Linked Open Data

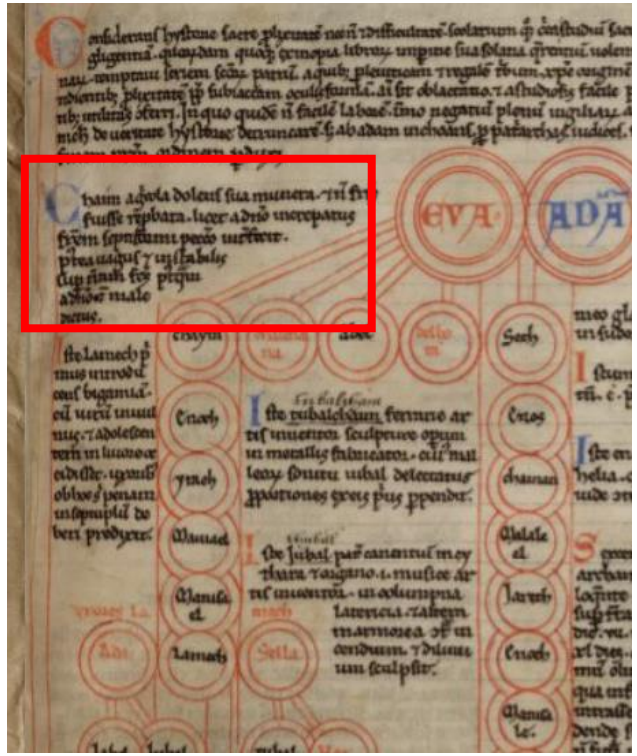
Master Class ‚Digital Scholarly Editing' 2024, Saarbrücken

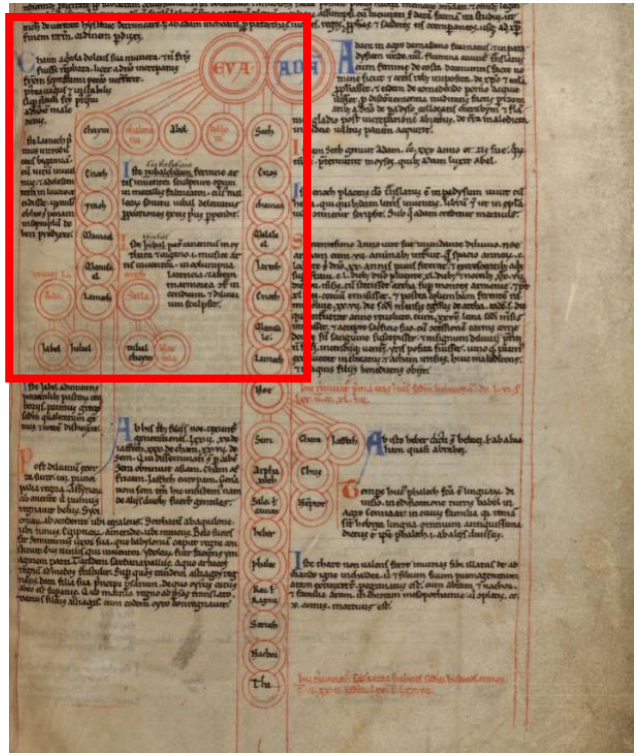
Saarbrücken, 19.-23.2.2024

Roman Bleier, roman.bleier@uni-graz.at

Tagging Persons

Chaim agricola dolens sua munera. et non fratris fuisse reprobata. licet a domino increpatus fratrem septiformi peccato interfecit. postea uagus et instabilis super terram factus postquam a domino etiam maledictus.





Cambridge, Corpus Christi College, MS 029, vir, https://parker.stanford.edu/parker/catalog/xj710dc7305

Referencing to a List

<persName ref="#pers-cain"> Chaim</persName> agricola dolens sua munera. et non fratris fuisse reprobata. licet a domino increpatus fratrem septiformi peccato interfecit. postea uagus et instabilis super terram factus postquam a domino etiam maledictus.

```
<listPerson>
  <person xml:id="pers-adam">
        <persName xml:lang="en">Adam</persName>
        <persName xml:lang="it">Adamo</persName>
        [...]
  </person>
  <person xml:id="pers-cain">
        <persName xml:lang="en">Cain</persName>
        <persName xml:lang="it">Caino</persName>
        <persName xml:lang="de">Kain</persName>
        [...]
  </person>
  [...]
</listPerson>
```

tei:standOff

- Functions as a container element for linked data, contextual information, and stand-off annotations embedded in a TEI document. (TEI Guidelines)
- tei:standOff is a direct child element of tei:TEI or tei:teiCorpus
- It is recommended to place lists with references to authority files (e.g. tei:listPlace or tei:listPerson) in the standOff-element

