



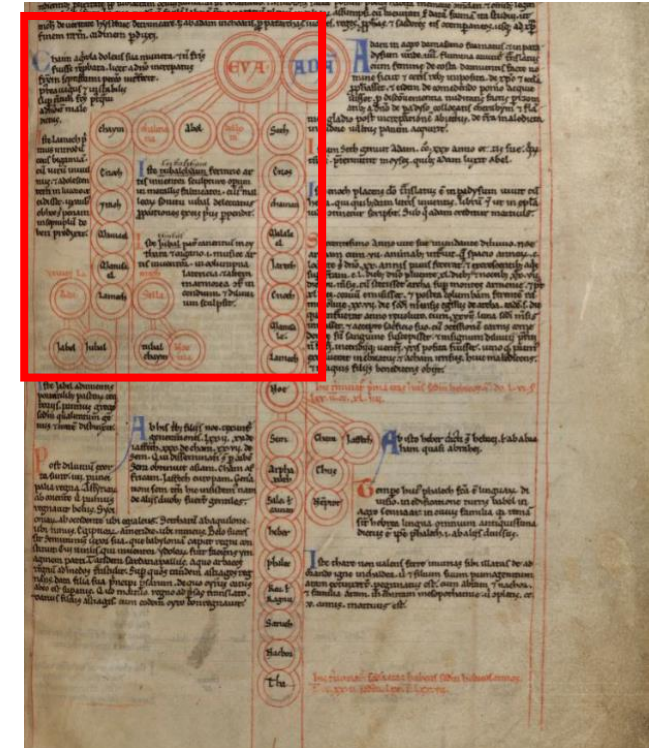
# Linked Open Data

Master Class ‚Digital Scholarly Editing‘ 2024, Saarbrücken  
Saarbrücken, 19.-23.2.2024

Roman Bleier, [roman.bleier@uni-graz.at](mailto: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>
      <div type="edition">
```

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

# Linking and Referencing Attributes

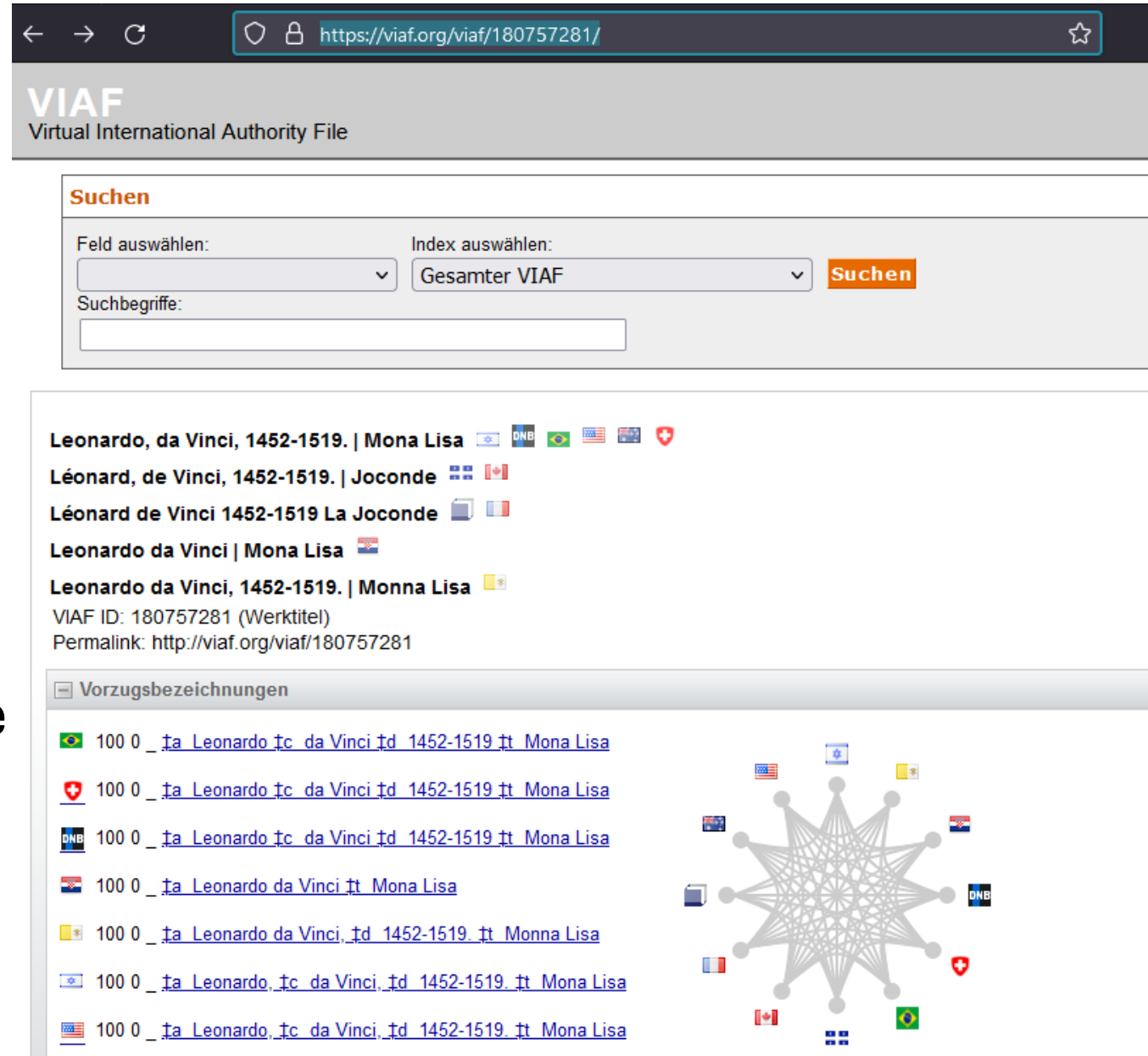
- TEI attributes for referencing, linking and pointing:
  - att.global.links: @corresp, @sameAs, @prev, @next, etc.
