Hands-on (Details)
Obtain Data Using SPARQL

A. Obtain data from DBpedia
B. Clean up DBpedia data
C. Visualize the data

Background:
- About DBpedia http://wiki.dbpedia.org/about
- Query examples http://wiki.dbpedia.org/OnlineAccess

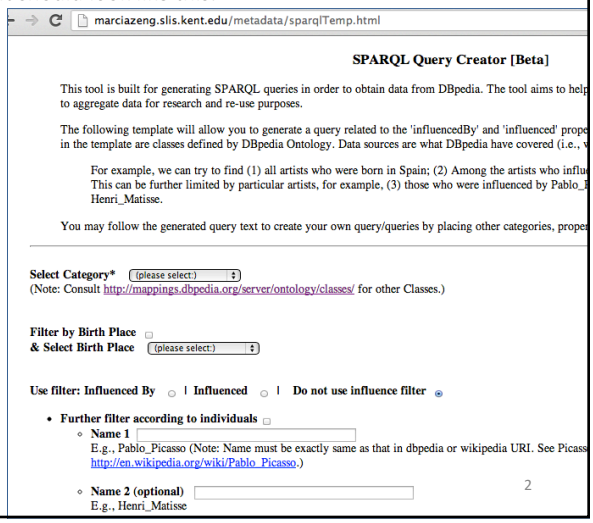
(Task A). Obtain data from DBpedia

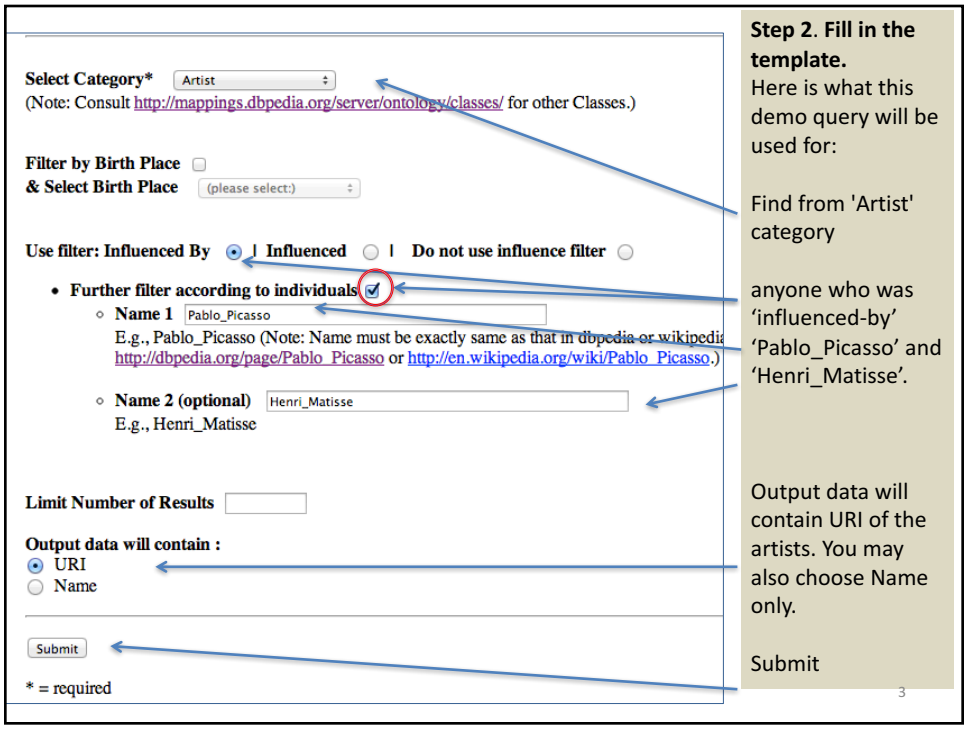
Use my SPARQL query generator to create a query text. (SPARQL is a RDF query language.)

