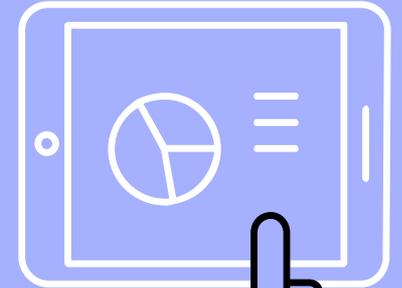
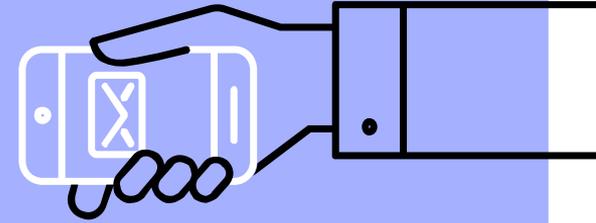


Linked Open Data Production and Publishing Workflow at the University of Washington Libraries

Benjamin Riesenber, Linked Data Specialist
Crystal Clements, Science Cataloger
Theodore Gerontakos, Metadata Librarian

University of Washington Libraries



Forming a Team

- ▶ Staff collaboration is helpful
- ▶ Assemble a team with diverse skill sets
- ▶ If possible, a person knowledgeable in linked data should be involved
- ▶ Skills can shape workflows
- ▶ Who's prepared for this work?
 - What is pertinent experience?
 - Willingness to learn on the fly



Learning Linked Data Skills On The Fly

- ▶ Mentorship
- ▶ Coursework
- ▶ Study groups
- ▶ Hands-on learning
- ▶ Guided development and study time
- ▶ Recycle!
- ▶ Consult helpful resources
- ▶ Develop skills based on tasks



Team Activities

- ▶ Meetings, meetings, meetings!
- ▶ Discuss models
- ▶ Updates
- ▶ Review and give feedback
- ▶ Forge a way forward and assign tasks
- ▶ Skill sharing and flexibility

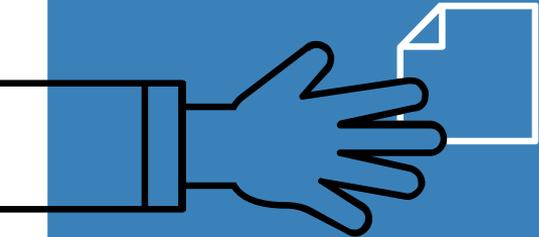
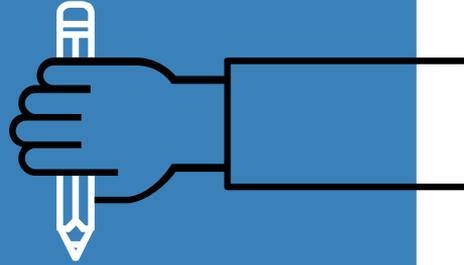


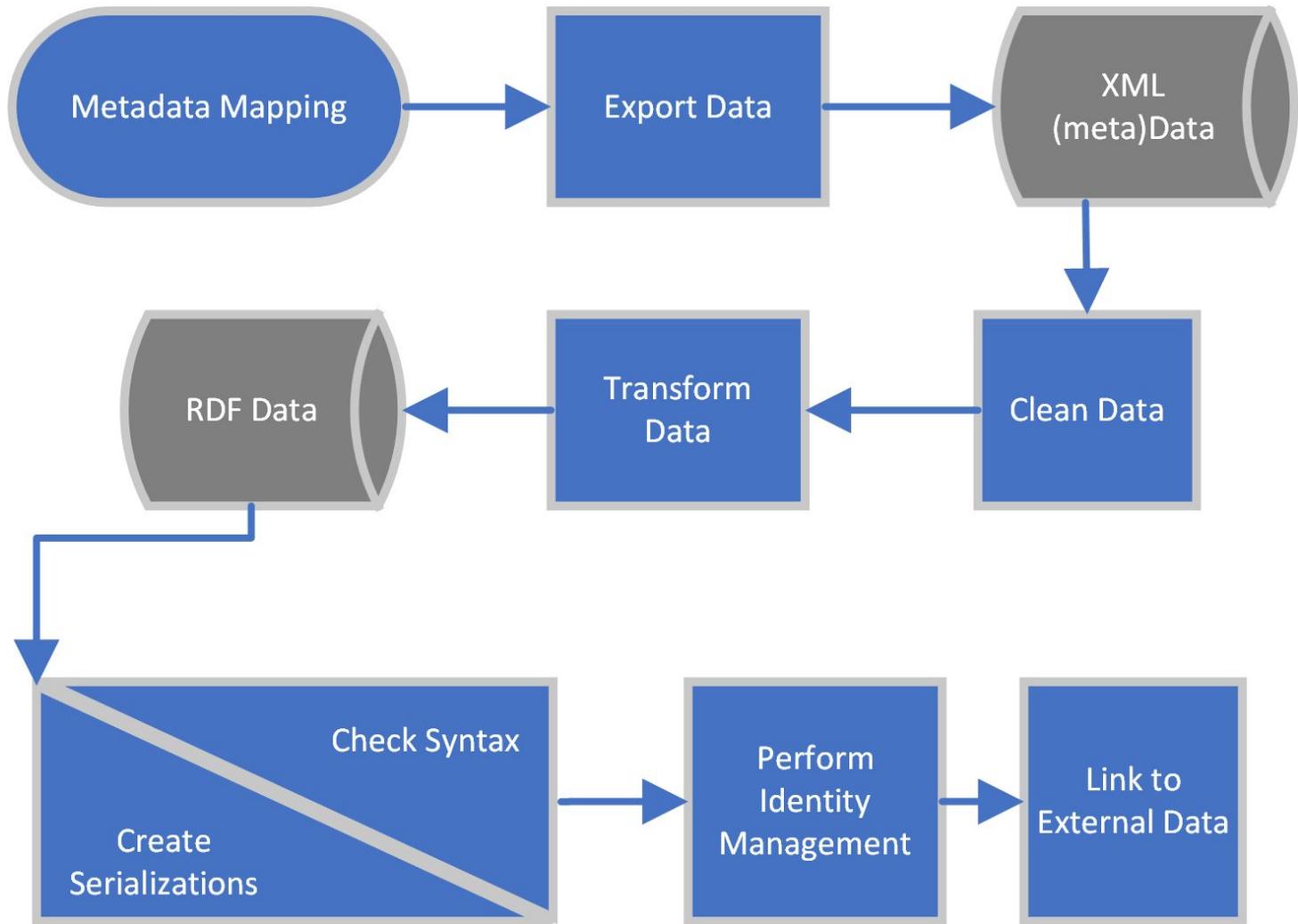
Application of Models

- ▶ Consult W3C Best Practices for Data on the Web
- ▶ Iterative process
- ▶ Flexibility
- ▶ External factors can change direction
- ▶ Teamwork
- ▶ It's never done!



