

# Data Management and Cleaning with Open Refine

Tuesday Morning  
MJC-BSANA-Data-Workshop-2024

# Session outline:

1. Recap from Session 1
2. Data acquisition and integration
3. Data management
4. Data cleaning and preparation
5. OpenRefine tutorial
  - a. Explore and evaluate the data to find inconsistencies.
  - b. Clean & transform the data for analysis
  - c. Undo/Redo & using operations templates
  - d. Export data for further analysis or visualisation

# Recap

1. Why is metadata important?
2. Types of databases
3. Examples of metadata schemas
4. Data validation
5. Controlled Vocabularies
6. VREs

# Data acquisition and integration

- Find your data sources
- Evaluate and review - What is in there? What is not?
- Define ways to extract/collect/integrate the data
- Define a data management plan
  - Where is the data going to sit?
  - How are you going to analyse/query it?
  - What questions do you want to ask?
  - How are you going to visualise/interpret the results.

# Data acquisition and integration

1. Identification of relevant data sources



2. Surveying of data sources



3. Data extraction

# Find your data sources

## **Electronic Resources Useful for Byzantinists:**

- Database of Byzantine Book Epigrams (<https://www.dbbe.ugent.be/>)
- Dumbarton Oaks Byzantine Seals Collection (<https://www.doaks.org/resources/seals>)
- Dumbarton Oaks Byzantine Coins Collection (<https://www.doaks.org/resources/coins>)
- Mapping Past Societies (<https://darmc.harvard.edu/>)
- Pinakes (<https://pinakes.irht.cnrs.fr/>)
- Prosopographie der mittelbyzantinischen Zeit Online (<https://www.degruyter.com/database/pmbz/html?lang=de>)
- Prosopography of the Byzantine World (<https://pbw2016.kdl.kcl.ac.uk/>)
- Pyle. A Gateway to Greek Manuscripts (<http://pyle.it/>)
- Suda On Line (<https://www.cs.uky.edu/~raphael/sol/sol-html/>)
- TLG (<https://stephanus.tlg.uci.edu/>)

## **Some Museums and Libraries with Important Byzantine Collections**

- Bibliothèque nationale de France
- The British Library
- The British Museum
- Dumbarton Oaks
- Louvre
- Metropolitan Museum of Art

# Data extraction

- Process of gathering and processing data from a range of sources (eg. databases, publications or APIs), and transforming it into a preferred machine-readable format such as JSON, CSV, and XML for further processing in the software of your choice.
  - Structured Data: highly organized and formatted
  - Unstructured Data: without a preferred data model or structure
  - Physical Sources: physical documents.
  - Digital Sources: data generated or stored in digital formats.

# Data extraction: data export

Place: **Byzantium** ✕

Hide filters

Sort by

Object name

On display only

Image only

Enter a keyword



Production date

Person/organisation (70 terms)

Place (28 terms)

Object name (11 terms)

Culture/period/dynasty  
(4 terms)

