

## INTRODUCTION

There are two activities this week. The first takes you through using the wikidata query service to enhance our data tables. The second introduces Palladio, a free visualization tool from Stanford University.

## ACTIVITY ONE – USE WIKIDATA TO ENHANCE DATA TABLES

This activity is based on the data tables you have created for previous sessions. You will need all the wikidata ids that you have entered to link the tables together.

Start with the people table. We want to add information about their date of birth and their place of birth, so we can plot that information on a map.

First, copy the authors column and paste it into NotePad++ (or TextMate). Do the same with the translators column. Sort and de-duplicate your list (manually, or go to edit – line operations in Notepad++). You should have a list of codes, all beginning Q.

For our SPARQL query, we will need to prefix the codes with wd: to show that they are wikidata references. Search for Q and replace with wd:Q. This is the SPARQL query:

```
SELECT ?person ?personLabel (YEAR(?birthdate) AS ?birthyear) ?birthplace
```

```
WHERE {  
VALUES ?person { wd:Q  
}
```

This is where we put the wikidata ids of the people in our table. If there are any problems with the query, try doing a few at a time.

```
OPTIONAL{ ?person wdt:P569 ?birthdate }  
OPTIONAL{ ?person wdt:P19 ?birthplace }
```

```
SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }  
} ORDER BY ?birthyear
```



@TAYOXFORD



EMMA.HUBER@BODLEIAN.OX.AC.UK



+44 1865 (2)78153

You should now have a table with information added to each person. Where the information was not available in wikidata, the cells are left blank. Download the resulting table as a csv file. Before you can use it further, you will need to do some data cleaning. Use find and replace to find <http://www.wikidata.org/entity/> and replace it with nothing so that only the codes are left. Add the new place codes to your places table. **You must add the new places to your places table before moving on to the next query.** Compare your table with the sample table in the resources for this session to check your looks as it should.

Next, we need to supplement the places table with the coordinates of each place. We'll also find some images of the places. Extract all the wikidata place codes as above, to create a deduplicated list of codes beginning with Q. As before we will need to replace the Q with wd:Q before doing the query. This is the SPARQL query for places:

```
SELECT ?place ?placeLabel ?placelocation ?image
WHERE {
VALUES ?place { wd:Q
}
```

This is where we put the wikidata ids of the places in our table. As before, if there are any problems with the query, try doing a few at a time.

```
OPTIONAL{ ?place wdt:P18 ?image }
OPTIONAL{ ?place wdt:P625 ?placelocation }
```

```
SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
} ORDER BY ?placeLabel
```

As before, you will need to clean the data by removing <http://www.wikidata.org/entity/>.

There is an added problem with the places table. To use the data with Palladio, we need to swap the coordinates around. There are various ways to do this.

Firstly, copy the coordinates column into NotePad++ (or TextMate). Use find and replace to remove the brackets and the word "Point". Now we are left with the numbers, separated by a space. We can express this as `^(\\S+)\\s(\\S+)$`. This regular expression means find the start of a line, followed by anything that is not a white space, and saving these characters as a set, then a white space, then a further set of non-space characters, followed by the end of a line.

Replace this with `\\2,\\s\\1` (The second set, a comma and a space, then the first set). The replacement expression is `$2,\\s$1` in TextMate.

Copy and paste the coordinates column back into the places csv file.

Again, compare your places.csv file with the sample file in the resources for this session.

You could append your data to the sample files before continuing with the second exercise.

**Please send your completed csv files to [emma.huber@bodleian.ox.ac.uk](mailto:emma.huber@bodleian.ox.ac.uk) so they can be included in the ongoing analysis of this dataset.**



@TAYOXFORD



EMMA.HUBER@BODLEIAN.OX.AC.UK



+44 1865 (2)78153

# ACTIVITY TWO – VISUALIZING DATA IN PALLADIO

Use the sample tables in the resources for this session. You could also append your own data (see exercise one).

Go to <https://hdlab.stanford.edu/palladio/> and click start. Drag and drop the works.csv file into the box and click load.

Drag and drop the works.csv file into the box and click load.

The screenshot shows the Palladio interface with a table titled 'Untitled' containing 4 rows. The columns are: ID (Text), Title (Text), Author (Number), Publisher (Number), Date of Publication (Date), Subject keywords (Text), Topic (Text), and pro-reform? (Text). A dialog box titled 'Verify special characters' is open, showing a 'Multiple values' field with a red highlight on a comma character. The dialog also includes a 'Verify special characters' button and a 'Multiple values' field with a red highlight on a comma character. Below the field, it says: 'If the dimension contains multiple values, insert the delimiter string above'.

Click on “Title” and then click on the red highlighted characters where it says “verify special characters”. This confirms that those characters have no special meaning in these fields. Move on to the author row in the works table. Click on it. This time, instead of verifying the special characters, you need to add a comma in the multiple values field, because commas are separating different authors of the same work. We also need to add a new table to link from these authors to our people table, so click on “add new table”, drag and drop the people table and click “load”.

The screenshot shows the 'Edit dimension' dialog in Palladio. The 'Title' field is 'Author', 'Data type' is 'Text', and 'Unique values' is 'Search'. A list of values is shown: Q1927120, Q38294, Q1234985, Q107698, Q317149, Q38370. The 'Multiple values' field is empty. The 'Extension' field is 'Choose a table' and 'Add a new table' button is visible. The dialog also includes a 'Verify special characters' button and a 'Multiple values' field with a red highlight on a comma character. Below the field, it says: 'If the dimension contains multiple values, insert the delimiter string above'.



@TAYOXFORD



EMMA.HUBER@BODLEIAN.OX.AC.UK



+44 1865 (2)78153

Work through all the tables in this way, as was demonstrated in the video for this session, until there are no red dots, and all the tables have been loaded. Make sure you rename the tables, to avoid getting confused later.

Click “Map” in the top bar, and then “New layer” in the box on the right. Click “Point to Point”, then in the source places select “Place of Birth” and in the Target places click “place of publication”.

Change the tooltip label to “place name”. Check the “size point” box, and choose a dimension to size the points by. Finally, click “add layer”.

If you wish, you can add another layer, and this time choose the tiles tab, and select terrain, to make your map more interesting.

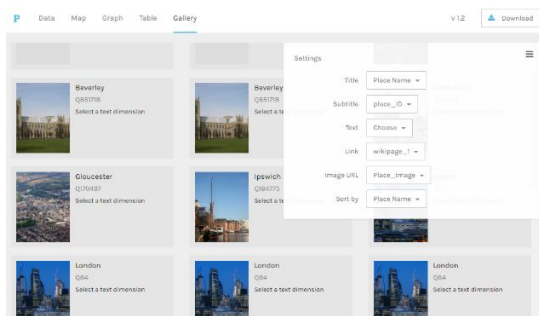
You could also add another layer, this time a points layer showing the places of publication, in a different colour.



On the graph tab, you can do simple network visualisations. Try setting the source to Author and the target to Translator or Publisher. Size the nodes by number of people:

Experiment with the timeline facets at the foot of the visualization.

Finally, have a look at the Gallery tab. Configure the settings to show the images included in the places csv file.



## FURTHER READING

There is a long, and from the comments, maintained list of DH GIS projects here:

<http://anterotesis.com/wordpress/mapping-resources/dh-gis-projects/>



@TAYOXFORD



EMMA.HUBER@BODLEIAN.OX.AC.UK



+44 1865 (2)78153