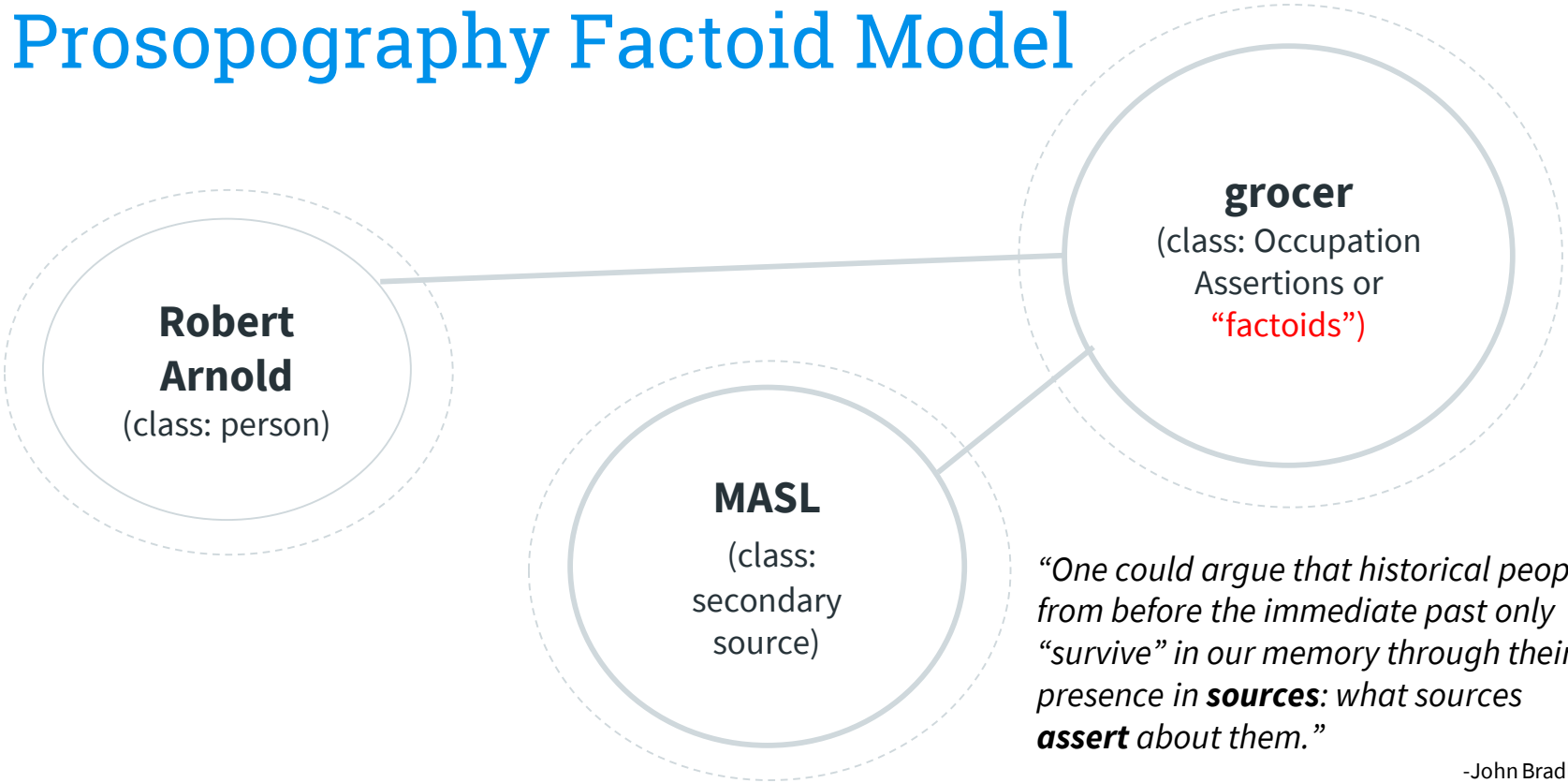
Linked Data and Digital Prosopography

Slide Presentation for the International Symposium on Digital Scholarship
Fordham London Centre, June 3-5, 2019
Katherina Fostano, kfostano@fordham.edu

Prosopography Factoid Model



*"One could argue that historical people from before the immediate past only "survive" in our memory through their presence in **sources**: what sources **assert** about them."*

-John Bradley,
A Semantic Web understanding of the Factoid Prosopography model."

