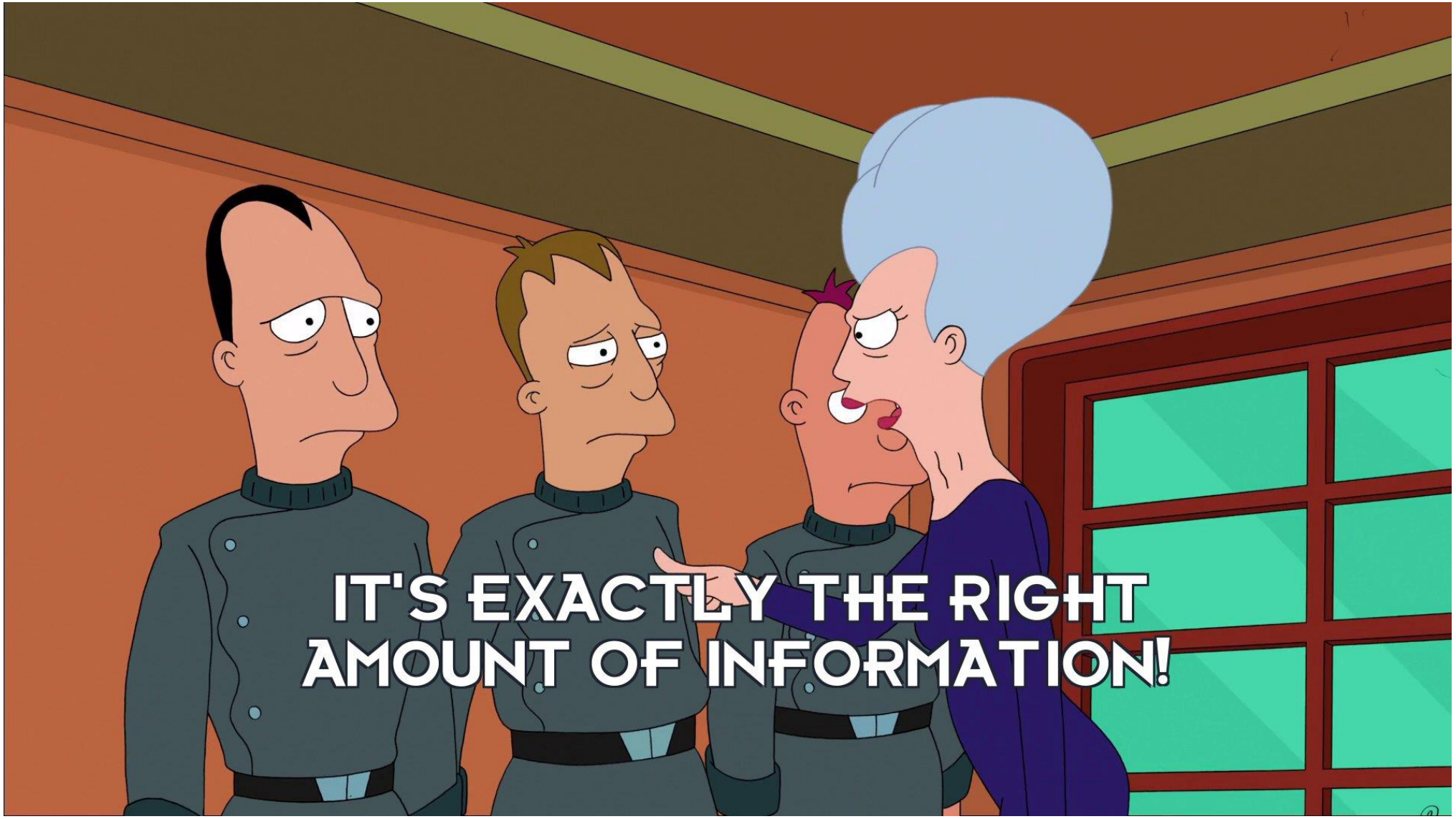
Metadata:
A Love Letter to the Future

LOVE DATA WEEK WORKSHOP

MAGGIE DULL, HEAD OF METADATA STRATEGIES, RIVER CAMPUS LIBRARIES


IT'S EXACTLY THE RIGHT
AMOUNT OF INFORMATION!

Overview

What is metadata?

What can we do with metadata?

How do we do the things?

How do we share the things?

Why does metadata matter?

What is metadata?

WHY IS IT IMPORTANT?

https://bigsnarf.wordpress.com/2014/06/24/metadata-bigdata/

Definition

Data about data.*

*Quite possibly the least helpful definition ever.

Try, Again

"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*. (3rd ed.) (pp. 1-20). Los Angeles, CA: Getty Publications. http://www.getty.edu/publications/intrometadata/

Or, Maybe

"Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an **information resource**."

From Hodge, G. M. (2004). Understanding Metadata. Bethesda, MD: NISO Press.

Information What?