coin



1853,0512.1 | Greek | 3rdC BC-  
2ndC BC | Minted in: Byzantium

coin



1921,0110.7 | Greek | 4thC BC-  
3rdC BC | Minted in: Byzantium

coin



1947,0606.1246 | Greek | 4thC BC-  
3rdC BC | Minted in: Byzantium

coin



1866,1201.759 | Greek | 240BC-  
220BC | Minted in: Byzantium

coin



1866,1201.758 | Greek | 1stC |  
Minted in: Byzantium

coin



1882,0506.3 | Greek | 3rdC BC-  
2ndC BC | Minted in: Byzantium



# Data extraction: data export

Image	Object type	Museum r Title	Denomin	Escapeme	Descriptio	Producer	School/st	State	Authority	Ethnic nar	Ethnic nar	Culture	Productio	Productio	Find spot	Materials	Ware
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: 1853,0512.1			Alloy coin.; Head.; Cornucopiae.							Greek	3rdC BC-2	Minted in: Byzantiur	alloy		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: 1921,0110.7			Alloy coin.; Head.; Bull on dolphin.							Greek	4thC BC-3	Minted in: Byzantiur	alloy		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: 1947,0606.1246			Alloy coin.; Head.; Bull on dolphin.							Greek	4thC BC-3	Minted in: Byzantiur	alloy		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: 1866,1201.758			Alloy coin.; Head.; Obelisk.							Greek	1stC	Minted in: Byzantiur	alloy		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: 1882,0506.3			Alloy coin.; Head.; Tripod.							Greek	3rdC BC-2	Minted in: Byzantiur	alloy		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: 1866,1201.759			Alloy coin.; Head.; Trident.							Greek	240BC-22C	Minted in: Byzantiur	alloy		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: 1896,0601.28	reduced tetradrachm		Silver coin.; Head.; Seated figure.							Greek	240BC-22C	Minted in: Byzantiur	silver		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: 1949,0411.243	triobol		Silver coin.; Forepart of bull on dolphin.; Trident.							Greek	387BC-34C	Minted in: Byzantiur	silver		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: HPB,p28.1.8	sigloi		Silver coin.; Bull on dolphin.; Incuse.							Greek	340BC-32C	Minted in: Byzantiur	silver		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: TC,p107.14.Lys			Silver coin.			Kingdom of Lysimachus				Greek		Minted in: Byzantiur	silver		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: 1910,1104.99			Silver coin.			Kingdom of Lysimachus				Greek		Minted in: Byzantiur	silver		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: 1920,0805.30			Copper alloy coin.; Head.; Trident.							Greek	240BC-22C	Minted in: Byzantiur	copper alloy		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: 1924,0708.11			Gold coin.			Kingdom of Lysimachus				Greek		Minted in: Byzantiur	gold		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: 1922,1115.16			Gold coin.			Kingdom of Lysimachus				Greek		Minted in: Byzantiur	gold		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: G.585			Silver coin.			Kingdom of Lysimachus				Greek		Minted in: Byzantiur	silver		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: 1841,B.425	triobol		Silver coin.; Forepart of bull on dolphin.; Trident.							Greek	387BC-34C	Minted in: Byzantiur	silver		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: RPK,p96F.2.Byz	diobol		Silver coin.; Bull on dolphin.; Incuse.							Greek	340BC-32C	Minted in: Byzantiur	silver		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: TC,p107.11.Lys			Silver coin.			Kingdom of Lysimachus				Greek		Minted in: Byzantiur	silver		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: 1935,0619.22			Silver coin.			Kingdom of Lysimachus				Greek		Minted in: Byzantiur	silver		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: 1949,0411.241	diobol		Silver coin.; Bull on dolphin.; Incuse.							Greek	340BC-32C	Minted in: Byzantiur	silver		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: 1923,1103.1			Silver coin.			Kingdom of Lysimachus				Greek		Minted in: Byzantiur	silver		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: BNK,G.214			Gold coin.			Kingdom of Lysimachus				Greek		Minted in: Byzantiur	gold		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: TC,p107.12.Lys			Silver coin.			Kingdom of Lysimachus				Greek		Minted in: Byzantiur	silver		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: 1911,0706.10			Silver coin.			Kingdom of Lysimachus				Greek		Minted in: Byzantiur	silver		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: 1866,1201.752	diobol		Silver coin.; Bull on dolphin.; Incuse.							Greek	340BC-32C	Minted in: Byzantiur	silver		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: 1896,0703.48	reduced tetradrachm		Silver coin.; Head.; Seated figure.							Greek	240BC-22C	Minted in: Byzantiur	silver		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: 1911,0704.174			Silver coin.			Kingdom of Lysimachus				Greek		Minted in: Byzantiur	silver		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: 1910,1105.52			Silver coin.			Kingdom of Lysimachus				Greek		Minted in: Byzantiur	silver		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: 1921,0110.5			Copper alloy coin.; Head.; Standing figure.							Greek	240BC-22C	Minted in: Byzantiur	copper alloy		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: BNK,G.228			Silver coin.			Kingdom of Lysimachus				Greek		Minted in: Byzantiur	silver		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: 1921,0520.2			Copper alloy coin.; Head.; Seated figure.							Greek	3rdC BC-2	Minted in: Calchedo	copper alloy		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: 1934,0523.3	tetradrachm		Silver coin.; Bull on dolphin, trident.; Incuse square.							Greek	387BC-34C	Minted in: Excavated	silver		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: RPK,Gre.42			Copper alloy coin.; Head of Apollo.; Tripod.							Greek	3rdC BC-2	Minted in: Byzantiur	copper alloy		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: 1896,0601.27	tetradrachm		Silver coin.; Bull on dolphin.; Incuse square.							Greek	387BC-34C	Minted in: Byzantiur	silver		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: BNK,G.213			Gold coin.			Kingdom of Lysimachus				Greek		Minted in: Byzantiur	gold		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: BNK,G.221			Silver coin.			Kingdom of Lysimachus				Greek		Minted in: Byzantiur	silver		
<a href="https://media.britishmuseum.org/me coin">https://media.britishmuseum.org/me coin</a>		No: 1866,1201.860			Silver coin.			Kingdom of Lysimachus				Greek		Minted in: Byzantiur	silver		

