

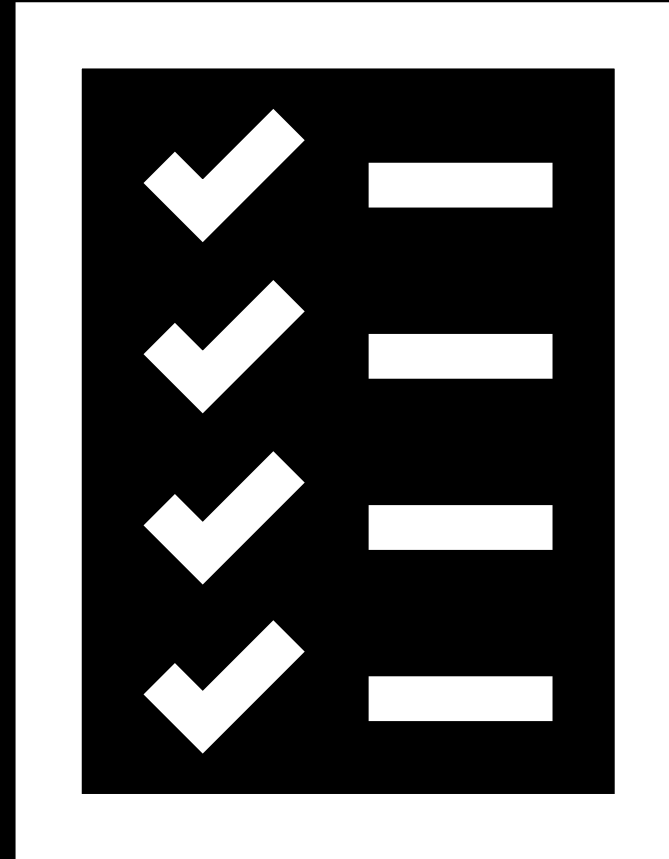
SunoikisisDC Digital Approaches to Cultural Heritage, Spring 2024

Session 9: Linked Open Data: Digital Collections

Paula Granados García (EMKP)

Contents:

1. Linked Open Data
2. Basic concepts
3. Applications
4. LOD resources



Main concepts:

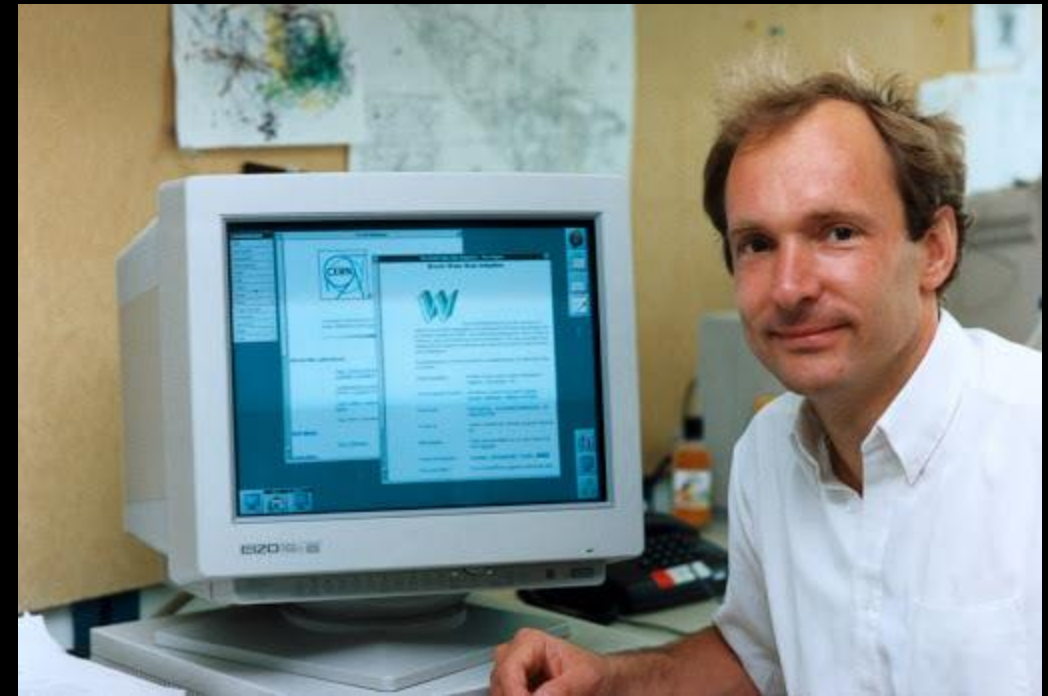
- Linked Open Data: theory, ecosystem, technologies and practices around openly licensed structured data interlinked with other data.
- Linked open data: specific datasets or files published on the web.
- Semantic Web: common space where data can be shared, processed, linked and reused.

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.



LOD mug



Tim Berners Lee, father of the web



A triple

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?

1. Accessibility

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

2. Openness Standards

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

3. Interoperability.

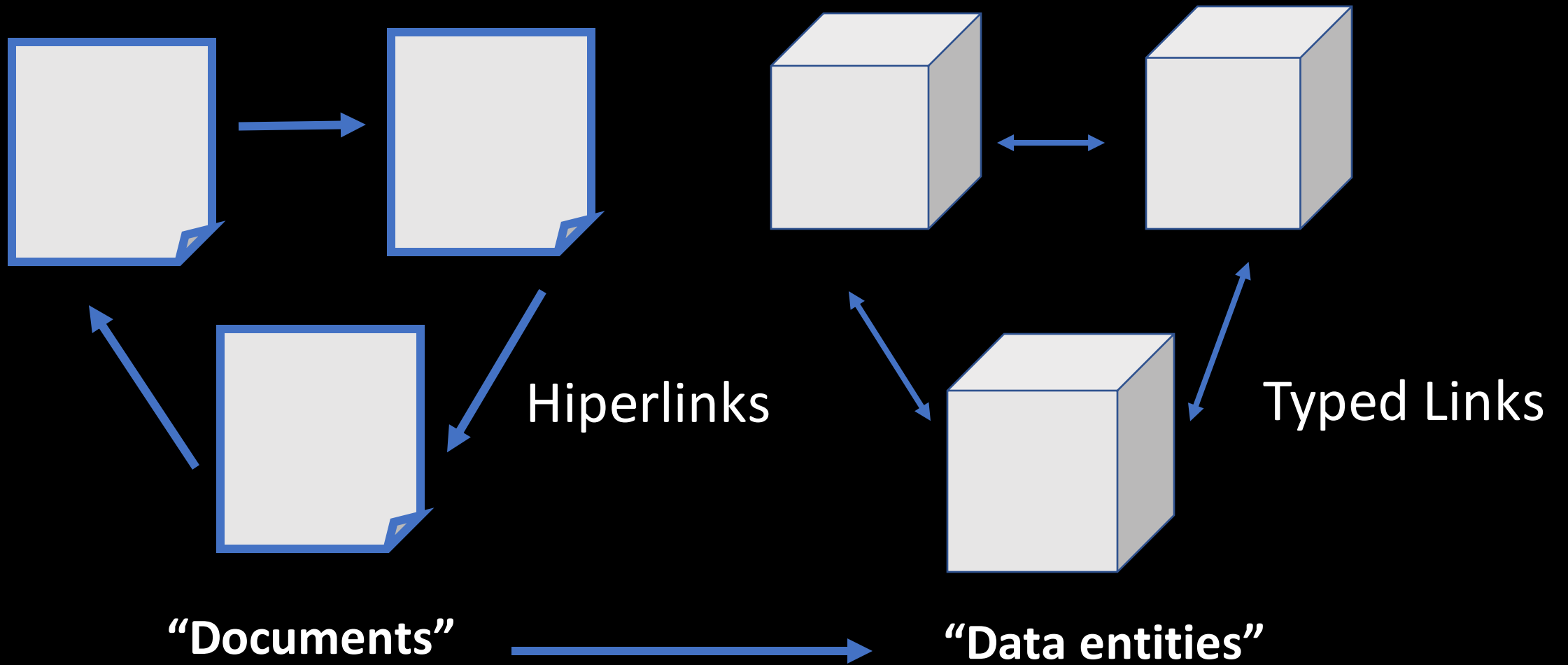
- Using HTTP URIs to interlink information.

4. Contextualisation. Granularity of detail.

- All data to be searched for and referenceable.

5. Reutilisation. Open Ended Potential.

The web of data



Linked Open Data recipe:

Model your
data
Using
existing
vocabularies

Ontologies

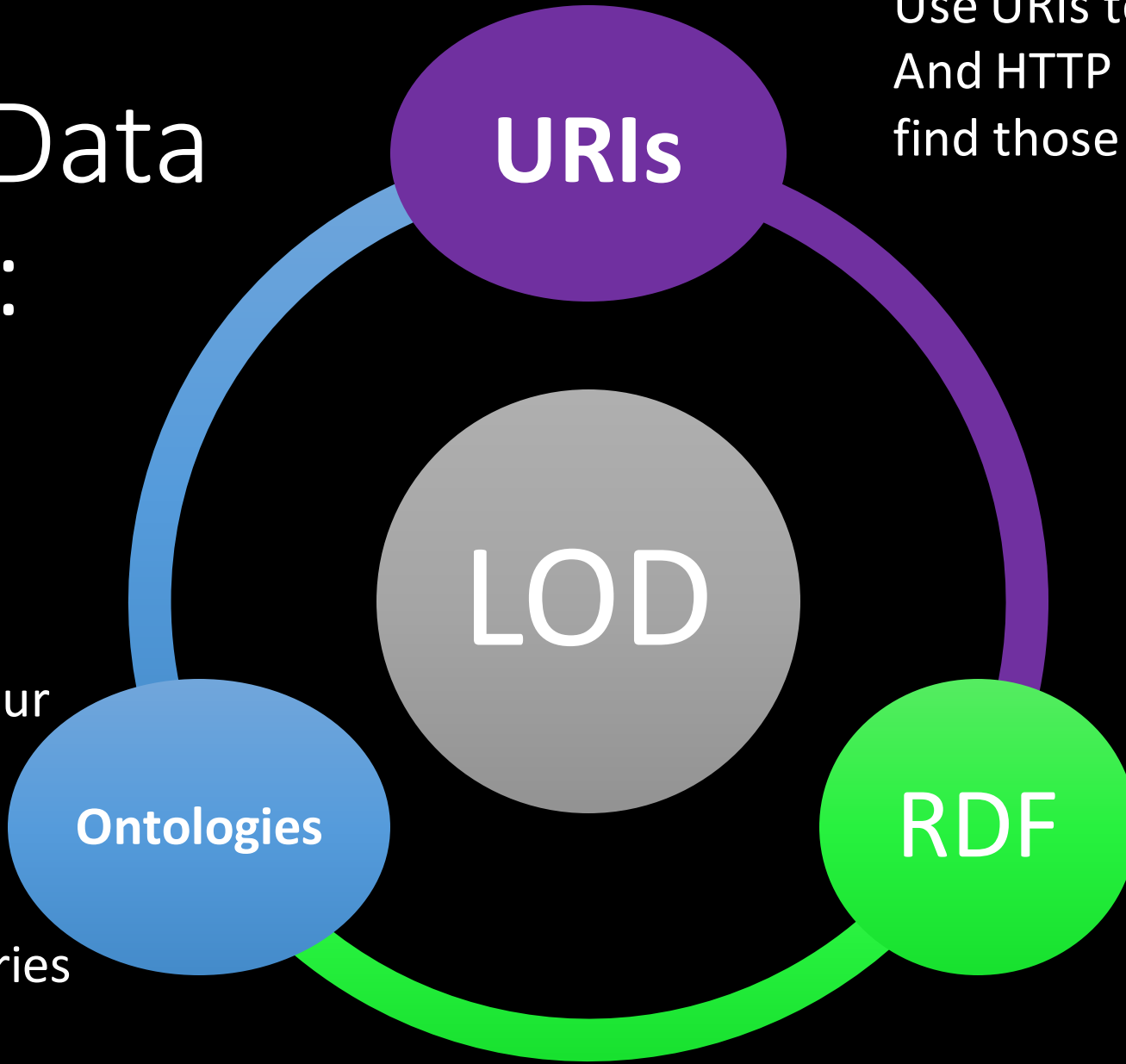
LOD

URIs

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

RDF

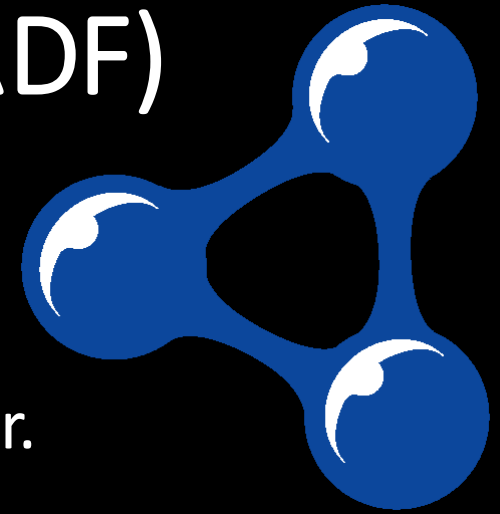
Provide useful RDF data
and include RDF
statements that link to
other data



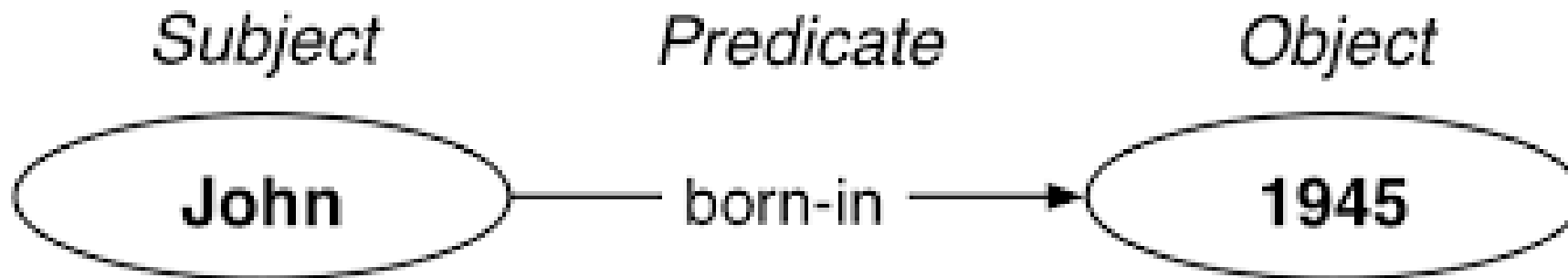
URIs

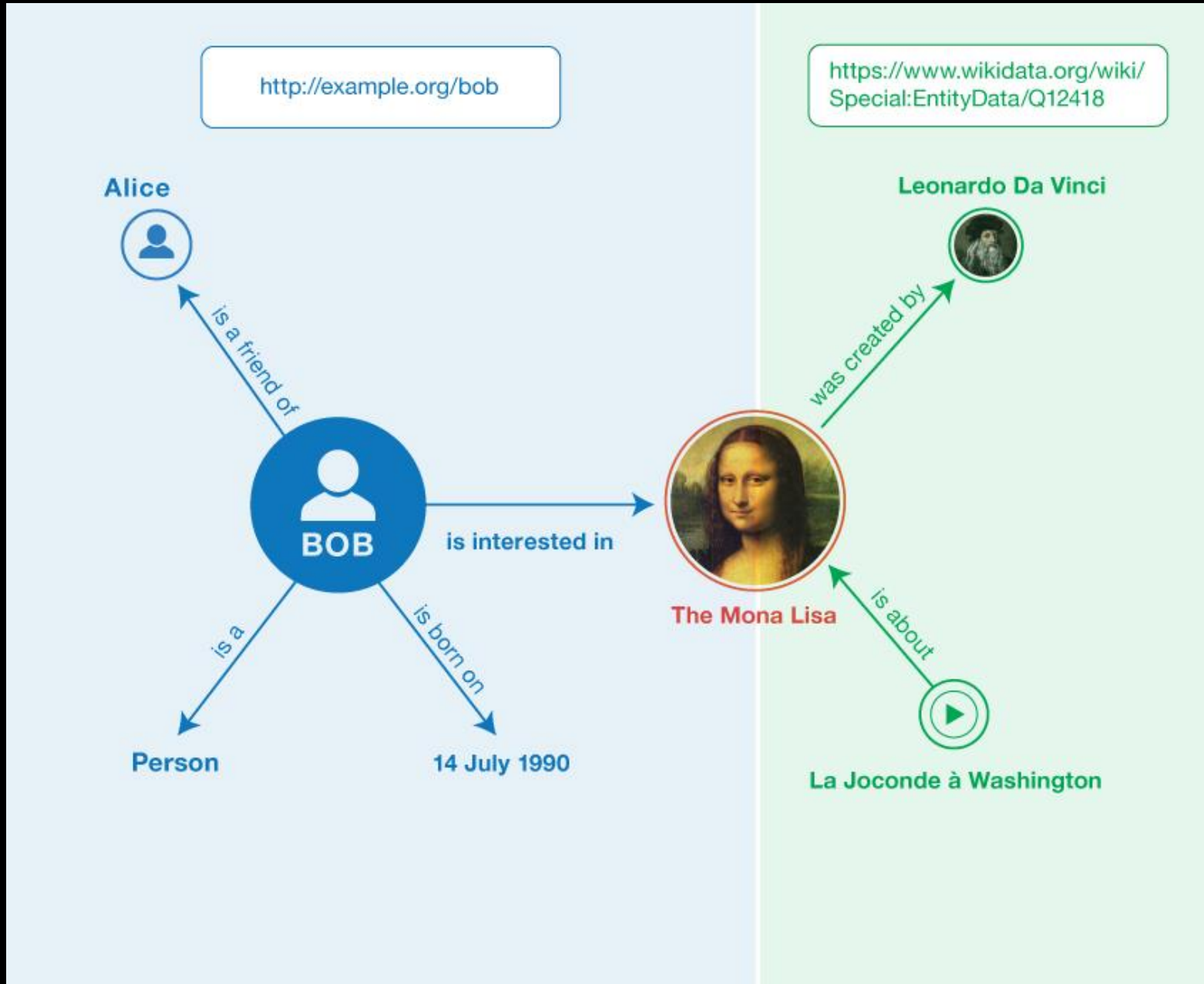
- URI (Uniform Resource Identifier)
 - A type of identifier.
 - String that can be used to refer to a resource on the web, an email address, a telephone number etc..
 - Should be dereferenceable (can be looked up by a browser)
- HTTP URIs (used to refer to entities)
 - http:// (Identifier in the scheme HTTP)
 - “namespace” = web domain = guarantees its uniqueness
- 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.





What are ontologies?