The Workflow







Metadata Mapping

Metadata Application Profile, version 5.0

12/7/2017

DOCUMENT URI: <http://dp.la/info/map>

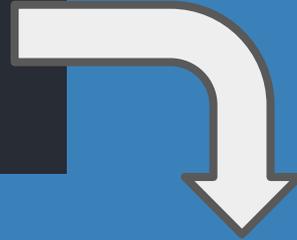
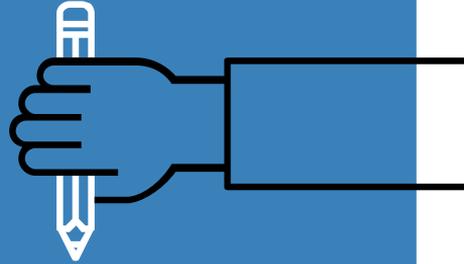
2.2. EDM Classes

2.2.1. Agent

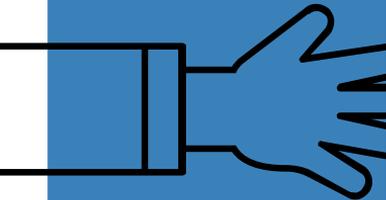
Class name: edm:Agent	
URI	http://www.europeana.eu/schemas/edm/Agent
Label	Agent
Definition	This class comprises people, either individually or in groups, who have the potential to perform intentional actions for which they can be held responsible.
Subclass of	edm:NonInformationResource
Equivalent class	E39_Actor (CIDOC CRM)

Exporting and Cleaning Data

```
<SubjectsLcsh>Simonds Manufacturing Company--Exhibits & </SubjectsLcsh>  
<SubjectsLcsh> </SubjectsLcsh>  
<SubjectsLcsh>displays</SubjectsLcsh>  
<SubjectsLcsh>Advertising--Washington (State)--Seattle</SubjectsLcsh>  
<SubjectsLcsh>Saws</SubjectsLcsh>  
<SubjectsLcsh>Alaska-Yukon-Pacific Exposition (1909 : Seattle,  
Wash.)</SubjectsLcsh>  
<SubjectsLcsh>Exhibitions--Washington (State)--Seattle</SubjectsLcsh>
```



```
<SubjectsLcsh>Simonds Manufacturing Company--Exhibits & displays;  
Advertising--Washington (State)--Seattle; Saws; Alaska-Yukon-Pacific  
Exposition (1909 : Seattle, Wash.); Exhibitions--Washington  
(State)--Seattle</SubjectsLcsh>
```



		nodes with no value for dpla:providedLabel)	
5	replace:	; </SubjectsLcsh>	
	with:	</SubjectsLcsh>	
	NOTE:	Some text values in these subjects field ended with a semicolon prior to the closing tag. Due to tokenizing in the XSL this was outputting empty blank nodes for LCSH and LCTGM subjects (blank nodes with no value for dpla:providedLabel)	
6	replace:	<Restrictions>5928,rights,	
	with:	<Restrictions>	
7	search:	<RepositoryCollection>.*
	
	manual fix:	[Create additional set of <RepositoryCollection> elements and enclose content after the break in them. Delete break characters "
"]	
	NOTE:	This search string uses regular expression syntax .* to allow for multiple characters between the XML element and the break characters.	

Data Transformation

```
<!-- template for agent file -->
```

```
<xsl:template match="met
```

```
<xsl:for-each-group
```

```
group-by="Photog
```

```
<rdf:Description
```



```
id text() != 'unidentified'
```

SAXON

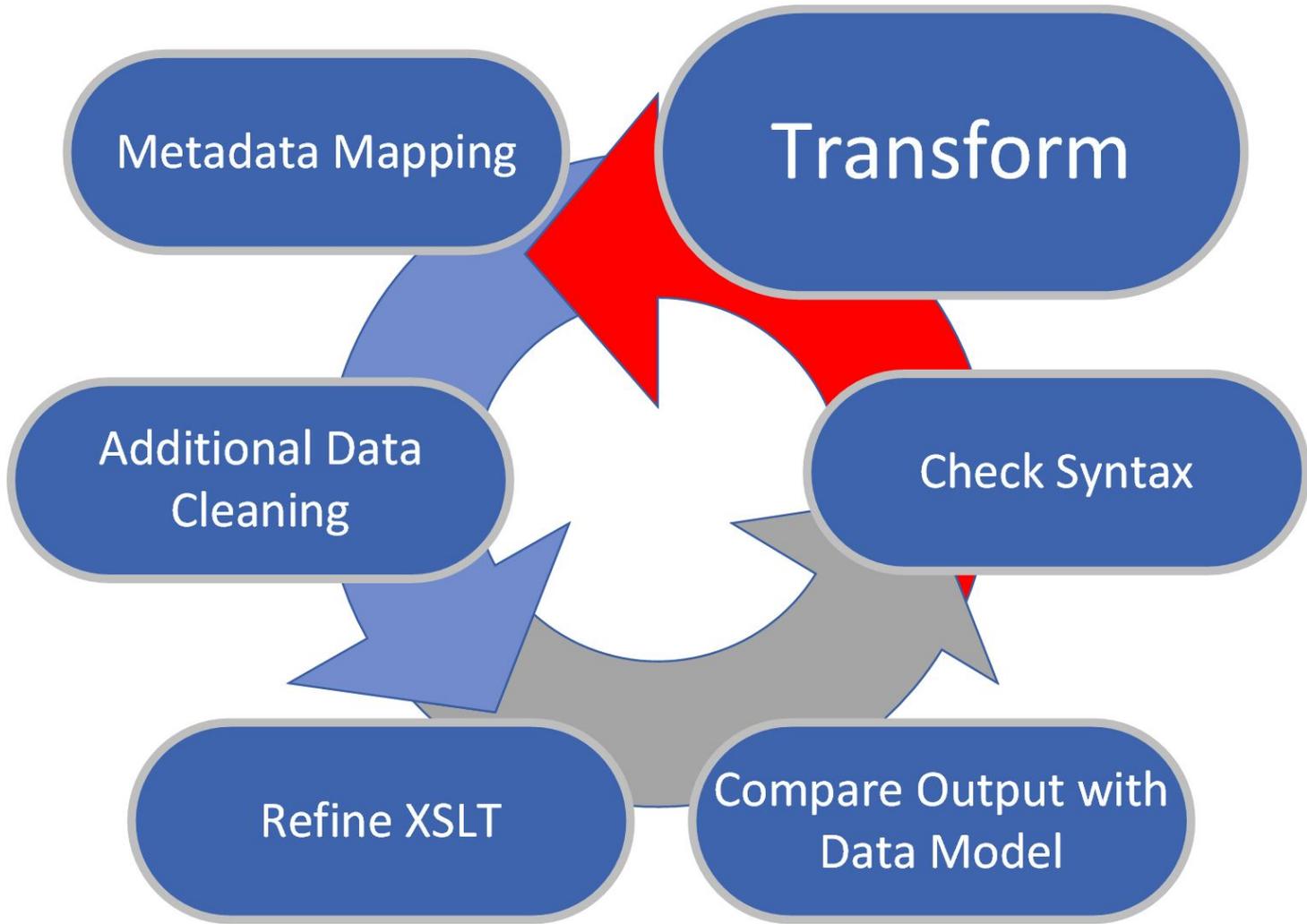
The XSLT and XQuery Processor

Contact Information: see <http://www.saxonica.com/about/contact.xml>

This is the home page for the Open Source SAXON XSLT processor developed by Saxonica Limited. It also commercial versions of the Saxon product available from Saxonica.

```
</rdf:Description>
```

```
</xsl:for-each-group>
```





In this commit: ...