  - att.pointing: @target, etc.
  - att.naming: @nymRef and @role
  - att.referencing: @cRef
  - att.canonical: @ref, @key
- **@ref: (reference)** provides an explicit means of locating 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 bibliographic 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



The screenshot displays the VIAF (Virtual International Authority File) website interface. At the top, the browser address bar shows the URL <https://viaf.org/viaf/180757281/>. The VIAF logo and the text 'Virtual International Authority File' are visible below the browser bar.

The main content area features a search section titled 'Suchen'. It includes two dropdown menus for 'Feld auswählen:' and 'Index auswählen:', with 'Gesamter VIAF' selected in the second. A search button labeled 'Suchen' is positioned to the right. Below these is a text input field for 'Suchbegriffe:'.

The search results section displays the following information:

- Leonardo, da Vinci, 1452-1519. | Mona Lisa** (with flags for DNB, Brazil, USA, and Switzerland)
- Léonard, de Vinci, 1452-1519. | Joconde** (with flags for France and Canada)
- Léonard de Vinci 1452-1519 La Joconde** (with flags for France and Italy)
- Leonardo da Vinci | Mona Lisa** (with flag for Hungary)
- Leonardo da Vinci, 1452-1519. | Monna Lisa** (with flag for Germany)

Below the results, the VIAF ID '180757281 (Werktitel)' and the Permalink 'http://viaf.org/viaf/180757281' are provided.

The 'Vorzugsbezeichnungen' section lists various preferred forms of the name, each preceded by a country flag and a code:

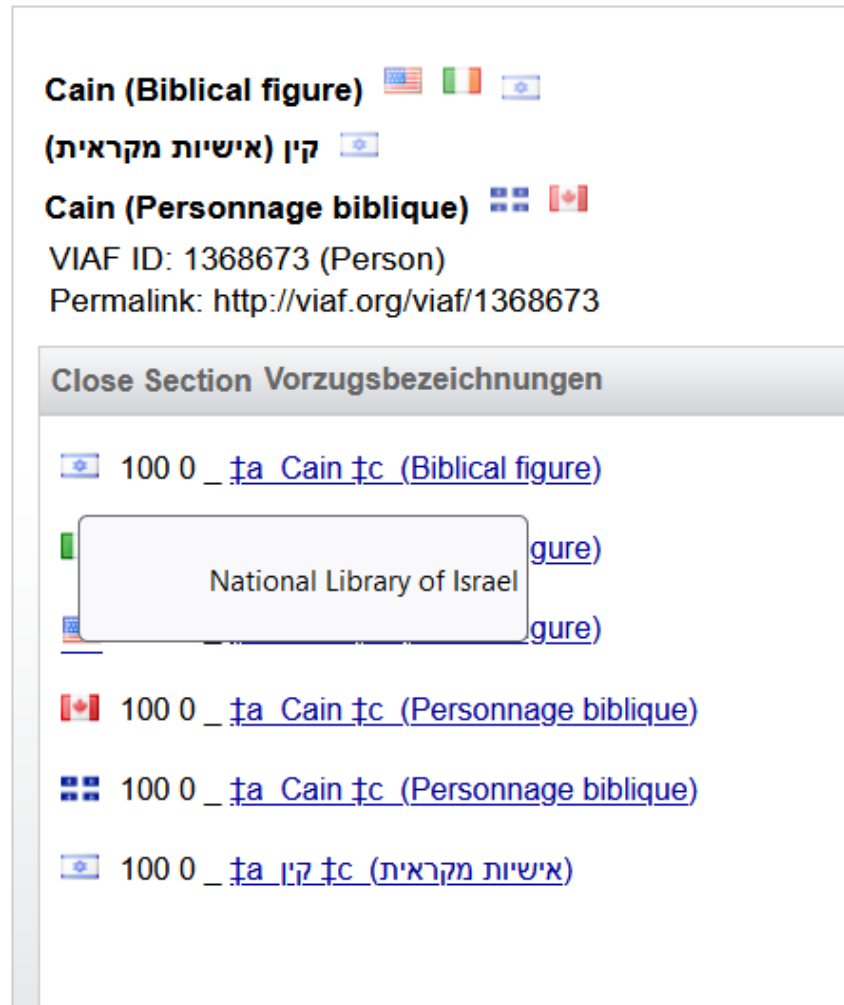
- 100 0 [\\_ta Leonardo \\_tc da Vinci \\_td 1452-1519 \\_tt Mona Lisa](#) (Brazil)
- 100 0 [\\_ta Leonardo \\_tc da Vinci \\_td 1452-1519 \\_tt Mona Lisa](#) (Switzerland)
- 100 0 [\\_ta Leonardo \\_tc da Vinci \\_td 1452-1519 \\_tt Mona Lisa](#) (DNB)
- 100 0 [\\_ta Leonardo da Vinci \\_tt Mona Lisa](#) (Hungary)
- 100 0 [\\_ta Leonardo da Vinci \\_td 1452-1519 \\_tt Monna Lisa](#) (Germany)
- 100 0 [\\_ta Leonardo \\_tc da Vinci \\_td 1452-1519 \\_tt Mona Lisa](#) (USA)
- 100 0 [\\_ta Leonardo \\_tc da Vinci \\_td 1452-1519 \\_tt Mona Lisa](#) (USA)



To the right of the list is a network diagram showing a central node connected to several other nodes, each represented by a flag, illustrating the interconnectedness of the authority file.