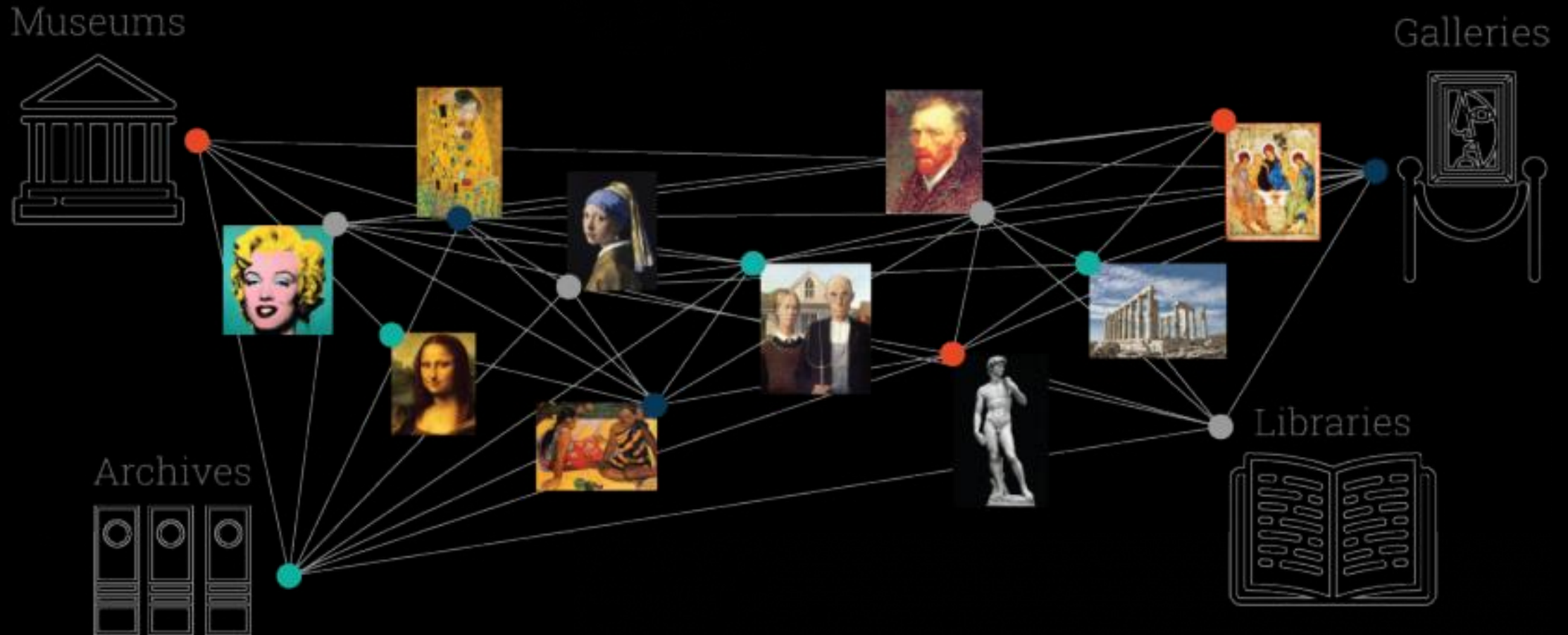
- The URIs are defined in vocabularies or ontologies.
- Processable files that enable us to define the complexity of the dataset.
- They define concepts, data categories and connections between data categories.
- File read by a piece of software.



Concept



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 → organic 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 pieces.
- Reaches across museums, libraries and archives.
- International and multi-disciplinary approach.
- Involves very different domains (collection description, archives, administration, science, scholarship, presentation..)

Resources

Evaluate your resources

1. Identification of relevant data sources

A light blue downward-pointing arrow with a white outline, indicating the flow from step 1 to step 2.

2. Surveying of data sources

A light blue downward-pointing arrow with a white outline, indicating the flow from step 2 to step 3.

3. Data extraction

| <i>Resource</i> | <i>Type</i> | <i>URL</i> | <i>Available in Open Access</i> | <i>Machine readable structured data</i> | <i>Non- proprietary format</i> | <i>W3C Recommendations</i> | <i>Links to similar resources</i> |
|-----------------|---------------------------|---|---|---|--|--------------------------------|---|
| Pleiades | GAZETEERS | https://pleiades.stoa.org/ | ★ | ★ | ★ | ★ | ★ |
| DARE | | https://imperium.ahlfeldt.se/ | ★ | ★ | ★ | ★ | ★ |
| TM Places | | http://www.trismegistos.org | ★ | ★ | ★ | ★ | ★ |
| EDH | DATABASES | http://edh-www.adw.uni-heidelberg.de/ | ★ | ★ | ★ | ★ | ★ |
| ResearchSpace | | https://public.researchspace.org/resource/Start | ★ | ★ | ★ | ★ | |
| Arachne | | https://arachne.dainst.org | ★ | ★ | ★ | ★ | |
| Nomisma | COLLABORATIVE PROJECTS | http://nomisma.org/ | ★ | ★ | ★ | ★ | ★ |
| Pelagios | | http://commons.pelagios.org/ | ★ | ★ | ★ | ★ | ★ |



Paula Granados García (EMKP)

pgranadosgarcia@britishmuseum.org

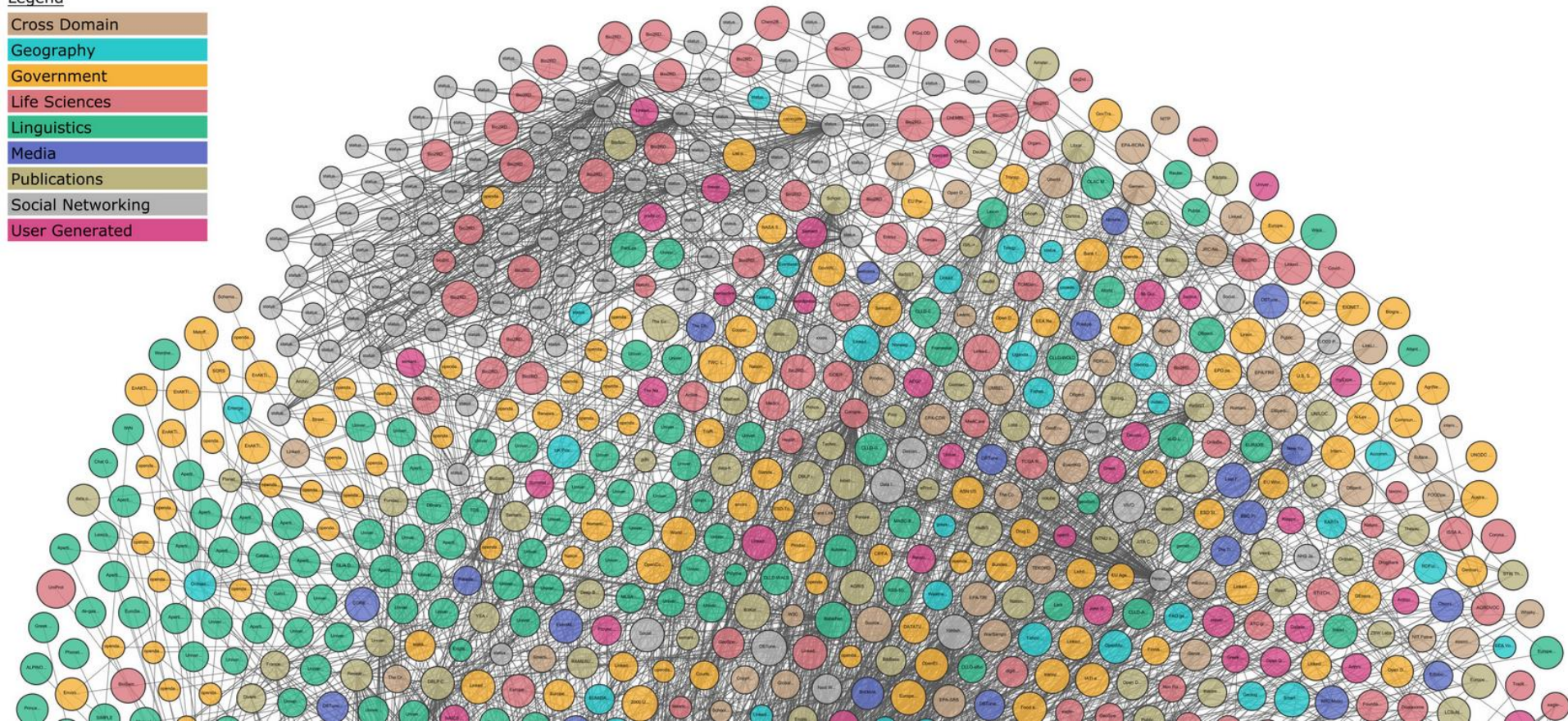
Linked Open Data: Digital collections

SunoikisisDC Digital Approaches to Cultural Heritage:
session 9

Vera Moitinho de Almeida, PhD
vmoitinho@letras.up.pt

The Linked Open Data Cloud

Legend





VIAF
Virtual International Authority File



Kunsthistorisches Museum ca. 194 m
Bécsi Szépművészeti Múzeum Kunsthistorisches
Austria • Vienna • Vienna • Inner city
museum
N 48° 12' 13" E 16° 21' 40"
48.20371 / 16.36126
GeoNameId: 6354990

[view map](#) [edit history](#) [view details](#) [alternates names](#)
[perma link](#) [perma semantic web id](#)
[part of](#) [contains](#)

| Name | Language Code | p | s | h | c | Action |
|---|---------------|-----------------|---|---|---|---|
| 빈 미술사 박물관 | ko | Korean | | | | edit / delete |
| พิพิธภัณฑ์ศิลปวัฒนธรรม | th | Thai | | | | edit / delete |
| Bécsi Szépművészeti Múzeum | hu | Hungarian | | | | edit / delete |
| http://en.wikipedia.org/wiki/Kunsthistorisches_Museum | link | link to website | | | | edit / delete |
| http://ru.wikipedia.org/wiki/%D0%9C%D1%83%D0%B7%D0%85%D0%B9_%D0%88%D1%81%D1%82%D0%B6%D1%80%D0%88%D0%88_%D0%88%D1%81%D0%8A%D1%83%D1%81%D1%81%D1%82%D0%82_%28%D0%92%D0%85%D0%BD%D0%80%29 | link | link to website | | | | edit / delete |
| Kunsthistorisches Museum Wien | nl | Dutch | | | | edit / delete |
| Kunsthistorisk Museum Wien | da | Danish | | | | edit / delete |
| Meno istorijos muziejus | lt | Lithuanian | | | | edit / delete |
| Musée d'histoire de l'art de Vienne | fr | French | | | | edit / delete |
| Museo de Historia del Arte de Viena | es | Spanish | | | | edit / delete |
| Museu de História da Arte em Viena | pt | Portuguese | | | | edit / delete |

Creators: E.W. Haley
Contributors: DARRAC, R. Talbert, Brady Kiesling, Sean Gillies, Johan Ahlstedt, Jeffrey Becker, Tom Elliott
Copyright © The Contributors. Sharing and reusing permitted under terms of the Creative Commons Attribution 3.0 License (cc-by).
Last modified Jun 07, 2018 05:11 PM — [History](#)

tags: [dare:ancient=1](#), [dare:major=1](#), [dare:feature=major settlement](#)

With settlement from the Neolithic onwards, Bracara Augusta was established as a civitas in 20 B.C. In the third century A.D. Diocletian advanced the settlement to the capital of the administrative area known as Conventus Bracarensis, forming part of the Roman province of Gallaecia.

