

Introduction to LOD and data reconciliation

...

Thursday Morning

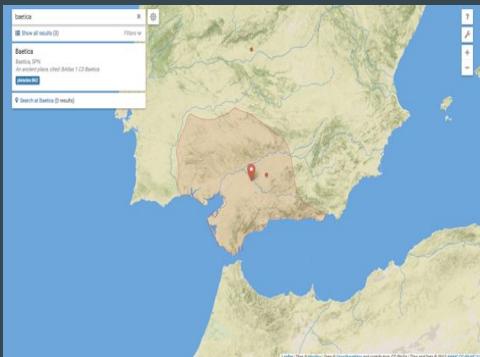
Session outline:

1. Linked Open Data
2. Basic concepts
3. Applications
4. LOD resources
5. Data reconciliation with OPeNRefine

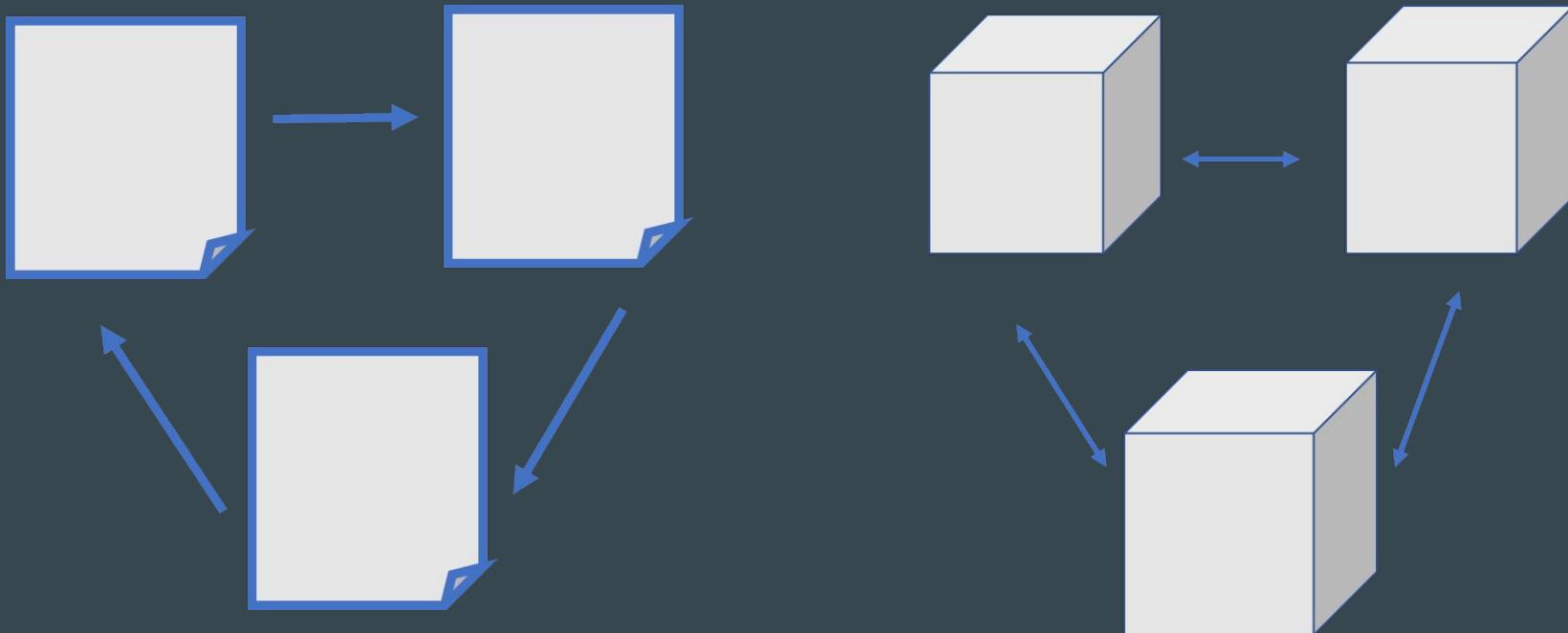
Research?



