New Discovery Tools

for Digital Humanities and Spatial Data
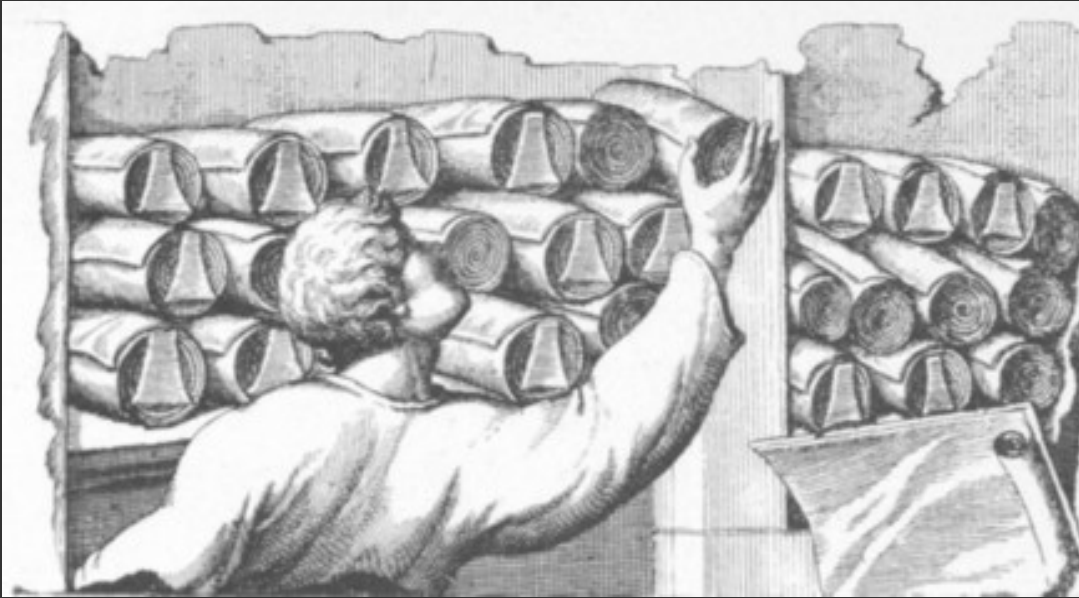
MIT July 2014

Merrick Lex Berman

Center for Geographic Analysis, Harvard

Physical discovery

John Willis Clark, *The Care of Books* (Cambridge, 1901).
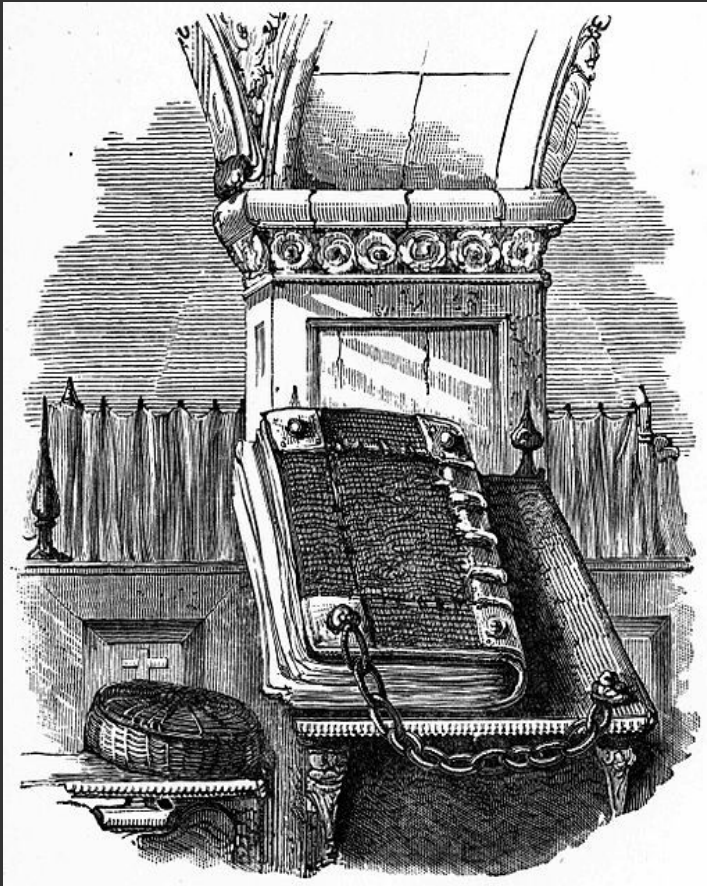*Paul Pelliot at the Mogao Caves, Dunhuang* (1908).

John Willis Clark, *The Care of Books* (Cambridge, 1901).
The Emperor and the Assassin (1999).

Bound and chained

St. Walburga Library, Zutphen, Netherlands 1564).
Andrews, William: "Curiosities of the Church: Studies of Curious Customs, Services and Records" (1891)
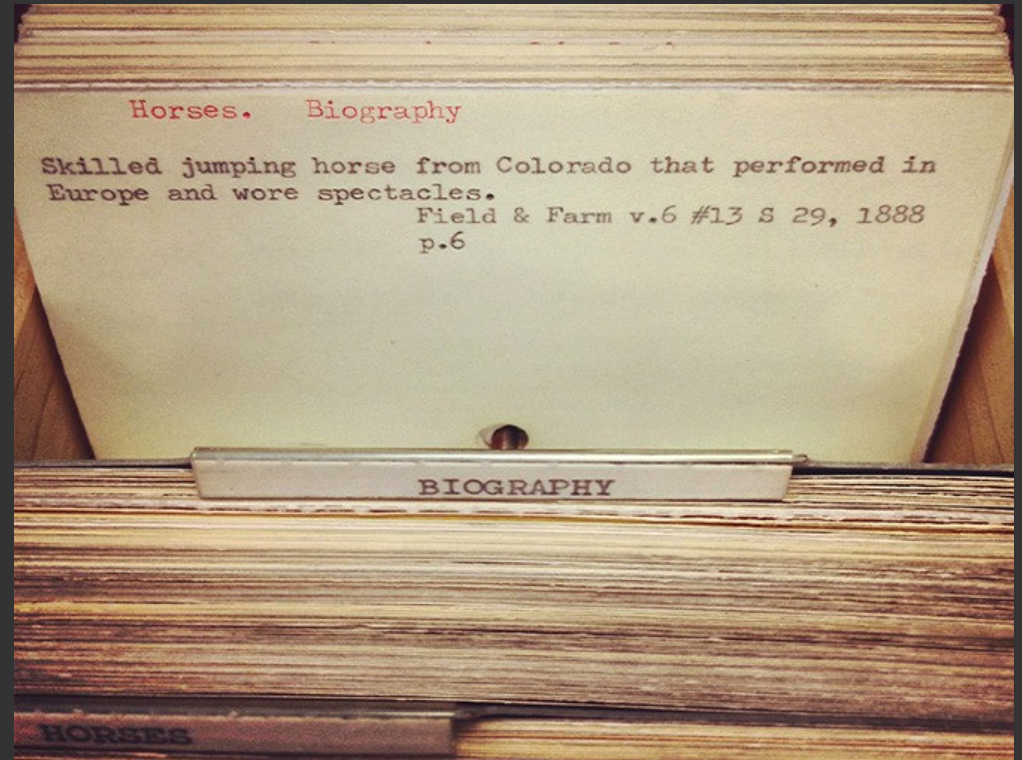
Binding pages: forming collections   systematizing {knowledge}

Abyssinian manuscript (ms 134, Goldschmidt no 21) Frankfurt University Library
Bibliotheque Nationale

Horses. Biography

Skilled jumping horse from Colorado that performed in Europe and wore spectacles.
Field & Farm v.6 #13 S 29, 1888
p.6

BIOGRAPHY

HORSES

Denver Public Library – Horses, biography.
John Allen, "Farewell Cards," *On Wisconsin Magazine*, Summer 2012.

Internal catalogs → federated catalogs → digital indexing {URIs}

accessing   {digital objects}

Project Gutenberg › 46,141 free ebooks › 5 by Charles Mackay

Memoirs of Extraordinary Popular Delusions and the Madness of Crowds by Mackay

No cover available

Download   **Bibrec**

Download This eBook

Format ?	Size	?	?	?
Read this book online: HTML	166 kB			
Read this book online: HTML	144 kB			
Read this book online: HTML	855 kB			
Read this book online: HTML	843 kB			
EPUB (with images)	4.2 MB	▣	▲	☁
EPUB (no images)	737 kB	▣	▲	☁
Kindle (with images)	9.4 MB	▣	▲	☁
Kindle (no images)	3.0 MB	▣	▲	☁
Plain Text UTF-8	1.6 MB			
More Files…				

MEMOIRS

OF

EXTRAORDINARY POPULAR DELUSIONS

AND THE

Madness of Crowds.

BY CHARLES MACKAY, LL.D.

AUTHOR OF "EGERIA," "THE SALAMANDRINE," ETC.
ILLUSTRATED WITH NUMEROUS ENGRAVINGS.

N'en déplaise à ces fous nommés sages de Grèce,
En ce monde il n'est point de parfaite sagesse;
Tous les hommes sont fous, et malgré tous leurs soins   BOILEAU.
Ne diffèrent entre eux que du plus ou du moins.

