Introduction to Schema.org and JSON-LD

Jacob Shelby

Goal of today's session

Learn about two key tools for increasing discoverability of information on the open web

- Schema.org
- Semantic Markup

Schema.org

What is it?

Schema.org is a set of vocabularies for describing things in an interoperable and semantic way on the web.

Historical Context

- Started in 2011 by Google, Yahoo, Bing
- Maintained by members of Google, Yahoo, Bing, Yandex
- Original scope was for businesses
- Community driven

Purpose

- “[To] create and support a common set of schemas for structured data markup on web pages.”*
- To make data on the web more meaningful to machines

* https://webmasters.googleblog.com/2011/06/introducing-schemaorg-search-engines.html

Item information

Title:
Brothers of the Omega Psi Phi Fraternity in step routine

Topics:
Student Life

Subjects:
African American fraternal organizations
College students -- Societies, etc.
Dance
Dogs
Fraternal organizations
Performances
Students -- Societies, etc.

Photograph →

Original Format:
Black-and-white print (photograph)

Extent:
8 x 10 inches

Item identifier:
0021721

Person →

Photographer:
Lodge, Rick

Created Date:
1981-10-01

Description:
Information from The Technician newspaper, dated October 7, 1981: Louis Cade [with dog] leads his Omega Psi Phi brothers in a step routine before 400 onlookers in front of the free expression tunnel.