Osuna



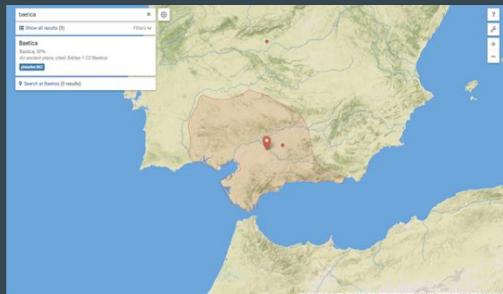
The web of data



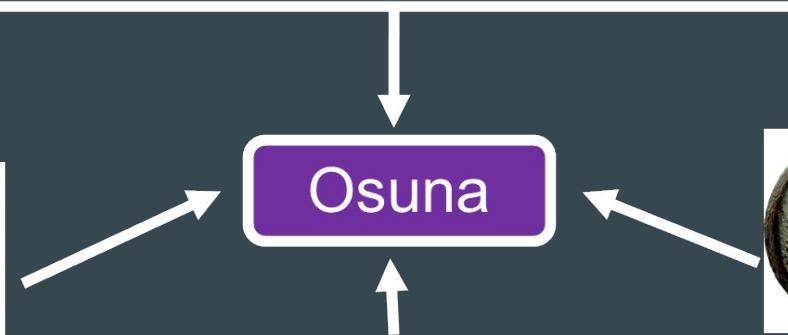
Osuna



Osuna



MAN



Gazetteer

Library

↑ Refers

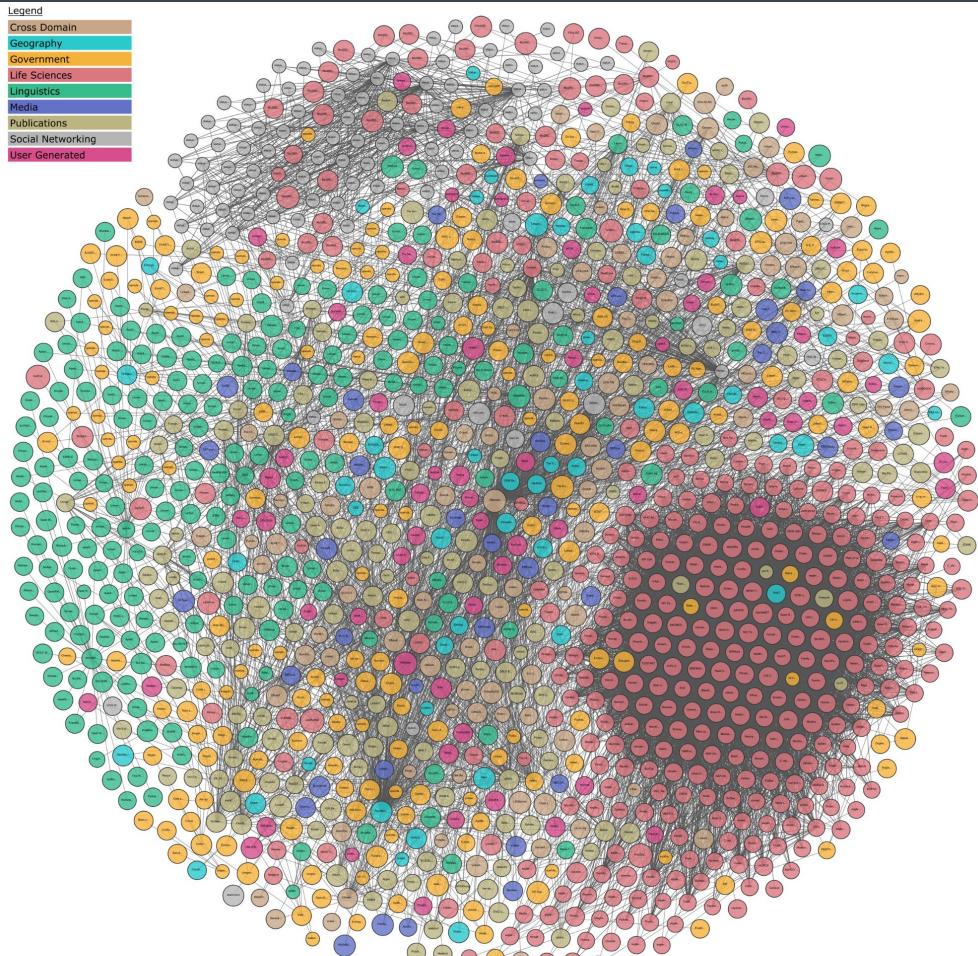
Book

↑ holds

MAN

↑ Refers

Legend
Cross Domain
Geography
Government
Life Sciences
Linguistics
Media
Publications
Social Networking
User Generated



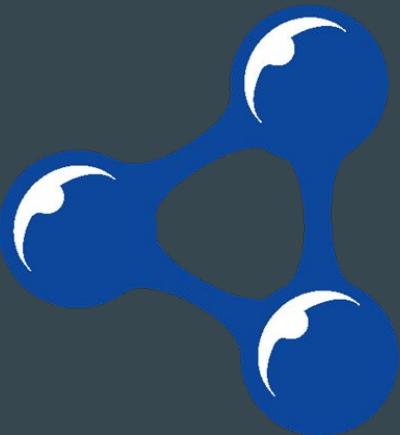
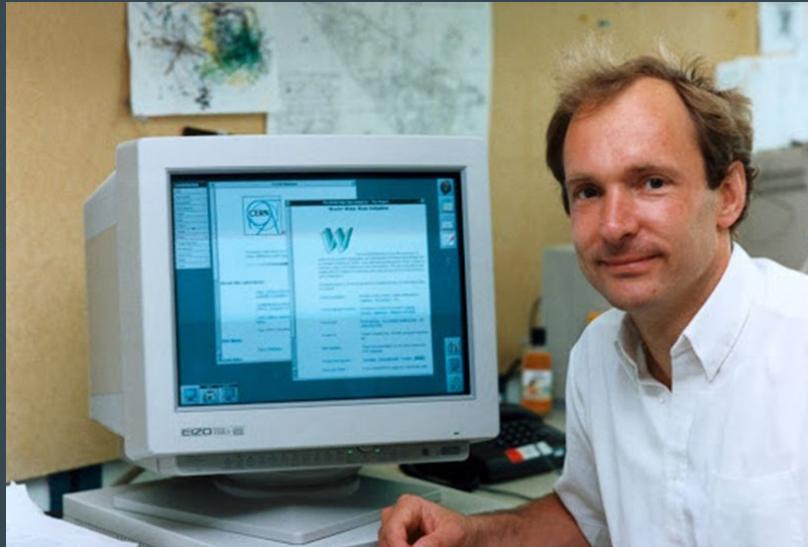
Linked Open Data cloud, available
at <https://lod-cloud.net/>

Main concepts:

- Linked Open Data: theory, ecosystem, technologies and practices around openly licensed structured data interlinked with other data.
- Linked open data: specific datasets or files published on the web.
- Semantic Web: common space where data can be shared, processed, linked and reused.

What is Linked Open Data?

- Method for publishing structured data on the web so that it can be interlinked.
- Builds upon standard web technologies (HTTP, RDF and URIs)
- Rather than using the web to serve web pages for human readers, it extends it to share information in a way that can be accessed automatically by computers.



Five-star criteria for LOD developed in 2010:



Available on the Web under an open license to be Open Data



Available as machine-readable structured data e.g. excel.



Available in a non-proprietary format (e.g., CSV file instead of excel)



Published following open standards from W3C (RDF and SPARQL)



All of the above and links to other people's data to provide context.

¿Why Linked Data?

- Accessibility

The user should be able to access, reuse and link to the data available.

- Openness

Standards For users to understand and reuse the formats in the long term.