LONDON:
OFFICE OF THE NATIONAL ILLUSTRATED LIBRARY,
227 STRAND.
1852.

LIST OF ENGRAVINGS IN VOL. I.

Frontispiece—Gardens of the Hotel de Soissons. (From a print in Mr. Hawkins' collection.)
Vignette—The Bubblers' Arms, Prosperity. (*Bubblers' Mirror, or England's Folly.*)
John Law. (From a rare print by Leon Schenk. 1720)
The Regent D'Orleans
Old Palais Royal from the Garden. (From a scarce print, *circa* 1720)
Law's House; Rue de Quincampoix. (From Nodier's *Paris*)
Humpbacked Man hiring himself as a Table
Hôtel de Soissons. (From Nodier's *Paris*)
The Coach upset
Murder of a Broker by Count D'Horn
John Law as Atlas. (From *England under the House of Hanover*)
Caricature—Lucifer's new Row Barge
Procession of Miners for the Mississippi
The Chancellor D'Aguesseau
Caricature—Law in a Car drawn by Cocks
M. D'Argenson
Caricature—Neck or Nothing, or Downfall of the Mississippi Company
The South-Sea House. (From a print, *circa* 1750)
Harley Earl of Oxford
Sir Robert Walpole
Cornhill. (Print, *circa* 1720)
Stock-jobbing Card, or the Humours of Change Alley. 1720. (From the *Bubblers' Medley*)
Caricature—People climbing the Tree of Fortune. (From the *Bubblers' Medley*)
The Gateway to Merchant Tailors' Hall. (Gateway from old print)
Mr. Secretary Craggs
Caricature—Beggars on Horseback. (From the *Bubblers' Medley*)

CARICATURE. [22]

During the progress of this famous bubble, England presented a singular spectacle. The public mind was in a state of unwholesome fermentation. Men were no longer satisfied with the slow but sure profits of cautious industry. The hope of boundless wealth for the morrow made them heedless and extravagant for to-day. A luxury, till then unheard-of, was introduced, bringing in its train a corresponding laxity of morals. The over-bearing insolence of ignorant men, who had arisen to sudden wealth by successful gambling, made men of true gentility of mind and manners blush that gold should have power to raise the unworthy in the scale of society. The haughtiness of some of these "cyphering cits," as they were termed by Sir Richard Steele, was remembered against them in the day of their adversity. In the parliamentary inquiry, many of the directors suffered

69

Are footnotes connections or stops?

citing   {references}

* The affection of cats for marum and valerian is well known; they will beat the stems down, mat them with their feet, and then roll upon them.

† Charon has orders to deprive all cats of their claws. Whence comes that saying in hopeless cases—" he has as much chance as a cat in Hell without claws?"

Mf.

"A Squib, from Dr. Jortin's Tabby-Cat," The Gentleman's Magazine, (April 1792).

From references to graphs

* The affection of cats for marum and valerian is well known; they will beat the ftems down, mat them with their feet, and then roll upon them.

† Charon has orders to deprive all cats of their claws. Whence comes that faying in hopeles cafes—" he has as much chance as a cat in Hell without claws?" Mf.

CHARON http://en.wikipedia.org/wiki/Charon_(mythology)

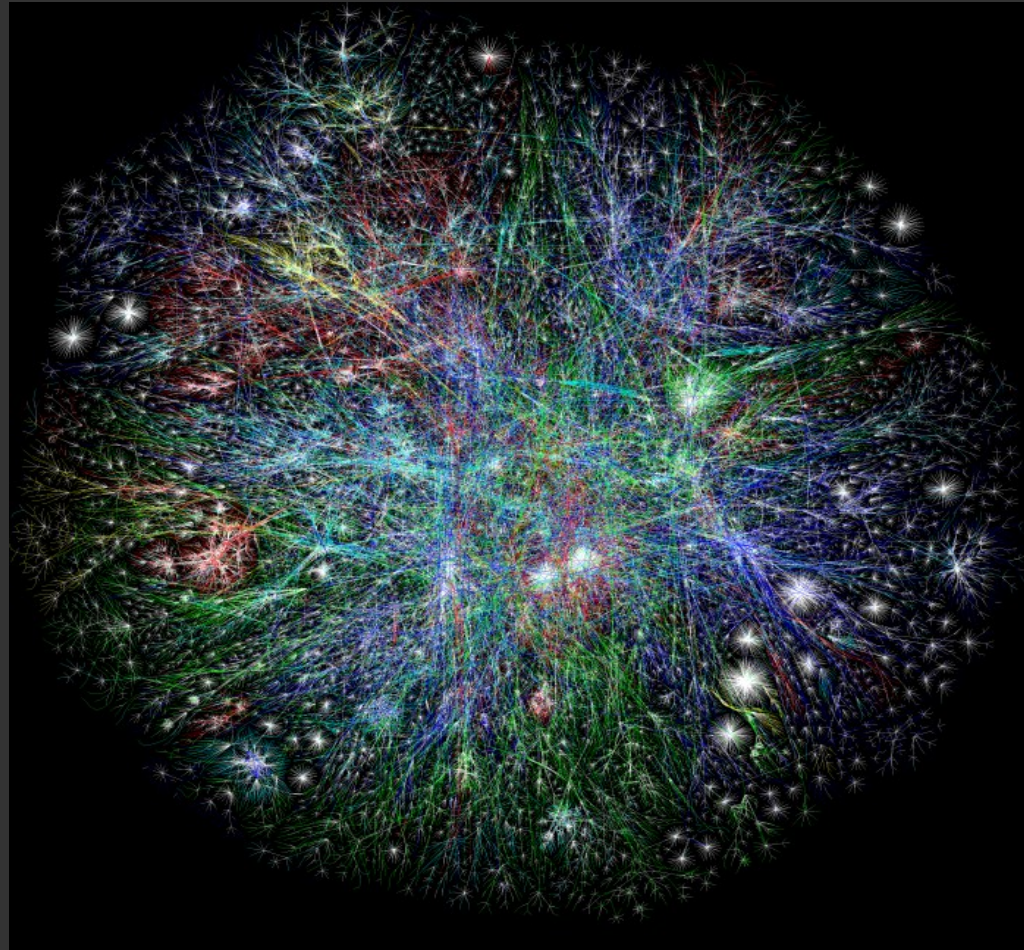http://en.wikipedia.org/wiki/Greek_mythology
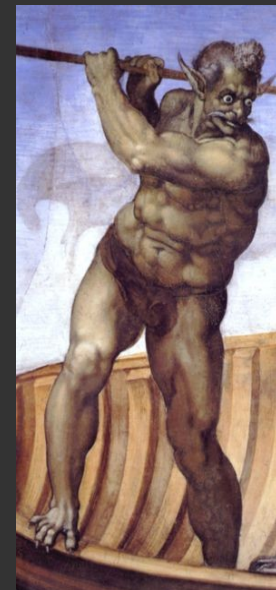
Rational structure
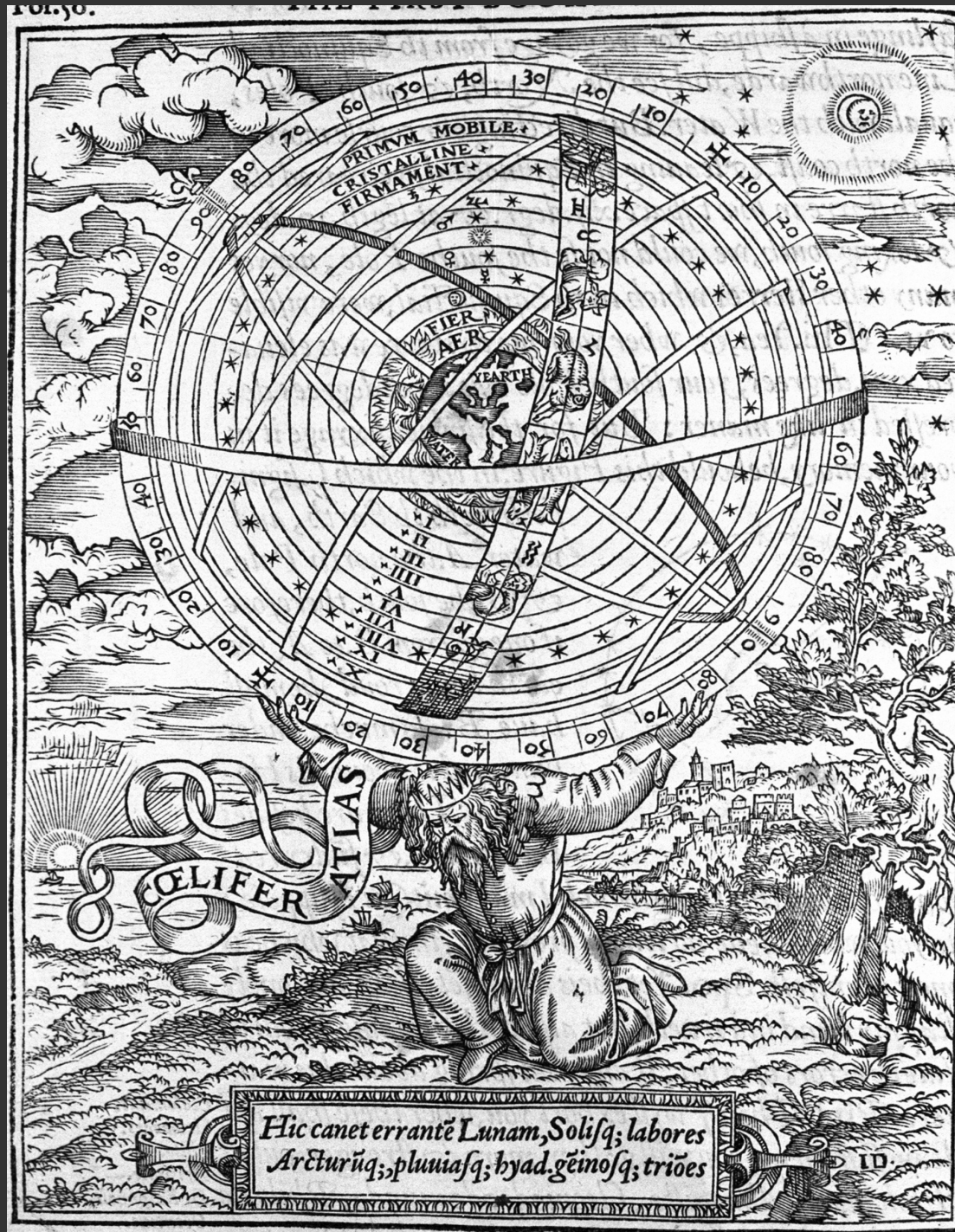
William Cunningham,
The Cosmographical Glasse
(John Day, 1559).