- Added triple to express `rdf:type = dpla:SourceResource` to `sourceResource` sub-template
- "Cleaned up" the transform, specifically
 - sub-templates for `sourceResource` and `webResource`:
 - Moving all `rdf:Description` elements down to `CONTENTdm` element templates
 - Agent, Aggregation, etc. files appear in alphabetical order in transform to match the order in which output files are saved to file

briesenberg07 commented on Jul 23, 2018

Contributor



Aggregation file

As with output web resources file, currently output aggregations do not have any `rdf:type` properties.
Add `rdf:type = ore:aggregation`?

Checking Syntax of Output RDF Data and Creating Alternate and Canonical Serializations

Redland RDF Libraries

OPTIONS

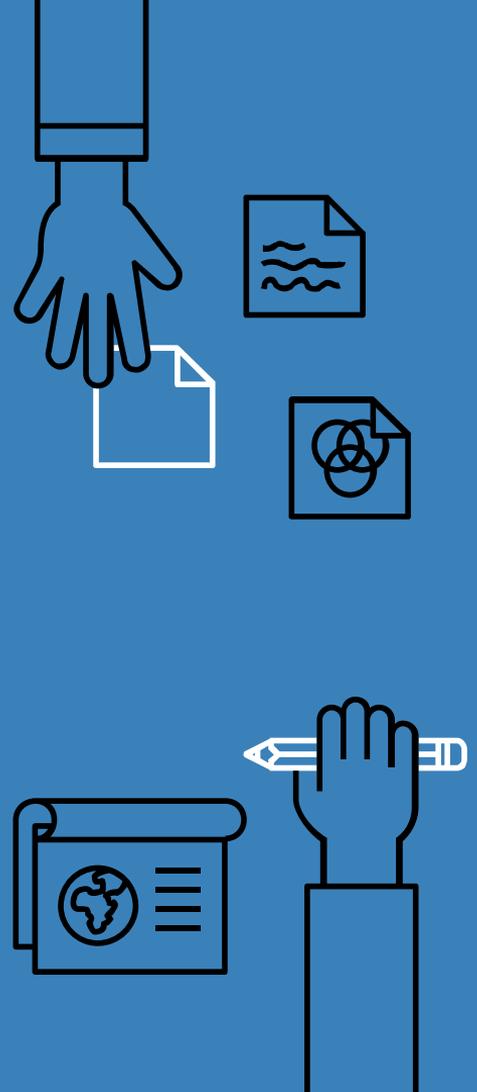
raper uses the usual GNU command line syntax, with long options starting with two dashes ('-') if supported by the getopt_long function. Otherwise the short options are only available.

-h, --help

Show a summary of the options.

-i, --input *FORMAT*

Set the input *FORMAT* to one of 'rdfxml' (RDF/XML, default), 'ntriples' (N-Triples, see below), 'turtle' (Turtle, see below) or 'rss-tag-soup' (RSS Tag Soup). The RSS Tag Soup parser can turn the many XML RSS formats and Atom 0.3 into RDF triples.



Identity Management

on



A free and open source Java framework for building [Semantic Web](#) and [Linked Data](#) applications.

W3C Rec

SPARQL 1.1 Update

Fuseki

Expose your triples as a SPARQL end-point accessible over HTTP. Fuseki provides REST-style interaction with your RDF data.

Linking to External Datasets

 **OpenRefine** aypSrTest01 [Permalink](#)



Item Discussion

Asahel Curtis (Q4803332)

