

Semantic Enrichment
for
Enhancing Historical and Cultural Heritage Data
to
Support Digital Humanities Research

Marcia L. Zeng

School of Information, Kent State University

HELSINKI CENTRE FOR DIGITAL HUMANITIES /

HELDIG DIGITAL HUMANITIES SUMMIT 2020

RESULTS OF TODAY – VISIONS FOR TOMORROW

Firefox



Outline

Introduction

Data resources
needed for DH

Semantic Enrichment Approaches

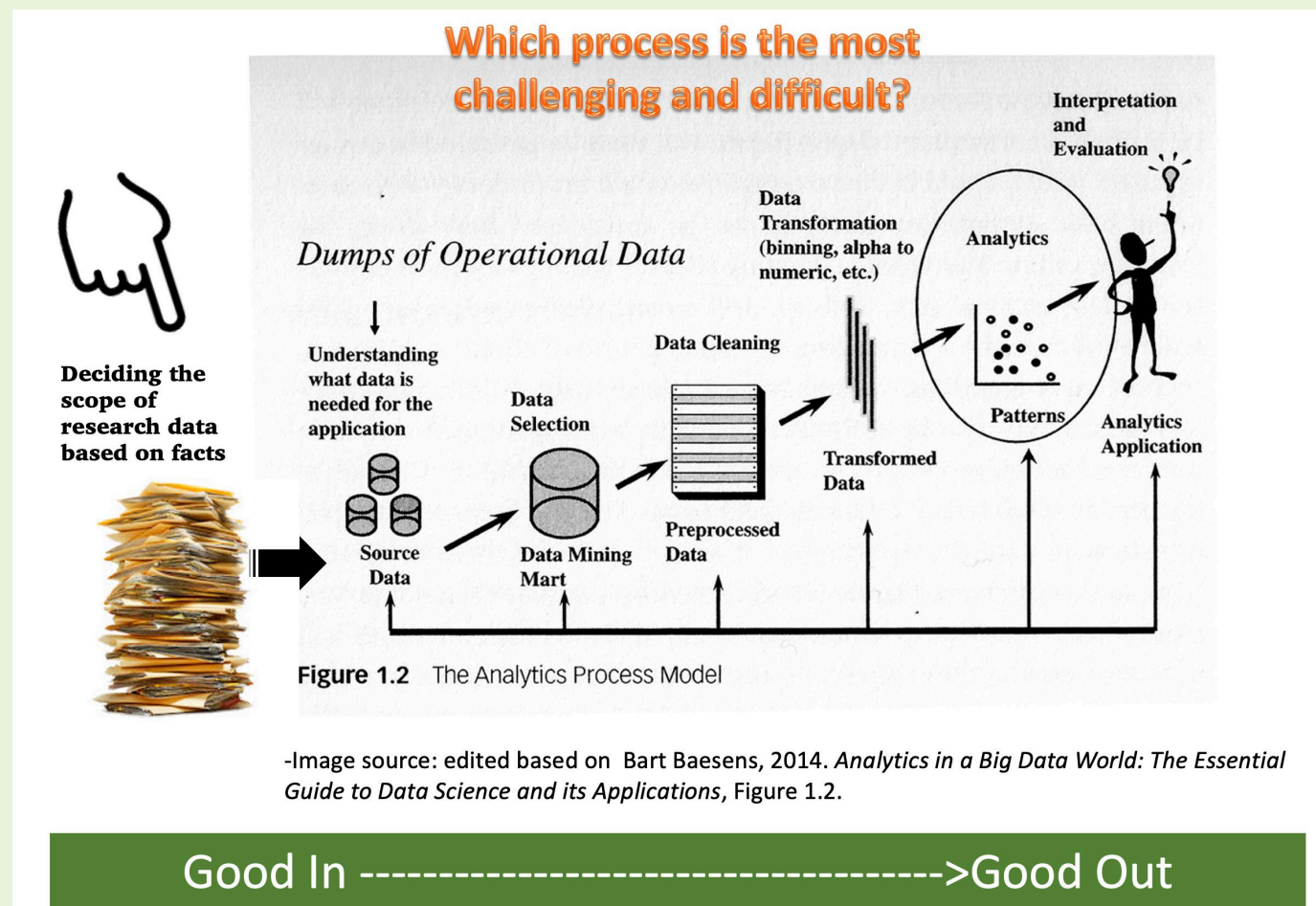
- Structured Data
- Semi-structured Data
- Unstructured Data

Summary and Conclusions



Introduction

- Demands for historical and cultural heritage data in DH research
- The needs of ensuring the FAIRness of historical and cultural heritage data



Source: Zeng, Marcia L. and James Lee. 2017. Smart Data Approaches to Exploring Independent Datasets across Disciplines, Media, and Perspectives for Research in the Humanities. *Digital Humanities 2017*, August 8-11, 2017, Montreal, Canada. [Slide 26.]





The Digging into Data Challenge (DiD) aims to address how "big data" changes the research landscape for the humanities and social sciences.

<https://diggingintodata.org/>

Round / (year)	DiD Funders	DiD Funder countries	Winner #
Round One / (2009)	NEH , NSF , SSHRC , Jisc (4)	US, Canada, UK.	8
Round Two/ (2011)	NEH , NSF , SSHRC , Jisc , IMLS , AHRC , ESRC , NWO (8)	US, Canada, UK. Netherlands	14
Round Three/ (2013)	NEH , NSF , SSHRC , Jisc , IMLS , AHRC , ESRC , NWO CFI , NSERC (10)	US, Canada, UK. Netherlands	14
Round Four / (2016) Renamed as the "T-AP Digging into Data Challenge"	NEH , NSF , SSHRC , Jisc , IMLS , AHRC , ESRC , NWO CFI , NSERC MINCyT , FAPESP , FRQ AKA , ANR , DFG , CONACYT , FCT . (18)	US, Canada, UK. Netherlands, Argentina, Brazil, Finland , France, Germany, Mexico, Portugal	14
Total: 4 rounds	18 funders from 11 countries		50 winners



Domains / Areas of Interests	Resources	Approaches
<ul style="list-style-type: none"> • activities in humanities & social science • ancient language • archaeology • biodiversity • child language development • Colonisation of America • comparative and epidemiological paradigm • criminal intent • debating • early modern common placing • economics • English speech • epidemiology • film and media history • financial system • history • human migration • human rights violations • information networks • information patterns and behaviors • journalism • language evolution • legal structures • linguistics • literary networks • manuscripts provenance • music • musicology • parliaments • policy • population • railroad • social science • sociological theory • standards of living • storytelling traditions and story repertoires • trading and financial markets • vocabularies 	<ul style="list-style-type: none"> • audio (music) recordings • cuneiform tablets (Mesopotamia) • folklore collections • GDP per capita • geographical data • GitHub • journals • Knowledge Graphs • Knowledge Organization Systems • letters • linguistics databases • manuscripts • manuscripts (pre-modern European) • maps • medical images • medieval charters • multilingual classic text • music info • news about terrorism • newspapers • open access publications • papyrus documents • passages • poetry • population databases • proceedings • quotations • records in indigenous style • records in Spanish • signs • social media • speech datasets • speech recordings • speeches • spoken language collections • tweets, political • video data • writing pieces 	<ul style="list-style-type: none"> • annotation • comparative analysis • computational analysis • computing • corpus building • cross datasets analysis • cross-datasets searching • cross-linguistic annotation • data management • data mining • image processing • indexing • linking • machine coding • machine learning • machine translation • metadata aggregation • metadata analysis • metadata auto-generation • metadata extraction • natural language processing (NLP) • protocols development • spatial-temporal correlation • speech mining • text analysis • visualization

Source: Compiled by M. Zeng based on the short descriptions available at <https://dev.diggingintodata.org/awards>



Funders: from more than 10 countries, including NEH, NSF, IMLS

Resources

- audio (music) recordings
- cuneiform tablets (Mesopotamia)
- folklore collections
- GDP per capita
- geographical data
- GitHub
- journals
- Knowledge Graphs
- Knowledge Organization Systems (KOS)
- letters
- linguistics databases
- manuscripts
- manuscripts (pre-modern European)
- maps
- medical images
- medieval charters
- multilingual classic text
- music info
- news about terrorism

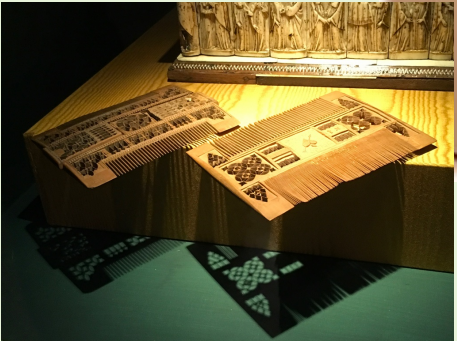
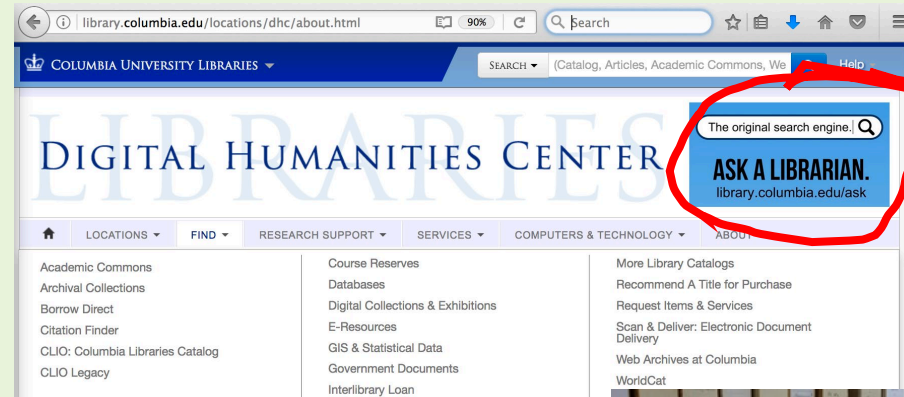
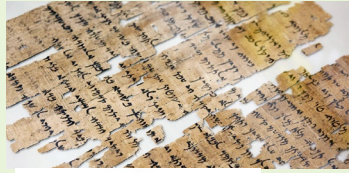
- newspapers
- open access publications
- papyrus documents
- passages
- poetry
- population databases
- proceedings
- quotations
- records in indigenous style
- records in Spanish
- signs
- social media
- speech datasets
- speech recordings
- speeches
- spoken language collections
- tweets, political
- video data
- vocabularies
- writing pieces

Domains / Areas of Interest	Resources	Approaches
<ul style="list-style-type: none"> • activities in humanities & social science • ancient language • archaeology • biodiversity • child language development • Colonization of America • comparative and epistemological paradigm • criminal intent • debating • early modern common phrasing • economics • English speech • epidemiology • film and media history • financial system • history • human migration • human rights violations • information networks • information patterns and behaviors • journalism • language evolution • legal structures • linguistics • literary networks • manuscript provenance • music • musicology • parliaments • poetry • population • regional • social science • sociological theory • standards of living • storytelling traditions and story repertoires • trading and financial markets • vocabularies 	<ul style="list-style-type: none"> • audio (music) recordings • cuneiform tablets (Mesopotamia) • folklore collections • GDP per capita • geographical data • GitHub • journals • Knowledge Graphs • Knowledge Organization Systems • letters • linguistics databases • manuscripts • manuscripts (pre-modern European) • maps • medical images • medieval charters • multilingual classic text • music info • news about terrorism • newspapers • open access publications • papyrus documents • passages • poetry • population databases • proceedings • quotations • records in indigenous style • records in Spanish • signs • social media • speech datasets • speech recordings • speeches • spoken language collections • tweets, political • video data • writing pieces 	<ul style="list-style-type: none"> • annotation • comparative analysis • computational analysis • computing • corpus building • cross datasets analysis • cross-datasets searching • cross-linguistic annotation • data management • data mining • image processing • indexing • linking • machine coding • machine learning • machine translation • metadata aggregation • metadata analysis • metadata auto-generation • metadata extraction • natural language processing (NLP) • network development • special temporal correlation • speech mining • text analysis • visualization

Domains/Areas of Interests || Resources || Approaches
Expressed in the Project Descriptions of *Digging into Data Challenge* Round 1-4, 2009-2016



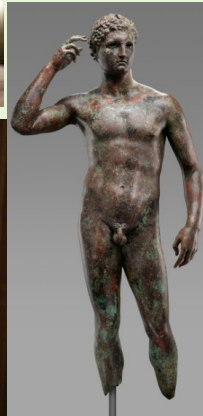
Data provided by LAMs and cultural heritage institutions are treasures for all humanities researchers.



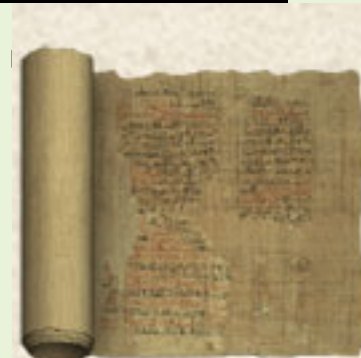
Music Treasures Consortium
Library of Congress



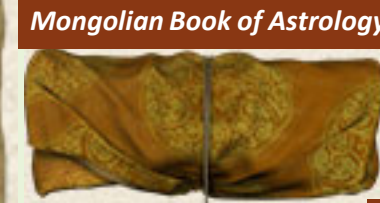
20th Century Press Archives, ZBW



Marcia L. Zeng, HELDIG Dr



Edwin Smith Surgical Papyrus



Mongolian Book of Astrology



The Marshall Nirenberg Charts: The "First Summary"



National Library of Medicine

Digitizing & Documenting → Datafying & Enriching → Contextualizing

Through LAMs: *unstructured data* found in documents and other information-bearing objects:

Transforming unstructured data into → structured data

Resources delivered on the web, including:

- Metadata
- Representative images
- Original documents' transcripts
- Media, etc.



National Library of Medicine "Turning the Pages" The Edwin Smith Surgical Papyrus

Case 5. A head wound with skull fracture (2,11 - 17)

Title
Practices for a gaping wound in his head that has fractured his skull.

Examination and Prognosis
If you treat a man for a gaping wound in his head, which has penetrated to the bone and split his skull, you have to probe that wound. Should you find that fracture that is in his skull deep and sunken under your fingers, and should the swelling that is on it be high, while he bleeds from his nostrils and his ears, suffers stiffness in his neck, and is unable to look at his shoulders and his chest, then you say about him: "One who has a gaping wound in his head, which has penetrated to the bone and fractured his skull, and who suffers from stiffness in his neck: an ailment for which nothing is done."

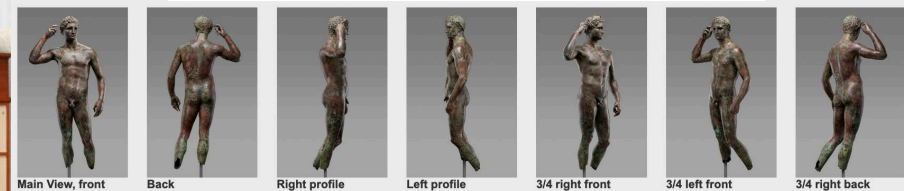
Treatment
You should not bandage him. He is to be put down on his bed until the time of his injury passes.

Explanations
As for "which has fractured his skull," it is fracturing his skull, with the bones that happen from that fracture sunken to the inside of his skull. The treatise "The Nature of Wounds" has said about it: it is the fracturing of his skull into many pieces, sunken to the inside of his skull.

Close **TEXT** **ZOOM**

<https://ceb.nlm.nih.gov/proj/ttp/flash/smith/smith.html>

J. Paul Getty Museum collection



Provenance Exhibitions Bibliography Education Resources Related Media

02:35

Audio: Statue of a Victorious Youth / Carol Mattusch (Highlights)

01:13

Audio: Statue of a Victorious Youth - Conservation / J. Podany

01:44

Audio: Victorious Youth (Descriptions)

01:39

Audio: Statue of a Victorious Youth, feat. Carole Mattusch

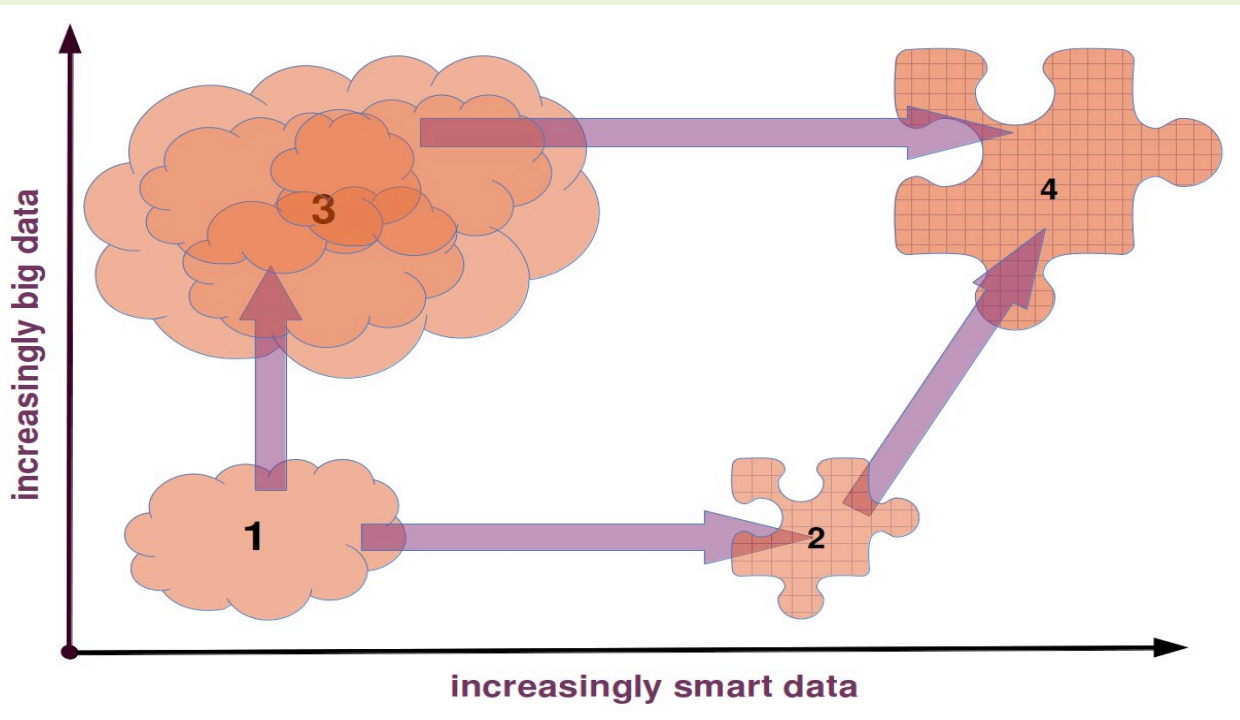
<https://www.getty.edu/art/collection/objects/7792>



“Big? Smart? Clean? Messy? Data in the humanities”

Schöch, Christof. *Journal for Digital Humanities*. 2(3): 2-13.

Data has to be cleaned, transformed, and analyzed to unlock its hidden potential.



The story of smart and big data.

<http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/>

Marcia L. Zeng. HELDIG DH Summit 2020

Once tamed through organizing and integrating processes, large volumes of unstructured, semi-structured, and structured data are turned into “smart data” that reflect the research priorities of a particular discipline or field.

Smart data inquiries can then be used to provide comprehensive analyses and generate new products and services.



Smart Data in the context of Big Data

Big Data

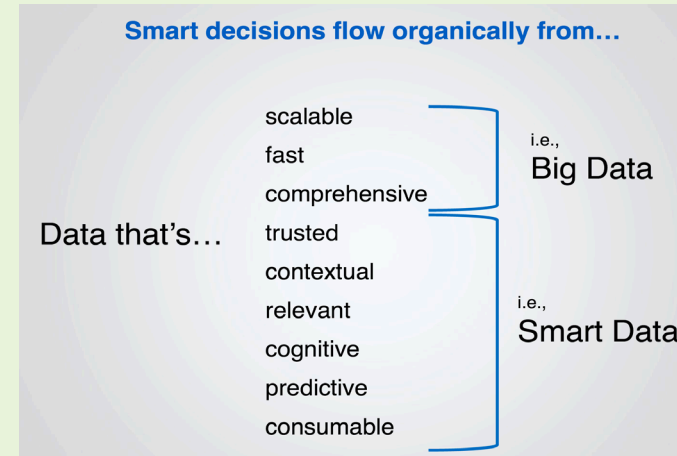
- Volume (data quantity)
- Velocity (data speed)
- Variety (data types & nature)
- Variability (data consistency)
- Veracity (data quality)



Smart Data

= The ability to achieve big insights from such data at any scale, great or small.

trusted
contextualized
relevant
cognitive
predictive
consumable



Kobielus, James. (2016)

Source: Zeng, M.L. 2017 [DOI: 10.1515/jdis-2017-0001](https://doi.org/10.1515/jdis-2017-0001)
Compiled based on Kobielus, James. (2016, June). *The Evolution of Big Data to Smart Data*.
Keynote at Smart Data Online 2016.



Outline

Introduction



Semantic Enrichment Approaches

1. Structured Data
2. Semi-structured Data
3. Unstructured Data

Summary and Conclusions





LAM Data Examples

1

Structured

- bibliographies
- indexing & abstracting databases
- citation indexes
- catalogs of all kinds
- special collection portals
- metadata registries
- curated research datasets
- name authorities

Semi-structured

-
- Text Encoding Initiative (TEI) files
- archival finding aids
- value added/tagged resources
- unstructured portion within metadata descriptions
- data from Web crawling that need to be cleaned
-

Unstructured

- documents, cultural artifacts, & original information-bearing objects
 - digitized or not-digitized
 - textual or non-textual
 - in all kinds of formats and media
 - Possibly of undetermined date and/or origin





1. Semantic Enrichment for **Structured Data**

Semantic Enrichment for **Structured Data**

➤ A common strategy in LAM data enhancement efforts in order to:

- Overcome challenges relating to data quality and discoverability in the digital age
- Provide more context and multilingual information for cultural heritage (CH) objects

[Refer to the most recent presentations at [SWIB20](#)]

➤ “Enrichment” can be used to refer to

- a process (e.g., the application of an enrichment tool);
or
- its result (the new metadata created at the end of the process).

