

# 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.

# A Functional Definition

---

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

From [Riley, J. \(2017\). Understanding Metadata: What is metadata and what is it for? Bethesda, MD: NISO Press.](#)

# A Different Perspective

---

“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.”

From Gilliland, A. (2016). Setting the Stage. In M. Baca (Ed.) *Introduction to Metadata*. (3<sup>rd</sup> ed.) (pp. 1-20). Los Angeles, CA: Getty Publications. <http://www.getty.edu/publications/intrometadata/>

# 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).





The screenshot shows a mobile application interface. At the top is a map from Apple Maps showing Fenway Park in Boston, with a red location pin. Below the map is a table of photo metadata.

|               |                            |
|---------------|----------------------------|
| Name          | IMG_6599.JPG               |
| Size          | 1.48 MB                    |
| Dimensions    | 3024x3024                  |
| Date Time     | Jun 26, 2017 at 6:24:41 PM |
| ISO Speed     | ISO 40                     |
| F Number      | f/2.2                      |
| Exposure Time | 1/120 sec                  |
| Exposure Mode | Auto Exposure              |

# De-identified COVID-19 NYU Langone Database

UID: 10391

## 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 Age

[Child \(2 years - 12 years\)](#)  
[Adult \(19 years - 64 years\)](#)  
[Senior \(65 years - 79 years\)](#)  
[Aged \(80 years and over\)](#)

## Keywords

[Comorbidities](#)  
[COVID-19](#)  
[Delivery of health care](#)  
[Electronic health records](#)

## Access

### 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 VPN. Questions or concerns about the NYULH COVID-19 de-identified clinical database can be directed to:  
[covid19\\_deid\\_db@nyulangone.org](mailto:covid19_deid_db@nyulangone.org).

[Access via NYULH](#)

## Associated Publications

Stachel A, Daniel K, Ding D, Francois F, Phillips M, Lighter J. Development and validation of a machine learning model to predict mortality risk in patients with COVID-19. *BMJ Health Care Inform.* 2021 May;28(1):e100235. doi: 10.1136/bmjhci-2020-100235.

Allen B, El Shahawy O, Rogers ES, Hochman S, Khan MR, Krawczyk N. Association of substance use disorders and drug overdose with adverse COVID-19 outcomes in New York City: January-October 2020. *J Public Health (Oxf)*. 2020 Dec 26;fdaa241. doi: 10.1093/pubmed/fdaa241.

Wang JM, Liu W, Chen X, McRae MP, McDevitt JT, Fenyö D. Predictive Modeling of Morbidity and Mortality in Patients Hospitalized With COVID-19 and its Clinical Implications: Algorithm Development and Interpretation. *J Med Internet Res.* 2021 Jul 9;23(7):e29514. doi: 10.1198/jmir.2021.23e29514.



# Canal System Locks

Recreation

[View Data](#)[Visualize](#)[Export](#)[API](#)

...

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 [More](#)

Updated  
October 31, 2021

Data Provided by  
New York State Canal Corporation

## About this Dataset

[Mute Dataset](#)

Updated

### October 31, 2021

|                   |                       |
|-------------------|-----------------------|
| Data Last Updated | Metadata Last Updated |
| October 31, 2021  | October 31, 2021      |

Date Created  
February 26, 2013

|       |           |
|-------|-----------|
| Views | Downloads |
| 4,926 | 2,064     |

Data Provided by Dataset

## Dataset Summary

|                     |                                  |
|---------------------|----------------------------------|
| Organization        | Maintenance and Operations       |
| Time Period         | 1910-Present                     |
| Posting Frequency   | As needed                        |
| Data Frequency      | Infrequent                       |
| Dataset Owner       | New York State Canal Corporation |
| Contact Information | jeffrey.gritsavage@canals.ny.gov |
| Coverage            | Statewide                        |
| Granularity         | Facility                         |
| Units               | Lock or Lock Combination         |

# 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, L. (2019). *L. E. L: The lost life and scandalous death of Letitia Elizabeth Landon, the celebrated "Female Byron"* (First ed.). New York: Alfred A. Knopf.

Miller, Lucasta. 2019. *L. E. L: The Lost Life and Scandalous Death of Letitia Elizabeth Landon, the Celebrated "Female Byron"*. First ed. New York: Alfred A. Knopf.

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?

---

CONSIDERING STATEMENTS



# 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

A solid blue horizontal bar at the bottom of the slide.