Canonical URI for this page:
<https://pleiades.stoa.org/places/236377>

Representative Point (Latitude, Longitude):
41.550875, -8.424973

Locations:

- DARMC Location 20067 (30 BC - AD 640)
- DARE Location (30 BC - AD 640)

Names:

- Bracara Augusta** (30 BC - AD 640)
- Bracari** (30 BC - AD 640)

Bracara Augusta makes connections with:

- Bracara Augusta → [connection](#) → Callaecia (unspecified date range)

Research Home » Tools » Art & Architecture Thesaurus » Full Record Display

Art & Architecture Thesaurus® Online
Full Record Display

[View Results](#) [Previous Page](#) [Help](#)

Click the icon to view the hierarchy.

Semantic View (JSON RDF N3 Turtle N-Triples) **Representative Images: 1**

ID: 300198866
Page Link: <http://vocab.getty.edu/page/aat/300198866>
Record Type: concept

oinochoai (vessels (containers), <containers by form>), ... Furnishings and Equipment (hierarchy name)

Note: Ancient Greek one-handed vessels used for lading and pouring wine or water; made in a variety of jug- and pitcherlike forms.

Terms:

- oinochoai (preferred C.U. English P.O.U.P.) (Dutch P.O.U.U.)
- oinochoe (C.U. English AB.U.S.N.) (Dutch AB.U.U.)
- oinochoai (C.U. English UF.U.N.)
- oinochoes (C.U. English UF.U.N.)
- oinochoe (C.U. English AB.U.S.N.)
- oinochoe (C.U. Spanish AB.U.S.N.)

Facet/Hierarchy Code: VTD

Hierarchical Position:

- Objects Facet
- Furnishings and Equipment (hierarchy name) (Q)
- containers (hierarchy name) (Q)
- containers (specifications) (Q)
- <containers by form> (Q)
- vessels (containers) (Q)
- oinochoai (Q)

Additional Notes:

Dutch ... Klassieke Griekse vaten met één handvat, die werden gebruikt voor het scheppen en schenken van wijn of water. Gemaakt in een reeks van vormen met het uiterlijk van kannen en kruiken.

Spanish ... Vasijas de una sola asa de la Antigua Grecia usadas para servir y verter el vino o el agua; hechas en una variedad de formas similares a jarros o cántaros.

Related concepts:

- [vra \(vessels\)](#) (vessels (containers), <containers by form>), ... Furnishings and Equipment (hierarchy name) (300043683)
- [pitchers \(vessels\)](#) (vessels (containers), <containers by form>), ... Furnishings and Equipment

SPARQL Queries [Any](#) [Search](#) [Brief](#) [+](#)

oinochoai

Source: <http://vocab.getty.edu/aat/300198866>

[Subject \(RZ\)](#) [Predicate](#) [Object](#) [All](#) [Website](#) [Hierarchy](#) [Download to: JSON](#) [JSONLD](#) [RDF](#) [N3/Turtle](#) [N-Triples](#)

Inference: [Explicit only](#)

Statements in which the resource exists as a subject.

| Predicate | Object |
|------------------------|--|
| rdf:type | gdp:Concept |
| rdfs:seeAlso | http://www.getty.edu/view/AT/AT/Display?find=&lang=&ANDNode=&subject=300198866 |
| dcterms:created | 1991-10-02T00:00:00 |
| dcterms:modified | 2001-07-26T22:03:39, 2001-07-26T22:15:13, 2007-06-14T13:59:41, 2007-07-06T16:01:01, 2009-06-02T14:44:23, 2010-04-23T06:54:29, 2010-04-23T06:54:30, 2011-01-07T12:19:15, 2011-11-14T13:26:15, 2012-07-20T16:18:05 |
| skos:changeNote | aat_rev3000057238, aat_rev5001145989, aat_rev5001145990, aat_rev5001145991, aat_rev5001145992, aat_rev5001175407, aat_rev5001175408, aat_rev5001865918, aat_rev5001870480, aat_rev5002045439, aat_rev5002512511, aat_rev5002512512, aat_rev5002638566, aat_rev50030354443, aat_rev5003255112 |
| gdp:parentString | vessels (containers), <containers by form>, containers (specifications), Containers (hierarchy name), Furnishings and Equipment (hierarchy name), Objects Facet |
| gdp:parentStringabbrev | vessels (containers), <containers by form>, ... Objects Facet |
| gdp:displayOrder | 46 |
| xl:prefLabel | aat_term1000198866-en, aat_term1000198866-ru, aat_term1000296919-es |
| xl:altLabel | aat_term1000199037-en, aat_term1000234021-en, aat_term1000296919-en, aat_term1000296919-ru, aat_term1000444919-es, aat_term1000444980-es |
| gdp:broaderGeneric | aat:300193015 |

Search for

Late Geometric I B ceramic (Nóto Aigalo (Periféria))

<http://chronontology.dainst.org/period/121pCUP510>

Information [Download JSON](#)

About the period

Names Spätgeometrisch I B (de) Late Geometric I B (en)

Period type ceramic

Provenance chronontology

Temporal extent Original test: 750 - 735 spätgeometrisch I b Begin: -750 (ca) End: -735 (ca) Source <https://archaeologie-erfgoed.nl/worden>

Has core area Nóto Aigalo (Periféria) Bóreas Aigalo (Periféria) Elada

Relations

Is a sense of Late Geometric I B (alle Bedeutungen) Arachon

Is part of Late Geometric I

Follows Late Geometric I A

Derived information



- Main page
- Community portal
- Project chat
- Create a new item
- Recent changes
- Random item
- Query Service
- Nearby
- Help
- Donate
- Lexicographical data
- Create a new Lexeme
- Recent changes
- Random Lexeme
- Tools
- What links here
- Related changes
- Special pages
- Permanent link
- Page information
- Concept URI
- Cite this page
- Get shortened URL
- Download QR code

museum (Q33506)

institution that holds artifacts and other objects of scientific, artistic, cultural, historical, or other importance

museum-style institution | museal institution

▼ In more languages

Configure

| Language | Label | Description | Also known as |
|------------|---------------------|---|--|
| English | museum | institution that holds artifacts and other objects of scientific, artistic, cultural, historical, or other importance | museum-style institution <div></div> museal institution <div></div> |
| Portuguese | museu | instituição que conserva artefatos de importância científica, artística, cultural ou histórica | Museus <div></div> Instituição museológica <div></div> |
| French | musée | institution qui conserve, collecte et expose des objets dans un souci d'enseignement et de culture | centres d'interprétation <div></div> musee <div></div> musées <div></div> |
| Spanish | museo | institución que conserva y expone colecciones con valor cultural | museos <div></div> |

All entered languages

Statements

| | |
|---|--|
| <div>subclass of</div> | <div><div><div><div><div></div><div>GLAM</div></div></div><div><div><div></div><div>▼ 0 references</div></div></div></div></div> |
| <div><div><div><div><div></div><div>organization</div></div></div><div><div><div></div><div>▼ 0 references</div></div></div></div></div> | |
| <div><div><div><div><div></div><div>architectural structure</div></div></div><div><div><div></div><div>▼ 0 references</div></div></div></div></div> | |

Identifiers

| | |
|---|--|
| <div>National Library of Spain ID</div> | <div><div><div><div></div><div>XX525827</div></div></div><div><div><div></div><div>► 1 reference</div></div></div></div> |
| <div><div><div><div></div><div>Bibliothèque nationale de France ID</div></div></div></div> | <div><div><div><div></div><div>11939565m</div></div></div><div><div><div></div><div>▼ 0 references</div></div></div></div> |
| <div><div><div><div></div><div>GND ID</div></div></div></div> | <div><div><div><div></div><div>4040795-0</div></div></div><div><div><div></div><div>► 1 reference</div></div></div></div> |
| <div><div><div><div></div><div>FAST ID</div></div></div></div> | <div><div><div><div></div><div>1030128</div></div></div><div><div><div></div><div>► 1 reference</div></div></div></div> |
| <div><div><div><div></div><div>National Library of Israel J9U ID</div></div></div></div> | <div><div><div><div></div><div>987007553199805171</div></div></div><div><div><div></div><div>► 1 reference</div></div></div></div> |

Wikipedia

(142 entries)

- af Museum
- als Museum
- ang Boldæfþæcraeften
- an Museu
- arc ܡܘܨܝܐ
- ar متحف
- arz متحف
- ast Muséu
- az Muzey
- bar Museen
- ba Музей
- bcl Museo
- be_x_old Музей
- be Музей
- bg Музей
- bh মুজিয়ম
- bjn Musium
- blk ᠮᠤᠰᠤᠢᠮᠤ
- bn জাদুঘর
- bo མེང་གླིང་ཁང་།
- br Mirdi
- bs Muzej
- ca Museu
- chr ႭႭႭႭ ႭႭႭႭႭႭ
- ckb مۆزەخانه
- csb Mùzeum
- cs Muzeum
- cv Музей
- cy Amgueddfa

Late Bronze Age (Q1758757)

Last period of Bronze Age

Later Bronze Age

▼ In more languages

Configure

| Language | Label | Description | Also known as |
|------------|-------------------------------------|--|---|
| English | Late Bronze Age | Last period of Bronze Age | Later Bronze Age <div></div> |
| Portuguese | No label defined | No description defined | |
| French | Bronze final | dernière période de l'Âge du bronze | Bronze récent <div></div> Âge du bronze final <div></div> bronze final <div></div> bronze récente <div></div> |
| Spanish | Edad del Bronce tardío | periodo arqueológico | |