- ✓ reconciliation,
- ✓ mapping,
- ✓ alignment,
- ✓ matching,
- ✓ massaging,
- ✓ merging,
- ✓ Interlinking
- ✓ ...

Three main stages

- 1) Analysis
- 2) Linking
- 3) Augmentation

Ref: -- Europeana Task Force on Enrichment and Evaluation. “Report on Enrichment and Evaluation” 29/10/2015



1. Semantic Enrichment for **Structured Data**

[Starting point: existing metadata components that are in a controlled/contrrollable form]

Outline

Approaches

A. Contextualize through entities

Creates typed relationships between resources of different types

B. “Massage”

(my) label → to → (their) URI(s)

(my) ID → to → (their) URI(s)

C. Connect to real “things”

Used by (examples)

- **Unified repositories**
 - Europeana
- **Collaborative projects**
(Suggest to visit):
 - Wikibase “[Project Passage](#)”
 - LD4P (Linked Data for Libraries)
 - Linked Data for Production: Pathway to Implementation ([LD4P2](#))
 - “Knowledge panels”
- **Individual LAMs and institutions**
 - Museum of Modern Art (MoMA)
 - Dictionary of Classic Mayan



1-A. Contextualize through entities

- ✓ Creates typed relationships between resources of different types
 - usually on those fields that are in a controlled form

[Europeana enriches xxx by aligning to (xxx)]

agent names →
places →
concepts →
time period → (Semium Time).

- ✓ Relate Objects to concepts, agents, places, etc., using the properties in EDM (e.g., *dc:subject*, *dc:creator*).

Update from Europeana

by Nov 15, 2020, enriched in this year:

- 1,429,242 for agents
- 15,206,861 for places
- 15,218,899 for concepts
- 18,597,882 for time period

Europeana Dereferenceable vocabularies

The Getty - Union List of Artist Names (ULAN)	edm:Agent	} agents
Virtual International Authority File (VIAF)	edm:Agent	
Wikidata	edm:Agent	
Gemeinsame Normdatei (GND)	edm:Agent, edm:Place, skos:Concept	} places
Getty Thesaurus of Geographic Names (TGN)	edm:Place	
Geonames	edm:Place	} concepts
The Getty - Art & Architecture Thesaurus (AAT)	skos:Concept	
IconClass	skos:Concept	
Israel Museum Jerusalem Concepts	skos:Concept	
Library of Congress Subject Headings (LCSH)	skos:Concept	
data.europeana.eu WWI Concepts from Library of Congress Subject Headings (LCSH)	skos:Concept	
Europeana Sounds Genres	skos:Concept	
UDC	skos:Concept	
UNESCO Thesaurus	skos:Concept	
YSO - General Finnish ontology	skos:Concept	
Fashion Thesaurus	skos:Concept	
MIMO Concepts	skos:Concept	

- Source: Europeana Semantic Enrichment Framework *Documentation*

Version: 17th November 2016 (updated 2017, 2018, 2020)

Available from <https://pro.europeana.eu/page/europeana-semantic-enrichment> --> [several vocabularies](#) (Compiled by MZ 2020-11-18)



1-B. "Massage"

An Example:

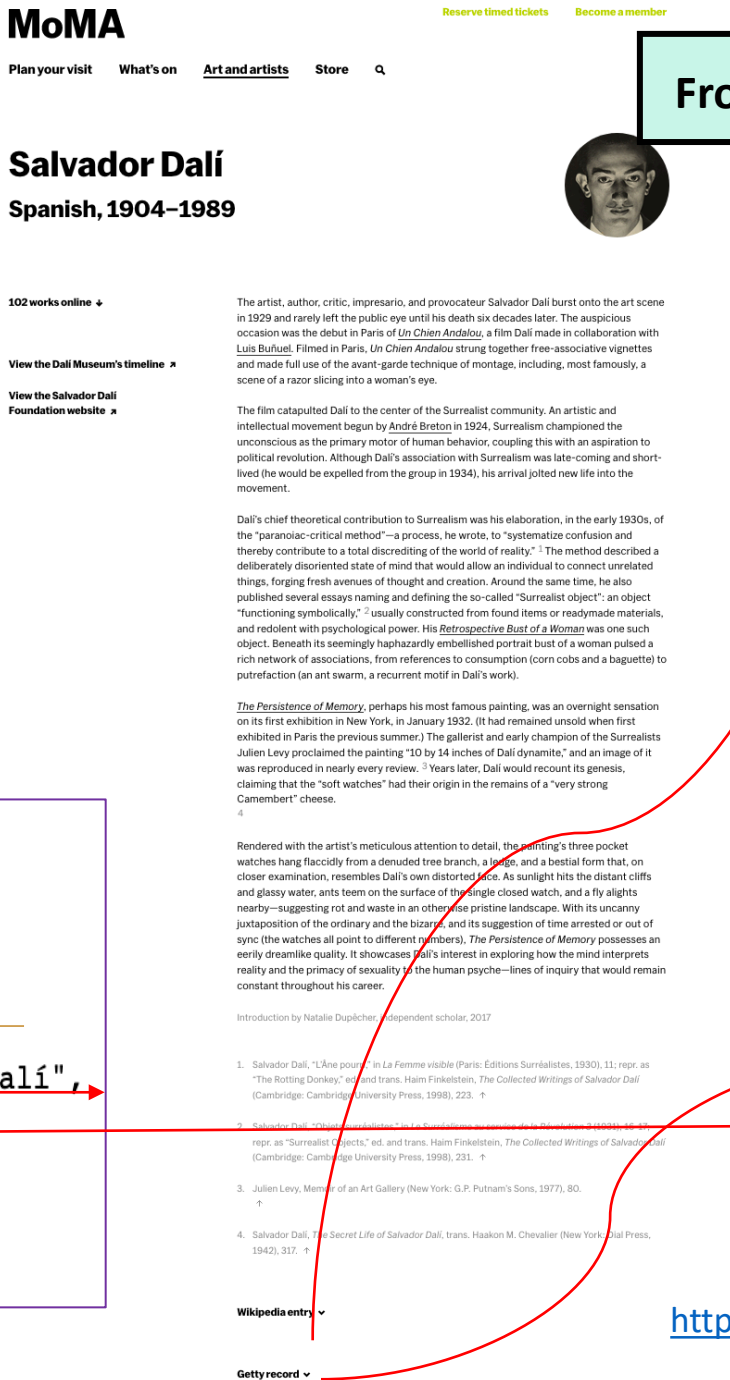
- (my) ID →to→ (their) URI(s)

(my) ID

"sameAs"
(their) URIs

```
37 <script type="application/ld+json">
38 {
39   "@context": "http://schema.org/",
40   "@type": "Person",
41   "url": "https://www.moma.org/artists/1364",
42   "sameAs": [
43     "https://en.wikipedia.org/wiki/Salvador_Dalí",
44     "http://vocab.getty.edu/ulan/500009365"
45   ],
46   "name": "Salvador Dalí"
47 }
48 </script>
49
50
```

Back-end



The screenshot shows the MoMA website for Salvador Dalí. At the top, there are links for "Plan your visit", "What's on", "Art and artists", "Store", and a search icon. The main heading is "Salvador Dalí" with the subtitle "Spanish, 1904–1989" and a portrait of Dalí. Below this, there are sections for "102 works online", "View the Dalí Museum's timeline", and "View the Salvador Dalí Foundation website". The main content area contains several paragraphs of text about Dalí's life and work, including his debut in Paris, his association with Surrealism, and his chief theoretical contribution. At the bottom, there is a "Wikipedia entry" and a "Getty record" section.

Front-end

Case: MoMA

Wikipedia entry ▾

Tool: OpenRefine

Getty record ^

Union List of Artist Names (ULAN)

Nationalities

Spanish, Catalan

Gender

Male

Roles

Artist, Writer, Illustrator, Painter, Sculptor

Names

Salvador Dalí, Salvador Dalí, Salvador Felipe Jacinto Dalí, Salvador Felip Jacint Dalí Domènech, Salvador Dalí y Domènech, Salvador Felipe Jacinto Dalí y Domènech, Salvador Dalí i Domènech, Salvador Felipe Jacinto Dalí Domènech, Salvator Dalí, Salvador Dalí Domènech, Salvador Dalm y Domenech, Салвадор Дали, Сальвадор Дали, 萨尔瓦多·达利, 达利萨尔瓦多

Ulan

500009365

View the full Getty record ↗

Information from Getty's Union List of Artist Names® (ULAN), made available under the [ODC Attribution License](#)

Images captured 2020-11-18

<https://www.moma.org/artists/1364>



Microsoft Bing **ulan:500009365**

ALL WORK IMAGES VIDEOS MAPS NEWS SHOPPING

16,000 Results Any time


Salvador Dalí | MoMA
<https://www.moma.org/artists/1364>

Spanish, 1904–1989. The artist, author, critic, impresario, and provocateur Salvador Dalí burst onto the art scene in 1929 and rarely left the public eye until his death six decades later. The auspicious occasion was the debut in Paris of Un Chien Andalou, a film Dalí made in collaboration with Luis Buñuel. Filmed in Paris, Un Chien Andalou strung together free-associative vignettes and ...

Salvador Dalí - Simple English Wikipedia, the free ...
https://simple.wikipedia.org/wiki/Salvador_Dalí

Salvador Dalí (11 May 1904 – 23 January 1989) was a Spanish painter who became famous for the unusual images he used in his paintings. He was born in Figueres, Catalonia, Spain. He was a key figure in surrealist art. His most famous work was The Persistence of Memory (1931), which is now in MoMA, the Museum of Modern Art in New York. It is a dream-like landscape with a soft, melted ...

Images of Ulan:500009365
<bing.com/images>



See all images >

File:A Christmas card to the Lucas family, a kneeling ...
 cas_family,_a...

Dec 31, 2017 · File:A Christmas card to the Lucas family, a kneeling angel and two other figures MET DP876056.jpg

Medium: etching **Date:** 1967date QS:P571,+1967-00-00T00:00:00Z/9
Description: Print; Prints **Title:** A Christmas card to the Lucas family, a kne...

Salvador Dalí – Wikipedia
https://fr.wikipedia.org/wiki/Salvador_Dalí

Salvador Felipe Jacinto Dalí i Domènech, efter 1982 Marqués de Púbol (* 11.Mei 1904 uun Figueres, Kataloonien; † 23. Janewoore 1989 uk diar), wiar en spoonsken mooler, graafiker an skulptör. Luke k diar. Det diar sidj as tullest di 8. Janewoore 2019, am a klook 00:07 feranert wurd.

Normdooten: WorldCat Identities, VIAF: ...

Salvador Dalí - Wikipedia
https://de.wikipedia.org/wiki/Salvador_Dalí

Results: Found websites for this 'thing' on the web.
[ulan: 500009365]

How big is the Microsoft Academic Knowledge Graph?

The Microsoft Academic Knowledge Graph as of 2018-11 contains, among others,

- 209,792,741 papers
- 253,641,783 authors
- 25,431 affiliations
- 1,380,196,397 references
- 146,257,535 citations
- 48,650 journals
- 15,704 conference instances
- 4,337 conference series
- 229,716 fields of study

<http://ma-graph.org/>

Potential use cases:

- Entity-centric exploration of papers, researchers, affiliations, etc. (e.g., concerning some research area)
- Easier data integration through use of RDF and by linking resources to other data sources (e.g., combining collections in RDF).
- Data analysis and knowledge discovery (e.g., measuring the popularity of papers and authors; recommending papers)

Schema view: <http://ma-graph.org/schema-linked-dataset-descriptions/>

Results

Bing

Ulan:500009365

Images Video

142 Results Any time

vocab.getty.edu
<vocab.getty.edu/ulan/500009365>
 We would like to show you a description here but the site won't allow us.

Salvador Dalí | MoMA
<https://www.moma.org/artists/1364?=&page=1&direction=>

Ulan 500009365 View the full Getty record Information from Getty's Union List of Artist Names © (ULAN), made available under the ODC Attribution License. View the Salvador Dalí Foundation website. Exhibitions Painting and Sculpture Changes 2013

Salvador Dalí – Wikipedia
https://fr.wikipedia.org/wiki/Salvador_Dalí

Salvador Felipe Jacinto Dalí i Domènech, efter 1982 Marqués de Púbol (* 11.Mei 1904 uun Figueres, Kataloonien; † 23. Janewoore 1989 uk diar), wiar en spoonsken mooler, graafiker an ...

Normdooten: WorldCat Identities, VIAF: 6400...

Салвадор Дали – Уикипедия
https://bg.wikipedia.org/wiki/Салвадор_Дали

Салвадор Дали е роден в 8:45 часа на 11 май 1904 година. във Фигерас, провинция Херона, Каталония, Испания в семейството на проспериращ нотариус.Фигерас е селскостопанско градче, намиращо се в подножието на Пиренеите ...

Академия: Кралска академия за изящни из... **Националност:** Испания
Починал: 23 януари 1989 г. (84 г.), Фигерас, ... **Роден:** 11 май 1904 г., Фигерас, Испания

Սալվադոր Դալի - Վիքիպեդիա՝ ազատ հանրագիտարան
https://hy.wikipedia.org/wiki/Սալվադոր_Դալի

Սալվադոր Դոմինգո Ֆելիպե Խասինտո Դալի ի Բոնիպարտե, Ֆիգերաս, Alt Empordà - հունվարի 23, 1904 (84 տարեկան), Ֆիգերաս, Կատալոնիա, Իսպանիա

Ծնվել է: մայիսի 11, 1904
Քաղաքացիություն: Իսպանիա

File:Dalí.Rinoceronte.JPG - Wikipedia
<https://en.wikipedia.org/wiki/File:Dalí.Rinoceronte.JPG>

The photographic reproduction of this work is covered under the article 35.2 of the Royal Legislative Decree 1/1996 of April 12, 1996, and amended by Law 5/1998 of March 6, 1998, which states that: Works permanently located in parks or on streets, squares or other public thoroughfares may be freely reproduced, distributed and communicated by painting, drawing, photography and audiovisual

<https://www.bing.com/>

Marcia L. Zen

Images captured 2020-11-18



fast.oclc.org/searchfast/?&limit=keywords&facet=all&query=Kennedy%2C John F. (John Fitzgerald)%2C 1917-

searchFAST FAST (Faceted Application of Subject Terminology)

Find FAST Subject Headings (Instructions)

SEARCH FAST

Keywords Kennedy, John F. (John Fitzgerald), 1917-1963 Search

FAST TERMS

Search results for: "Kennedy, John F. (John Fitzgerald), 1917-1963"

Limit Results by: All

Displaying 1 to 2 of 2

Heading	Facet	Uses
Kennedy, John F. (John Fitzgerald), 1917-1963	person	13191
Kennedy, John F. (John Fitzgerald), 1917-1963 (Spirit)	person	1

TERM DETAILS

Kennedy, John F. (John Fitzgerald), 1917-1963 [Find in WorldCat](#)

USED FOR:

- Gannaidi, 1917-1963
- JFK (John Fitzgerald Kennedy), 1917-1963
- Kan-nai-ti, 1917-1963
- Kanadī, Jūn Fītz Jīrāld, 1917-1963
- Kanīdī, Jūn F., 1917-1963
- K'enedi, 1917-1963
- Kenedi, Dzhon F., 1917-1963
- Kenedi, Džon Fridžerald, 1917-1963
- Kenedi, G'on F., 1917-1963
- Kenedijs, Džons F., 1917-1963
- Kennedi, Dzhon Fītsžerald, 1917-1963
- Kennedy, Jack, 1917-1963
- Kennedy, John Fitzgerald, 1917-1963
- Kennedy, Ken, 1917-1963

USAGE:

- LC (2017) Subject Usage: 1,530
- WC (2017) Subject Usage: 13,191

RECORD ID: fst00035588

SOURCES AND OTHER LINKS:

- [Kennedy, John F. \(John Fitzgerald\), 1917-1963--\(DLC\)n 79055297](#)
- [John F. Kennedy--http://en.wikipedia.org/wiki/John_F._Kennedy](#)
- [Kennedy, John F. \(John Fitzgerald\), 1917-1963--https://viaf.org/viaf/68910251](#)

LINKS TO FULL RECORD:

- Permanent Link <http://id.worldcat.org/fast/35588>
- MARC-21 record <http://id.worldcat.org/fast/35588.html>
- MARC-21 UTF-8 (raw data) <http://id.worldcat.org/fast/35588.mrc>
- MARC-21 xml (raw data) <http://id.worldcat.org/fast/35588.mrc.xml>
- RDF record (raw data) <http://id.worldcat.org/fast/35588.rdf.xml>

view MARC

Front-end

Resources and Other Links

1-C. Connect to real "things"

[Other projects using this approach:

- LD4P2
- "Knowledge panels"
- Cornell Univ. Library,
- Stanford Univ. library, etc.]

Images captured 2020-11-18
<http://fast.oclc.org/searchfast/>



John F. Kennedy's entry in FAST is enriched with other sources

Case: FAST

foaf:focus allows FAST terms (*skos:Concept*) to be connected to URIs that identify real-world entities, to include detailed information that is usually excluded in authority records.

schema:sameAs allows FAST terms (*skos:Concept*) to take advantage of all the various string values included in VIAF (containing dozens multilingual name authorities) without having to manually include the values in the RDF triples for the specific term.

correct coding of properties

Front-end

RECORD ID:

fst00035588

SOURCES AND OTHER LINKS:

Kennedy, John F. (John Fitzgerald), 1917-1963--(DLC)n 79055297

John F. Kennedy--http://en.wikipedia.org/wiki/John_F._Kennedy

Kennedy, John F. (John Fitzgerald), 1917-1963--<https://viaf.org/viaf/68910251>

LINKS TO FULL RECORD:

Permanent Link <http://id.worldcat.org/fast/35588>

MARC-21 record <http://id.worldcat.org/fast/35588/marc21.xml>

RDF record <http://id.worldcat.org/fast/35588/rdf.xml>

```
<foaf:focus>
<rdf:Description rdf:about="http://en.wikipedia.org/wiki/John_F._Kennedy">
<rdfs:label>John F. Kennedy</rdfs:label>
</rdf:Description>
</foaf:focus>
```

```
<schema:sameAs>
<rdf:Description rdf:about="https://viaf.org/viaf/68910251">
<rdfs:label>Kennedy, John F. (John Fitzgerald), 1917-1963</rdfs:label>
</rdf:Description>
</schema:sameAs>
```

Back-end

Ref: O'Neill, Ed, and Jeff Mixer 2013. (1) The case for faceting (2) FAST Linked Data mechanics. In 76th Annual Meeting of the American Society for Information Science and Technology (ASIS&T), Montreal, Canada, Nov. 2-6, 2013.



Possible situations for each of the datasets

- If it has used local controlled vocabularies
 - The terms used or the form representing the concepts and named entities are local.
- If it has used a pre-LOD vocabulary
 - There might be no URIs/IRIs yet.
- If a mapping decision is to be made
 - In a subject domain there could be more than one standard vocabulary.
- If it needs to map its local lists to a standardized LOD KOS*
 - Human resources and quality control are most critical and could be challenging.
 - In addition to the normal standard vocabularies, other special vocabularies might be needed.
 - Suggest checking:
 - FINTO <https://finto.fi/en/>
 - Mix'n'Match <https://tools.wmflabs.org/mix-n-match/#/>

For a dataset formed through aggregation

- in addition to the above issues, synonyms and acronyms occur in the data provided by different sources.
- Heavy disambiguation and semantic conflict controls are needed.

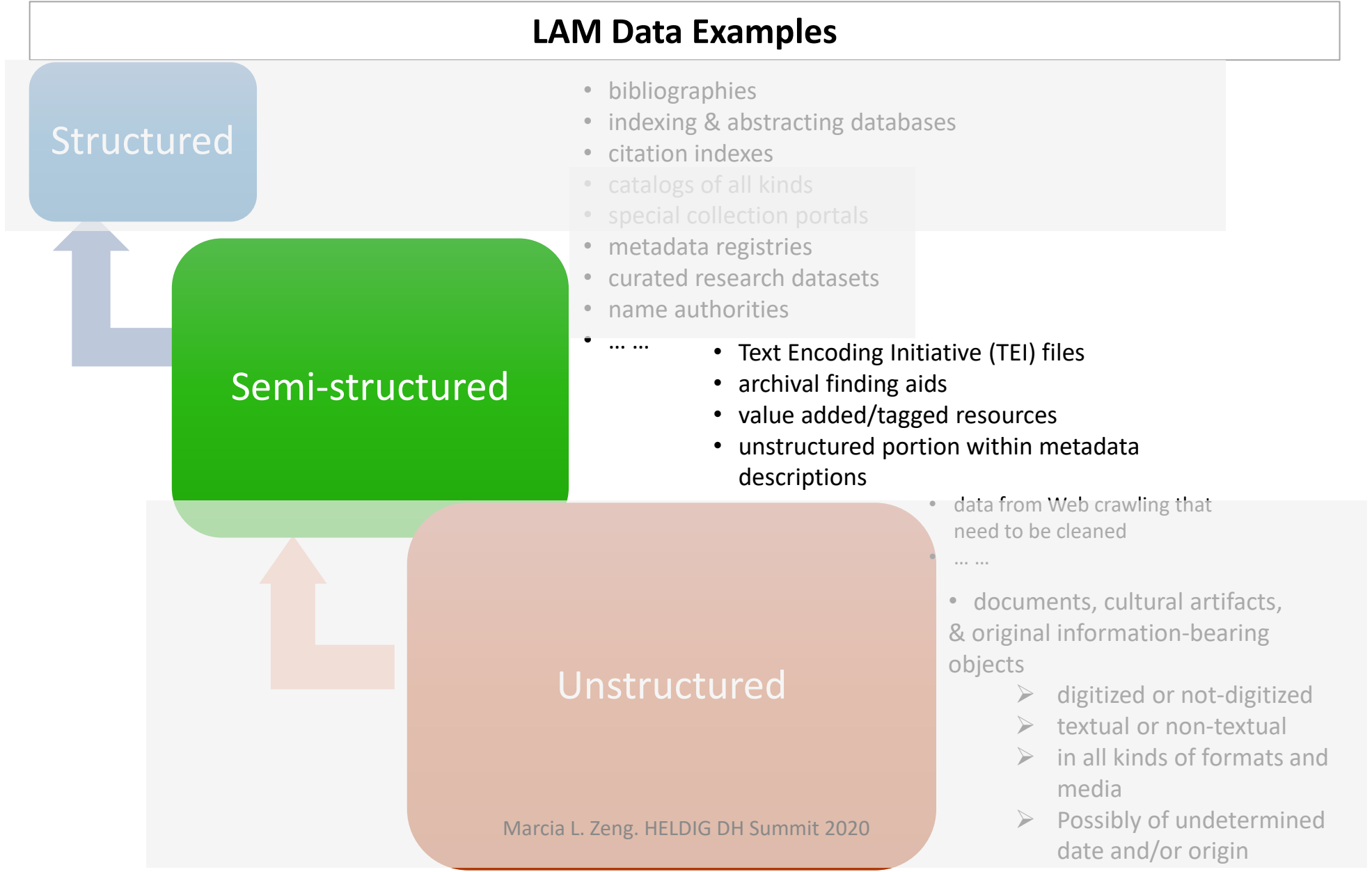


2. Semantic Enrichment for **Semi-structured Data**





2



Why use data from semi-structured data resources?