# 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

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

Assignment1(2).pdf

Assignment1(3).pdf

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

Dull\_Assignment1.pdf

2022\_Dull\_Assignment1.pdf

“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?*

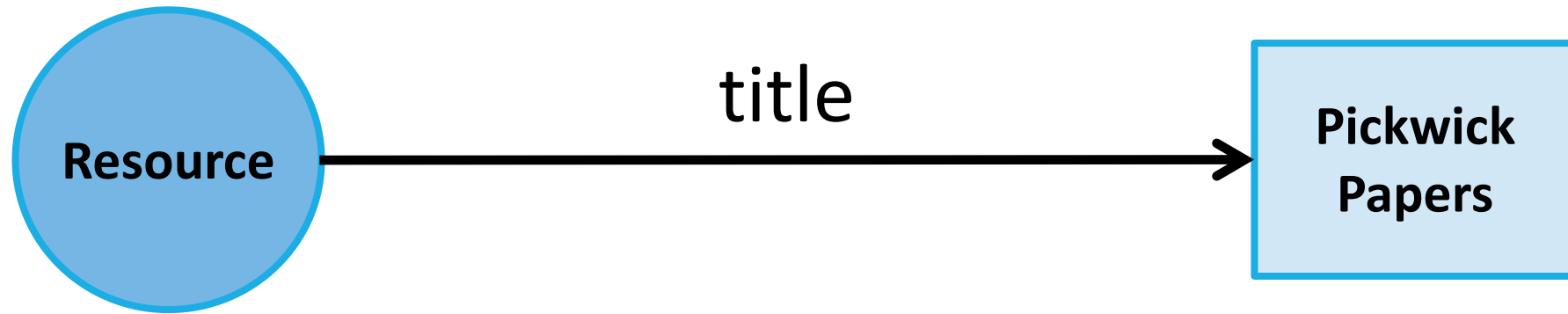
Schema create common vocabulary and semantics that allows for a shared understanding your metadata and its structure.

# 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)*



# 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       |  | <a href="#">More details</a> |
|-----------------------|--|------------------------------|
| <b>URI</b>            | http://purl.org/dc/terms/date  |                              |
| <b>Label</b>          | Date   |                              |
| <b>Definition</b>     | A point or period of time associated with an event in the lifecycle of the resource.   |                              |
| <b>Comment</b>        | 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 [ <a href="#">ISO 8601-1</a> ] or a published profile of the ISO standard, such as the W3C Note on Date and Time Formats [ <a href="#">W3CDTF</a> ] or the Extended Date/Time Format Specification [ <a href="#">EDTF</a> ]. 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. |                              |
| <b>Type of Term</b>   | Property   |                              |
| <b>Has Range</b>      | http://www.w3.org/2000/01/rdf-schema#Literal   |                              |
| <b>Subproperty of</b> | <ul style="list-style-type: none"> <li><a href="#">Date</a> (http://purl.org/dc/elements/1.1/date)</li> </ul>  |                              |

| Term Name: title      |   |
|-----------------------|---|
| <b>URI</b>            | http://purl.org/dc/terms/title  |
| <b>Label</b>          | Title   |
| <b>Definition</b>     | A name given to the resource.   |
| <b>Type of Term</b>   | Property  |
| <b>Has Range</b>      | http://www.w3.org/2000/01/rdf-schema#Literal  |
| <b>Subproperty of</b> | <ul style="list-style-type: none"> <li><a href="#">Title</a> (http://purl.org/dc/elements/1.1/title)</li> </ul> |

# 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?

# Document, Document, Document

---

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.

|                              |   |
|------------------------------|---|
| <b>Project Element</b>       | Title   |
| <b>Mapping</b>               | <a href="#">Title</a>   |
| <b>Required/Optional</b>     | Required – ALL items in the collection must have a title.   |
| <b>Controlled Vocabulary</b> | N   |
| <b>Definition</b>            | A name given to the resource.   |
| <b>Usage</b>                 | <p>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.</p> <p>Only the first word and proper nouns in the title should be capitalized.</p> <p>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.</p> |
| <b>Example</b>               | <p>Busy art room</p> <p>Professor and class looking at pictures on floor, different angle</p>   |

Metadata allows you to  
achieve your goals.

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

# Questions??

---

[mdull@library.rochester.edu](mailto:mdull@library.rochester.edu)

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