ECC058 – Web Data Collection Assignment 1

# Purpose

The purpose of this assignment is to understand the resource and effort required to collect data from the web.

# Overview

Each student will collect data on 20 distinct inscriptions from at least 2 different websites according to the requirements below, format the data using the sample spreadsheet (Epi Data Collection Sample.xlsx) as a guide and track the time required to collect all 20 inscriptions. To ensure uniqueness the queried data ranges will be assigned to individuals in the form of URLs where each URL results in 100 inscriptions to collect from. Each student must only use the browser to find and copy (ability to cut and paste is assumed) the data into the required format.

# Requirements

Each inscription must have:

1. Resource URL used to access the data for the inscription
2. Website Identification Number.
3. Interpretive transcription of the Latin text on the artifact with minimum 5 lines and with a minimum of 40 engraved characters (excluding characters add by the researcher)
4. The material type of the artifact
5. Chronological information (date range) where inscriptions are select from a minimum of 5 different date ranges.
6. Geographic information on the findspot of the artifact in lat/long format.
7. Image URL to a somewhat orthogonal image of the text of the inscription with medium to high resolution. Preference IIIF image URLs (Mondays class will explain IIIF standard ).

Optional Data:

1. Alternative identifiers
2. Type of inscription
3. Type of monument/artifact
4. Diplomatic transcription (a one to one mapped transcription of just what is easily visible on the artifact)

**EDH Website Search URL**

**Where the “start=0” can be adjusted by 100 to get the following set of 100 results (i.e. start=100 or start=200 etc.)**

<https://edh-www.adw.uni-heidelberg.de/inschrift/erweiterteSuche?&beleg=c&compFundjahr=eq&compHoehe=eq&compBreite=eq&compJahre=eq&compMonate=eq&compTage=eq&compStunden=eq&bool=AND&beleg89=ja&nurMitFoto=ja&anzahl=100&addFeldMaterial=ja&addFeldDTyp=ja&addFeldIGat=ja&addFeldKommentar=ja&addFeldMasse=ja&addFeldBH=ja&start=0>

**EDR Website Search URL**

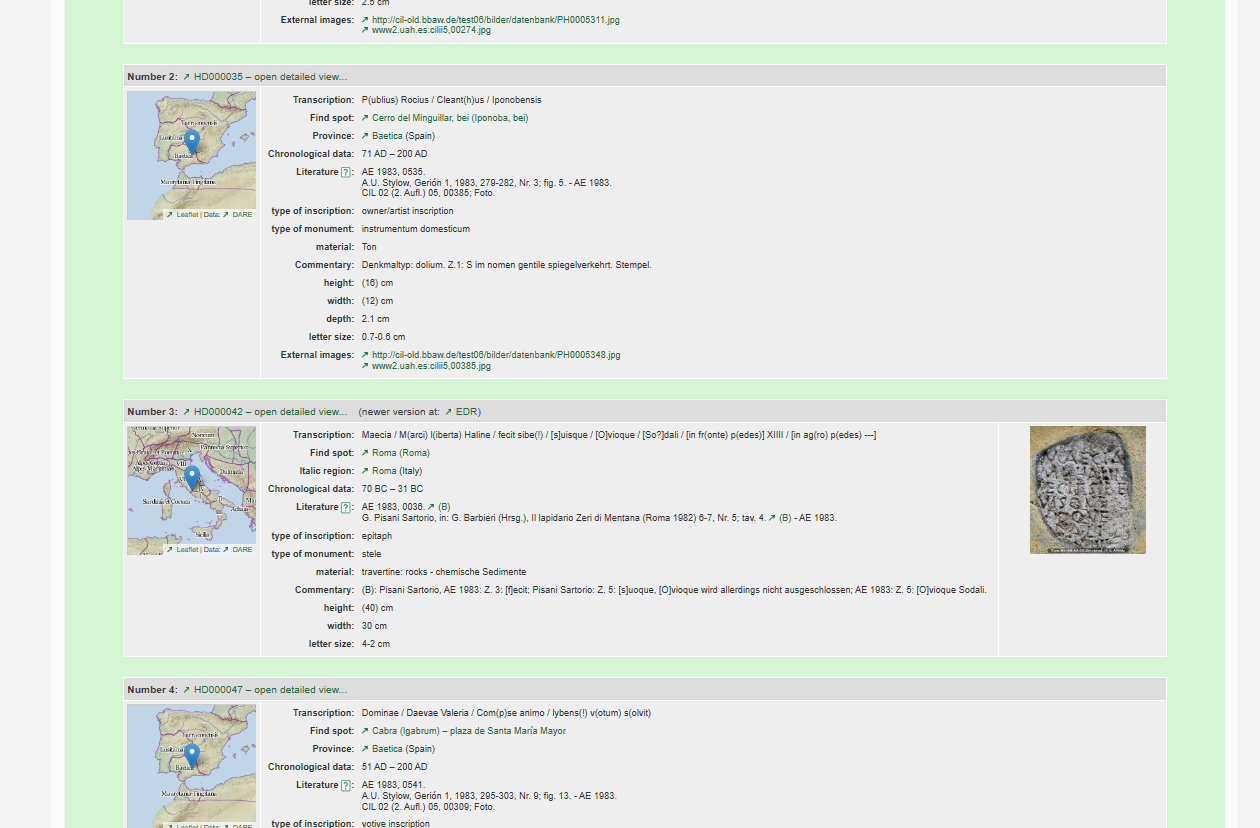
**Where the “p=1” can be adjusted by 1 to get the following page of 100 results (i.e. p=2 or p=3 etc.)**