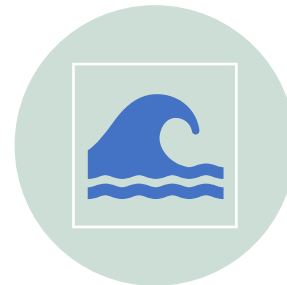
An important feature of semi-structured data resources that should be recognized, is that **they are the products of information processing.**



These semi-structured data **represent the accumulated time, knowledge, and experience of the creators** who generated them through a formal workflow which conforms to professional standards and best practices.



With semantic enrichment processes, the data values in semi-structured data are **contextualized through the metadata elements/fields;** hence, the function and meaning are clearly implied.



By parsing these data through advanced information technologies, these LAM data are dramatically **enriched and are converted into new access points.**





2. Semantic Enrichment For **Semi-structured Data** Outline

[Starting point: existing metadata components that are in free-form]

Investigations and findings (use entity extraction)

- A. MARC 5xx fields, unstructured notes, etc.
- B. Archival Finding aids' descriptions
- C. Special Collections and others
 - 1) Photograph collections' metadata (about the collection)
 - 2) Records of theses and dissertations
 - 3) Museum objects captions

Semi-structured → to → Structured Data



2- A. MARC 5xx fields, unstructured notes, TOC

Demo: entity extraction from 5xx fields

Language: English

Topics:

Entertainment Culture	84%
Film (M:H1)	70%
Israel (G:3S)	56%
Music (M:H0)	33%
Marine Port Services (TRBC) (B:64)	33%
Slovakia (G:1C)	11%
Transportation, Ground (TRBC) (B:65)	9%

Entities:

- City**
 - London, Greater
- Person**
 - Allan Rouse
 - George Harrison
 - John Lennon
 - Kevin Howlett
 - Mike Heatley
 - Paul McCartney
 - Ringo Starr
 - Tony Barrow

Social Tags:

After

500 Program notes by Kevin Howlett, Mike Heatley, and Allan Rouse, and original liner notes by Tony Barrow (18 pages : ports.) inserted in container

500 Includes Please please me mini-documentary

500 Sound recording

505 00 |tI saw her standing there --|tMisery --|tAnna (Go to him) --|tChains --|tBoys --|tAsk me why --|tPlease please me --|tLove me do --|tP.S. I love you --|tBaby it's you --|tDo you want to know a secret --|tA taste of honey --|tThere's a place --|tTwist and shout

511 0 The Beatles (John Lennon, vocals, guitar, harmonica ; George Harrison, vocals, guitar ; Paul McCartney, vocals, bass ; Ringo Starr, vocals, drums, tambourine, maracas) ; with additional musicians

518 Recorded 1962 and February 11, 1963, at Abbey Road Studios

London

Before

Tool used: Open Calais
Note: Only for assistant extraction; still need human cleaning process.



OCLC: "Mining MARC's Hidden Treasures: Initial investigations into how notes of the past might shape our future."

WHAT

Approximately 19 million records for musical resources in WorldCat were analyzed in 2016.

- Generated during the 45-year history of WorldCat;
- Comprised both musical sound recordings and musical scores;
- Approximately 2.5 million names that can be identified as distinct.

HOW

- Associating performer names with authority data
- Identifying role terms and phrases with controlled vocabularies
- Extended work has been conducted in multiple languages for
 - the performer roles
 - medium of performance terms
 - associating the name of an instrument with its performer
 - and more.

WHERE

--from uncontrolled occurrences in notes and/or statements of responsibility in records

The **extra descriptive information** may be found in such fields as:

- 245 subfield \$c (Statement of Responsibility)
- 500 (General Note)
- 505 (Formatted Contents Note)
- 508 (Creation/Production Credits Note)
- 511 (Participant or Performer Note)
- 520 (Summary, Etc.)

Weitz, Jay, Jenny Toves, Diane Vizine-goetz, Nannette Naught, and Robert Bremer. "Mining MARC's Hidden Treasures: Initial investigations into how notes of the past might shape our future." *Journal of Library Metadata* 16, no. 3-4 (2016): 166-180.



2-B. Archival Finding Aids' descriptions

Portions of a finding aid and explanation of the text used in the semantic analysis process.



**Finding Aid to the Artificial Collection:
Pearl Harbor Attack (Dec 6 – Dec 8, 1941)**

Size: (.5 cu.ft.)

Dates: December 6, 1941 – December 8, 1941

Location of Repository: Franklin D. Roosevelt Presidential Library

Name of Finding Aid Author: Ali Caron & Georgina Garcia

Date of Creation: Summer 2011

Copyright Notice: The writings of Franklin D. Roosevelt within this collection. The official writings of United States government officials within this collection. The writings of Eleanor Roosevelt within this collection are subject to Mrs. R. Other materials are subject to the United States Copyright law, 17 U.S.C. 101

Administrative Note: Franklin D. Roosevelt Presidential Library is the first presidential library used by a sitting president. The library houses documents, photographs pertaining and relating to President Franklin D. Roosevelt (FDR), the events of Pearl Harbor and the library houses materials about the attack

Scope and Content: This artificial collection is composed of photocopies of

Finding Aids to the Artificial Collection: Pearl Harbor Attack (Dec 6-Dec 8, 1941)

Series Descriptions: The collection is organized in 2 series:

- Series I: Documents – The items selected for this series remain within December 8, 1941, date range. The items reflect the Pearl Harbor attack that incident. The contents found under the container list portion are title; and folder title, found in quotations.
- Series II: Still Photographs – Images comprising this series are selected from the FDR Library's General Photograph Collection, folder: WWII: Hawaii: Attack on Pearl Harbor. Listed here are original captions taken from the photographs themselves, along with a Library control number. Unless copyright information is stated in the image caption, all of the photographs in this series belong in the public domain. This means that, to the best of our knowledge, the materials may be freely used by the researcher. However, for copyrighted materials, it is the researcher's responsibility to determine the limits of Fair Use as defined by sections 107 to 118 of the copyright law and to obtain permission from the copyright holder for further use.

Arrangement: Series I is arranged alphabetically after the president's papers and Series II is arranged numerically.

Container Lists:

SERIES I: DOCUMENTS

- OF400: Appointments; Hawaii, 1941
- OF4675: World War II; General, 1941-1942
- PPF200b: Nov. 11, 1941- Jan. 6, 1942; Public Reactions

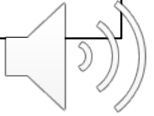
Source:

http://www.fdrlibrary.marist.edu/archives/pdfs/findingaids/findingaid_pearlharborattack.pdf

Pages

- Text from an archival finding aids
 - Descriptive information
 - creator histories
 - Scope and content notes
 - detailed description of contents, including folder and item titles
 - Abstracts from these descriptions

WHERE



Research sample

WHAT

- 43 archival record groups
- from 16 institutions, including:
 - university archives
 - government records archives
 - manuscript/special collections repositories in various LAMs

Tools used:

HOW

- OpenCalais (demo)
- COGITO Intelligent API (demo)
- MachineLinking
- Zemanta
- OpenRefine

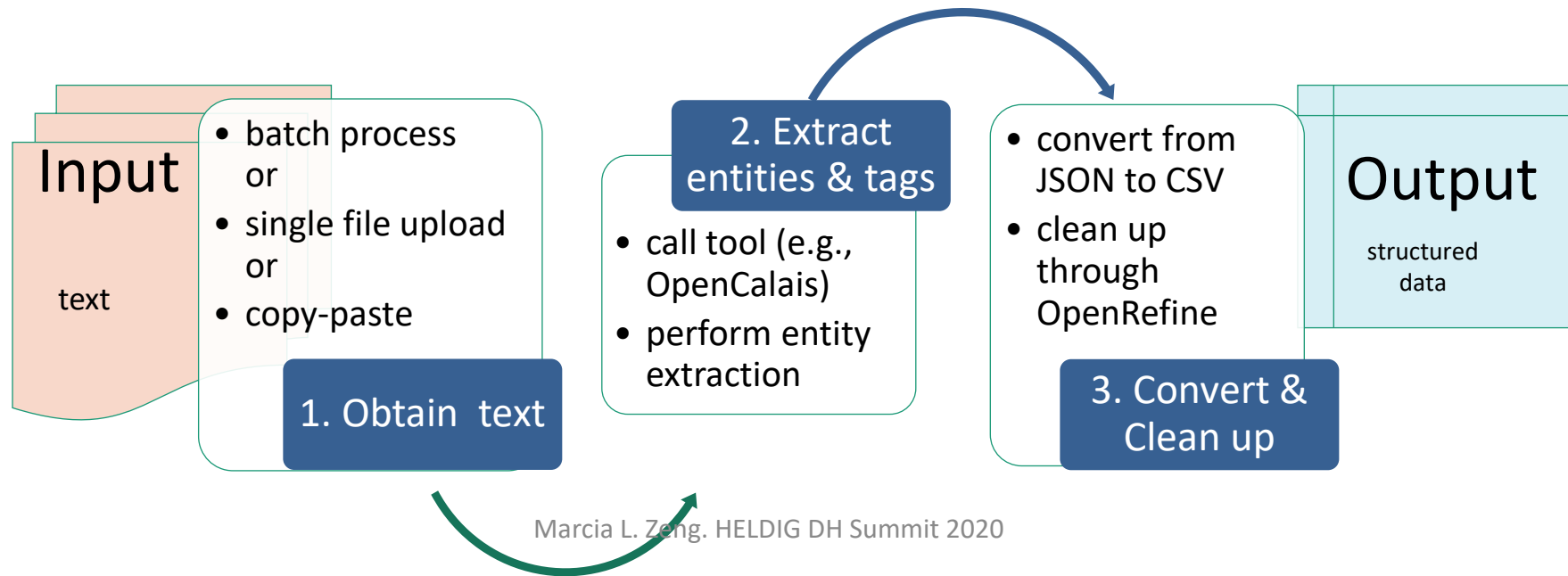
Functions:

- Entity extraction
- Tagging
- Categorization
- Semantic reasoning
- Fact mining
- Data visualization

<http://lod-lam.slis.kent.edu/SemanticAnalysis.html>

KSU iSchool
LOD-LAM team,
2013-14

HOW



The Calais initiative is about enabling semantic applications by providing a metadata generation web service, sample applications using that service to jumpstart development efforts, and support for developers.

The Calais Web Service

The Calais web service automatically attaches rich semantic metadata to the content you submit. Using natural language processing, machine learning and other methods

Enter text here:

Finding Aid to the Artificial Collection:
 Pearl Harbor Attack (Dec 6 – Dec 8, 1941)
 Size: (.5 cu.ft.)
 Dates: December 6, 1941 – December 8, 1941
 Location of Repository: Franklin D. Roosevelt Presidential Library
 Name of Finding Aid Author: Ali Caron & Georgina Garcia
 Date of Creation: Summer 2011
 Copyright Notice: The writings of Franklin D. Roosevelt within this collection are in the public domain. The official writings of United States government officials within this collection are in the public domain. The writings of Eleanor Roosevelt within this

Before

Example from the semantic analysis results using OpenCalais demo tool, indicating the entities and social tags generated.

Social Tags:

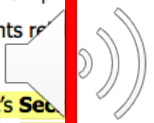
Hawaii	☆☆☆
Film	☆☆☆
Attack on Pearl Harbor	☆☆☆
Pearl Harbor advance-knowledge conspiracy theory	☆☆☆
National Pearl Harbor Remembrance Day	☆☆☆
USS Arizona Memorial	☆☆☆
Battleship Row	☆☆☆
Pearl Harbor	☆☆☆
Geography of the United States	☆☆☆

Finding Aid to the Artificial Collection:
 Pearl Harbor Attack (Dec 6 – Dec 8, 1941)
 Size: (.5 cu.ft.)
 Dates: December 6, 1941 – December 8, 1941
 Location of Repository: **Franklin D. Roosevelt** Presidential Library
 Name of **Finding Aid Author:** **Ali Caron** & **Georgina Garcia**
 Date of Creation: Summer 2011
 Copyright Notice: The writings of **Franklin D. Roosevelt** within this collection are in the public domain. **The official writings** of **United States** are in the public domain. The writings of **Eleanor Roosevelt** within this collection are subject to **Mrs. Roosevelt's** literary estate. All other materials are in the public domain.
 Administrative Note: **Franklin D. Roosevelt Presidential Library** is the first presidential library and the only presidential library used by a president, and photographs pertaining and relating to **President Franklin D. Roosevelt** (FDR). FDR was in office during the events of **Pearl Harbor**.
 Scope and Content: This artificial collection is composed of photocopies of original documents identified from the holdings of the Franklin D. Roosevelt Presidential Library, specifically the events of three days: December 6, 7, and 8, 1941. The items are a collection of documents gathered from other collections found in the library. The criterion for selecting the historical content is solely based on the date range—December 6, 1941 to December 8, 1941. Selected materials include: documents, diaries, telegrams, letters, memoranda, and photographs. Library staff has endeavored to make this research collection as comprehensive as possible; this collection does not represent the entirety of records of **Pearl Harbor**. There is a vast amount of documents relating to the lead up to **Pearl Harbor** attack itself, and the aftermath. To simplify access, only the most relevant documents are available within this collection. **The Pearl Harbor Guide** is available for researchers seeking additional information, including documents related to the attack and WWII.
 Provenance: The Pearl Harbor artificial collection includes: **President's Official File** (OF), **President's Personal File** (PPF), **President's Secret**

Entities:

- City
- Company
- Country
- Date
- Facility

After



Research Findings

From 43 archival record groups from 16 institution
Extracted **8,096** entities and **336** suggested social tags

- Entities correctly identified via Calais analysis included:
 - personal names (Person)
 - corporate names (Company, Facility, Organization)
 - geographic names (City, Continent, Country, Natural Feature, ProvinceOrState, Region)
 - events (Holiday, PoliticalEvent)
- The scores for each identified entity may be used as a valuable clue about the importance of that entity to the overall scope of the archival collection.
- The “Social Tags,” “IndustryTerm,” and “Products” categorizations were the least reliable in terms of accuracy.

The screenshot displays a software interface with several panels. The 'Entities' panel on the left lists categories like City, Company, Country, Date, Facility, Industry Term, Organization, and Person, each with a list of items and a blue progress bar. A red arrow points from the 'Country' list to a detailed popup for 'United States (Country)', which shows 'Relevance: 54%', 'Count: 2', and geographic coordinates. The 'Events & Facts' panel on the right lists categories such as 'Armed Attack', 'Generic Relations', and 'Person Career', with a list of related items. The 'Social Tags' panel at the bottom right shows a list of tags like 'Hawaii', 'Film', and 'Attack on Pearl Harbor', each with a star rating. A speaker icon is visible in the bottom right corner of the interface.

Case: Finding Aids

Additional notes (about using those tools for entity extraction)

Suggestions based on the Finding Aids study

- It would be well worth the effort for institutions to experiment with semantic analysis methods
 - as an initial step to suggest key entities and topics,
or
 - as a final check to ensure that important concepts or entities have not been overlooked.
- For certain types of records, particularly those for which subject indexing is not common, semantic analysis may provide entity-based entry points to archival records that were not previously available.
- Such techniques will enhance subject analysis at the first two levels (description and identification) but are unlikely to be useful for interpretation of the material.

Ref:

Zeng, Marcia Lei, Karen Gracy, and Maja Zumer. 2014. Using a semantic analysis tool to generate subject access points: A study using Panofsky's theory and two research samples. *Knowledge Organization* 41(6): 440-451.

Gracy, Karen and Marcia Zeng. 2015. Creating Linked Data within Archival Description: Tools for Extracting, Validating, and Encoding Access Points for Finding Aids. [poster] *DH 2015*, June 29–July 3, 2015, Sydney, Australia.





2. Semantic Enrichment For **Semi-structured Data** Outline

[Starting point: existing metadata components that are in free-form]

Investigations and findings (use entity extraction)

- A. MARC 5xx fields, unstructured notes, etc.
- B. Archival Finding aids' descriptions
- C. **Special Collections and others**
 - 1) Photograph collections' metadata (about the collection)
 - 2) Records of theses and dissertations
 - 3) Museum objects' captions

Semi-structured → to → Structured Data



2-C. a) Theses and Dissertations

Another testing by KSU iSchool LOD-LAM team, 2013-14.

Sample: **WHAT**

44 philosophy theses

- a selected sub-sample (22) from KentLINK; and
- a random sample (22) from OhioLINK.

Focus on: **WHERE**

- **abstracts**
- titles
- keywords
- **introduction paragraphs**

Result: New structured data generated from semi-structured data

- Semantic analysis based on the abstracts generated more successful tags than those based on the titles.

Process **HOW**

Submitted to OpenCalais separately to obtain the results.

1. All of the candidate terms were counted according to Agent Names, Geographic Names, Corporate Name, and Topic Terms.
2. They were manually validated (by a philosophy master's student, with MLIS) to determine:
 - ✓ the relevance to the thesis,
 - ✓ the type of a term (e.g., named entity, tag, or general heading),
 - ✓ its availability in
 - LCNAF,
 - LCSH,
 - Wikipedia (as an entry),
 - Stanford Encyclopedia of Philosophy.



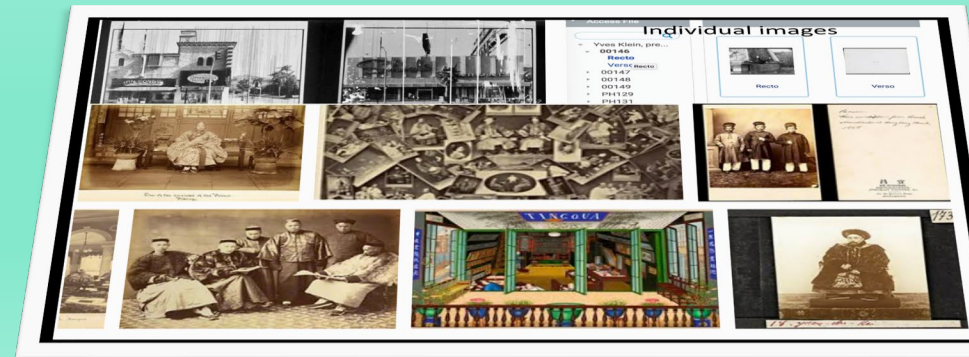
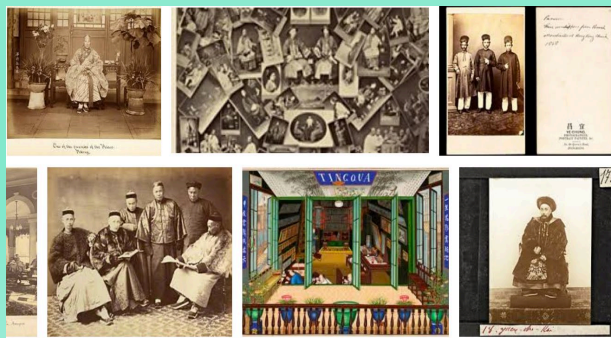
2-C. b) Photograph collections' metadata (about the collection)

R&D Derivatives

- Bibliographies
- Finding aids/Documentation
- Books, articles ...
- Exhibitions
- Portals
- Documentaries, media, ...

Collections

Individual images



A photograph collection's metadata record

Semi-structured data [inside of a structured, metadata record]

915 boxes

WHERE

Summary: Max Hutzel's "Foto arte minore" project comprises thorough photographic documentation of art historical development in Italy up to the 18th century, including objects of the Etruscans and the Romans, as well as early Medieval, Romanesque, Gothic, Renaissance and Baroque monuments. Consonant with Hutzel's belief that throughout Italy there are minor artistic centers that deserve attention, sites depicted are frequently obscure and previously undocumented. Hutzel's work is typified by a feeling for place that goes beyond the purely documentary.

Included are thorough interior and exterior documentation of secular buildings, museum holdings, ancient ruins, and religious institutions covering a broad range of artistic forms and styles, including architecture, painting, frescoes, sculpture, manuscripts, metalwork and other minor arts, ranging in date from Antiquity to late Baroque. The regions most heavily represented are: the Abruzzi, Lazio (including Rome), the Marches, and Umbria. Additional photos cover sites in: Basilicata, Campania, Emilia-Romagna, Friuli-Venezia Giulia, Lombardy, Piedmont, Puglia, Sardinia, Tuscany, and Veneto.