Technical Goals

1. To **support** prosopographical studies of the people of Medieval London

1. To facilitate **access** to our data about Medieval Londoners

2. To **connect** our data to other published of Medieval Londoners online

3. To accomplish these ends at a low cost using **open source** technology

Digital Strategy: Collection and Preparation

Harvesting

Aggregating Data from digitized sources online, **MASL**, **BEAVEN**.

Structuring

Standardising to fit the **metadata schema** created by Dr.Kowaleski.

Presenting

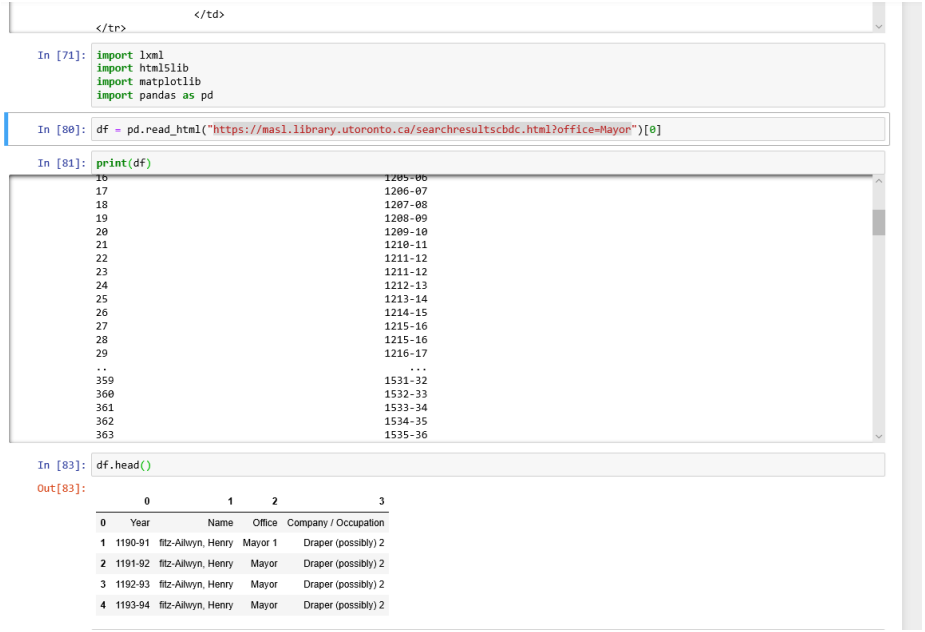
Creating an **interface** where people can **access** our data and also **contribute** to our data.

Harvesting

Harvesting or web scraping is the process where a small script is used to automatically extract a large amount of data from a website.

MASL Dataset
Harvested with Python using the following libraries; Beautiful Soup, Pandas, Numpy

Beaven Dataset
Harvested with a custom **Perl** script programmed
by Dr. Liz Duchovni.



```
                    </td>
</tr>

In [71]: import lxml
         import html5lib
         import matplotlib
         import pandas as pd

In [80]: df = pd.read_html("https://masl.library.utoronto.ca/searchresultscbdc.html?office=Mayor")[0]

In [81]: print(df)
16                                    1205-06
17                                    1206-07
18                                    1207-08
19                                    1208-09
20                                    1209-10
21                                    1210-11
22                                    1211-12
23                                    1211-12
24                                    1212-13
25                                    1213-14
26                                    1214-15
27                                    1215-16
28                                    1215-16
29                                    1216-17
..                                        ...
359                                   1531-32
360                                   1532-33
361                                   1533-34
362                                   1534-35
363                                   1535-36

In [83]: df.head()
Out[83]:
```

	0	1	2	3
0	Year	Name	Office	Company / Occupation
1	1190-91	fitz-Ailwyn, Henry	Mayor 1	Draper (possibly) 2
2	1191-92	fitz-Ailwyn, Henry	Mayor	Draper (possibly) 2
3	1192-93	fitz-Ailwyn, Henry	Mayor	Draper (possibly) 2
4	1193-94	fitz-Ailwyn, Henry	Mayor	Draper (possibly) 2

Web Scraping Tutorials:
https://programminghistorian.org/en/lessons/intro-to-beautiful-soup

