td>Only the first word and proper nouns in the title should be capitalized.</td>
</tr>
<tr>
<td></td>
<td>Information regarding the potential sources of titles (backs of images or constructed by digitization staff) will be included on the page describing the overall collection.</td>
</tr>
<tr>
<td>Example</td>
<td>Busy art room</td>
</tr>
<tr>
<td></td>
<td>Professor and class looking at pictures on floor, different angle</td>
</tr>
</tbody>
</table>
Metadata allows you to achieve your goals.

**But**, you can only leverage the metadata that you have.
Questions??

mdull@library.rochester.edu

https://www.library.rochester.edu/services/metadata-outreach