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
 <teiHeader>
<!-- ... -->
 </teiHeader>
 <standOff>
  <listPlace>
<!-- ... -->
   <place xml:id="Cilicia">
    <placeName>Cilicia</placeName>
    <idno type="URI">https://pleiades.stoa.org/places/658440</idno>
   </place>
   <place xml:id="Creta">
    <placeName xml:lang="la">Creta</placeName>
    <placeName xml:lang="en">Crete</placeName>
    <idno type="URI">https://pleiades.stoa.org/places/589748</idno>
   </place>
<!-- ... -->
   <place xml:id="Rhodus">
    <placeName xml:lang="la">Rhodus</placeName>
    <placeName xml:lang="en">Rhodes</placeName>
    <idno type="URI">https://pleiades.stoa.org/places/590031</idno>
   </place>
   <place xml:id="Syria">
    <placeName>Syria</placeName>
    <idno type="URI">https://pleiades.stoa.org/places/1306</idno>
   </place>
<!-- ... -->
  </listPlace>
 </standOff>
 <text>
  <body>
```

https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-standOff.html

Linking and Referencing Attributes

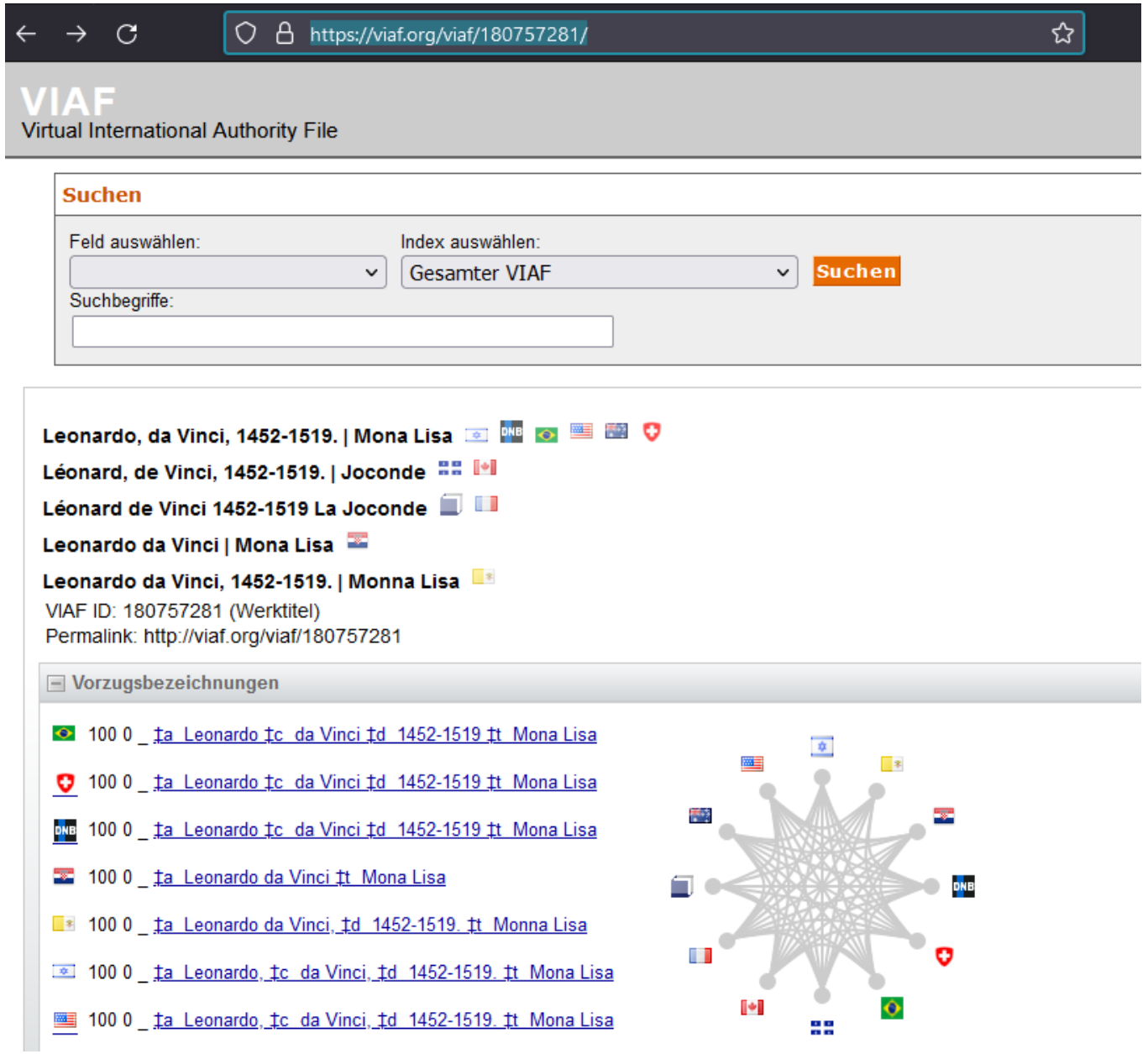
- TEI attributes for referencing, linking and pointing:
  - att.global.linking: @corresp, @sameAs, @prev, @next, etc.
  - att.pointing: @target, etc.
  - att.naming: @nymRef and @role
  - att.referencing: @cRef
  - att.canonical: @ref, @key
- **@ref: (reference)** provides an explicit means of locating a full definition or identity for the entity being named by means of one or more URIs.

```
<name ref="http://viaf.org/viaf/109557338"
 type="person">Seamus Heaney</name>
```

Authority Files

Authority Files

- In a library context authority control is used to organise bibliograpic information about people, places, and subjects
- It is a controlled vocabulary
- An authority file is a database to manage these data
- Authority files (such as GND or VIAF) provide stable URIs for reference



VIAF

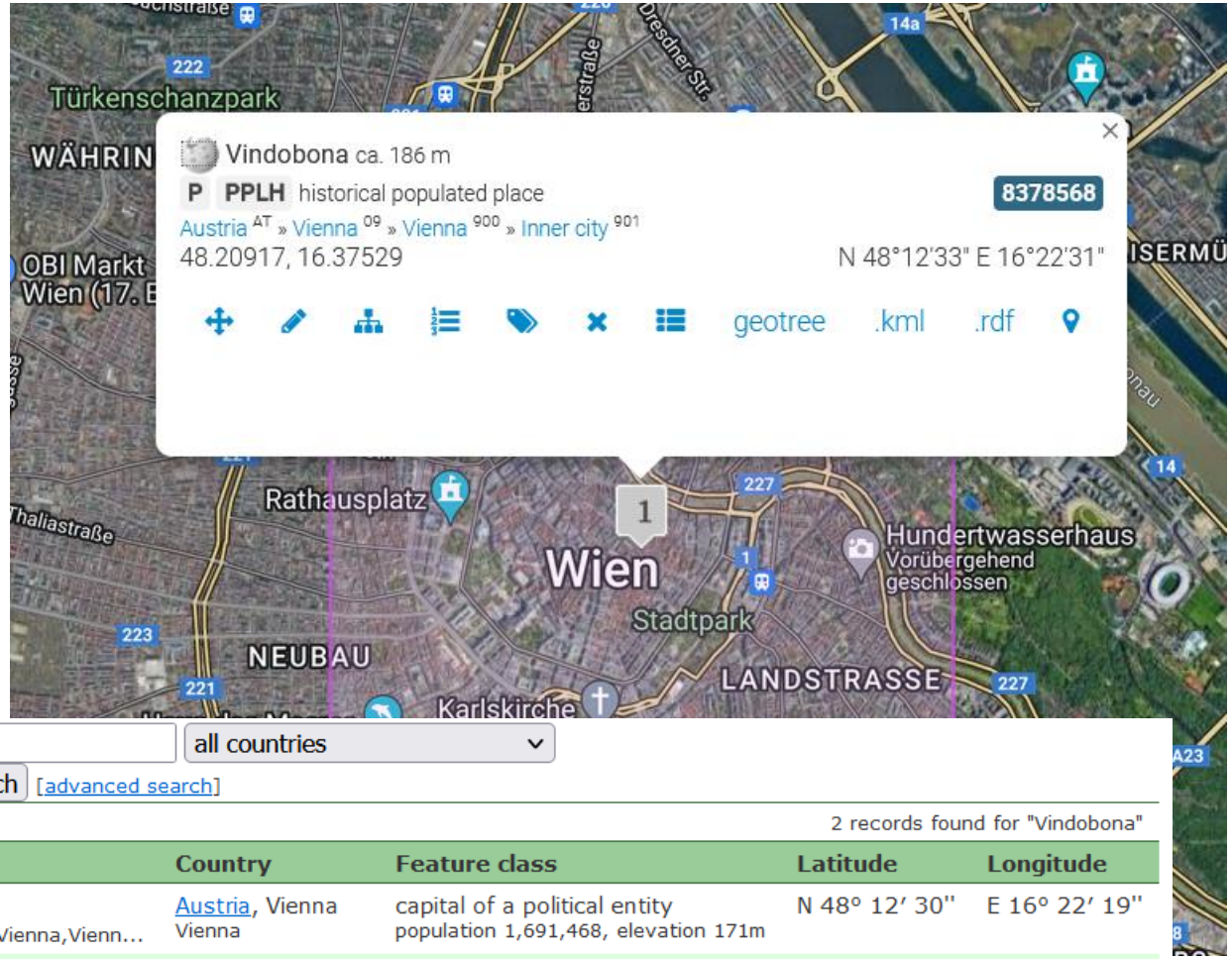
- Virtual International Authority File (VIAF)
- Started in 1998 as an attempt to link the authority files of the German National Library and the Library of Congress
- VIAF links now more than 50 authority files
- Provides stable URIs for referencing

http://viaf.org/viaf/1368673

GeoNames

https://www.geonames.org/8378568

- GeoNames geographical database covers all countries and contains over eleven million placenames
- Some historical placenames, but it is not a priority of the project



Vindobona | all countries | search | [advanced search]

2 records found for "Vindobona"

	Name	Country	Feature class	Latitude	Longitude
1	Vienna Bec,Bech,Becs,Beç,Beč,Bienne,Bécs,Dunaj,Fiena,VIE,Vena,Viden,Viden',Vieden,Viedeň,Viena,Vienna,Vienn...	Austria, Vienna Vienna	capital of a political entity population 1,691,468, elevation 171m	N 48° 12′ 30″	E 16° 22′ 19″
2	Vindobona Vindobona	Austria, Vienna Vienna > Inner city	historical populated place	N 48° 12′ 33″	E 16° 22′ 31″

How to Reference Authority Files in the TEI

Direct reference to an authority files entry:

```
<persName ref="http://viaf.org/viaf/1368673 ">Chain</persName>
<persName ref="viaf:1368673">Chain</persName>

<placeName ref="https://www.geonames.org/2964574">
                                                    Dublin</placeName>
<placeName ref="geo:2964574">Dublin</placeName>
```

Defining Prefixes

```
<listPrefixDef>
    <prefixDef ident=„viaf"
        matchPattern="([0-9]+)"
        replacementPattern="http://viaf.org/viaf/$1">
        <p> In the context of this project, the prefix "viaf" point to
            the VIAF authorty file.
        </p>
    </prefixDef>
   <prefixDef ident= "geo" [...]

</listPrefixDef>⚓
```

References to Authority Files in listPerson, etc.

- tei:idno supplies any form of identifier used to identify an object
- Often it contains an authority file identifier

```
<person xml:id="pers-cain">
        <persName xml:lang="en">Cain</persName>
        <persName xml:lang="it">Caino</persName>
        <persName xml:lang="de">Kain</persName>
        [...]
        <idno type="VIAF"> 1368673 </persName>
  </person>
```

Hands-on

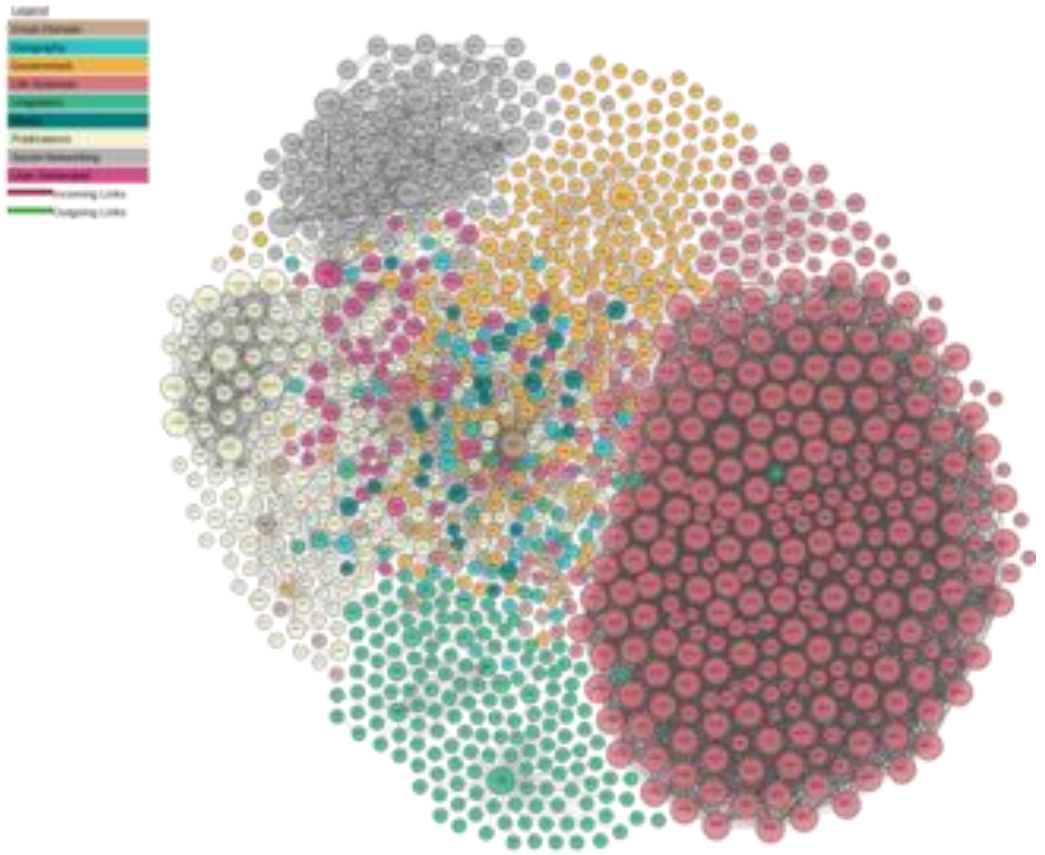