To a noisy interconnected universe of knowledge   {graph}

a tiny slice of the Internet (showing 5 million edges)

The center will not hold!
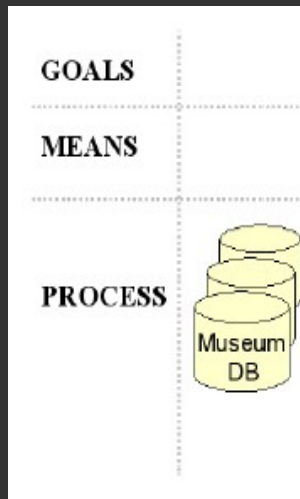
So what happens to the library?

Philology Library at the Free University of Berlin

On the road from DH 1.0 to DH 2.0

monolithic database projects
  - *custom data models*
  - *custom user interfaces*

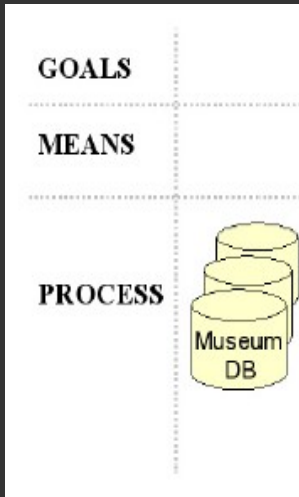Patrik Svensson,   *The Landscape of Digital Humanities*

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

open APIs
- *machine readable*
- *metadata and content*

monolithic database projects
- *custom data models*
- *custom user interfaces*

monolithic database projects
- *custom data models*
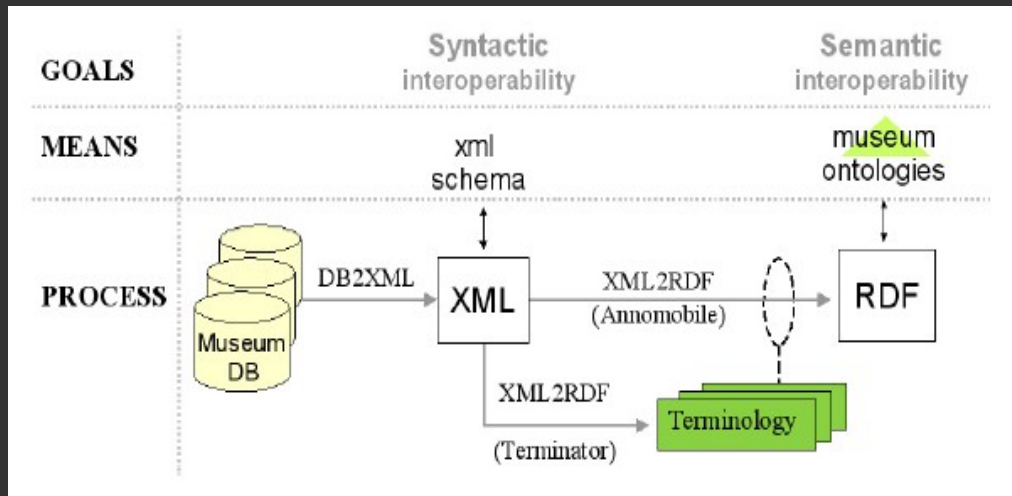- *custom user interfaces*

open APIs
- *machine readable*
- *metadata and content*

linked open data
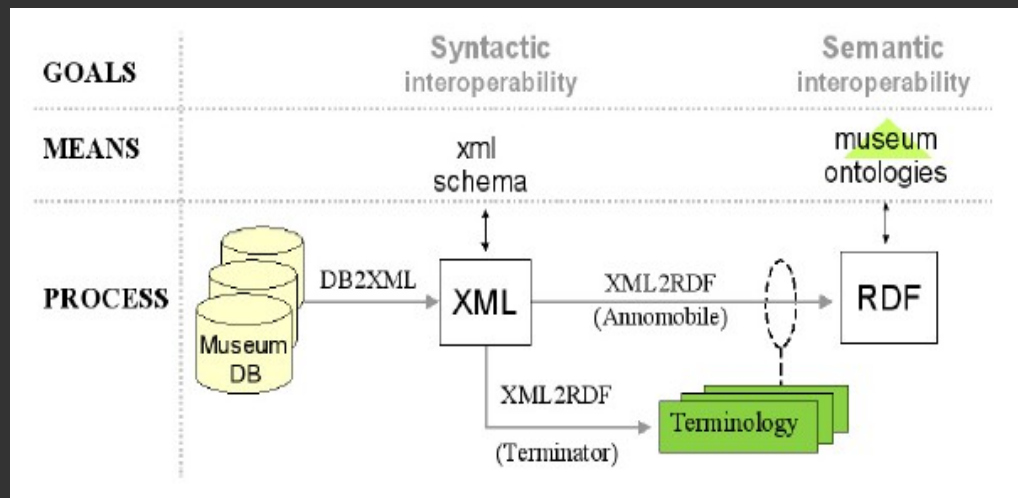- *machine readable*
- *disaggregated*

interoperability
- *ontologies*

monolithic database projects
- *custom data models*
- *custom user interfaces*

open APIs
- *machine readable*
- *metadata and content*

linked open data
- *machine readable*
- *disaggregated*

interoperability
- *ontologies*

open scholarship
- *digital publication*
- *annotation*
- *interpretation*
- *curation + narrative*
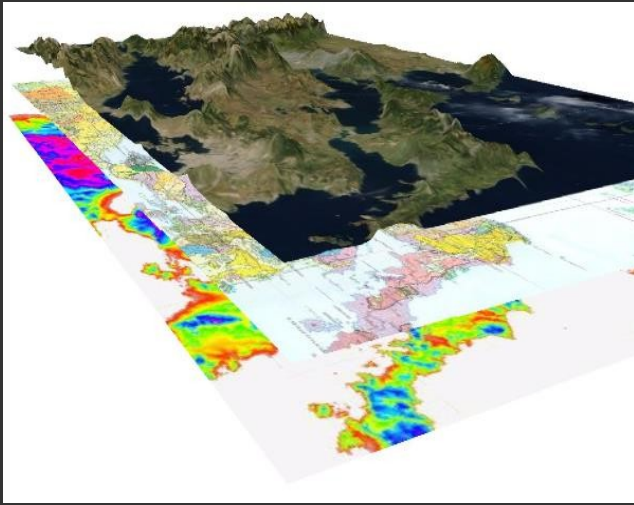- *collaborative*
- *open-ended*