- Interoperability

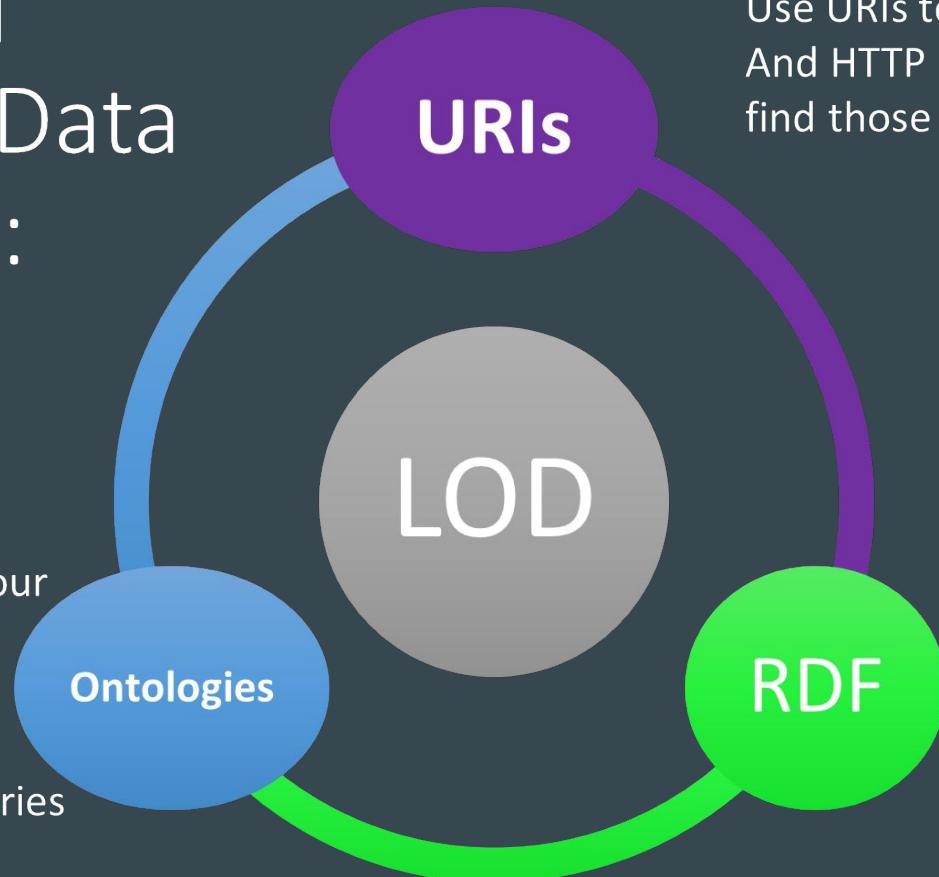
Using HTTP URIs to interlink information.

- Contextualisation

Granularity of detail. All data to be searched for and referenceable.

- Reuse. Open-Ended Potential.

Linked Open Data recipe:



Use URIs to name things
And HTTP URLs so that people can
find those things

Model your
data
Using
existing
vocabularies

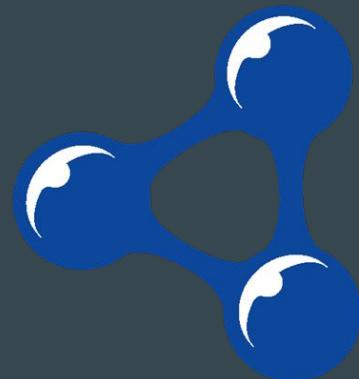
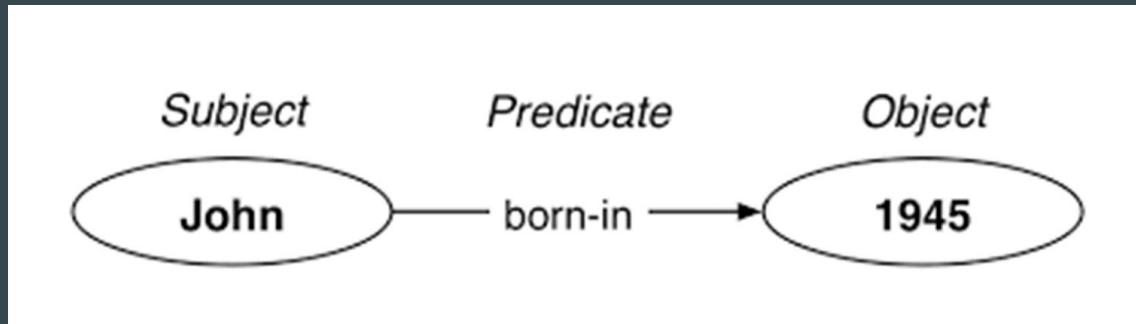
Provide useful RDF data
and include RDF
statements that link to
other data