Information Object or Information Resource

= **Stuff You Care About**

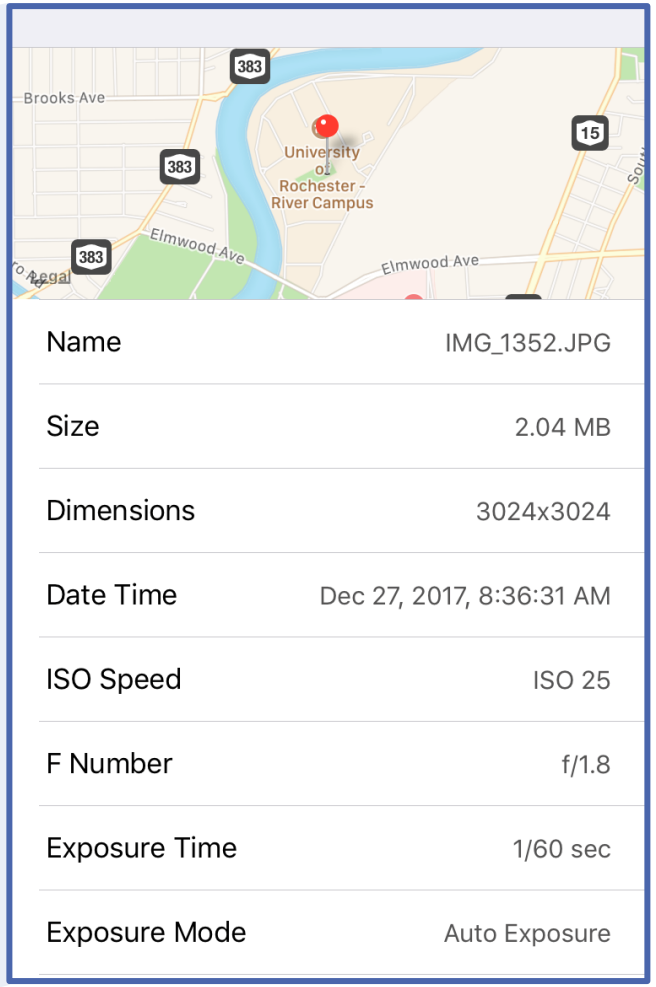
What is metadata?

WHAT IS IT MADE OF?

