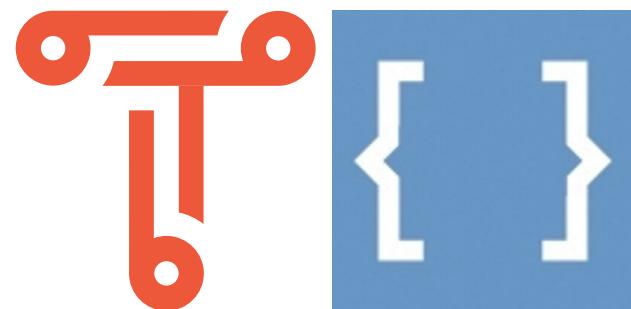


NLGIS: A USE CASE IN LINKED HISTORIC GEODATA

Wouter Beek wouter@triply.cc, VU University Amsterdam, Triply

Richard Zijdeman richard.zijdeman@iisg.nl, International Institute for Social History

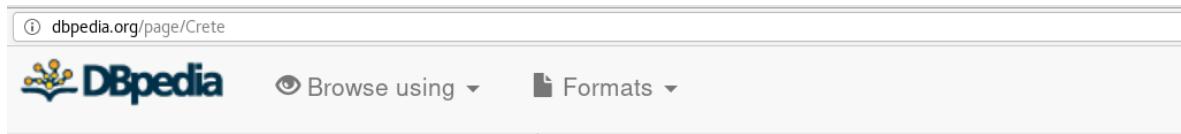
June 3rd, 2018



PROBLEM STATEMENT

Example of an artifact with poor geodata: [link](#)

Geodata in today's LOD Cloud...



The screenshot shows the DBpedia page for the entity "Crete". The top navigation bar includes links for "dbpedia.org/page/Crete", "Browse using", and "Formats". The main content area displays a list of geodata properties and their values:

- [geodata:Crete](#)
- [geodata:Crete](#)
- [freebase:Crete](#)
- <http://sw.cyc.com/concept/Mx4rvViCppwpEbGdrcN5Y29ycA>
- [yago-res:Crete](#)
- [yago-res:Crete](#)
- <http://d-nb.info/gnd/4073791-3>

Below this, specific coordinate values are listed:

geo:geometry	▪ POINT(24.909999847412 35.209999084473)
geo:lat	▪ 35.209999 (xsd:float)
geo:long	▪ 24.910000 (xsd:float)
skos:exactMatch	▪ http://rdfdata.eionet.europa.eu/ramon/nuts/EL43

NLGIS DATASETS

<https://druid.datalegend.net/nlgis>

Dataset	No statements	Main concepts	No geometries	Timeframe
CShapes	6,120	countries, cities	510	1920-present
Mint Authorities	6,987	authorities, houses	950	565-present
Gemeentegeschiedenis	46,929	municipalities, provinces	3,219	1813-present
nLGis	60,036	features, geometries	4,679	

LESSONS LEARNED

1. Combine what belongs together
2. Do not use ambiguous ‘null’ values
3. No perfect tool for data transformation
4. No perfect triple store for geo
5. Direct feedback helps a lot
6. Use interoperable representations

[1] COMBINE WHAT BELONGS TOGETHER

DATE/TIME

```
:Greece iisg:cowStartYear "1946"^^xsd:gYear ;  
        iisg:cowStartMonth "1"^^xsd:gMonth ;  
        iisg:cowStartDay "1"^^xsd:gDay .
```

```
:Greece iisg:cowStart "1946-01-01"^^xsd:date
```

(In CShapes, 'cow' stands for Correlates of War.)

COMBINING WHAT BELONGS TOGETHER PREVENTS BUGS

Longitude/latitude

```
:somewhere wgs84:lat "..."  
          wgs84:lat "..." ;  
          wgs84:long "..." ;  
          wgs84:long "..." .
```

There is also `wgs84:lat_long`, but it is almost never used.

THERE ARE MANY INSTANCES OF THIS!

OCLC VIAF [example](#)

```
:EmmaGoldman schema:givenName "Ema" ;  
             schema:givenName "Emma" ;  
             schema:familyName "Gol'dman" ;  
             schema:familyName "Gōrudoman" .
```

There are many other instances of this problem, e.g.,
`foaf:firstName` and `foaf:lastName`.

[2] DO NOT USE AMBIGUOUS ‘NULL’ VALUES

CShapes uses -1 to denote an unknown year.

In the context of CShapes (countries after 1920) this makes sense.