All entered languages

Statements

| | |
|------------------------|--|
| <div>instance of</div> | <div><div><div><div></div><div>archaeological period</div></div></div><div><div><div></div><div>▼ 0 references</div></div></div></div> <div><div><div></div><div>► 0 references</div></div></div> <div><div><div></div><div>► 0 references</div></div></div> |
|------------------------|--|

amphora (Q178401)

type of storage container

amphorae |

▼ In more languages

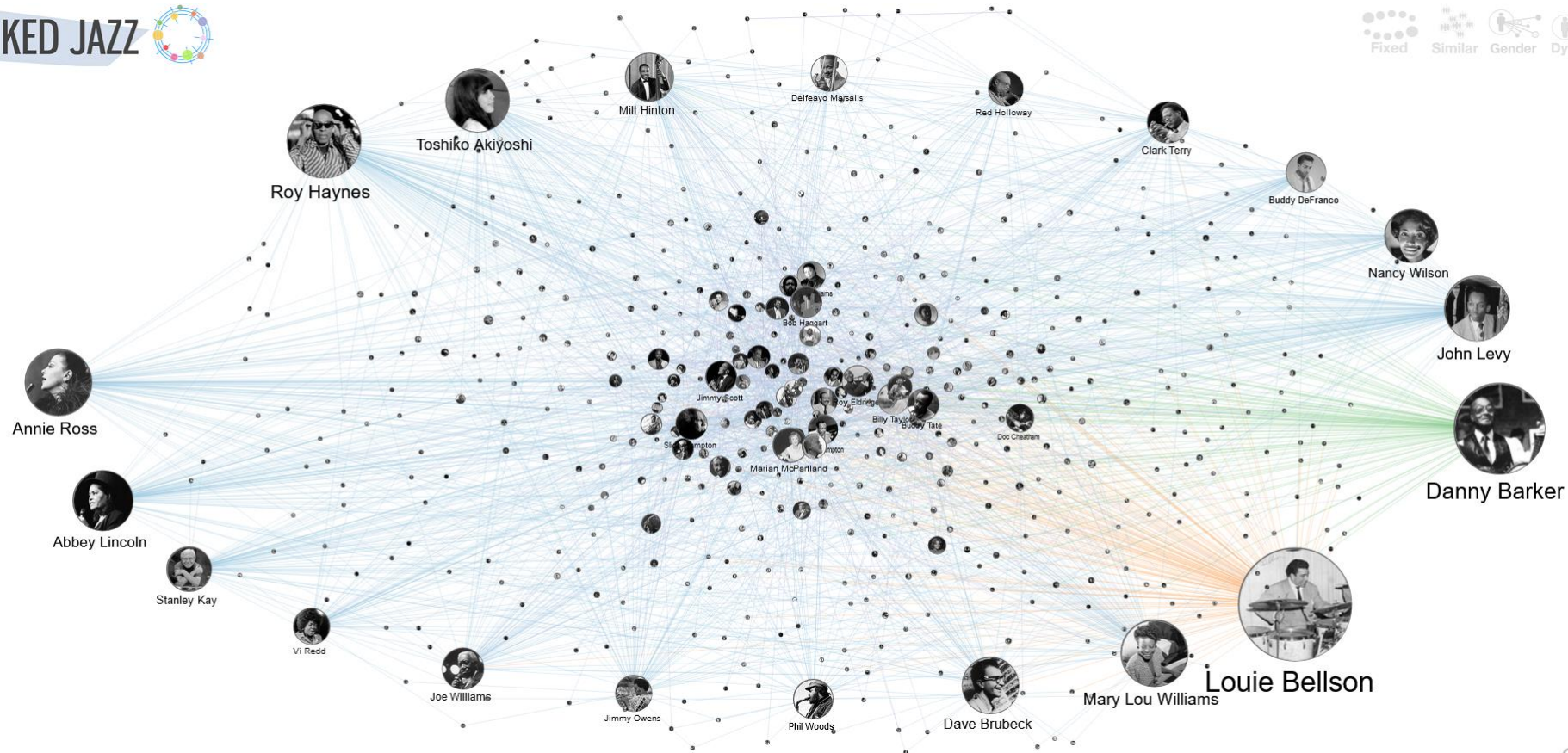
Configure

| Language | Label | Description | Also known as |
|------------|----------------------|---|---|
| English | amphora | type of storage container | amphorae <div></div> <div></div> |
| Portuguese | ânfora | No description defined | antôra <div></div> |
| French | amphore | récipient de stockage pourvu de deux anses | amphores <div></div> |
| Spanish | ánfora | Recipiente de cerámica, barro, etc., de cuello más o menos largo y delgado, generalmente con dos asas, y cuerpo ovalado u oblongo, que en las culturas antiguas se usaba para guardar y transportar líquidos o granos. | anfora <div></div> |

All entered languages

Statements

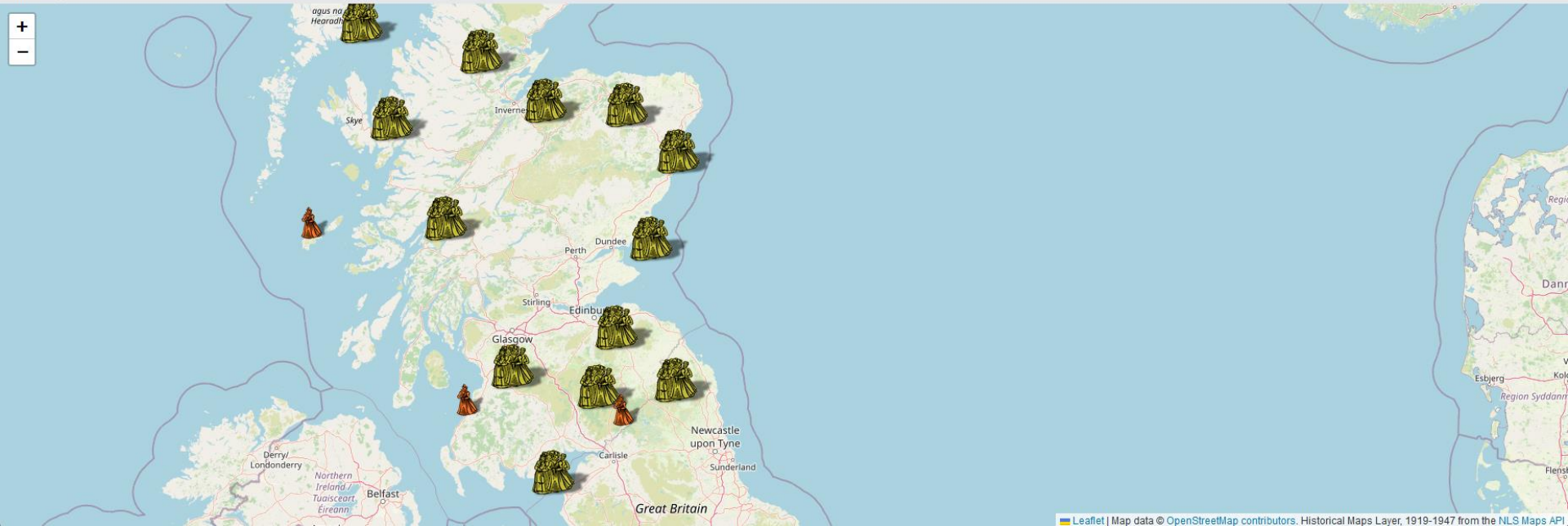
| | |
|------------------------|--|
| <div>instance of</div> | <div><div><div><div></div><div>archaeological artifact</div></div></div><div><div><div></div><div>▼ 0 references</div></div></div></div> <div><div><div></div><div>► 0 references</div></div></div> <div><div><div></div><div>► 0 references</div></div></div> |
|------------------------|--|



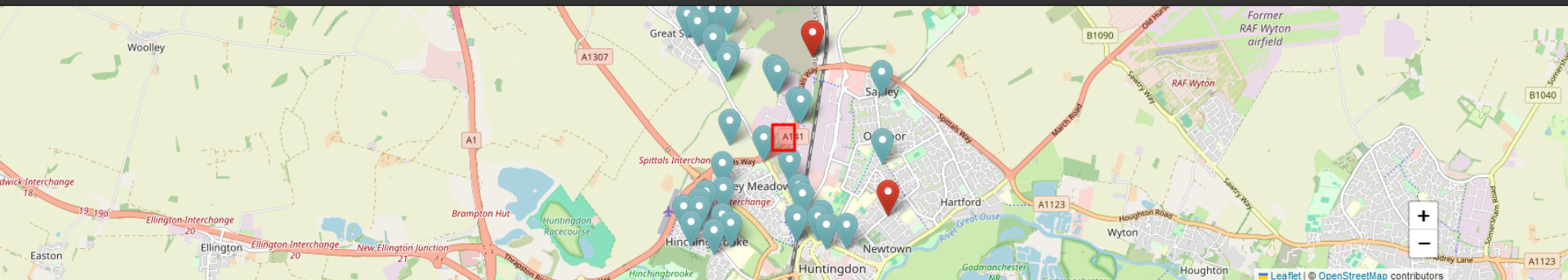


Places of Residence for Accused Witches (total named accused witches: 3141) ⓘ

Filters ▾



Back to search results



Current resource
Nearby resource (hide)

Images and GIS from an Archaeological Evaluation at Unit 2, St Johns Business Park, Huntingdon, Cambridgeshire 2022



Description

This collection comprises images and GIS data from an archaeological evaluation by Cotswold Archaeology at Unit 2, St Johns Business Park, Huntingdon, Cambridgeshire. The work was undertaken in June 2022, at the request of SLR Consulting.

A total of 16 trenches, each measuring 30m long by 1.8m wide trenches were excavated across the c.2.9ha site.

The results of the fieldwork broadly confirmed those of a preceding programme of geophysical survey. Archaeological features in the form of ditches, as well as a small number of pits, were encountered at the northern end and within the western portion of the Site and appeared likely to related to a known Late Iron Age/Romano-British settlement previously investigated immediately to the north of the Site.