Genre:
Photographs

Names:
North Carolina State University -- Students
Omega Psi Phi Fraternity

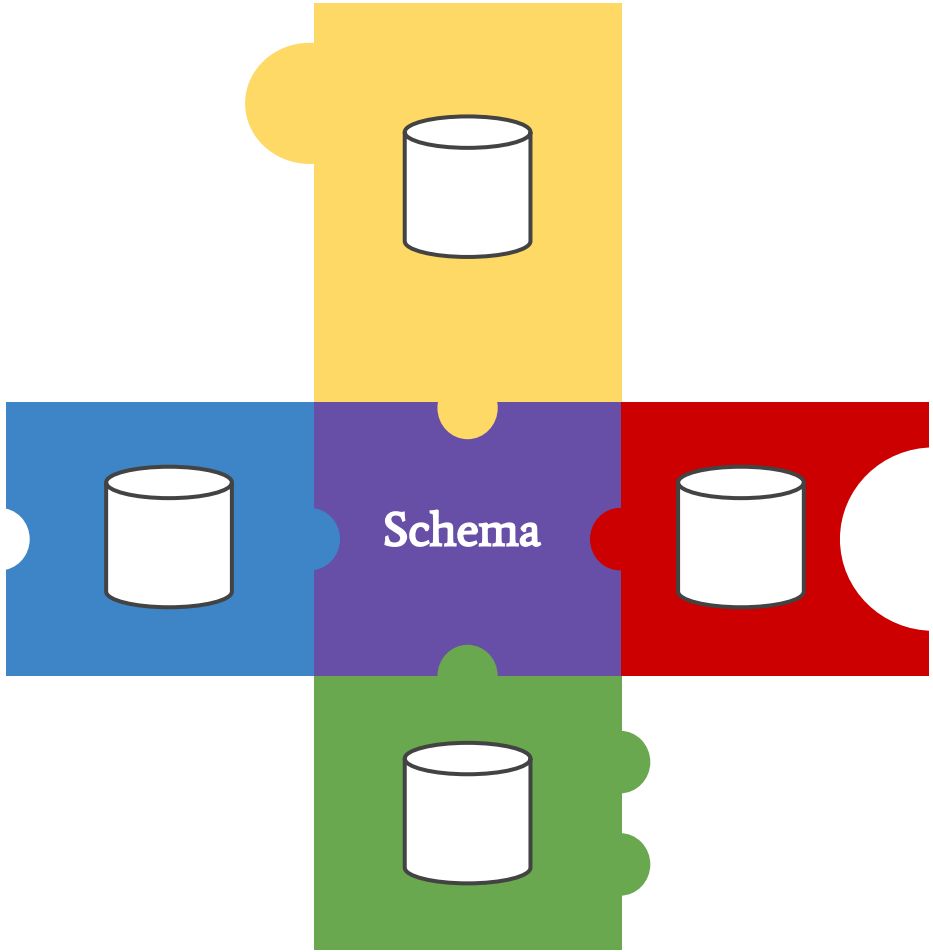
Place →

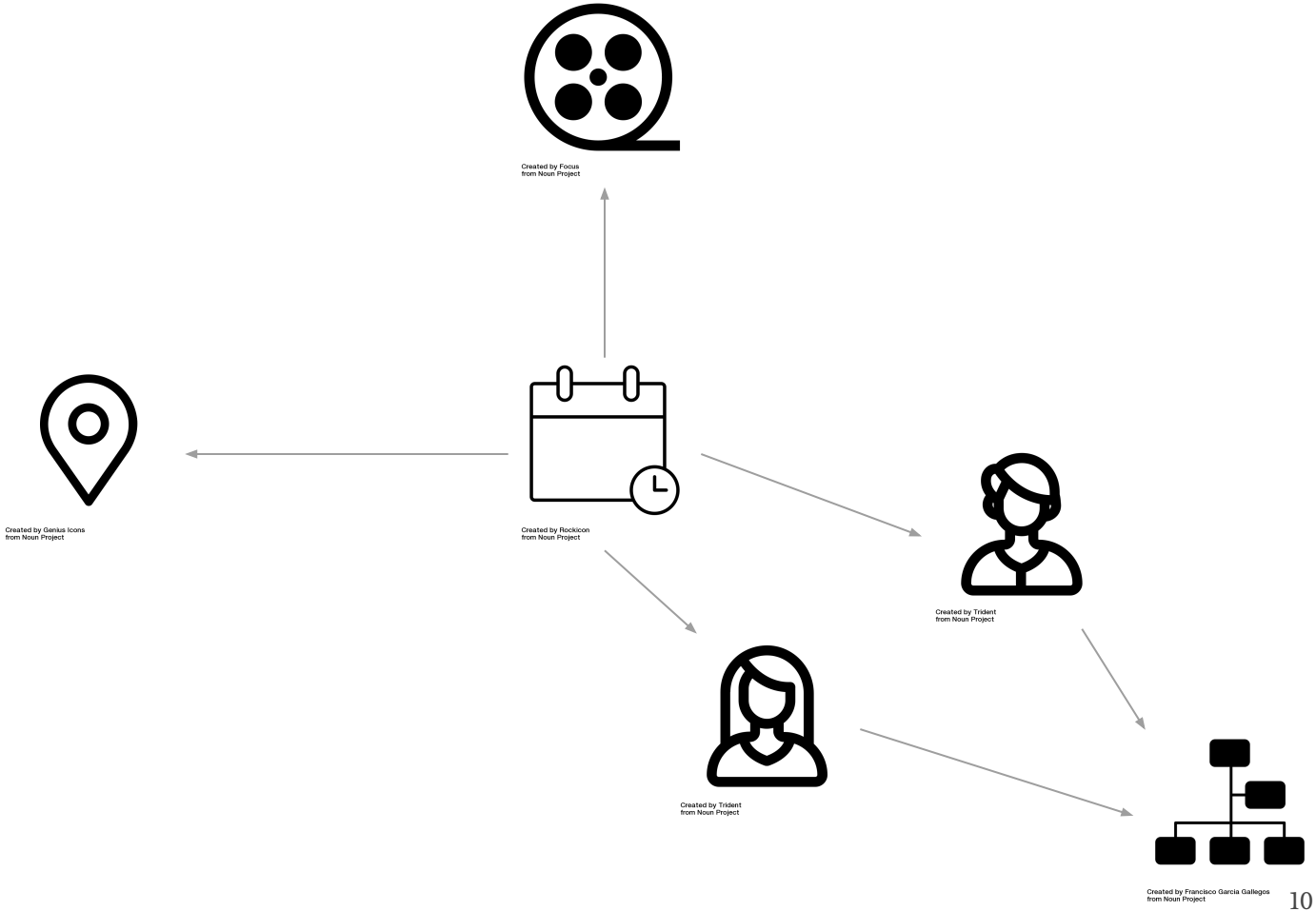
Location:
Raleigh (N.C)