The collection contains more than 67,000 black-and-white prints and approximately 86,400 negatives. Because Hutzel carefully cropped his images for printing, the prints more accurately represent his style than the negatives. Also included are circa 825 photographs of medieval buildings and art in Campania, in the ancient region called "Campi Flegrei," made in 1990 by Roberto Sigismondi, Hutzel's long-time assistant (some 800 additional images from this campaign have been removed and interfiled in the Photo Study Collection's Antiquities section.)

primo.getty.edu/primo_library/libweb/action/dlDisplay.do?vid=GRI&afte

Books, Journals, Archives, Digital Coll

All items keywords anywhere in the record

Foto arte minore / Max Hutzel.
Hutzel Max.

Available at Special Collections L3 : PHOTO ARCHIVE - HUTZEL (N6911 .H88)

Request Details

Title: Foto arte minore / Max Hutzel.
Variant Title: Max Hutzel photographs of art and architecture in Italy.
Author/Creator: Hutzel Max.
Creation Date: 1960-1990

Biographical/Historical Note: German-born photographer and scholar Max Hutzel (1911-1988) photographed in Italy from the early 1960s until his death, resulting in a vast body of photographs referred to by Hutzel as Foto Arte Minore.

Arrangement: Arranged by geographic region in Italy, then by province, city, site complex and monument. Some Ancient and Medieval material have been filed in the core collection of the repository's Photo Archive.

Physical Desc.: 915 boxes..
ca. 67,275 photographic prints : b&w ; 24 x 18 cm..
ca. 86,400 negatives : b&w ; 10 x 13 cm. or smaller..

Cumulative Index/Finding Aid: Finding aid: Online database HUTZEL, in the repository's STAR system, provides a monument-level finding aid for the collection.

Notes: Title from associated documentation.

Form/Genre: Black-and-white prints (photographs)
Black-and-white negatives

Subjects: Art, Italian
Art -- Italy
Architecture -- Italy
Architecture, Ancient -- Italy
Architecture, Medieval -- Italy
Architecture, Renaissance -- Italy
Architecture, Baroque -- Italy
Cities and towns -- Italy
Streets -- Italy
Painting, Italian
Sculpture, Italian
Decorative arts -- Italy

Contributors: Sigismondi Roberto.

LC Call Number: N6911 .H88

ID/Accession Number: 94-F92
86.P.8

Access/Rights: Open for use by qualified researchers.
Photographs and permission to publish must be obtained from copyright holder(s).

OCLC Record Number: 406995970

Persistent Link: http://primo.getty.edu/GRI:GETTY_ALMA21124005810001551

Back to results list

http://primo.getty.edu/primo_library/libweb/action/dlDisplay.do?vid=GRI&afterPDS=true&institution=GETTY&docId=GETTY_ALMA21124005810001551



Before

Original descriptive metadata: →

Cumulative Index/Finding Aid: Finding aid: Online database HUTZEL, in the system, provides a monument-level finding aid.

Notes: Title from associated documentation.

Form/Genre: Black-and-white prints (photographs)
Black-and-white negatives

Subjects:
 Art, Italian
 Art -- Italy
 Architecture -- Italy
 Architecture, Ancient -- Italy
 Architecture, Medieval -- Italy
 Architecture, Renaissance -- Italy
 Architecture, Baroque -- Italy
 Cities and towns -- Italy
 Streets -- Italy
 Painting, Italian
 Sculpture, Italian
 Decorative arts -- Italy

Contributors: Sigismondi Roberto.

LC Call Number: N6911 .H88

ID/Accession Number: 94-F92
86.P.8

Access/Rights: Open for use by qualified researchers.
Photographs and permission to publish must be obtained from the copyright holder(s).

OCLC Record Number: 406995970

Persistent Link: http://primo.getty.edu/GRI:GETTY_ALM/21

[Back to results list](#)

FOUND IN DOCUMENT

- Music Album
- Person
- Position
- Province Or State
 - Basilicata,Italy 20%
 - Campania,Italy 20%
 - Emilia-Romagna,Italy 20%
 - Lazio,Italy 20%
 - Lombardy,Italy 20%
 - Piedmont,Oregon,United... 20%
 - Sardinia,Italy 20%
 - Tuscany,Italy 20%
 - Umbria,Italy 20%
 - Veneto,Italy 20%

DOCUMENT VIEW

Max Hutzel's "Foto arte minore" project comprises thorough photographic documentation of art historical development in Italy up to the late 19th century. The collection includes photographs of Medieval, Romanesque, Gothic, and Renaissance architecture, painting, fresco, and sculpture. Throughout Italy there are many ancient ruins, and religious architecture, painting, fresco, and sculpture. The collection includes Antiquity to late Baroque. The collection includes photographs of medieval buildings and art in Campania, in the ancient region called "Campi Flegrei," made in 1990 by Roberto Sigismondi, Hutzel's long-time assistant (some 800 additional images from this campaign have been included in the Photo Study Collection's Antiquities section).

ans, as well as early Hutzel's belief that frequently obscure and the purely s, museum holdings, including going in date from including Rome), the

Relevance 20%

shortname	Basilicata
latitude	40.5908
longitude	16.0933
containedbycountry	Italy
forenduserdisplay	false

Additional photos cover sites in: Basilicata, Campania, Emilia-Romagna, Friuli-Venezia Giulia, Lombardy, Piedmont, Puglia, Sardinia, Tuscany, and Veneto. The collection contains more than 67,000 black-and-white prints and approximately 86,400 negatives. Because Hutzel carefully cropped his images for printing, the prints more accurately represent his style than the negatives. Also included are circa 825 photographs of medieval buildings and art in Campania, in the ancient region called "Campi Flegrei," made in 1990 by Roberto Sigismondi, Hutzel's long-time assistant (some 800 additional images from this campaign have been included in the Photo Study Collection's Antiquities section).

Each place name embeds the full name, short name, latitude & longitude, and what country contains this place.

```
<rdf:Description rdf:about="http://d.opencalais.com/er/geo/provinceorstate/ra1g-geo1539693a2-eb41-f766-dd0e-c8b9595c90a2">
  <rdf:type rdf:resource="http://s.opencalais.com/1/type/er/Geo/ProvinceOrState"/>
  <c:docId rdf:resource="http://d.opencalais.com/dochash-1/dd44e538-f3a0-35fc-8422-7d7f76aa6c04"/>
  <c:name>Basilicata,Italy</c:name>
  <c:shortname>Basilicata</c:shortname>
  <c:latitude>40.5908</c:latitude>
  <c:longitude>16.0933</c:longitude>
  <c:containedbycountry>Italy</c:containedbycountry>
  <!--Basilicata-->
  <c:subject rdf:resource="http://d.opencalais.com/genericHasher-1/00aa24c0-6b83-3042-88a8-9aed5a25efdc"/>
</rdf:Description>
```

Demo: Photograph collection

COGITO

http://www.getty.edu/research/tools/guides/bibliographies/photography_china/

Before

www.intelligenceapi.com/demo/

COGITO® Intelligence API

Home Preview Tagging Categorization Text Mining Semantic Reasoning Fact Mining Emotions Time Reference

Select one or MORE articles from Live RSS

- Mon May 14 10:52:00 CDT 2018: Ramadan fast: Should children give up food and water? (www.bbc.co.uk)
- Sun May 13 18:57:02 CDT 2018: Volcano Kilauea: What stops eruptions of lava? (www.abc.co.uk)

Or enter a webpage URL to analyze

Enter some text to analyze here

www.getty.edu/research/tools/guides_bibliographies/photography_china/index.html

Search Tools & Databases

- Primo Search
- Getty Research Portal
- Collection Inventories & Finding Aids
- Photo Archive
- Research Guides & Bibliographies
 - Photography in China, 1839-ca. 1911
 - History of Photography
 - Interpreting History through Photographs
 - Published Collections
- Digital Collections
- Article & Research Databases
 - Collecting & Provenance Research
 - BHA & RILA
 - Getty Vocabularies

History of Photography in China, 1839-ca. 1911: Selected Annotated Bibliography

Created in conjunction with the exhibition Brush & Shutter: Early Photography in China at the J. Paul Getty Museum, February 8-May 1, 2011, and the accompanying catalog of the same name

Created 2011
Compilers: Shi Chen, Julia Grimes, Tiffany Lee, Jia Tan, Linlin Wang
Editors: Jeffrey W. Cody and Frances Terpak

Note to the Reader

This bibliography represents the state of research into the field of Chinese photography through March 2010. It was created in conjunction with the exhibition Brush & Shutter: Early Photography in China at the J. Paul Getty Museum, February 8-May 1, 2011, and the accompanying catalog of the same name.

In this bibliography, pinyin will usually be used to romanize Chinese characters. For example, Guangzhou is the name for the city formerly known as Canton. However, in the cases of Macau and Hong Kong, we have retained the earlier accepted romanizations.

All Chinese names have been alphabetized according to their pinyin spellings and follow the Chinese custom of placing the surname (*xing*) first, unless the individual lives in the West and places the surname last.

Translations of Chinese-language titles have been provided where appropriate. Those translations given by the authors or publishers of a work are prefaced by an "s=" sign, while parentheses indicate a translation by the compilers of the bibliography.

A Chinese-language version of this bibliography may be found in the translation of the catalog published by Hong Kong University Press.

Introduction

Photography in China had been an overlooked area of studies, in both China and abroad, until the last decade of the twentieth century. However, in the late 1990s, possibly due to the popularity of the *Lao Zhaopian* (Old photographs) series, there was a proliferation of Chinese publications related to China and photography, particularly in a series format. Since the late 1970s, as the legitimacy of the socialist ideology merged with the reality of a market economy,

Peking Street View (detail), William Saunders (1832-1892), ca. 1860-70s, albumen print. Clark

Description about an annotated bibliography

After

COGITO® Intelligence API

Home Preview Tagging Categorization Text Mining Semantic Reasoning Fact Mining Emotions Time Reference People Organizations Places Writprint Original Text

PREVIEW displays a visual summary of Cogito Intelligence Api's features. Please click on one of the available menus (Tagging, Categorization...) to view greater detail.

Top categories & emotions

Language	High
Politics	High
History	High
China	Very High
Consideration	Medium
Desire	Medium
Dynamic	Medium

Map showing CHINA and AFRICA highlighted.

Keywords: "China's Communist Party", photo, "dissertations full text database", publication, catalog, bibliography, photograph, photography, "photography in China", "exhibition catalog", "Chinese-language version", Chinese, "full text database", article, Canton, "photographs of China", China, "Hong Kong", "press photograph", shutter, "Communist Party", exhibition

History of Photography in China, 1839-ca. 1911: Selected Annotated Bibliography

Created in conjunction with the exhibition Brush & Shutter: Early Photography in China at the J. Paul Getty Museum, February 8-May 1, 2011, and the accompanying catalog of the same name

During this period of dramatic change in China and the West, a growing curiosity about China resulted in the establishment of photographic studios, along with images taken by foreigners living or traveling in China.

Using another tool, COGITO Intelligence API

[Paragraph in an annotated bibliography]

<http://www.intelligenceapi.com/>



Canton ?

GEO Coordinate » 23.12861/113.25889

Facts in this document » [China](#)

Region » [Guangdong Province](#)

Country » [China](#)

Continent » [Asia](#)

Sentiment » [Neutral](#)

[China](#)

[Beijing](#)

[Shanghai](#)

[Cat](#)

[Southeast Asia](#)

[Europe](#)

[Macau](#)

[Hong Kong](#)

[View Woo-Chow City](#)

[Lai Afong](#)

? This window displays the **Places** entities with their geographic coordinates (left-hand column) along with 2 graphics, the semantic relationships and the Fact Mining details (right-hand column). You can switch between the relationships and the Fact Mining graphs with the buttons below the graphical map.

Canton



No relations found

Relations

capital

After

Demo: Photograph collection

COGITO

China (Asia) ?

GEO Coordinate » 35.0/105.0

Continent » [Asia](#)

People » [William Saunders](#), [Clark Worswick](#)

Organizations » [University Press](#), [View](#), [Getty Research Institute](#), [Qing](#), [Communist Party](#)

Principal limited companies » [University Press](#)

Places » [China](#), [Canton](#), [Macau](#), [Hong Kong](#), [Beijing](#), [Southeast Asia](#), [Shanghai](#), [Europe](#)

? Fact mining identifies and groups the facts with the relative entities involved. The left-hand column lists the facts extracted from the analyzed text. The different colors depend on the taxonomies of reference (see Categorization tab). For each single fact selected in the left column, the right-hand column displays the related topics and entities above each sentence (atomic knowledge unit). With the new time references extraction (beta version), you can locate facts on a time span providing a new temporal analysis perspective.

? The left column shows the "Old", "Recent", "Present" and "Future" Time References including the retrieved temporal wordings divided by subtypes (i.e. "Absolute", "Relative" and "Range" forms). On the right panel instead, all the sentences respectively containing the temporal collocations are shown with all of the domain entities and topics thus gathered in the phrases. (beta vers.)

OLD

Absolute » [1911](#), [Feb-8-May-1-2011](#), [2003](#), [1839](#), [After 1949](#).

Range » [1852-1892](#), [until the last decade of the twentieth century](#), [1840-1911](#), [in the early twenty-first century](#), [1856-1860](#), [1937-1945](#), [From 1949](#)

Relative » [formerly](#), [The time period](#)

? [China](#), [1911](#), Main elements (China)
 History of Photography in China, 1839-ca. 1911: Selected Annotated Bibliography

[China](#), [Getty Museum](#), [Feb-8-May-1-2011](#), [Feb-8](#), [May-1-2011](#), Main elements (China)
 Created in conjunction with the exhibition Brush & Shutter: Early Photography in China at the J. Paul Getty Museum, February 8-May 1, 2011, and the accompanying catalog of the same name

[China](#), [Getty Museum](#), [Feb-8-May-1-2011](#), [Feb-8](#), [May-1-2011](#), Main elements (China)
 It was created in conjunction with the exhibition Brush & Shutter: Early Photography in China at the J. Paul Getty Museum, February 8-May 1, 2011, and the accompanying catalog of the same name.

[Clark Worswick](#), [China](#), [Southeast Asia](#), [Getty Research Institute](#), [2003](#), Main elements (China)
 Clark Worswick Collection of Photographs of China and Southeast Asia, the Getty Research Institute, 2003.R.22.27

[The time period](#), [1839](#), [after 1839](#), [China](#), [of 1839](#), [1856-1860](#), [1856](#), [1860](#), [treaty port](#), [Qing](#), [half of the nineteenth century](#), [in 1911](#), [1911](#), Main elements (China)
 The time period for this bibliography - 1839 to ca.1911 - was chosen because it coincides roughly with the blossoming of the photography medium after 1839, the burgeoning power of the photographic process in the West after 1839, and the 1850s-1860s when the

ures » [treaty port](#)

nterest » [Getty Museum](#)

» [Getty Museum](#), [Getty Research Institute](#), [treaty port](#)

old (beta vers.) » [1911](#), [Feb-8-](#)

[1](#), [formerly](#), [1832-1892](#), [2003](#), [until the last decade of the twentieth century](#), [1840-1911](#), [in the early twenty-first century](#), [The time period](#), [1839](#), [after 1839](#), [of 1839](#), [1856-1860](#), [half of the nineteenth century](#), [in 1911](#), [1937-1945](#), [1945-1949](#), [After 1949](#), [From 1949](#)

[Feb-8](#), [May-1-2011](#), [1832](#), [1892](#), [2003](#), [1840](#), [1839](#), [1856](#), [1860](#), [1937-1945](#)

[China](#), [1911](#) (OLD), Main elements (China)
 History of Photography in China, 1839-ca. 1911: Selected Annotated Bibliography

[China](#), [Getty Museum](#), [Feb-8-May-1-2011](#) (OLD), [Feb-8](#), [May-1-2011](#), Main elements (China)
 Created in conjunction with the exhibition Brush & Shutter: Early Photography in China at the J. Paul Getty Museum, February 8-May 1, 2011, and the accompanying catalog of the same name

[China](#), [Getty Museum](#), [Feb-8-May-1-2011](#) (OLD), [Feb-8](#), [May-1-2011](#), Main elements (China)
 It was created in conjunction with the exhibition Brush & Shutter: Early Photography in China at the J. Paul Getty Museum, February 8-May 1, 2011, and the accompanying catalog of the same name.

[Canton](#), [formerly](#) (OLD), Main elements (Guangzhou, Canton)
 For example, Guangzhou is the name for the city formerly known as Canton.

[Macau](#), [Hong Kong](#), Main elements (Macau, Hong Kong)
 However, in the cases of Macau and Hong Kong, we have retained the earlier accepted romanizations.

[Hong Kong](#), [University Press](#), Main elements (Hong Kong)
 A Chinese-language version of this bibliography may be found in the translation of the catalog published by Hong Kong University Press.

The Writeprint of the COGITO estimates the **readability level** of a provided document collecting and forging a full set of readability indexes as well as **grammatical, lexical and semantic analysis scores**.

primo.getty.edu/primo_library/libweb/action/dlDisplay.do?vid=GRI&aft...

Books, Journals, Archives, Digital Col...

All items keywords anywhere in the record

Foto arte minore / Max Hutzel.
Hutzel Max.

Available at Special Collections L3 : PHOTO ARCHIVE - HUTZEL (N6911 .H88)

Request Details

Title: Foto arte minore / Max Hutzel.
Variant Title: Max Hutzel photographs of art and architecture in Italy.
Author/Creator: Hutzel Max.
Creation Date: 1960-1990

Biographical/Historical Note: German-born photographer and scholar Max Hutzel (1911-1988) photographed in Italy from the early 1960s until his death, resulting in a vast body of photographs referred to by Hutzel as Foto Arte Minore.

Arrangement: Arranged by geographic region in Italy, then by province, city, site complex and monument. Some Ancient and Medieval material have been filed in the core collection of the repository's Photo Archive.

Physical Desc.: 915 boxes..
ca. 67,275 photographic prints : b&w ; 24 x 18 cm..
ca. 86,400 negatives : b&w ; 10 x 13 cm. or smaller..

Summary: Max Hutzel's "Foto arte minore" project comprises thorough photographic documentation of art historical development in Italy up to the 18th century, including objects of the Etruscans and the Romans, as well as early Medieval, Romanesque, Gothic, Renaissance and Baroque monuments. Consonant with Hutzel's belief that throughout Italy there are minor artistic centers that deserve attention, sites depicted are frequently obscure and previously undocumented. Hutzel's work is typified by a feeling for place that goes beyond the purely documentary.

Included are thorough interior and exterior documentation of secular buildings, museum holdings, ancient ruins, and religious institutions covering a broad range of artistic forms and styles, including architecture, painting, frescoes, sculpture, manuscripts, metalwork and other minor arts, ranging in date from Antiquity to late Baroque. The regions most heavily represented are: the Abruzzi, Lazio (including Rome), the Marches, and Umbria. Additional photos cover sites in: Basilicata, Campania, Emilia-Romagna, Friuli-Venezia Giulia, Lombardy, Piedmont, Puglia, Sardinia, Tuscany, and Veneto.

The collection contains more than 67,000 black-and-white prints and approximately 86,400 negatives. Because Hutzel carefully cropped his images for printing, the prints more accurately represent his style than the negatives. Also included are circa 825 photographs of medieval buildings and art in Campania, in the ancient region called "Campi Flegrei," made in 1990 by Roberto Sigismondi, Hutzel's long-time assistant (some 800 additional images from this campaign have been removed and interfiled in the Photo Study Collection's Antiquities section.)

Before

Writeprint estimates the readability level of a provided document collecting and forging a full set of readability indexes as well as grammatical, lexical and semantic analysis scores. The left-hand column shows a few graphs to concretely assess the document's reading difficulty, the writer's vocabulary level, the minimum schooling grade needed to read the document flawlessly, the spotted slangs and registers, and a table showing accurate and focused text analysis. The right-hand column shows a graph which associates a set of highly validating readability indexes and marks the author's style and efficacy. The two graphs below the former display a narrow grammatical inquiry centered on the verbs' genre and their tense analysis.

Readability Index
50.38

Vocabulary richness
31.61

Grade

Registers and Slangs

Index	Value	Standard Values
Sentences Count	9.0	-
Words per Sentence	31.0	10.06-16.92
Characters per Sentence	187.89	55.34-109.92
Short Words Count	14.79	5.20-9.50 %
Different Words Count	79.41	66.10-81.26 %
Uncommon Words Count	7.35	5.22-9.62 %
Common Vocabulary Weight	65.44	42.04-100.00 %
High Vocabulary Weight	7.03	0.96-1.15 %
Technical Vocabulary Weight	3.68	58.70-68.44 %
Nouns Count	23.3	8.32-11.67 %
Verbs Count	8.6	2.94-4.96 %
Adjectives Count	11.83	9.92-12.97 %
Conjunctions Count	4.66	2.36-4.58 %
Adverbs Count	5.02	2.06-4.68 %

Writeprint

Grammatical Tenses

Verb Classes

After



2-C. c) Museum objects' captions

Demo: Museum object captions

Blue and White plum vase of the four loves in Yuan Dynasty



Before

元青花四爱图梅瓶
 元代。2006年钟祥市郢靖王墓出土。高38.7厘米，口径6.4厘米，底径13厘米。梅瓶腹部分别绘王羲之爱兰图，陶渊明爱菊图，周敦颐爱莲图，林和靖爱梅、鹤图。

The stories of four ancients and their favorites are painted on the belly of the vase, which are respectively Wang Xizhi loves the orchid, Tao Yuanming loves the chrysanthemum, Zhou Dunyi loves the lotus and Lin Jing loves plum blossom and the crane.