# Data extraction: web scraping

## Search the Byzantine Seals Collection

[More filters](#)

### General Search

☒ all items

Search

☐ in current results

### Century

☐ 5th (49)

☐ 6th (462)

☐ 7th (235)

☐ 8th (102)

☐ 9th (100)

☐ 10th (100)

☐ 11th (100)

☐ 12th (27)

☐ 13th (27)

☐ 14th (7)

☐ 15th (2)

### Location

All (15947) ▼

### Title



[John protospatharios, judge of the Hippodrome and skribas \(eleventh century\)](#)  
BZS.1951.31.5.1607



[Constantine, azim.anos \(eleventh/twelfth century\)](#)  
BZS.1951.31.5.816



[Leo Kritharas, patrikios \(eleventh century, second half\)](#)  
BZS.1951.31.5.806



[Basil Chrysomalles \(thirteenth century, first half\)](#)  
BZS.1951.31.5.780



[Himerios vestarches \(?\) \(tenth/eleventh century\)](#)  
BZS.1951.31.5.885



[Nicholas Alex., protospatharios \(tenth/eleventh century\)](#)  
BZS.1951.31.5.855



[Michael patrikios \(eighth century\)](#)  
BZS.1951.31.5.770



[Georgakios imperial spatharios \(eighth/ninth century\)](#)  
BZS.1951.31.5.774



[George imperial protospatharios \(tenth century\)](#)  
BZS.1951.31.5.777

# Data extraction: web scraping

Python is popular for web scraping, some of the best libraries are:

- **BeautifulSoup** allows you to parse HTML and XML documents. Using API, you can easily navigate through the HTML document tree and extract tags, meta titles, attributes, text, and other content.
- **Requests** simple yet powerful Python library for making HTTP requests. It is designed to be easy to use and intuitive, with a clean and consistent API. With Requests, you can easily send GET and POST requests, and handle cookies, authentication, and other HTTP features.
- **Selenium** allows you to automate web browsers such as Chrome, Firefox, and Safari and simulate human interaction with websites. You can click buttons, fill out forms, scroll pages, and perform other actions. It is also used for testing web applications and automating repetitive tasks.
- **Pandas** allow storing and manipulating data in various formats, including CSV, Excel, JSON, and SQL databases. Using Pandas, you can easily clean, transform, and analyze data extracted from websites.