Geographic data and the library
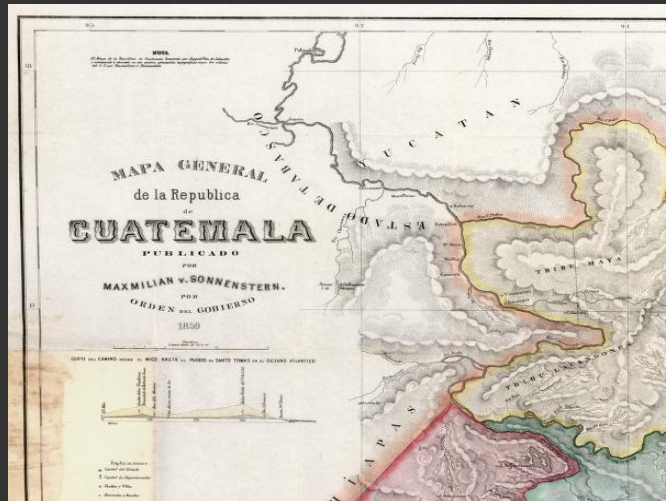
Geospatial datasets - GIS
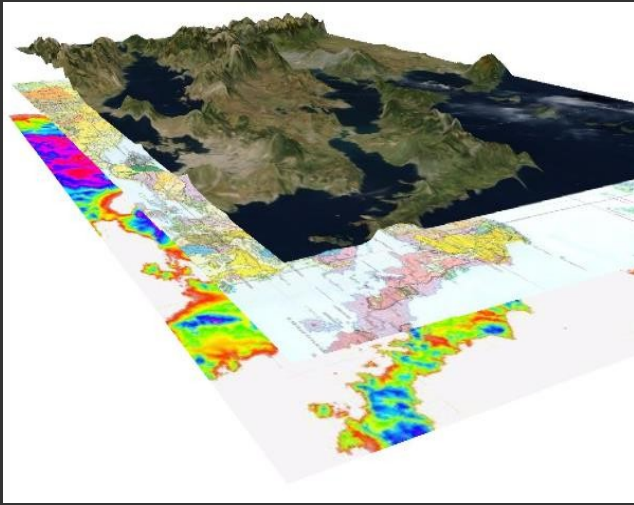
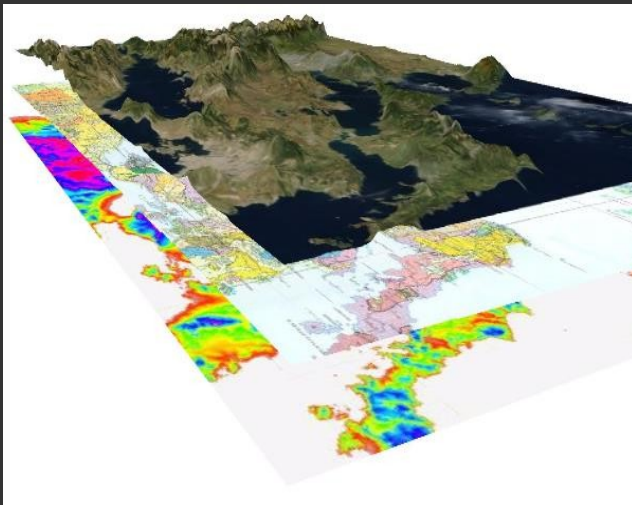# Geographic data and the library
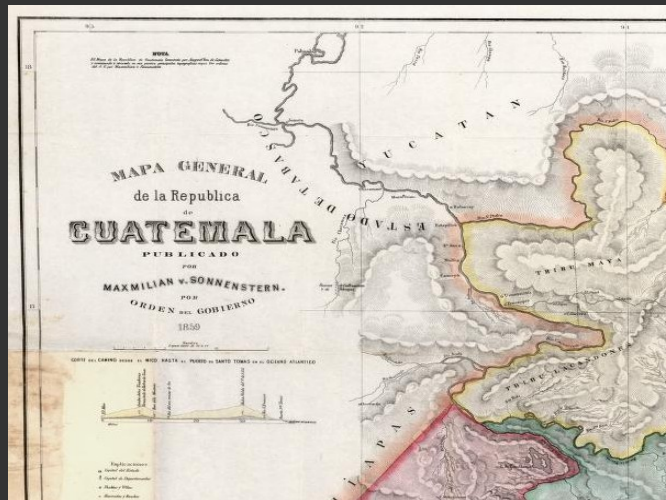
Geospatial datasets - GIS

Maps and scanned maps

Geographic data and the library

Geospatial datasets - GIS

The catalog as geodata

Maps and scanned maps

Geospatial datasets

Maps and scanned maps

OpenGeoportal

Federated search system

Digital Repository

The catalog as geodata

Figure 1-3. Example of a MARC 21 Record for a Book

Leader	01495cam 2200325 a 450
005	20041228210354.0
008	030122s2004 nyu c 000 1 eng
010 __	|a 2003000737
020 __	|a 0689856393
040 __	|a DLC |c DLC |d DLC |d TxGeoBT
042 __	|a lcac
050 00	|a PZ7.K1166 |b Ki 2004
082 00	|a [Fic] |2 22
100 1	|a Kadohata, Cynthia.
245 10	|a Kira-kira / |c Cynthia Kadohata
250	|a 1st ed.
260	|a New York : |b Atheneum Books for Young Readers, |c c2004.
300	|a 244 p. ; |c 19 cm.
520	|a Chronicles the close friendship between two Japanese-American sisters growing up in rural Georgia during the late 1950s and early 1960s, and the despair when one sister becomes terminally ill.
586	|a Newbery Award, 2004
650 1	|a Sisters |v Fiction.
650 1	|a Friendship |v Fiction.
650 1	|a Japanese Americans |v Fiction.
650 1	|a Death |v Fiction.
651 1	|a Georgia |x History |y 20th century |v Fiction.
856 42	|3 Contributor biographical information: |u http://www.loc.gov/catdir/bios/simon053/2003000737.html
856 42	|3 Publisher description: |u http://www.loc.gov/catdir/description/simon052/2003000737.html

651 - Subject Added Entry-Geographic Name

https://maps.googleapis.com/maps/api/geocode/json?address=Georgia

```
{
   "results" : [
      {
         "address_components" : [
            {
               "long_name" : "Georgia",
               "short_name" : "GA",
               "types" : [ "administrative_area_level_1", "political" ]
            },
            {
               "long_name" : "United States",
               "short_name" : "US",
               "types" : [ "country", "political" ]
            }
         ],
         "formatted_address" : "Georgia, USA",
         "geometry" : {
            "bounds" : {
               "northeast" : {
                  "lat" : 35.0006589,
                  "lng" : -80.8407866
               },
               "southwest" : {
                  "lat" : 30.3555908,
                  "lng" : -85.60516490000001
               }
            },
            "location" : {
               "lat" : 32.1656221,
               "lng" : -82.9000751
            },
```

Figure 1-3. Example of a MARC 21 Record for a Book

|	
Leader	01495cam 2200325 a 450
005	20041228210354.0
008	030122s2004 nyu c 000 1 eng
010 __	|a 2003000737
020 __	|a 0689856393
040 __	|a DLC |c DLC |d DLC |d TxGeoBT
042 __	|a lcac
050 00	|a PZ7.K1166 |b Ki 2004
082 00	|a [Fic] |2 22
100 1	|a Kadohata, Cynthia.
245 10	|a Kira-kira / |c Cynthia Kadohata
250	|a 1st ed.
260	|a New York : |b Atheneum Books for Young Readers, |c c2004.
300	|a 244 p. ; |c 19 cm.
520	|a Chronicles the close friendship between two Japanese-American sisters growing up in rural Georgia during the late 1950s and early 1960s, and the despair when one sister becomes terminally ill.
586	|a Newbery Award, 2004
650 1	|a Sisters |v Fiction.
650 1	|a Friendship |v Fiction.
650 1	|a Japanese Americans |v Fiction.
650 1	|a Death |v Fiction.
651 1	|a Georgia |x History |y 20th century |v Fiction.
856 42	|3 Contributor biographical information: |u http://www.loc.gov/catdir/bios/simon053/2003000737.html
856 42	|3 Publisher description: |u http://www.loc.gov/catdir/description/simon052/2003000737.html

245 – Title

260 – Place of Publication

651 - Geographic Name

Mapping options: georss → openlayers

serializing {feeds}
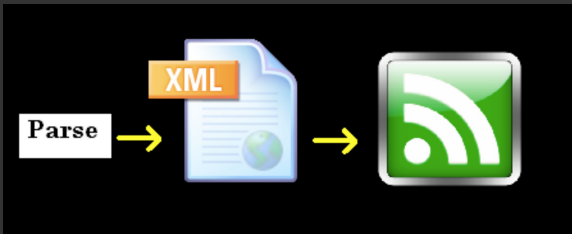

```
1  <?xml version="1.0" encoding="UTF-8"?>
2
3  <rss version="2.0"
   xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#"
   xmlns:dc="http://purl.org/dc/elements/1.1/">
4
5  <channel>
6  <title>Asian Newspapers in Harvard College Library</title>
7  <description></description>
8  <link>http://www.fas.harvard.edu/~chgis</link>
9  <dc:publisher>China Historical GIS</dc:publisher>
10 <pubDate>2007</pubDate>
11
12
13 <item>
14 <title>Air Force Academy press</title>
15 <pubDate>2008-02-01</pubDate>
16 <description><![CDATA[공사학보<br> language=Korean <br> place of
   publication=Ch&apos;ungbuk Cheongweon-gun <br> newspaper
   id=102]]></description>
17 <link>http://hollisclassic.harvard.edu/F/?func=find-
   b&amp;request=9452557&amp;find_code=SYS</link>
18 <geo:lat>36.75</geo:lat>
19 <geo:long>127.75</geo:long>
20 </item>
21 <item>
22 <title>Akahata</title>
23 <pubDate>2008-02-01</pubDate>
24 <description><![CDATA[赤旗<br> language=Japanese <br> place of
   publication=Tokyo <br> newspaper id=689]]></description>
25 <link>http://hollisclassic.harvard.edu/F/?func=find-
   b&amp;request=7773470&amp;find_code=SYS</link>
26 <geo:lat>35.68</geo:lat>
27 <geo:long>139.75</geo:long>
28 </item>
29 <item>
30 <title>Akahata</title>
31 <pubDate>2008-02-01</pubDate>
32 <description><![CDATA[赤旗縮刷版<br> language=Japanese <br> place
   of publication=Tokyo <br> newspaper id=111]]></description>
33 <link>http://hollisclassic.harvard.edu/F/?func=find-
   b&amp;request=8266710&amp;find_code=SYS</link>
34 <geo:lat>35.68</geo:lat>
```