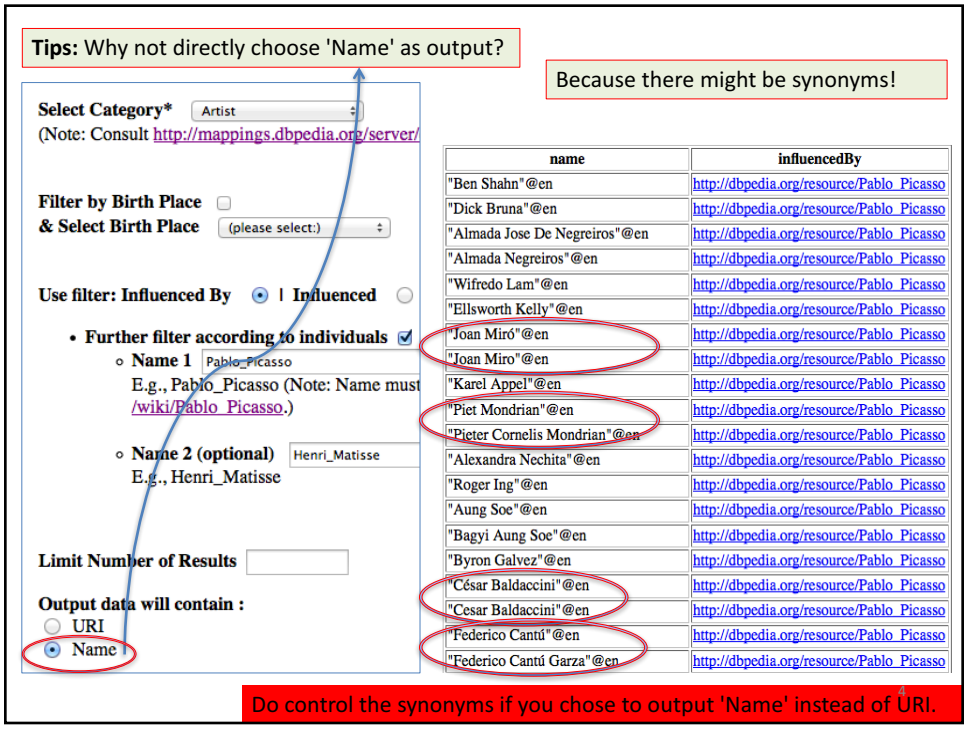
Step 1. Go to: http://marciazeng.slis.kent.edu/metadata/sparqlTemp.html

SPARQL language is not so easy to use if you are not an everyday user. So, I created a tool to make this step easier. It is the first time to be used by others besides me, so, be patient and let me know if you run into any problems.

It should look like this:




Select Category* Artist
(Note: Consult http://mappings.dbpedia.org/server/ontology/classes/ for other Classes.)
Filter by Birth Place
& Select Birth Place (please select:)
Use filter: Influenced By | Influenced | Do not use influence filter
Further filter according to individuals
Name 1 Pablo_Picasso
E.g., Pablo_Picasso (Note: Name must be exactly same as that in dbpedia or wikipedi
http://dbpedia.org/page/Pablo_Picasso or http://en.wikipedia.org/wiki/Pablo_Picasso.)
Name 2 (optional) Henri_Matisse
E.g., Henri_Matisse
Limit Number of Results
Output data will contain :
URI
Name
Submit
* = required
Step 2. Fill in the template.
Here is what this demo query will be used for:
Find from 'Artist' category
anyone who was 'influenced-by' 'Pablo_Picasso' and 'Henri_Matisse'.
Output data will contain URI of the artists. You may also choose Name only.
Submit
3


Tips: Why not directly choose 'Name' as output?
Because there might be synonyms!
Select Category* Artist
(Note: Consult http://mappings.dbpedia.org/server/
Filter by Birth Place
& Select Birth Place (please select:)
Use filter: Influenced By | Influenced
Further filter according to individuals
Name 1 Pablo_Picasso
E.g., Pablo_Picasso (Note: Name must
/wiki/Pablo_Picasso.)
Name 2 (optional) Henri_Matisse
E.g., Henri_Matisse
Limit Number of Results
Output data will contain :
URI
Name
name
influencedBy
"Ben Shahn"@en http://dbpedia.org/resource/Pablo_Picasso
"Dick Bruna"@en http://dbpedia.org/resource/Pablo_Picasso
"Almada Jose De Negreiros"@en http://dbpedia.org/resource/Pablo_Picasso
"Almada Negreiros"@en http://dbpedia.org/resource/Pablo_Picasso
"Wifredo Lam"@en http://dbpedia.org/resource/Pablo_Picasso
"Ellsworth Kelly"@en http://dbpedia.org/resource/Pablo_Picasso
"Joan Miró"@en http://dbpedia.org/resource/Pablo_Picasso
"Joan Miro"@en http://dbpedia.org/resource/Pablo_Picasso
"Karel Appel"@en http://dbpedia.org/resource/Pablo_Picasso
"Piet Mondrian"@en http://dbpedia.org/resource/Pablo_Picasso
"Pieter Cornelis Mondrian"@en http://dbpedia.org/resource/Pablo_Picasso
"Alexandra Nechita"@en http://dbpedia.org/resource/Pablo_Picasso
"Roger Ing"@en http://dbpedia.org/resource/Pablo_Picasso
"Aung Soe"@en http://dbpedia.org/resource/Pablo_Picasso
"Bagyi Aung Soe"@en http://dbpedia.org/resource/Pablo_Picasso
"Byron Galvez"@en http://dbpedia.org/resource/Pablo_Picasso
"César Baldaccini"@en http://dbpedia.org/resource/Pablo_Picasso
"Cesar Baldaccini"@en http://dbpedia.org/resource/Pablo_Picasso
"Federico Cantú"@en http://dbpedia.org/resource/Pablo_Picasso
"Federico Cantú Garza"@en http://dbpedia.org/resource/Pablo_Picasso
Do control the synonyms if you chose to output 'Name' instead of URI.
4