# 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





Cain (Biblical figure)   

קין (אישיות מקראית) 

Cain (Personnage biblique)  

VIAF ID: 1368673 (Person)  
Permalink: <http://viaf.org/viaf/1368673>

Close Section Vorzugsbezeichnungen

-  100 0 \_ [ⴱa Cain ⴱc \(Biblical figure\)](#)
-   [gure\)](#)
-   [gure\)](#)
-  100 0 \_ [ⴱa Cain ⴱc \(Personnage biblique\)](#)
-  100 0 \_ [ⴱa Cain ⴱc \(Personnage biblique\)](#)
-  100 0 \_ [ⴱa קין ⴱc \(אישיות מקראית\)](#)

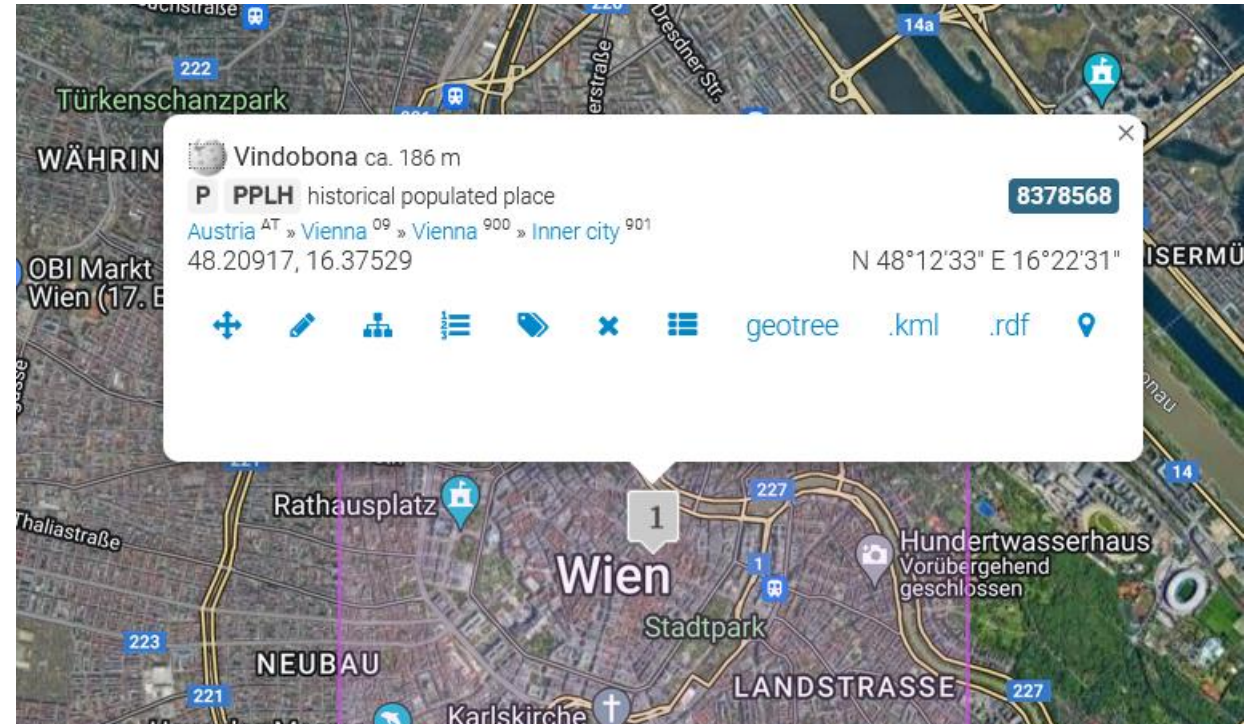
<http://viaf.org/viaf/1368673>



# 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 <a href="#">Vienna</a> Bec, Bech, Beacs, Beç, Beč, Bienne, Bécs, Dunaj, Fiena, VIE, Vena, Viden, Viden', Vieden, Viedeň, Viena, Vienna, Vienn...	<a href="#">Austria</a> , Vienna Vienna	capital of a political entity population 1,691,468, elevation 171m	N 48° 12' 30"	E 16° 22' 19"
2 <a