URIs

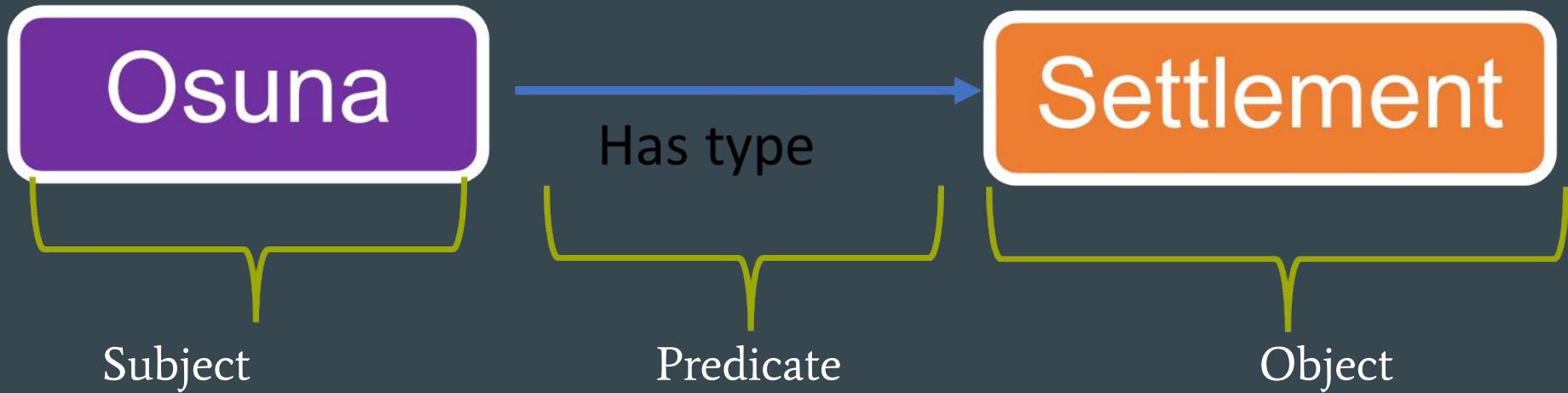
- URI (Uniform Resource Identifier)
 - A type of identifier.
 - String that can be used to refer to a resource on the web, an email address, a telephone number etc..
 - Should be dereferenceable (can be looked up by a browser)
- URL (Uniform Resource Locator)
 - Web address
 - Points to definition/information resource

Resource Description Framework (RDF)

- RDF: Abstract data model at the centre of the method.
- Based on three parts “triple”
- The subject of one triple may become the object of another.



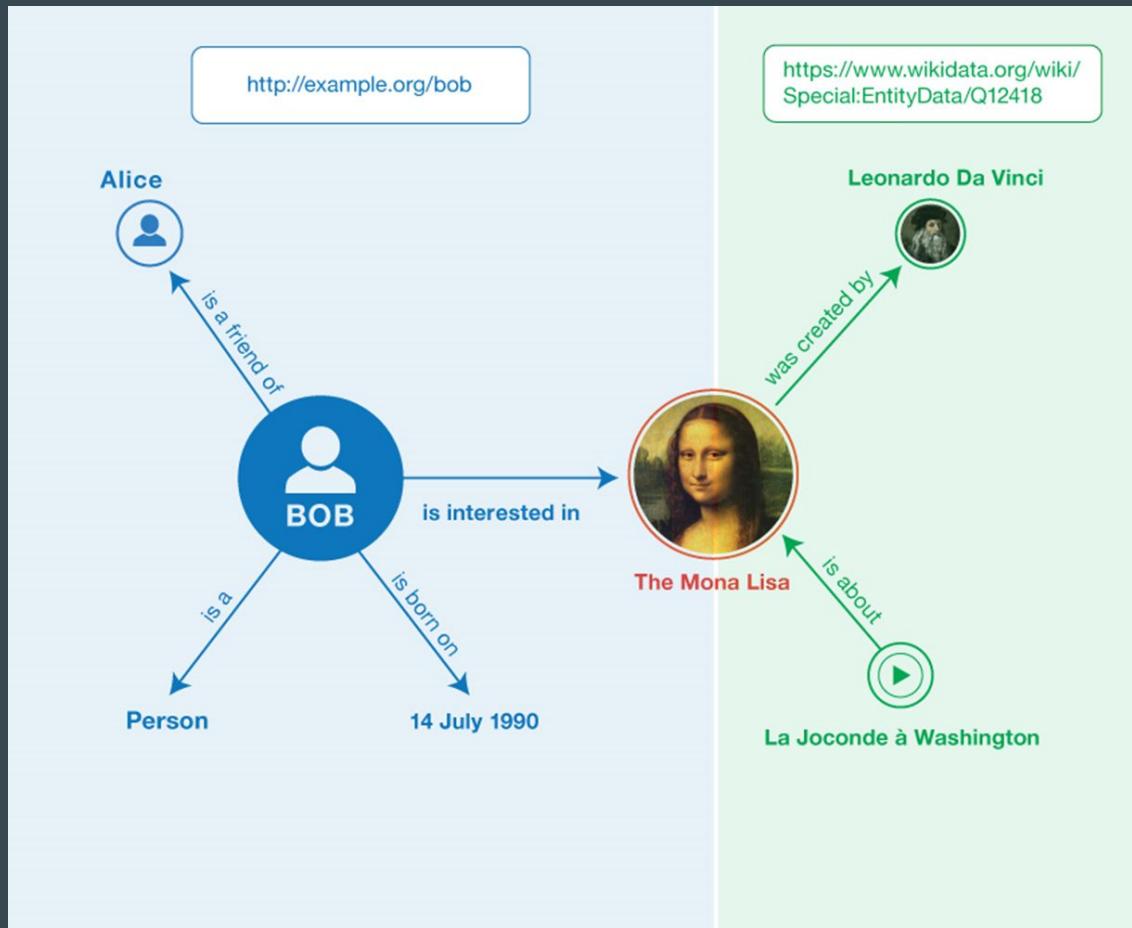
Resource Description Framework (RDF)



Osuna (<https://pleiades.stoa.org/places/256503>)