<http://www.edr-edr.it/edr_programmi/res_complex_comune.php?do=book&boolfundstelle=AND&boolaufbewahrung=AND&booldenkmaltyp=OR&ope_mis_alto=%3C%3D&ope_mis_largo=%3C%3D&ope_mis_spesso=%3C%3D&lingua%5B%5D=latina&boollivelloSocialeTot=OR&boolBibliografia=AND&boolTesto=AND&boolApparato=AND&ope_dat_jahr_a=%3D&ope_dat_jahr_e=%3D&bool=AND&ordinamento=id_nr&javasi=javascriptsi&se_foto=tutte&lang=en&p=1>

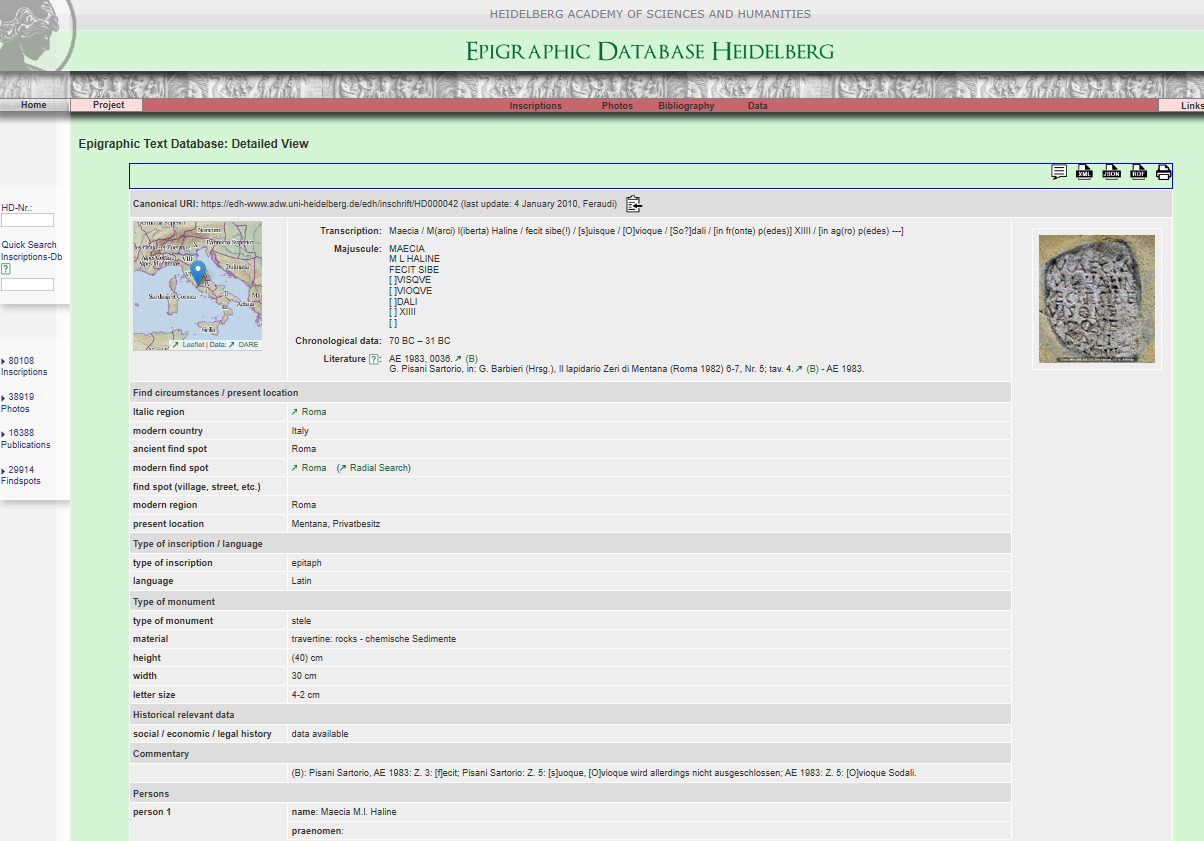
**Assigned selection ranges (by name for 500 each):**

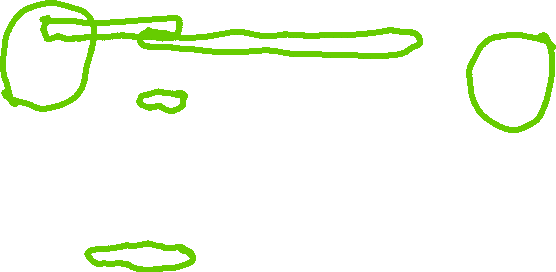
Elisa Brunetta 1st EDH (0,100,200,300,400) EDR(1,2,3,4,5)  
Marco Del Din 2nd EDH (500,600,700,800,900) EDR(6,7,8,9,10)  
Serena Guida 3rd EDH (1000,1100,1200,1300,1400) EDR(11,12,13,14,15)  
Linda Pietrasanta 4th EDH (1500,1600,1700,1800,1900) EDR(16,17,18,19,20)   
Melina Russo 5th EDH (2000,2100,2200,2300,2400) EDR(21,22,23,24,25)