href="#">Vindobona</a> Vindobona	<a href="#">Austria</a> , 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 authority 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

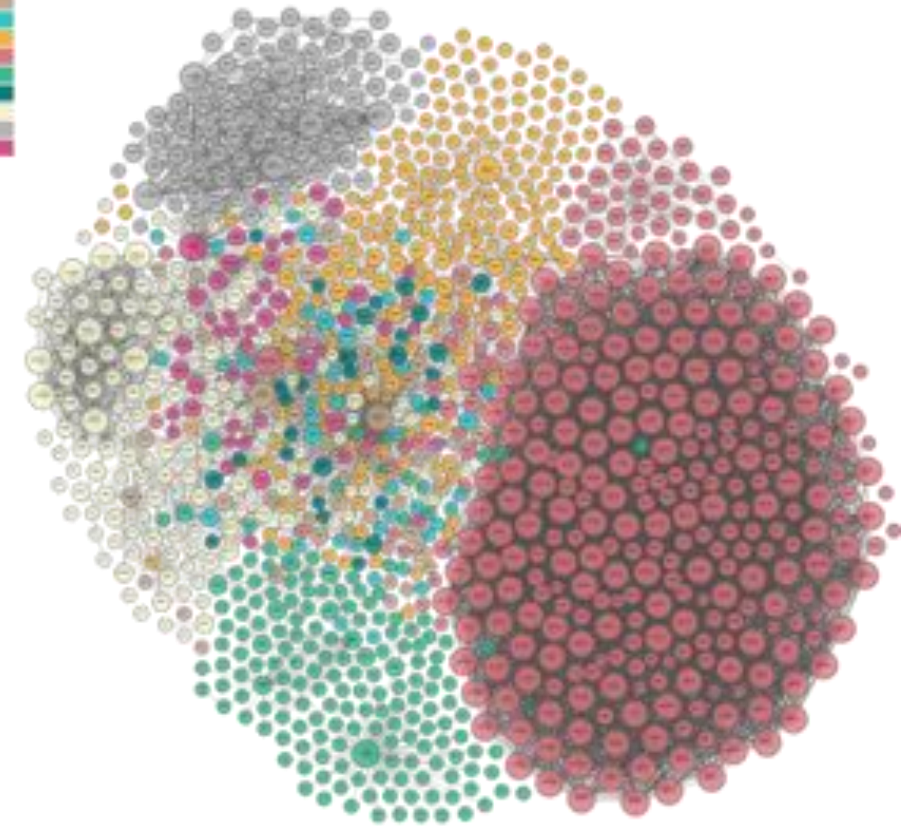
- You may continue with the files from yesterday
- But I have also added a file with an XML sample from the project *History as a visual concept*
- Use the authority files VIAF and Geonames to add the correct identifier to your
- The websites <https://www.geonames.org> and <http://viaf.org> should allow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

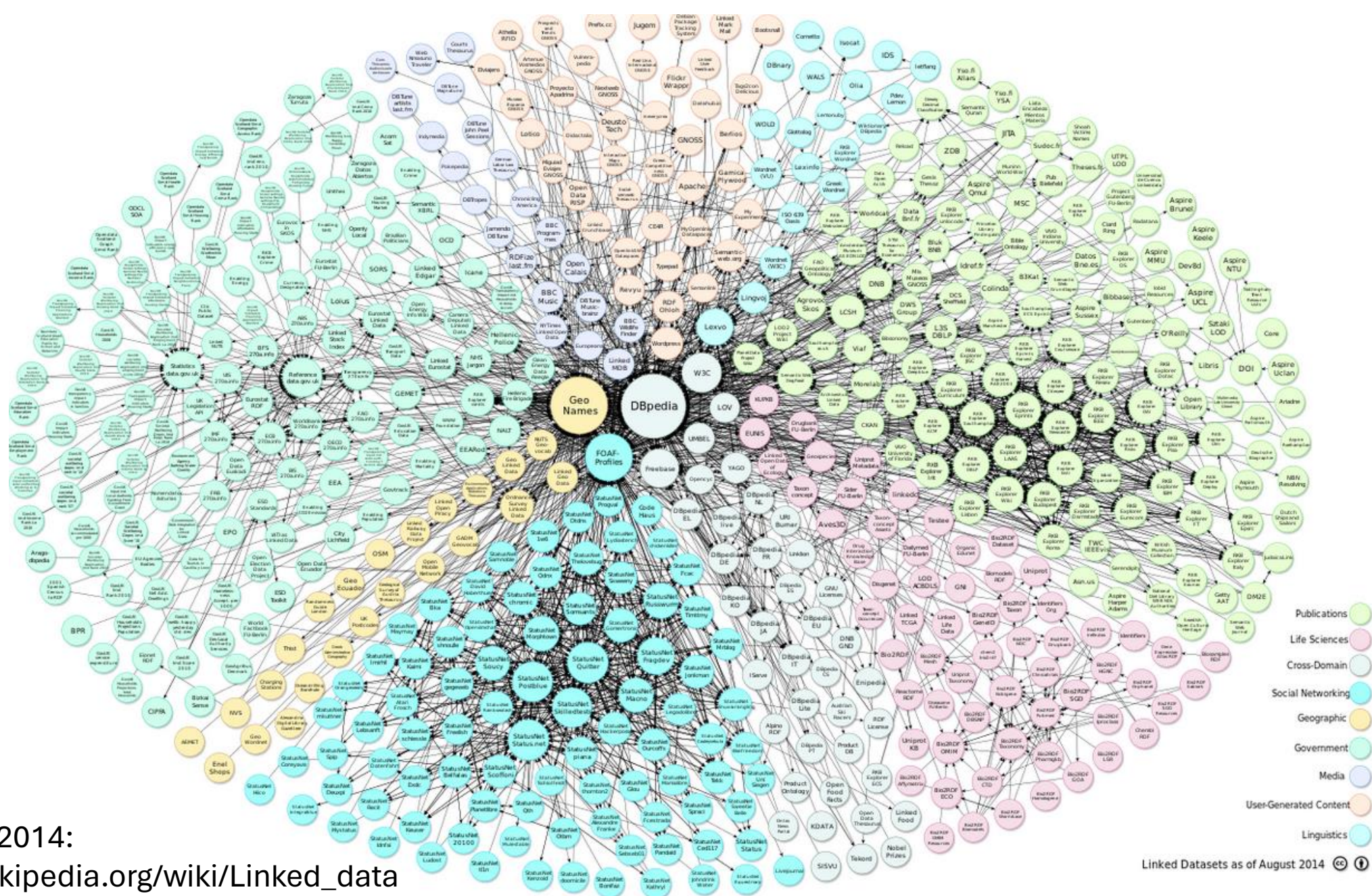


LOD-Cloud, 2017:  
[https://en.wikipedia.org/wiki/Linked\\_data](https://en.wikipedia.org/wiki/Linked_data)



# TED-Talk: Tim Berners-Lee, 2009

- [https://www.ted.com/talks/tim\\_berners\\_lee\\_the\\_next\\_web](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