Yuan Dynasty (1271-1368)

After

Tool: Boson

实体识别: [查看文档](#) [结果不正确](#)

更多 较少 平衡 准确 更准确

实体类别图示: 时间 地点 人名

← Entity types

TIME, PLACE, AGENT

Entities →

keywords

关键词提取: [查看文档](#) [结果不正确](#)

名称	权重	名称	权重	名称	权重
图梅瓶	35	郢靖王	24	王羲之	19
梅瓶	33	厘米	24	元代	19
钟祥市	32	周敦颐	23	出土	18
底径	32	青花	22	口径	18
林和靖	24	陶渊明	20	腹部	17



Blue and White plum vase of the four loves in Yuan Dynasty

Source: hubei.gov.cn 08/26/2016 09:08:27



Time: Yuan Dynasty.

Unearthed from :King Yingjing's Tomb in Chongxiang city in 2006.

Height: 38.7cm,

Surface diameter: 6.4cm,

Bottom diameter: 13cm.

The stories of four ancients and their favorites are painted on the belly of the vase, which are respectively Wang Xizhi loves the orchid, Tao Yuanming loves the chrysanthemum, Zhou Dunyi loves the lotus and Lin Jing loves plum blossom and the crane.

http://en.hubei.gov.cn/culture/heritage/201608/t20160826_889291.shtml

Before

After

Tool: OpenCalais

FOUND IN DOCUMENT

- ENTITIES
 - Person
 - Lin Jing 80%
 - Tao Yuanming 80%
 - Wang Xizhi 80%
- SOCIAL TAGS
 - Chinese people 100%
 - Chinese culture 100%
 - Tao Yuanming 100%
 - Archaeplastida 66%
 - Prunus mume 66%
 - Chrysanthemum 66%
 - Zhou Dunyi 66%
 - Wang 66%
 - Padma 66%

Tool: COGITO

MAIN ELEMENTS

belly of the vase vase favorite story love the orchid orchid

crane white lotus darling abdomen plum Lin Jing

Tao Yuanming blossom lotus Wang Xizhi ancient

love plum blossom belly Zhou Dunyi love the lotus

love the chrysanthemum chrysanthemum plum tree



Semantic analysis tools

- taxonomy and ontology-supported
- with machine learning and natural language processing behind

[Tools tested]

- **Intelligent Tagging** (previously known as **Open Calais**) Demo
 - <https://permid.org/onecalaisViewer>
- **Cogito Intelligence API** free demo version
 - <https://www.intelligenceapi.com/demo/>

[Other tools]

- **Ambiverse Natural Language Understanding - AmbiverseNLU**
 - <https://github.com/ambiverse-nlu/ambiverse-nlu>
 - → Try the [demo](http://ambiversenlu.mpi-inf.mpg.de) at <http://ambiversenlu.mpi-inf.mpg.de>
- **Gate cloud**
 - <https://cloud.gate.ac.uk/shopfront/sampleServices>
- **Spacy**
 - <https://spacy.io/usage/linguistic-features#entity-linking>

Cogito Intelligence API:

- **5 specific taxonomies** of terms (in over 1,000 different categories) for Intelligence, Terrorism, Cyber Crime, Crime, and Geographic domains
- **A domain ontology (updated regularly)** with a wide range of diverse topics, for example: weapons, crimes, cyber attacks, points of interest, chemical weapons, controlled substances, terrorist groups, critical infrastructure, world leaders, public companies and more

<https://www.intelligenceapi.com/>





The take-aways so far

- Semantic analytics, one of the advanced semantic enrichment methods, has been used for analyzing, searching, and presenting information by using **explicit semantic relationships** between known entities.
- The tools used in the experiments are powered by multiple **taxonomies and domain ontologies**, and benefit from **machine learning** and other new **artificial intelligence (AI)** technologies, far beyond normal natural language processing.
- On top of rule-based systems, embedding-based systems **for Knowledge Graph** completion has become a dominating focus in research and development during recent years.



The take-aways so far (cont.)

- The examples reveal that, ultimately, additional useful data can be derived from **large digital collections** as well as from **individual item-centered information clusters**.
- These activities can be managed **case-by-case**.
 - from the top-down or the bottom-up
 - collectively or independently
 - with or without significant project funding
- Aggregation can be based on the pieces/chunks of information one needs from a dataset, without integrating a whole database or converting full metadata records.



Revisit

Why use data from semi-structured data resources?



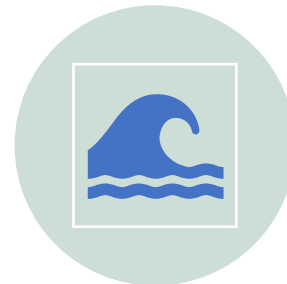
An important feature of semi-structured data resources that should be recognized, is that **they are the products of information processing.**



These semi-structured data **represent the accumulated time, knowledge, and experience of the creators** who generated them through a formal workflow which conforms to professional standards and best practices.



With semantic enrichment processes, the data values in semi-structured data are **contextualized through the metadata elements/fields;** hence, the function and meaning are clearly implied.



By parsing these data through advanced information technologies, these LAM data are dramatically **enriched and are converted into new access points.**

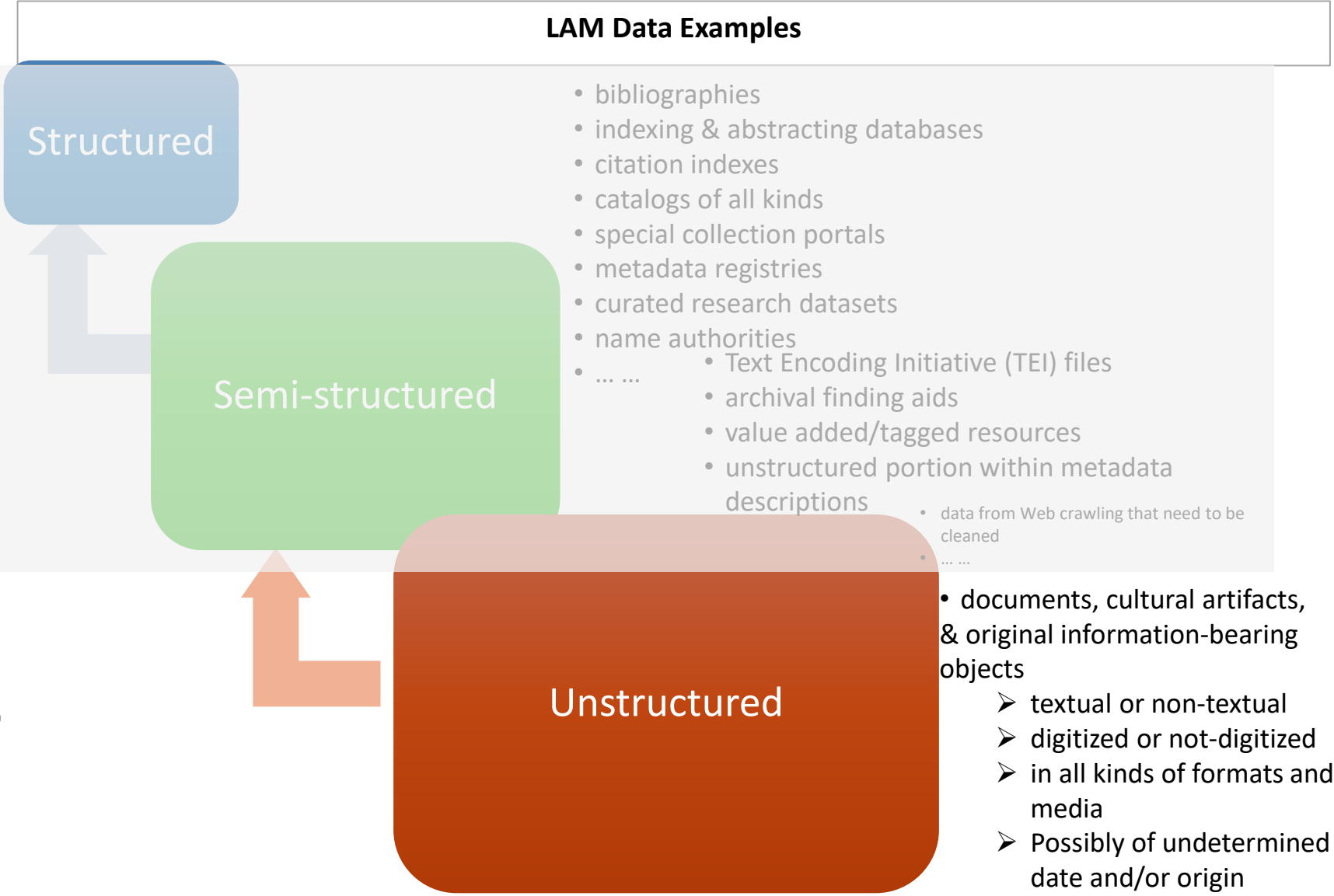


3. Semantic Enrichment for **Un-structured Data**



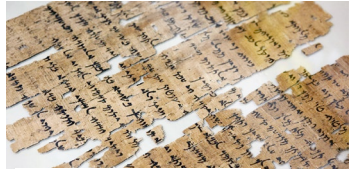


3



Through LAMs: ***unstructured data*** found in documents and other information-bearing objects:

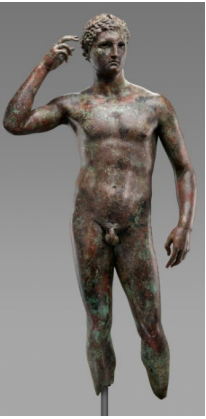
- Are available in the largest quantity.
- Have the most diversity in type, nature, and quality.
- Are the most challenging to process.



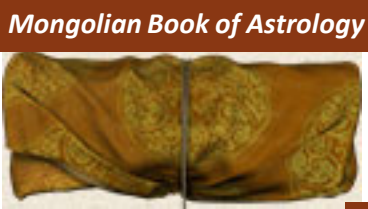
Music Treasures Consortium
Library of Congress



20th Century Press Archives, ZBW



Edwin Smith Surgical Papyrus



Mongolian Book of Astrology

Calculating the Cycle	Generating the Cycle	Planning for Cycle
21.99	0.05	0.02
1.44	0.01	0.01
1.08	0.02	0.02
0.28	0.05	0.04
gly	his	tlso leo

The Marshall Nirenberg Charts: The "First Summary"



3. Semantic Enrichment for Un-structured Data

[Outline](#)

Examples:

- A. Oral history transcripts
- B. Images
- C. Maps
- D. Murals
- E. Cultural objects
- F. Intangible Cultural Heritages
- G. ...

New structured data generated from **unstructured data** supporting knowledge discovery

Content-based, Semantic-based (vs. word-based)

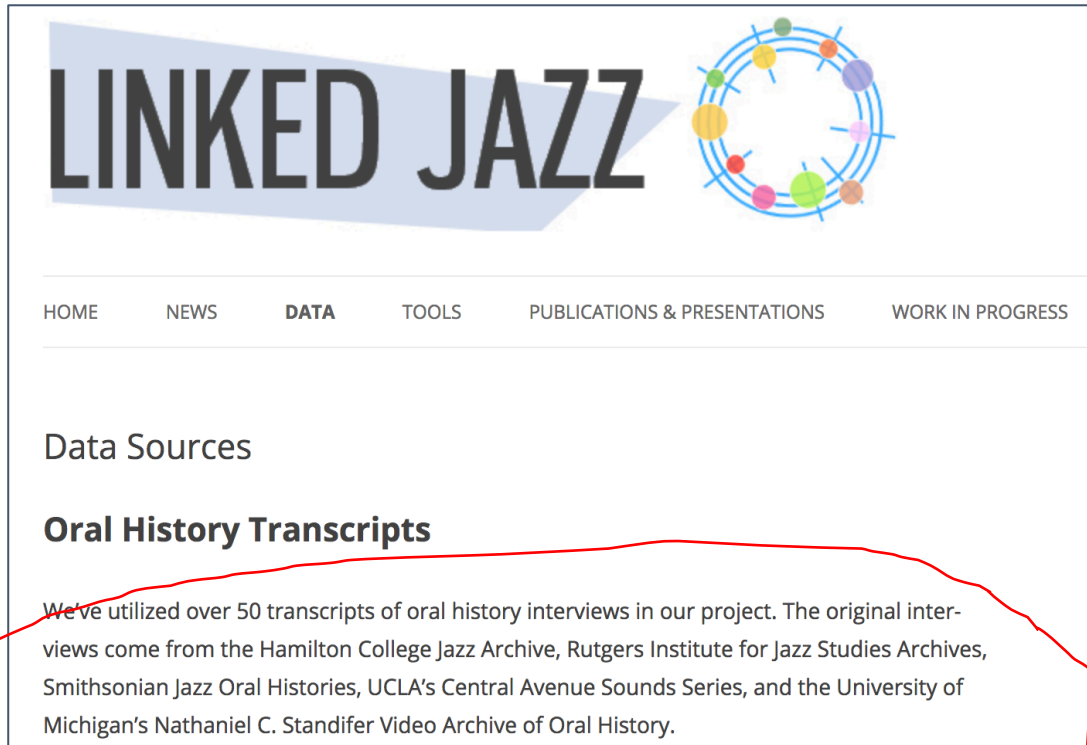


3-A. Oral history transcripts

Case: Linked Jazz

<http://linkedjazz.org/>

The project focuses on digitalized archives of jazz history to expose relationships between musicians and reveal their community's network.



The screenshot shows the Linked Jazz website interface. At the top, the text 'LINKED JAZZ' is displayed in a large, bold, sans-serif font. To the right of the text is a circular logo composed of several colored dots (yellow, green, blue, purple, red, pink) connected by thin lines, forming a network. Below the header is a navigation menu with the following items: HOME, NEWS, DATA, TOOLS, PUBLICATIONS & PRESENTATIONS, and WORK IN PROGRESS. The main content area is titled 'Data Sources' and features a sub-section 'Oral History Transcripts'. A red hand-drawn circle highlights the following text: 'We've utilized over 50 transcripts of oral history interviews in our project. The original interviews come from the Hamilton College Jazz Archive, Rutgers Institute for Jazz Studies Archives, Smithsonian Jazz Oral Histories, UCLA's Central Avenue Sounds Series, and the University of Michigan's Nathaniel C. Standifer Video Archive of Oral History.'

<http://linkedjazz.org/network/>

Marcia L. Zeng. HELDIG DH Summit 2020

Highlights

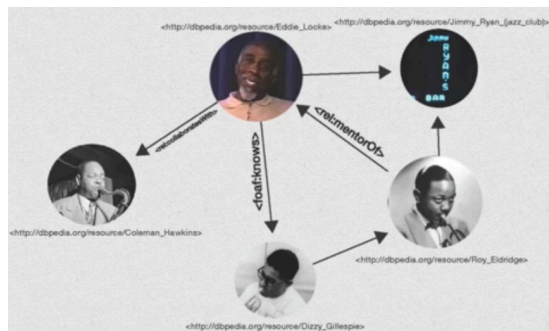
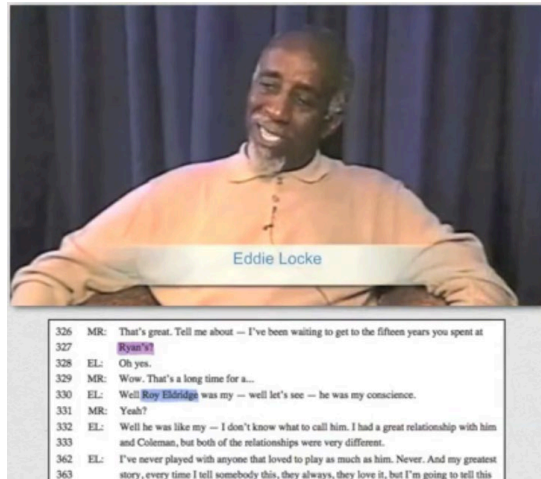
- Pioneer LOD project
- Powered by OCR and NLP

- 50+ transcripts of oral history interviews
- from 5 archives/ special collections



Methodology Summary:

- A natural language processing tool pulls excerpts from transcripts of interviews **with jazz musicians** that mention a *relationship* with **another jazz musician**.
- After the process of controlling synonyms and eliminating ambiguity, aligning with name authorities, the musician names were mapped to the DBpedia, and data about each person was obtained.
- The relationships were presented based on an ontology.
- A visualization tool was used to present a unique interactive interface.



LINKED JAZZ

Perfect (960) High (1,319) Medium (2,501) Low (1,727) Many (1,256) None (1,089) Verified (7) Deleted(1) Search

Name	Years	Perfect	High	Medium	Low	Many	None	Verified	Deleted
Paul Barbarin	1899 - 1909								
Barber	1920 - 2007								
Ross Barbour	1928 - 2011								
Eddie Barefield	1909 - 1991								
Polo Barnes	1901 - 1981								
Walter Barnes	1905 - 1940								
Bill Baron	1927 - 1989								
Harry Barris	1905 - 1962								
Court Basie	1904 - 1984								
Bob Bates	1923 - 1981								

Eddie Barefield
Eddie Barefield (December 12, 1909, Iowa - January 4, 1991, New York City) was an American jazz saxophonist, clarinetist and arranger most noteworthy for his work with Bernie Macken, Fletcher Henderson, Don Redman, Coleman Hawkins, Sonny Phlox, Bernie Young, and Ben Webster. Barefield's musical career also included work with ABC and WOR radio as well as appearances in several films. Barefield died of a heart attack at Mount Sinai Hospital in New York on January 4, 1991.
http://dbpedia.org/resource/Eddie_Barefield

LINKED JAZZ network visualization showing relationships between musicians like Oscar Peterson, Louis Bellson, J. J. Johnson, Billy Taylor, and others.

LINKED JAZZ network visualization showing relationships between musicians like Charlie Christen, Charlie Johnson, Charlie Parker, Chick Webb, Coleman Hawkins, Court Basie, Dizzy Gillespie, Drum Boogie, Duff Smith, Duke Ellington, Ella Fitzgerald, and Elmer Snowden.

LINKED JAZZ network visualization showing relationships between musicians like Chico Hamilton, Mary Lou Williams, Danny Barker, Buddy DeFranco, Gerry Mulligan, Duke Ellington, and others.

Marcia L. Zeng. HELDIG DH Summit 2020

Highlights

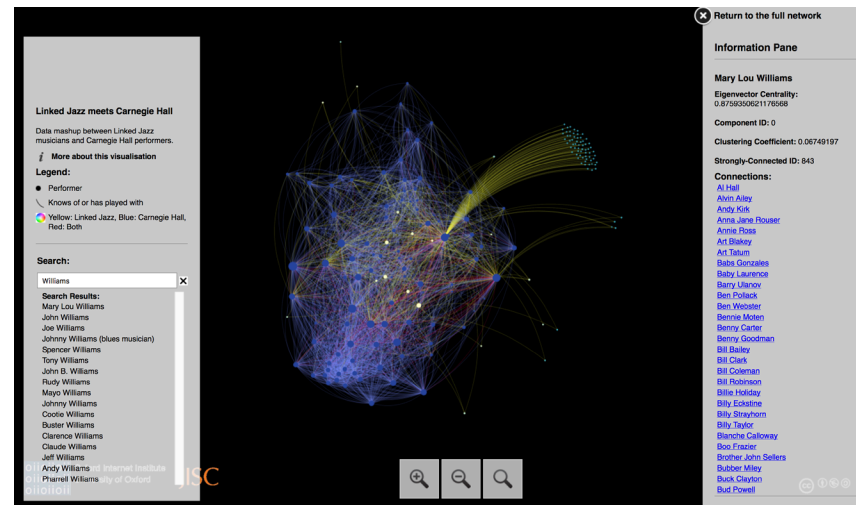
- Pioneer LOD project
- Powered by OCR and NLP
- Named entity centered
- Contextualization supported by ontology
- Recognized by music community



Data Interlinking: Linked Jazz and Carnegie Hall

- The Carnegie Hall Archives - performance history, Carnegie Hall, 1891-
- [Performance History Search](#) online, 2013 –
 - ~ 50,000 unique events encompassing roughly 90,000 people/performers
 - a significant contribution to researchers of music and cultural history.
- Two data sources selected from ‘jazz’ Carnegie Hall 1912-1955 events, RDF triples describing:
 1. 19197 people and their associated data (instrument played, birth/death date and location, profession)
 2. 154 jazz [top-level]events (performer and group names, date, place (e.g. main hall), and title (top of concert program))

- [Visualized result](#) in Gephi. (Data mashup between Linked Jazz musicians and Carnegie Hall performers.)
 - E.g., Mary Lou Williams



http://pfch.nyc/linked_jazz_meets_carnegie_hall/CH-LJ_network/index.html

http://pfch.nyc/linked_jazz_meets_carnegie_hall/CH-LJ_network/index.html#Mary%20Lou%20Williams

Highlights

- Pioneer LOD project
- Powered by OCR and NLP
- Named entity centered
- Contextualization supported by ontology
- Recognized by music community
- Connected with / reused by Carnegie Hall project





<https://iiif.io/>

IIIF Image API 3.0

IIIF Presentation API 3.0

IIIF Authentication API 1.0

IIIF Search API 1.0

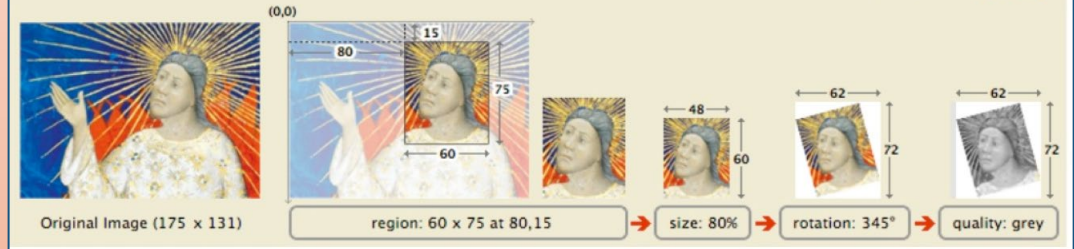
Base URL: {scheme}://{host}/{prefix}/{identifier}

Image Resource:

{base}/{region}/{size}/{rotation}/{quality}.{format}

Order of Implementation

http://www.example.org/image-service/abcd1234/80,15,60,75/pct:80/345/grey.jpg
region size quality rotation format



IIIF: Extend DMS To...



