How, When, and Why to use Linked Data

Starting and Sustaining an Institutional Conversation

Museums and the Web: 8 April 2015

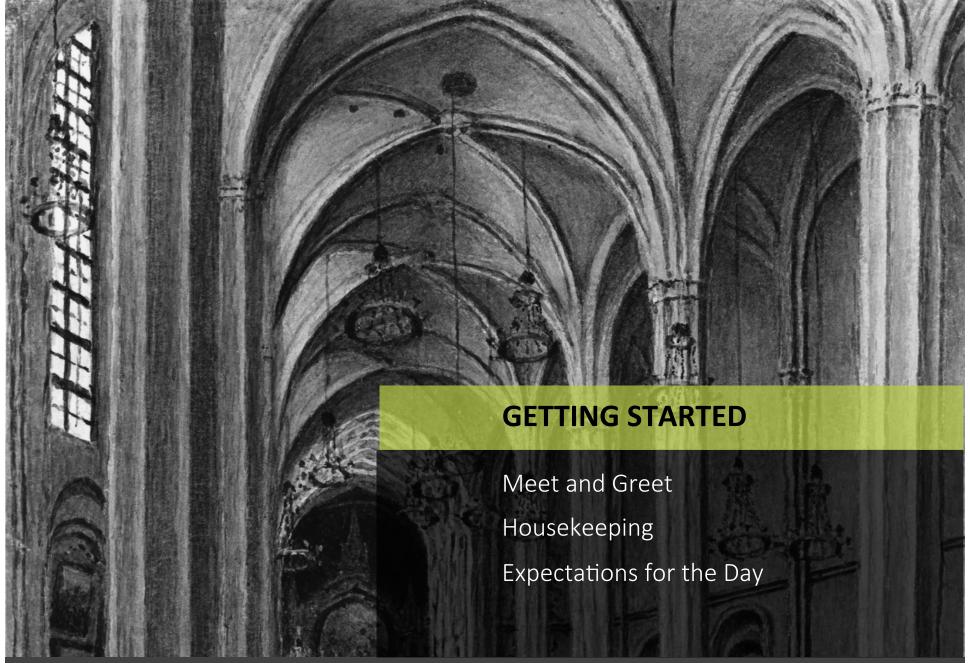


Duane Degler

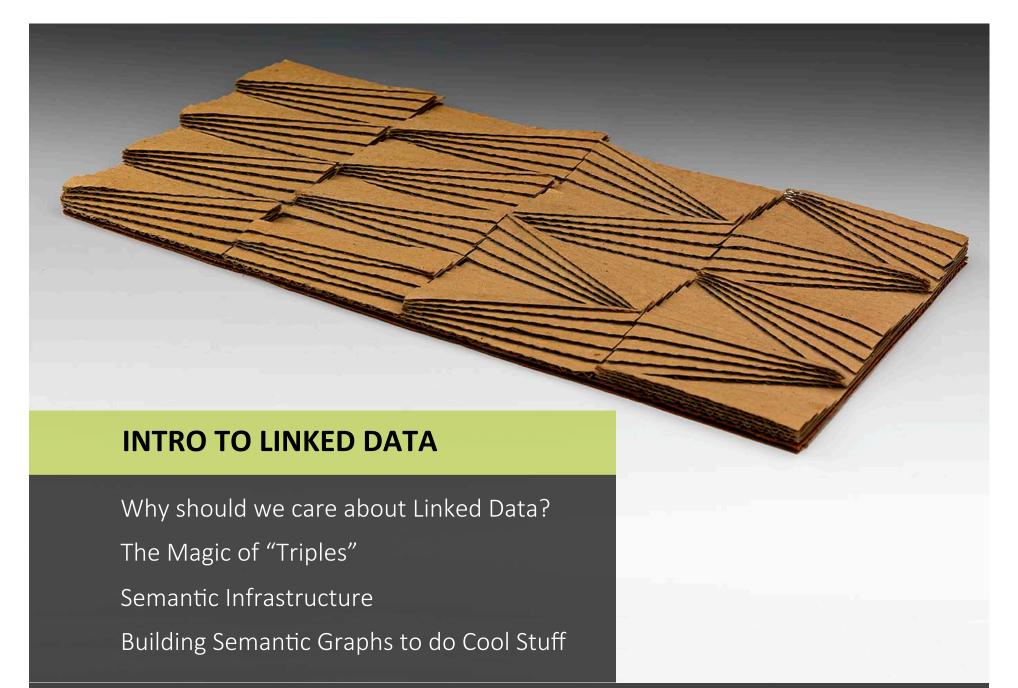
- @ddegler
- ✓ duane@designforcontext.com

Neal Johnson

- @vanWinkleTunes
- □ neal@designforcontext.com



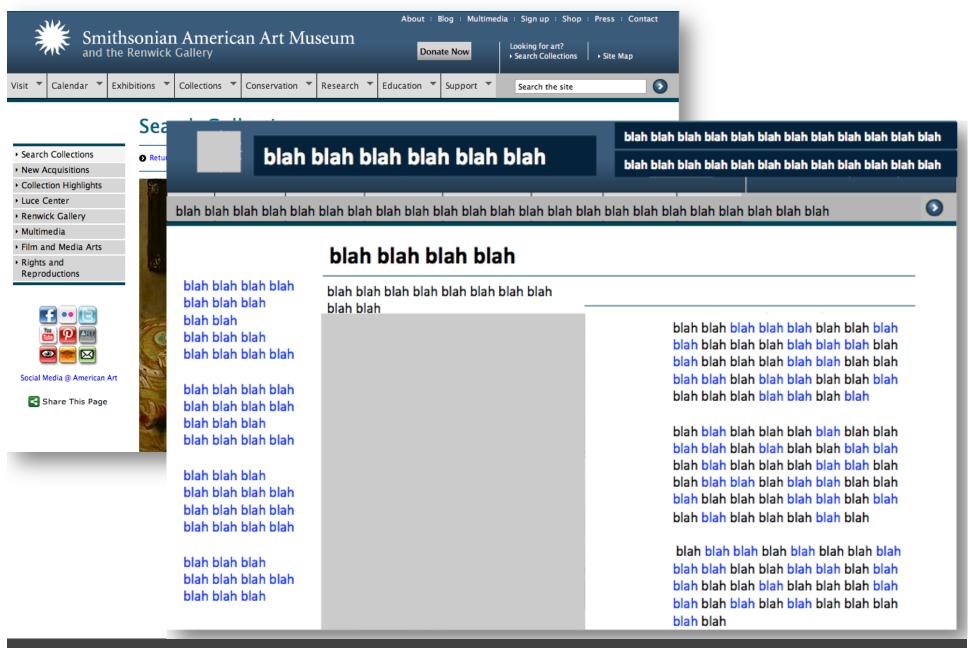
Jakob Alt. "Interior of a Gothic Church". 1864. The Walters Art Museum http://bit.lv/1wQpaiر



INTRO TO LINKED DATA

• Why should we care about Linked Data?

The Web with/without Linked Data



Carnegie Museum of Art program idea: Art Tracks



Carnegie Museum of Art provenance app: Elysa



Documentation



Madonna and Child Enthroned with SS. Leonard, Jerome, John the Baptist, and Francis [1472]

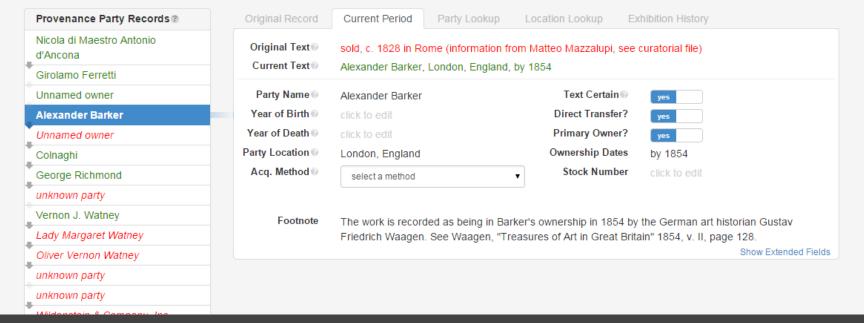
Nicola di Maestro Antonio d'Ancona

71.4

Nicola di Maestro Antonio d'Ancona; Possibly commissioned by Girolamo Ferretti, San Francesco delle Scale, Ancona, Italy. Unnamed owner, Rome, Italy, by 1828 [1].

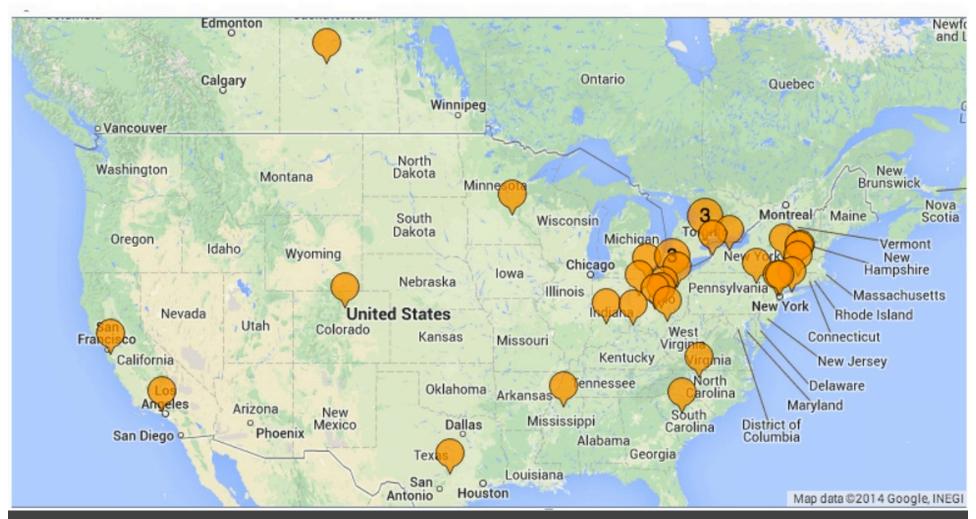
Alexander Barker, London, England, by 1854 [2]; Unnamed owner, Hall's Place, Canterbury, England, until 1880 [3]; (sold at Colnaghi, London, England); George Richmond, R.A; sale, Christie, Manson & Woods, London, Iot, May 1, 1897, Iot 46. Vernon J. Watney, Cornbury Park, 1899-?; Lady Margaret Watney, Cornbury Park, Charlbury; Oliver Vernon Watney, Cornbury Park, Charlbury; sale, Christie, Manson & Woods, Ltd., London, in "Highly Important Pictures by Old Masters", Iot, June 23, 1967. , 31 CE; Wildenstein & Company, Inc., New York; purchased by Museum, February 1971.

NOTES: [1] This information comes from correspondence with Matteo Mazzalupi, see curatorial file. [2] The work is recorded as being in Barker's ownership in 1854 by the German art historian Gustav Friedrich Waagen. See Waagen, "Treasures of Art in Great Britain" 1854, v. II, page 128. [3] Hall's Place, in Harbledown near Canterbury, was a country home that underwent several changes in ownership and had many residents. It is uncertain which individual or family purchased this work.