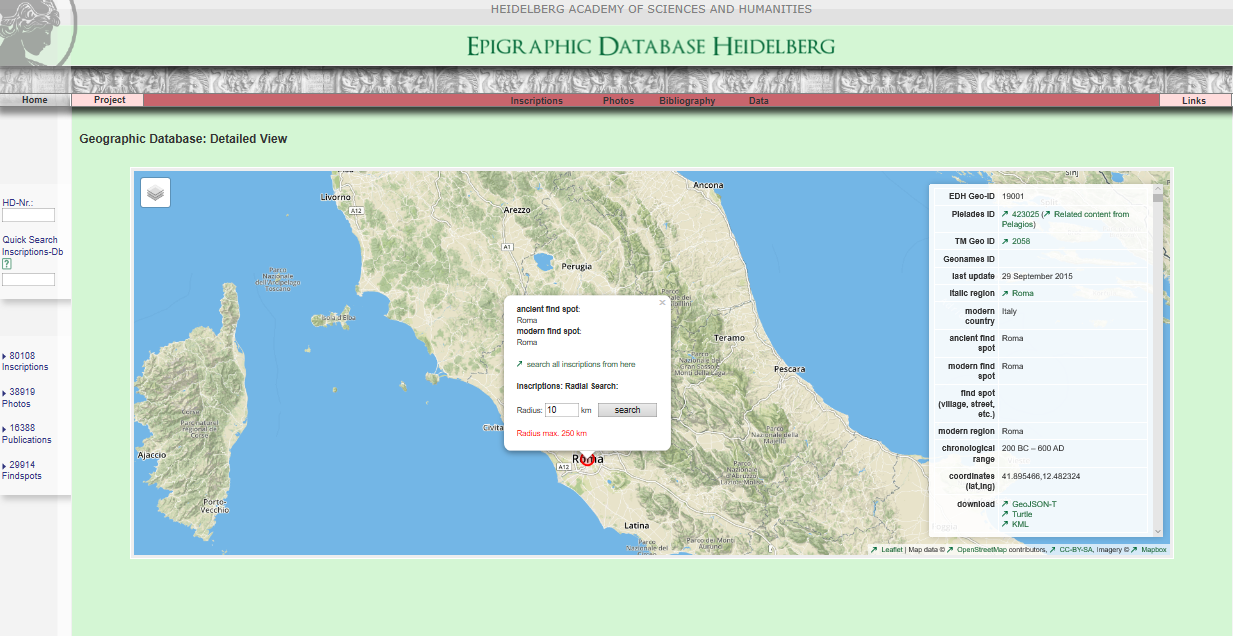
# Worked example using EDH

1. Enter URL   
   <https://edh-www.adw.uni-heidelberg.de/inschrift/erweiterteSuche?&beleg=c&compFundjahr=eq&compHoehe=eq&compBreite=eq&compJahre=eq&compMonate=eq&compTage=eq&compStunden=eq&bool=AND&beleg89=ja&nurMitFoto=ja&anzahl=100&addFeldMaterial=ja&addFeldDTyp=ja&addFeldIGat=ja&addFeldKommentar=ja&addFeldMasse=ja&addFeldBH=ja&start=0>
2. Scroll to inscription that meets requirements, gather ref number and click on link to inscription detail page  
   