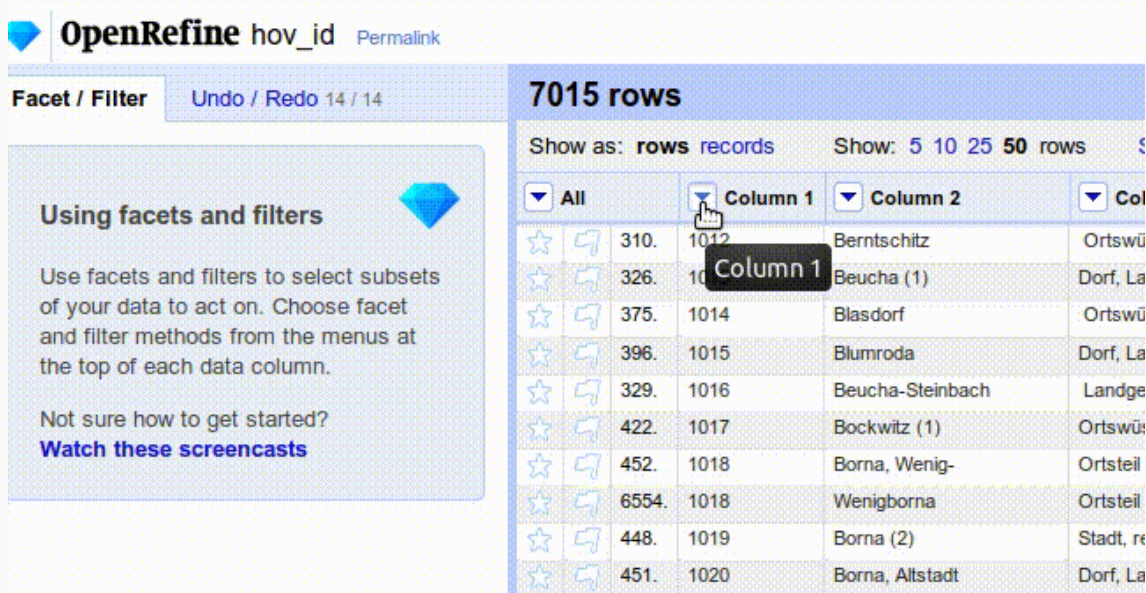
Edited from: <https://nanonets.com/blog/web-scraping-with-python-tutorial/>

# Data extraction: text to data

Optical Character Recognition (OCR): method to convert an image of text into a machine-readable text format.

- Google Docs OCR free alternative offered by Google Drive to convert images to text. The process behind Google Docs OCR is nothing more than uploading images from which you need the data, into Google Docs, and exporting the data as a text format into your computer.
- FreeOCR is a Windows OCR application that uses the Tesseract OCR engine. It can handle PDF files and images.

# Data management with OpenRefine



**OpenRefine** hov\_id Permalink

Facet / Filter Undo / Redo 14 / 14

**7015 rows**

Show as: rows records Show: 5 10 25 50 rows

▼ All ▼ Column 1 ▼ Column 2 ▼ Col

		Column 1	Column 2	Col
☆	310.	1012	Berntschitz	Ortswü
☆	326.	1012	Beucha (1)	Dorf, La
☆	375.	1014	Blasdorf	Ortswü
☆	396.	1015	Blumroda	Dorf, La
☆	329.	1016	Beucha-Steinbach	Landge
☆	422.	1017	Bockwitz (1)	Ortswü:
☆	452.	1018	Borna, Wenig-	Ortsteil
☆	6554.	1018	Wenigborna	Ortsteil
☆	448.	1019	Borna (2)	Stadt, re
☆	451.	1020	Borna, Altstadt	Dorf, La

**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](#)

- Powerful tool to work with messy data: clean it, transform it from one format to another and extend it with web services and external data.

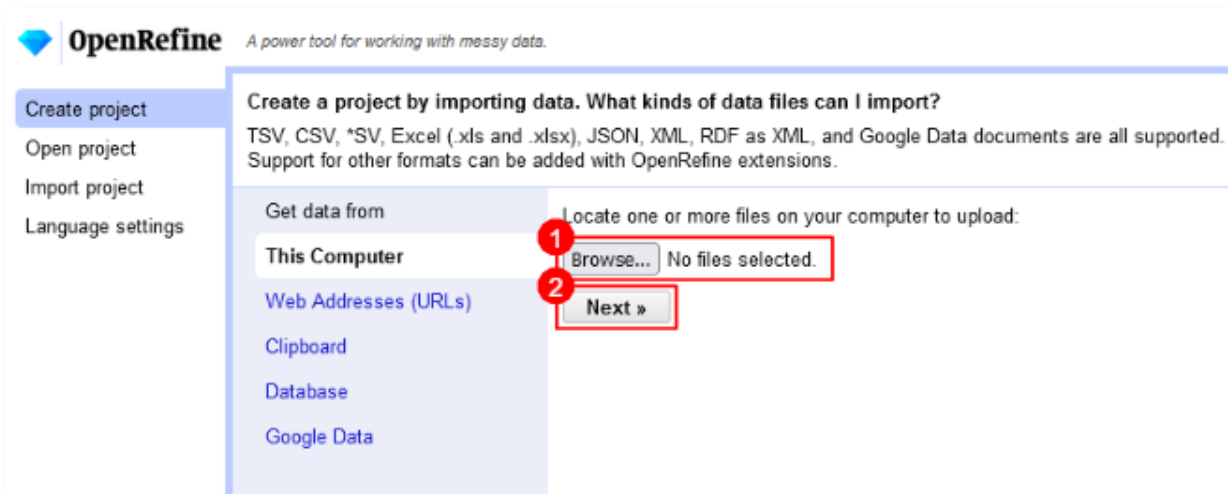
- It works by running a small server on your computer and tuning the search engine to execute the orders and interact with the tool.