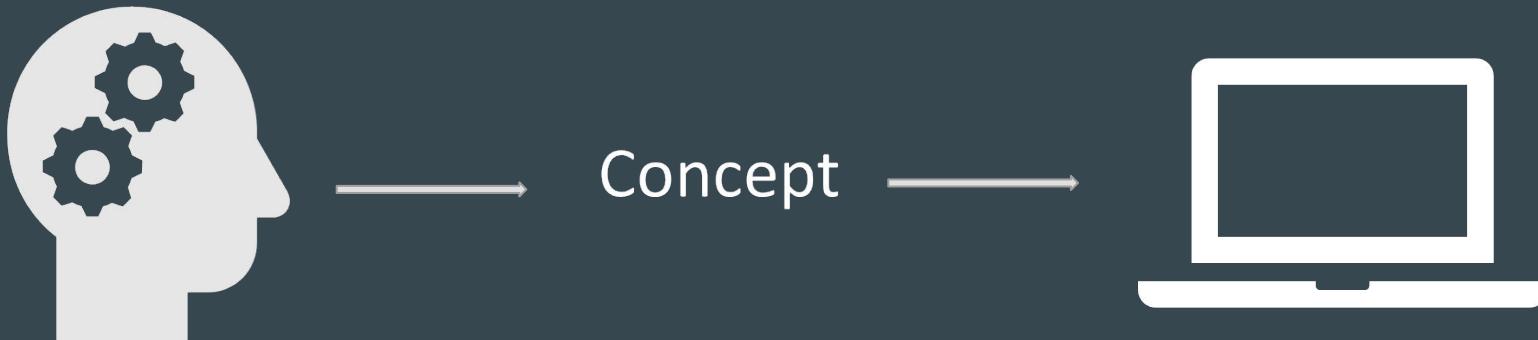
has type (<https://pleiades.stoa.org/vocabularies/place-types>)

settlement (<https://pleiades.stoa.org/vocabularies/place-types/settlement>)

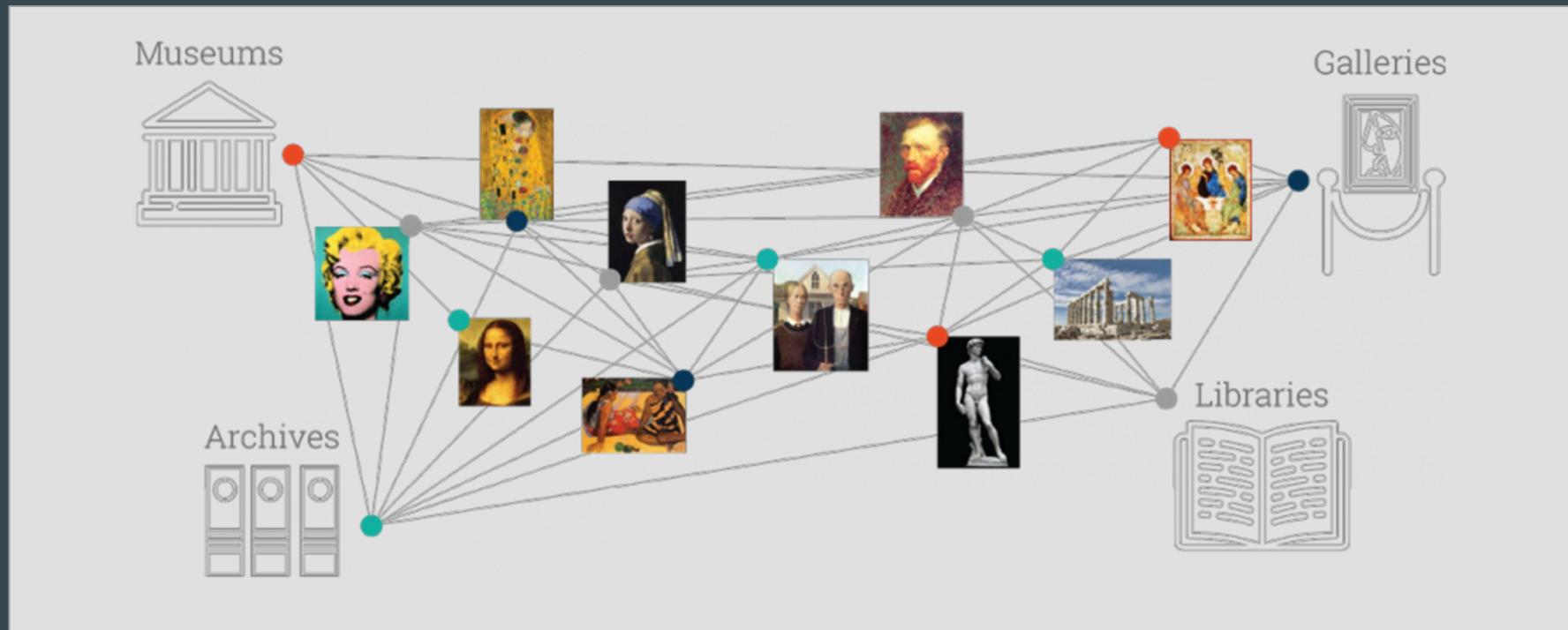


What are ontologies?

- The URIs are defined in vocabularies or ontologies.
- Processable files that enable us to define the complexity of the dataset.
- They define concepts, data categories and connections between data categories.
- File read by a piece of software.



LODLAM (LOD in Libraries, Archives and Museums)



What is LODLAM?

- Linked Open Data in Libraries, Archives and Museums.
 - Cultural institutions are recognising Linked Data and Open Access in their Digital strategies.
 - Wide range initiatives to generate, publish and sustain LOD.
 - Emerging from domain-relevant use cases.
 - Requires experimentation àorganic and dynamic process.
-
- What are the relevant elements? We need the right metadata.
 - How do we connect these entities? We need the right relationships.

LODLAM Motivations

- Interlink cultural information from a variety of sources.
- Aimed to achieve semantic interoperability in cultural heritage.
- Allows you to ask bigger questions that couldn't be asked from individual sites/objects.
- Reaches across museums, libraries and archives.
- International and multidisciplinary approach.
- Involves very different domains (collection description, archives, administration, science, scholarship, presentation..)

LOD resources

PLEIADES



Geonames

GeoNames About ▾ Browse ▾ Download ▾ API ▾ Help ▾ Paris, Mount Everest, New York Search X +

anonymous ▾

Map Satellite

Layers Layers

Refresh Refresh

Zoom In Zoom In

Zoom Out Zoom Out

Coordinates: N 37°13'20" W 5°06'49"

6361023

Osuna ca. 252 m
A ADM3 third-order administrative division
Spain ES » Andalucía SE » Seville SE » Osuna 41068
population : 17973
37.22229, -5.11371

geotree .kml .rdf

1

Map showing the satellite view of the area around Osuna, Spain. A callout box provides detailed information about the location, including its name, elevation, administrative status, population, coordinates, and download links for geotree, .kml, and .rdf formats. The map also displays nearby towns like Lantejuela, Complejo, Marinaleda, El Rubio, Herrera, La Mina, and others, along with roads labeled A-92 and A-318.