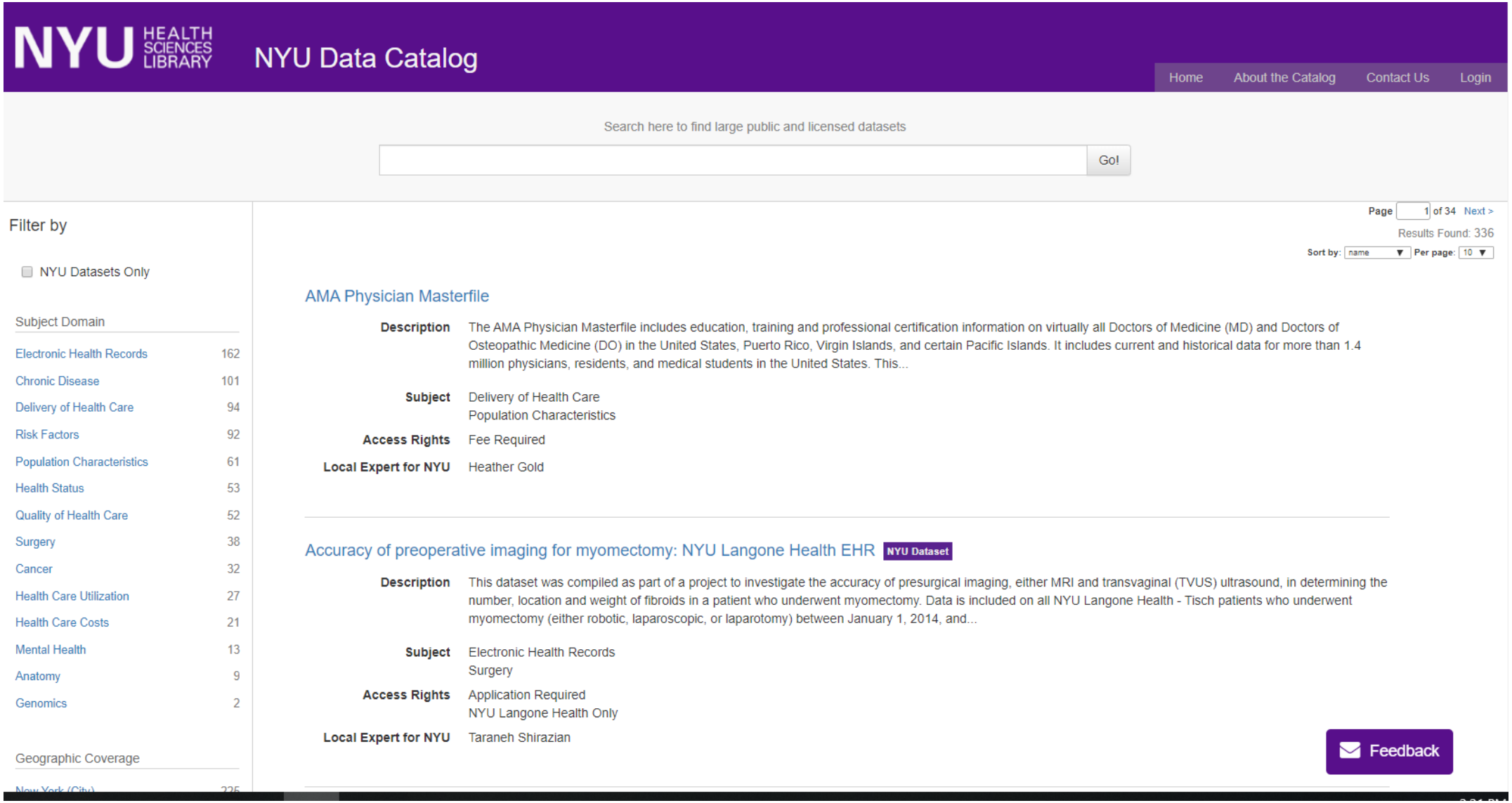
Statements About Resources

Metadata is comprised of

structured statements about resources



Name	IMG_1352.JPG
Size	2.04 MB
Dimensions	3024x3024
Date Time	Dec 27, 2017, 8:36:31 AM
ISO Speed	ISO 25
F Number	f/1.8
Exposure Time	1/60 sec
Exposure Mode	Auto Exposure


NYU HEALTH SCIENCES LIBRARY
NYU Data Catalog
Home
About the Catalog
Contact Us
Login
Search here to find large public and licensed datasets
Go!
Filter by
NYU Datasets Only
Subject Domain
Electronic Health Records 162
Chronic Disease 101
Delivery of Health Care 94
Risk Factors 92
Population Characteristics 61
Health Status 53
Quality of Health Care 52
Surgery 38
Cancer 32
Health Care Utilization 27
Health Care Costs 21
Mental Health 13
Anatomy 9
Genomics 2
Geographic Coverage
Page 1 of 34 Next >
Results Found: 336
Sort by: name Per page: 10
AMA Physician Masterfile
Description The AMA Physician Masterfile includes education, training and professional certification information on virtually all Doctors of Medicine (MD) and Doctors of Osteopathic Medicine (DO) in the United States, Puerto Rico, Virgin Islands, and certain Pacific Islands. It includes current and historical data for more than 1.4 million physicians, residents, and medical students in the United States. This...
Subject Delivery of Health Care
Population Characteristics
Access Rights Fee Required
Local Expert for NYU Heather Gold
Accuracy of preoperative imaging for myomectomy: NYU Langone Health EHR NYU Dataset
Description This dataset was compiled as part of a project to investigate the accuracy of presurgical imaging, either MRI and transvaginal (TVUS) ultrasound, in determining the number, location and weight of fibroids in a patient who underwent myomectomy. Data is included on all NYU Langone Health - Tisch patients who underwent myomectomy (either robotic, laparoscopic, or laparotomy) between January 1, 2014, and...
Subject Electronic Health Records
Surgery
Access Rights Application Required
NYU Langone Health Only
Local Expert for NYU Taraneh Shirazian
Feedback

What can we **do** with metadata?

Metadata Allows Us To **Do** Things

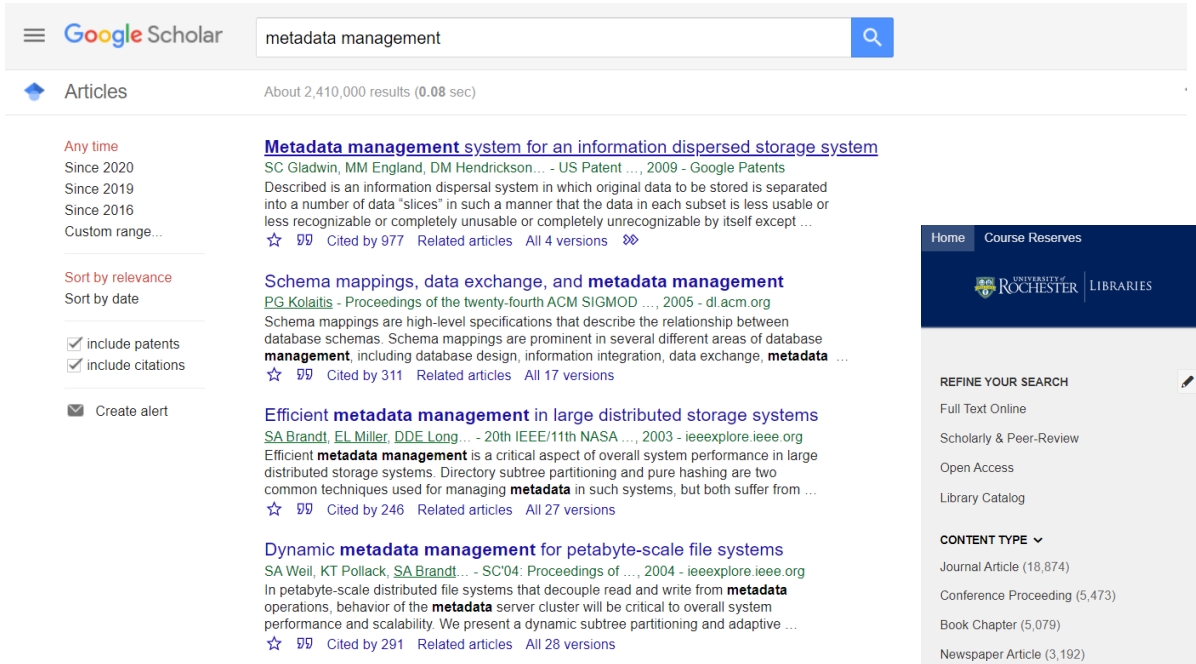
…with information and stuff:

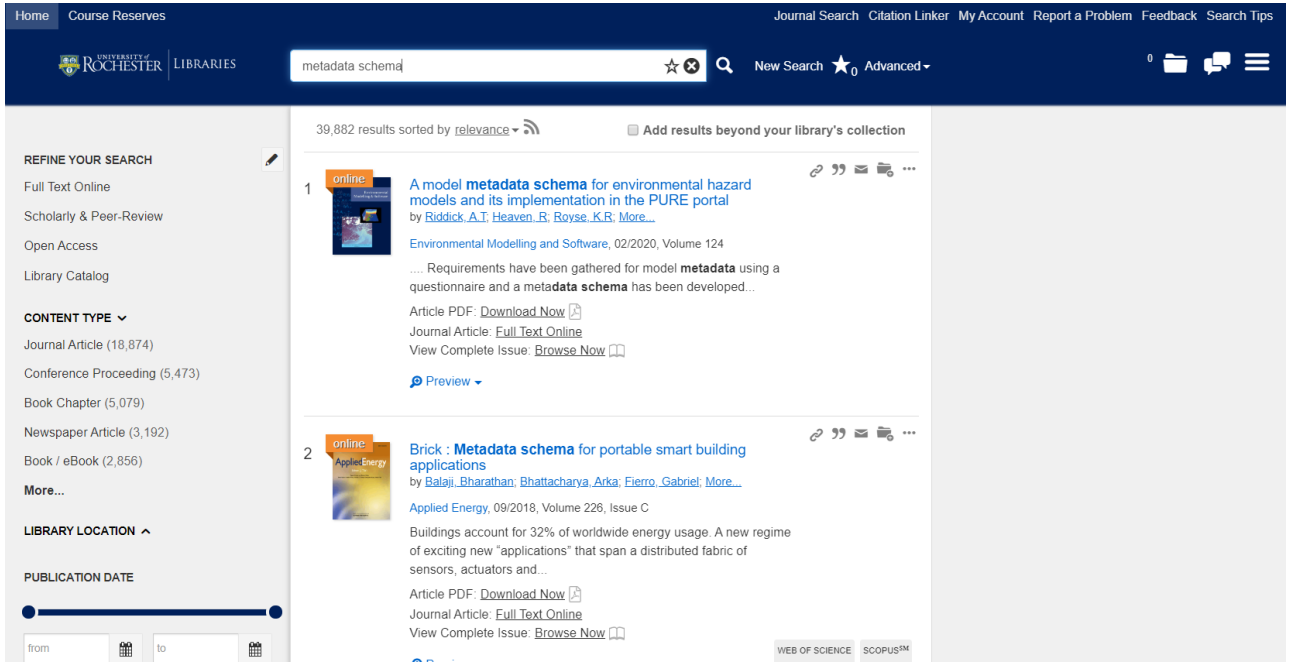
Discover

Connect

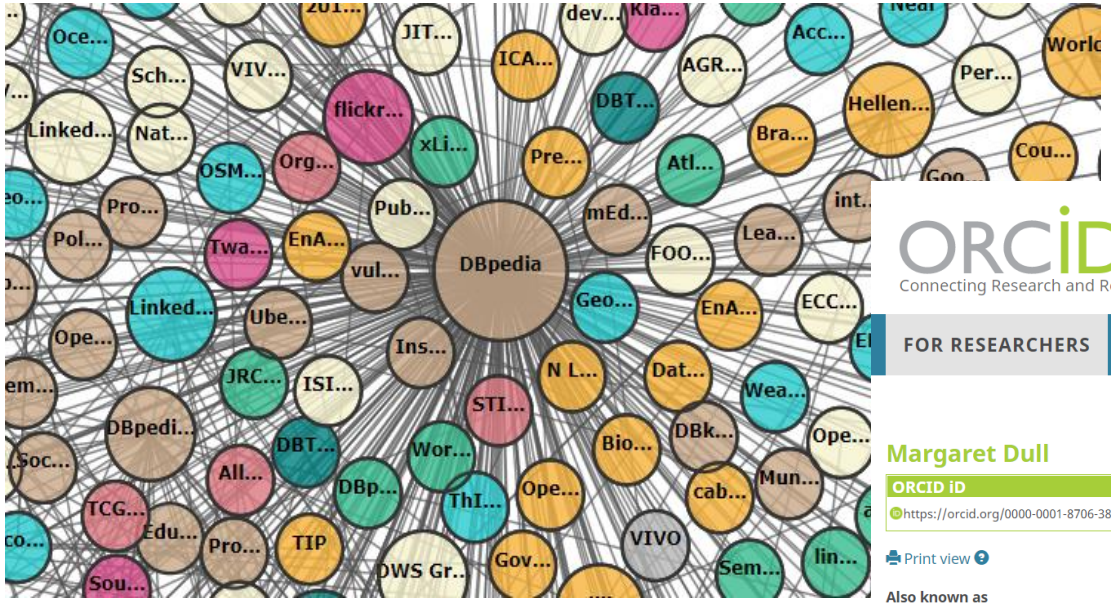
Preserve and Manage

Discover

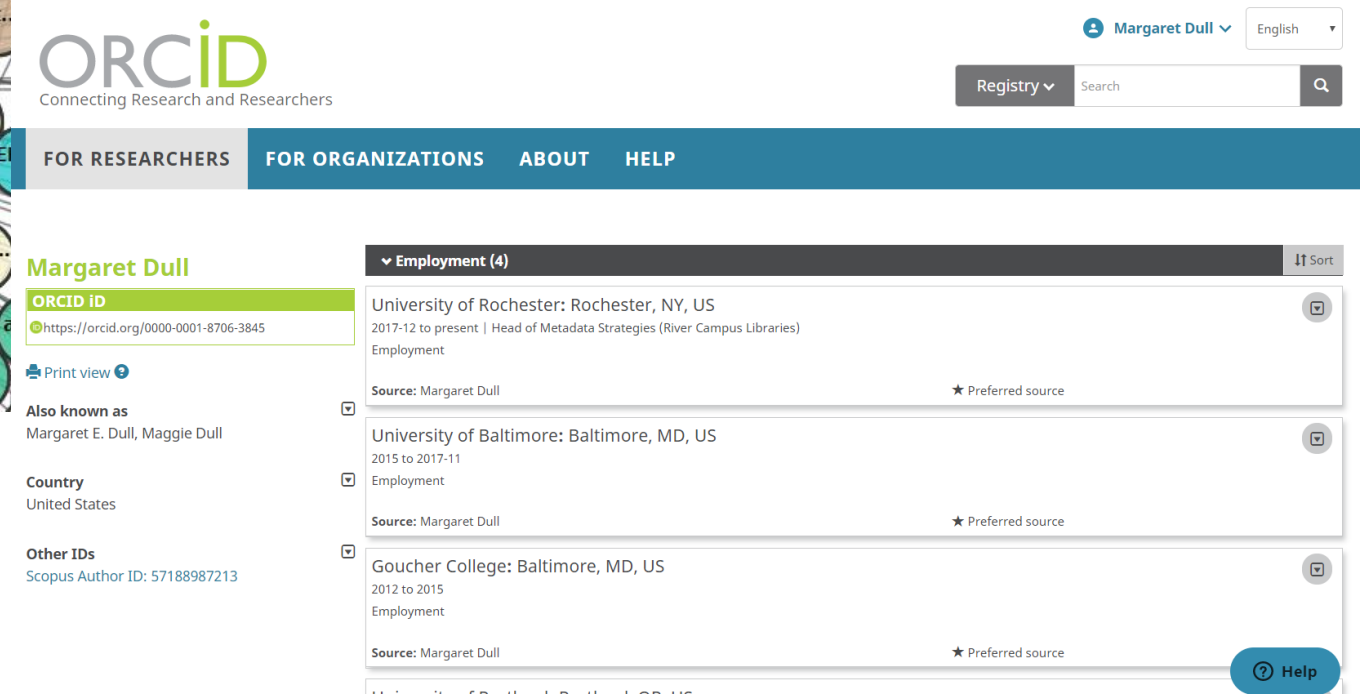

Google Scholar
metadata management
Articles
About 2,410,000 results (0.08 sec)
Any time
Since 2020
Since 2019
Since 2016
Custom range...
Sort by relevance
Sort by date
include patents
include citations
Create alert
Metadata management system for an information dispersed storage system
SC Gladwin, MM England, DM Hendrickson... - US Patent ..., 2009 - Google Patents
Described is an information dispersal system in which original data to be stored is separated into a number of data "slices" in such a manner that the data in each subset is less usable or less recognizable or completely unusable or completely unrecognizable by itself except ...
Cited by 977 Related articles All 4 versions
Schema mappings, data exchange, and metadata management
PG Kolaitis - Proceedings of the twenty-fourth ACM SIGMOD ..., 2005 - dl.acm.org
Schema mappings are high-level specifications that describe the relationship between database schemas. Schema mappings are prominent in several different areas of database management, including database design, information integration, data exchange, metadata ...
Cited by 311 Related articles All 17 versions
Efficient metadata management in large distributed storage systems
SA Brandt, EL Miller, DDE Long... - 20th IEEE/11th NASA ..., 2003 - ieeexplore.ieee.org
Efficient metadata management is a critical aspect of overall system performance in large distributed storage systems. Directory subtree partitioning and pure hashing are two common techniques used for managing metadata in such systems, but both suffer from ...
Cited by 246 Related articles All 27 versions
Dynamic metadata management for petabyte-scale file systems
SA Weil, KT Pollack, SA Brandt... - SC'04: Proceedings of ..., 2004 - ieeexplore.ieee.org
In petabyte-scale distributed file systems that decouple read and write from metadata operations, behavior of the metadata server cluster will be critical to overall system performance and scalability. We present a dynamic subtree partitioning and adaptive ...
Cited by 291 Related articles All 28 versions