Standardizing

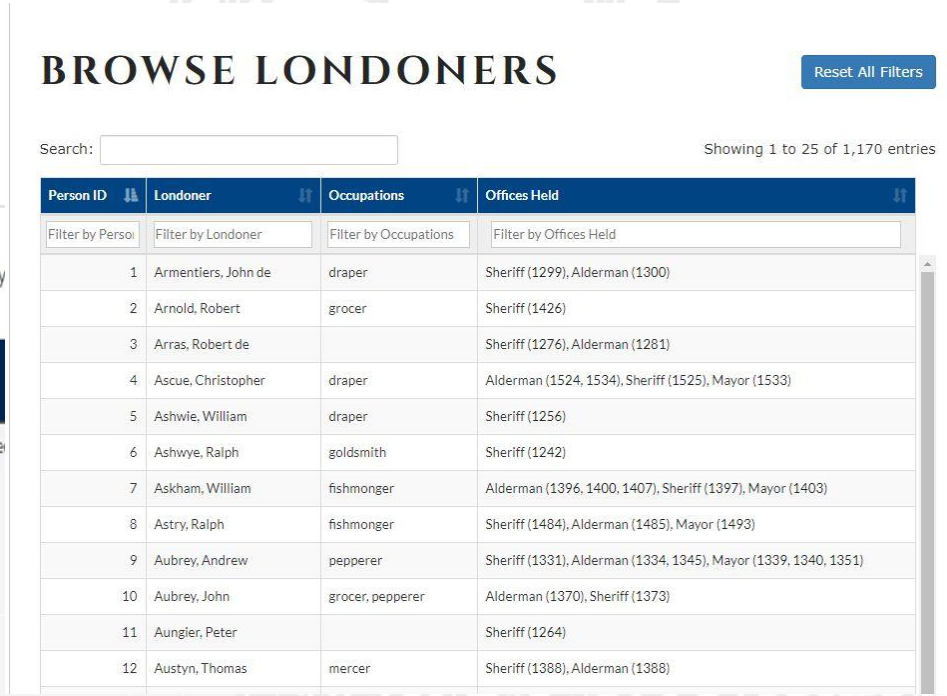
Metadata	Data Type	Description
Rec_ID	int	A unique whole number, no decimals, randoml
Pers_ID	int	A whole number, randomly assigned. Each pers
Pers_Rec	int	A unique number with three decimal places. Th
Title	varchar	Note if this person is given a title (Sir, Master,
Doc_Forename	varchar	Person's forename as it appears in the (manusc
Doc_Surname	varchar	Person's surname as it appears in this documen
Standard_Name	varchar	Standardized (modern English) version of the p
Alias	varchar	Alias of individual given in this record. The ali
Identifier	char(1)	A=alien; C=clergy; M=married couple; W=wor
Citizen	varchar	Insert Y if person noted as a citizen of London
Guild	varchar	Note gild (capitalized) if the individual is name
Occupation	varchar	Note occupation (modernized, lower case) spec
Occ_Flag	char(1)	Code indicating quality of data in Occupation i D=occupation explicitly stated in this documen E=occupation noted elsewhere (source is specif O=occupation surmised from another source by P=some evidence in this record that the individ S=strong evidence in this record that the indivi T=this record notes when individual translated Number: when a number is given after the D, C
Activity	text	Summarize the circumstances of the individual For MASL, Beaven, and many other secondary For Beaven entries noting service of an Alderm

Schema by Dr. Maryanne Kowaleski

Rec_ID	Pers_ID	Pers_Rec	Title	Doc_Forer	Doc_Surname	Standard_Name	Alias	Identifier	Citizen
2850	1047	1047.001				—, John			Y
2852	1048	1048.001				—, William			Y
972	100	100.001				Abbot, John			Y
1225	100	100.002				Abbot, John			Y
3132	101	101.002				Abraham, William			Y
990	101	101.001				Abraham, William			Y
2856	945	945.001			Abyndone	Abyndon, Nicholas de			Y
2293	102	102.003				Abyndon, Simon de			Y

Presentation

Londoner View



BROWSE LONDONERS

Reset All Filters

Search:

Showing 1 to 25 of 1,170 entries

Person ID	Londoner	Occupations	Offices Held
Filter by Perso	Filter by Londoner	Filter by Occupations	Filter by Offices Held
1	Armentiers, John de	draper	Sheriff (1299), Alderman (1300)
2	Arnold, Robert	grocer	Sheriff (1426)
3	Arras, Robert de		Sheriff (1276), Alderman (1281)
4	Ascue, Christopher	draper	Alderman (1524, 1534), Sheriff (1525), Mayor (1533)
5	Ashwie, William	draper	Sheriff (1256)
6	Ashwye, Ralph	goldsmith	Sheriff (1242)
7	Askham, William	fishmonger	Alderman (1396, 1400, 1407), Sheriff (1397), Mayor (1403)
8	Astry, Ralph	fishmonger	Sheriff (1484), Alderman (1485), Mayor (1493)
9	Aubrey, Andrew	pepperer	Sheriff (1331), Alderman (1334, 1345), Mayor (1339, 1340, 1351)
10	Aubrey, John	grocer, pepperer	Alderman (1370), Sheriff (1373)
11	Aungier, Peter		Sheriff (1264)
12	Austyn, Thomas	mercer	Sheriff (1388), Alderman (1388)