Digital Project:
University Archives Photographs

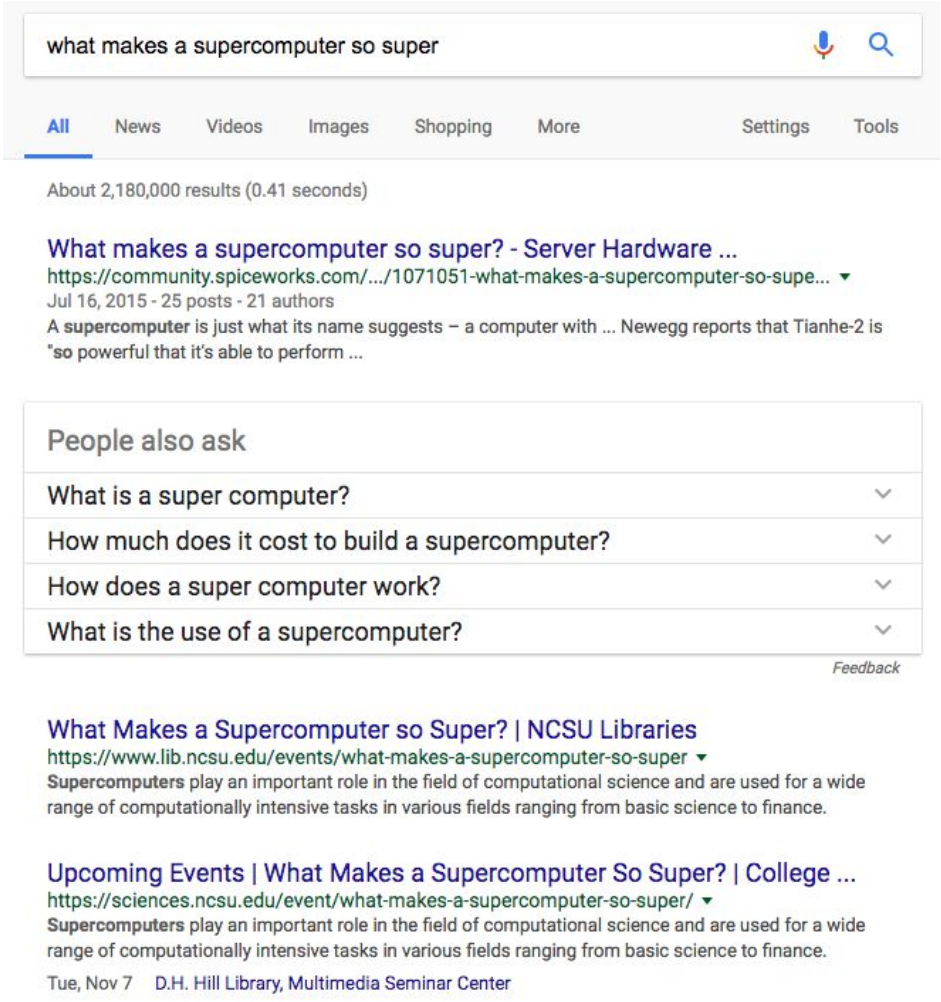
Image

follows linked data principles


Schema

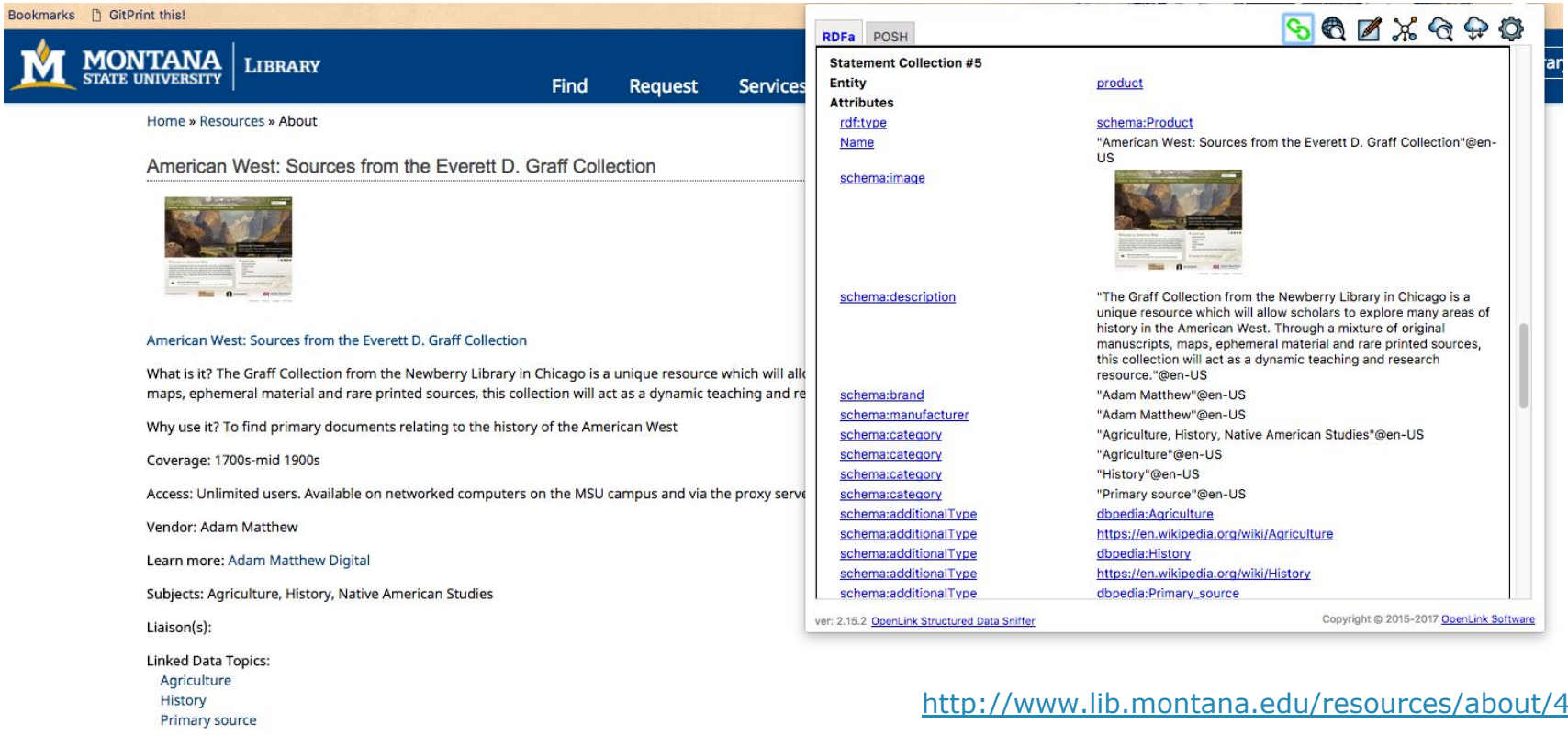

Created by Focus
from Noun Project
Created by Genius Icons
from Noun Project
Created by Rockicon
from Noun Project
Created by Trident
from Noun Project
Created by Trident
from Noun Project
Created by Francisco Garcia Gallegos
from Noun Project