The pottery assemblage consisted largely of coarsewares and utilitarian jar forms consistent with low status domestic activity, although few fragments of finer dishes, cups, and flagons were also recovered, confirming the consumption or service of food and/or liquids in the vicinity. The presence of some fineware products from the Lower Nene Valley and both South and Central Gaulish samian production centres suggests a level of access to markets supplying both regional and imported wares.

Crop processing and food preparation waste recovered from bulk soil samples further confirm the close proximity of the Site to core areas of domestic

Resource links

[View resource at provider](#)
[Json](#) [Xml](#) [Rdf](#) [Cite](#) [Report an issue](#)

Resource is part of

[ADS Archives](#)

Thematically similar

Thematically similar resources based on terms in common of:

[Copper alloy Nummus \(AE 1 - AE 4\).](#)
[Copper alloy Nummus \(AE 1 - AE 4\).](#)
[Copper alloy Nummus \(AE 1 - AE 4\).](#)

[Get data](#)[How-to](#)[Tools](#)[Community](#)[About](#)[Login](#)

GBIF | Global Biodiversity Information Facility

Free and open access to biodiversity data

[OCCURRENCES](#)[SPECIES](#)[DATASETS](#)[PUBLISHERS](#)[RESOURCES](#)[What is GBIF?](#)[About GBIF Portugal](#)

Hare-foot clover (*Trifolium arvense* L.) collected in Kalo Vig, Denmark by G.P. Lewis & B.B. Klitgård. Via Royal Botanic Gardens, Kew (CC BY 4.0)



2,644,578,047

Occurrence records



93,064

Datasets



2,194

Publishing institutions



10,229

Peer-reviewed papers
using data



CHECKLIST DATASET | REGISTERED SEPTEMBER 5, 2011

English Wikipedia - Species Pages

Published by [Wikimedia Foundation](#)

Döring M

564,195 CITATIONS 14 CITATIONS

[DATASET](#) [TAXONOMY](#) [METRICS](#) [DOWNLOAD](#) [HOME PAGE](#)

Species pages extracted from the English Wikipedia article XML dump from 2022-08-02. Multimedia, vernacular names and textual descriptions are extracted, but only pages with a taxobox or speciesbox template are recognized. See <https://github.com/mdoering/wikipedia-dwca> for details.

Publication date: August 2, 2022

Metadata last modified: August 23, 2022

Hosted by: [GBIF Secretariat](#)

Licence: Unspecified

How to cite: [DOI: 10.1146/r3k3q](#)

440,889
Accepted names

123,303
Synonyms

28%
Overlap with GBIF Backbone

24%
Overlap with Catalogue of Life

Sign in



[mdoering / wikipedia-dwca](#) Public

A parser for wikipedia species pages using the taxobox template

7 stars 2 forks 2 branches 0 tags 0 activity

Star Notifications

Code Issues Pull requests Actions

master Code

mdoering 2 years ago

src 2 years ago

gitignore 8 years ago

README.md 2 years ago

pom.xml 2 years ago

README

A Wikipedia Parser generating a [Darwin Core Archive](#) for species pages using the [taxobox](#) or [speciesbox](#) template and their derivatives. The parser focuses on the English, German, Spanish and French wikis currently and works on the [article xml dumps](#)



1 (Input a SPARQL query or choose a query example)

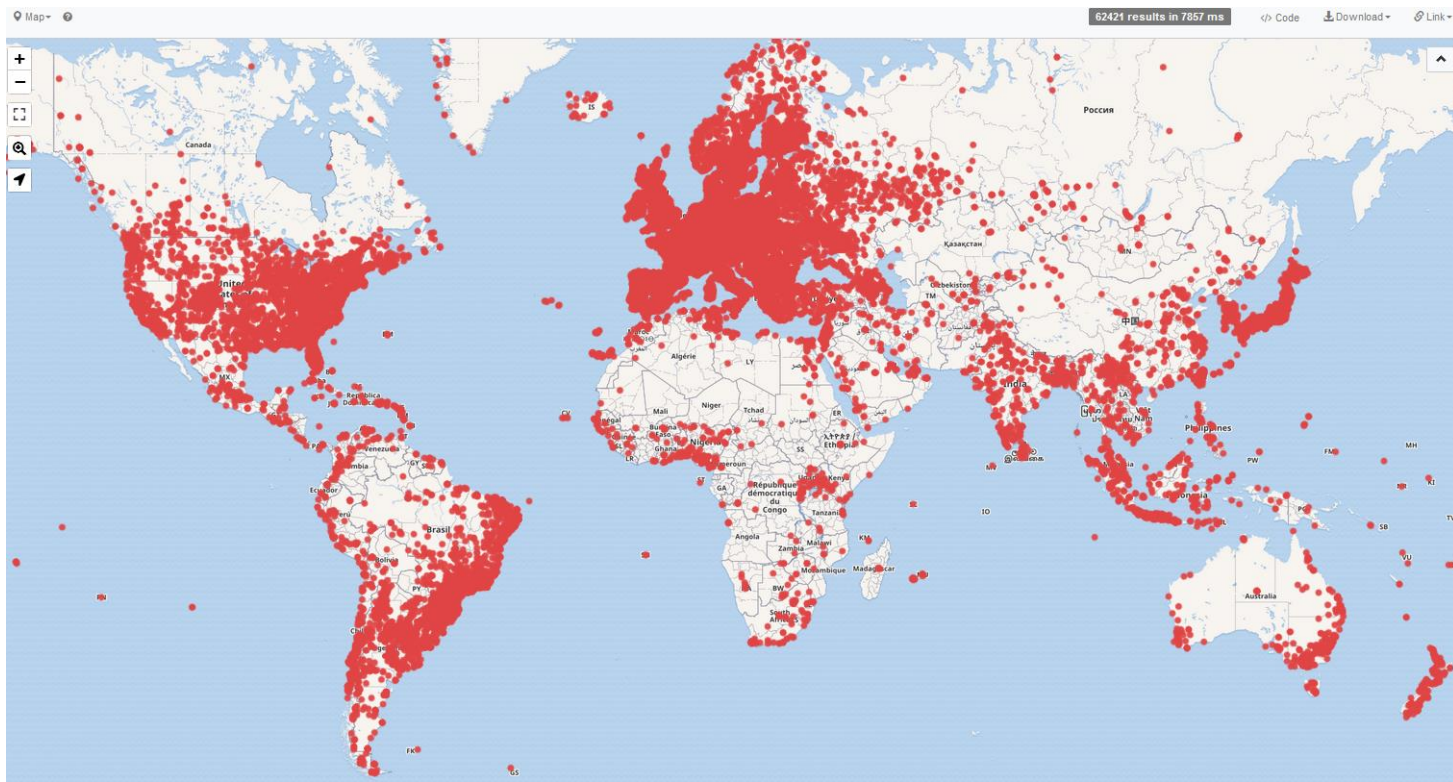




Map of Museums

#defaultview:Map

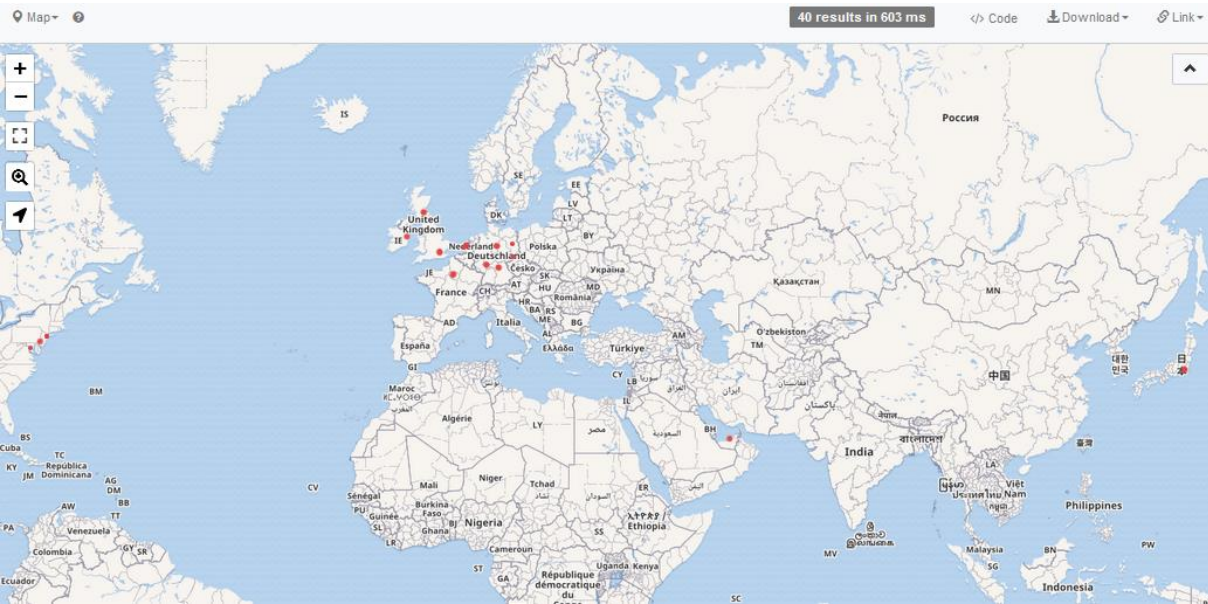
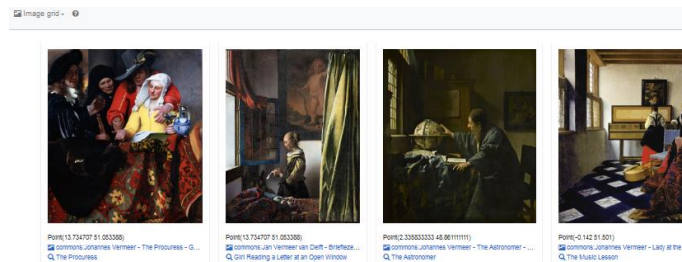
```
SELECT DISTINCT * WHERE {  
  ?item wdt:P31/wdt:P279* wd:Q33506;  
  wdt:P625 ?geo.  
}
```