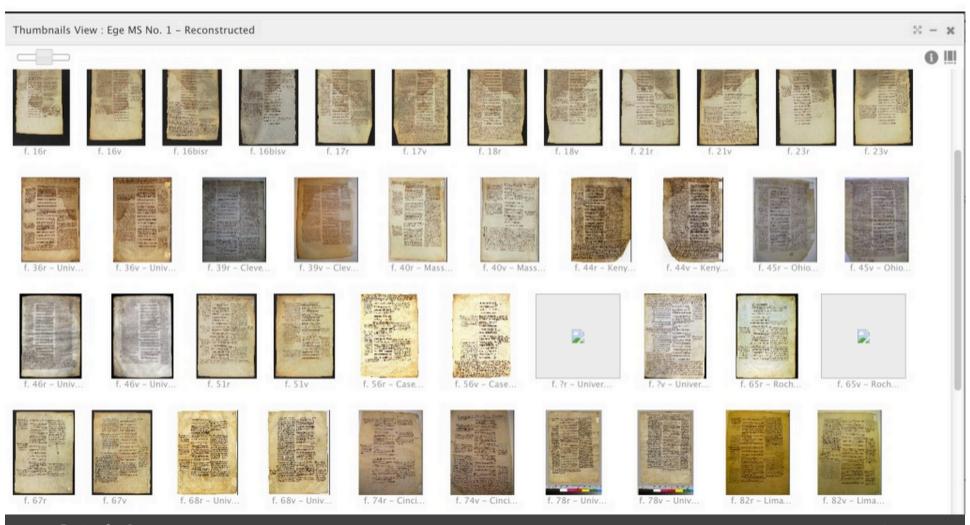
Connecting the diaspora of cultural materials:

Remnants of Otto Ege MS 1, as of 2014

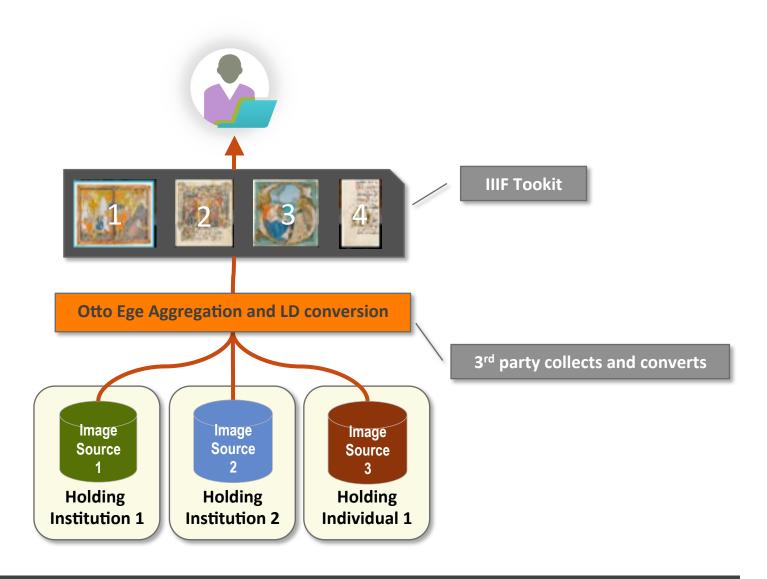


Connecting the diaspora of cultural materials:

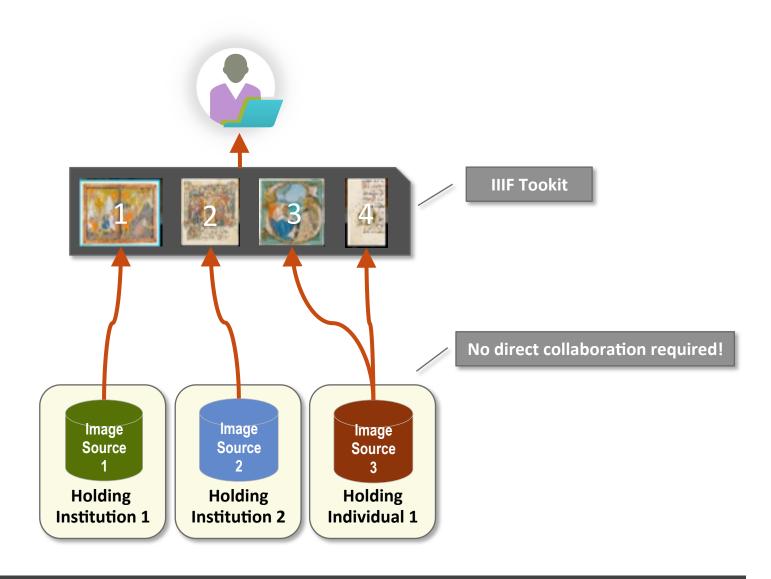
Otto Ege MS 1, Reconstructed, in 2014



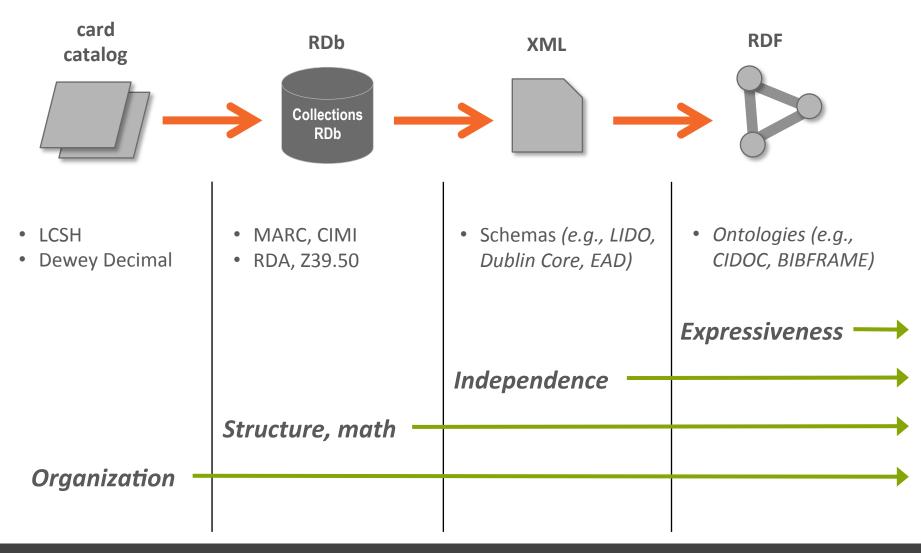
Otto Ege MS1: Aggregated model



Otto Ege MS1: Federated model



Linked Data is a "Next Gen" data standard



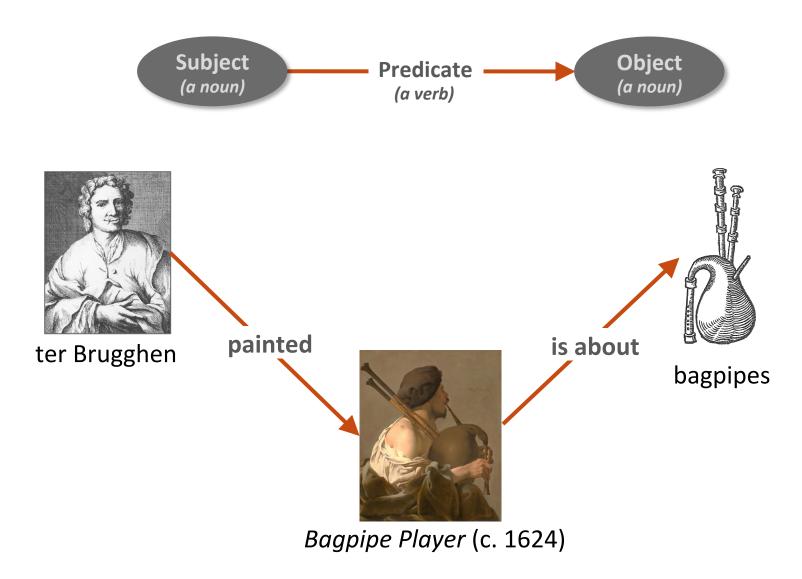
Linked Data value for cultural heritage

- Do more with my content
 - Eliminate silos
 - Enhance scholarship
 - Enhance the museum experience
 - Serve new audiences
 - Manage content more effectively
- Stay relevant and visible (discoverable)
 - In my domain AND at web scale
- Avoid one-off, single-use content and applications
- Enhance/extend my institutional relationships
- ...and more!

INTRO TO LINKED DATA

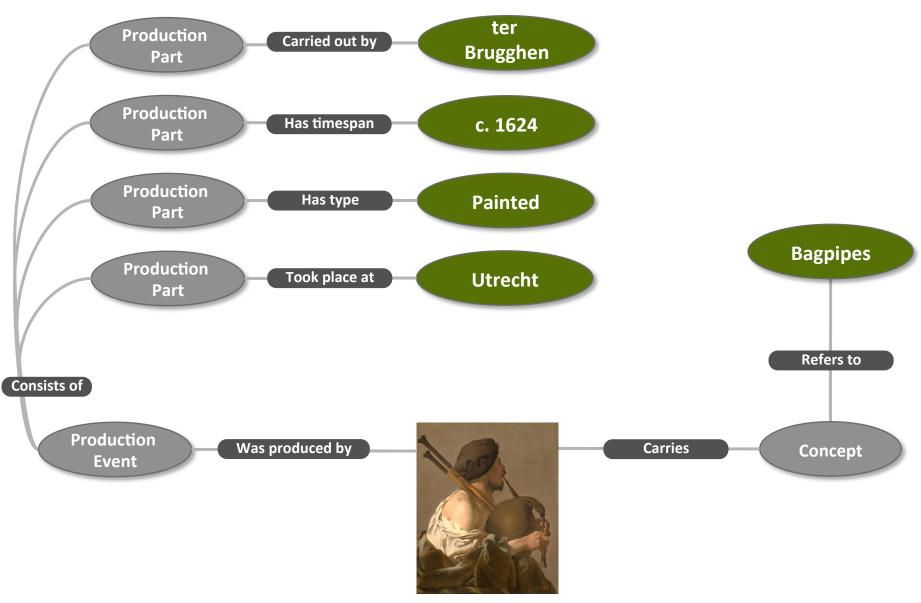
- Why should we care about Linked Data?
- You need a few key ingredients:
 - RDF Triples: Assert facts, create "Knowledge Graphs" using URIs

Semantic Triples: Asserting "facts"



http://en.wikipedia.org/wiki/File:P._Bodart_Portrait_of_Henric_Ter_Brugghen.jpg http://www.music.iastate.edu/antiqua/bagpipe.htm http://www.nga.gov/content/ngaweb/Collection/art-object-page.144298.html

Linking many concepts creates a "Knowledge Graph"



http://www.nga.gov/content/ngaweb/Collection/art-object-page.144298.html

URIs establish the unique ID of an instance of a concept

These "resources" are at the heart of RDF



Hendrick ter Brugghen: https://rkd.nl/en/explore/artists/13338

http://purl.org/nga/collection/constituent/9714

https://www.freebase.com/m/07x8fh



Bagpipe Player: http://purl.org/nga/collection/artobject/144298

https://www.freebase.com/m/04rhvt5



bagpipes: http://en.wikipedia.org/wiki/Bagpipes

https://www.freebase.com/m/0192l

RDF = Resource Description Framework...a flavor of XML

```
<?xml version="1.0"?>
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:cd="http://www.nga.gov/object/144298">
<rdf:Production
 <rdf:has_title>Bagpipe Player</rdf:has_title>
 <rdf:carried out by>Hendrick ter Brugghen</rdf:carried out by>
 <rdf:has timespan>c. 1624</rdf:has timespan>
 <rdf:has type>painted</rdf:has type>
 <rdf:took place at>Utrecht</rdf:took place at>
</rdf:Production>
</rdf:RDF>
```

RDF...the Resource Description Framework

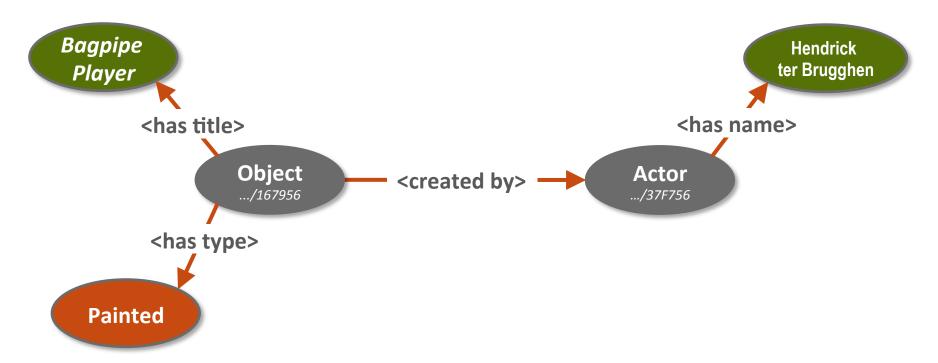
- There are many flavors of RDF syntax
 - RDF-XML
 - Turtle
 - N-Triples
 - JSON-LD
 - RDFa
 - ...etc.
- Many tools and platforms do syntax conversion on demand, so this isn't a concern.
- Choose what works for your internal processing, and know you can change or add if needed!