exact match



<https://doi.org/10.6069/uwlib.55.A.3.6#CurtisAsahel18741941>

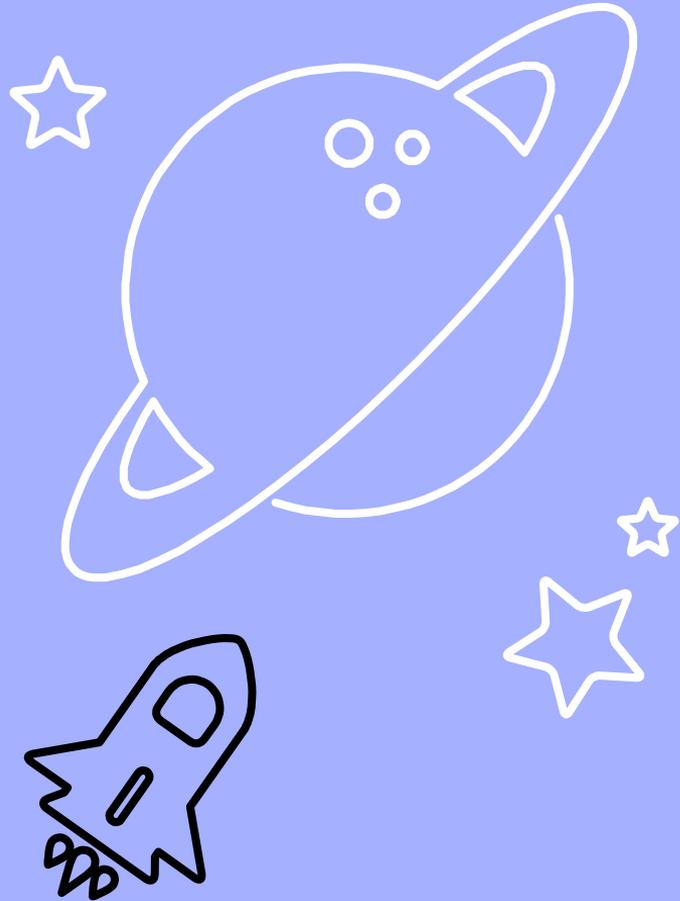
 edit

▼ 0 references

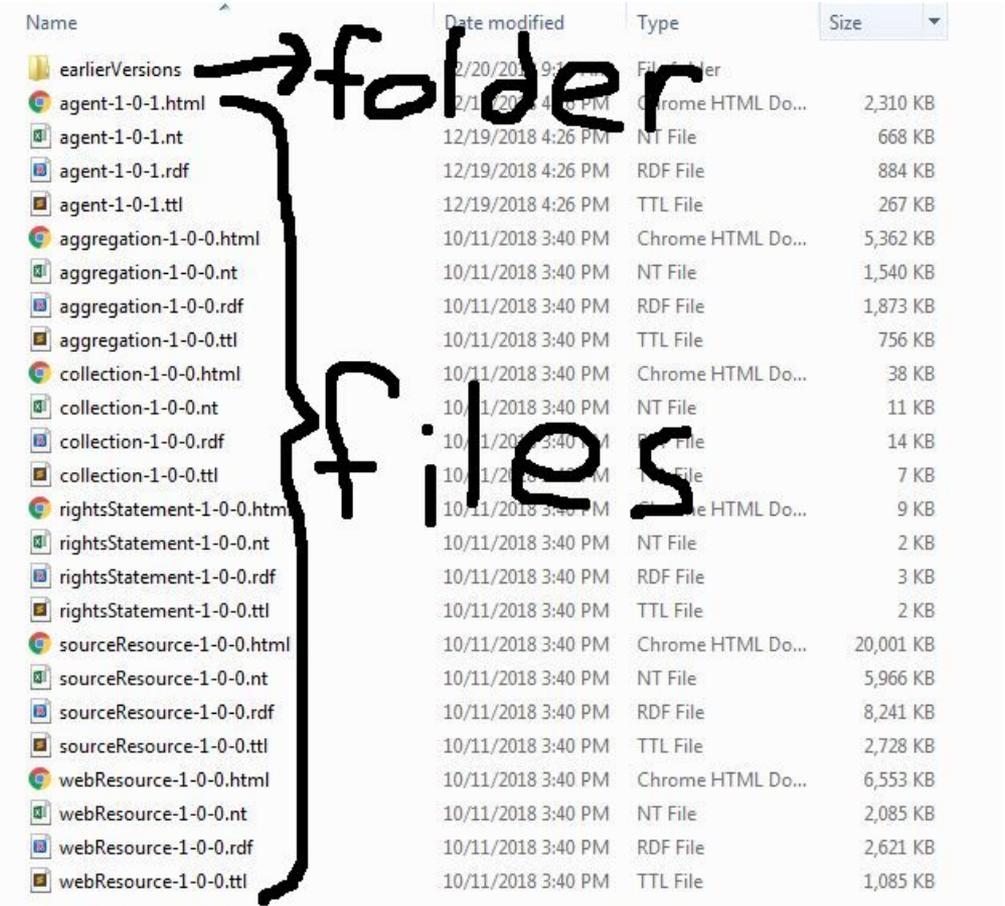
+ add reference

+ add value

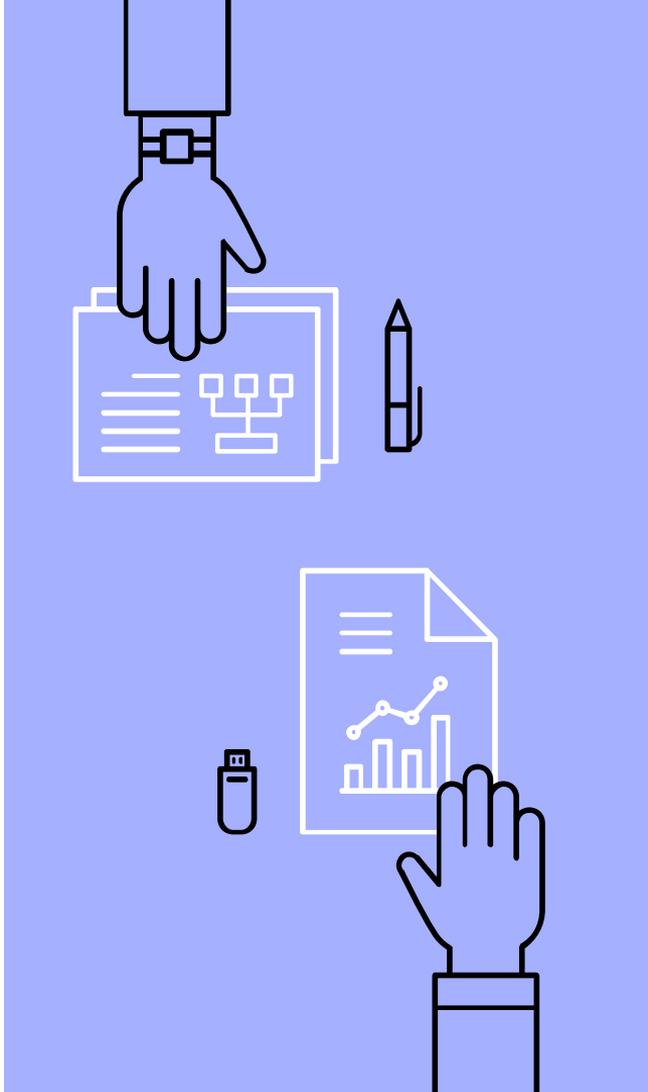
Publishing Model



We Publish Files



Name	Date modified	Type	Size
earlierVersions	12/20/2018 9:30 AM	Folder	
agent-1-0-1.html	12/19/2018 4:26 PM	Chrome HTML Do...	2,310 KB
agent-1-0-1.nt	12/19/2018 4:26 PM	NT File	668 KB
agent-1-0-1.rdf	12/19/2018 4:26 PM	RDF File	884 KB
agent-1-0-1.ttl	12/19/2018 4:26 PM	TTL File	267 KB
aggregation-1-0-0.html	10/11/2018 3:40 PM	Chrome HTML Do...	5,362 KB
aggregation-1-0-0.nt	10/11/2018 3:40 PM	NT File	1,540 KB
aggregation-1-0-0.rdf	10/11/2018 3:40 PM	RDF File	1,873 KB
aggregation-1-0-0.ttl	10/11/2018 3:40 PM	TTL File	756 KB
collection-1-0-0.html	10/11/2018 3:40 PM	Chrome HTML Do...	38 KB
collection-1-0-0.nt	10/11/2018 3:40 PM	NT File	11 KB
collection-1-0-0.rdf	10/11/2018 3:40 PM	RDF File	14 KB
collection-1-0-0.ttl	10/11/2018 3:40 PM	TTL File	7 KB
rightsStatement-1-0-0.html	10/11/2018 3:40 PM	Chrome HTML Do...	9 KB
rightsStatement-1-0-0.nt	10/11/2018 3:40 PM	NT File	2 KB
rightsStatement-1-0-0.rdf	10/11/2018 3:40 PM	RDF File	3 KB
rightsStatement-1-0-0.ttl	10/11/2018 3:40 PM	TTL File	2 KB
sourceResource-1-0-0.html	10/11/2018 3:40 PM	Chrome HTML Do...	20,001 KB
sourceResource-1-0-0.nt	10/11/2018 3:40 PM	NT File	5,966 KB
sourceResource-1-0-0.rdf	10/11/2018 3:40 PM	RDF File	8,241 KB
sourceResource-1-0-0.ttl	10/11/2018 3:40 PM	TTL File	2,728 KB
webResource-1-0-0.html	10/11/2018 3:40 PM	Chrome HTML Do...	6,553 KB
webResource-1-0-0.nt	10/11/2018 3:40 PM	NT File	2,085 KB
webResource-1-0-0.rdf	10/11/2018 3:40 PM	RDF File	2,621 KB
webResource-1-0-0.ttl	10/11/2018 3:40 PM	TTL File	1,085 KB