Books



Manuscripts



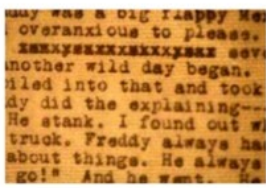
Newspapers



(Sheet) Music



Art / Vis. Resources



Archival Materials



Maps



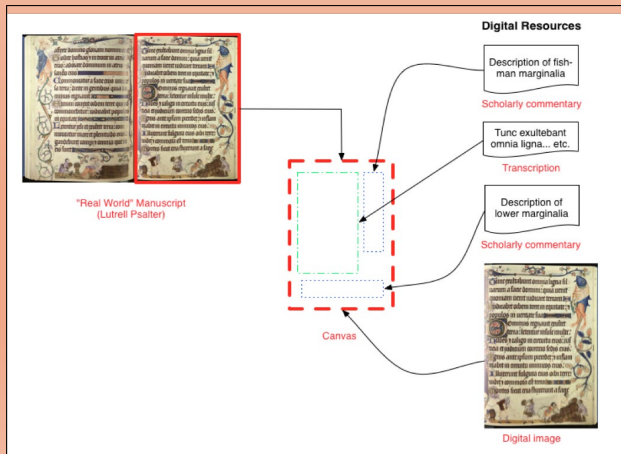
STEM Imagery



Scrolls



Architecture



Using the supported tools, annotations by experts or other contributors can be added to the canvas. Annotations, annotation lists, and content all have unique identifiers that can be processed by machines without confusion.



Structural and Contextual Views on the WCD (Wooden Slips Character Dictionary) System

Academia Sinica Center for Digital Cultures

- Core functions in WCD platform
 - * **Metadata Search** by object and by character
 - * **Image Research** on objects and characters
 - * **Image annotation** by classified categories
 - * **Cross-database query** for Chinese characters



Wooden Slips Item H01537

The written characters are living cultural evidence on the

- military and legal systems,
- educational practices,
- economy, beliefs, and
- everyday life

of military personnel and civilians in the Han dynasty (206 BC-AD 220).



Image comparison of the characters “Jia chu” (甲渠), name of a military base

。其中以
威刺麻灣
形，具有
《居延漢
形、書寫

中央研究院
歷史語言研究所
Institute of History and Philology, Academia Sinica

Digital
Cultures
Academia Sinica
Center for
Digital Cultures
中央研究院數位文化中心

歷史文字資料庫統合檢索系統
史的文字データベース連携検索システム

WCD Website: <https://wcd-ihp.ascdc.sinica.edu.tw/woodslip/index.php>

WCD (Wooden Slips Character Dictionary): Image Annotation



- Annotation under different motivations and sub-types (an extension to IIF APIs).

Motivation types

- (1) Tagging
- (2) Describing
- (3) Commenting
- (4) Classifying

- The method of classified annotation in different sub-categories makes the **annotated text more reusable for scholars' needs** by saving object data or interpreting the form and meaning of the characters.

The character “Jia” (甲) can be annotated under different motivations such as tagging, describing, commenting, or classifying (Item H04737)

See on the webpage: <https://wcd-ihp.ascdc.sinica.edu.tw/woodslip/item.php?id1=H04737>



IIF-based Union Catalog of WCD

歷史文字資料庫統合檢索系統

繁體中文

《歷史文字資料庫統合檢索系統》介紹 使用方法

查詢文字：甲

重新查詢 查詢文字 開始查詢

· 請輸入欲查詢的文字。(限中文單字)

Input field for cross-datasets retrievals

Academia Sinica, R.O.C.

聯珠字典—中央研究院歷史語言研究所|中央研究院數位文化中心

查詢結果：41筆 顯示全部圖檔

Nara Research Institute, Japan

木簡庫—奈良文化財研究所

查詢結果：12筆

University of Tokyo, Japan

草字字典數位資料庫—東京大學史料編纂所

查詢結果：7筆

Character in original database

Display in IIF Manifest JSON-LD

Display on Mirador

- “**Union Catalog**” for searching historical Chinese Characters in cooperation with international research communities in Asia.
- Functions of WCD’s Union Catalog
 - * **Character retrieval across institutes**
 - * Redirection to original database
 - * Access to the IIF Manifest structure of retrieved characters
 - * Presentation of retrieved characters in Mirador viewer
- System functions based on **IIF APIs** and **customized API for sharing of the search results**

Ref: Chen, S. & Lu, L. 2020, Linked Data as Method for Supporting DH-Research on the Cultural Resources of Chinese Wooden Slips and the Interpretation of Ancient Chinese Characters. DCMI 2020.

WCD Union Catalog : <https://wcd-ihp.ascdc.sinica.edu.tw/union/>





OCLC IIF Explorer (with CONTENTdm)

OCLC ResearchWorks IIF Explorer











Search for images of people, places, events, and more.

Welcome to the IIF Explorer

<https://researchworks.oclc.org/iif-explorer/>

OCLC ResearchWorks has created an index of all of the images in the CONTENTdm digital content management systems hosted by OCLC. The IIF Explorer offers a way to search across those collections. This prototype is called the "IIF Explorer" because the CONTENTdm images are associated with documents that use the IIF (International Image Interoperability Framework) "Presentation Manifest" standard. IIF Presentation Manifests and related images are presented with a viewer provided by Project Mirador. The IIF Explorer is an experimental prototype system, available for testing and evaluation. Please note that data in the IIF Explorer index is not guaranteed to be current with the source CONTENTdm collections, facets for type, audience, creator, etc. are partially representative of the source data based on preliminary metadata analysis and reconciliation, and the user interface is in early stages of development and testing.

Some sample searches

 The Lockheed Vega airplane	 1906 San Francisco Earthquake and Fire	 Illuminated Medieval Manuscripts	 Labor Strike Photographs and Texts	 Maps of Paris, France and its environs
 The Yellowstone Expedition	 Photographs by Carleton Watkins	 Woodcut Works and Illustrations	 Monorail Photographs and Illustrations	 Tanglewood Music Center

The IIF Explorer is an OCLC ResearchWorks prototype application provided under these terms of use.

Aggregators

- CONTENTdm
- Artstor
- DPLA (Digital Public Library of America)
- Europeana
- Internet Archive
- Wikimedia Foundation


Endless showcases!

<https://www.youtube.com/channel/UClcQIkLdYra7ZnOmMJnC5OA/videos>

(AI OCR)

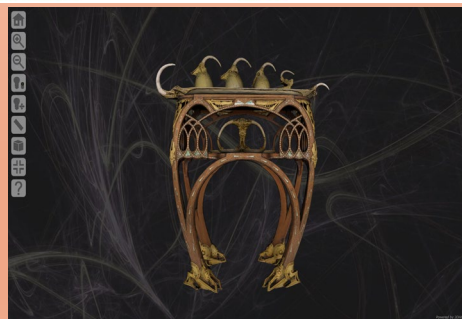
KuroNet Kuzushiji Recognition Service
<http://codh.rois.ac.jp/kuronet/>

Crop the rectangle region and export it to KuroNet Service.



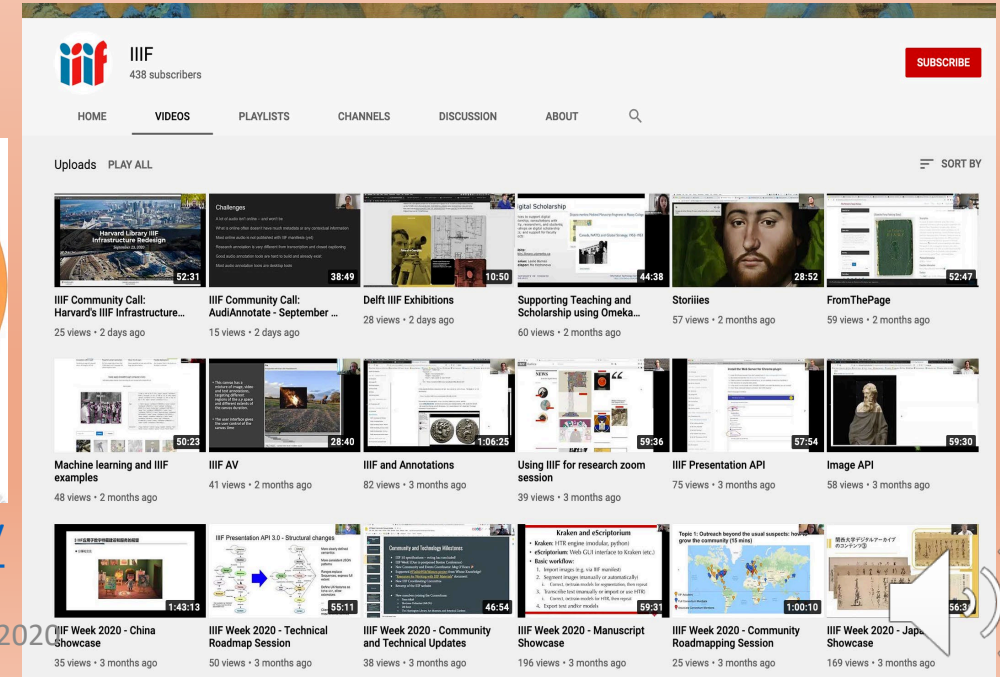
Kuzushiji OCR result is shown on IIF Curation Viewer.

<http://codh.rois.ac.jp/kuronet/>



New: 3D Community

iif.io/community/groups/3d



The screenshot shows the IIF YouTube channel page with 438 subscribers. The channel name is "IIF" and the logo is visible. The page lists several video uploads, including:

- IIF Community Call: Harvard's IIF Infrastructure... (25 views • 2 days ago)
- IIF Community Call: AudiAnnotate - September... (15 views • 2 days ago)
- Delft IIF Exhibitions (28 views • 2 days ago)
- Supporting Teaching and Scholarship using Omeka... (60 views • 2 months ago)
- Stories! (57 views • 2 months ago)
- FromThePage (59 views • 2 months ago)
- Machine learning and IIF examples (48 views • 2 months ago)
- IIF AV (41 views • 2 months ago)
- IIF and Annotations (82 views • 3 months ago)
- Using IIF for research zoom session (39 views • 3 months ago)
- IIF Presentation API (75 views • 3 months ago)
- Image API (58 views • 3 months ago)
- IIF Week 2020 - China Showcase (35 views • 3 months ago)
- IIF Week 2020 - Technical Roadmap Session (50 views • 3 months ago)
- IIF Week 2020 - Community and Technical Updates (38 views • 3 months ago)
- IIF Week 2020 - Manuscript Showcase (196 views • 3 months ago)
- IIF Week 2020 - Community Roadmapping Session (25 views • 3 months ago)
- IIF Week 2020 - Japan Showcase (169 views • 3 months ago)

Maphub

a historical map annotation portal

- Annotation
- Multilingual searching
- Geo-referencing
- Map overlays



Highlights

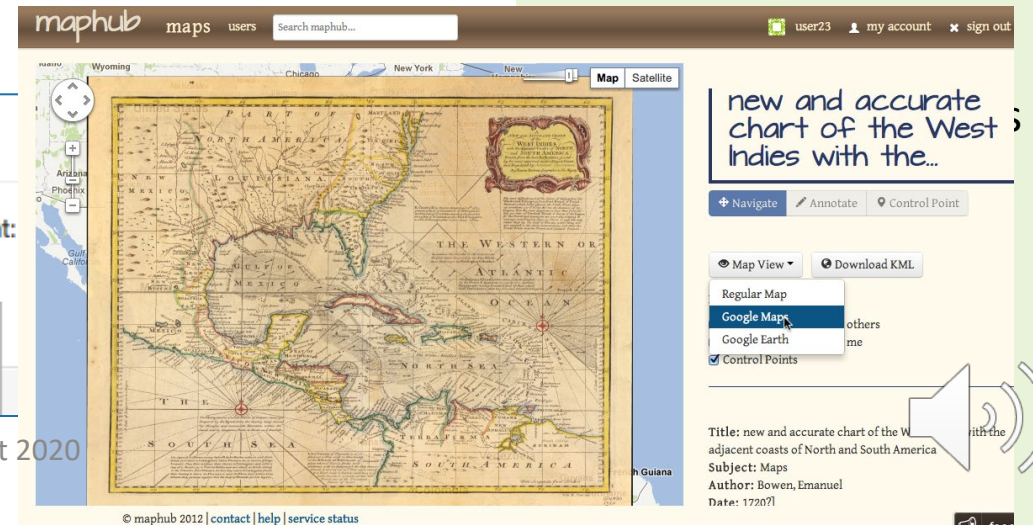
- Annotation through mash-up culture
- Supporting multilingual retrieval
- Beyond documenting maps
- Contextualization supported geo-referencing



Add Control Point

Enter a location that belongs to this point:

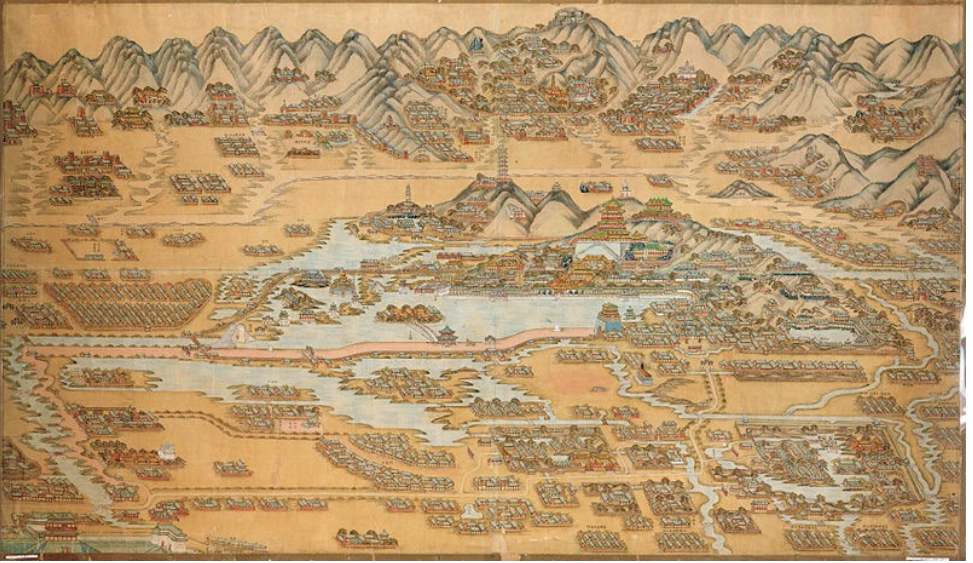
- Havana, Ciudad de La Habana, Cuba
- Havana, Florida, United States



<http://maphub.github.io/>

Marcia L. Zeng. HELDIG DH Summit 2020

My testing, 2013



"Eight Banners" Brigade barracks, temples, villages, bridges, mountains, and the Summer Palace in Beijing. Produced in Pen-and-ink and watercolor at later Qing Dynasty.

maphub maps users **北京**

After my annotation, it is searchable by Chinese.

Search Results Your search for '北京' returned 1 results.

[Beijing Yi he yuan he ba qi bing ying]

Date: [After 1888]
 Author:
 Subject: Maps, Manuscript
 Updated: 25 days ago
 1 Annotation (show all)

Highlights

- Annotation through mash-up culture
- Supporting multilingual retrieval
- Beyond documenting maps
- Contextualization supported geo-referencing

Add Annotation

Shows the "Eight Banners" Brigade barracks, temples, villages, bridges, mountains, and the Summer Palace in Beijing. Produced in Pen-and-ink and watercolor at later Qing Dynasty.

颐和园八旗兵营图
 收藏于美国国会图书馆的《颐和园八旗兵营图》，绘制于晚清时期。全图以颐和园为中心，着重绘出了遍布于颐和园周边的八旗兵营。当中也对所涉及地区的园林、寺庙有所描绘。

Tags
 Summer Palace Beijing Qing Dynasty Eight Banners

Click on a tag to accept it. Click once more to reject it.

Save annotation

Add Control Point

Enter a location that belongs to this point:
 Beijing

- Beijing, Beijing, China
- Beijingzi, Liaoning, China
- Beijing, Jiangxi, China
- Beijing, Guangdong Province, China
- Beijing, Chongqing Shi, China
- Beijing, Shanxi Sheng, China
- Beijingzui, Hubei, China
- Beijingtou, Shaanxi, China
- Beijingshan, Anhui Sheng, China
- Beijingtang, Guangdong Province, China
- Beijingling, Hainan, China
- Beijinggang, Hunan, China

Save control point

Add Control Point

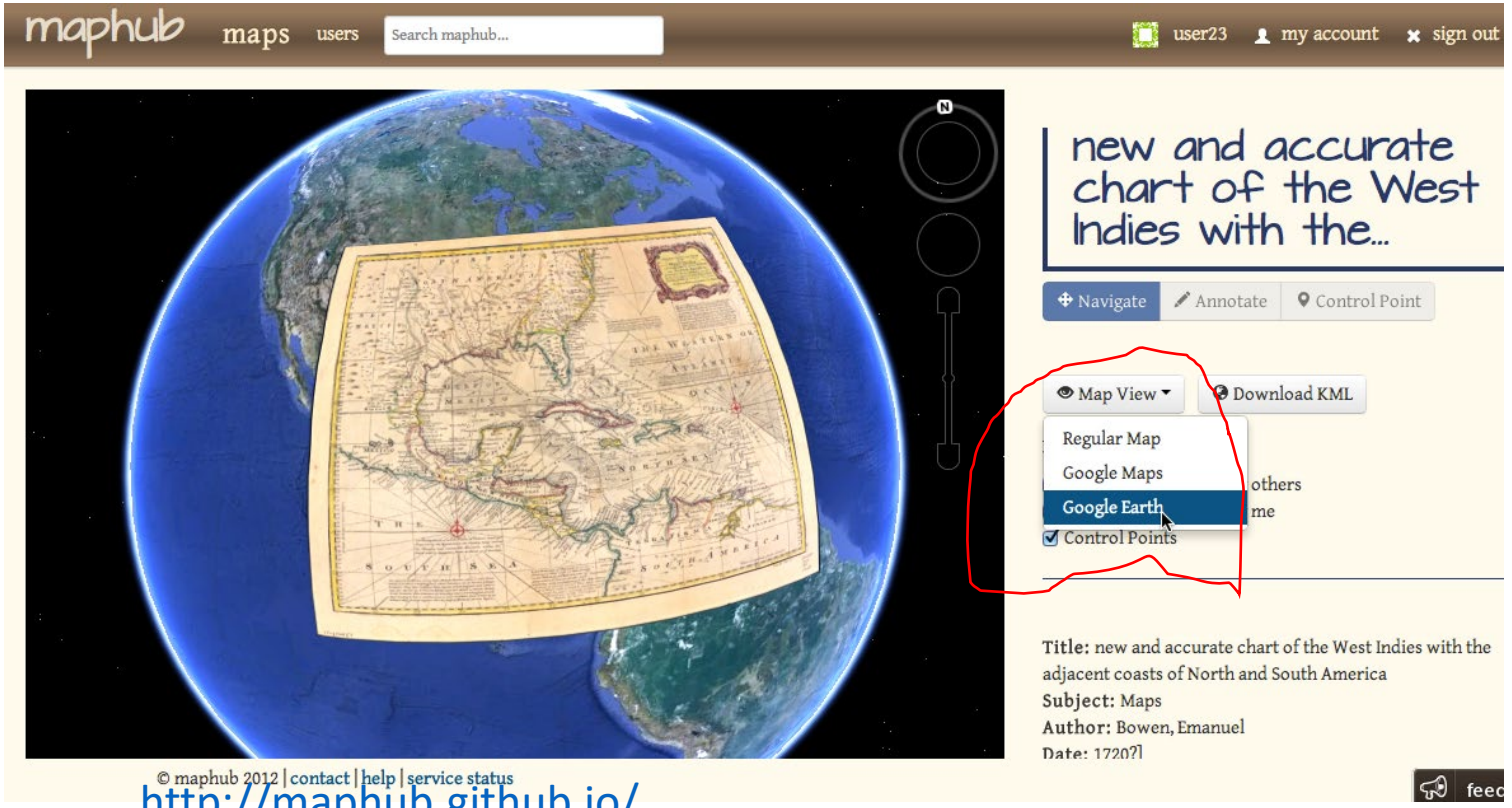
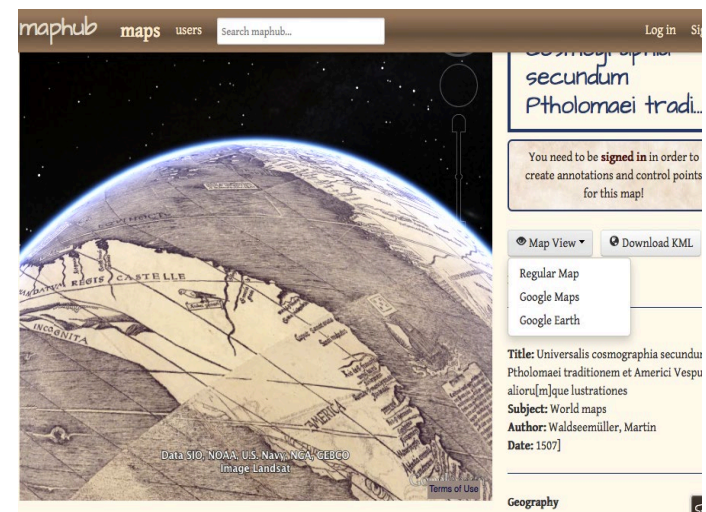
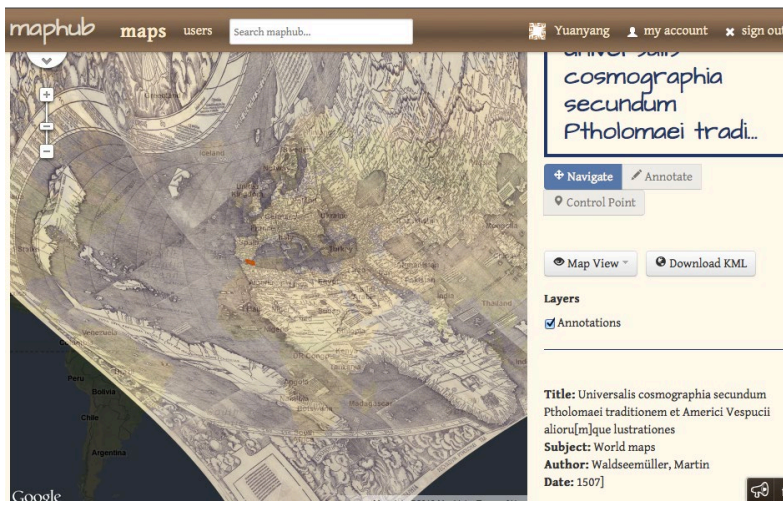
Enter a location that belongs to this point:
 Kunming Lake, Beijing, Beijing, China

Save control point



Highlights

- Annotation through mash-up culture
- Supporting multilingual retrieval
- Beyond documenting maps
- Contextualization supported geo-referencing
- Revealing history through map overlays



3-D. Murals

Dunhuang Mogao caves

(also known as the
Thousand Buddha Grottoes)

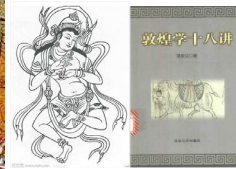
Semantic Enrichment and Thesaurus projects



Center for Digital Humanities
Wuhan University, China



- Located south-east of the Dunhuang oasis on the Silk Road, in Gansu province, China.
- Started in AD 366.
- With 492 caves, the total size of the murals reaches more than 45,000 square meters .