- You may continue with the files from yesterday
- But I have also added a file with an XML sample from the project *History as a visual concept*
- Use the authority files VIAF and Geonames to add the correct identifier to your
- The websites https://www.geonames.org and http://viaf.org should allow you to search for terms and names

Web of Data

Linked Open Data

„The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data.”

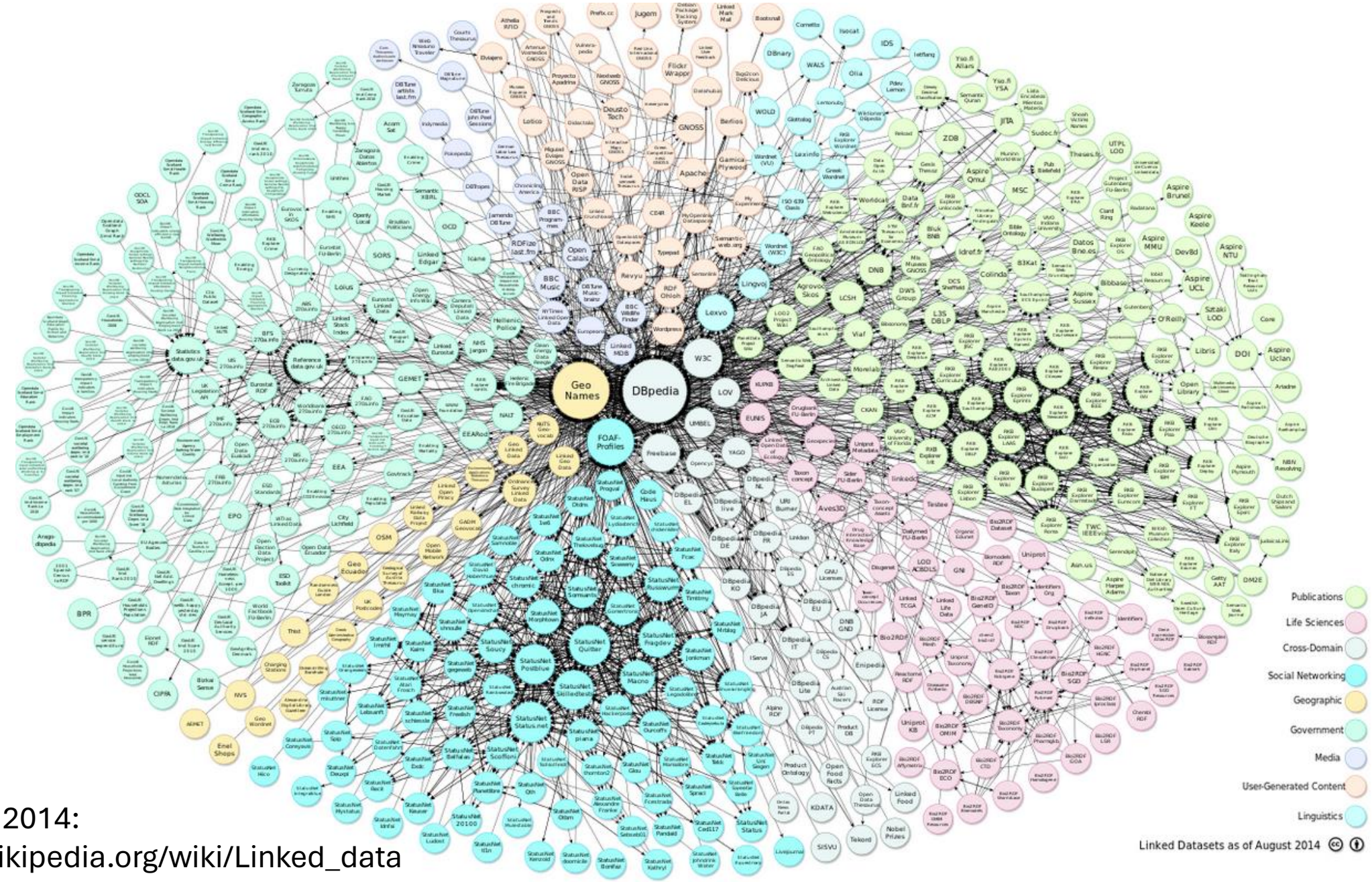
(Tim Berners-Lee, 2006, https://www.w3.org/DesignIssues/LinkedData.html)



LOD-Cloud, 2017: https://en.wikipedia.org/wiki/Linked_data

TED-Talk: Tim Berners-Lee, 2009

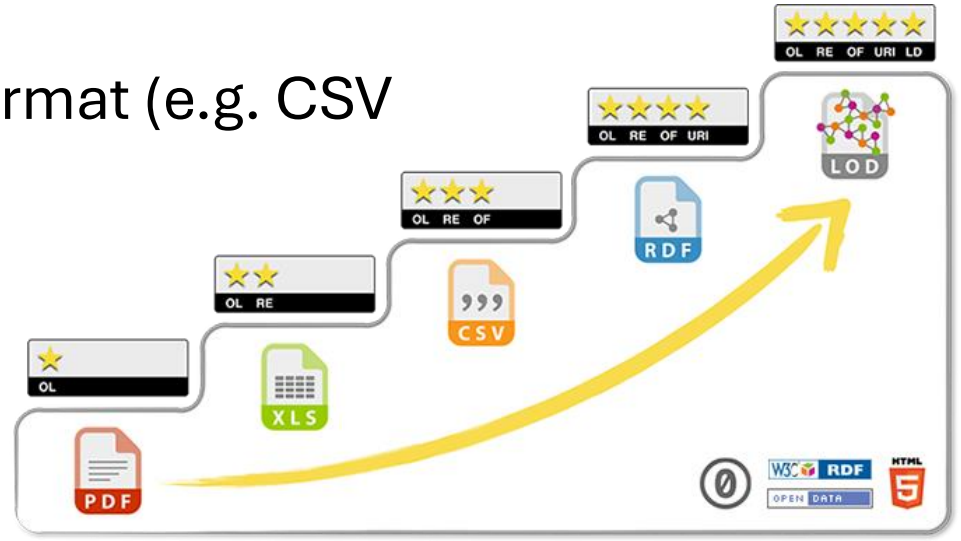
- https://www.ted.com/talks/tim_berners_lee_the_next_web
- Put data your on the internet!
- Not as nice web pages, but as structured data.
- Raw data now!
- Data is about relationships => Linked Data



LOD-Cloud, 2014:
https://en.wikipedia.org/wiki/Linked_data

Five-Star-Linked Open Data

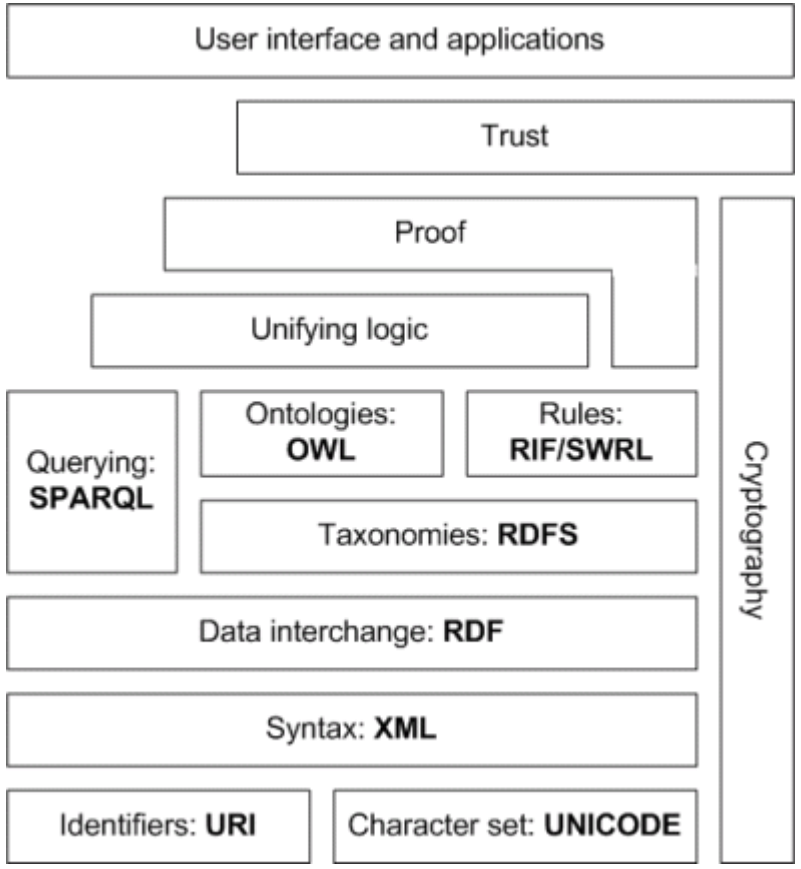
- make your stuff available on the Web (whatever format) under an open license