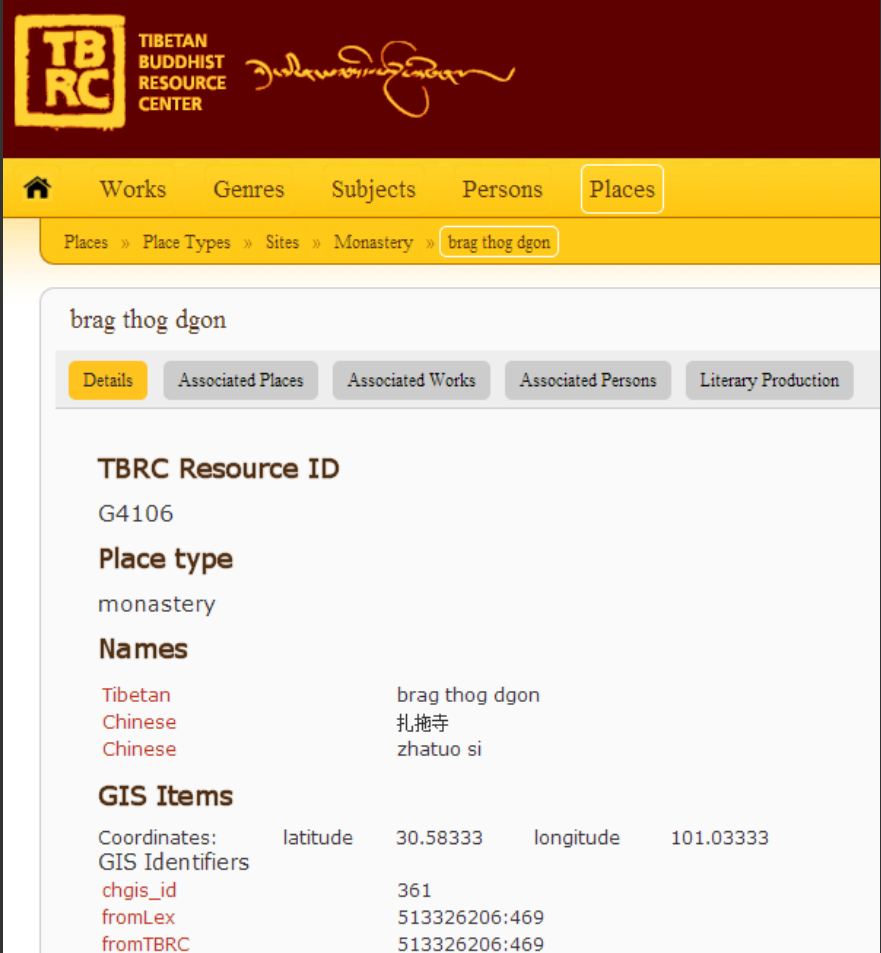**Asian Language Newspapers in Harvard Yenching Library**

http://goo.gl/UTpwyt

Mapping options: json → leaflet

```
{
    type: "FeatureCollection",
  ▾ features: [
      ▾ {
            type: "Feature",
            id: "tbrc_major_v1_twh.1",
          ▾ geometry: {
                type: "Point",
              ▾ coordinates: [
                    79.64,
                    33.41
                ]
            },
            geometry_name: "the_geom",
          ▾ properties: {
                TBRC_GID: "G2266",
                M_NM_WY: "lhun grub chos sdings dgon",
                M_NM_CH: "伦珠曲丁寺",
                LONG: 79.64,
                LAT: 33.41,
                LINK: "<a href='http://tbrc.org/#library_place_Object-G2266'>TBRC Details</a>"
            }
        },
```
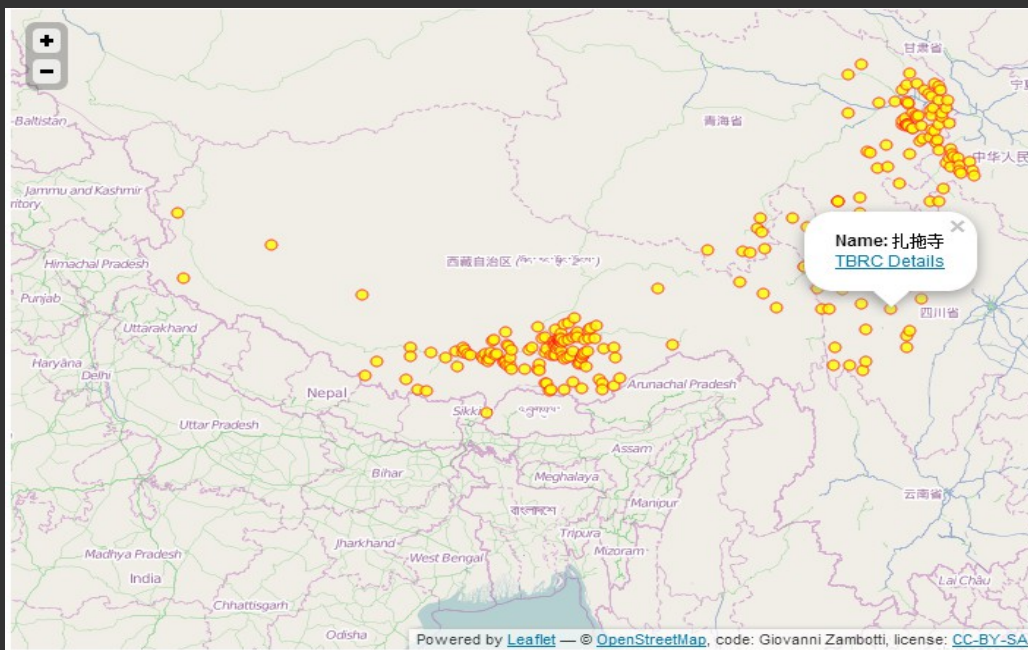
Serializing { json }

http://goo.gl/NGLfh6

Mapping options: customized from Metacarta geocoding product

Mapping options: geoHollis beta demonstration

http://goo.gl/jorMvG
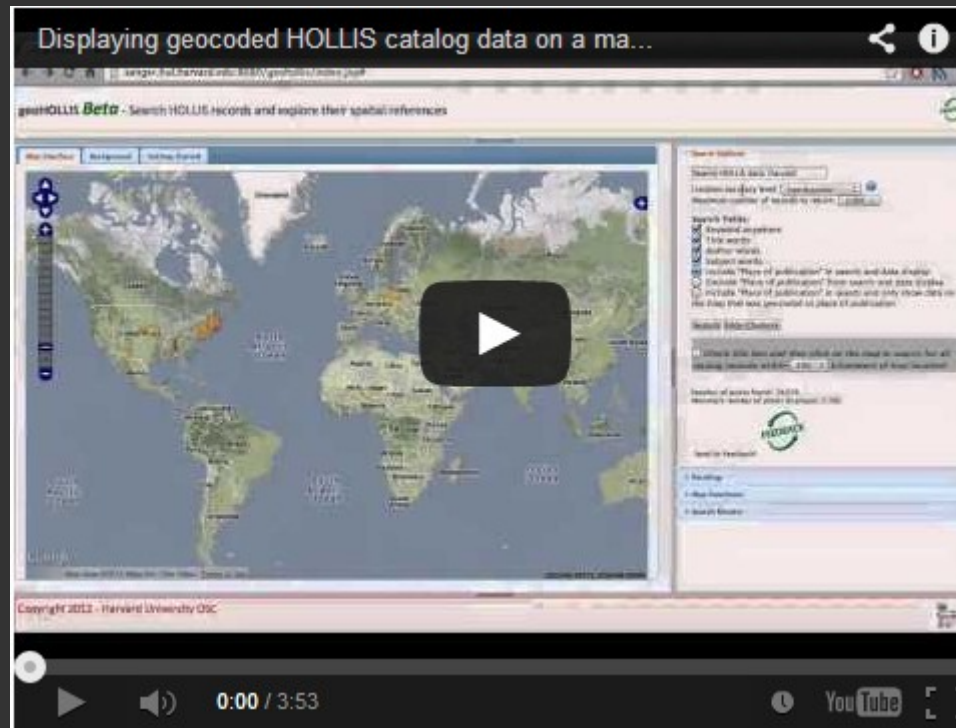
Issues raised: geoHollis beta

geocoding

- Identification of placenames in metadata varies in accuracy
- Ambiguous matches cannot be easily fixed
- Geocoding many types of information can be confusing
- No explicit way to match geocoded words to point on map

clustering

- Max number of search results greatly impacts the clusters
- The geocoded words don't explain how the cluster was decided