Pleiades

Log in

PLEIADES

Home Places Credits Participate Blog Documentation Downloads

You are here: Home → Ancient Places → Urso/Col. Genetiva Iulia

Urso/Col. Genetiva Iulia

a Pleiades place resource

Creators: Jr., F.H. Stanley, R.C. Knapp
Contributors: Brady Kaeling, Sean Gillies, Johan Åhlén, Jeffrey Becker, Tom Elliott, DARMC, R. Warner, R. Talbert, María Jesús Redondo
Copyright © The Contributors. Sharing and remixing permitted under terms of the Creative Commons Attribution 3.0 License (cc-by).
Last modified Mar 09, 2024 02:05 PM — History

tags: dare:ancient=1, dare:major=1, dare:feature=major_settlement

At the modern town of Osuna in Spain's Seville province, archaeological investigation has revealed an extensive Roman-era settlement. In 2022, an earlier "Phoenician" necropolis was discovered.

Canonical URI for this page:
<https://pleiades.stoa.org/places/256503>

Representative Point (Latitude, Longitude):
37.246621, -5.099275

Locations:

- Representative Locations:
 - DARE Location (330 BC - AD 640) accuracy: +/- 10 meters.

Names:

- Geographic Names:
 - Col. *Genetiva Iulia* (330 BC - AD 640)
 - Osuna (Spanish, modern)
 - Osuna (Spanish, AD 640 - AD 1453)
 - Vrso (Urso; Latin, 330 BC - AD 640)

Urso/Col. Genetiva Iulia makes connections with:
None

Urso/Col. Genetiva Iulia receives connections from:
Urso theater → located at → Urso/Col. Genetiva Iulia (unspecified date range)

Place type:
settlement

References:

- Evidence:
 - CIL II, 5439

Search

Search Site

Advanced Search...

Photos

This place has no portrait photo. One from the list of related photos could be suggested for the Pleiades Places group on Flickr.

0 other related photos...

Use this tag in Flickr to mark depictions of this place's site(s):
`pleiades:depict=256503`

or this one to mark objects found here:
`pleiades:findspot=256503`



200 m

mapbox

Display location accuracy buffer(s)
Show place in Google Earth.
Show area in Geohnames, Google Maps, or OpenStreetMap.

Trismegistos

Th

Places

Attestations (GeoRef)

About

Report an error

Urso (Osuna)

TM Geo 26236

Spain (Hispania) - provincia: Baetica - modern region: Andalucía -
Conventus Astigitanus (admin.) - located in the provincia Sevilla

37.246621,-5.099275 (Lat,Long)

Links: Digital Atlas of the Roman Empire, EDH, EDH [2], EDH [3], EDH [4], EDH [5], EDH [6], EDH [7], EDH [8], EDH [9], EDH [10], EDH [11]; and further links, Nomisma, Pleiades, ToposText, Wikipedia, Wikipedia [2], Wikipedia [3]

Status: city: oppidum; civitas; colonia

Latin name(s): Urso - Ursao - Genetiva Urbanorum

Ethnic(s): Ursonensis - Ursaonensis - Genetivus

Modern name(s): Osuna

Variants: Urso (Ursao) - Colonia Genetiva Iulia - Genetiva Urbanorum - Osuna

Map: Barrington Atlas, 2000, pl. 26 E4

Arachne Idai Objects

Osuna, Relief, Sitzender Stier

Museo Arqueológico Nacional, Madrid

arachne.dainst.org/entity/1146879

Arachne ID:	1146879
Category:	Einzelobjekte
Old serial number:	153469

Informationen zum Objekt

Lokalisierung

Museo Arqueológico Nacional, Madrid, Spanien,
Art der Ortsangabe: Aufbewahrungsort

Herkunft

Spanien, Osuna / Sevilla(P) / Andalucía, 1903, Aus den Ausgrabungen von Arthur Engel und Pierre Paris.
Der Stein fand sich nicht an seinem ursprünglichen Ort, sondern war als Spolie in einer römischen Mauer verbaut.

Datierung

Eisenzeit II (Chronontology) 2. Hälfte, 5. Jh. v. Chr.-4. Jh. v. Chr./ nach: Martín Almagro-Gorbea

Gattung/Funktion/Kulturepoche

Relief

Grab

Grabbau: Relief

Kulturepoche: iberisch

Antiker Aufstellungsort: vermutlich stammt der Stein von einem iberischen Grabbau

Antike Römische Provinz: Baetica

Image



Linked objects (5)

Literatur	3
Orte	1

Idai Chronontology

≡ IDAI chronontology

Roman Republic

<http://chronontology.dainst.org/period/YRaaGdvlgwMF>

The timeline chart displays various historical periods along a horizontal axis. The Roman Republic period is shown in purple, spanning from approximately -500 to 200 BC. Other periods visible include the Qatabanian, Pre-Roman Iron Age, Late Iron Age, Middle History, Roman Iron Age, Roman Empire, and Proto-Roman Iron Age. The chart uses a color-coded legend where green represents the Roman Republic period.