Record View

Browse All Search Records

1 of 108 Next Page

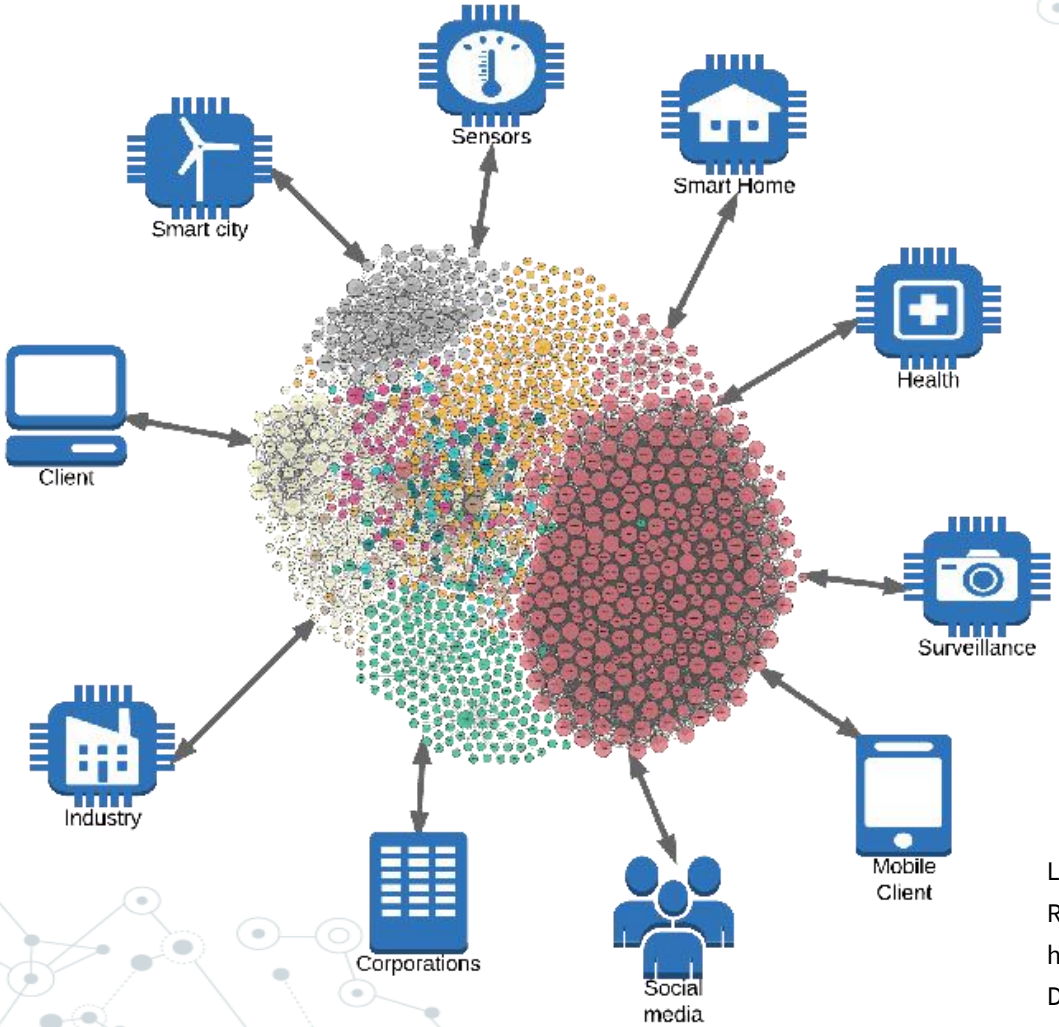
Sort by

Person ID	Londoner	Occupation	Civic Office	Ward	Year	Activity
1047	–, John		Serjeant of the Channel (Surveyor of Streets and Lanes)		1385	Served as Serjeant of the Channel (Surveyor of Stre 275)
1048	–, William		Common Clerk		1307	Served as Common Clerk. (LBB, 197)
						364.
100	Abbot, John	vintner	Sheriff		1428	Elected Sheriff; chosen by Mayor. MASL. Full Record