facets

- adding and removing facets is essential
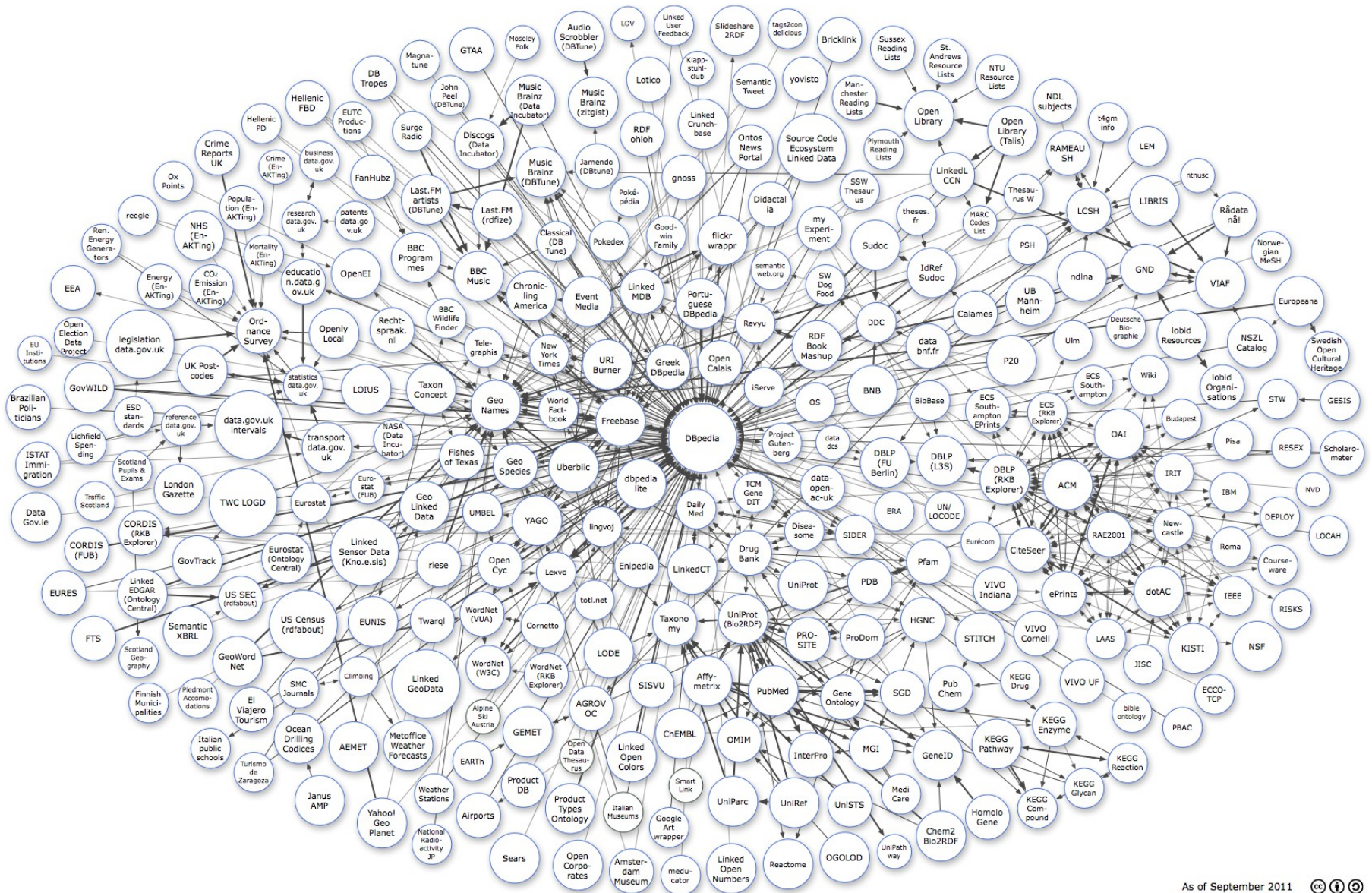
Richard Cyganiak, *Linked Open Data Cloud* (2011).

Linked Open Data Cloud

geographic sources

by Richard Cyganiak

http://richard.cyganiak.de/2007/10/lod/

As of September 2011

Locations in Classical Texts and Atlases
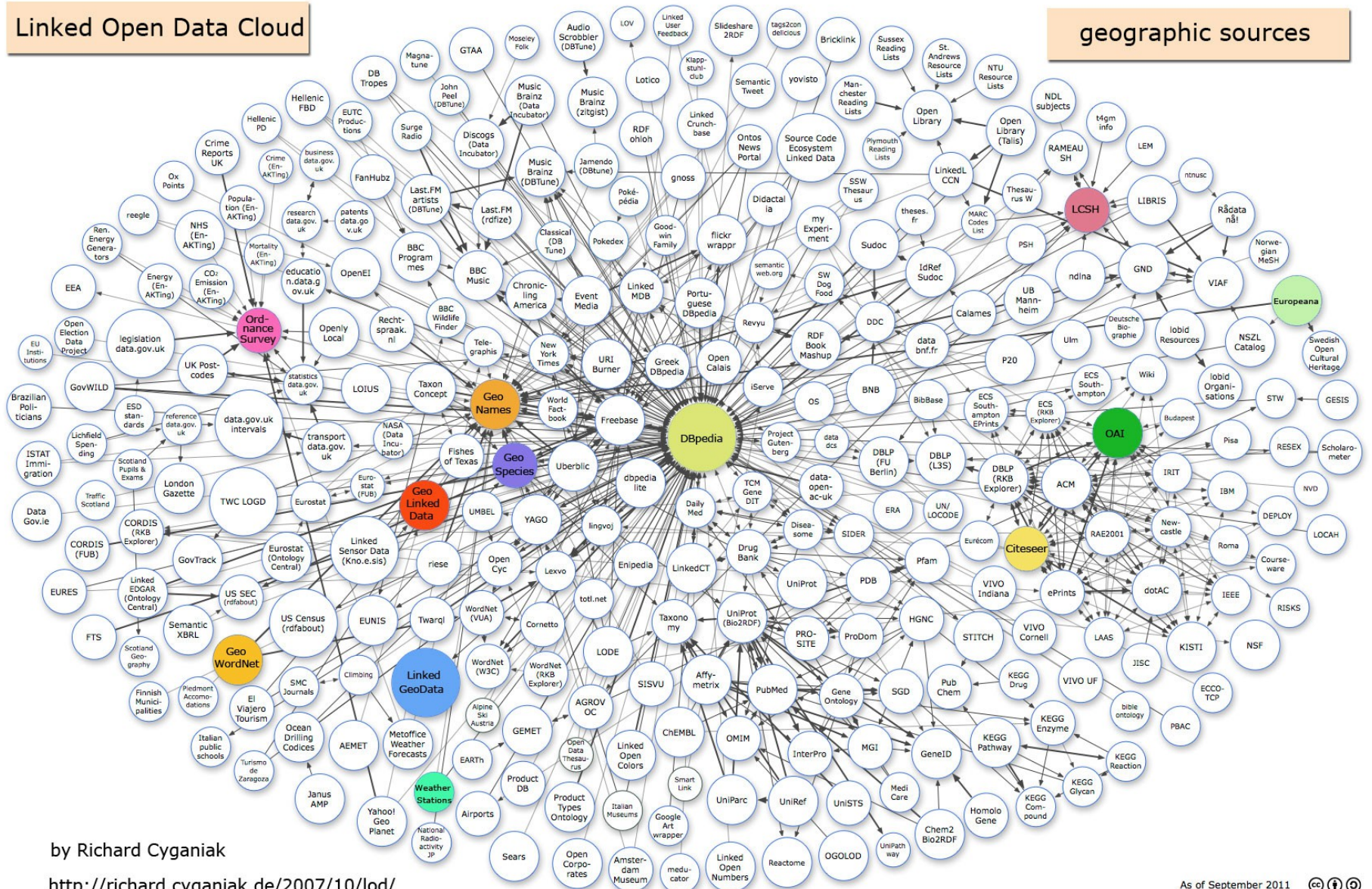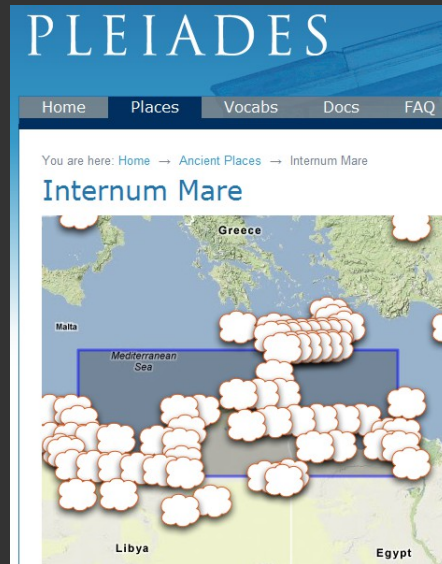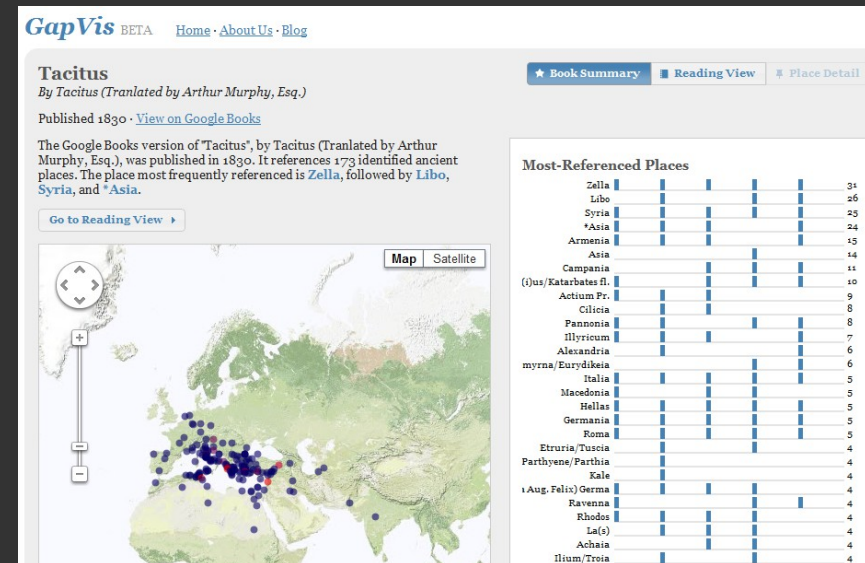