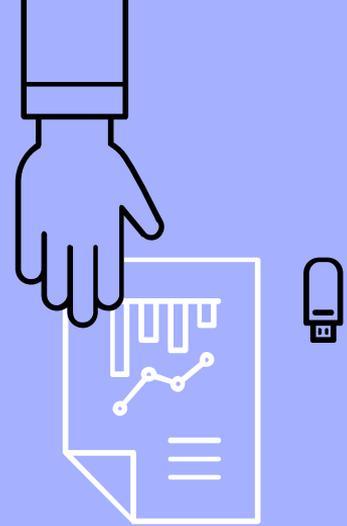
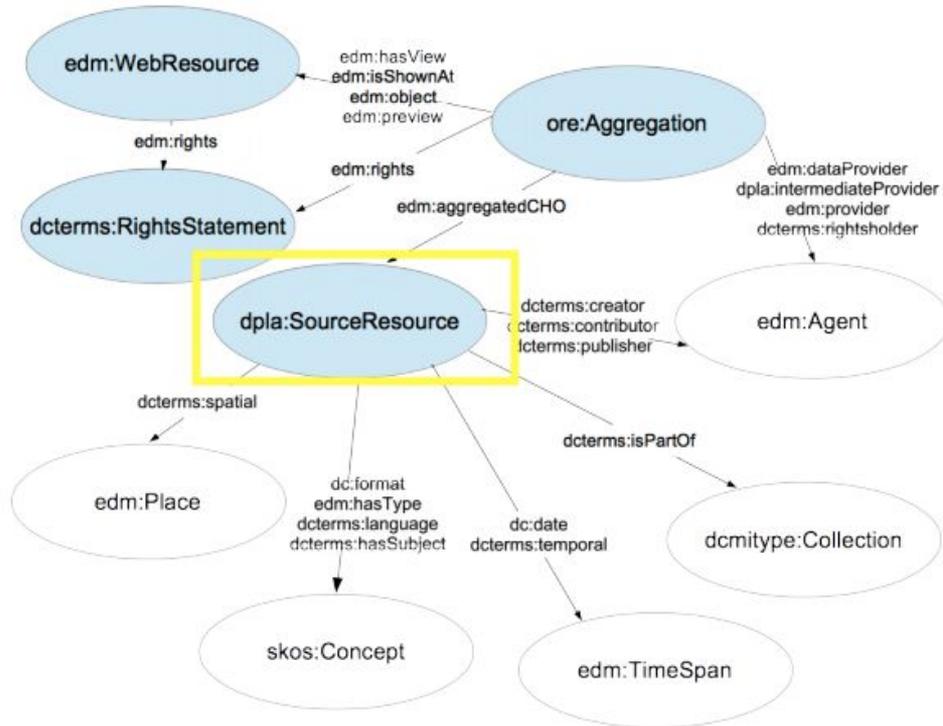


File Contents Determined by DPLA Types

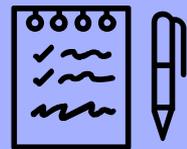
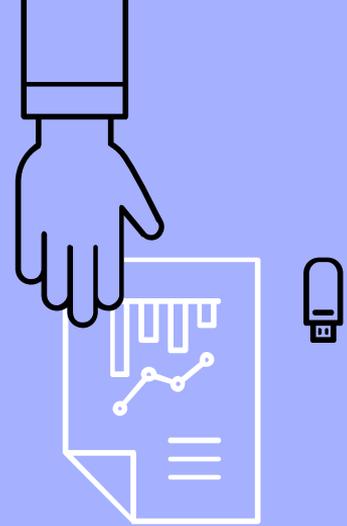
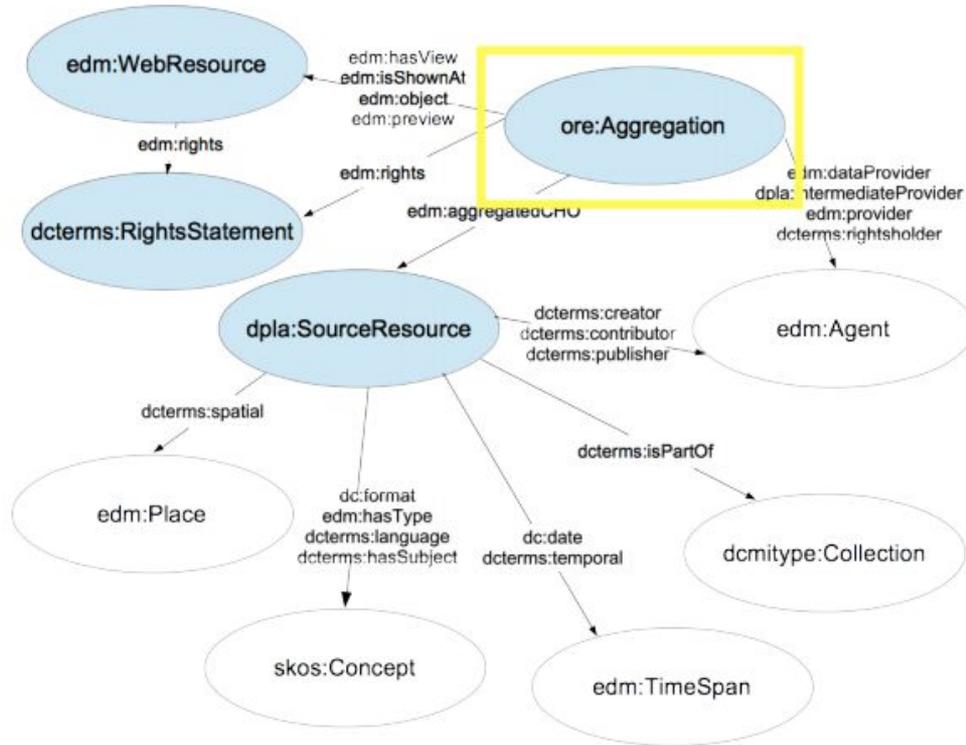
Files are organized by DPLA types, including:

- `dpla:SourceResource`
- `edm:WebResource`
- `ore:Aggregation`

dpla:SourceResource

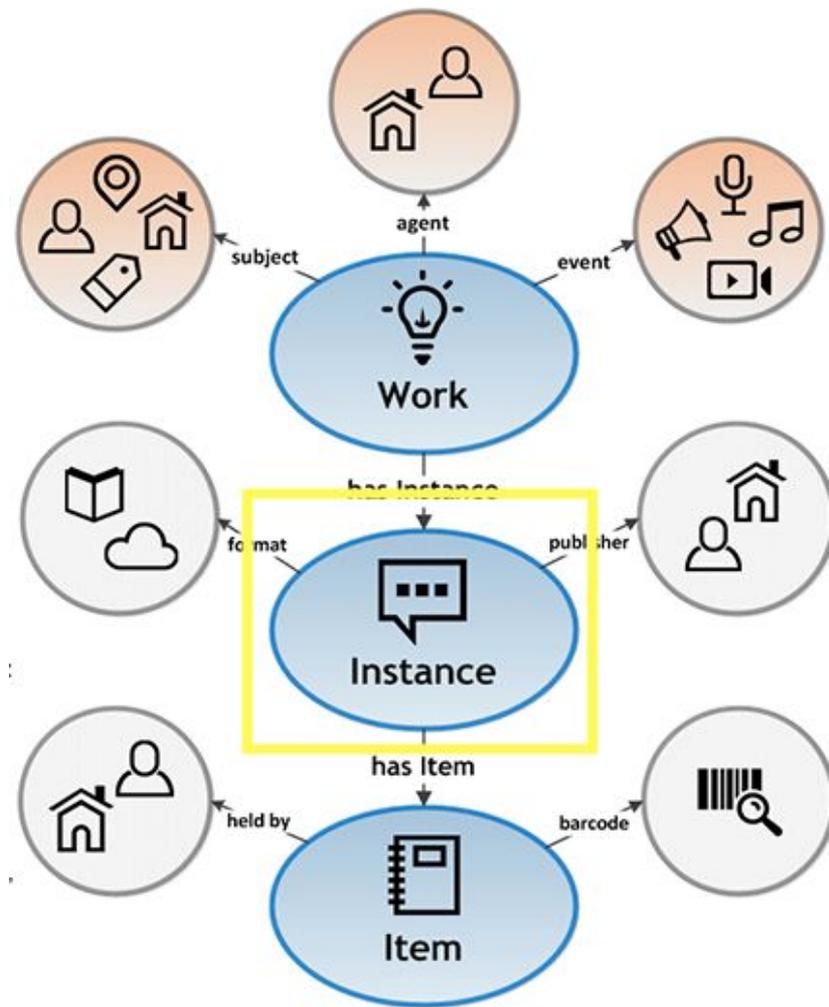


ore:Aggregation



If we use BIBFRAME...