# Create Project

- **Create Project:** loads a dataset into OpenRefine. There are various supported formats, and you can import data in different ways:
  - This Computer: Select a file stored on your local machine
  - Web Addresses (URLs): Import data directly from an online source
  - Clipboard: Copy-paste your data into a text field
  - Google Data: Enable access to a Google Spreadsheet or Fusion Table.

# Create Project

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.



# Open Project

- **Open Project:** This option helps you go back to an existing project created during a former session.
- **Import Project:** directly import an existing OpenRefine project archive.



# Explore your data

999 rows								
Show as: <b>rows</b> records		Show: 5 10 25 50 100 500 1000 rows						
<input type="checkbox"/> All	<input type="checkbox"/> ID	<input type="checkbox"/> Settlement	<input type="checkbox"/> Architectural/Urban Remains	<input type="checkbox"/> Ethnicity_A1	<input type="checkbox"/> Ethnicity_A2	<input type="checkbox"/> Ethnicity_B1	<input type="checkbox"/> Ethnicity_B2	
<input type="checkbox"/>	1.	1	Abdera		Phoenician	Phoenician		
<input type="checkbox"/>	2.	2	Abra		Turdetani?	Iberian		
<input type="checkbox"/>	3.	3	Acci	Dinastic Temple with corinthian chapitel; termaet aqueductt theatret sleners	Roman	Roman		
<input type="checkbox"/>	4.	4	Acinipo	Necropoleis; theatre	Oretani?	Iberian		
<input type="checkbox"/>	5.	5	Ad Aras					
<input type="checkbox"/>	6.	6	Ad Decumo 1					
<input type="checkbox"/>	7.	7	Ad Decumo 2					
<input type="checkbox"/>	8.	8	Ad Gemellas					
<input type="checkbox"/>	9.	9	Ad Herculem					
<input type="checkbox"/>	10.	10	Ad Lucos					

- Number of rows
- Display options
- Column headers and menus
- Cell contents

# Manipulating columns

999 rows						
Show as: <b>rows</b> records		Show: 5 10 25 50 100 500 1000 rows				
▼ All	▼ ID	▼ Settlement	▼ Architectural/Urban Remains	▼ Ethnicity_A1	▼ Ethnicity_A2	
☆	1.	1	Facet Text filter Edit cells Edit column Transpose Sort... View Reconcile	Phoenician	Phoenician	
☆	2.	2		Turdetani?	Iberian	
☆	3.	3		Temple with corinthian	Roman	
☆	4.	4			Iberian	
☆	5.	5				
☆	6.	6	Ad Decumo 1			
☆	7.	7	Ad Decumo 2			
☆	8.	8	Ad Gemellas			
☆	9.	9	Ad Herculem			
☆	10.	10	Ad Lucos			