Dunhuang murals depict various aspects of

- medieval politics
- economics
- culture
- arts
- religion
- ethnic relations
- daily dress in Western China.





The rich content of the Dunhuang caves makes image representation a challenge.

- Identification of an image's internal objects—its “ofness”, is often ignored or lacks sufficient granularity;
- Content contained in these images continues to evade adequate semantic disclosure and connection ;
- The neglect of non-expert and novice users' needs during the process of developing and using CH digital images.

Source: Wang, et. al. 2020.



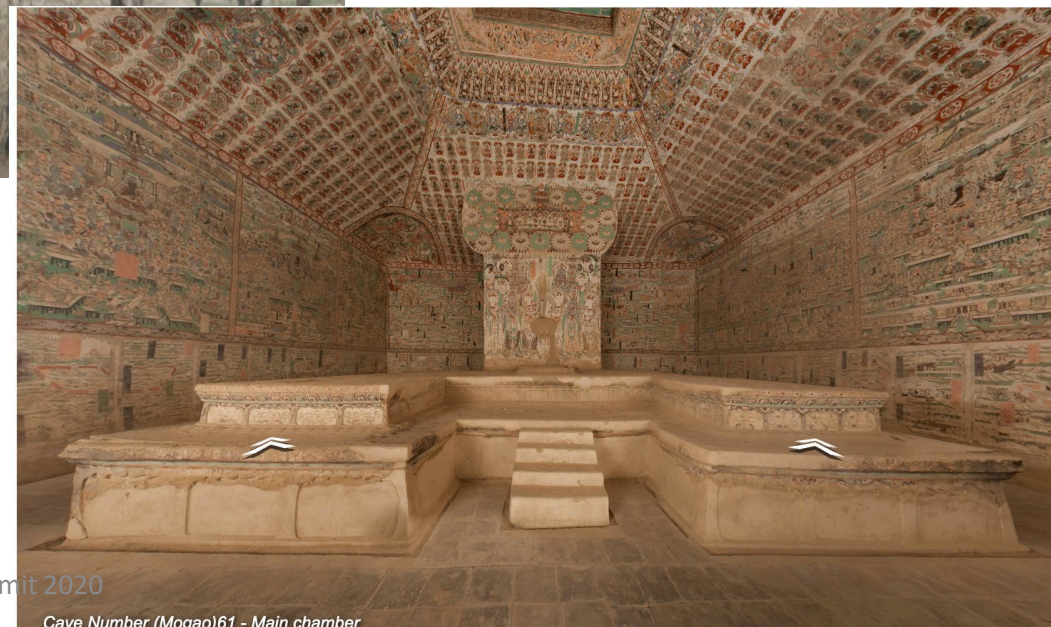
Zoom in:



Wutai Mountain Map 《五台山图》
13.45 meters long x 3.42 meters wide.

Located on the west wall of the main chamber of cave
61 of the Dunhuang Mogao Grottoes →

Image source: <https://www.e-dunhuang.com/cave/10.0001/0001.0001.0061>



Marcia L. Zeng. HELDIG DH Summit 2020

Cave Number (Mogao)61 - Main chamber



WUHAN UNIVERSITY

标注图片
POI Annotation
✕

The annotation on Mural Content

选区属性(Region) -

图片属性(ImageFile) +

The annotation on Digital image of mural

- + = x

URI for Each POI

ID

编辑	1	http://dh.whu.edu.cn/IDAMSservice/adcd121ea5a2496287701b6a	河东道山门西南	<p style="color: red; font-weight: bold;">⇒ Title of Each POI</p>
编辑	2	http://dh.whu.edu.cn/IDAMSservice/adcd121ea5a2496287701b6a	资福和尚庵	
编辑	3	http://dh.whu.edu.cn/IDAMSservice/adcd121ea5a2496287701b6a	无著和尚塔	

IIIF

{scheme}://{server}/{prefix}/{identifier}/{region}/{size}/{rotation}/{quality}.{format}

It supports users to annotate POI manually according to their metadata framework or ontology model

Marcia L. Zeng. HELDIG DH Summit 2020

Source: Wang, et. al. 2020.



Dunhuang Mural Thesaurus

Case: Digital Dunhuang

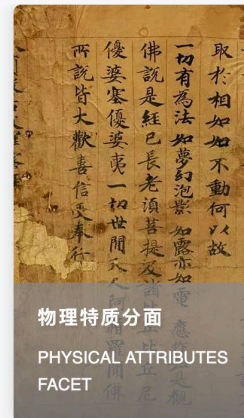
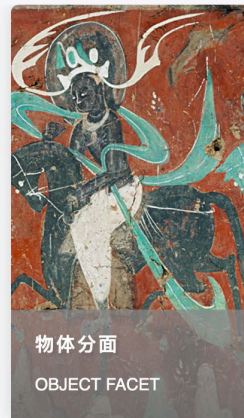
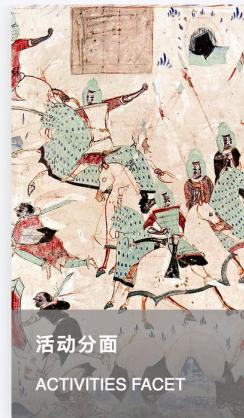
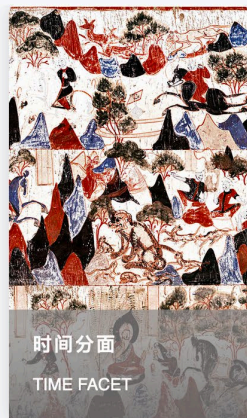
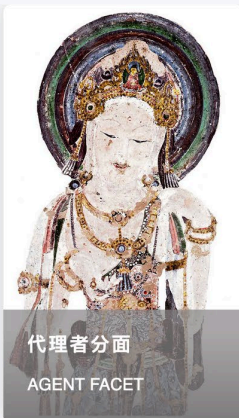
Covering:

- Mural protection and restoration
- religion
- iconography
- cave archaeology
- humanity, culture, and other research perspectives.

Type	Number
Facets	5
Hierarchy terms	83
Concepts	3199
Instances	989
Total Terms	4276

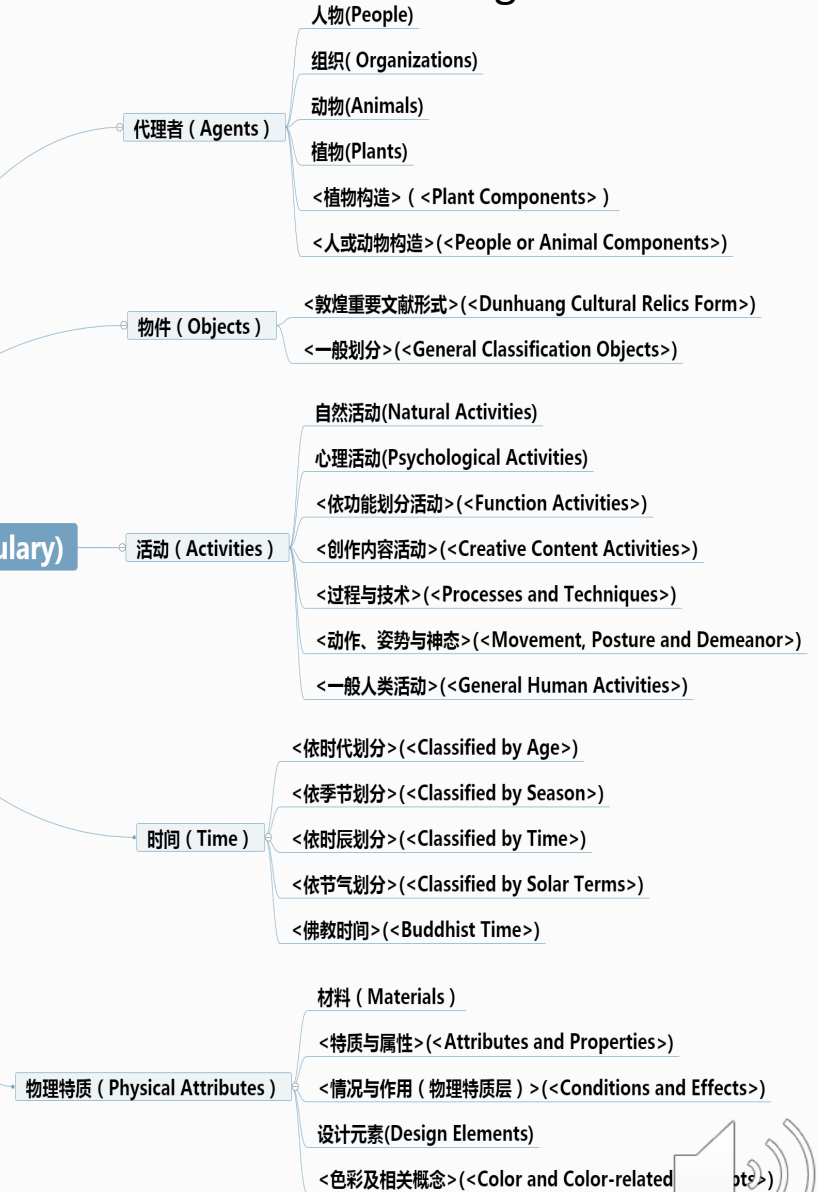
FIVE FACETS

敦煌壁画主题词表五大分面



Facets and second-level categories

敦煌词表(Dunhuang Vocabulary)



<http://dh.whu.edu.cn/dhvocab/home>

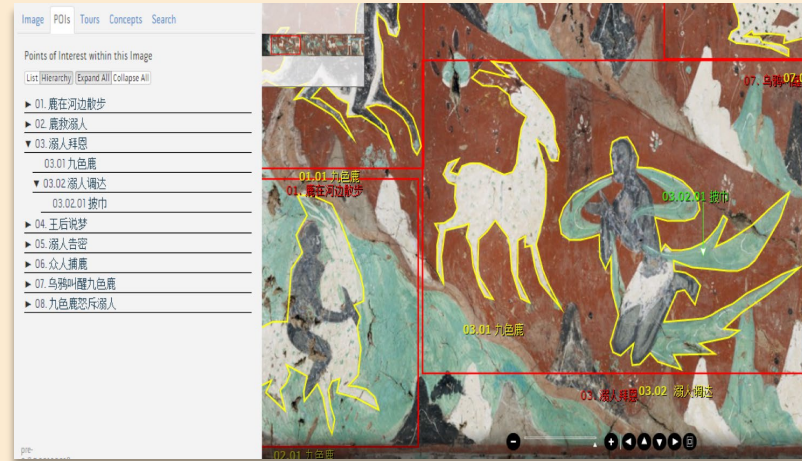
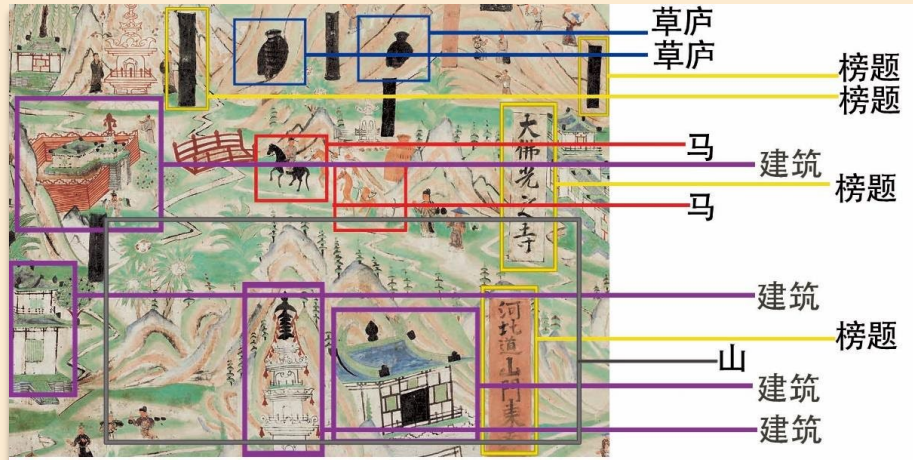
Image captured 2020-11-25

Dunhuang Mural Thesaurus used in Semantic annotation

Case: Digital Dunhuang

Highlights

- Semantic annotation (deeper than normal tagging)
- SKOSified annotated figures
- Datifying after digitizing
- Contextualization supported by thesaurus
- Re-story-telling



Source: Wang, et. al. 2020.

Marcia L. Zeng. HELDIG DH Summit 2020



Online Coins of the Roman Empire (OCRE)

<http://numismatics.org/ocre/>

A revolutionary new tool designed to help in the identification, cataloging, and research of the rich and varied coinage of the Roman Empire.

All coin types from Augustus in 31 BC to Zeno in AD 491 (representing five centuries of Roman imperial numismatics) have been published.

OCRE incorporated 107,000+ physical coins related to these coin types from 21 different datasets, originated from

- large collections a
- smaller civic or university museums,
- archaeological databases, and
- the Domuztepe excavations published through OpenContext which publishes research data on the web (Gruber 2017).



Online Coins of the Roman Empire (OCRE), a joint project of the [American Numismatic Society](#) and the [Institute for the Study of the Ancient World](#) at New York University, is a revolutionary new tool designed to help in the identification, cataloging, and research of the rich and varied coinage of the Roman Empire. The project records every published type of Roman Imperial Coinage from Augustus in 31 BC, until the death of Zeno in AD 491. This is an easy to use digital corpus, with downloadable catalog entries, incorporating over 43,000 types of coins.

As of April 2017, OCRE provides links to examples present in nearly 20 American and European databases (both archaeological and museum in context), including the [ANS collection](#), the [Münzkabinett of the State Museum of Berlin](#), and the [British Museum](#), now totalling over 100,000 physical specimens. Between these collections, OCRE is now able to illustrate 50% of the imperial coin types that it contains. Moving forward, as more collections join the project, it will eventually incorporate and display almost all recorded Roman Imperial coin types. Furthermore, it draws findspot information from another ANS-developed resource, [Coin Hoards of the Roman Republic](#), enabling the mapping of the

Language ▾

- Arabic
- Bulgarian
- Danish
- German
- Greek
- English
- Spanish
- French
- Hungarian
- Italian
- Dutch
- Polish
- Romanian
- Russian
- Swedish
- Turkish
- Ukrainian



Highlights

A pioneering, revolutionary new tool

- Ontology /knowledge base design
- Following Linked Data principles
- Innovative, user-friendly accesses

Symbols

Identify



Visualize your queries on-the-fly

<http://numismatics.org/ocre/>

Images captured 2020-12-03

Case: OCRE

Highlights

A pioneering, revolutionary new tool

- Ontology /knowledge base design
- Extremely user-friendly accesses
- Innovative research-oriented functions
- Smart data in DH



OCRE Browse Search Maps Symbols Identify a Coin Contributors Visualize Queries Feedback APIs About Language

Visualize Queries

Use the data selection and visualization options below to generate a chart

Typological Analysis

Select Category for Analysis

Select a category below to generate a graph showing the quantitative d

Material

Numeric response type

- Percentage
- Count

Compare Queries

You can compare multiple queries to generate a more complex chart. N category, the drop-down menu will include only those mints that produ

Group **+** Add Query Field

Authority (Person) Constantine I

Group **+** Add Query Field

Authority (Person) Constantine II

Group **+** Add Query Field

Authority (Person) Constantine III

Generate Chart

Select...

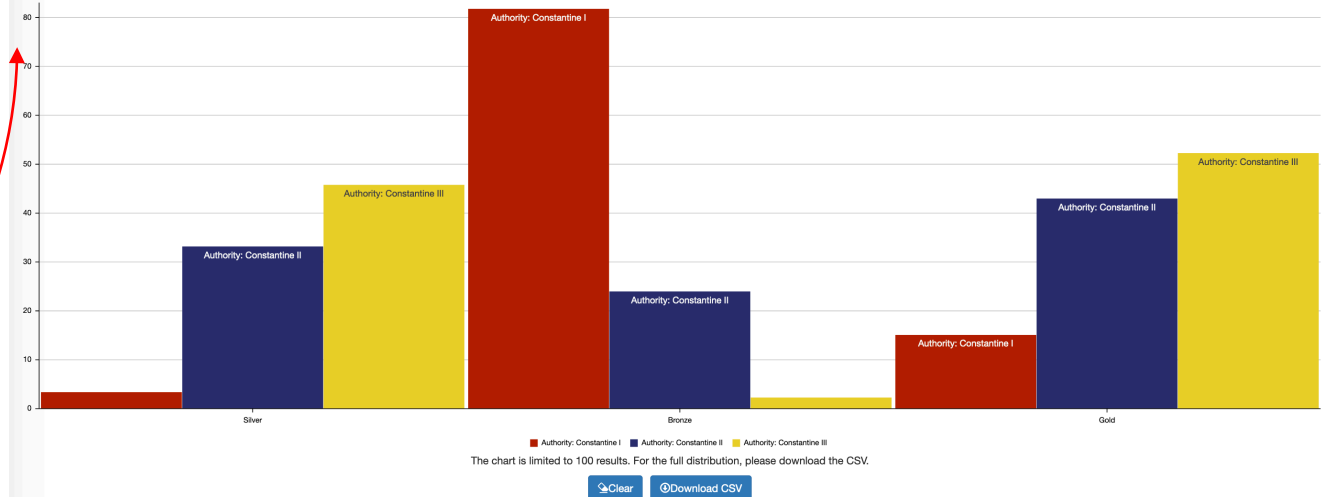
- Authority (Person)
- Authority (State)
- Coin Type
- From Date
- To Date
- Denomination
- Deity
- Issuer
- Manufacture
- Material
- Mint
- Object Type
- Portrait
- Region
- Stated Authority

OCRE Browse Search Maps Symbols Identify a Coin Contributors Visualize Queries Feedback APIs About Language

Visualize Queries

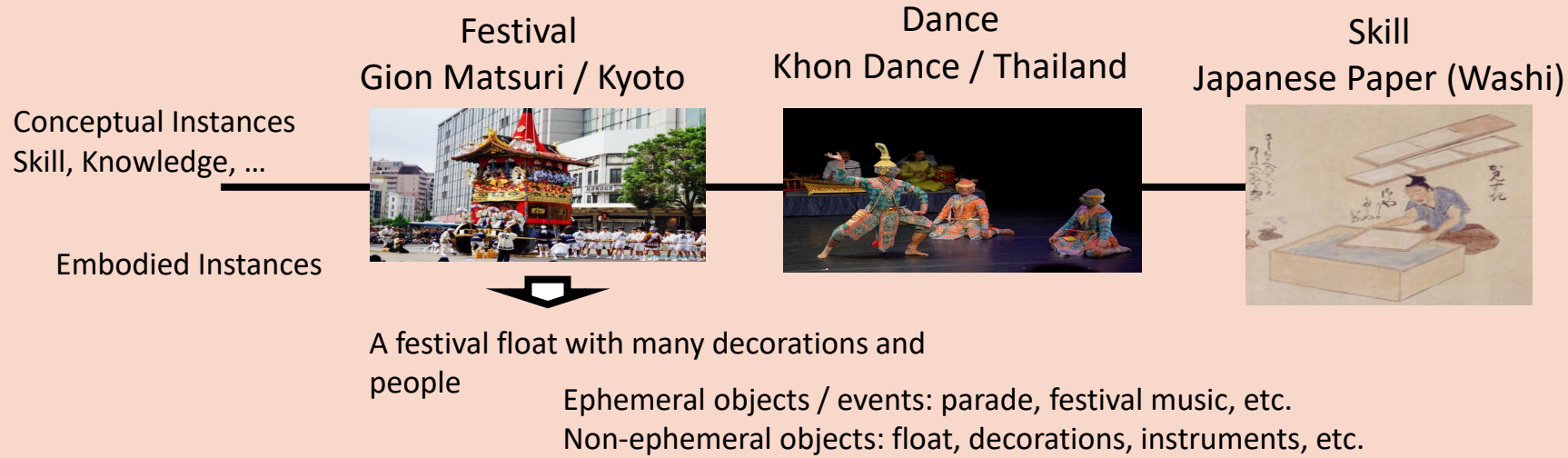
Use the data selection and visualization options below to generate a chart based on selected parameters. Instructions for using this feature can be found here: <http://wiki.numismatics.org/numishare:visualize>.