Home
Course Reserves
Journal Search
Citation Linker
My Account
Report a Problem
Feedback
Search Tips
University of Rochester Libraries
metadata schema
New Search
Advanced
REFINE YOUR SEARCH
Full Text Online
Scholarly & Peer-Review
Open Access
Library Catalog
CONTENT TYPE
Journal Article (18,874)
Conference Proceeding (5,473)
Book Chapter (5,079)
Newspaper Article (3,192)
Book / eBook (2,856)
More...
LIBRARY LOCATION
PUBLICATION DATE
from
to
39,882 results sorted by relevance
Add results beyond your library's collection
1
online
A model metadata schema for environmental hazard models and its implementation in the PURE portal
by Riddick, A.T; Heaven, R; Royse, K.R; More...
Environmental Modelling and Software, 02/2020, Volume 124
.... Requirements have been gathered for model metadata using a questionnaire and a metadata schema has been developed...
Article PDF: Download Now
Journal Article: Full Text Online
View Complete Issue: Browse Now
Preview
2
online
Brick : Metadata schema for portable smart building applications
by Balaji, Bharathan; Bhattacharya, Arka; Fierro, Gabriel; More...
Applied Energy, 09/2018, Volume 226, Issue C
Buildings account for 32% of worldwide energy usage. A new regime of exciting new "applications" that span a distributed fabric of sensors, actuators and...
Article PDF: Download Now
Journal Article: Full Text Online
View Complete Issue: Browse Now
WEB OF SCIENCE
SCOPUS

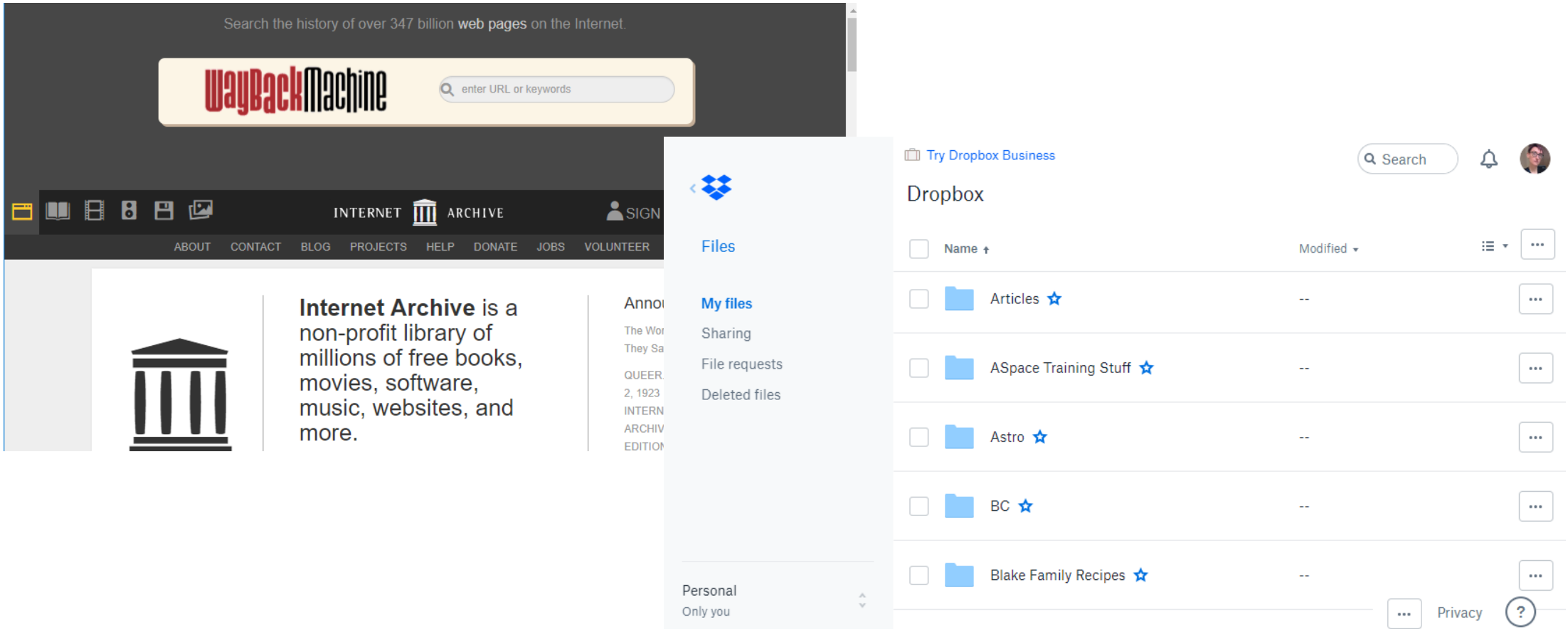
Connect



http://lod-cloud.net



Preserve and Manage



How do we do the things?