# Move columns around

999 rows								
Show as: <b>rows</b> records			Show: 5 10 25 50 100 500 1000 rows Sort ▼					
▼ All	▼ ID	▼ Settlement	▼ Architectural/Urban Remains	▼ Ethnicity_A1	▼ Ethnicity_A2	▼ Ethnicity_B1	▼ Ethnicity_B2	
☆	16.	16	Aipora	Facet	Celtic	Celtic	Roman republican	Roman
☆	24.	24	Arucci	Text filter	Celtic	Celtic	Roman republican	Roman
			Edit cells					
☆	25.	25	Arunda	Edit column	Split into several columns...		Roman republican	Roman
			Transpose	Join columns...				
☆	174.	174	Segida Restituta Iulia	Sort...	Add column based on this column...		Roman republican	Roman
☆	192.	192	Turobriga	View	Add column by fetching URLs...		Roman republican	Roman
			Reconcile	Add columns from reconciled values...				
☆	198.	198	Ugultunia		Rename this column...		Roman republican	Roman
					Remove this column			
☆	2.	2	Abra		Move column to beginning			
☆	4.	4	Acinipo	Necropoleis; theatre	Move column to end			
					Move column left			
☆	18.	18	Alba	Necropoleis; mausoleum	Move column right		Roman republican	Roman
☆	26.	26	Arva	columns with Corinthian capitals; mural painting; Roman thermal infrastructures; swimming pools	Iberian	Iberian	Punic	Punic

# Merge several columns into one

## Join columns

### Select and order columns to join

☒ Architectural/Urban Remains

☐ ID

☐ Settlement

☐ Ethnicity\_A1

☐ Ethnicity\_A2

☐ Ethnicity\_B1

☐ Ethnicity\_B2

☐ Ethnicity\_C1

☐ Ethnicity\_C2

Select all

Deselect all

### Select options

Separator between the content of each column   
Enter one or more characters, or keep blank to join the columns without separator.

☒ Replace nulls with

Enter one or more characters, or keep blank to replace nulls with blank strings.

☐ Skip nulls.

☐ In separator and nulls substitutes, use \n for new lines, \t for tabulation, \\n for \n, \\t for \t.

☒ Write result in selected column.

☐ Write result in new column named

☐ Delete joined columns.

OK

Cancel

# Manipulating cells

999 rows

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

▼ All	▼ ID	▼ Settlement	▼ Architectural/Urban Remains	▼ Ethnicity_A1	▼ Ethnicity_A2	▼ Ethnicity_B1	▼ Ethnicity_B2
☆ ↻	1.	1	Facet	Phoenician	Phoenician		
☆ ↻	2.	2	Text filter	Turdetani?	Iberian		
☆ ↻	3.	3	Edit cells	an	Roman		
☆ ↻	4.	4	Edit column				
☆ ↻	5.	5	Transpose				
☆ ↻	6.	6	Sort...				
☆ ↻	7.	7	View				
☆ ↻	8.	8	Reconcile				
☆ ↻	9.	9	Ad Decumo 1				
☆ ↻	10.	10	Ad Decumo 2				
☆ ↻			Ad Gemellas				
☆ ↻			Ad Herculem				
☆ ↻			Ad Lucos				

Transform...

Common transforms

Trim leading and trailing whitespace

Collapse consecutive whitespace

Unescape HTML entities

Replace smart quotes with ASCII

To titlecase

To uppercase

To lowercase

To number

To date

To text

To null

To empty string

# Manipulating cells

Facet	▶		Roman republican	Roman		
Text filter	▶		Roman	Roman		
Edit cells	▶	Transform...				
Edit column	▶	Common transforms		▶ Trim leading and trailing whitespace		
Transpose	▶	Fill down		Collapse consecutive whitespace		
Sort...		Blank down		Unescape HTML entities		
View	▶	Split multi-valued cells...		Replace smart quotes with ASCII		
Reconcile	▶	Join multi-valued cells...		To titlecase		
Celtic	Celtic	Cluster and edit...		To uppercase		
		Replace...		To lowercase		
Turdetani?	Iberian			To number		
Oretani?	Iberian			To date		
				To text		
Bastetani	Iberian	Roman republican		To null		
				To empty string		