Period	Approximate Dates
Qatabanian	-600 to -500 BC
Pre-Roman Iron Age	-500 to -200 BC
Late Iron Age	-500 to -200 BC
Middle History	-500 to -200 BC
Roman Iron Age	-500 to -200 BC
Roman Empire	-500 to -200 BC
Proto-Roman Iron Age	-500 to -200 BC
Roman Republic	-500 to -200 BC
Roman Republic	-500 to -200 BC
Roman Republic	-500 to -200 BC
50 B.C.	-500 to -200 BC
Late Iron Age	-500 to -200 BC
Pre-Roman Iron Age	-500 to -200 BC

Information [Download JSON](#)

About the period

Names Roman Republic (en)

Period type [PeriodO](#) (p083p5rj8d7)

Digital provenance [PeriodO](#) (p083p5rj8d7)

Temporal extent Begin: 508 BCE
End: 26 BCE
Source: Gates, Charles, 1950: Ancient cities : the archaeology of urban life in the Ancient Near East and Egypt, Greece and Rome
Original text: ca. 509 BC - 27 BC

Spatially part of region <http://dbpedia.org/resource/Lazio>

No map data available

Leaflet | © OpenStreetMap contributors

Nomisma



Browse IDs About Who We Are Research Tools Documentation Get Data

Search IDs

Urso (Mint, Concept)

Canonical URI: <http://nomisma.org/id/urso>

Labels

Preferred Label Urso (*en*), Osuna (*fr*), Osuna (*es*), Osuna (*it*), Osuna (*de*) Additional labels ▾

Definitions

en The mint at the ancient site of Urso in Hispania Ulterior.

Geospatial Data

URI: <http://nomisma.org/id/urso#this>

Latitude 37.236984

Longitude -5.102780

Part Of <http://nomisma.org/id/hispania#this>

Relations

Broader Concept <http://nomisma.org/id/hispania>

Close Match <http://catalogue.bnf.fr/ark:/12148/cb11937275m>

Close Match <http://collection.brictishmuseum.org/id/place/x117226>

Close Match <http://dbpedia.org/resource/Osuna>

Close Match <http://sws.geonames.org/2513052/>

Close Match <http://vifaf.org/vifaf/136075560>

Close Match <http://vocab.getty.edu/tgn/7008228>

Close Match <http://www.wikidata.org/entity/Q586018>

Close Match <https://ikmk.smb.museum/ndp/ort/2655>

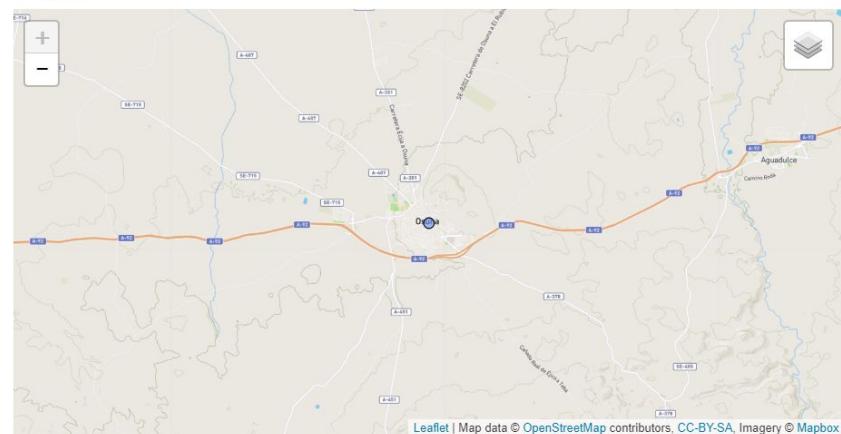
Close Match <https://pleiades.stoa.org/places/256503>

Close Match <https://www.firebaseio.com/m/07grq6>

Export

Linked Data GitHub File RDF/XML RDF/TTL JSON-LD

Geographic Data KML GeoJSON



SPARQL protocol and RDF query language.

- SQL query language for RDF data
- It allows the user to ask questions to the dataset at issue
- SPARQL queries search the dataset for the specific subset of data that meets certain conditions.
- The anatomy of SPARQL queries is constituted by several sections which require a certain order.

Nomisma SPARQL endpoint

SPARQL Endpoint

For examples, see [SPARQL Examples](#). A basic tutorial on SPARQL is available from Apache Jena.

```
PREFIX crm: <http://www.cidoc-crm.org/cidoc-crm/>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX dcatype: <http://purl.org/dc/dcatype/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>
PREFIX nm: <http://nomisma.org/id/>
PREFIX nmo: <http://nomisma.org/ontology#>
PREFIX org: <http://www.w3.org/ns/org#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

SELECT * WHERE {
  ?s ?p ?o
} LIMIT 100
```

Additional prefixes

bio	crmarchaeo	crmdig	crmgeo	crmsci	osgeo	prov	rdac	sd
spatial	un	void	wordnet					

Query structure:

List of namespace definitions.

PREFIXES

Query form + Variables

SELECT, CONSTRUCT, ASK, DESCRIBE

Query patterns and filters

WHERE

Solution modifiers

ORDER BY, LIMIT, OFFSET

Understand the RDF

```
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#> .
@prefix nm: <http://nomisma.org/id/> .
@prefix nmo: <http://nomisma.org/ontology#> .
@prefix prov: <http://www.w3.org/ns/prov#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

nm:urso a nmo:Mint,
    skos:Concept ;
    dcterms:isPartOf nm:greek_numismatics ;
    geo:location <http://nomisma.org/id/urso#this> ;
    skos:broader nm:hispania ;
    skos:changeNote <http://nomisma.org/id/urso#provenance> ;
    skos:closeMatch <http://catalogue.bnf.fr/ark:/12148/cb11937275m>,
        <http://collection.britishmuseum.org/id/place/x117226>,
        <http://dbpedia.org/resource/Osuna>,
        <http://sws.geonames.org/2513052/>,
        <http://viaf.org/viaf/136075560>,
        <http://vocab.getty.edu/tgn/7008228>,
        <http://www.wikidata.org/entity/Q586018>,
        <https://ikmk.smb.museum/ndp/ort/2655>,
        <https://pleiades.stoa.org/places/256503>,
        <https://www.freebase.com/m/07grq6> ;
    skos:definition "The mint at the ancient site of Urso in Hispania Ulterior."@en ;
    skos:inScheme nm: ;
    skos:prefLabel "Osuna"@an,
        "أشونة"@ar,
        "Osuna"@ca,
        "Osuna"@cs,
```

Query

PREFIX nmo: <<http://nomisma.org/ontology#>>

SELECT * WHERE {

?s a nmo:Mint

} LIMIT 100

Sparql Exercise

1. Go to the Nomisma database and look at the RDF behind two or three mints from different provinces. How much of the RDF can you understand (or guess at)?
2. Now visit the [Nomisma.org triplestore](#) and explore their data.
3. Using some of their [sample Sparql queries](#) and the knowledge you have acquired, try to come up with some queries to explore their data a little more.
4. Can you understand what is going on in the sample queries they provide?