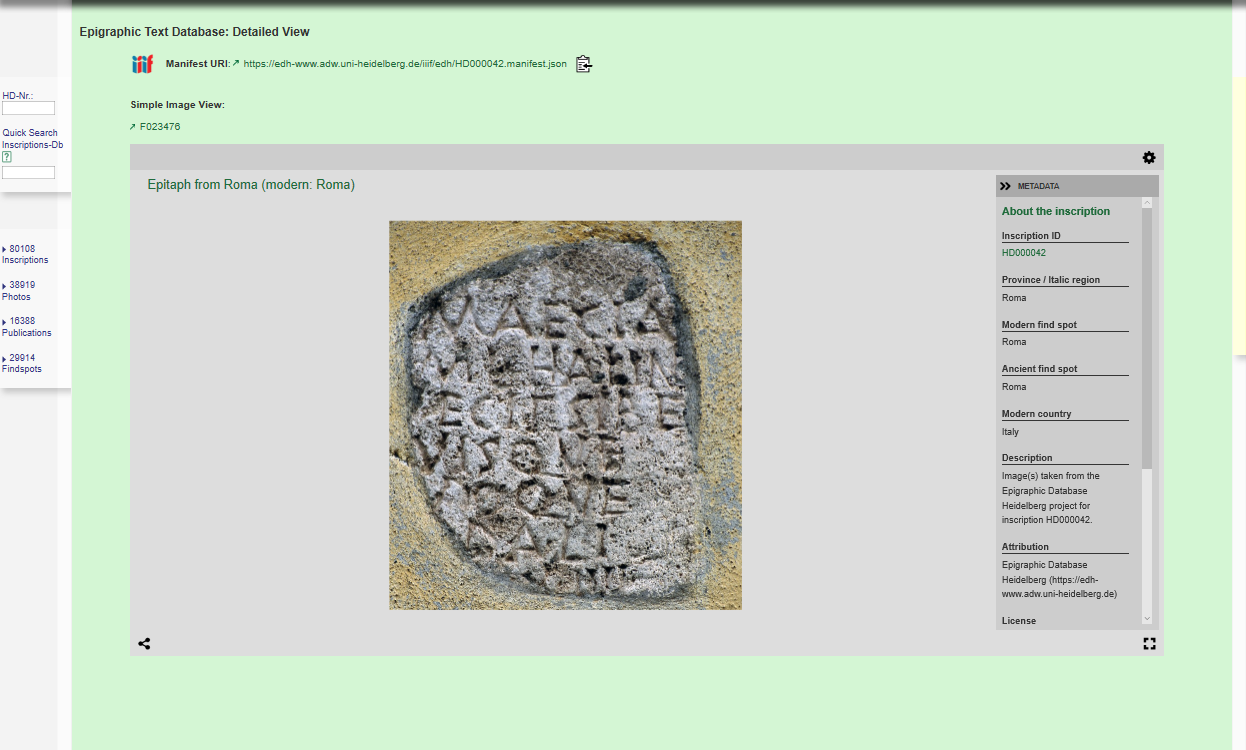


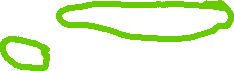
1. Gather ref URL, inscription text, date, material type and click map to get geo info or image to get image URL  
   

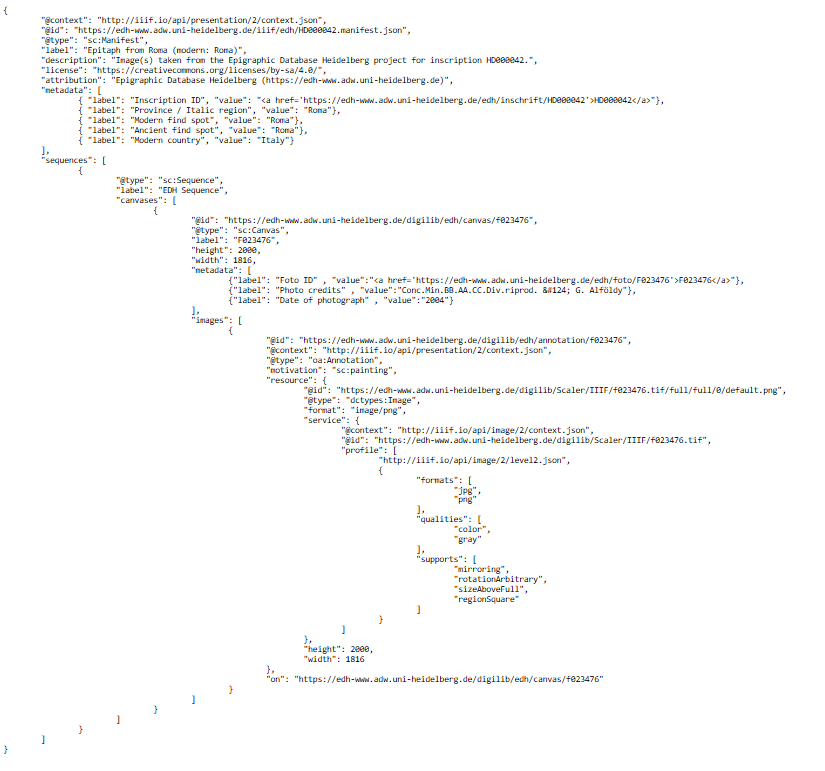


1. From map page gather Lat/Long  
   



1. From image page click to navigate to IIIF manifest or navigate to simple image page  
   



1. From IIIF Manifest gather the image URL from the sequences/canvases/images/resource/@id  
   



1. From the simple image page select the json link and gather the image URL  
   