INTRO TO LINKED DATA

- Why should we care about Linked Data?
- You need a few key ingredients:
 - RDF triples: Assert facts, create "Knowledge Graphs" using URIs
 - Ontologies and Vocabularies to establish a common language

Ontologies

- They codify meaning and structure around a set of concepts
 - Define entities and predicates that connect them
 - Guidance for linking to formulate assertions
 - Guidance for mapping instances of data (names/types) to concepts



Ontologies

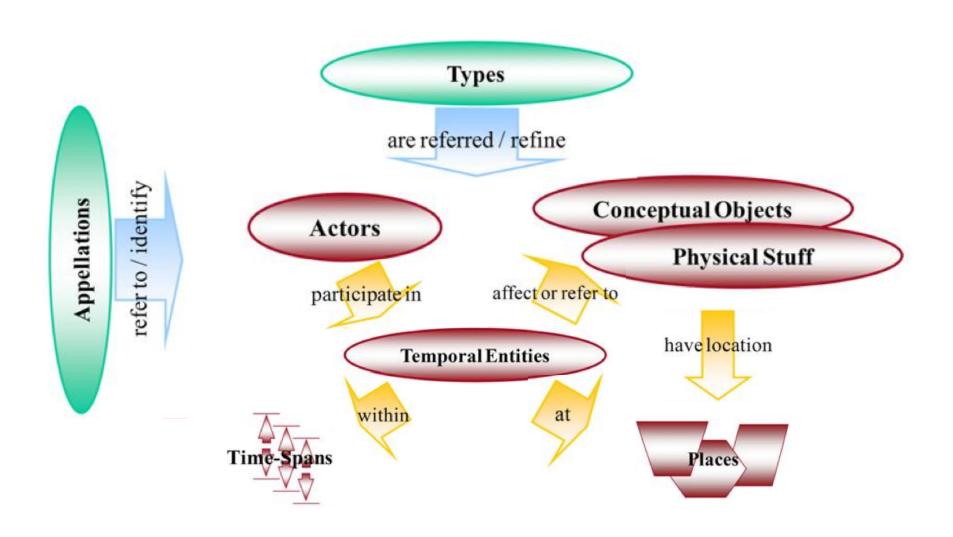
- General-purpose ontologies
 - schema.org
 - Commonly used in the industry: skos, bibo, foaf, geonames, prov, etc.

- Domain-specific ontologies
 - Conceptual Reference Model (CRM) for cultural heritage writ large
 - FRBR and BIBFRAME for libraries
 - A few contenders in development for archives

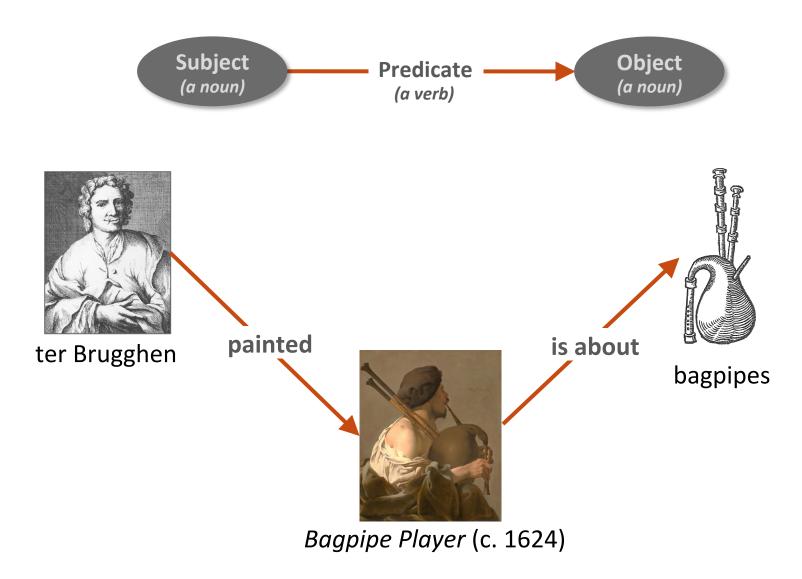
CIDOC-CRM, an ontology for cultural heritage

- Created by the International Council of Museums
- The product of more than 20 years of work
- An ISO standard
- Only 86 classes and 137 properties
- Event-based (both persistent things and temporary things)
- Easily extended
- Proving robust in the face of modeling a broad array of cultural heritage data

Ontologies: CIDOC-CRM – a relatively simple model



Semantic Triples: Asserting "facts"



http://en.wikipedia.org/wiki/File:P._Bodart_Portrait_of_Henric_Ter_Brugghen.jpg http://www.music.iastate.edu/antiqua/bagpipe.htm http://www.nga.gov/content/ngaweb/Collection/art-object-page.144298.html

Exercise: *Predicate Playground*

- Connect the dots!
 - Split into small groups
 - Use the elements on the page to create a graph of triples
 - Use your personal knowledge and/or the web to build your graph. Don't worry about the CRM for now...and name your predicates as you see fit.
 - Brief out your examples. Can you divine a general topic from your elements?
 - Wouldn't it be a more interesting and complete picture if this data were connected into one big graph? With a resource like VIAF you can do it easily!
 - What if a teach could make use of this graph to discover the availability of resources?

Subject: All

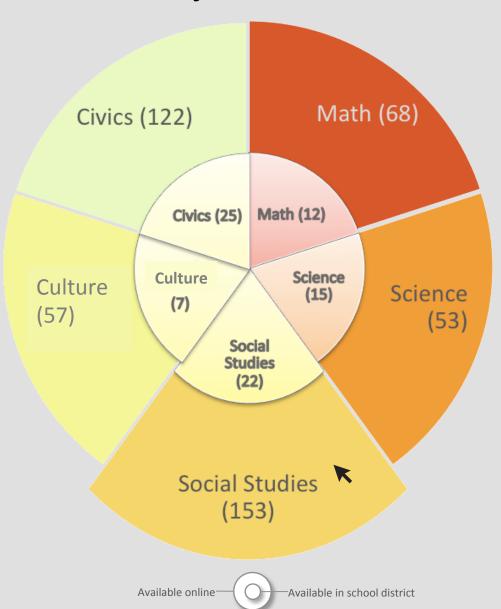
Curricula: Core

•

Grade Level: 6-12

lacktriangleright

Results for Benjamin Franklin 534 results total





Social Studies

Texts

11 Internal 74 External

Videos

2 Internal 12 External

Visuals

8 Internal 59 External

Activity Guides

1 Internal 8 External

See details

My Lesson Plans

Benjamin Franklin: A Great Thinker

Math

Worksheet

Science

- On seasonal time (Hudson)
- Electricity Worksheet

Civics

.

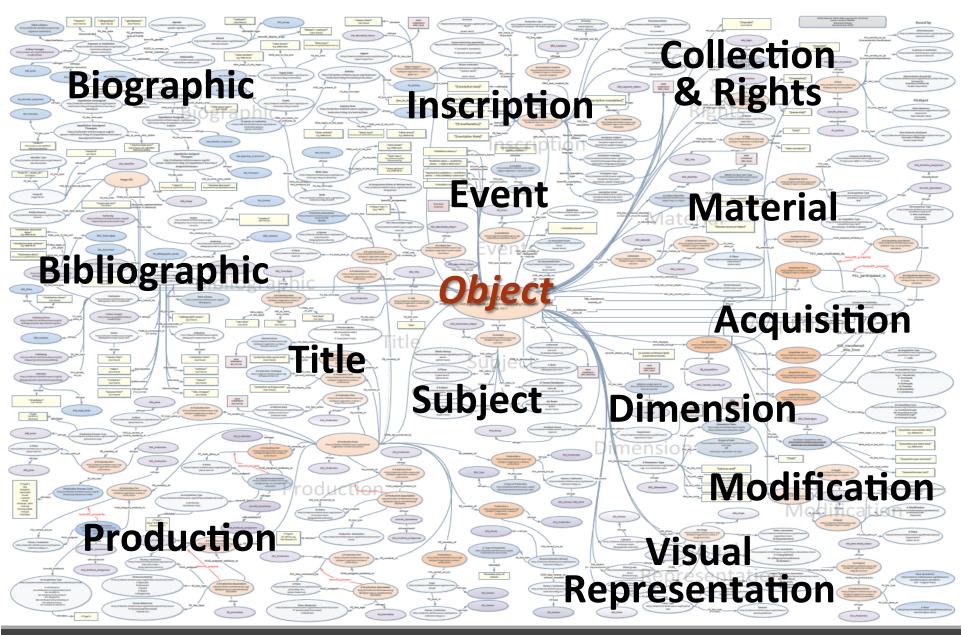
Social Studies

•

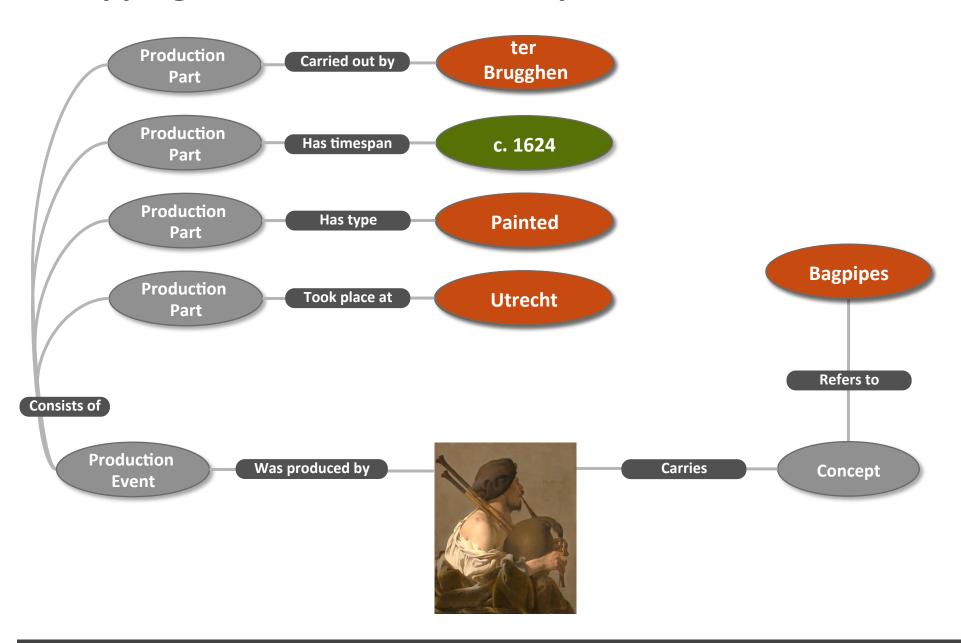
Culture

- Vie de Franklin, écrite par lui-même - Tome I (French)
- Vie de Benjamin Franklin, écrite par luimême - Tome II (French)