Map of the locations of all paintings by Johannes Vermeer with an image

```
#defaultview:Map
```

```
SELECT ?painting ?paintingLabel ?location ?image where {
    ?painting wdt:P276 ?collection;
        wdt:P170 wd:Q41264;
        wdt:P18 ?image.
    ?collection wdt:P625 ?location.
    SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
```

40 results in 27 ms

| painting | paintingLabel | location | image |
|----------------------------|-----------------------------|--------------------------------|--|
| wd:Q45130 | The Geographer | Point(8.6737 50.1032) | commons:Johannes Vermeer, Der Geograf, 1669.png |
| wd:Q45130 | The Geographer | Point(8.673028 50.103243) | commons:Johannes Vermeer, Der Geograf, 1669.png |
| wd:Q63692 | The Procuress | Point(13.734707 51.053388) | commons:Johannes Vermeer - The Procuress - Google Art Project.jpg |
| wd:Q94006 | Saint Praxedis | Point(139.7758498 35.7153567) | commons:Vermeer saint praxedis.jpg |
| wd:Q94006 | Saint Praxedis | Point(139.775677 35.71546) | commons:Vermeer saint praxedis.jpg |
| wd:Q151143 | Woman with a Pearl Necklace | Point(13.36472222 52.50061111) | commons:Jan Vermeer van Delft - Young Woman with a Pearl Necklace - Google Art Project.jpg |

Exercise outline

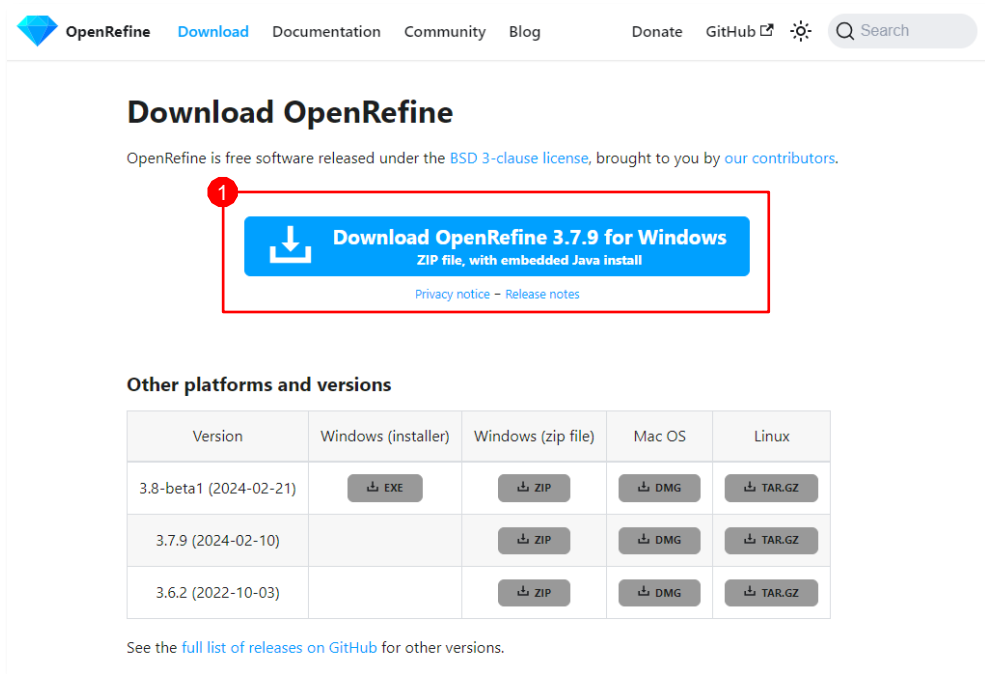
1. Download & install OpenRefine
2. Create a project by importing data
3. Clean & transform data for analysis
4. Undo/Redo & using operations templates
5. Reconcile & match data
6. Export data for further analysis or visualisation

OpenRefine helps you clean your data to prepare it for analysis.

Your data might need cleaning if
it's entered inconsistently,
comes from several sources or
needs to be formatted in a different way.

1. Download & install Open Refine

<https://openrefine.org/download.html>

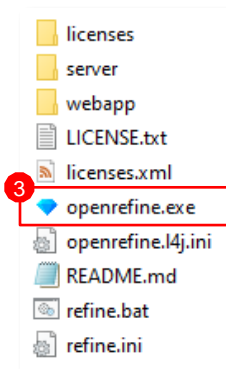


The screenshot shows the OpenRefine website's download page. At the top is a navigation bar with links for OpenRefine, Download, Documentation, Community, Blog, Donate, GitHub, and a search bar. The main heading is "Download OpenRefine". Below it, a text line states: "OpenRefine is free software released under the [BSD 3-clause license](#), brought to you by [our contributors](#)." A red circle with the number "1" highlights a blue button that says "Download OpenRefine 3.7.9 for Windows" with a download icon and the text "ZIP file, with embedded Java install". Below the button is a link for "Privacy notice - Release notes". Underneath the button section is a heading "Other platforms and versions" followed by a table of download links for different operating systems and versions.

| Version | Windows (installer) | Windows (zip file) | Mac OS | Linux |
|------------------------|---------------------|---------------------|---------------------|------------------------|
| 3.8-beta1 (2024-02-21) | EXE | ZIP | DMG | TAR.GZ |
| 3.7.9 (2024-02-10) | | ZIP | DMG | TAR.GZ |
| 3.6.2 (2022-10-03) | | ZIP | DMG | TAR.GZ |

See the [full list of releases on GitHub](#) for other versions.

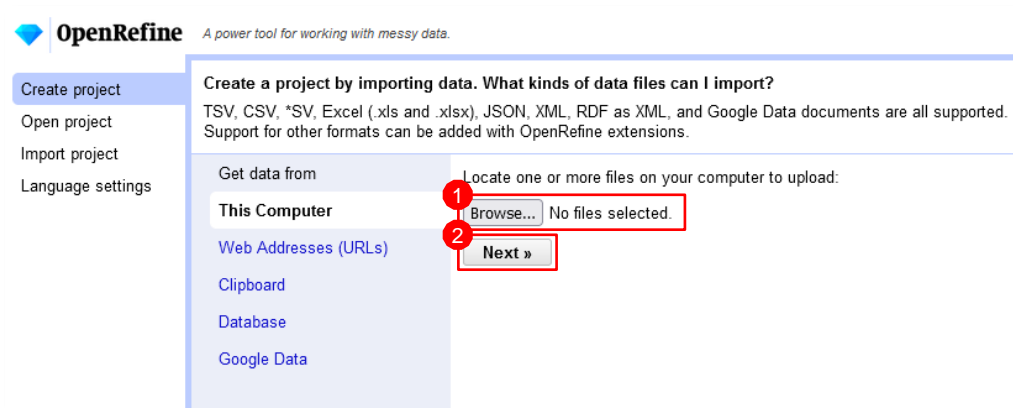
2
(unzip folder)



2. Create a project by importing data

On the browser page, import the *.CSV file provided and create a project.

A subset of data for archaeological settlements and sculptures from the Roman province of Baetica.



« start over
Configure parsing options
Project name: OpenRefine exercise Sunoikisis20.
Tags:

3 Create project »

| | | | | | | | | | | | | | |
|----|------------|---|------------|------------|------------------|-------|------------------|-------|------|-----|----------|---------|-----------------------|
| 21 | Aratispi | Wall with rectangular bastions; ceramic furnace | Punic | Punic | Roman republican | Roman | | | -30 | 640 | | Baetica | Municipium |
| 22 | Arsa | | | | | | | | -330 | 640 | | Baetica | Municipium flavium |
| 23 | Artigi | | | | | | | | -330 | 640 | | Baetica | Mansio |
| 24 | Arucci | | Celtic | Celtic | Roman republican | Roman | | | -30 | 640 | | Baetica | Estipendiaria |
| 25 | Arunda | | Celtic | Celtic | Roman republican | Roman | | | -30 | 300 | | Baetica | Oppida |
| 26 | Arva | columns with Corinthian capitals; mural painting; Roman thermal infrastructures; swimming pools | Iberian | Iberian | Punic | Punic | Roman republican | Roman | -30 | 300 | | Baetica | Municipium |
| 27 | Asido | Roman sewers | Turdetani | Iberian | Roman republican | Roman | | | -330 | 640 | | Baetica | Municipium |
| 28 | Asta Regia | stucco fragments | Phoenician | Phoenician | Punic | Punic | Roman republican | Roman | -750 | 640 | Ulterior | Baetica | Colonia Augusta Firma |

Parse data as

CSV / TSV / separator-based files

Line-based text files
Fixed-width field text files
PC-Axis text files
JSON files
MARC files
JSON-LD files

Character encoding: WINDOWS-1252

Update preview

☐ Disable auto preview

Columns are separated by:
☒ commas (CSV)
☐ tabs (TSV)
☐ custom:
☒ Use character " " to enclose cells containing column separators
☐ Trim leading & trailing whitespace from strings
Escape special characters with \

☐ Ignore first 0 line(s) at beginning of file
☒ Parse next 1 line(s) as column headers
☐ Column names (comma separated)

☐ Attempt to parse cell text into numbers
☒ Store blank rows
☒ Store blank cells as nulls
☐ Store file source
☐ Store archive file

☐ Discard initial 0 row(s) of data
☐ Load at most 0 row(s) of data

Create project

Open project

Import project

Language settings

All

| | Last modified | Name | Tags | Creator | Subject | Description | Row Count |
|---------|---------------------|--|------|---------|---------|-------------|-----------|
| ✕ About | 2024-03-13 13:49 PM | OpenRefine exercise Sunoikisis2024 csv | | | | | 999 |

3. Clean & transform data for analysis

- a) Change the case in a column and transform the value type to date in the column with dates.