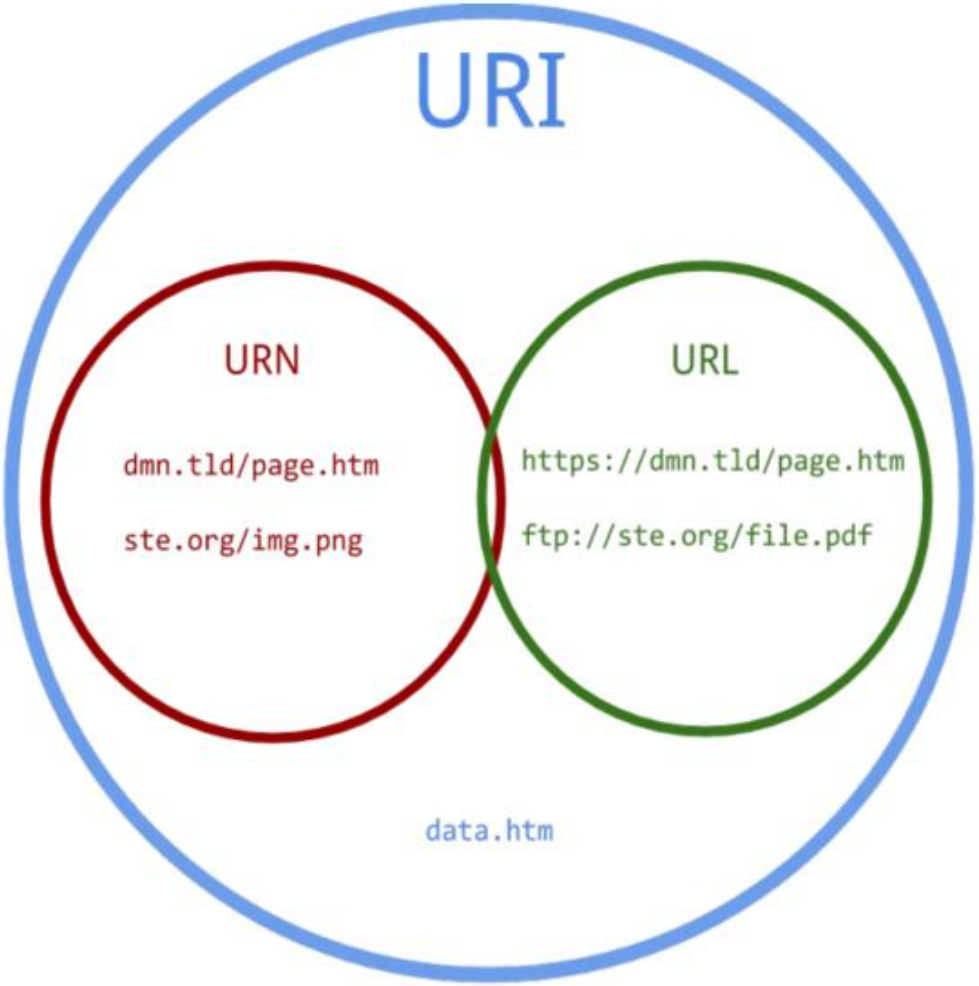
Semantic enrichment is adding a layer of topical metadata to content so that machines can make sense of it and build connections to it

- *Michael Clarke*

Pause and smile



Linked Data Authorization platform - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Linked-Open-Data-Cloud-and-Data-publishers_fig1_321328847 [accessed 30 May, 2019]


URI
URN
dmn.tld/page.htm
ste.org/img.png
URL
https://dmn.tld/page.htm
ftp://ste.org/file.pdf
data.htm

Linked data *is a framework for describing, expressing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using.*

-Tim Berners-Lee

John Bradley: Why Linked Data is Relevant in Prosopography

Published prosopography offers an almost ideal kind of research that could be expressed as linked data.