- make it available as structured data (e.g. Excel instead of image scan of a table)
- make it available in a non-proprietary open format (e.g. CSV instead of Excel)
- use the RDF Standard and URIs to denote things, so people can point at your stuff
- link your data to other data to provide context



https://5stardata.info/en/

Web of Data: Technologie Stack

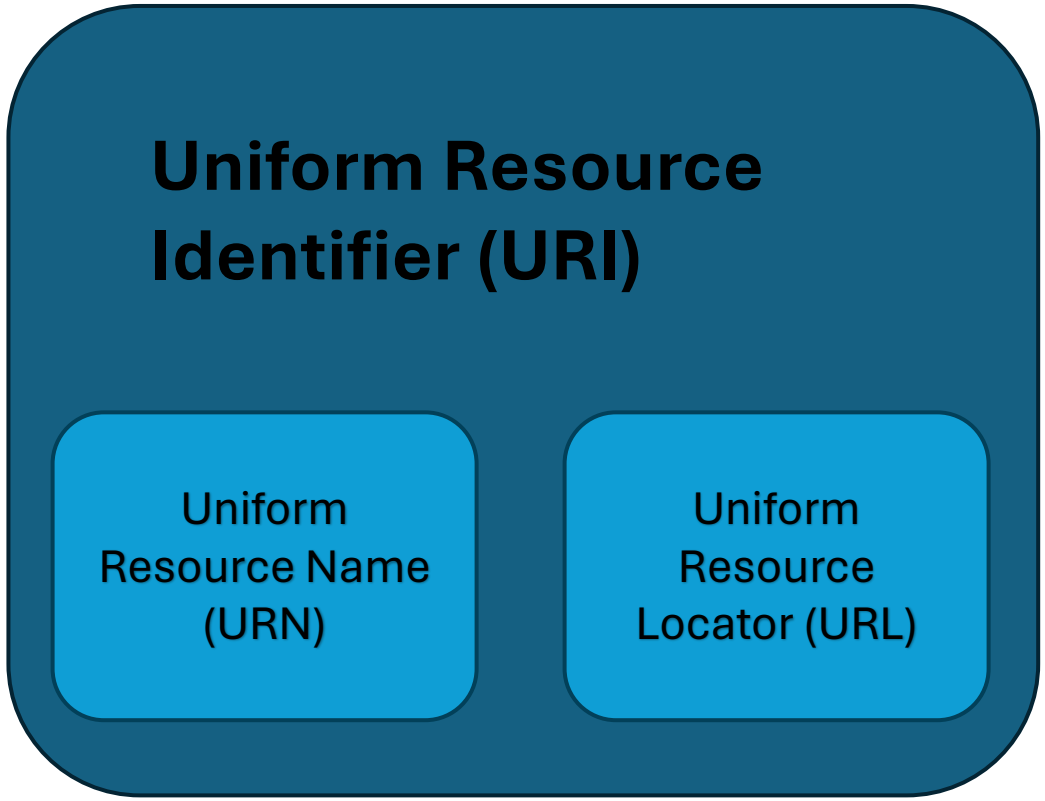
- Different standards are used to build the Web of Data
- URI
- RDF
- OWL and RDFs
- SPARQL



https://en.wikipedia.org/wiki/File:Semantic-web-stack.png

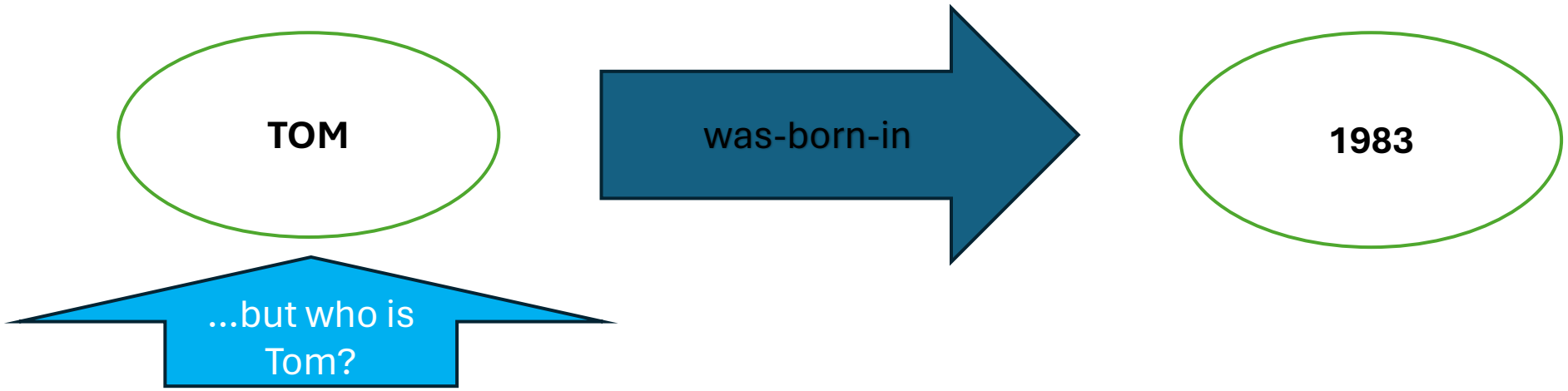
Uniform Resource Identifier

- URI is a unique string that identifies a resource (anywhere !)
- Subsets of URIs are:
  - URN: naming
  - URL: locating on the Internet



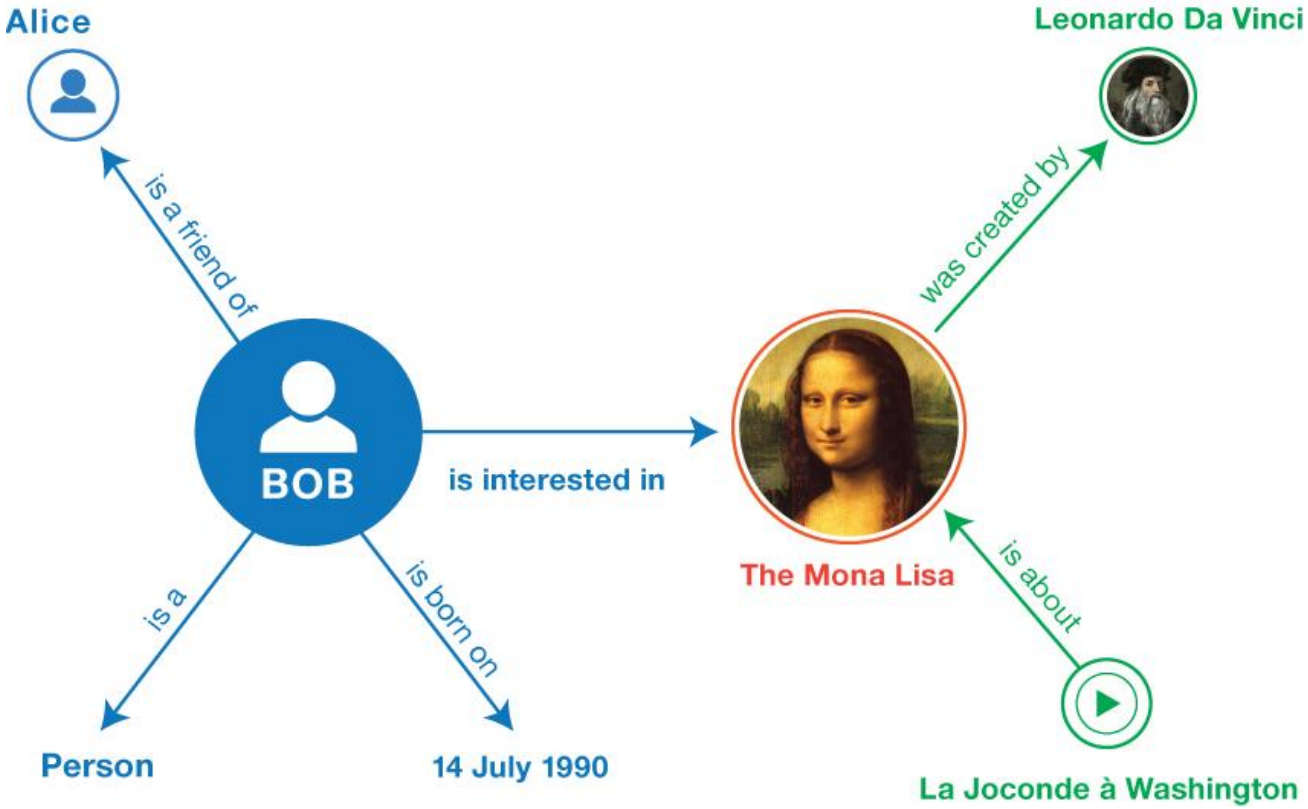
The Resource Description Framework (RDF)

- Semantic triples or RDF triples
- Triples are like simple sentences
- Subject – Predicate - Object
- A simple triple statement



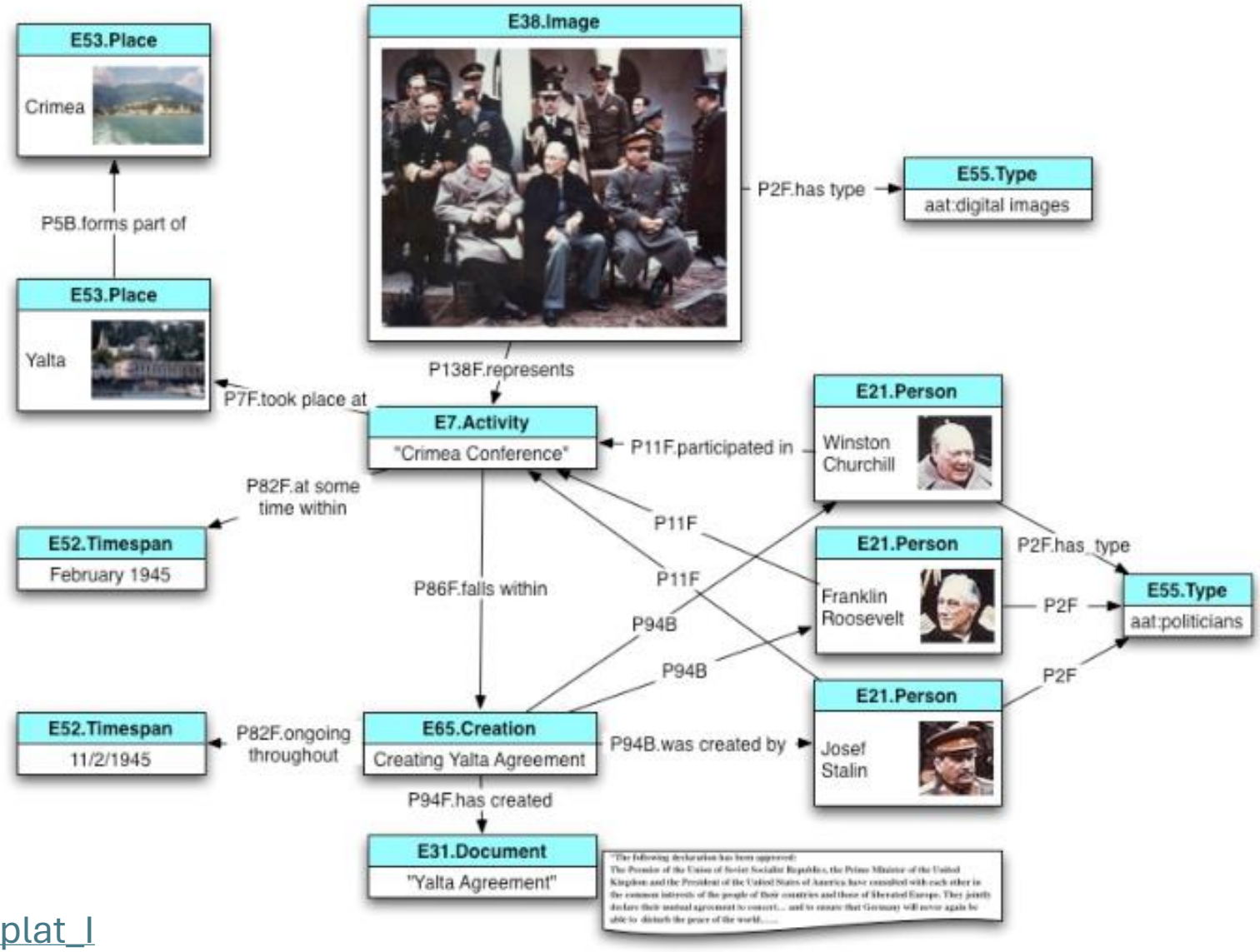
From Triple to Knowledge Graph

- Adding more triple connections provides more information and context