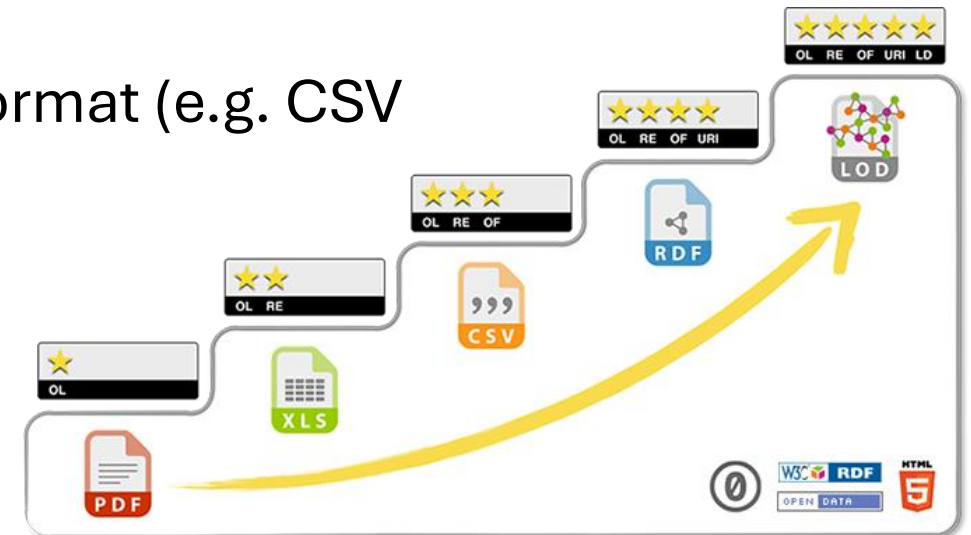


LOD-Cloud, 2014:  
[https://en.wikipedia.org/wiki/Linked\\_data](https://en.wikipedia.org/wiki/Linked_data)



# Five-Star-Linked Open Data

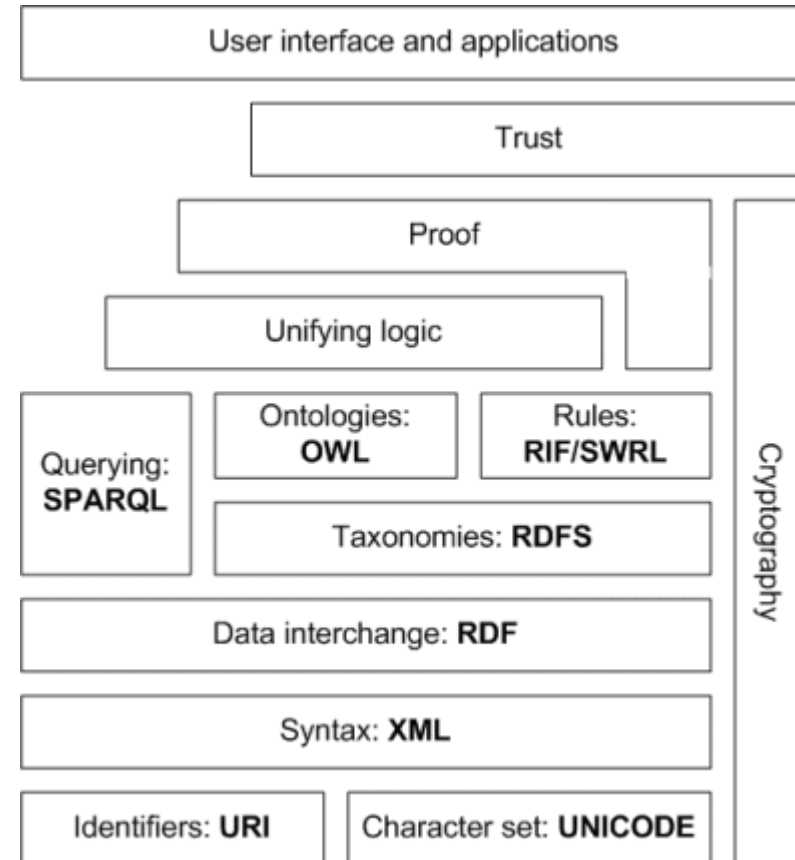
- make your stuff available on the Web (whatever format) under an open license
- make it available as structured data (e.g. Excel instead of image scan of a table)
- make it available in a non-proprietary open format (e.g. CSV instead of Excel)
- use the RDF Standard and URIs to denote things, so people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



<https://5stardata.info/en/>

# Web of Data: Technologie Stack

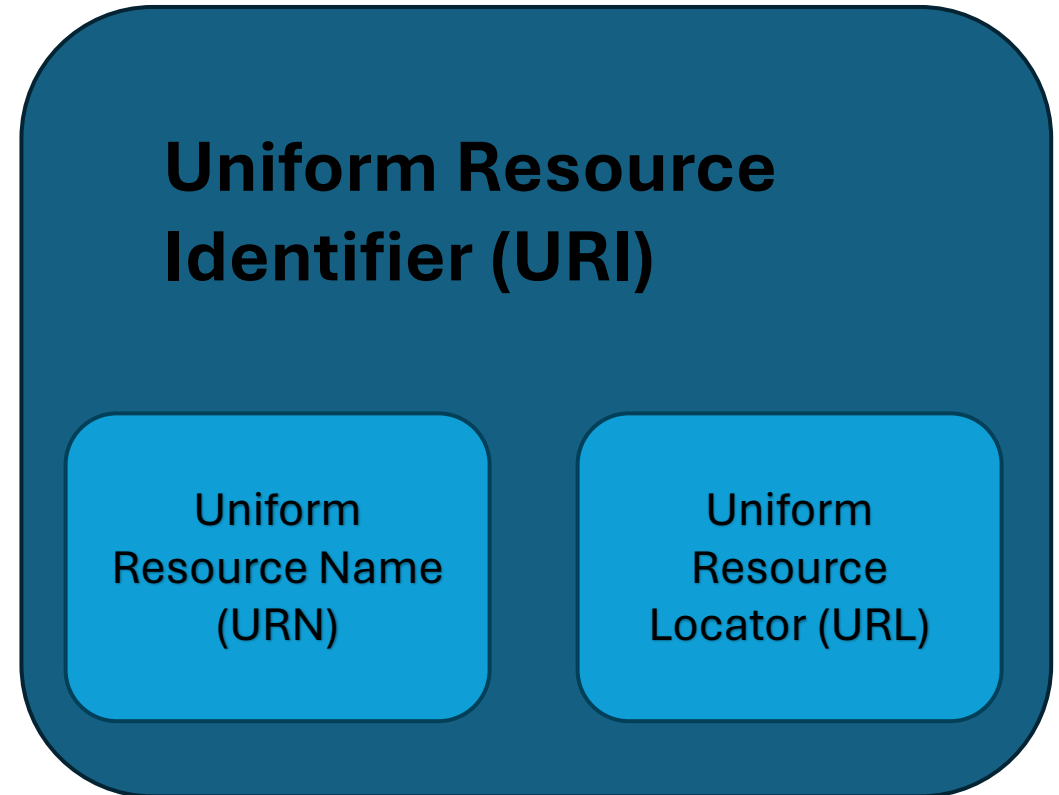
- Different standards are used to build the Web of Data
- URI
- RDF
- OWL and RDFs
- SPARQL



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

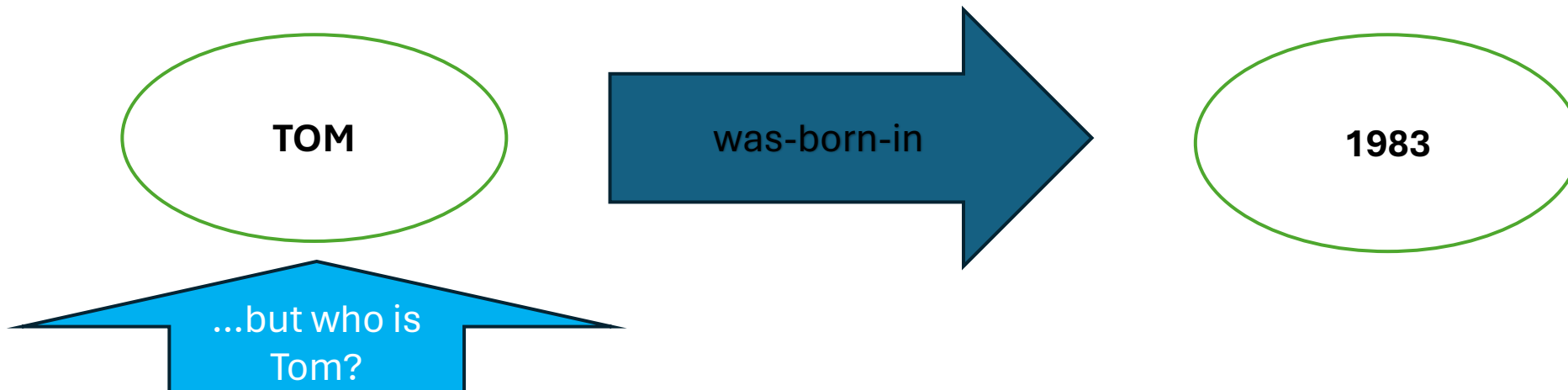