Schema.org

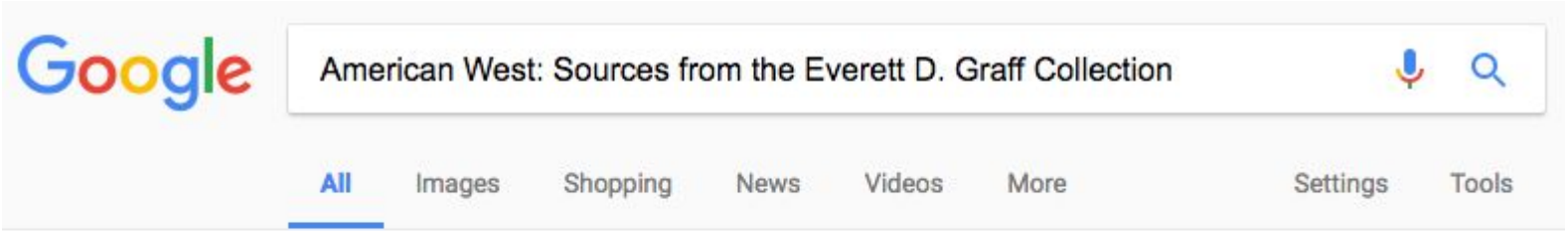


https://www.google.com/search?q=what+makes+a+supercomputer+so+super

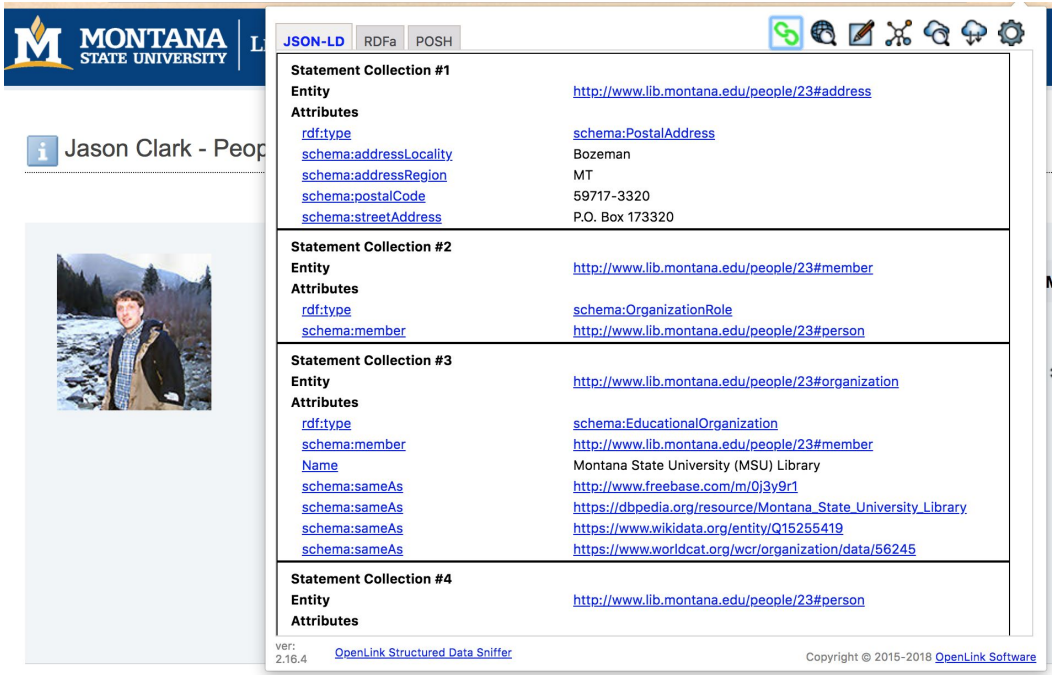
Montana State University Databases



http://www.lib.montana.edu/resources/about/482


Google
American West: Sources from the Everett D. Graff Collection
All
Images
Shopping
News
Videos
More
Settings
Tools
About 17,900 results (1.21 seconds)
American West: Sources from the Everett D. Graff Collection : Montana ...
www.lib.montana.edu › Resources
Rating: 3.5 - 4,380 votes
Aug 1, 2017 - The Graff Collection from the Newberry Library in Chicago is a unique resource which will allow scholars to explore many areas of history in the ...

Montana State University People



http://www.lib.montana.edu/people/about/23

Impact on Library SEO

The Open SESMO (Search Engine & Social Media Optimization) Project: Linked and Structured Data for Library Subscription Databases to Enable Web-scale Discovery in Search Engines

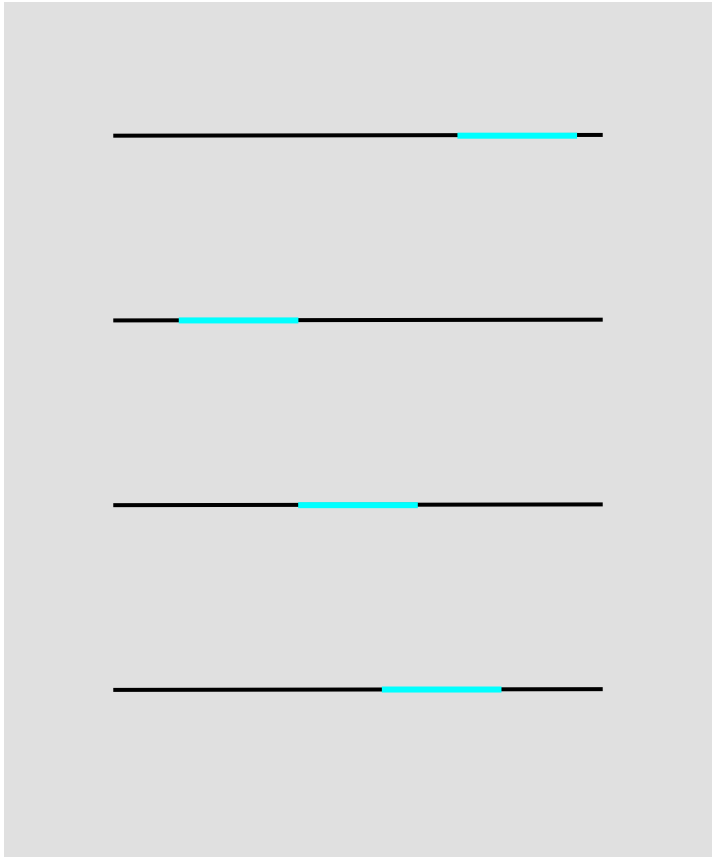
Linked Data is People: Building a Knowledge Graph to Reshape the Library Staff Directory

Usage by Big Engines

Schema.org: Evolution of Structured Data on the Web

Google - Introduction to Structured Data

Semantic Markup



Person Example: Without Markup

George Bush, the 41st President of the United States is the father of George W. Bush, the 43rd President of the United States.