Edit cells > Common transforms

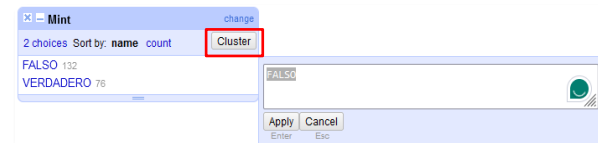
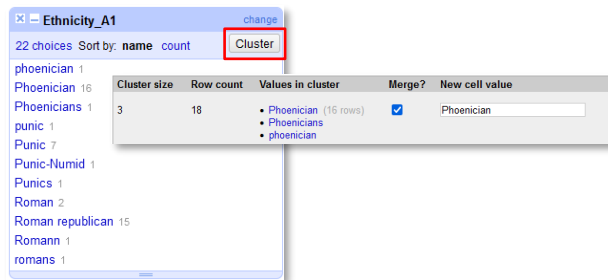
3. Clean & transform data for analysis

- a) Change the case in a column and transform the value type to date in the column with dates.

Edit cells > Common transforms

- b) Cluster cell values.

Facet > Text Facet



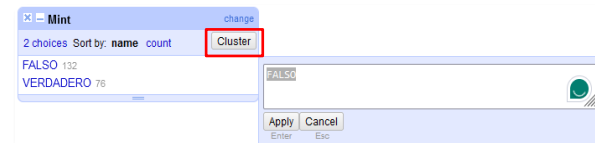
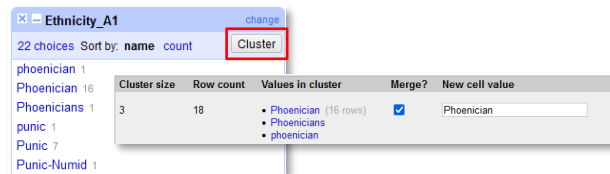
3. Clean & transform data for analysis

- a) Change the case in a column and transform the value type to date in the column with dates.

Edit cells > Common transforms

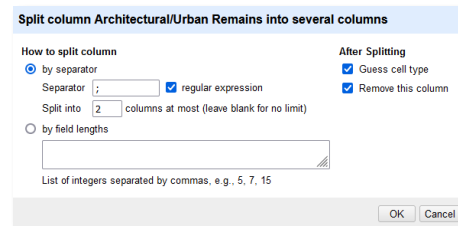
- b) Cluster cell values.

Facet > Text Facet



- c) Split a column into two or more columns and merge them again.

Edit column > Split into several columns



4. Undo/Redo & using operations templates

Facet / Filter **Undo / Redo 11 / 11** <

Extract... Apply...

Filter

8. Edit single cell on row 52, column Ethnicity_A1

9. Mass edit 24 cells in column Ethnicity_A1

10. Edit single cell on row 1, column End Date

11. Edit single cell on row 3, column Architectural/Urban Remains

999 rows

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

| | | | ID | Settlement | Architectural/Urban Remains | Ethnicity_A1 | Ethnicity_A2 |
|---|---|----|----|------------|--|--------------|--------------|
| ☆ | 🗨 | 1. | 1 | Abdera | | Phoenician | Phoenician |
| ☆ | 🗨 | 2. | 2 | Abra | | Turdetani? | Iberian |
| ☆ | 🗨 | 3. | 3 | Acci | dynastic temple with corinthian chapel; termæ; aqueduct; theatre; sewers | romans | Roman |
| ☆ | 🗨 | 4. | 4 | Acinipo | necropoleis; theatre | Oretani? | Iberian |

4. Undo/Redo & using operations templates

Facet / Filter **Undo / Redo 11 / 11** **999 rows**

Extract... Apply...

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

| | All | ID | Settlement | Architectural/Urban Remains | Ethnicity_A1 | Ethnicity_A2 |
|---|-----|----|------------|--|--------------|--------------|
| ☆ | 1. | 1 | Abdera | | Phoenician | Phoenician |
| ☆ | 2. | 2 | Abra | | Turdetani? | Iberian |
| ☆ | 3. | 3 | Acci | dynastic temple with corinthian chapel; termæ; aqueduct; theatre; sewers | romans | Roman |
| ☆ | 4. | 4 | Acinipo | necropoleis; theatre | Oretani? | Iberian |

8. Edit single cell on row 52, column Ethnicity_A1

9. Mass edit 24 cells in column Ethnicity_A1

10. Edit single cell on row 1, column End Date

11. Edit single cell on row 3, column Architectural/Urban Remains

3 Extract operation history

Extract and save parts of your operation history as JSON that you can apply to this or other projects in the future

☒ Text transform on cells in column Architectural/Urban Remains using expression value.toLowerCase()

Edit single cell on row 19, column Ethnicity_A1

Edit single cell on row 21, column Ethnicity_A1

Edit single cell on row 25, column Ethnicity_A1

Edit single cell on row 36, column Ethnicity_A1

Edit single cell on row 32, column Ethnicity_A1

Edit single cell on row 3, column Ethnicity_A1

Edit single cell on row 52, column Ethnicity_A1

☒ Mass edit cells in column Ethnicity_A1

Edit single cell on row 1, column End Date

Edit single cell on row 3, column Architectural/Urban Remains

Select all Deselect all Export

```
{
  "op": "core/text-transform",
  "engineConfig": {
    "facets": {},
    "mode": "row-based"
  },
  "columnName": "Architectural/Urban Remains",
  "expression": "value.toLowerCase()",
  "onError": "keep-original",
  "repeat": false,
  "repeatCount": 10,
  "description": "Text transform on cells in column Architectural/Urban Remains using expression value.toLowerCase()"
},
{
  "op": "core/mass-edit",
  "engineConfig": {
    "facets": {},
    "mode": "row-based"
  },
  "columnName": "Ethnicity_A1",
  "expression": "value",
  "edits": [
    {
      "from": [
        "Phoenician",
        "Phoenician",
        "Phoenician"
      ],
      "to": [
        "Phoenician",
        "Phoenician",
        "Phoenician"
      ]
    }
  ],
  "fromBlank": false,
}
```

Close

- . Select and copy the operations 3
- . Open another project
- . Apply the operations template 4

5. Reconcile & match data

a) Associate names to database keys

Reconcile > Start Reconciling

Reconcile column "Settlement_WikiData"

Services

Wikidata reconci link (en)

Add standard service... Discover s

Reconcile each cell to an entity of one of these types:

- ☒ ancient city Q15661340
- ☐ human settlement Q486972
- ☐ archaeological site Q839954
- ☐ encyclopedia article Q13433827
- ☐ titular see Q15217609
- ☐ diocese of the Catholic Church Q3146899
- ☐ province of the Philippines Q24748
- ☐ Roman archaeological site Q21752094

Also use relevant details from other columns:

| Column | Include? As property |
|---------------------|-------------------------------------|
| ID | <input type="checkbox"/> |
| Settlement | <input type="checkbox"/> |
| Architectural/Urban | <input type="checkbox"/> |
| Remains | <input type="checkbox"/> |
| Ethnicity_A1 | <input checked="" type="checkbox"/> |
| Ethnicity_A2 | <input checked="" type="checkbox"/> |
| Ethnicity_B1 | <input type="checkbox"/> |
| Ethnicity_B2 | <input type="checkbox"/> |
| Ethnicity_C1 | <input checked="" type="checkbox"/> |
| Ethnicity_C2 | <input checked="" type="checkbox"/> |
| Start Date | <input type="checkbox"/> |
| End Date | <input type="checkbox"/> |

Reconcile against type:

☐ Reconcile against no particular type

☒ Auto-match candidates with high confidence

Maximum number of candidates to return

Add standard service... Discover services...

Start reconciling... Cancel

b) Add data automatically

Edit column > Add columns from reconciled values

Add columns from reconciled column Settlement_WikiData

Add property

Preview

Reset


Suggested properties

- Bibliothèque nationale de France ID
- coat of arms image
- contains the administrative territorial entity
- coordinate location
- country
- flag image
- GND ID
- head of government
- image
- inception
- instance of
- local dialing code
- located in or next to body of water
- located in the administrative territorial entity
- located in time zone

| Oscas | Spain | Ancient Rome |
|------------------|------------------|--|
| <not reconciled> | <not reconciled> | <not reconciled> |
| Oscas | Spain | 42 13916690642792 -0 40850935232113816 |
| <not reconciled> | Roman Empire | |
| Oset | Spain | |
| Ossonoba | Portugal | 37 094259833333 -7 894882 |
| <not reconciled> | Ancient Rome | |
| <not reconciled> | | |
| Pax Julia | Portugal | 38 01777778 -7 86527778 |
| <not reconciled> | Ancient Rome | |
| <not reconciled> | | |
| Præsidium | Ancient Rome | |
| <not reconciled> | | |
| <not reconciled> | | |
| <not reconciled> | | |

OK Cancel

6. Export data for further analysis or visualisation

 **OpenRefine** OpenRefine exercise Sunoikis2024 csv [Permalink](#) Open... **Export** Help

Facet / Filter Undo / Redo 20 / 20 **1007 rows** Extensions Wikibase

Show as: rows records Show: 5 10 25 50 100 500 1000 rows « first < previous 1 next > last »

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?
[Watch these screencasts](#)

| | | ID | Settlement | Settlement_WikiData | country | coordinate location | Architectural/Urban Remains | Ethnicity_A1 | Ethnicity_A2 |
|---|---|----|------------|--|----------------------------------|---------------------|---|----------------------|--------------|
| ☆ | 🗨 | 1. | 1 | Abdera | Greece | 40.95,24.98305556 | | Phoenician | Phoenician |
| | | | | Choose new match | Choose new match | | | | |
| ☆ | 🗨 | 2. | | | | 40.932628,24.975539 | | | |
| ☆ | 🗨 | 3. | 2 | Abra | | | | Turdetani? | Iberian |
| | | | | amyloid-beta related angitis (50) ABRA (50) ABRA (50) Create new item Search for match | | | | edit | |
| ☆ | 🗨 | 4. | 3 | Acci | Spain | 37.300275,-3.134537 | dinastic temple with corinthian chapitel; termæ; aqueduct; theatre; slewers | romans | Roman |
| | | | | Choose new match | Choose new match | | | | |

ThanksObrigada!

Vera Moitinho de Almeida, PhD
vmoitinho@letras.up.pt