# Uniform Resource Identifier

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



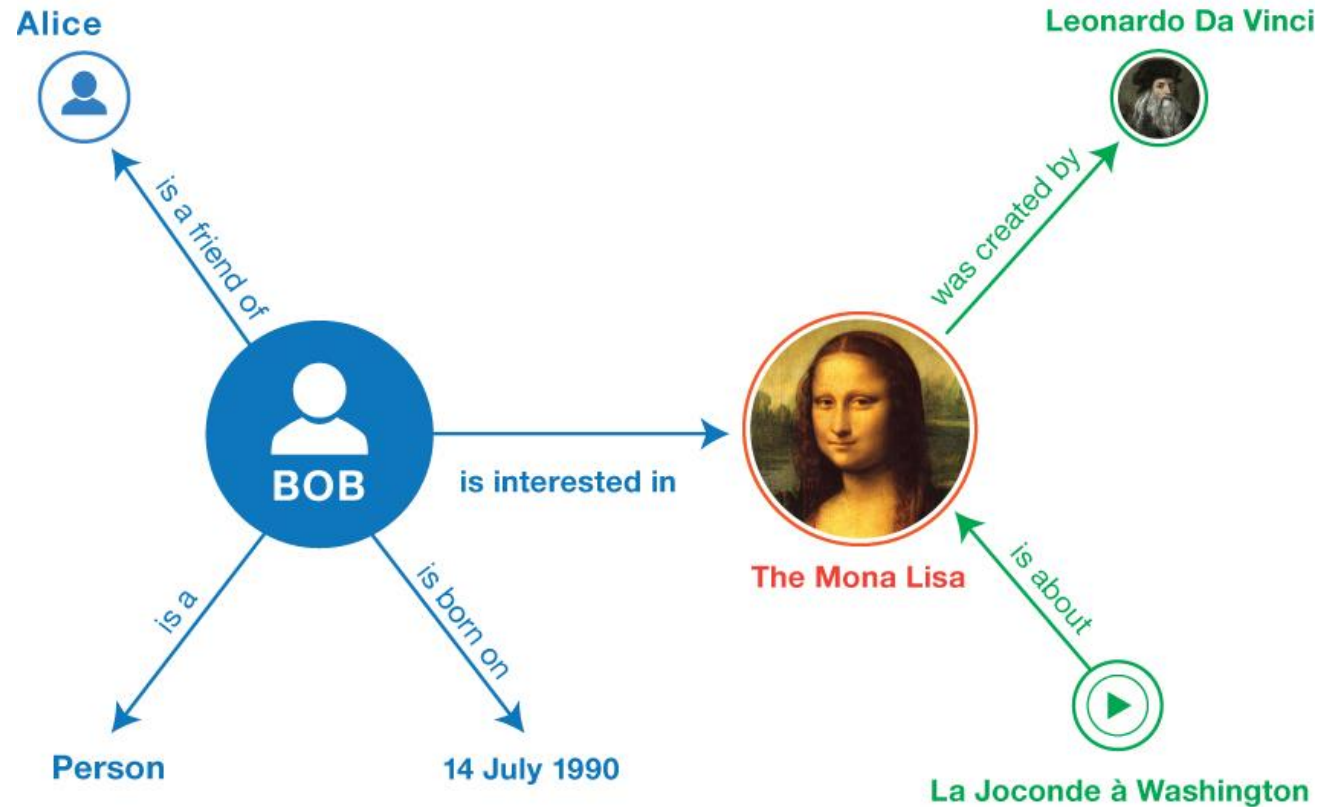
# The Resource Description Framework (RDF)

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



# From Triple to Knowledge Graph

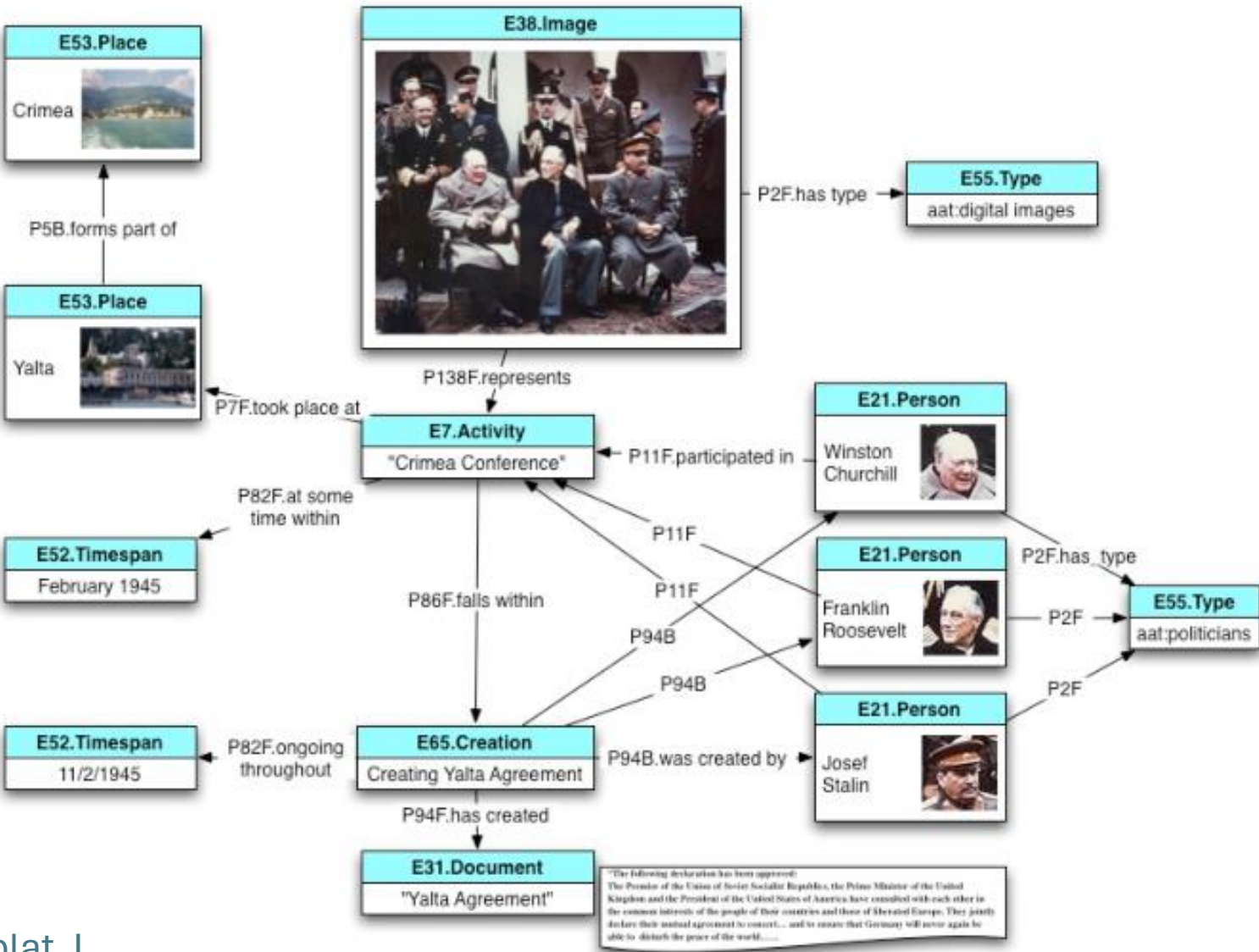
- Adding more triple connections provides more information and context





# CIDOC CRM

Ontologies like CIDOC CRM provide concepts to describe things



Introduction video:  
[https://www.youtube.com/watch?v=XP58cplat\\_I](https://www.youtube.com/watch?v=XP58cplat_I)

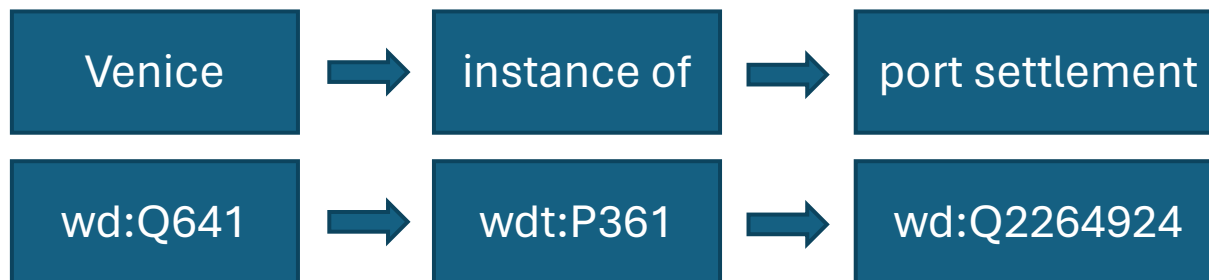
# Wikidata

# Wikidata



# 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



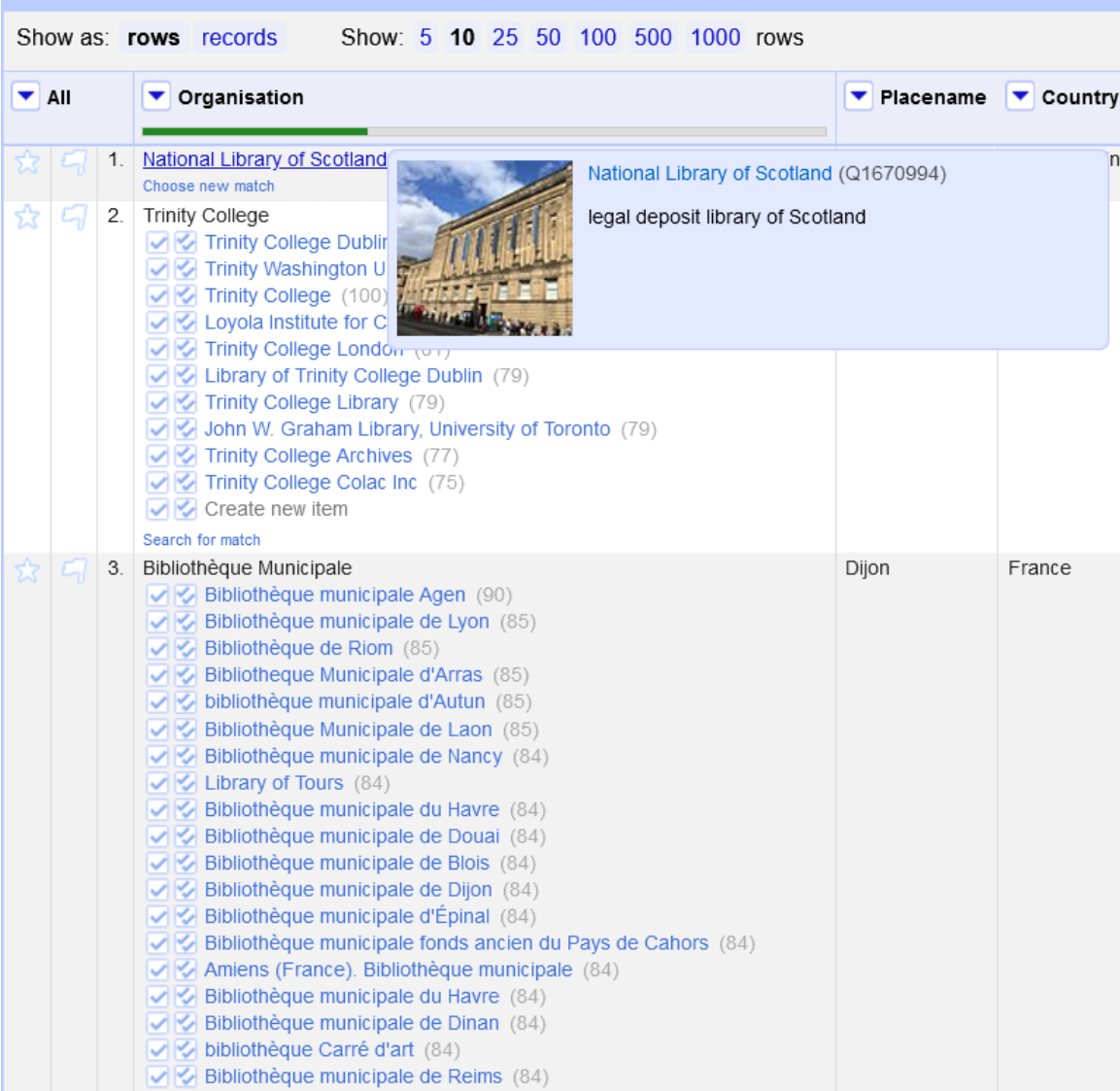
Map of geographical bias in Wikidata (by Adam Shoreland, via Wikimedia Commons, CC0),  
<https://commons.wikimedia.org/wiki/File:Wikidata-map-2021-10-18-items-intensity-100.png>