First

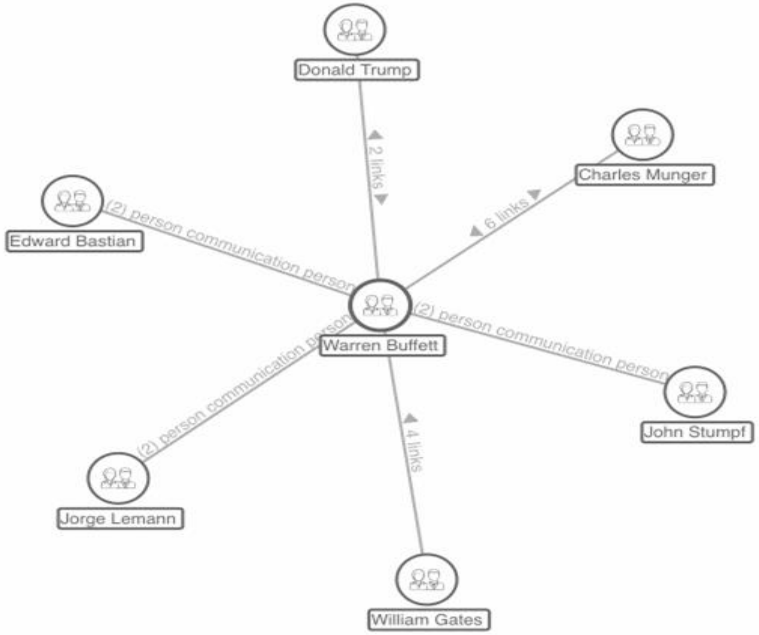
prosopography aims to develop the **identity** of historical **people** in a way that crosses multiple historical **sources**, these identified historical people act, by their very nature, as a kind of **interlinking** between these different sources.

Second

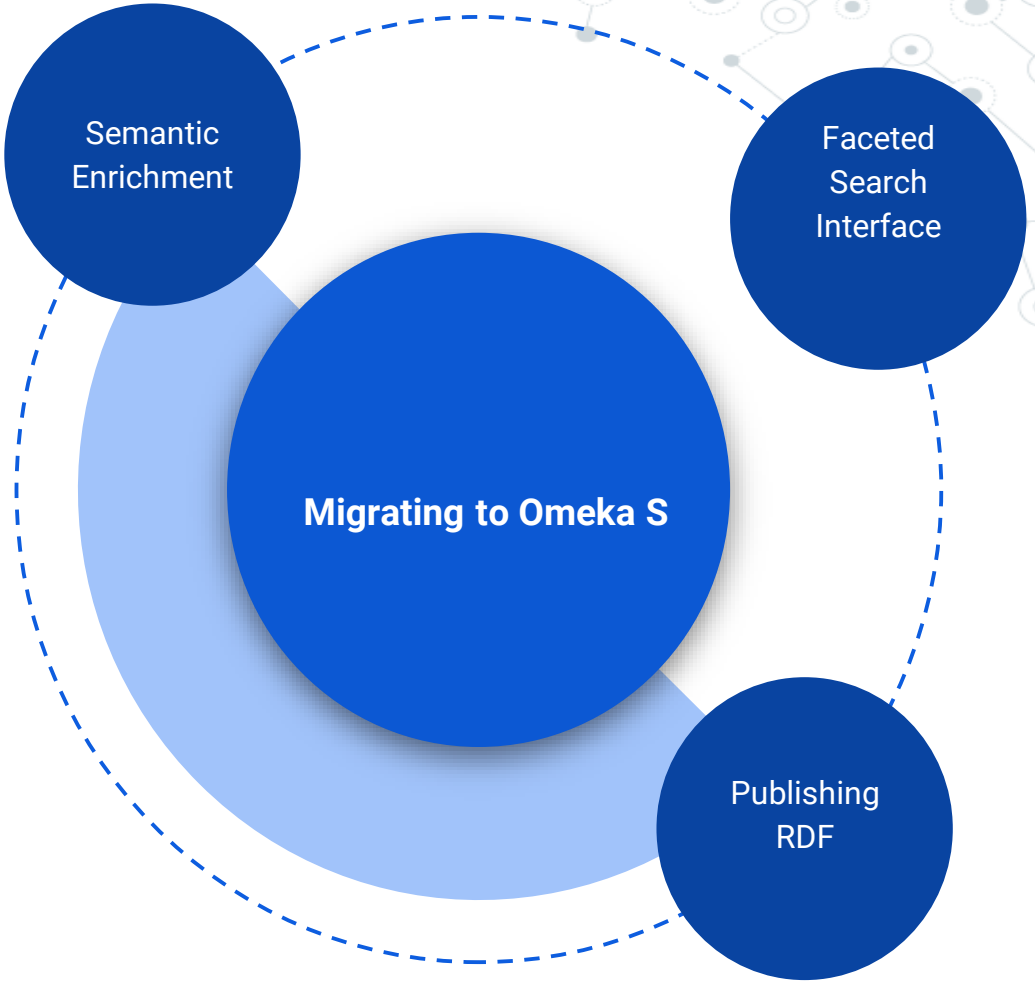
prosopography is, at least potentially, a **global object** — something used by other researchers throughout the world as a source for identities for historical people.