More Tips

1. The first name and last name of a person need to be bound together by the dash (e.g., **Pablo_Picasso**).
   - This will be important for **visualization**. Otherwise the visualization software will treat as two separate names.
   - If you requested the data from URIs, the dash for a name should be automatically handled.

2. Different domains may have different data for 'influenced' and 'influencedBy'. If no luck for a name you choose, try a philosopher or an artist.

5

Step 3. After you finish that form and "submit," you should get a new page like this, which give you the formed query. Copy the query text (copy everything in this box as a whole chunk; do not miss anything).
Then, go to DBpedia's SPARQL endpoint by clicking the link.

After you have submitted your choices in that template, you will be given this query text list.

1. Copy the text from the box.

2. Go to DBpedia's SPARQL endpoint.

Step 1: Copy the query text from the following box

```
PREFIX dataset: <http://dbpedia.org/ontology/>

SELECT ?uri ?influencedBy
WHERE
{
 {
 ?uri a dataset:Artist .
 ?uri dataset:influencedBy ?influencedBy .
 filter regex(?influencedBy, 'Pablo_Picasso', 'i') .
 }
 UNION
 {
 ?uri a dataset:Artist .
 ?uri dataset:influencedBy ?influencedBy .
 filter regex(?influencedBy, 'Henri_Matisse', 'i') .
 }
}
```

1 Make sure no space after his name.

2

Step 2: Go to dppedia.org/sparql and paste into the query text box

6

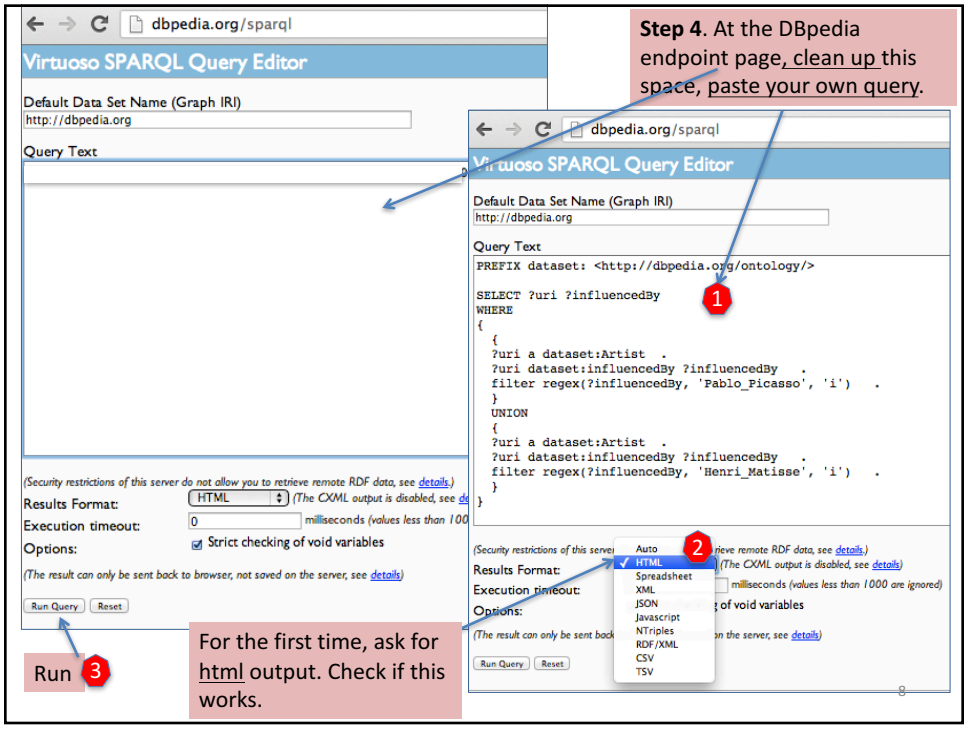
Now you are at DBpedia's SPARQL endpoint "**Virtuoso SPARQL Query Editor**"
(to be continued in next slide →)