Perseus Project

Pleiades

Google Ancient Places

Pelagios - linking historical objects to places - using Pleiades

Mycenae

An ancient place, cited: BAtlas 58 D2 Mycenae
http://pleiades.stoa.org/places/570491

Alternative Names: Mycenae
Coverage: Mykenes

[JSON] [RDF/XML] [Turtle]

770 References to Mycenae

Dataset	No. of References
Arachne (in 2 subsets)	133 (of 128180 total in set)
SquinchPix	164 (of 10782 total in set)
CLAROS (in 1 subsets)	140 (of 204474 total in set)
Regnum Francorum Online (in 1 subsets)	1 (of 16427 total in set)
Pleiades Annotations in the Perseus Digital Library (in 2 subsets)	126 (of 61738 total in set)
SAWS - Sharing Ancient Wisdoms	1 (of 723 total in set)
AWMC - Ancient World Mapping Center	1 (of 34608 total in set)
Ure Museum	133 (of 2344 total in set)
Google Ancient Places (in 3 subsets)	20 (of 41053 total in set)
Omnes Viae	1 (of 2983 total in set)
Ancient History Encyclopedia	50 (of 1098 total in set)

[JSON] [RDF/XML] [Turtle]

Pleiades

Pelagios API   http://pelagios.dme.ait.ac.at/api

Documentation http://pelagios-project.blogspot.com/

Pelagios - linking historical objects to places - using CHGIS

Mycenae

An ancient place, cited: BAtlas 58 D2 Mycenae
http://pleiades.stoa.org/places/570491

Alternative Names: Mycenae
Coverage: Mykenes

JSON | RDF/XML | Turtle

770 References to Mycenae

Dataset	No. of References
Arachne (in 2 subsets)	**133** (of 128180 total in set)
SquinchPix	**164** (of 10782 total in set)
CLAROS (in 1 subsets)	**140** (of 204474 total in set)
Regnum Francorum Online (in 1 subsets)	**1** (of 16427 total in set)
Pleiades Annotations in the Perseus Digital Library (in 2 subsets)	**126** (of 61738 total in set)
SAWS - Sharing Ancient Wisdoms	**1** (of 723 total in set)
AWMC - Ancient World Mapping Center	**1** (of 34608 total in set)
Ure Museum	**133** (of 2344 total in set)
Google Ancient Places (in 3 subsets)	**20** (of 41053 total in set)
Omnes Viae	**1** (of 2983 total in set)
Ancient History Encyclopedia	**50** (of 1098 total in set)

JSON | RDF/XML | Turtle

Pleiades

CHGIS

Pelagios API   http://pelagios.dme.ait.ac.at/api

Documentation http://pelagios-project.blogspot.com/

XML web service for Chinese historical placenames:

```xml
<?xml version="1.0" encoding="UTF-8"?>
<query>
    <resultsDescription>CHGIS database query for Placename/Year: 方
广/1820</resultsDescription>
    <resultsTotal>3</resultsTotal>
    <resultsTransmitted>3</resultsTransmitted>
    <queryExecutionTimeSeconds>0.66</queryExecutionTimeSeconds>
    <queryDate>Fri, 27 Sep 2013 00:24:17 -0400</queryDate>
    <publisher>China Historical GIS</publisher>
    <publisherURL>http://www.fas.harvard.edu/~chgis/</publisherURL>

  <item id="relig_1714">
      <placename>
          <name_romanized>Fangguang Si</name_romanized>
          <name_vernacular>方广寺</name_vernacular>
          <name_alternate>方廣寺</name_alternate>
      </placename>
      <feature_type>
          <type_english>buddhist temple</type_english>
          <type_romanized>fo si</type_romanized>
          <type_vernacular>佛寺</type_vernacular>
          <type_id>113</type_id>
      </feature_type>
      <temporal>
          <begin_year>1820</begin_year>
          <begin_year_rule>9</begin_year_rule>
          <end_year>1820</end_year>
          <end_year_rule>9</end_year_rule>
      </temporal>
      <spatial>
          <object_type>point</object_type>
          <coordinate_type>point location</coordinate_type>
          <degrees_latitude>27.87</degrees_latitude>
          <latitude_direction>N</latitude_direction>
          <degrees_longitude>112.91</degrees_longitude>
          <longitude_direction>E</longitude_direction>
          <present_location>湖南湘潭市</present_location>
      </spatial>
```
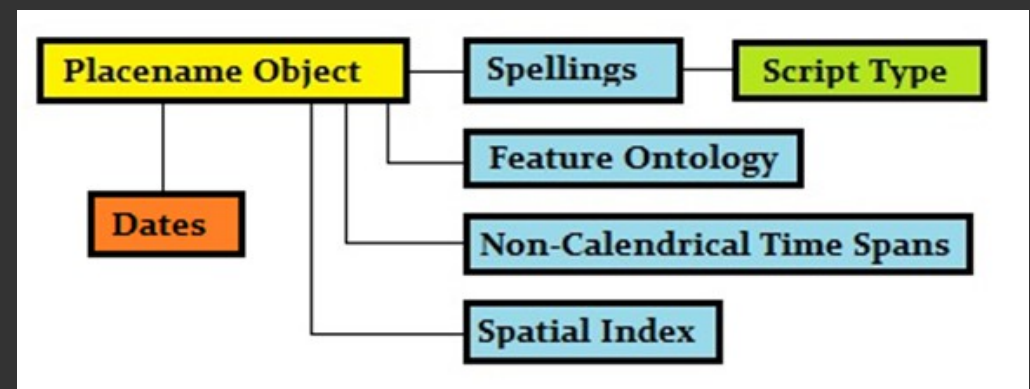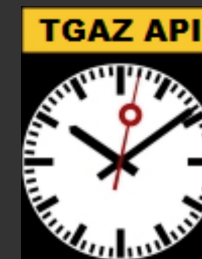
Faceted search web service for historical placenames

CHGIS Gazetteer Web Service


```xml
<item id="32084">
    <placename>
        <name_romanized>Fu Zhou</name_romanized>
        <name_vernacular>福州</name_vernacular>
        <name_alternate>福州</name_alternate>
    </placename>
    <feature_type>
        <type_english>prefecture</type_english>
        <type_romanized>Zhou</type_romanized>
        <type_vernacular>州</type_vernacular>
        <type_id>20</type_id>
    </feature_type>
</item>
```
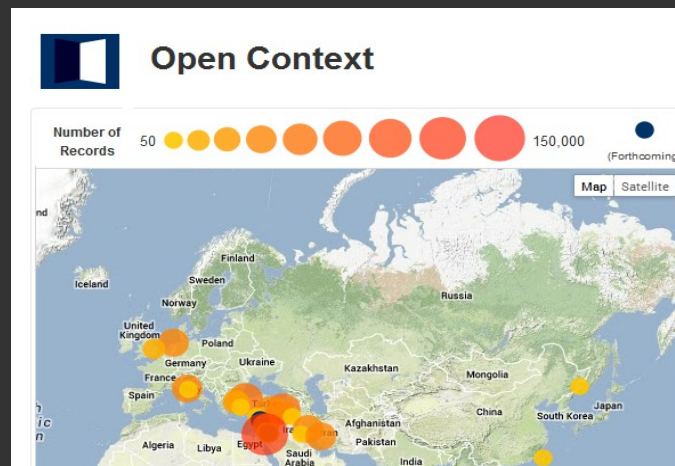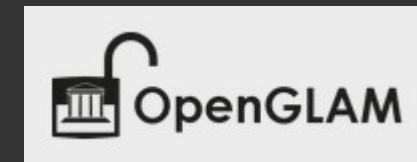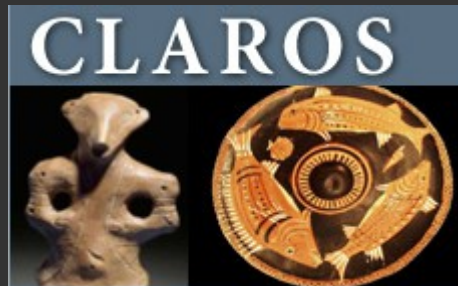
http://chgis.hmdc.harvard.edu/xml/