...we might need a
file describing
resources typed as
bf:Instance



File Format: HTML+RDFa



HTML+RDFa 1.1 - Second Edition

Support for RDFa in HTML4 and HTML5

W3C Recommendation 17 March 2015



Additional Format: N-Triples



RDF 1.1 N-Triples

A line-based syntax for an RDF graph

W3C Recommendation 25 February 2014



Additional Format: RDFXML

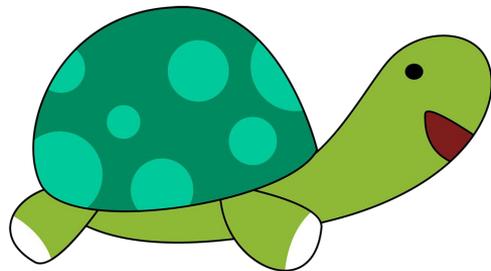


RDF 1.1 XML Syntax

W3C Recommendation 25 February 2014



Additional Format: Turtle



RDF 1.1 Turtle

Terse RDF Triple Language

W3C Recommendation 25 February 2014



HTML+RDFa File

Assigned a DOI

 Update DOI (Form)

 Update DOI (File Upload)

Findable

Metadata Export

DataCite XML

DataCite JSON

Schema.org JSON-LD

BibTeX

DOI created

August 18, 2018, 03:43:15 UTC

DOI registered

August 18, 2018, 03:44:17 UTC

DOI last updated

January 2, 2019, 03:32:29 UTC

Schema

DataCite Metadata Schema 4

10.6069/uwlib.55.b.2

URL

<https://www.lib.washington.edu/static/public/cams/data/vocabularies/reconVocab.html>

Metadata

Summary View ▾

RDF Vocabulary for Linked Data Processes Dataset

University Of Washington Libraries

LibraryData published 2018 via University of Washington Libraries

The RDF triples below define a vocabulary for describing processes run on linked data datasets. This vocabulary was originally intended as administrative and provenance data for RDF datasets produced at the University of Washington Libraries.



 <https://doi.org/10.6069/uwlib.55.b.2>

Citation

APA ▾

University Of Washington Libraries. (2018). *RDF Vocabulary for Linked Data Processes* [Data set]. University of Washington Libraries.

<https://doi.org/10.6069/UWLIB.55.B.2>

DOIs

- Each file contains many resources of a given type
- Each resource IRI is a hash IRI with the DOI used as the base

```
<https://doi.org/10.6069/uwlib.55.A.3.6#AbramychevPI>  
  a vra:Person, edm:Agent .
```

```
<https://doi.org/10.6069/uwlib.55.A.3.6#AbulmanovN>  
  a vra:Person, edm:Agent .
```

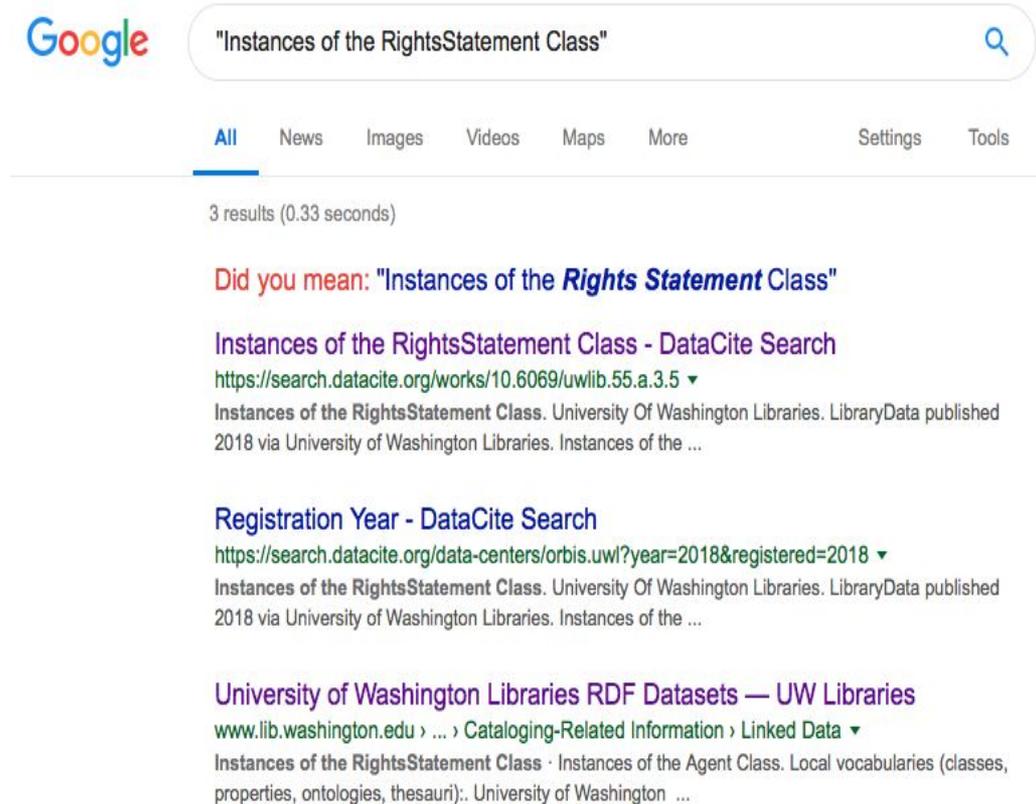
```
<https://doi.org/10.6069/uwlib.55.A.3.6#AgishevMKh>  
  a vra:Person, edm:Agent .
```

```
<https://doi.org/10.6069/uwlib.55.A.3.6#AidarovSS>  
  a vra:Person, edm:Agent .
```

```
<https://doi.org/10.6069/uwlib.55.A.3.6#AkulovEA>  
  a vra:Person, edm:Agent .
```

DataCite

- Increases visibility
- Datacite requires XML metadata
- Datacite's XML metadata can be found by search engines



Google

"Instances of the RightsStatement Class" 🔍

All News Images Videos Maps More Settings Tools

3 results (0.33 seconds)

Did you mean: "Instances of the *Rights Statement Class*"

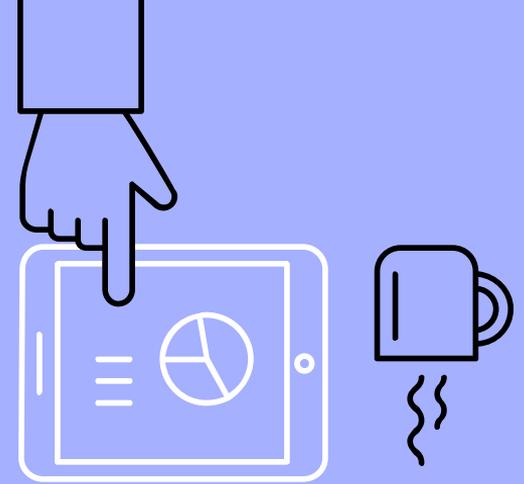
Instances of the RightsStatement Class - DataCite Search
<https://search.datacite.org/works/10.6069/uwlib.55.a.3.5> ▼
Instances of the RightsStatement Class. University Of Washington Libraries. LibraryData published 2018 via University of Washington Libraries. Instances of the ...