Item [Discussion](#)

Read

[View history](#)

Search Wikidata



Osuna (Q586018)

municipality in Seville, Spain

[edit](#)

▼ In more languages

Configure

Language	Label	Description	Also known as
English	Osuna	municipality in Seville, Spain	
Spanish	Osuna	municipio de la provincia de Sevilla, España	
French	Osuna	commune espagnole	
German	Osuna	Gemeinde in Spanien	

All entered languages

Statements

instance of



municipality of Spain

[edit](#)

▼ 0 references

+ add reference

+ add value

part of



Q100593719

[edit](#)

► 1 reference

Wikipedia (51 entries) [edit](#)

an Osuna

ar أشونة

arz اشونه

ast Osuna

azb اسونا

bg Осуна

br Osuna

ca Osuna

ceb Osuna (munisipyo)

ce Осуна

dag Osuna

de Osuna (Sevilla)

diq Osuna

en Osuna

eo Osuna

es Osuna

eu Osuna (Sevilla)

fa اسونا

fr Osuna

ga Osuna

gl Osuna

hu Osuna

hy Օսոնա

Wikidata query service

 Do you need help creating a query? You can build queries without having to write SPARQL in the new [Query Builder](#).



Wikidata Query Service

Examples

Help

More tools

Query Builder

1 #Map of the locations of all paintings by Johannes Vermeer with an image
2 #defaultView:Map
3 **SELECT** ?painting ?paintingLabel ?location ?image **where** {
4 ?painting wdt:P276 ?collection;
5 wdt:P170 wd:Q41264;
6 wdt:P18 ?image.
7 ?collection wdt:P625 ?location.
8 **SERVICE** wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
9 }

Data reconciliation with OpenRefine

What is Data Reconciliation?

Data reconciliation involves matching records in your dataset to unique identifiers in a standardized dataset, such as a database of geographical locations, bibliographic records, or any controlled vocabulary. This helps in ensuring data consistency and accuracy.

Steps previous to reconciliation

1. Install OpenRefine
2. Launch OpenRefine
3. Load your data
4. Create a new project
5. Preview and Parse Data

Data reconciliation with OpenRefine

Reconciling Data

1. Choose a Column to Reconcile

- Identify the column in your dataset that you want to reconcile with an external dataset. For example, if you have a column of city names, you might want to reconcile this with a database of standardized city identifiers.

2. Start Reconciliation

- Click on the arrow next to the column name and select "Reconcile" > "Start reconciling."

3. Select a Reconciliation Service

- OpenRefine provides built-in reconciliation services and allows you to add custom services. For this example, let's use the Wikidata reconciliation service:

- If it's not already listed, you can add it by selecting "Add Standard Service" and entering <https://wikidata.reconcil.link/en/api>.

Data reconciliation with OpenRefine

4. Configure Reconciliation Settings

1. Choose the type of data you are reconciling against. For city names, you might select "Q515" (the Wikidata item type for cities).
2. Click "Start Reconciling."

5. Examine Suggested Matches

- OpenRefine will attempt to match your data to entries in the external dataset. The results will be displayed in the column with a match score.
- Review these suggestions. Matches with higher scores are more likely to be correct.

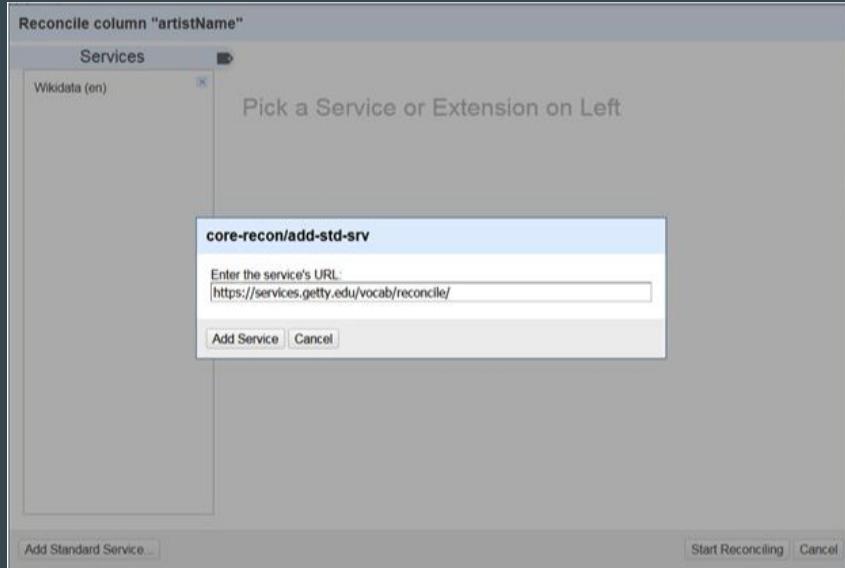
6. Manually Confirm or Correct Matches

- Click on a cell to see more details about the suggested match.
- Confirm the match, select a different match, or mark it as "None of the above" if no match is correct.

Export reconciled data from OpenRefine

1. Generate a new column with reconciled IDs
2. Explore the URIs
3. Merge in a separate column the URI body with the ID
4. Now you have a column with the new extracted URIs

Set up Getty Vocabularies reconciliation service



<https://services.getty.edu/vocab/reconcile/>