Temporal Gazetteer Web Service

http://chgis.hmdc.harvard.edu/tgaz/api

RDF gazetteer interchange format


```
← → C  🗋 chgis.hmdc.harvard.edu/placename/rdf/hvd_90144

@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix osgeo: <http://data.ordnancesurvey.co.uk/ontology/geometry/> .
@prefix pelagios: <http://pelagios.github.io/vocab/terms#> .
@prefix pleiades: <http://pleiades.stoa.org/places/vocab#> .
@prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix spatial: <http://geovocab.org/spatial#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix gn: <http://www.geonames.org/ontology#> .
<http://chgis.hmdc.harvard.edu/placename/hvd_90144> a pelagios:PlaceRecord ;
  dcterms:title "Fu Zhou" ;
  dcterms:description "prefecture 州" ;
  dcterms:temporal "start=725; end=740" ;
  gn:countryCode "cn" ;
  pleiades:hasName [ rdfs:label "福州" ] ;
  pleiades:hasName [ rdfs:label "福州" ] ;
  pleiades:hasName [ rdfs:label "Fu Zhou" ] ;
  pleiades:hasLocation [ geo:lat "0.00000"^^xsd:double ; geo:long "0.00000"^^xsd:double ] ;
  .
```

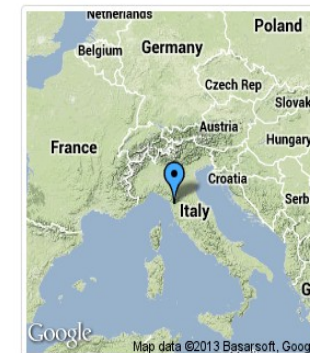# Potential connections: LOD projects in DH world

CLAROS

CHARISMA

OpenGLAM

Pelagios

Open Context

Number of Records: 50 ... 150,000 (Forthcoming)

europeana
think culture

Luca

An ancient place, cited: BAtlas 41 C2 Luca
http://pleiades.stoa.org/places/403234

Alternative Names: Luca, Lucca
Coverage: Lucca

JSON | RDF/XML | Turtle

18 References to Luca

Dataset	No. of References
Regnum Francorum Online (in 1 subsets)	1 (of 16427 total in set)
Pleiades Annotations in the Perseus Digital Library (in 1 subsets)	5 (of 61738 total in set)
AWMC - Ancient World Mapping Center	2 (of 34608 total in set)
Google Ancient Places (in 4 subsets)	8 (of 41053 total in set)
Omnes Viae	1 (of 2983 total in set)
ORBIS: The Stanford Geospatial Network Model of the Roman World	1 (of 402 total in set)

Implications for the future

Transformation of the catalog:

- metadata describing resources → direct access to digital resources

- users expect instant, machine-readable information

- human readable forms & facets → APIs

- users expect both discovery and deeper analysis of content

- aggregated data and visualizations of data rather than lists

- linked open data and the semantic web will encroach into library science

- graph databases and structures will get applied to search tools

For Geographic data:

- search and visualization using maps will be more commonplace

- catalog metadata will be geocoded (internally or on-the-fly)

Map to catalog
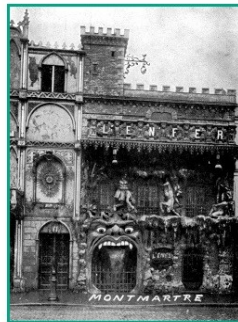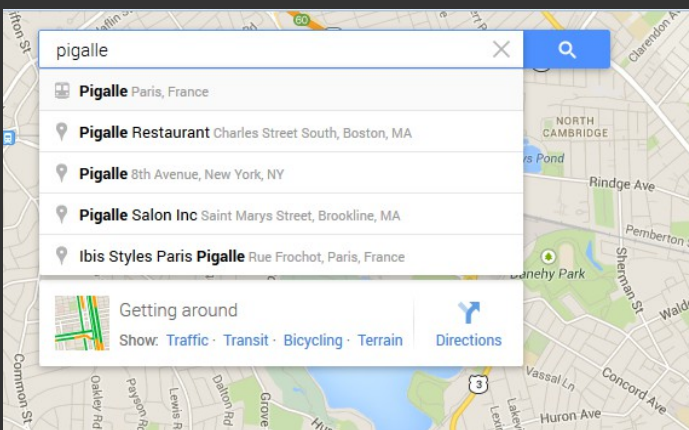
spatial indexing {metadata}

Unbound and linked

The Sentiment of the World Throughout History Through Wikipedia

1860

1864

http://goo.gl/VUWsvO
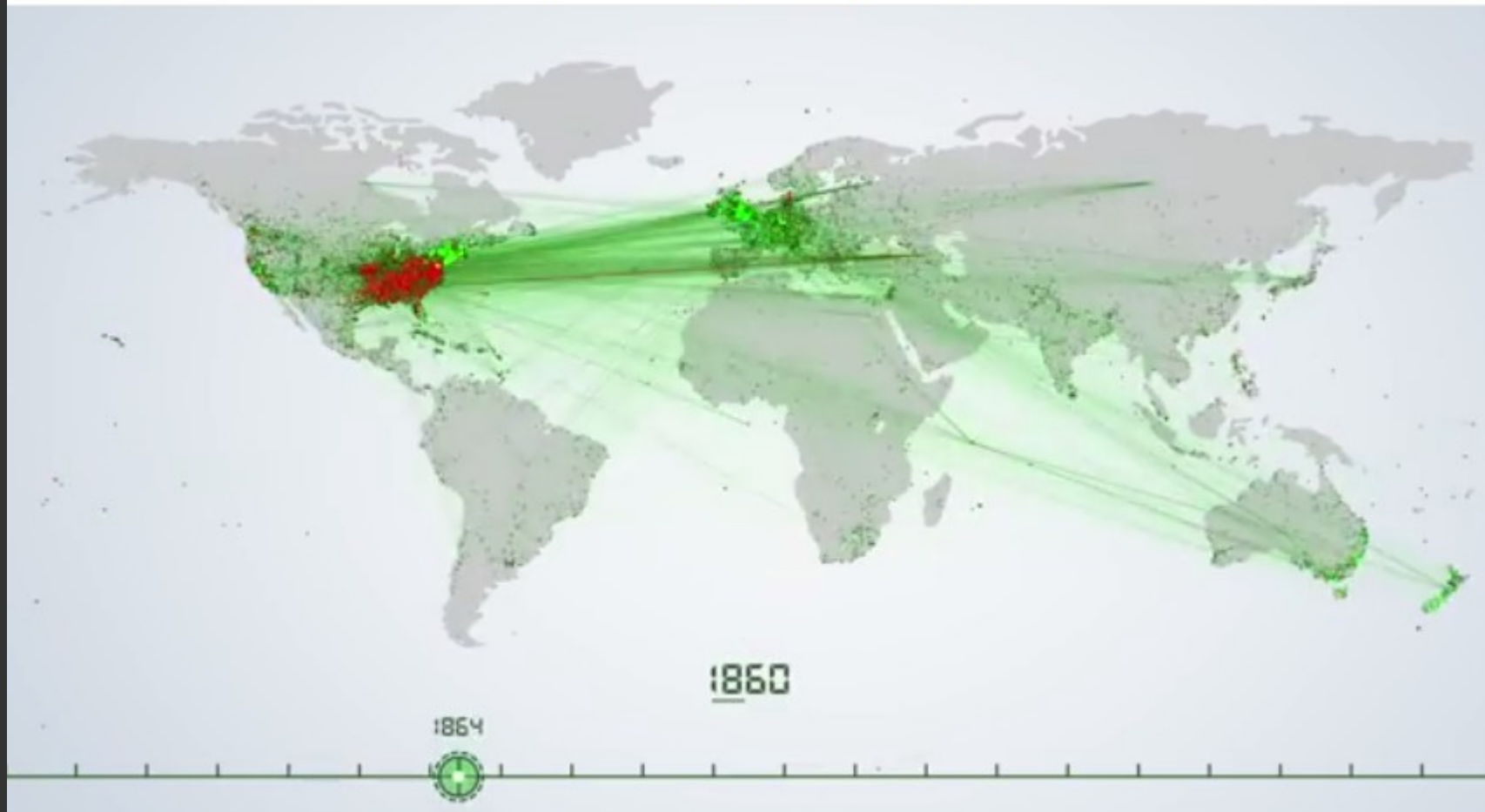
Kalev Leetaru, "Fulltext Geocoding Versus Spatial Metadata for Large Text Archives: Towards a Geographically Enriched Wikipedia" *D-Lib Magazine*, (Sep-Oct 2012)

Interesting examples using spatial search:

- New York Public Library – Maps Division

- Ed Summers   ici webmap for wikipedia

- Europeana.eu

- Spatial History Project - Stanford

- Spatial Humantities – Scholar's Lab - UVA

- MapRankSearch – Klokan Tech

- Digital Silk Road – Stein Placename Database

- GeoCLEF 2008

- Biodiversity of the Hengduan Mountains

Temporal Gazetteers and Linked Geodata

http://www.fas.harvard.edu/~chgis/gazetteer

Bibliography on big data and spatial humanities

https://www.zotero.org/groups/dh_and_big_data/items