But on the web we can query CShapes and Pleiades.

[3] NO PERFECT TOOL FOR DATA TRANSFORMATION

Requirements:

1. Support multiple source formats
2. Scale to datasets of arbitrary size

No currently available data transformation tool implements these two core requirements.

SUPPORT MULTIPLE SOURCE FORMATS

- CSV
- (Geo)JSON
- XML
(GML/MARCXML/EAD)
- relational DB
- RDF

Proprietary formats can sometimes be transformed into open formats, e.g., ESRI ShapeFile.

SCALE TO DATASETS OF ARBITRARY SIZE

Be able to stream through the data at the required granularity level.

```
# , name      , population , shape
1 , Amsterdam , 1.3M      , MultiPolygon(..)
2 , Athens     , 3.1M      , MultiPolygon(..)
...
```

[4] NO PERFECT TRIPLE STORE FOR GEO

GeoSPARQL support is either absent, not standards-compliant, or not performant.

- Most stores do not implement GeoSPARQL syntax, but some do.
- Most stores have miserable/unusable performance, but some have good performance.
- Some stores change the data merely by loading it.
- Some stores cannot load larger shapes.
- Commercial stores are not necessarily better than FOSS (if fact: they are very often worse).

[5] DIRECT FEEDBACK HELPS A LOT

When writing GeoSPARQL queries, a table of results is not enough.

PDOK Data Platform

Query Another Query My Favorite Query +

```
1 v prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
2 v prefix top: <http://brt.basisregistraties.overheid.nl/def/top10nl#>
3 v prefix geo: <http://www.opengis.net/ont/geosparql#>
4 v select ?feature ?wktLiteral {
5   ?feature rdf:type top:Heemtuin .
6   ?feature geo:hasGeometry ?geometry .
7   ?geometry geo:asWKT ?wktLiteral .
8 }
9 limit 100
10
```

Tip: Add a label variable prefixed with the geo variable name to show popups on the map. E.g. `wktLiteralLabel`

[6] USE INTEROPERABLE REPRESENTATIONS

OPTIONS FOR REPRESENTING GEODATA IN LOD

- WGS84 Geo Positioning Vocabulary (W3C)
- GeoSPARQL (OGC)
 - Well-Known Text (WKT)
 - Geography Markup Language (GML)
- Make up your own vocabulary
- GeoJSON + JSON-LD

PLEIADES

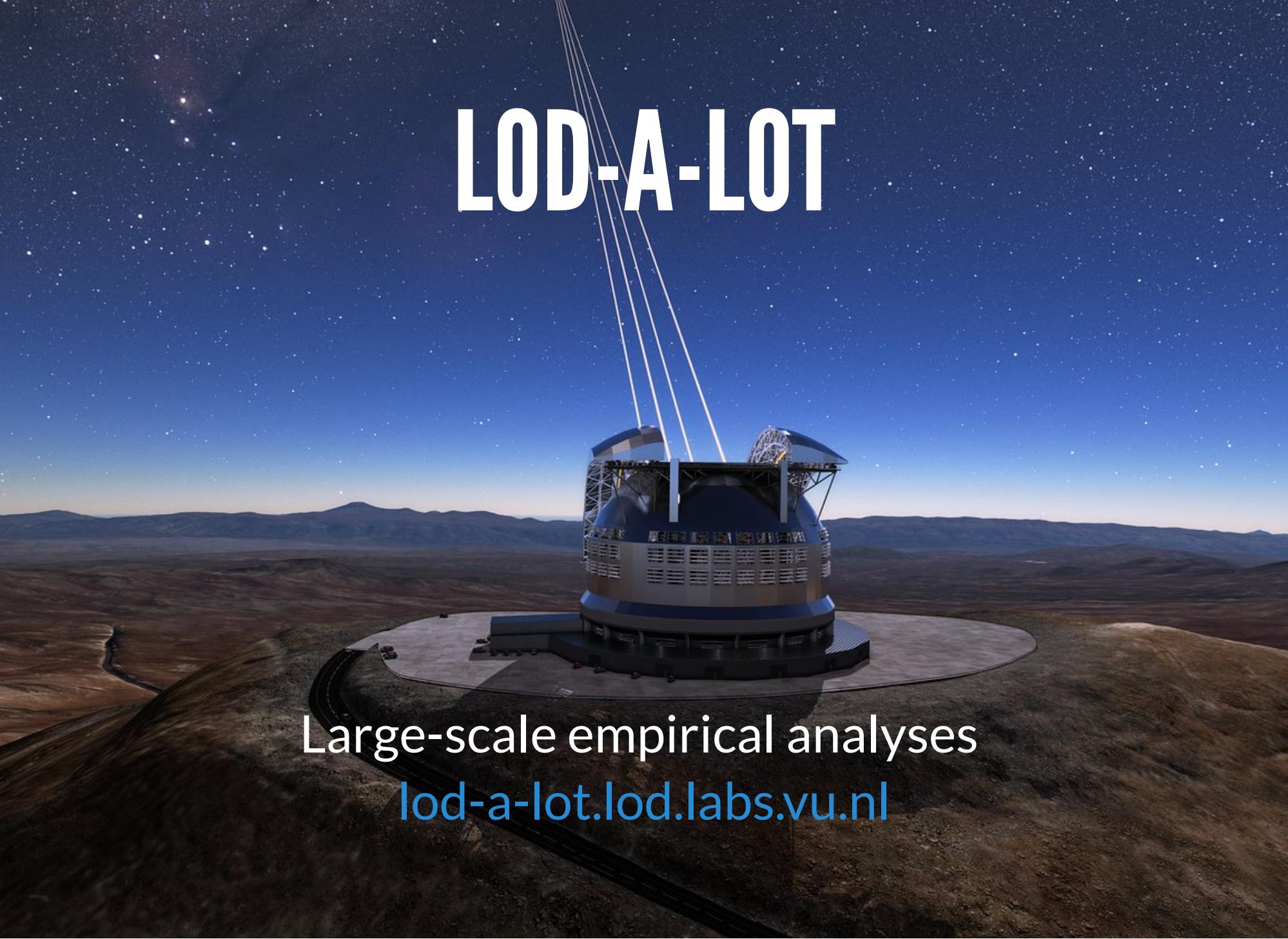
```
prefix geo: <http://data.ordnancesurvey.co.uk  
place:Athens a lawd:Place ;  
    geo:hasGeometry [  
        geo:asWKT "LineString(5.16 52.05, ...)" ;  
    ] ;
```