http://schema.org/Person

Person Example: JSON-LD Syntax

Example 4

Without Markup | Microdata | RDFa | **JSON-LD**

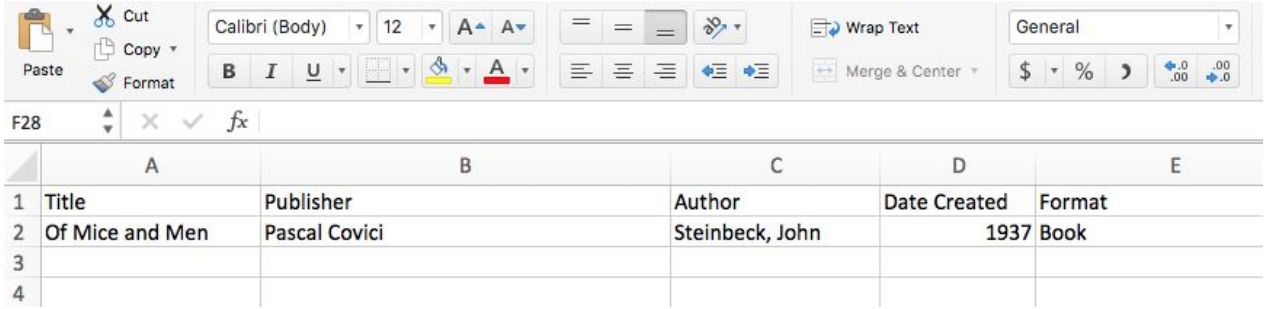
```
<script type="application/ld+json">
{
  "@context": "http://schema.org",
  "@type": "Person",
  "name": "George Bush",
  "disambiguatingDescription": "41st President of the United States",
  "children": {
    "@type": "Person",
    "name": "George W. Bush",
    "disambiguatingDescription": "43rd President of the United States"
  }
}
</script>
```

http://schema.org/Person

JSON Basics

- JavaScript Object Notation is a data format
- Based on a subset of the JavaScript Programming Language
- Text based and language independent

JSON Example



	A	B	C	D	E
1	Title	Publisher	Author	Date Created	Format
2	Of Mice and Men	Pascal Covici	Steinbeck, John	1937	Book
3					
4					

```
{
"Title": "Of Mice and Men",
"Publisher": "Pascal Covici",
"Author": "Steinbeck, John",
"DateCreated": "1937",
"Format": "Book"
}
```

JSON-LD Extension

- Lightweight linked data format
- Extends JSON
- JSON-LD homepage

Describing a Painting

with schema.org in JSON-LD

JSON

{ }

JSON-LD

```
{
      "@context": "http://schema.org",
      "@type": "Painting"
}
```

http://schema.org/docs/jsonldcontext.json

JSON-LD with Properties

```
{
    "@context": "http://schema.org",
    "@type": "Painting",
    "name": "The Scream",
    "url": "https://en.wikipedia.org/wiki/The_Scream",
    "creator":
}
```

Nested Properties

creator	Organization or Person	The creator/author of this CreativeWork. This is the same as the Author property for CreativeWork.

JSON-LD with Nesting

```
{
    "@context": "http://schema.org",
    "@type": "Painting",
    "name": "The Scream",
    "url": "https://en.wikipedia.org/wiki/The_Scream",
    "creator": {
          "@type": "Person",
          "name": "Edvard Munch",
          "sameAs": "https://en.wikipedia.org/wiki/Edvard_Munch"
    }
}
```