# 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](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](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>



The screenshot displays the OpenRefine reconciliation interface. At the top, it shows 'Show as: rows records' and 'Show: 5 10 25 50 100 500 1000 rows'. Below this, there are filters for 'All', 'Organisation', 'Placename', and 'Country'. The main area shows a list of suggestions for the selected item 'National Library of Scotland'. The first suggestion is 'National Library of Scotland (Q1670994)' with a thumbnail image of the building and the description 'legal deposit library of Scotland'. Other suggestions include 'Trinity College' with various sub-suggestions like 'Trinity College Dublin', 'Trinity Washington U', etc. The third suggestion is 'Bibliothèque Municipale' with a long list of sub-suggestions like 'Bibliothèque municipale Agen', 'Bibliothèque municipale de Lyon', etc. The interface also includes a 'Search for match' button and a 'Choose new match' option.



< 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
	<a href="#">remove</a> <a href="#">configure</a>
<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	48.208333333333,16.3725
<not reconciled>	

OK Cancel

# Use of Wikidata: Histropedia

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



# 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](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](https://www.wwp.northeastern.edu/outreach/conference/kodm2012/flanders_jannidis_datamodeling.pdf)

This work is licensed under a [Creative Commons  
Namensnennung 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



All works of other author cited here are their intellectual property and are used for academic purpose only.