Without interoperable representations:

- clients do not know what to do with your data
- triple store cannot index your geometries
- reasoners arrive at contradictions

After applying standards...

LOD-A-LOT

A large telescope is mounted on a circular concrete platform on a hillside. The telescope has a large blue and silver dome. Four white lines extend from the top of the dome towards the upper left corner of the image, suggesting a beam splitter or light path. The background is a dark blue night sky filled with stars. In the distance, there are mountains and a winding road on the hillside.

Large-scale empirical analyses
lod-a-lot.lod.labs.vu.nl

LINKED GEODATA VOCABULARY USE

Property	Nº statements	Nº documents
wgs84:alt	2,349,607	9,843
wgs84:lat	42,883,363	11,134
wgs84:lat_long	283	173
wgs84:location	14,688,561	117
wgs84:long	42,916,785	11,134
geo:asGML	0	1
geo:asWKT	188,427,329	50
geo:hasGeometry	28,366,268	7

Based on the LOD-a-lot data collection (Fernández et al. 2017).

GEOJSON + JSON-LD

Unfortunately, these two popular formats are incompatible:

- GeoJSON uses square brackets to denote (nested) lists of geographic coordinates.
- JSON-LD uses square brackets as summarized syntax for repeated property assertion.

This may be fixed in future a version of the JSON-LD standard.

NLGIS DATASETS

<https://druid.datalegend.net/nlgis>

Dataset	No statements	Main concepts	No geometries	Timeframe
CShapes	6,120	countries, cities	510	1920-present
Mint Authorities	6,987	authorities, houses	950	565-present
Gemeentegeschiedenis	46,929	municipalities, provinces	3,219	1813-present
nLGis	60,036	features, geometries	4,679	

Dutch Cultural Heritage institutes already use this to
annotate their collection with ([example](#)).

FUTURE WORK

- Cover more places & times.
- Annotate Cultural Heritage objects with detailed geographic information.
- Create a standardized vocabulary for how geolocations change through time.
- Improve GeoSPARQL support in triple stores.
- Explore new ways of displaying Cultural Heritage objects in space and time ([example](#)).

THANK YOU!

nlGis: <https://druid.datalegend.net/nlgis/>

Wouter Beek wouter@triply.cc, VU University Amsterdam, Triply

Richard Zijdeman richard.zijdeman@iisg.nl, International Institute for Social History

June 3rd, 2018