NCSU Schema.org Digital Collections Project

Contact us about this resource

Item information

Title:
Going Places: Students walking by the Bell Tower

Image from https://d.lib.ncsu.edu/collections/catalog/0227879

Publishing Workflow

Extract

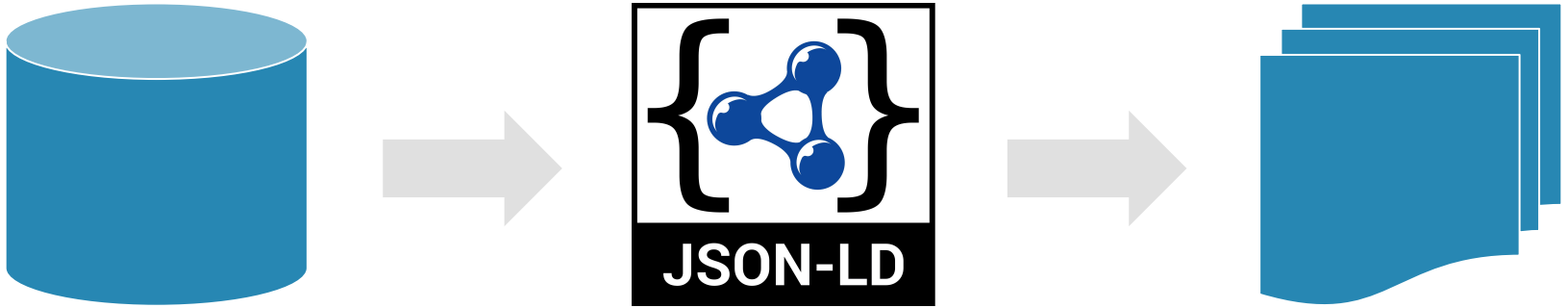
- Data from source

Transform

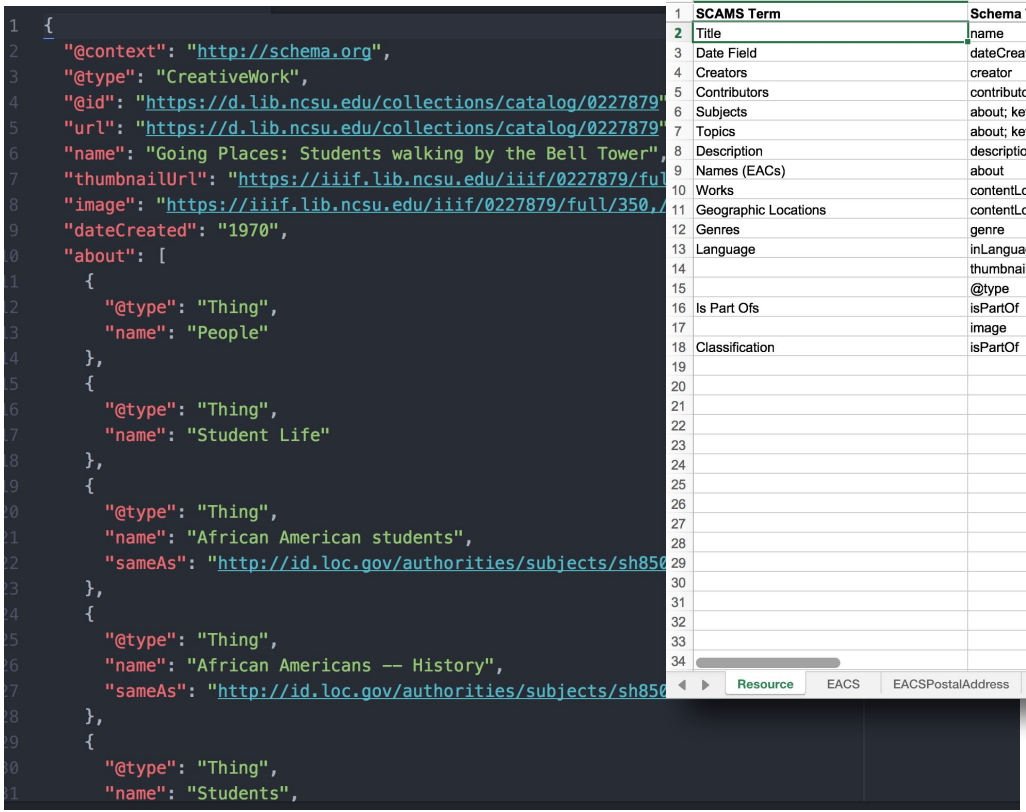
- Data into schema.org + semantic markup