British Museum Collection Mapping using CIDOC-CRM



Mapping to the CRM model for "production"



Vocabularies and Instance Data

- 3rd party authorities, e.g.
 - Getty Vocabularies
 - GeoNames
 - Virtual International Authority File (VIAF)
 - DBpedia

- Local authorities
 - Vocabularies unique to your institution (e.g., department names)
 - Your variation on an existing authority (e.g., we prefer a variant name of the artist)

INTRO TO LINKED DATA

- Why should we care about Linked Data?
- You need a few key ingredients: RDF triples, Ontologies and Vocabularies
- Modeling and Mapping

Modeling and Mapping

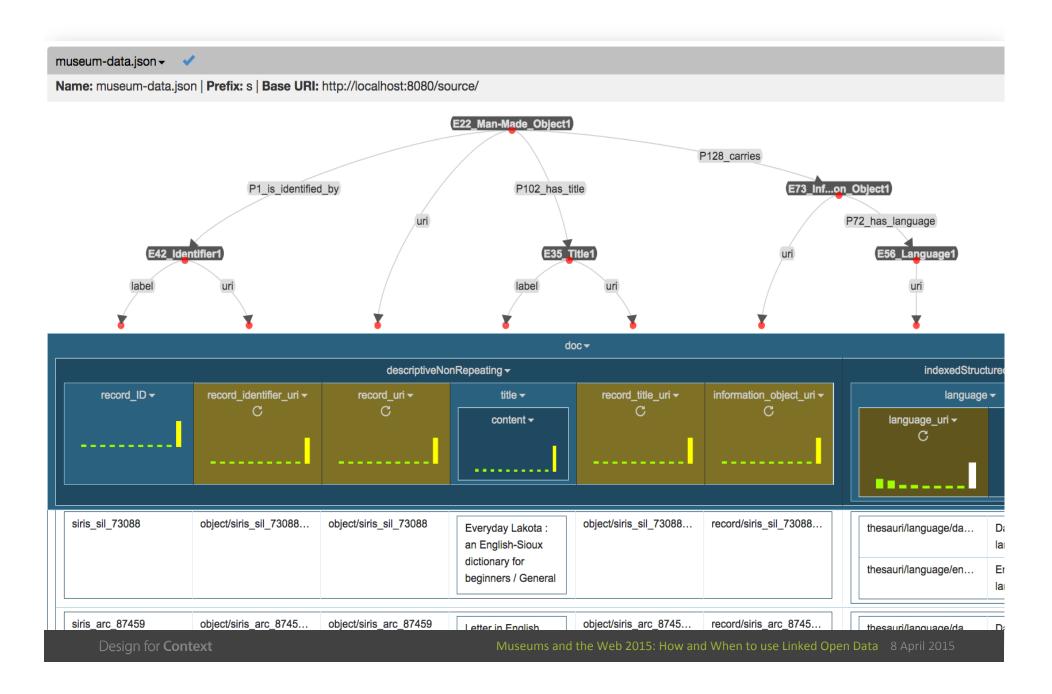
Modeling

- Defining the concepts in our domain of interest
 - What to call them, how to type them
 - Their relationships to one another
- Rules for linking "things" together
- Based on analysis of your existing data and your knowledge of the domain