# Manipulating cells - data types

999 rows

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

		<input type="checkbox"/> All	<input type="checkbox"/> ID	<input type="checkbox"/> Settlement	<input type="checkbox"/> Architectural/Urban Remains	<input type="checkbox"/> Ethnicity_A1	<input type="checkbox"/> Ethnicity_A2	<input type="checkbox"/> Ethnicity_B1	<input type="checkbox"/> Ethnicity_B2	
<input type="checkbox"/>	<input type="checkbox"/>	1.	1	Abdera		Phoenician	Phoenician			Ror repi
<input type="checkbox"/>	<input type="checkbox"/>	2.	2	Abra		Turdetani?	Iberian			Ror repi
<input type="checkbox"/>	<input type="checkbox"/>	3.	3	Acci						
<input type="checkbox"/>	<input type="checkbox"/>	4.	4	Acinipo						Ror repi
<input type="checkbox"/>	<input type="checkbox"/>	5.	5	Ad Aras						
<input type="checkbox"/>	<input type="checkbox"/>	6.	6	Ad Decumo 1						
<input type="checkbox"/>	<input type="checkbox"/>	7.	7	Ad Decumo 2						

Data type: text

Acci

Apply   Apply to all identical cells   Cancel

Enter   Ctrl-Enter   Esc

The data types supported are:

- string (one or more text characters)
- number (one or more characters of numbers only)
- boolean (values of "true" or "false")
- [date](#) (ISO-8601-compliant extended format with time in UTC: YYYY-MM-DDTHH:MM:SSZ)

# Dates

Start Date	End Date	Roman Province 1	Roman Province 2	Legal Status A	Legal Status B	Legal Status C
-750	Facet		Baetica	Civitas		
-330	Text filter		Baetica	Opiddum de Tosiria?		
-330	Edit cells			Colonia Iulia		
-330	Edit column					
-330	Transpose					
-330	Sort...					
-330	View					
-30	Reconcile					
-30	640		Baetica			
-30	640		Baetica			
-330	640		Baetica			

Transform...

Common transforms

Fill down

Blank down

Split multi-valued cells...

Join multi-valued cells...

Cluster and edit...

Replace...

Trim leading and trailing whitespace

Collapse consecutive whitespace

Unescape HTML entities

Replace smart quotes with ASCII

To titlecase

To uppercase

To lowercase

To number

To date

To text

To null

To empty string



# Sorting Data

**Sort by ID**

Sort cell values as

☒ text ☐ case-sensitive

☐ numbers

☐ dates

☐ booleans

☒ a - z ☐ z - a

Position blanks and errors

Valid values

Errors

Blanks

Drag and drop to re-order

OK

Cancel

# Sorting Data

1000 rows **Sort** ▼ « first < prev

s	▼ Ethnicity	Ethnicity_B2	▼ Ethnicity_C1	▼ Ethnicity_C2	▼ Start Date	▼		
	Celtic				-30	6.		
	Celtic	Celtic	Roman republican	Roman	-30	6.		
	Celtic	Celtic	Roman republican	Roman	-30	30		
	Celtic	Celtic	Roman republican	Roman	-330	30		
	Celtic	Celtic	Roman republican	Roman	-330	30		
	Celtic	Celtic	Roman republican	Roman	-30	6.		
	Turdetani?	Iberian		Roman republican	Roman	-330	30	
	Oretani?	Iberian		Roman republican	Roman	-330	30	
	Bastetani	Iberian	Roman republican	Roman	-30	6.		
	Iberian	Iberian	Punic	Punic	Roman republican	Roman	-30	30

Remove sort

Reorder rows permanently

By Ethnicity\_A2 ▶

Sort...

Reverse

Remove sort

# Control Data view

Show: 5 10 25 50 100 500 1000 rows Sort ▾				
ment	Architectural/Urban Remains	Ethnicity_A1	Ethnicity_A2	Eth
	Facet ▶	Celtic	Celtic	Roman republic
	Text filter	Celtic	Celtic	Roman republic
	Edit cells ▶			
	Edit column ▶			
	Transpose ▶	Celtic	Celtic	Roman republic
lia	Sort...	Celtic	Celtic	Roman republic
	View ▶	<div>Collapse this column</div> <div>Collapse all other columns</div> <div>Collapse all columns to left</div> <div>Collapse all columns to right</div> <div>Expand all columns</div> <div>Expand all columns to the left</div> <div>Expand all columns to the right</div>		Roman republic
	Reconcile ▶			
				Roman republic
	Necropoleis; theatre			
	Necropoleis; mausoleum	Bastetani	Iberian	Roman republic