Load

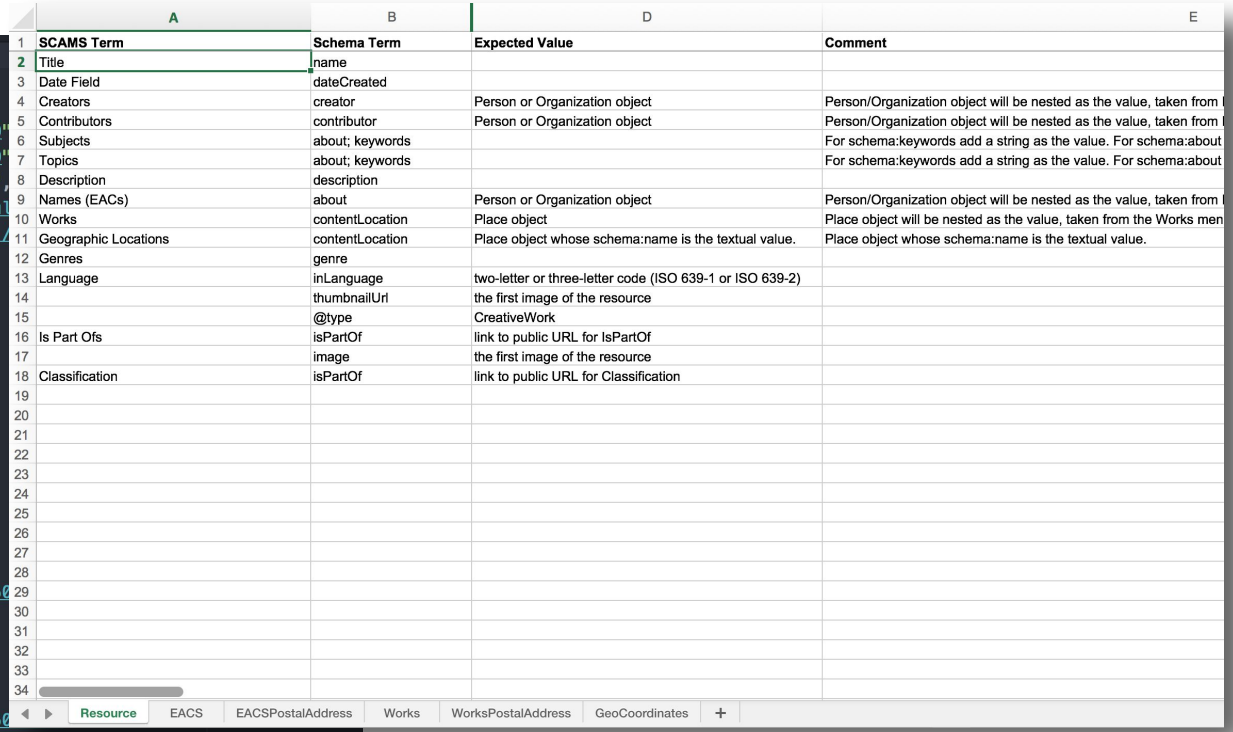
- Transformed data into HTML documents



Key Components for ETL Process



```
{
  "@context": "http://schema.org",
  "@type": "CreativeWork",
  "@id": "https://d.lib.ncsu.edu/collections/catalog/0227879"
  "url": "https://d.lib.ncsu.edu/collections/catalog/0227879"
  "name": "Going Places: Students walking by the Bell Tower",
  "thumbnailUrl": "https://iiif.lib.ncsu.edu/iiif/0227879/ful
  "image": "https://iiif.lib.ncsu.edu/iiif/0227879/full/350,/
  "dateCreated": "1970",
  "about": [
    {
      "@type": "Thing",
      "name": "People"
    },
    {
      "@type": "Thing",
      "name": "Student Life"
    },
    {
      "@type": "Thing",
      "name": "African American students",
      "sameAs": "http://id.loc.gov/authorities/subjects/sh850
    },
    {
      "@type": "Thing",
      "name": "African Americans -- History",
      "sameAs": "http://id.loc.gov/authorities/subjects/sh850
    },
    {
      "@type": "Thing",
      "name": "Students",
```



	A	B	D	E
1	**SCAMS Term**	**Schema Term**	**Expected Value**	**Comment**
2	Title	name		
3	Date Field	dateCreated		
4	Creators	creator	Person or Organization object	Person/Organization object will be nested as the value, taken from
5	Contributors	contributor	Person or Organization object	Person/Organization object will be nested as the value, taken from
6	Subjects	about; keywords		For schema:keywords add a string as the value. For schema:about
7	Topics	about; keywords		For schema:keywords add a string as the value. For schema:about
8	Description	description		
9	Names (EACs)	about	Person or Organization object	Person/Organization object will be nested as the value, taken from
10	Works	contentLocation	Place object	Place object will be nested as the value, taken from the Works men
11	Geographic Locations	contentLocation	Place object whose schema:name is the textual value.	Place object whose schema:name is the textual value.
12	Genres	genre		
13	Language	inLanguage	two-letter or three-letter code (ISO 639-1 or ISO 639-2)	
14		thumbnailUrl	the first image of the resource	
15		@type	CreativeWork	
16	Is Part Ofs	isPartOf	link to public URL for IsPartOf	
17		image	the first image of the resource	
18	Classification	isPartOf	link to public URL for Classification	
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30				
31				
32				
33				
34				

Resource | EACS | EACSPostalAddress | Works | WorksPostalAddress | GeoCoordinates | +

Now to apply what we've learned

Activity Instructions

1. Download materials from GitHub
   a. https://github.com/eightBitter/lita-schema
2. Open `metadataFlattened.xlsx`
3. Look up the schema.org CreativeWork and Person
   a. CreativeWork: http://schema.org/CreativeWork
   b. Person: http://schema.org/Person
4. Open `sourceToSchema_mapping.xlsx`
5. Using the schema.org pages, create a metadata map from the source metadata to Schema.org
6. Open `schemaCreativeWork_template.json`
7. Using the source metadata and the map you just created, create a JSON-LD template
8. You can validate the using the Google Structured Data Testing Tool:
   a. https://search.google.com/structured-data/testing-tool

Questions?

Jacob Shelby

@jacob_shelby421

jtshelby@ncsu.edu