CONSIDERING STATEMENTS

Types of Metadata Statements

Descriptive

Administrative

Structural

Descriptive Metadata

Helps you identify or describe your stuff.

- Title, author, edition…
- Date, time, duration, geographic location…
- File name, file type, who created the file…

Administrative Metadata

Helps you manage and administer your stuff.

- **Technical** metadata often relates to information about the file itself.
- **Rights** metadata helps you understand the intellectual property rights regarding the resource.
- **Preservation** metadata helps you manage the files over a period of time.

Structural Metadata

Helps you define the relationship between parts.

- Defines chapters or sections within chapters.
- Relationship between subsets of a dataset.

How do we do the things?

PAUSE TO CONSIDER YOUR STUFF

Ask Yourself

- What do you have?
- What do you want to do with it?
- What can you say about it?
- Who is your audience?

Know Your Audience

STUDENT ASSIGNMENTS

Assignment1.pdf

Assignment1(2).pdf

ASKING FOR HELP

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

You can only leverage
metadata that you **have**
or that you can **understand**.

Title	**Letters of Matthew Vassar available in the Vassar College Library / compiled by the Bibliography Department, Vassar College Library, 1963**
Imprint	[Poughkeepsie, New York] : [Vassar College], [1965]

LOCATION	CALL #	STATUS
Archive/SpecColl (VC)	LD7182.65.V3 L48 1965 1	LIB USE ONLY
Archive/SpecColl (VC)	LD7182.65.V3 L48 1965 2	LIB USE ONLY
Archive/SpecColl (VC)	LD7182.65.V3 L48 1965 3	LIB USE ONLY
Archive/SpecColl (VC)	LD7182.65.V3 L48 1965 4	LIB USE ONLY

Descript	4 volumes (ciii, that is cxii, 538 leaves) ; 29 cm
Subject	Vassar, Matthew, 1792-1868 -- Correspondence
	Vassar College -- History
Add Na/Ti	Vassar, Matthew, 1792-1868, author
	Vassar College. Library, compiler

Reminder

"Metadata is **structured information** that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource."

From Hodge, G. M. (2004). Understanding Metadata. Bethesda, MD: NISO Press.

Date: 06-08-10

How do we share the things?

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 so that your audience understands your metadata and its structure.

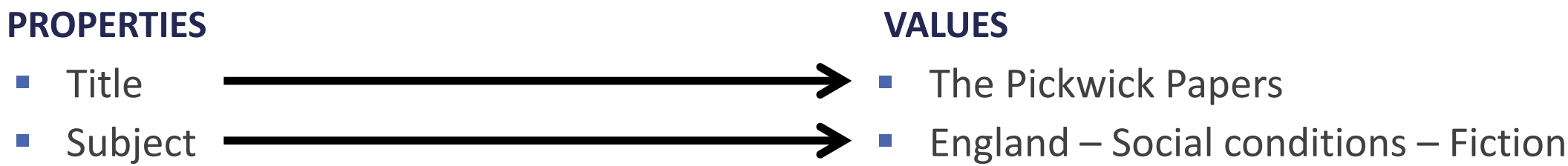
Making Statements with Metadata Schemas

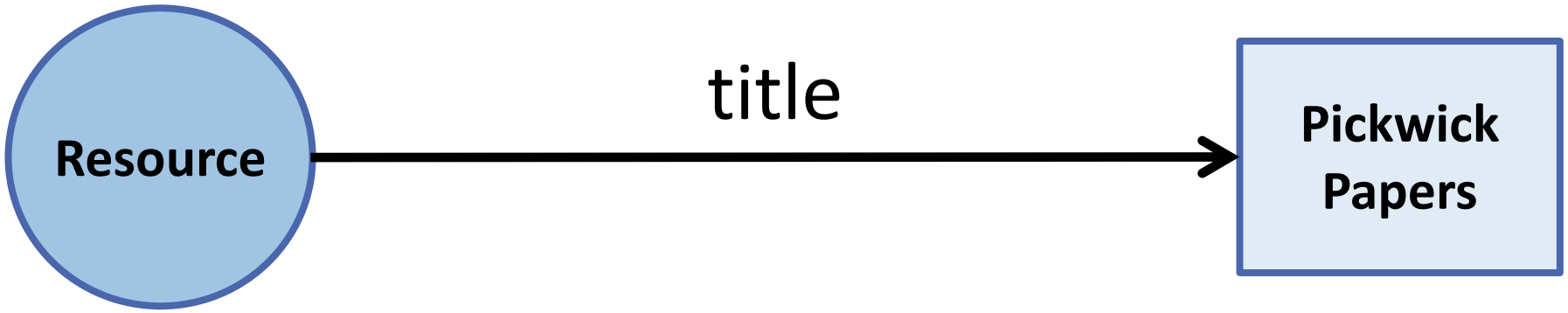
"Metadata schemes (also called schema) are sets of **metadata elements** designed for a **specific purpose**, such as describing a particular type of information resource. The definition or meaning of the elements themselves is known as the **semantics** of the scheme. The values given to metadata elements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

(Hodge's *Understanding Metadata*)

Creating Statements

Statements about resources are made up of **properties** and **values**.