Registration Year - DataCite Search
<https://search.datacite.org/data-centers/orbis.uwl/?year=2018®istered=2018> ▼
Instances of the RightsStatement Class. University Of Washington Libraries. LibraryData published 2018 via University of Washington Libraries. Instances of the ...

University of Washington Libraries RDF Datasets — UW Libraries
www.lib.washington.edu > ... > [Cataloging-Related Information](#) > [Linked Data](#) ▼
Instances of the RightsStatement Class · Instances of the Agent Class. Local vocabularies (classes, properties, ontologies, thesauri): University of Washington ...

HTML+RDFa File Purpose

- The HTML+RDFa file serves as
 - Dataset landing page
 - Instance of the dataset itself



HTML+RDFa File as a Landing Page

Instances of the Agent Class → TITLE

DESCRIPTION→Instances of the class <http://www.europeana.eu/schemas/edm/Agent> represented in the University of Washington digital collections; the class is taken from the Digital Public Library of America Metadata Application Profile 5.0

Backlink → heading for “backlink,” a term from W3C’s “Describing Linked Datasets with the VOID Vocabulary,” section 6.3

This dataset is part of the dataset University of Washington Libraries' Semantic Web Data

Links to Alternate Serializations for Instances of the Agent Class

- N-Triples → link to dataset as N-Triples
- RDF/XML → link to dataset as RDF/XML
- Turtle → link to dataset as Turtle

Version information

Version 1-0-1 issued 2018 → dataset version information

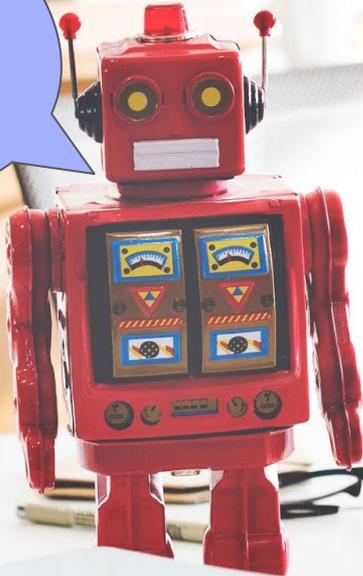
HTML+RDFa File as a Dataset for Humans

RDF Triples for Instances of the Agent Class

Subject	Predicate	Object
https://doi.org/10.6069/uwlib.55.A.3.6	rdf:type	http://rdfs.org/ns/void#Dataset
https://doi.org/10.6069/uwlib.55.A.3.6	void:inDataset	http://doi.org/10.6069/uwlib.55.a#uwSemWeb
https://doi.org/10.6069/uwlib.55.A.3.6#BannerPhotoCo	dpla:providedLabel	Banner Photo Co.
https://doi.org/10.6069/uwlib.55.A.3.6#BannerPhotoCo	dct:description	Studio
https://doi.org/10.6069/uwlib.55.A.3.6#BannerPhotoCo	rdf:type	http://www.europeana.eu/schemas/edm/Agent
https://doi.org/10.6069/uwlib.55.A.3.6#BannerPhotoCo	foaf:based_near	United States--Washington (State)--Seattle
https://doi.org/10.6069/uwlib.55.A.3.6#BannerPhotoCo	uwprop:inDigiColl	https://doi.org/10.6069/uwlib.55.A.3.4#AlaskaYukonPacificExpositionCollection
https://doi.org/10.6069/uwlib.55.A.3.6#BradleyLevi	dpla:providedLabel	Bradley, Levi
https://doi.org/10.6069/uwlib.55.A.3.6#		http://www.europeana.eu/schemas/edm/Agent

HTML+RDFa File as a Dataset for Machines

Oh, that makes sense now. Thx.



```
<h2>RDF Triples for Instances of the Agent Class</h2>
<table border="1" cellpadding="6">
  <tr>
    <th>Subject</th>
    <th>Predicate</th>
    <th>Object</th>
  </tr>
  <tr about="https://doi.org/10.6069/uwlib.55.A.3.6">
    <td><a href="https://doi.org/10.6069/uwlib.55.A.3.6">https://doi.org/10.6069/uwlib.55.A.3.6</a></td>
    <td><a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#type">rdf:type</a></td>
    <td property="rdf:type" resource="http://rdfs.org/ns/void#Dataset"><a href="http://rdfs.org/ns/void#Dataset">http://rdfs.org/ns/void#Dataset</a></td>
  </tr>
  <tr about="https://doi.org/10.6069/uwlib.55.A.3.6">
    <td><a href="https://doi.org/10.6069/uwlib.55.A.3.6">https://doi.org/10.6069/uwlib.55.A.3.6</a></td>
    <td><a href="http://rdfs.org/ns/void#inDataset">void:inDataset</a></td>
    <td property="void:inDataset" resource="http://doi.org/10.6069/uwlib.55.a#uwSemWeb"><a href="http://doi.org/10.6069/uwlib.55.a#uwSemWeb">http://doi.org/10.6069/uwlib.55.a#uwSemWeb</a></td>
  </tr>
  <tr about="https://doi.org/10.6069/uwlib.55.A.3.6#BannerPhotoCo">
    <td id="BannerPhotoCo"><a href="https://doi.org/10.6069/uwlib.55.A.3.6#BannerPhotoCo">https://doi.org/10.6069/uwlib.55.A.3.6#BannerPhotoCo</a></td>
    <td><a href="http://dp.la/about/map/providedLabel">dpla:providedLabel</a></td>
    <td property="dpla:providedLabel" xml:lang="">Banner Photo Co.</td>
  </tr>
  <tr about="https://doi.org/10.6069/uwlib.55.A.3.6#BannerPhotoCo">
    <td><a href="https://doi.org/10.6069/uwlib.55.A.3.6#BannerPhotoCo">https://doi.org/10.6069/uwlib.55.A.3.6#BannerPhotoCo</a></td>
    <td><a href="http://purl.org/dc/terms/description">dct:description</a></td>
    <td property="dct:description" xml:lang="">Studio</td>
  </tr>
  <tr about="https://doi.org/10.6069/uwlib.55.A.3.6#BannerPhotoCo">
    <td><a href="https://doi.org/10.6069/uwlib.55.A.3.6#BannerPhotoCo">https://doi.org/10.6069/uwlib.55.A.3.6#BannerPhotoCo</a></td>
```