# Facing Data

Refresh

Reset all

Remove all

×

Mint

change

2 choices

Sort by: name count

Cluster

FALSO 132

VERDADERO 76

(blank) 791

Facet by choice counts

Show as: rows records

Show: 5 10 25 50 100 500 1000 rows

Legal Status C	Via	Conventus	Mint	Excavated	M
		Conventus Gaditanus	VERDADERO	VERDADERO	Adra
		Conventus Astigitanus/border	FALSO		La Ca

FALSO

Apply Enter

Cancel Esc

# Transposing cells across columns

999 rows

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

▼ All	▼ ID	▼ Settlement	▼ Architectural/Urban Remains	▼ Ethnicity_A1	▼ Ethnicity_A2	▼ Ethnicity_B1
☆	16.	16	Aipora	Celtic	Celtic	Roman republican
☆	24.	24	Arucci	Celtic	Celtic	Roman republican
☆	25.	25	Arunda	Celtic	Celtic	Roman republican
☆	174.	174	Segida Restituta Iulia			Roman republican
☆	192.	192	Turobriga			Roman republican
☆	198.	198	Ugultunia	Celtic	Celtic	Roman republican
☆	2.	2	Abra	Turdetani?	Iberian	
☆	4.	4	Acinipo	Necropoleis; theatre	Oretani?	Iberian

# Cluster and edit similar cells

## Cluster and edit column "Ethnicity\_A1"

Find groups of different cell values that might be other representations of the same thing. For example, "New York" and "new york" likely refer to the same concept and just differ by capitalization, and "Gödel" and "Godel" probably refer to the same person. [Find out more...](#)

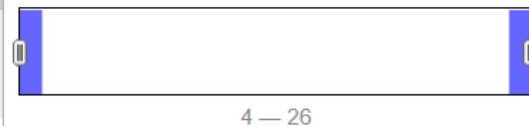
Method Key collision

Keying function Daitch-Mokotoff

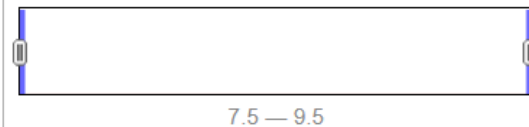
2 clusters found

Cluster size	Row count	Values in cluster	Merge?	New cell value
2	26	<ul style="list-style-type: none"><li>Turdetani (24 rows)</li><li>Turdetani? (2 rows)</li></ul>	<input type="checkbox"/>	<input type="text" value="Turdetani"/>
2	4	<ul style="list-style-type: none"><li>Oretani (3 rows)</li><li>Oretani?</li></ul>	<input type="checkbox"/>	<input type="text" value="Oretani"/>

# Rows in cluster



Average length of choices



# Using the project History

Facet / Filter

Undo / Redo 1 / 2

Extract... Apply...

Filter

0. Create project

1. Text transform on 0 cells in column End Date: value.toDate()

2. Transpose every 2 cells in column Architectural/Urban Remains into separate columns

999 rows

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

▼ All	▼ ID	▼ Settlement	▼ Architectural/Urban Remains	▼ Ethnicity_A1	▼ Ethnicity_A2
☆ ↻ 16.	16	Aipora		Celtic	Celtic
☆ ↻ 24.	24	Arucci		Celtic	Celtic
☆ ↻ 25.	25	Arunda		Celtic	Celtic
☆ ↻ 174.	174	Segida Restituta Iulia		Celtic	Celtic
☆ ↻ 192.	192	Turobriga		Celtic	Celtic
☆ ↻ 198.	198	Ugultunia		Celtic	Celtic

# Exercise

1. Change the order of rows and columns.
2. Edit cell contents within a particular column
3. Change cells type.
4. Transform rows into columns, and columns into rows
5. Add new columns based on existing data, with fetching new information.
6. Convert your rows of data into multi-record rows.