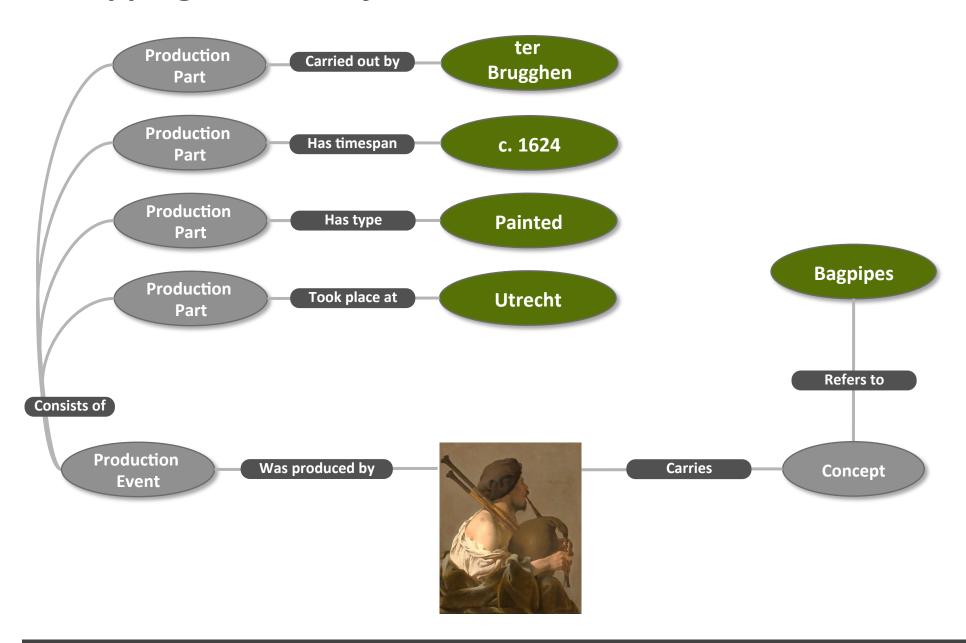
Mapping

 Translating your data to triples using the model, linking rules and vocabularies

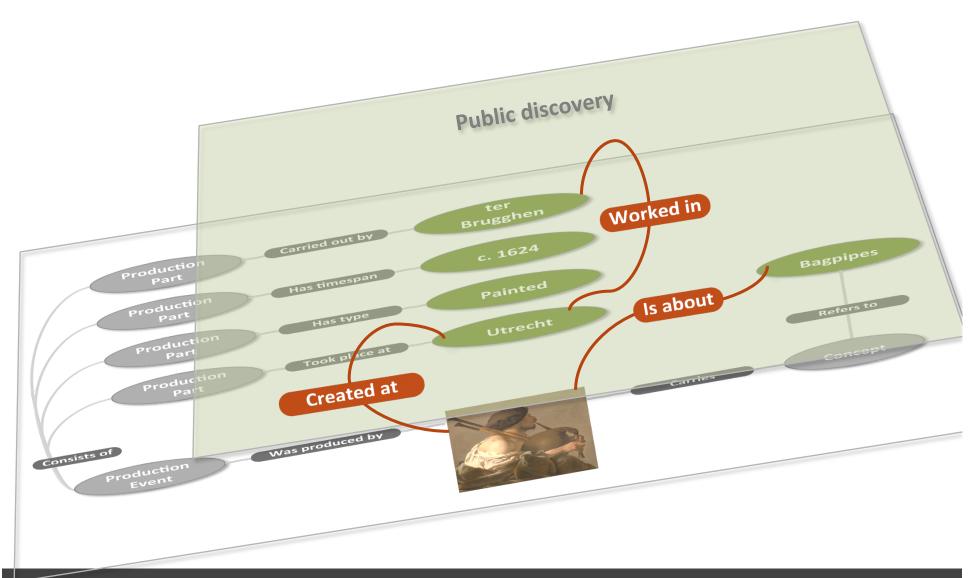
Mapping view in KARMA tool



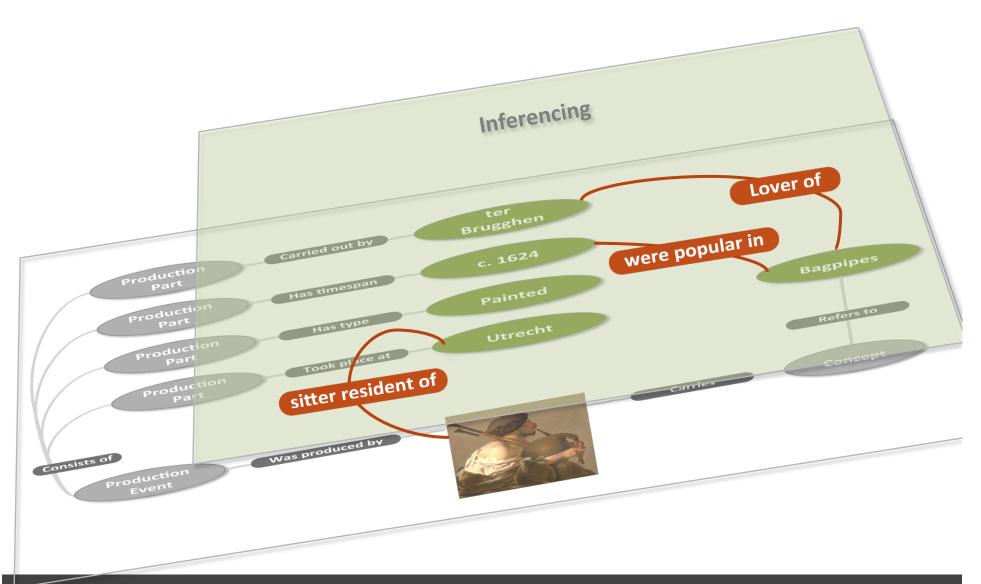
Mapping: Shortcuts for user convenience



Shortcuts and non-destructive contexts

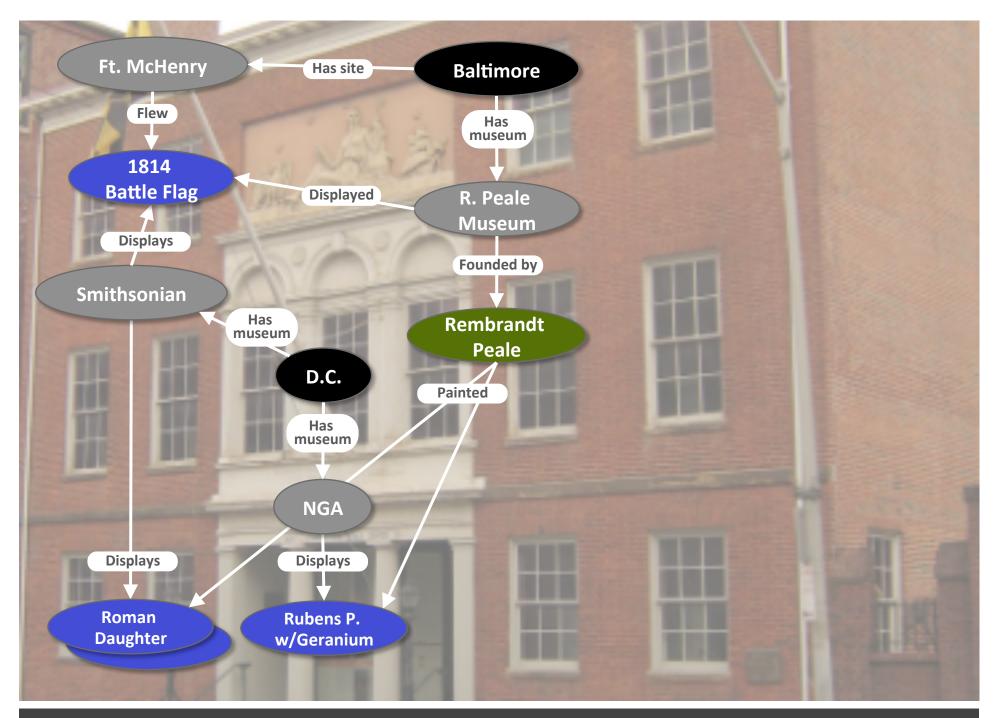


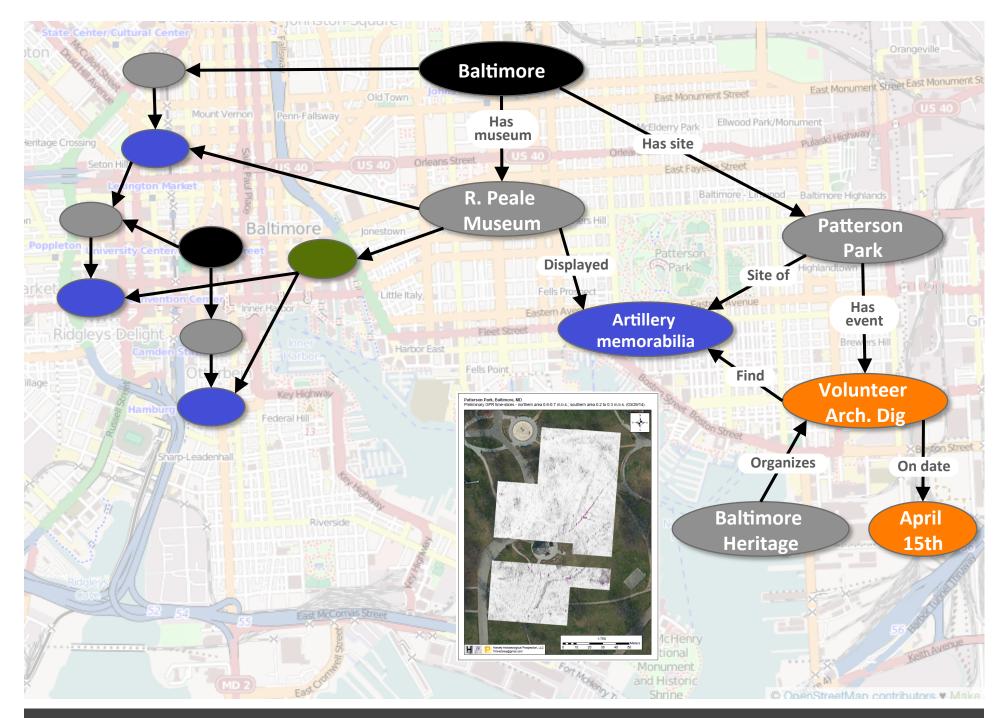
Inferencing: Eliciting facts without linking explicitly



INTRO TO LINKED DATA

- Why should we care about Linked Data?
- You need a few key ingredients: RDF triples, Ontologies and Vocabularies
- Modeling and Mapping
- Enhancing capacity for "serendipitous discovery" by linking between LD sources





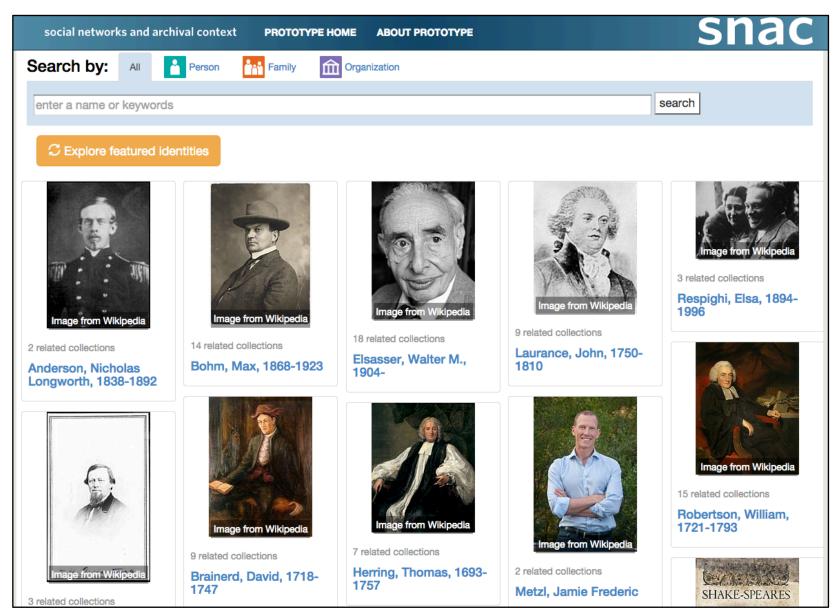
List of potential data sources for Peale example

- Cultural data
 - Cultural object information
 - Historical exhibition information
 - Images/media
- Geographic data
 - Placenames
 - Geolocations
 - Maps
 - Building registry

- Public data
 - Public events
 - Organizational events
- Historical data
 - State and city history
 - Art history
 - American, British, Canadian History
 - Historical biography



Social Networks & Archival Context (SNAC)



http://socialarchive.iath.virginia.edu/snac/search

Social Networks & Archival Context (SNAC)

social networks and archival context

PROTOTYPE HOME

ABOUT PROTOTYPE

snac



Bohm, Max, 1868-1923 Alternative names

Dates:

Authority Source: WorldCat, VIAF, LC

birth 1868 death 1923 Nationality: Language:

Gender:

Biographical notes:



Max Bohm (1868-1923) was a painter in Provincetown, Mass.

Max Bohm was born on January 21, 1868, in Cleveland, Ohio. He began his study of art in 1887 and studied in artist communities in Brittany and in Paris at the Academie Julian with Boulanger, Lefebvre, and Benjamin Constant. He and his family spent the next several decades moving between France, England, and various places in the U.S. Eventually, they settled in Bronxville, N.Y. and

had a cottage in Provincetown, Mass., where Bohm died on September 19, 1923.

From the description of Max Bohm papers, 1873-1970, bulk 1880-1959. (Unknown). WorldCat record id: 430380032

From the description of Max Bohm papers, 1870-1970, bulk 1880-1959. (Unknown). WorldCat record id: 422875086

Painter: Provincetown, Mass.

Born in Cleveland, Ohio.

From the description of Max Bohm papers, 1856-1964, (Unknown), WorldCat record id: 227190065

American painter.

From the description of Receipt, 1898 Aug. 13. (Getty Research Institute). WorldCat record

Links to collections

♂ Archival Collections

14

creator of

referenced in

BOHM, MAX, Artist file: miscellaneous

uncataloged material.

Museum of Modern Art (MOMA)

Bohm, Max. 1868-1923, Artist file.

Brooklyn Museum Libraries & Archives Bohm, Max, 1868-1923. Max Bohm papers, 1856-

Unknown (ISIL:DSI-AAA)

Bohm, Max, 1868-1923. Max Bohm papers, 1870-

1970, bulk 1880-1959. Unknown (ISIL:DSI-AAA)

Bohm, Max. 1868-1923, Max Bohm papers, 1873-

1970. bulk 1880-1959.

Unknown (ISIL:DSI-AAA)

Bohm, Max. 1868-1923, Receipt, 1898 Aug. 13.

Getty Research Institute

Bohm, Max. [Max Bohm]: artist file

John and Mable Ringling Museum of Art Library

Milch Gallery. Milch Gallery records, 1911-1980.

Unknown (ISIL:DSI-AAA)

Milch Gallery. Milch Gallery records, 1911-1980.

Unknown (ISIL:DSI-AAA)

Visualize:

List collection locations

Radial Graph Demo

View source EAC-CPF

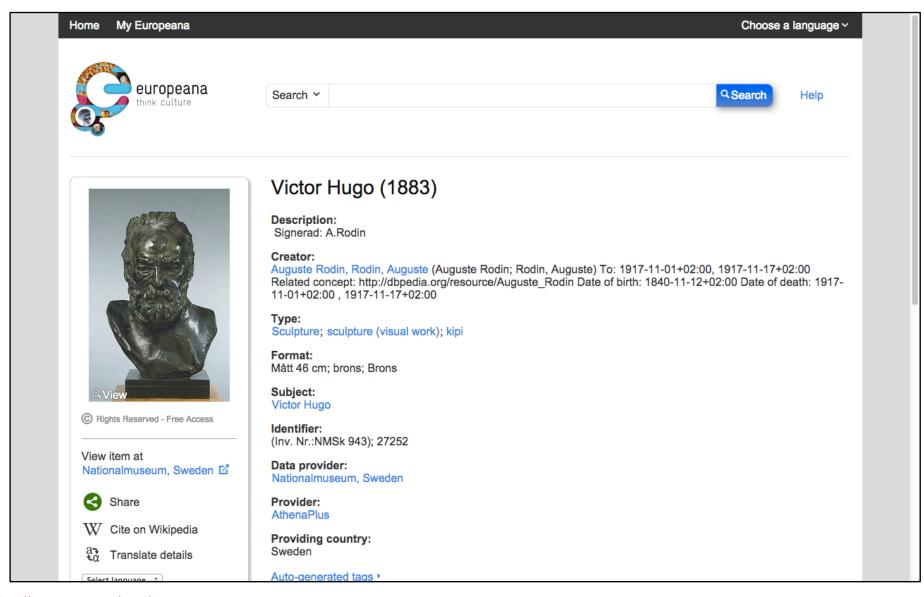
Subjects:

- · Painting, American
- · Painting--Study and teaching

Occupations:

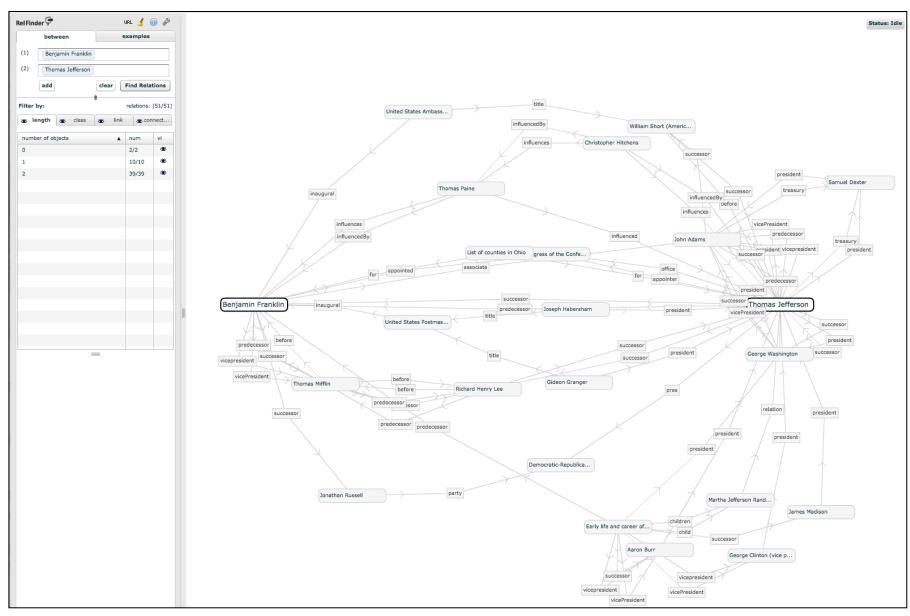
not available for this record

Europeana



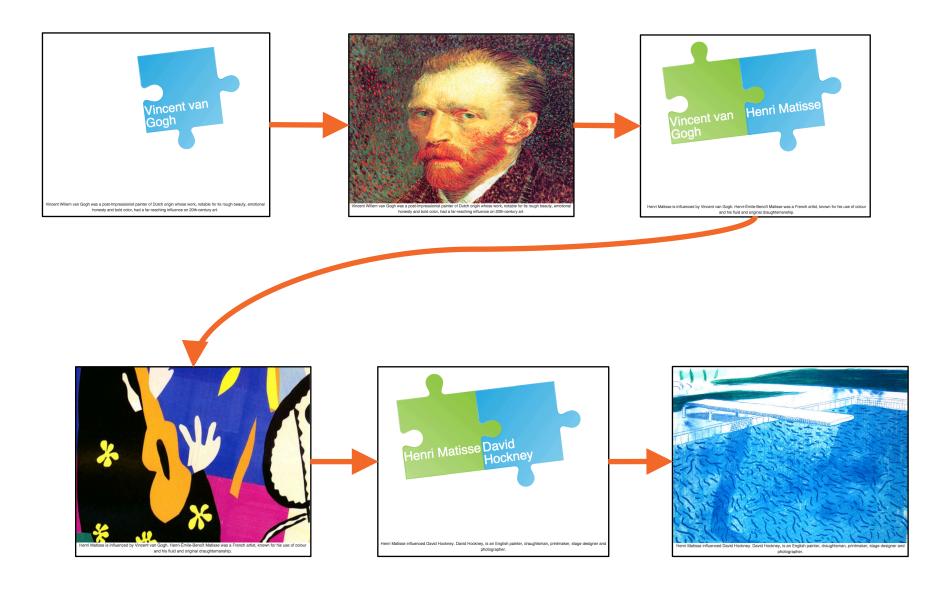
http://www.europeana.eu/portal/

RelFinder



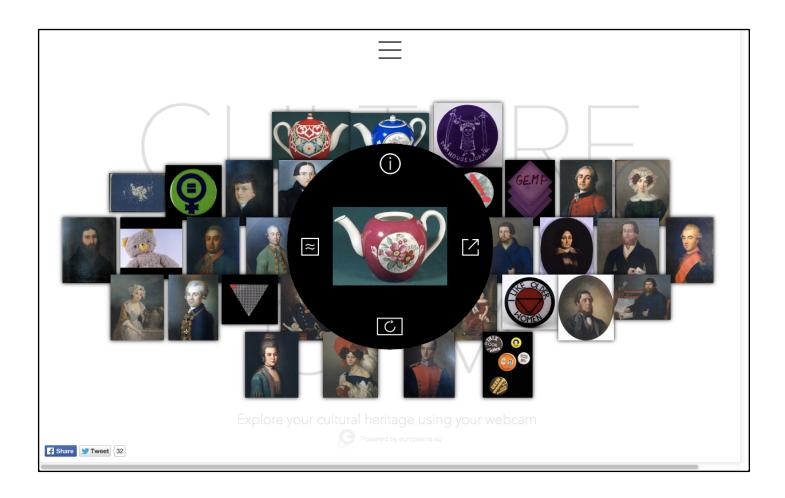
http://www.visualdataweb.org/relfinder/relfinder.php

USC Information Sciences Institute: LOD Stories app



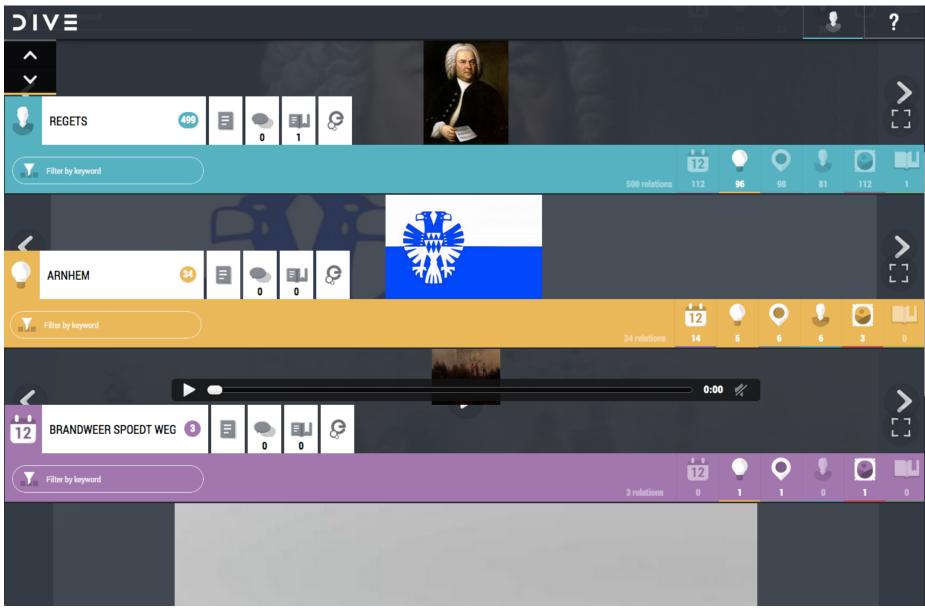
http://lodstories.isi.edu/LODStories/html/linkeddataeduapp.html

Culture Cam



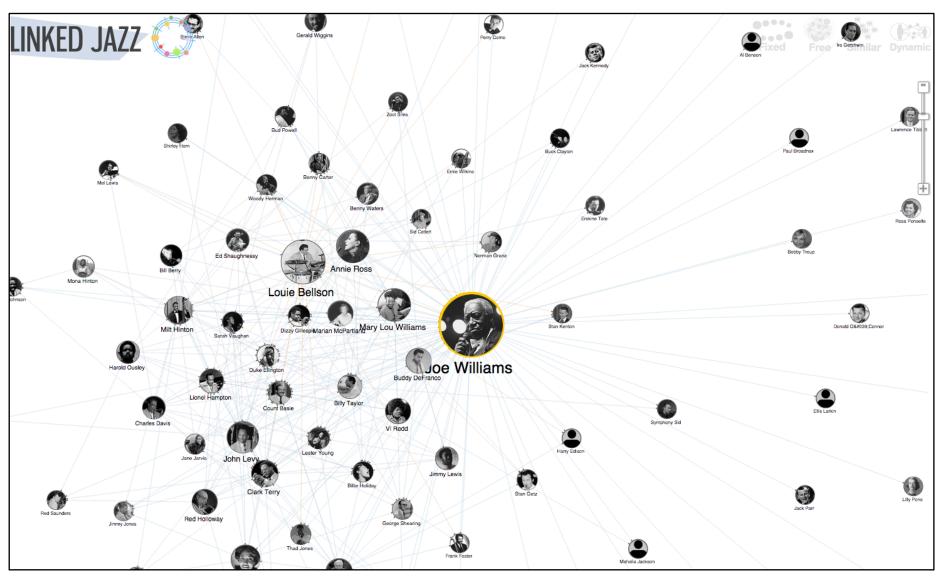
http://culturecam.eu/

DIVE



http://dive.beeldengeluid.nl/

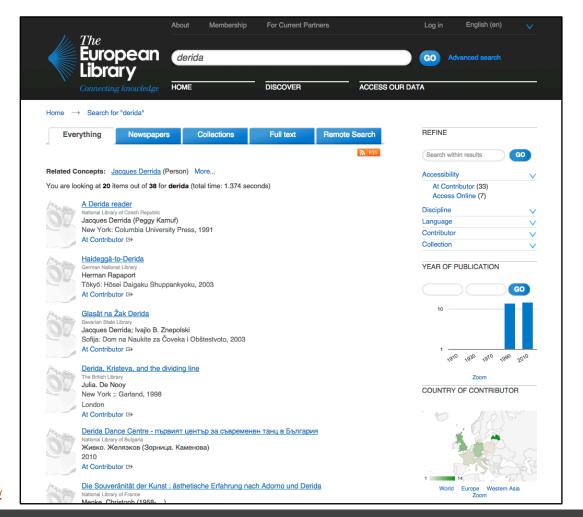
Network Visualization: *Linked Jazz*



https://linkediazz.org/

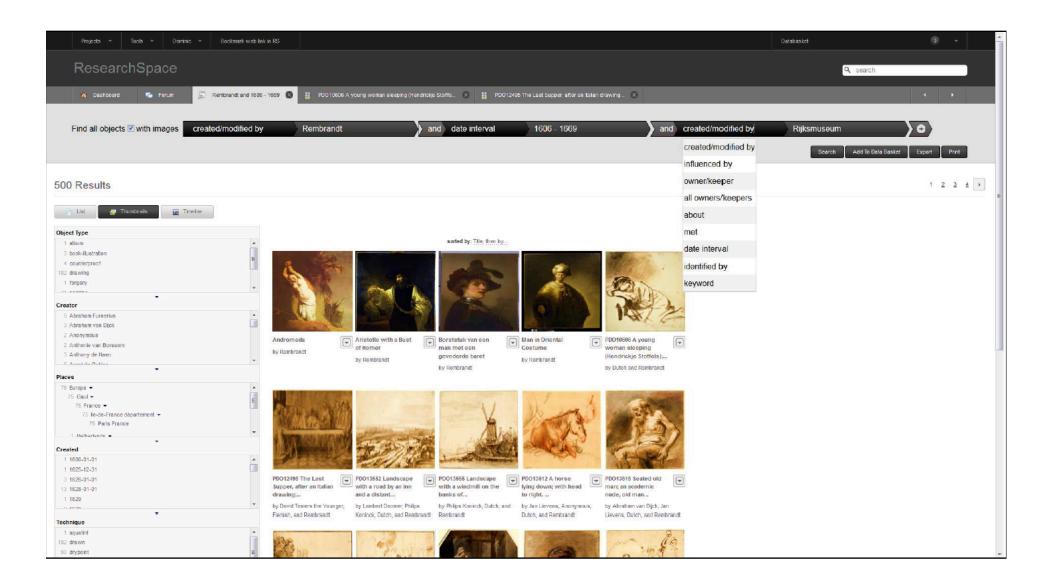
OCLC and The European Library

 LD-driven access to multiple sources, faceted search/browse, and more



https://www.oclc.org http://www.theeuropeanlibrary.org/tel4/

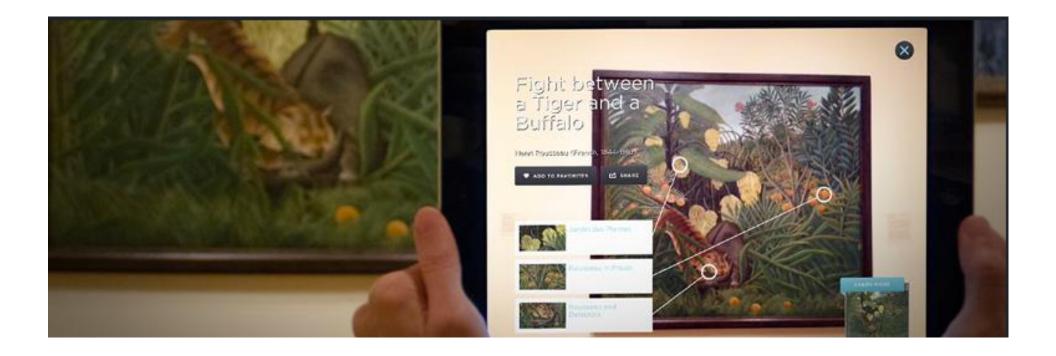
ResearchSpace



http://www.researchspace.org/

Personalization and Federation

Cleveland Museum of Art "Gallery One"



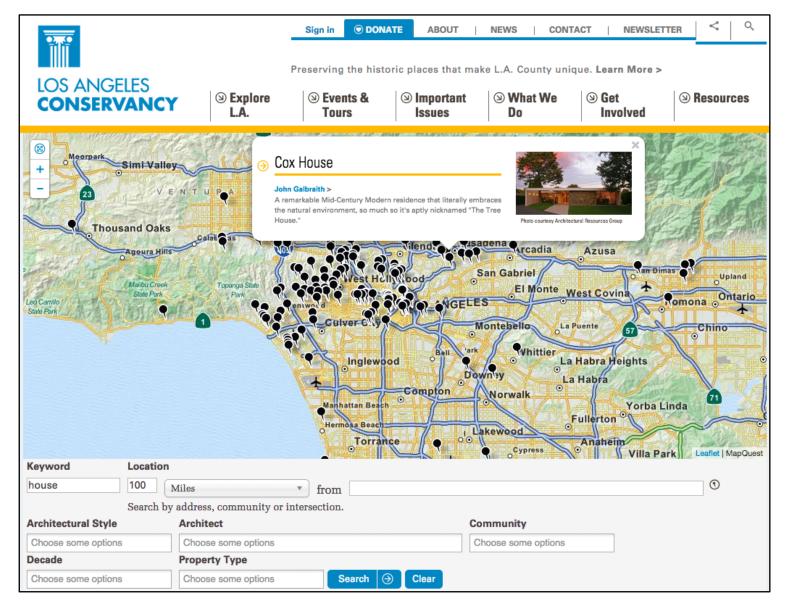
Users take their selections with them through the museum's galleries

...but what if user selections were persistent and available to other applications?