Tips: some web browsers will not work well with SPARQL.
Always use Firefox or Chrome.

It should look like this:

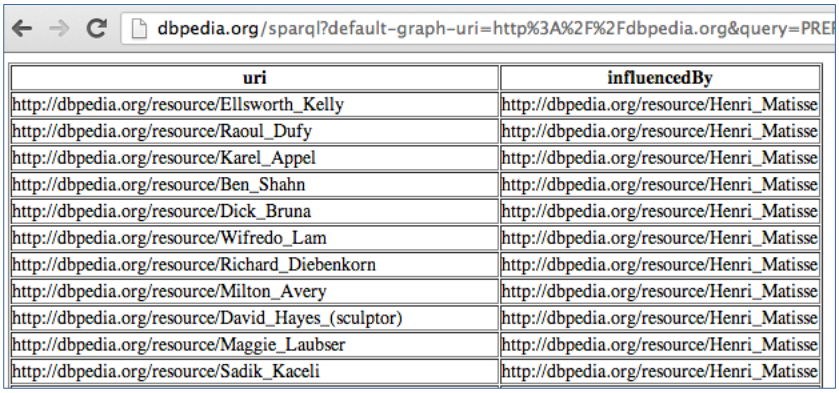
← clean up this space

7



8

Assume your html output worked well, like below. [If you choose output names, not URIs, you will not have those http:// part and will just have the names.



dbpedia.org/sparql?default-graph-uri=http%3A%2F%2Fdbpedia.org&query=PREF

uri	influencedBy
http://dbpedia.org/resource/Ellsworth_Kelly	http://dbpedia.org/resource/Henri_Matisse
http://dbpedia.org/resource/Raoul_Dufy	http://dbpedia.org/resource/Henri_Matisse
http://dbpedia.org/resource/Karel_Appel	http://dbpedia.org/resource/Henri_Matisse
http://dbpedia.org/resource/Ben_Shahn	http://dbpedia.org/resource/Henri_Matisse
http://dbpedia.org/resource/Dick_Bruna	http://dbpedia.org/resource/Henri_Matisse
http://dbpedia.org/resource/Wifredo_Lam	http://dbpedia.org/resource/Henri_Matisse
http://dbpedia.org/resource/Richard_Diebenkorn	http://dbpedia.org/resource/Henri_Matisse
http://dbpedia.org/resource/Milton_Avery	http://dbpedia.org/resource/Henri_Matisse
http://dbpedia.org/resource/David_Hayes_(sculptor)	http://dbpedia.org/resource/Henri_Matisse
http://dbpedia.org/resource/Maggie_Laubser	http://dbpedia.org/resource/Henri_Matisse
http://dbpedia.org/resource/Sadik_Kaceli	http://dbpedia.org/resource/Henri_Matisse

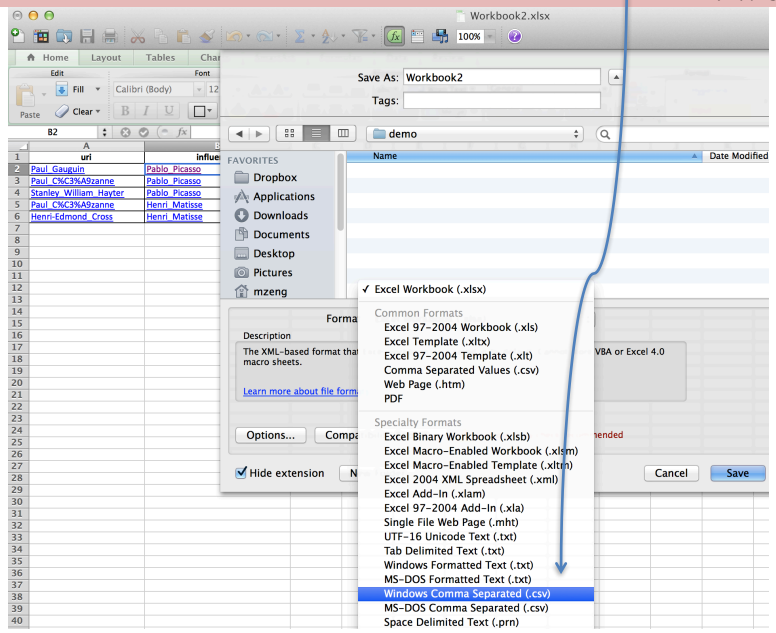
Step 5. If it looks correct, then:
"select all" the table showing on the html page, copy-paste to your spreadsheet (see next slide).