HTML+RDFa as a Carrier of Datasets

Contains metadata to improve visibility on the web

Google Search supports structured data in the following formats, unless documented otherwise:

-“Understanding How Structured Data Works”
<https://developers.google.com/search/docs/guides/intro-structured-data>



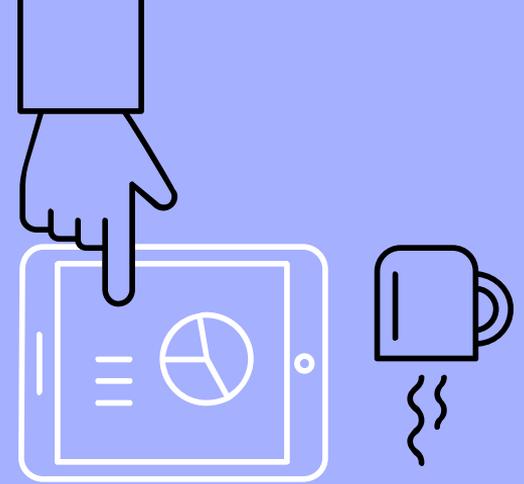
Format
JSON-LD* (Recommended)
Microdata
RDFa

HTML/head/script

```
<script type="application/ld+json">
{
  "@context" : "http://schema.org" ,
  "@type" : "Dataset" ,
  "@id" : "https://doi.org/10.6069/uwlib.55.a.3.6" ,
  "creator" : {
    "@type" : "Organization" ,
    "url" : "https://doi.org/10.6069/uwlib.55.A.3.6#UniversityofWashingtonLibraries"
  } ,
  "name" : "Instances of the Agent Class" ,
  "datePublished" : "2018" ,
}
</script>
```

Additional Serializations

- NTriples
- RDFXML
- Turtle
- Still no JSON-LD



Additional Serializations

- No DOIs
 - DOIs do not allow direct access to the data

<https://www.lib.washington.edu/static/public/cams/data/datasets/uwSemWebParts/agent-1-0-1.nt>

<https://www.lib.washington.edu/static/public/cams/data/datasets/uwSemWebParts/agent-1-0-1.rdf>

<https://www.lib.washington.edu/static/public/cams/data/datasets/uwSemWebParts/agent-1-0-1.ttl>

Metadata

- Already mentioned:
 - JSON-LD embedded in the HTML/head
- Additional metadata:
 - HTML <LINK> metadata embedded in the HTML/head
 - VoID metadata



DERI Vocabularies

Home

Vocabulary of Interlinked Datasets (VoID)

Authors:

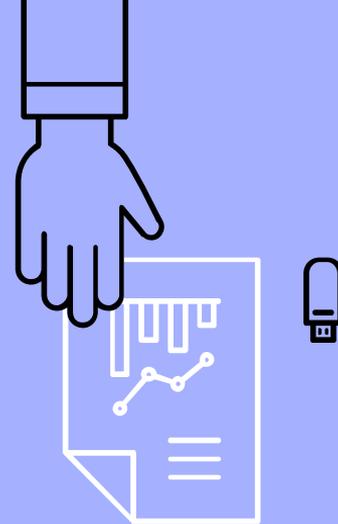
Richard Cyganiak (Digital Enterprise Research Institute, NUI Galway) (Email)
Jun Zhao (Department of Zoology, University of Oxford) (Email)
Keith Alexander (Talis) (Email)
Michael Hausenblas (LIDRC) (Email)

Last update:

06 March 2011

Namespace URI:

<http://rdfs.org/ns/void#>



VoID Metadata

Most often just 2 triples in each class-based file...

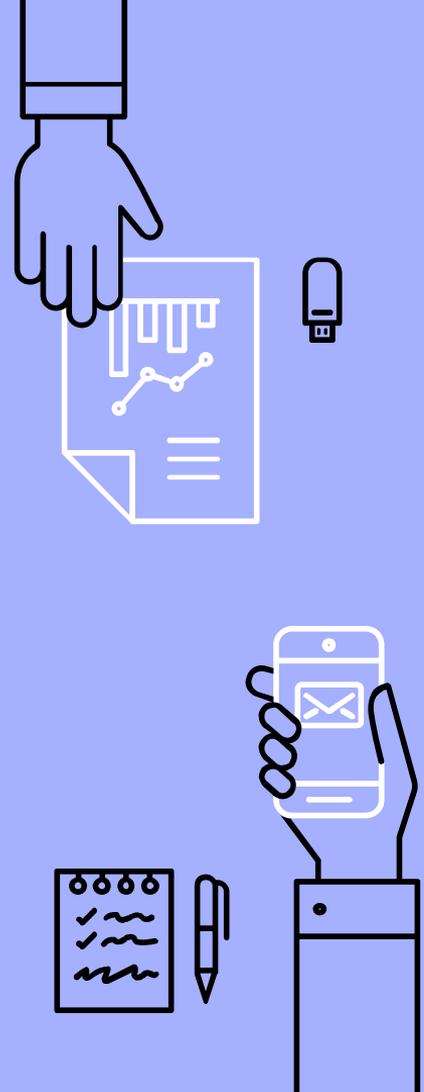
```
<https://doi.org/10.6069/uwlib.55.A.3.6>  
  void:inDataset  
<http://doi.org/10.6069/uwlib.55.a#uwSemWeb> ;  
  a          void:Dataset .
```



VoID Metadata

...referencing the a fuller description contained in the parent dataset description:

```
<https://doi.org/10.6069/uwlib.55.a.3.6>  
  dct:issued "2018" ;  
  dct:title "Instances of the Agent Class"@en ;  
  void:class edm:Agent ;  
  owl:version "1-0-1" ;  
  dct:hasFormat  
<https://www.lib.washington.edu/static/public/cams/data/data  
sets/uwSemWebParts/agent-1-0-1.nt>,  
<https://www.lib.washington.edu/static/public/cams/data/data  
sets/uwSemWebParts/agent-1-0-1.rdf>,  
<https://www.lib.washington.edu/static/public/cams/data/data  
sets/uwSemWebParts/agent-1-0-1.ttl> ;  
...
```



Create Links on the Libraries website at
<https://www.lib.washington.edu/msd/pubcat/ld>

University of Washington Libraries RDF Datasets

Hyperlinks to linked data datasets

Resources associated with University of Washington Digital Collections:

- [University of Washington Libraries' Semantic Web Data](#)
 - [Instances of the SourceResource Class](#)
 - [Instances of the AggregationClass](#)
 - [Instances of the WebResource Class](#)
 - [Instances of the Collection Class](#)
 - [Instances of the RightsStatement Class](#)
 - [Instances of the Agent Class](#)

Local vocabularies (classes, properties, ontologies, thesauri):

- [University of Washington Libraries rdfs:Class Definitions](#)
- [Linked Data Platform Vocabulary](#)
- [Linked Data Platforms](#)
- [Linked Data Vocabularies: Childrens Literature Categories](#)
- [RDF Vocabulary for Linked Data Processes](#)



Benjamin Riesenber,
Linked Data Specialist
ries07@uw.edu

Crystal Clements, *Science
Cataloger* cec23@uw.edu

Theodore Gerontakos,
Metadata Librarian
tgis@uw.edu

University of Washington Libraries



Thank You!