Historic Places LA (Arches Platform)



https://www.laconservancv.org/



VINCENT VAN GOGH

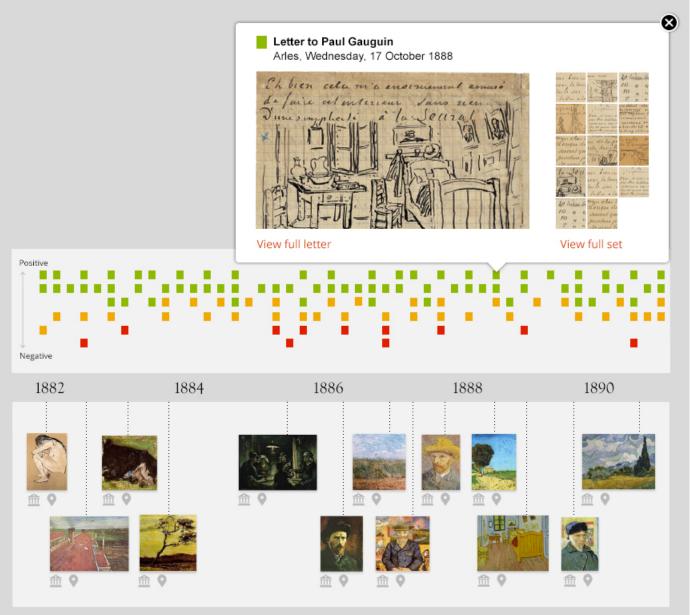
1853-1890

BIRTHPLACE: ZUNDERT, NL

PLACES WORKED: NUENE, ANTWERP, PARIS, ARLES, more...

Letters from the artist provide insight into his mood. A sentiment analysis on those letters identifies the frequency of positive and negative sentiments expressed by the artist over time.









Who uses linked data effectively out in the world?

Big Aggregators

- Google knowledge graph
- Google Culture (emerging)
- FB open graph for embedding
- IBM Watson

Companies

- BBC
- Disney
- NBC SNL
- Best Buy
- Nature

Open Data Repositories

- DBpedia
- Freebase
- Wikidata
- Open Street Map
- Data.gov, data.gov.uk, etc.

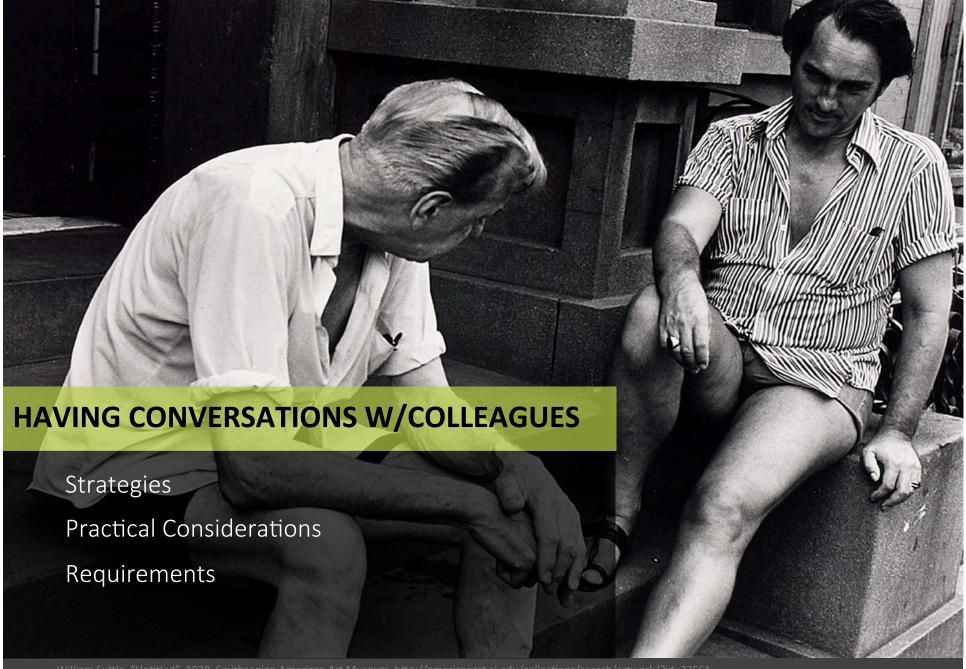
Cultural Heritage

- Europeana
- Various Dutch Institutions
- Yale Center for British Art
- Mellon ConservationSpace & ResearchSpace
- European Libraries
- IIIF
- HistoricPlaces LA (Arches)
- Kultursampo

Vocabularies

- Getty
- Library of Congress Subject Headings (LCSH)
- OCLC
- GeoData

Not a complete list!

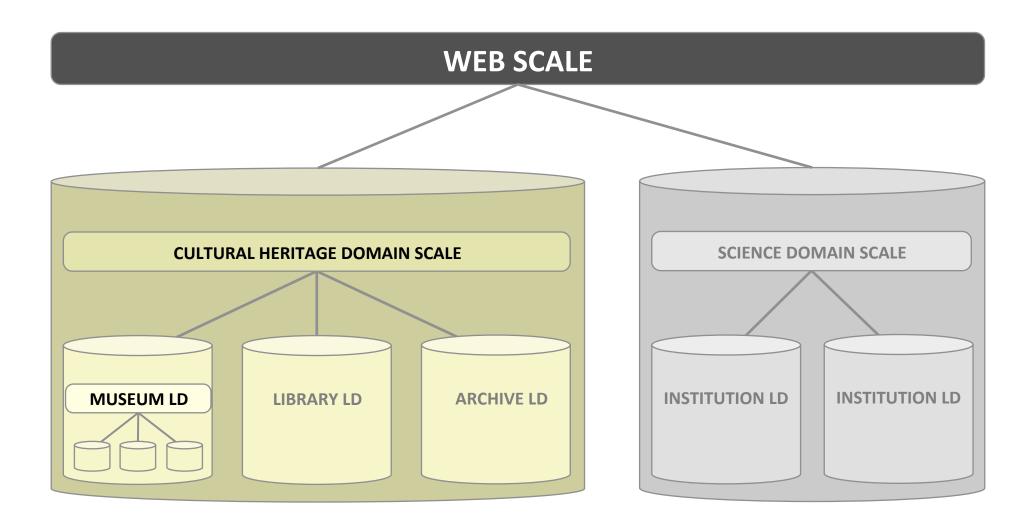


William Suttle, "Untitled", 1938, Smithsonian American Art Museum http://americanart.si.edu/collections/search/artwork/?id=2356

HAVING CONVERSATIONS W/COLLEAGUES

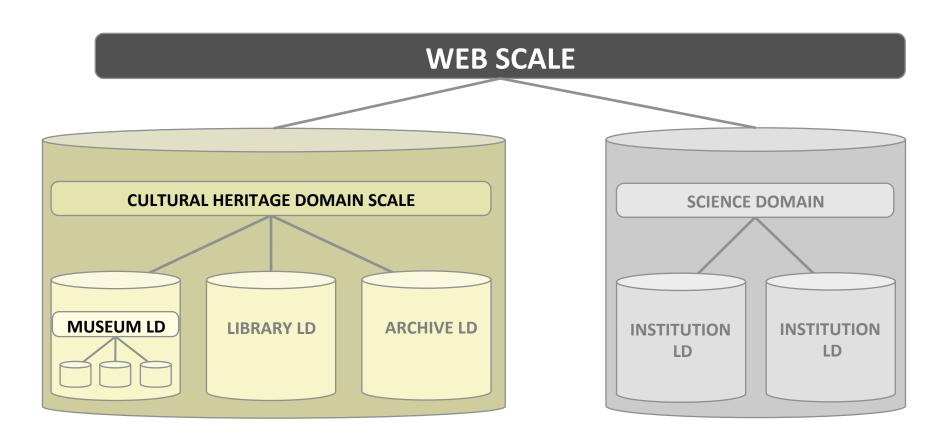
- Institutional strategies
 - Levels of influence and effect

Levels of influence and effect: *Gaining experience AND exposure*



Quick survey...

- What are your aspirations at this moment?
- What level of potential for LD do you see in your institutions in the near term? Long term?



HAVING CONVERSATIONS W/COLLEAGUES

- Institutional strategies
 - Levels of potential influence and effect
 - The Holy Grail: 5-star linked (open) data...and ease of integration

Achieving 5-Star Linked (Open) Data

1-star: make stuff available on the Web (any format) under open license

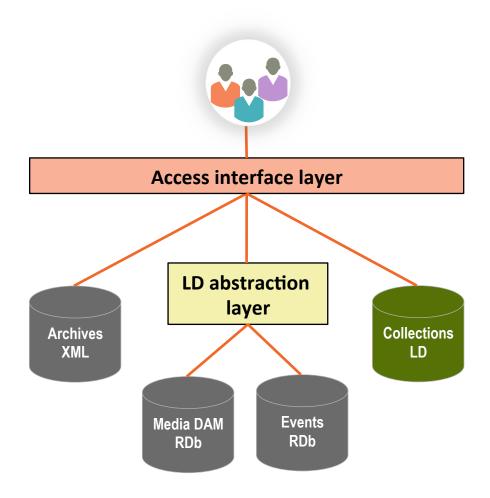
★★ 2-star: make it available as structured data (e.g., Excel)

★★★ 3-star: use non-proprietary formats (e.g., CSV instead of Excel)

Achieving 5-Star Linked (Open) Data

- You can achieve 5 stars incrementally...but keep your eyes on the prize!
- There's some value in achieving only 1–3 stars
- You can target your star goals uniquely for each content type depending on:
 - Complexity (e.g., structured vs. semi-/unstructured)
 - Resource availability
 - Internal
 - 3rd party linked data resources to be linked to/from
 - Program requirements

Linked Data can be *flexibly* adopted and integrated

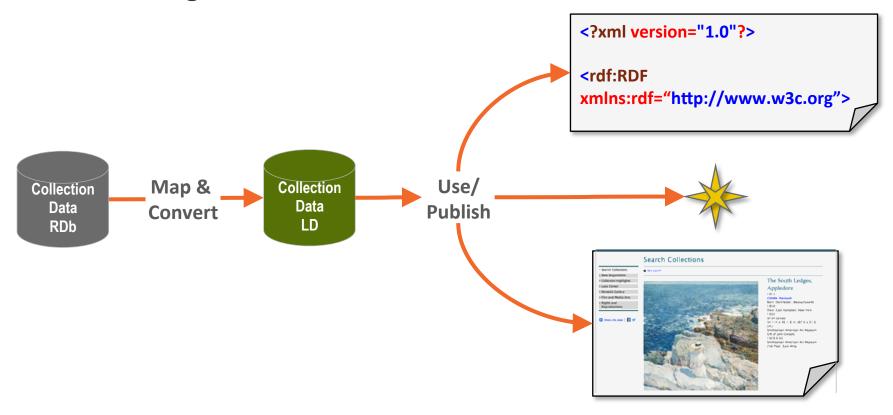


HAVING CONVERSATIONS W/COLLEAGUES

- Institutional strategies
 - Levels of potential influence and effect
 - The Holy Grail: 5-star linked (open) data
 - Inclusion of a range of data sources

Most start with structured collection data/media

- Mapping & converting
- Publishing
- Maintaining



How broad a range of data is useful and feasible?