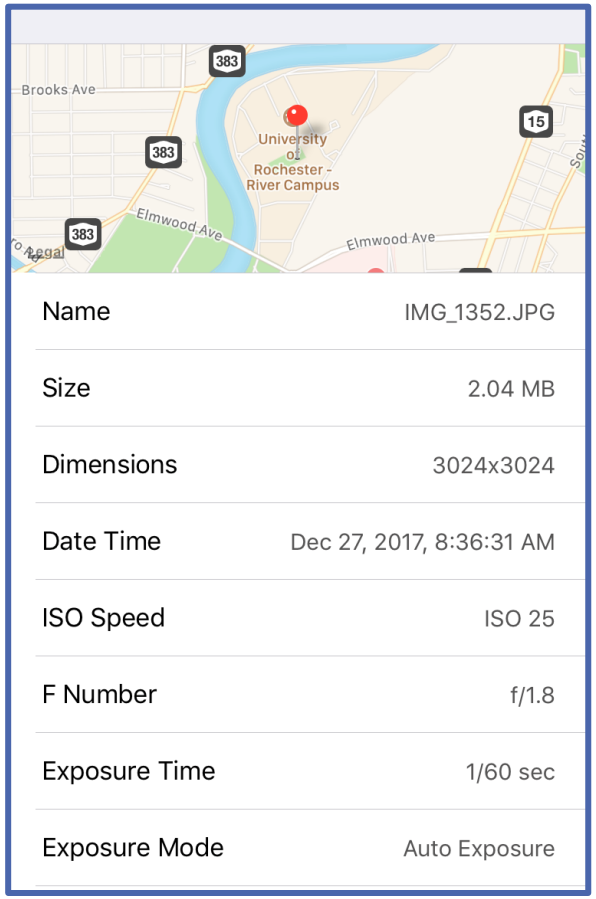



Resource
title
Pickwick Papers

Properties and Values

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

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



Term Name: date	More details
URI	http://purl.org/dc/terms/date
Label	Date
Definition	A point or period of time associated with an event in the lifecycle of the resource.
Comment	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 [W3CDTF] 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.
Type of Term	Property
Has Range	http://www.w3.org/2000/01/rdf-schema#Litera
Subproperty of	• Date (http://purl.org/dc/elements/1.1/da

Term Name: title	
URI	http://purl.org/dc/terms/title
Label	Title
Definition	A name given to the resource.
Type of Term	Property
Has Range	http://www.w3.org/2000/01/rdf-schema#Literal
Subproperty of	• Title (http://purl.org/dc/elements/1.1/title)

Reduce, Reuse, Recycle

Many communities of practice have established metadata schema that you can adopt.

- Reduce: do not re-invent the wheel
- Reuse: borrow and use statements created by others
- Recycle: others can borrow and use your statements

Appendix B: Metadata Elements

Metadata Element	Description	Dublin Core Value
Dataset Name / Title	The name of the dataset as it will appear on the platform.	Title
Dataset Description	Short description that explains the purpose of the Dataset and the data within	Description
Category	The general category that the dataset is included in on the site (The categories include: Economic Development, Education, Energy & Environment, Government & Finance, Health, Human Services, Public Safety, Recreation, Transparency, and Transportation).	Type
Tags / Keywords	Keywords about the dataset used for searching purposes.	Subject
Data Provided By	The Agency that provided the data.	Contributor
URL to Dataset Program Web Page	The URL to the program area web pages.	n/a
Responsible Organization Within Agency	The organization that the dataset owner is a part of	Creator
Time Period	The timeframe of data available in the associated data file (e.g., Beginning 2005).	Coverage (temporal)
Create Date	The date the resource was made available in its present form – auto generated	Date
Posting Frequency	How often the Dataset will be refreshed (e.g., Annually, Monthly, Daily).	n/a
Contact e-mail information	The email address the viewers of the data can use to ask questions about the dataset	n/a

NEW YORK STATE

OPEN DATA HANDBOOK

data.ny.gov

http://ny.github.io/open-data-handbook/OpenDataHandbook.pdf

Why does metadata matter?

IN CONCLUSION

Metadata allows you to achieve your goals.

BUT, you can only leverage
the metadata that you have.

Questions

MAGGIE DULL | MDULL@LIBRARY.ROCHESTER.EDU