CIDOC CRM

Ontologies like CIDOC CRM provide concepts to describe things



Introduction video:
https://www.youtube.com/watch?v=XP58cplat_I

Wikidata

Wikidata


And every piece of data in Wikidata can be described in any of the nearly

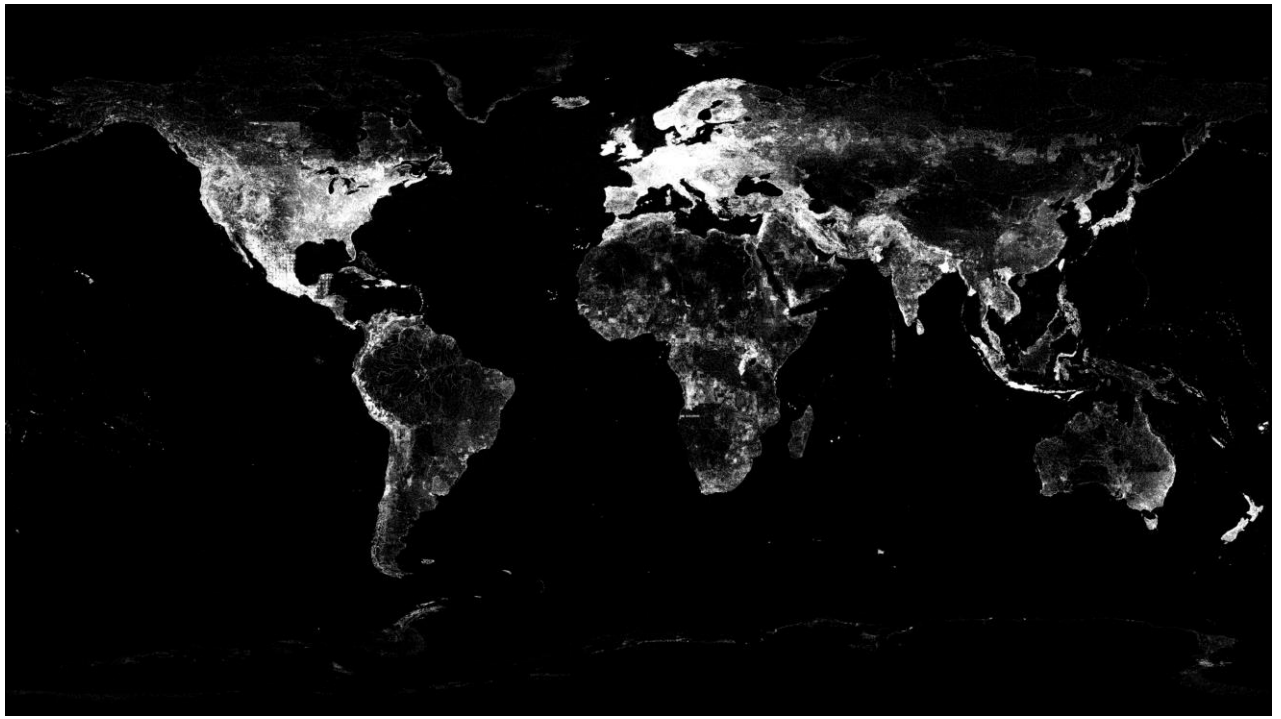
Wikidata: entry or item

- Wikidata: huge global database / LOD
- Wikidata is Linked Open Data
- Statements are represented as triples
- Wikidata distinguishes between:
  - Items: Q-identifier QID, e.g. „Venice“ Q641
  - Properties: P-identifier, e.g. „instance of“ P31

Namespaces:
wd: http://www.wikidata.org/entity/
wdt: http://www.wikidata.org/prop/direct/

Bias?

- The bias question can be asked for all authority files
- All authority files are created in a certain (often institutional) context



Map of geographical bias in Wikidata (by Adam Shoreland, via Wikimedia Commons, CC0), https://commons.wikimedia.org/wiki/File:Wikidata-map-2021-10-18-items-intensity-100.png

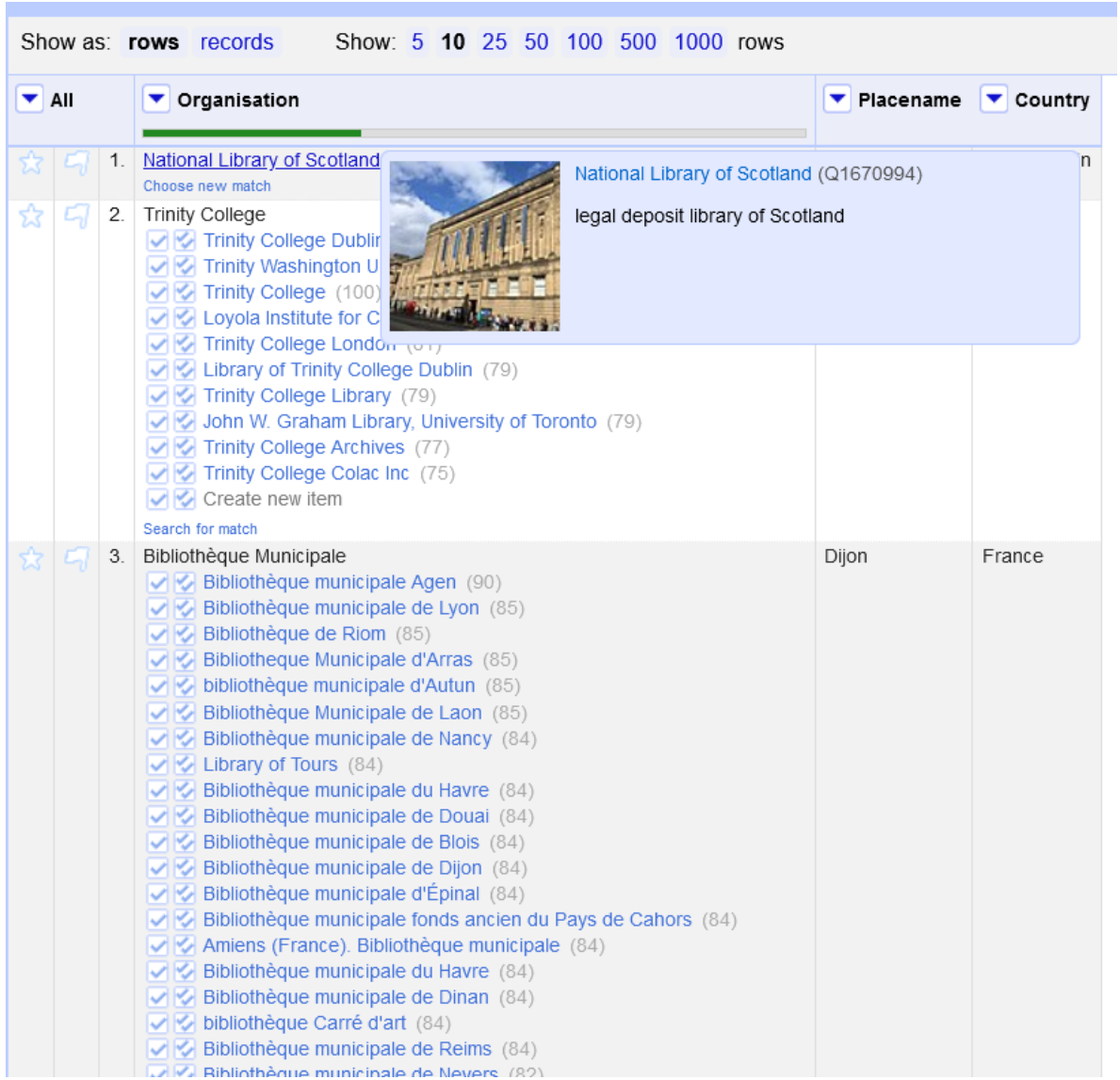
Wikidata: query service

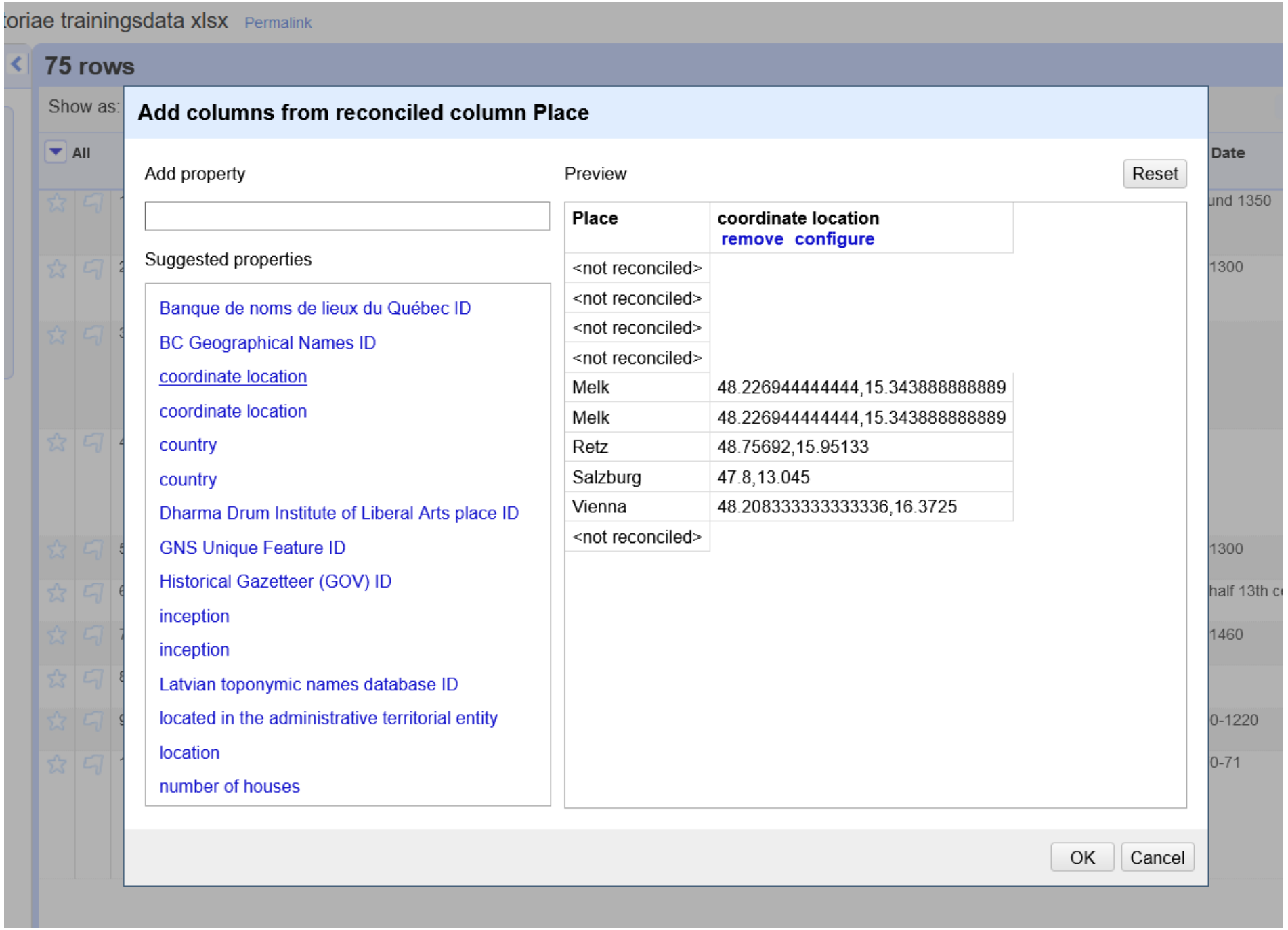
- Wikidata provides a SPARQL-endpoint:
  - https://query.wikidata.org/
- SPARQL can be used to query Wikidata
- Wikidata SPARQL tutorial:
  - https://www.wikidata.org/wiki/Wikidata:SPARQL_tutorial
- Wikidata collection of sample queries about all kinds of topics:
  - https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service/queries/examples

Reconciliation with OpenRefine

- OpenRefine „reconciliation“ of your dataset
- „Reconciliation“: match your dataset with external sources
- Wikidata is included as default service for reconciliation
- OpenRefine documentation:

https://docs.openrefine.org/manual/reconciling




toriae trainingsdata xlsx
Permalink
75 rows
Show as:
All
Add columns from reconciled column Place
Add property
Suggested properties
Banque de noms de lieux du Québec ID
BC Geographical Names ID
coordinate location
coordinate location
country
country
Dharma Drum Institute of Liberal Arts place ID
GNS Unique Feature ID
Historical Gazetteer (GOV) ID
inception
inception
Latvian toponymic names database ID
located in the administrative territorial entity
location
number of houses
Preview
Reset
Place
coordinate location
remove configure
<not reconciled>
<not reconciled>
<not reconciled>
<not reconciled>
Melk
48.226944444444,15.343888888889
Melk
48.226944444444,15.343888888889
Retz
48.75692,15.95133
Salzburg
47.8,13.045
Vienna
48.208333333333336,16.3725
<not reconciled>
OK
Cancel
Date
und 1350
1300
1300
half 13th c
1460
0-1220
0-71

Use of Wikidata: Histropedia

- a project that uses Wikidata as an information source and explores events through timelines
- https://www.youtube.com/watch?v=6hutIjDvGZE

Further Reading

- #dariahTeach: An introduction to Conceptual Modelling - Part 4: From models to ontologies, and back. https://www.youtube.com/watch?v=oD-XTd2UY1Y
- George Bruseker: The uses and limits of formal ontologies (Cidoc CRM) for history. https://www.youtube.com/watch?v=3U6WS_M_SFg
- Øyvind Eide and Christian-Emil Ore: Ontologies and data modeling. In: The Shape of Data in the Digital Humanities (2018).
- Julia Flanders, Fotis Jannidis: Knowledge Organization and Data Modeling in the Humanities (2015). https://www.wwp.northeastern.edu/outreach/conference/kodm2012/flanders_jannidis_datamodeling.pdf