Note: If you know how to deal with other format, you may ask for output in CSV directly. Save the csv file into your own computer. Name with your [file-name].csv. However, sometimes, the auto saved CSV file might not work on your computer.

9

After copying onto a spreadsheet, save the spreadsheet as a CSV file, for visualization use.
File → Save As – > [choose .CSV from the save as format list.]

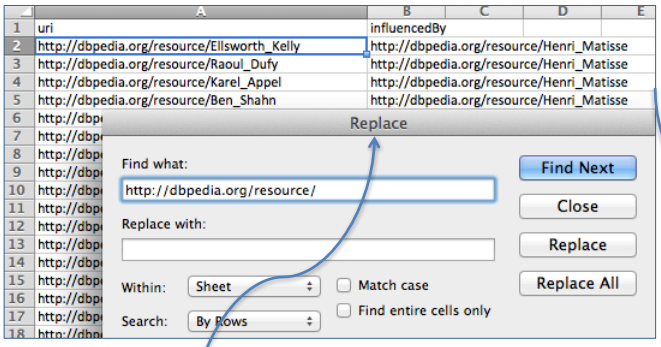
FINISHED.



10

Task B. Clean up DBpedia data

Open your *.csv file in Microsoft Office Excel (spreadsheet), clean up the data. Be ready for visualization.



	A	B
1	uri	influencedBy
2	http://dbpedia.org/resource/Ellsworth_Kelly	http://dbpedia.org/resource/Henri_Matisse
3	http://dbpedia.org/resource/Raoul_Dufy	http://dbpedia.org/resource/Henri_Matisse
4	http://dbpedia.org/resource/Karel_Appel	http://dbpedia.org/resource/Henri_Matisse
5	http://dbpedia.org/resource/Ben_Shahn	http://dbpedia.org/resource/Henri_Matisse

Note, the first name and last name are bound together by the dash (e.g., Pablo_Picasso) . This will be important to visualize.

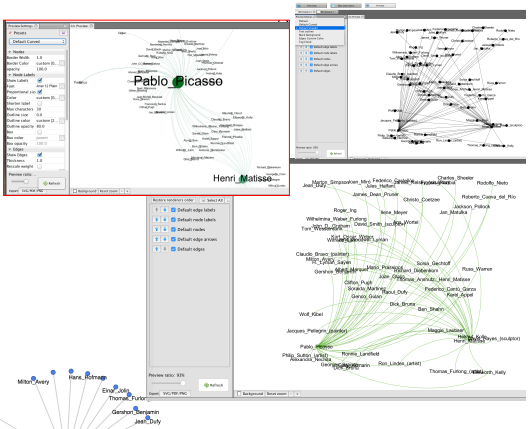
- For better visualization results, I usually delete the http://dbpedia.org/resource/ part. (In Excel, choose Edit → Replace".)
- You may clean up other unnecessary words, e.g., those qualifiers such as "_(painter)".
- Also find original name such as "Joan_Miró" to replace "Joan_Mir%C3%B3" in the spreadsheet.
- Sort by Column A. Delete any real duplicate line.

	A	B
1	uri	influencedBy
2	Ellsworth_Kelly	Henri_Matisse
3	Raoul_Dufy	Henri_Matisse
4	Karel_Appel	Henri_Matisse
5	Ben_Shahn	Henri_Matisse
6	Albert_Marquet	Henri_Matisse
7	Dick_Bruna	Henri_Matisse
8	Wifredo_Lam	Henri_Matisse
9	Richard_Diebenkorn	Henri_Matisse
10	Milton_Avery	Henri_Matisse

11

Task C. Visualize the data

- Tools can be used:
  - Gephi
    http://gephi.org/
  - Google Fusion Tables



https://support.google.com/fusiontables

12