John Bradley, DPRR RDF Services Documentation, http://romanrepublic.ac.uk/rdf/doc/why.html

Global Graph



Migrating to Omeka S

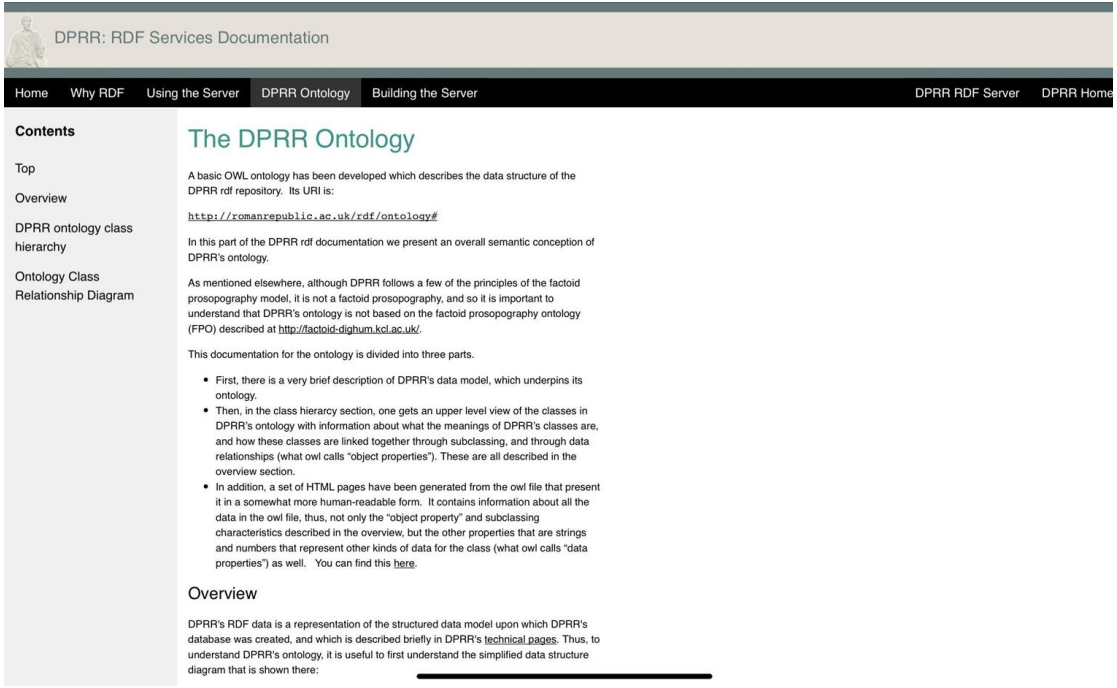


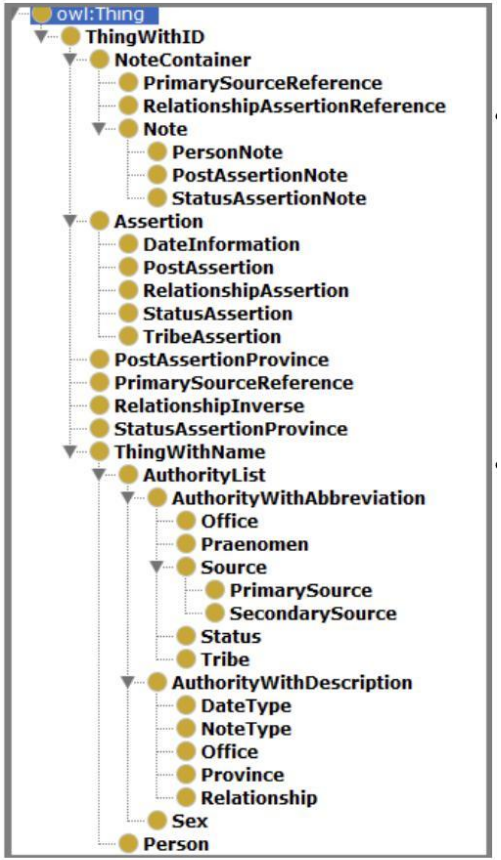
Omeka Semantic

A free open source content management system that lets users create and manage websites, and publish items with linked open data.

- **Linked Data Vocabularies:** semantic vocabularies are the basic building blocks for inference techniques on the Semantic Web
- **Computer Ontologies**: Formal specification of the terms in a domain and the relationships among them; the basic building block of ontologies are **triples**; Subject-predicate-object.

