

# Semantic Web Systems 2015 - 2016

## Solution Guide to Assignment 1

*This solution guide summarises one possible approach for completing Assignment 1.*

*NOTE 1: this is not a complete solution by itself,*

*NOTE 2: this is not the only correct approach, and others are possible.*

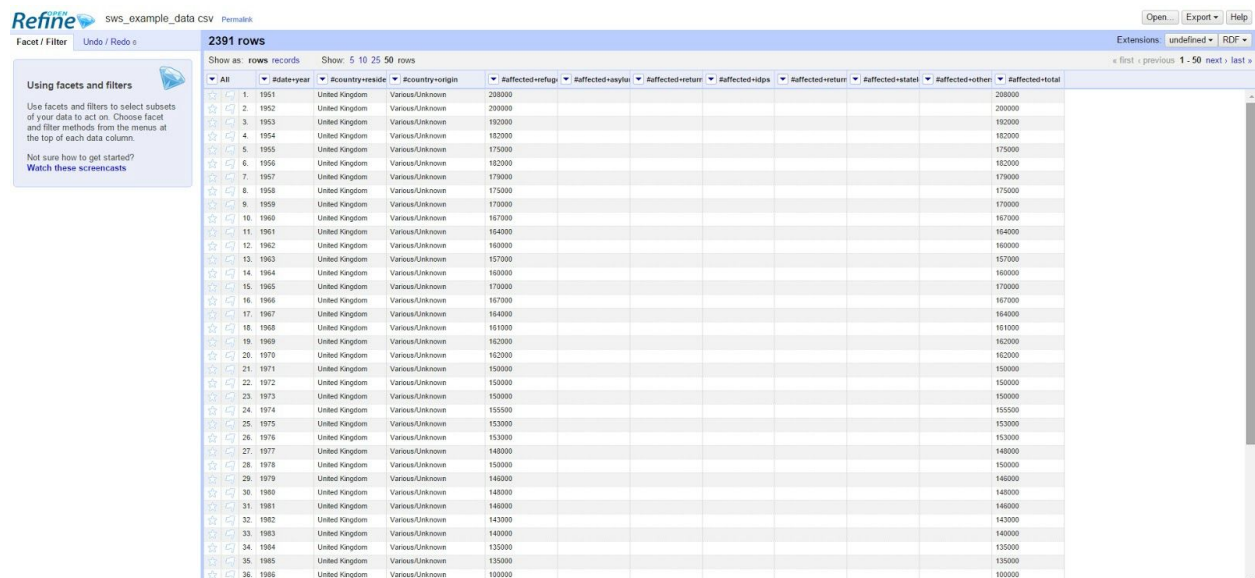
*This solution guide will use the following dataset as an example:*

<https://data.hdx.rwllabs.org/dataset/refugees-residing-gbr/resource/4a6d0850-f1e8-4ebd-88c6-1b9b21d2419f>

## CSV to RDF transformation process

(1) I have installed Open Refine (http://openrefine.org/) and its RDF refine extension (http://refine.deri.ie/)

(2) I have loaded the CSV file in Open Refine



The screenshot shows the OpenRefine web interface. At the top, it says 'Refine' and 'sws\_example\_data.csv'. Below this, there's a 'Facet / Filter' section on the left with a 'Using facets and filters' tip. The main area displays a table with 2391 rows. The table has columns: #date-year, #country-reside, #country-origin, #affected-refuge, #affected-asylum, #affected-return, #affected-ridps, #affected-return, #affected-retail, #affected-retail, and #affected-total. The data shows counts for various years from 1951 to 1986, all originating from 'Various:Unknown' and residing in 'United Kingdom'.

	#date-year	#country-reside	#country-origin	#affected-refuge	#affected-asylum	#affected-return	#affected-ridps	#affected-return	#affected-retail	#affected-retail	#affected-total
1	1951	United Kingdom	Various:Unknown	208000							208000
2	1952	United Kingdom	Various:Unknown	209000							209000
3	1953	United Kingdom	Various:Unknown	192000							192000
4	1954	United Kingdom	Various:Unknown	182000							182000
5	1955	United Kingdom	Various:Unknown	175000							175000
6	1956	United Kingdom	Various:Unknown	182000							182000
7	1957	United Kingdom	Various:Unknown	179000							179000
8	1958	United Kingdom	Various:Unknown	175000							175000
9	1959	United Kingdom	Various:Unknown	176000							176000
10	1960	United Kingdom	Various:Unknown	167000							167000
11	1961	United Kingdom	Various:Unknown	164000							164000
12	1962	United Kingdom	Various:Unknown	160000							160000
13	1963	United Kingdom	Various:Unknown	157000							157000
14	1964	United Kingdom	Various:Unknown	160000							160000
15	1965	United Kingdom	Various:Unknown	170000							170000
16	1966	United Kingdom	Various:Unknown	167000							167000
17	1967	United Kingdom	Various:Unknown	164000							164000
18	1968	United Kingdom	Various:Unknown	161000							161000
19	1969	United Kingdom	Various:Unknown	162000							162000
20	1970	United Kingdom	Various:Unknown	162000							162000
21	1971	United Kingdom	Various:Unknown	150000							150000
22	1972	United Kingdom	Various:Unknown	150000							150000
23	1973	United Kingdom	Various:Unknown	150000							150000
24	1974	United Kingdom	Various:Unknown	155000							155000
25	1975	United Kingdom	Various:Unknown	153000							153000
26	1976	United Kingdom	Various:Unknown	153000							153000
27	1977	United Kingdom	Various:Unknown	148000							148000
28	1978	United Kingdom	Various:Unknown	150000							150000
29	1979	United Kingdom	Various:Unknown	146000							146000
30	1980	United Kingdom	Various:Unknown	148000							148000
31	1981	United Kingdom	Various:Unknown	146000							146000
32	1982	United Kingdom	Various:Unknown	143000							143000
33	1983	United Kingdom	Various:Unknown	140000							140000
34	1984	United Kingdom	Various:Unknown	135000							135000
35	1985	United Kingdom	Various:Unknown	135000							135000
36	1986	United Kingdom	Various:Unknown	130000							130000