Typological Analysis



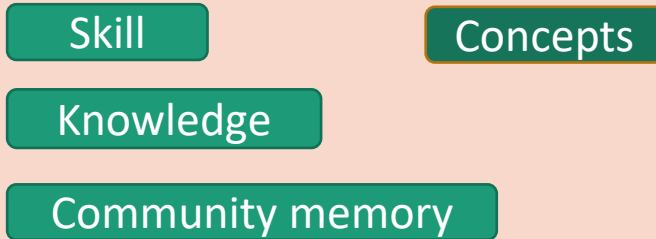
SMART data

- Adequately represent a sufficient number of relevant features of humanistic objects of inquiry to enable the necessary level of precision and nuance required by humanities scholars
- Provide users with a sufficient amount of data to enable quantitative methods of inquiry, helping researchers to surpass the limitations inherent in methods based on close reading strategies (Schöch, 2013).

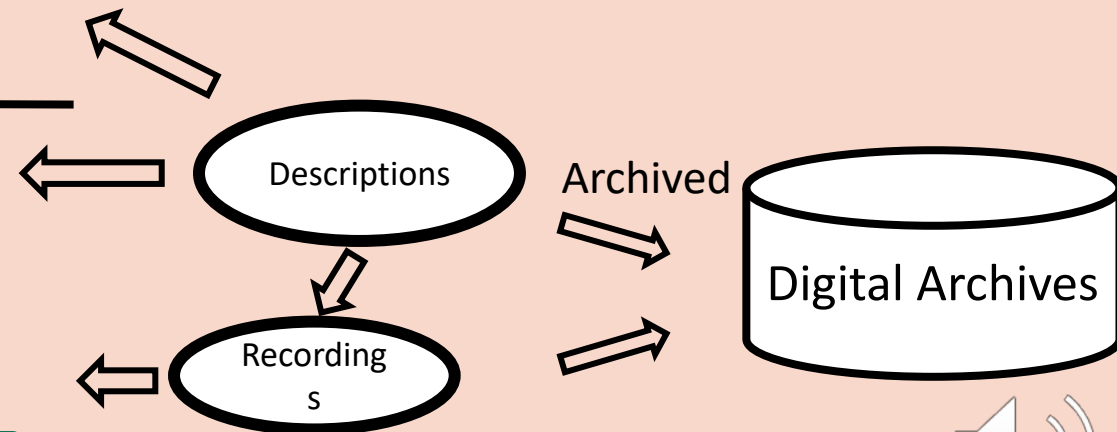
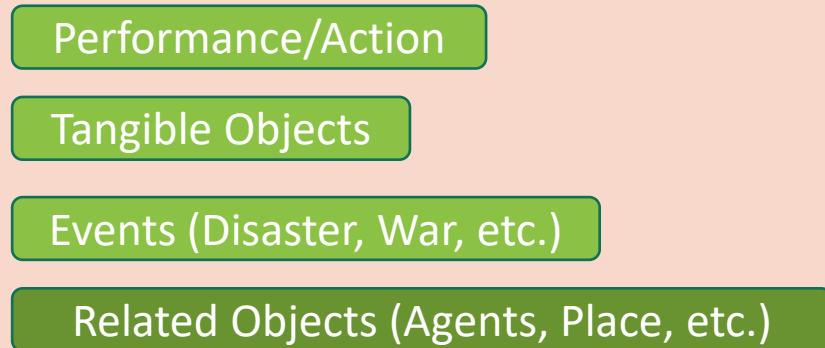


Digital Archiving Process:

Conceptual / Abstract entities



Embodied World
Physical / Digital entities



Metadata Models for Organizing Digital Archives on the Web

Metadata-Centric Projects at Tsukuba

Digital Archive = a collection of digital resources, mainly of cultural and historical resources

• Metadata-Centric Projects at Tsukuba

1. Great East Japan - Enhancing digital archives' usability
2. **Japan Earthquake Archive**
- Aggregating metadata within and across archives after the Great East Japan Earthquake and Tsunami (2011-03-11)
3. Manga - Aggregating resources on the web
4. Cultural heritage objects for digital archives – tangible and intangible cultural heritage

• Common Research Goals

- **Enrich values of digital archives by metadata aggregation** within and across digital archives and linking institutional digital archives and web resources

Great East Japan Earthquake Archive Challenges

- Quality of metadata (especially photographs and videos)
- Item-based metadata – relationships among the items during one event
- What do users want to find?
 - A specific photo?
 - A group of photos?
 - Around a place or an event?
 - From one community or across all archives and web?

Aggregation by:

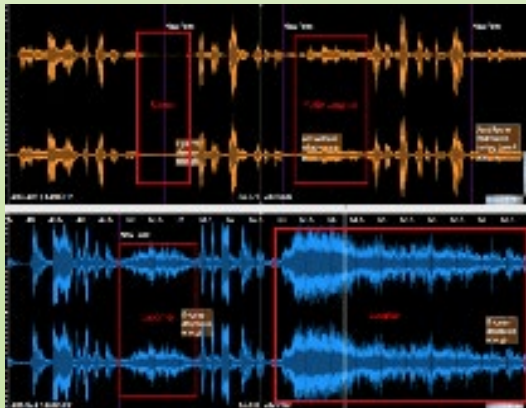
- Temporal
- Special
- Agent
- Subject information

Other notable LAM projects for cultural heritage and historical resources

Audio

Linked Reading project, U. Cincinnati

- Elliston Poetry Archive: 700+ recordings of poetry or poetry-related content
- a linked data infrastructure and sound analysis platform
- using semantic analysis of the printed text, coupled with sonic analysis of the audio archives



<http://dsc.uc.edu/projects/elliston-poetry-archive-sonification-experiments>

Scents

“Odeuropa” Project

- 1st step: develop AI to screen historical texts in seven languages for descriptions of odours – and their context – as well as to spot aromatic items within images, such as paintings.

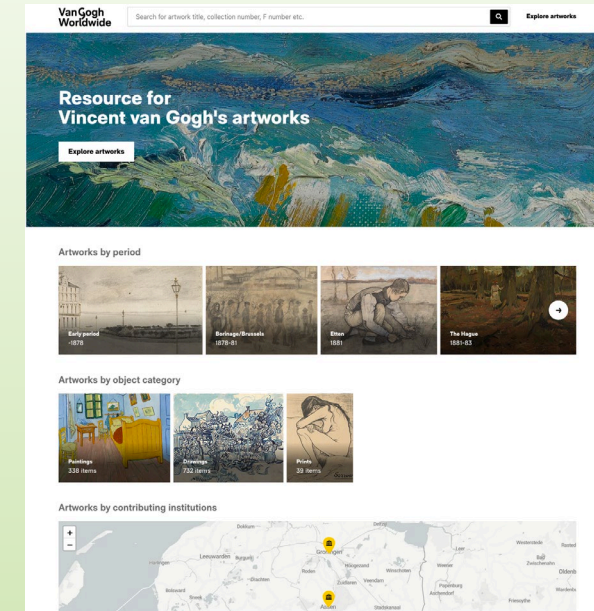


<https://www.theguardian.com/science/2020/nov/17/scents-of-history-study-hopes-to-recreate-smells-of-old-europe>

Art-historical + technical info

Van Gogh Worldwide

- Launched 2020-11-05
- Uses *Linked Art Model*
- Provides all published scholarly research results about the artworks, inter-linked by means of thesauri.
- Art-historical + technical information *about* the work of Vincent van Gogh (1853-1890).



<https://vangoghworldwide.org/>

Vincent van Gogh
The Letters

by period
by correspondent
by place
with sketches

Search
keyword or number(s) >>
Advanced search
Search results

Van Gogh as a letter-writer
Correspondents
Biographical & historical context
Publication history

About this edition
Chronology
Concordance, lists, bibliography
Book edition

316 216 160 << 315 | 317 >>

To Theo van Gogh. The Hague, on or about Thursday, 15 February 1883.

SEARCH THIS LETTER

original text + line endings facsimile translation notes artworks

316 | Show metadata

My dear Theo,
Sincerest thanks for your letter, and the enclosure was most welcome; it really does help me. I begin by telling you that it's a great relief to me that the past of the woman you write about is entirely different from what I first instinctively thought. Namely that she has known not only misery and straitened circumstances, but also other things, so I believe that she'll appreciate you fully, also as regards civilization and broader views, much more so than a woman who has been hurt by misery from an early age and knows no better than to

1. This phrasing indicates that Theo sent extra money; this is confirmed by a remark in letter 323, ll. 66-67. See also Date.

2. A year before Van Gogh had noted these words Michelet's *La femme* (in a slightly different form) drawing *Sorrow* - see letter 216, n. 3.

3. For this pronunciation by Theo, see letter 291

original text + line endings facsimile translation no

1. This phrasing indicates that Theo sent extra money confirmed by a remark in letter 323, ll. 66-67. See also Date.

2. A year before Van Gogh had noted these words Michelet's *La femme* (in a slightly different form) drawing *Sorrow* - see letter 216, n. 3.

3. For this pronunciation by Theo, see letter 291

<http://www.vangoghletters.org/vg/letters/let316/letter.html#original>

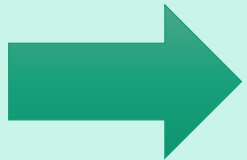
Outline

Introduction

Semantic Enrichment Approaches

- Structured Data
- Semi-structured Data
- Unstructured Data

Summary and Conclusions



Advanced semantic technologies now allow researchers to:

- access and reuse large volumes of diverse data
 - unearth patterns and connections formerly hidden from view
 - reconstruct the past
 - discover impacts in real and virtual environments
 - bring the complex intricacies of innovations to light
- all as never before --



Image source: Gandon, F. 2018. A Survey of the First 20 Years of Research on Semantic Web and Linked Data.



Enhancing Historical and Cultural Heritage Data to Support Digital Humanities Research

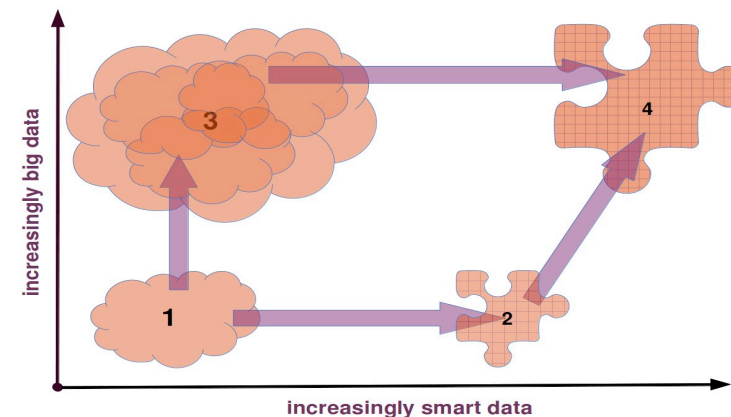
- Enhancing subject access through controlled vocabularies that have embraced Linked Data
- Transforming the semi-structured and unstructured data into structured data through semantic analysis and aligning with standardized controlled vocabularies
- Linking and contextualizing existing structured data across data silos
- Enabling one-to-many usages of LAM data in supporting digital humanities

In DH, emphasis has been on:

➤ transforming unstructured data to → structured data

Structured Data Trending:

- Machine ~~readable~~-understandable data
- Machine ~~readable~~ actionable data
- Accurate (no error) data in the processes of interlinking, citing, transferring, rights-permission, use and reuse
- One → to → Many uses and high efficiency processing data



5-star Linked Data

The baseline of our work is the [5-star Linked Data model](#), proposed [originally](#) by Tim Berners-Lee.

- ★ Make data available on the Web in whatever format.
- ★★ Make data available as structured data (e.g., Excel instead of an image scan of a table).
- ★★★ Use non-proprietary formats (e.g., CSV instead of Excel format).
- ★★★★ Use URIs to denote things, so that people can point at your data.
- ★★★★★ Link your data to other data to provide context.

7-star Linked Data Service

However, in our opinion, providing 5-star Linked Data is just the beginning. To actually make use of the datasets, consumers need more support in getting to know and access them, as well as a better grasp of their quality and provenance. To this end, we extend the model with two additional stars:

- ★★★★★★ Provide your data with a schema and documentation so that people can *understand and re-use* your data easily.
- ★★★★★★ Validate your data and denote its provenance so that people can *trust the quality* of your data.

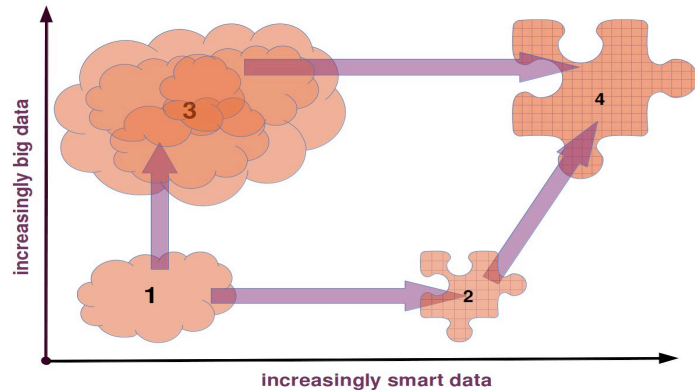
"7-star" [Linked Data Finland portal](http://www.ldf.fi/) <http://www.ldf.fi/>

Marcia L. Zeng. HELDIG DH Summit 2020



Results of Today – Visions for Tomorrow

Enhancing Historical and Cultural Heritage Data to Support Digital Humanities Research



Data Silos → Data Lakes → Data Planets → Data Universe

“Sampo” Model and Sampo-UI Framework

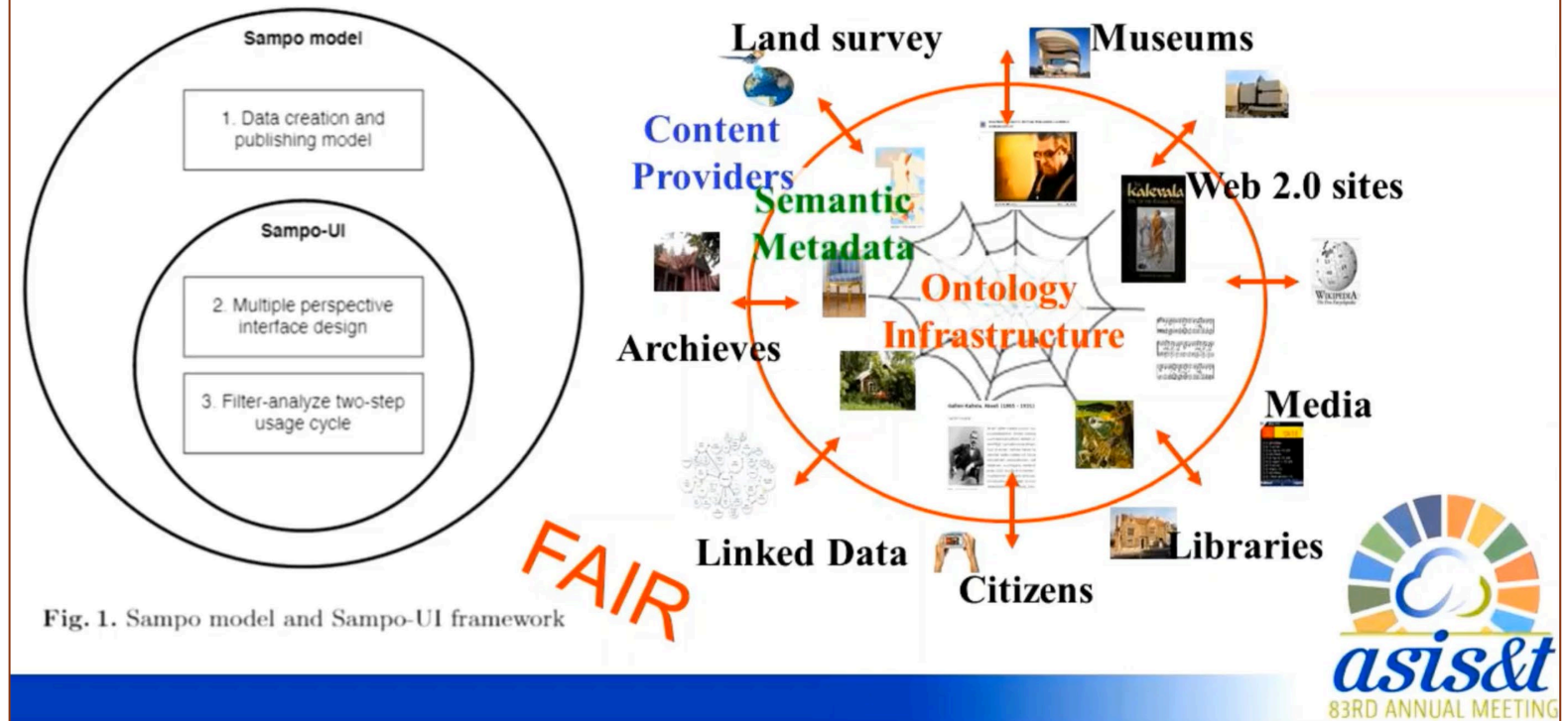


Fig. 1. Sampo model and Sampo-UI framework

Schöch, C. 2013. Big? Smart? Clean? Messy? Data in the humanities. *Journal for Digital Humanities*. 2(3): pp.2-13.

Hyvönen E. 2020. Building a National Level Linked Open Data Infrastructure for Digital Humanities in Finland. 83rd Annual Meeting of the Association for Information Science and Technology, Oct. 22- Nov.1, 2020. <https://vimeo.com/460086143>



References (1)

Borgman, C.L., 2015. *Big data, little data, no data: Scholarship in the networked world*. MIT press.

Chen, S. & Lu, L. 2020, Linked Data as Method for Supporting DH-Research on the Cultural Resources of Chinese Wooden Slips and the Interpretation of Ancient Chinese Characters. DCMI Virtual 2020. <http://shorturl.at/otzX4>

Daquino, M., Mambelli, F., Peroni, S., Tomasi, F. and Vitali, F., 2017. Enhancing semantic expressivity in the cultural heritage domain: exposing the Zeri Photo Archive as Linked Open Data. *Journal on Computing and Cultural Heritage (JOCCH)*, 10(4), pp.1-21.

Gracy, K. and Zeng, M. 2015. Creating Linked Data within Archival Description: Tools for Extracting, Validating, and Encoding Access Points for Finding Aids. *DH 2015*, June 29–July 3, 2015, Sydney, Australia.

Europeana Task Force on Enrichment and Evaluation. 2015. Report on Enrichment and Evaluation. 29/10/2015. http://pro.europeana.eu/files/Europeana_Professional/EuropeanaTech/EuropeanaTech_taskforces/Enrichment_Evaluation/FinalReport_EnrichmentEvaluation_102015.pdf

Europeana Semantic Enrichment Framework, Documentation. 17th November 2016 (updated 2017, 2018, 2020). Available from <https://pro.europeana.eu/page/europeana-semantic-enrichment>

Hyvönen, E., Tuominen, J., Alonen, M. and Mäkelä, E., 2014, May. Linked Data Finland: A 7-star model and platform for publishing and re-using linked datasets. In *European Semantic Web Conference* (pp. 226-230). Springer, Cham.

Hyvönen, E. 2020. Linked Open Data infrastructure for Digital Humanities in Finland. In: *Proc. Of the Digital Humanities in the Nordic Countries* (DHN 2020) (pp. 254-259). <https://seco.cs.aalto.fi/publications/2020/hyvonen-lodi4dh-dhn-2020.pdf>

Hyvönen, E. 2020. “Sampo” model and semantic portals for Digital Humanities on the Semantic Web. In: *Digital Humanities in Nordic Countries* (DHN 2020) (pp. 373-378). <https://seco.cs.aalto.fi/publications/2020/hyvonen-sampos-dhn-2020.pdf>

Hyvönen, E. 2020. Using the Semantic Web in Digital Humanities: Shift from data publishing to data-analysis and serendipitous knowledge discovery. *Semantic Web* 11(1) pp. 187–193. <http://semantic-web-journal.net/system/files/swj2310.pdf>



References (2)

O'Neill, Ed, and Jeff Mixter. 2013. "Maximizing the Usage of Value Vocabularies in the Linked Data Ecosystem." *76th Annual Meeting of the American Society for Information Science and Technology (ASIS&T), Montreal, Canada, Nov. 2-6, 2013.*

<http://nkos.slis.kent.edu/ASIST2013/ONeill-Mixter.pptx>

Schöch, C. 2013. Big? Smart? Clean? Messy? Data in the humanities. *Journal for Digital Humanities*. 2(3): pp.2-13.

<http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/>

Sugimoto, S., Kiryakos, S., Wijesundara, C., Monika, W., Mihara, T. and Nagamori, M., 2018. Metadata models for organizing digital archives on the Web: Metadata-centric projects at Tsukuba and lessons learned. In *International Conference on Dublin Core and Metadata Applications, 2018* (pp. 95-105).

Sugimoto, S. 2019. Modeling Culture – a perspective from digital archives and metadata. In *International Conference on Dublin Core and Metadata Applications, 2019*.

Svensson, P. 2010. The Landscape of Digital Humanities. *Digital Humanities Quarterly*. 4(1).

<http://digitalhumanities.org/dhq/vol/4/1/000080/000080.html>

Weitz, J., Toves, J., Vizine-Goetz, D., Naught, N. and Bremer, R., 2016. Mining MARC's hidden treasures: initial investigations into how notes of the past might shape our future. *Journal of library metadata*, 16(3-4), pp.166-180..

Wang, X., et al. 2020. Representation and Display of Digital Images of Cultural Heritage: A Semantic Enrichment Approach. *Knowledge Organization* (forthcoming).

Zeng, M.L., Gracy, K.F. and Žumer, M., 2014. Using a semantic analysis tool to generate subject access points: A study using Panofsky's theory and two research samples. *Knowledge Organization*. 41(6), pp.440-451.

Zeng, M.L., 2019. Semantic enrichment for enhancing LAM data and supporting digital humanities. Review article. *El profesional de la información*, 28(1). <http://www.elprofesionaldelainformacion.com/contenidos/2019/ene/03.html> [open access]