DPRR Ontology



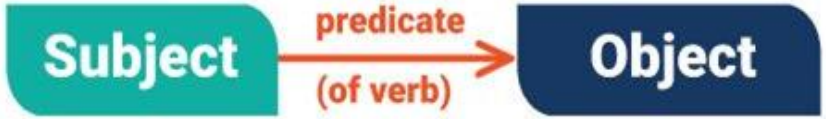


http://romanrepublic.ac.uk/rdf/doc/ontology.html

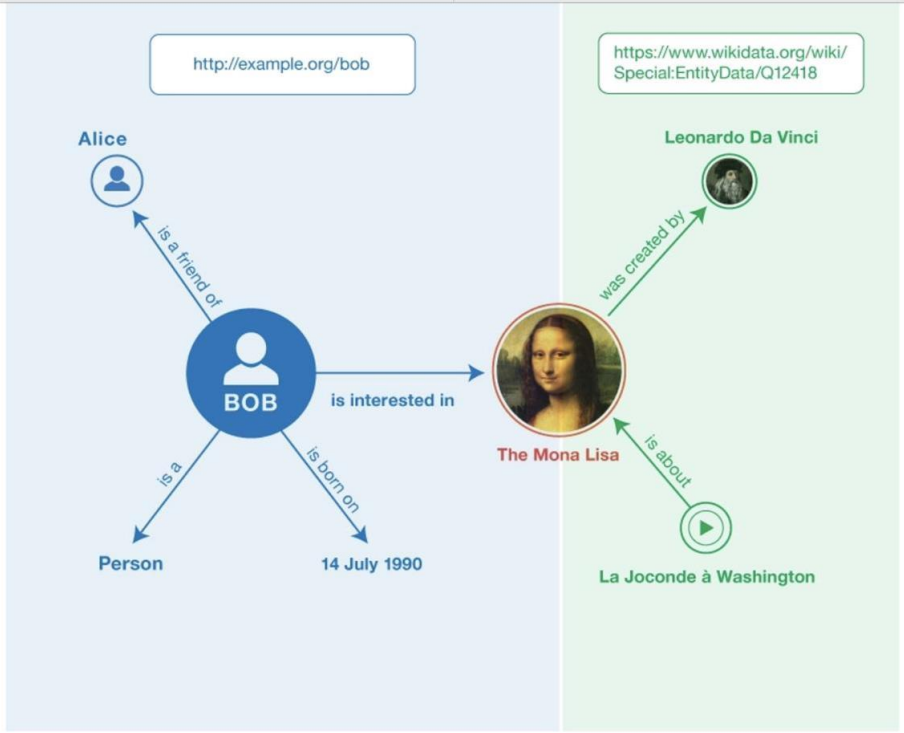
Resource Description Framework (RDF)

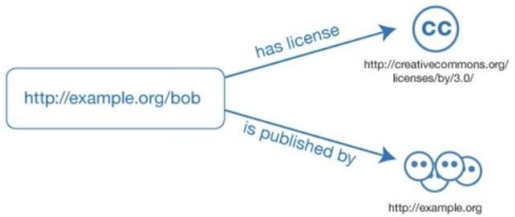
RDF is a general method for describing data by **defining relationships** between **data objects.**

An RDF statement expresses a relationship between two resources. The subject and the object represent the two resources being related; the predicate represents the nature of their relationship. The relationship is phrased in a directional way (from subject to object) and is called in RDF a property. Because RDF statements consist of three elements they are called triples.



World Wide Web consortium (https://www.w3.org/TR/rdf11-primer/#section-triple).





<Bob> **<is a>** <person>.
<Bob> **<is a friend of>** <Alice>.
<Bob> **<is born on>** <the 4th of July 1990>.
<Bob> **<is interested in>** <the Mona Lisa>.

URI for Identifying Historical People

<URI-of-HenryBarton> :aPerson;
:hasName "HenryBarton";
:hasOccupationAssertion <skinner>;
:hasAuthorityRecord
<http://www.medeivallondoners.org/record/18333>

<http://www.medeivallondoners.org/person/barton-henry>
<owl:sameAs>
<https://www.historyofparliamentonline.org/volume/1386-1421/member/barton-henry-1435>

Sharing

Wikidata: Andrew Gray's Wiki Project British Politicians

This project is aiming to compile structured data for all Members of Parliament in the United Kingdom, and its predecessors.

https://www.wikidata.org/wiki/Q5717856

MLD Dataset

Thanks!

You can find me at:

@kfostano

kfostano@fordham.edu