(3) I have loaded the following reconciliation service in Open Refine, in order to reconcile data with Freebase:

<http://reconcile.freebaseapps.com/reconcile>

(more info here: <https://github.com/OpenRefine/OpenRefine/wiki/Reconciliation>)

(4) I started the reconciliation by clicking on the column menu, then selecting “Reconcile” and “Start reconciling”.

(5) Then I selected the Freebase reconciliation service and specified I was interested in reconciling only entities of type “Country”.

(6) After the reconciliation finished, I selected “Match each cell to its best candidate” from the “Reconcile” -> “Action” menu.

(7) I have create a new column with the reconciled data using the “Edit column” -> “Add column based on this column” function, and extracted the reconciled data using the “cell.recon.match.id” GREL expression.

For example the following country strings were matched to the following URIs:

Brazil	<a href="http://rdf.freebase.com/ns/m.015fr">http://rdf.freebase.com/ns/m.015fr</a>
Cameroon	<a href="http://rdf.freebase.com/ns/m.01nln">http://rdf.freebase.com/ns/m.01nln</a>
Somalia	<a href="http://rdf.freebase.com/ns/m.06tgw">http://rdf.freebase.com/ns/m.06tgw</a>
Sierra Leone	<a href="http://rdf.freebase.com/ns/m.06s_2">http://rdf.freebase.com/ns/m.06s_2</a>
Rwanda	<a href="http://rdf.freebase.com/ns/m.06dfg">http://rdf.freebase.com/ns/m.06dfg</a>

(8) I have created an RDF skeleton for the conversion:

The screenshot shows the Refine interface with a CSV file named 'sws\_example\_data.csv' open. The 'RDF Schema Alignment' dialog is displayed, showing the 'RDF Skeleton' tab. The 'Base URI' is set to 'http://vocab.inf.ed.ac.uk/sws/s1054760/sws/data/'. The dialog lists available prefixes and shows mappings for various RDF nodes. The background shows a table of data with columns for 'date+year', 'country+origin', 'affected+refugees', 'reconciled\_country URI', and 'affected+total'.

In this data model, I chose to divide each row of the CSV file into several RDF nodes, each one representing the population measurement for one particular type of population (e.g. asylum seekers from country X measured in year YYYY).

(9) I have added this base URI for the creation of new instances:

<http://vocab.inf.ed.ac.uk/sws/s1054760/sws/data/>

(10) I have connected each population type with the other cells as follows:

With this relation, I linked the population type to the year their numbers have been measured:

<http://www.europeana.eu/schemas/edm/year>

With this relation, I linked the population type with the country of residence:

<http://purl.org/healthcarevocab/v1#CountryOfResidence>

The country of residence is always the same (UK), I have reused a URI for it, instead of the label.

With this relation, I linked the population type with the country of origin:

<http://dbpedia.org/ontology/countryOrigin>

With this relation, I linked the population type with the number of people in that population for that year:

<http://dbpedia.org/ontology/populationTotal>

(11) To differentiate the various population types, I have assigned them to the corresponding class:

<http://vocab.inf.ed.ac.uk/sws/s1054760/sws/vocab/Refugees>

<http://vocab.inf.ed.ac.uk/sws/s1054760/sws/vocab/AsylumSeekers>

<http://vocab.inf.ed.ac.uk/sws/s1054760/sws/vocab/ReturnedRefugees>

<http://vocab.inf.ed.ac.uk/sws/s1054760/sws/vocab/IDPs>

<http://vocab.inf.ed.ac.uk/sws/s1054760/sws/vocab/ReturnedIDPs>

<http://vocab.inf.ed.ac.uk/sws/s1054760/sws/vocab/StatelessPersons>

<http://vocab.inf.ed.ac.uk/sws/s1054760/sws/vocab/OtherPopulationOfConcern>

<http://vocab.inf.ed.ac.uk/sws/s1054760/sws/vocab/TotalPopulationOfConcern>

*\*(IDP = Internally Displaced Persons)*

These are classes I created, as I could not find a comprehensive and stable vocabulary that included all of those concepts. It should be noted, for example, that I could not have reused the DBpedia "<http://dbpedia.org/resource/Refugee>" class, as it refers to the concept of a single refugee. In my case, however, I refer to "Refugees", defined as a population type, measured at a specific time, which has a country of origin and a country of residence.

(12) I generated the data by clicking "Export" -> "RDF as Turtle"

(13) I created a Turtle vocabulary file listing all the classes and relations I used using the RDFS data model.

(14) I used the easyrdf service to convert my Turtle vocabulary into RDF/XML  
(<http://www.easyrdf.org/converter>)

(15) I generated a graph visualisation of my vocabulary using the W3 validator service  
(<https://www.w3.org/RDF/Validator/>) and my vocabulary in RDF/XML format.