- Tombstone and broader collections data
- Other semi-structured and unstructured data
 - Provenance
 - Constituents
 - Exhibitions
 - Conservation
 - Research files
 - Video, audio
 - Interviews
 - Manuscripts

What else?

Benefits for sourcing data of all types

- Break down silos
 - Simplified access across data collections
 - Visibility into the sum total of "meaning" you've created
- Enable structural consistency
 - Ease of construction and inspection when creating or updating the data
 - Conform to industry standards
 - Ease of reuse in publishing

- Normalize entities (people, organizations, etc.) and dates
 - Standardize vocabularies across your data contains entities
 - Use dates in calculations, information visualization
- Add depth to your LD holdings
 - More resources you can search across and publish to the world

Supplementing technologies

- Enhancing data
 - Annotation
 - Inferencing and semantic search engines
- Converting unstructured data to structured data
 - Natural language processing and text parsers
 - Artificial intelligence, machine learning and automated analysis
 - Entity and concept extraction to support vocabulary enhancement

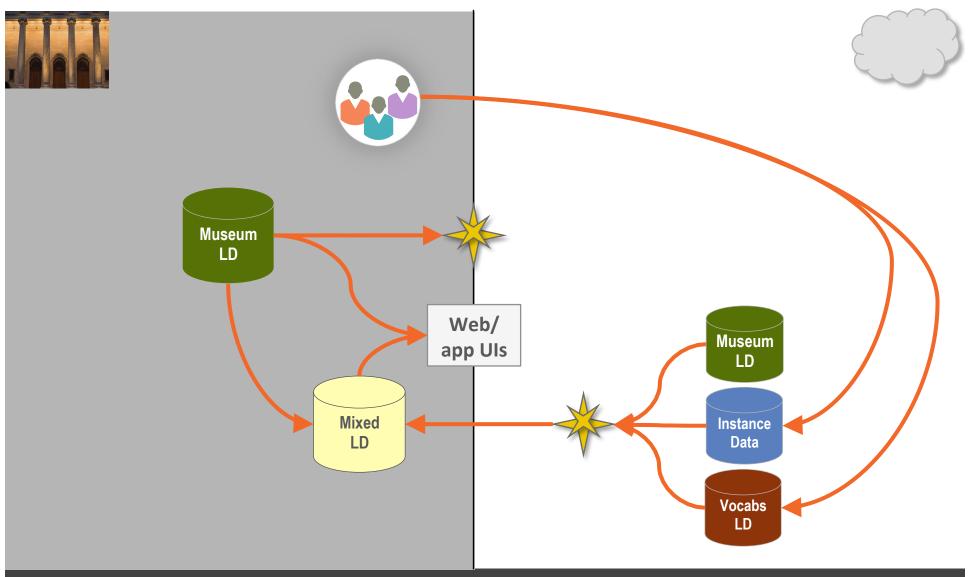
Supplementing technologies

- Managing
 - Content management
 - Content presentation
 - Persistent URL formats and management services
 - Image management and addressing
 - Terminology management and shared vocabularies/libraries
- Visualizing data
 - Information visualization
 - Interface and application frameworks

- Institutional strategies
 - Levels of potential influence and effect
 - The Holy Grail: 5-star linked (open) data
 - Inclusion of a range of data sources
 - Institutional role in LD production, consumption, collaboration

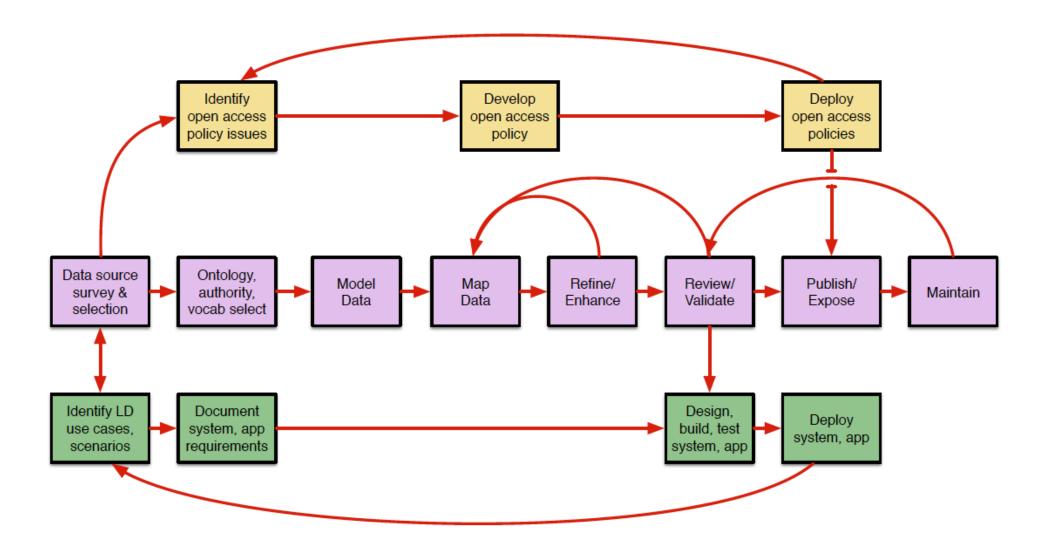
Roles to play in the LD community

Producer - Consumer - Collaborator



- Institutional strategies
 - Levels of potential influence and effect
 - The Holy Grail: 5-star linked (open) data
 - Inclusion of a range of data sources
 - Institutional role in LD production, consumption, collaboration
 - LD Lifecycle: A process, not a project

Linked Data Lifecycle



- Institutional strategies
 - Levels of potential influence and effect
 - The Holy Grail: 5-star linked (open) data
 - Inclusion of a range of data sources
 - Institutional role in LD production, consumption, collaboration
 - LD Lifecycle: A process, not a project
 - LD design and implementation requirements

LD design and implementation requirements

Discovery & Serendipity

 Discover relevant information objects – known and unknown – through search/browse (and foster that serendipitous, "A-ha!" moment)

Extend use

- Take away information objects relevant to my interest
- Conveniently re-use information in my own work products

Persistence

- Conveniently link/reference my work and source information objects
- Trust that information objects to which I link/reference will remain available

Sustainability

Flexibility in design and data modeling to adapt to future capabilities and topics

Scalability

Comfortably adapt to ever-growing collections and different working styles across institutions

- Institutional strategies
 - Levels of potential influence and effect
 - The Holy Grail: 5-star linked (open) data
 - Inclusion of a range of data sources
 - Institutional role in LD production, consumption, collaboration
 - LD Lifecycle: A process, not a project
 - LD design and implementation requirements

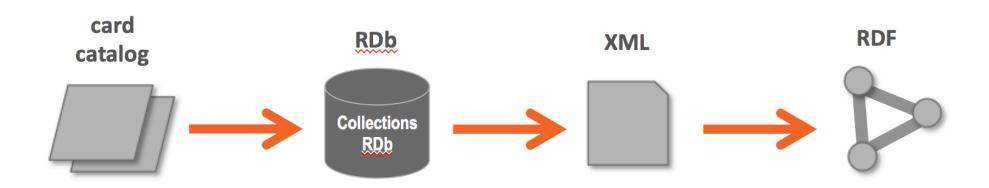
Exercise

- Jot down basic ideas for a programmatic use of linked data in your own institution.
 - Presume no constraints... you have all the data, money, personnel, and time you need
 - Don't over elaborate or get hung up on question of medium/platform
 - Focus on one of: Internal, partnership/public program, blue sky
 - What data would be required?
 - What problems would it solve?
- Get into groups and share your ideas. The group discusses and enriches the idea.



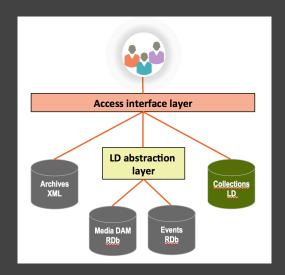
Bartholomeus van der Helst. The Governors of the Kloveniersdoelen. Amsterdam Museum http://www.amsterdammuseum.nl/

- Maturity, flexibility, and a critical mass of participants
 - A next-generation data standard



- Maturity, flexibility, and a critical mass of participants
 - A next-generation data standard
 - An active, world-spanning community

- Maturity, flexibility, and a critical mass of participants
 - A next-generation data standard
 - An active, world-spanning community
 - Flexibility in adoption, integration, change management



- Maturity, flexibility, and a critical mass of participants
- Do more with your content
 - Eliminate silos
 - Avoid one-off single-use content and applications, encourage re-use
 - Manage content more effectively
 - Improve analytical capability

- Maturity, flexibility, and a critical mass of participants
- Do more with your content
- Enhance the museum experience
 - Increase discoverability and serendipity (e.g., enhance scholarship)
 - Push the boundaries on interaction
 - Serve existing audiences better and attract new ones
 - Bridge the physical and digital realms

- Maturity, flexibility, and a critical mass of participants
- Do more with your content
- Enhance the museum experience
- A pioneering opportunity
 - Demonstrate leadership, define a legacy
 - Exercise a spirit of adventure and discovery

- Maturity, flexibility, and a critical mass of participants
- Do more with your content
- Enhance the museum experience
- A pioneering opportunity
- Increase your reach
 - The promise and challenge of "openness"
 - Visibility and authority in your community and at web scale
 - Enhance and extend institutional relationships



Create

Effectively manage and use your data

Extend

Share data seamlessly among partners

Enrich

Integrate with broader humanities and societal data

The value propositions for linked data

Maturity, flexibility, and a critical mass of participants

- A next-generation data standard
- An active, world-spanning community
- Flexibility in adoption, integration, change management

Do more with your content

- Eliminate silos
- Avoid one-off single-use content and applications
- Manage content more effectively
- Improve analytical capability

Enhance the museum experience

- Increase discoverability and serendipity
- Push the boundaries on interaction
- Serve existing audiences better, attract new ones
- Bridge the physical and digital realms

A pioneering opportunity

- Demonstrate leadership, define a legacy
- Exercise a spirit of adventure and discovery

Increase your reach

- The promise and challenge of "openness"
- Visibility and authority in your community and at web scale
- Enhance/extend institutional partnerships



Pierre Puvis de Chavannes, "Work", c 1863, National Gallery of Art http://www.nga.gov/content/ngaweb/Collection/art-object-page.1194.htm

How, When, and Why to use Linked Data

Starting and Sustaining an Institutional Conversation

Museums and the Web: 8 April 2015



Duane Degler

- ✓ duane@designforcontext.com

Neal Johnson

- @vanWinkleTunes
- □ neal@designforcontext.com