

CT5166 Assignment 2

Zhe Jing Chin 22221970

Task 1

In this task, we design an ontology with the dataset “COVID-19 HPSC County Statistics Historic Data” that consists of a total of 16 columns as shown below:

```
'OBJECTID', 'ORIGID', 'CountyName', 'PopulationCensus16', 'TimeStamp',  
'IGEasting', 'IGNorthing', 'Lat', 'Long', 'UGI', 'ConfirmedCovidCases',  
'PopulationProportionCovidCases', 'ConfirmedCovidDeaths', 'ConfirmedCovidRecovered', 'SHAPE_Length', 'SHAPE_Area'
```

The OBJECTID is just to indicate the number of rows in the dataset, hence it is not necessary to be in the ontology. While the ORIGID is just an identifier for the CountyName for this particular dataset, thus it is not necessary to be in the ontology too. Initially, the ontology consists of 1 main class and its subclass: County and CovidRecord, while the rest are properties. However, to merge the dataset in Task 4, we join them with City, another subclass of County.

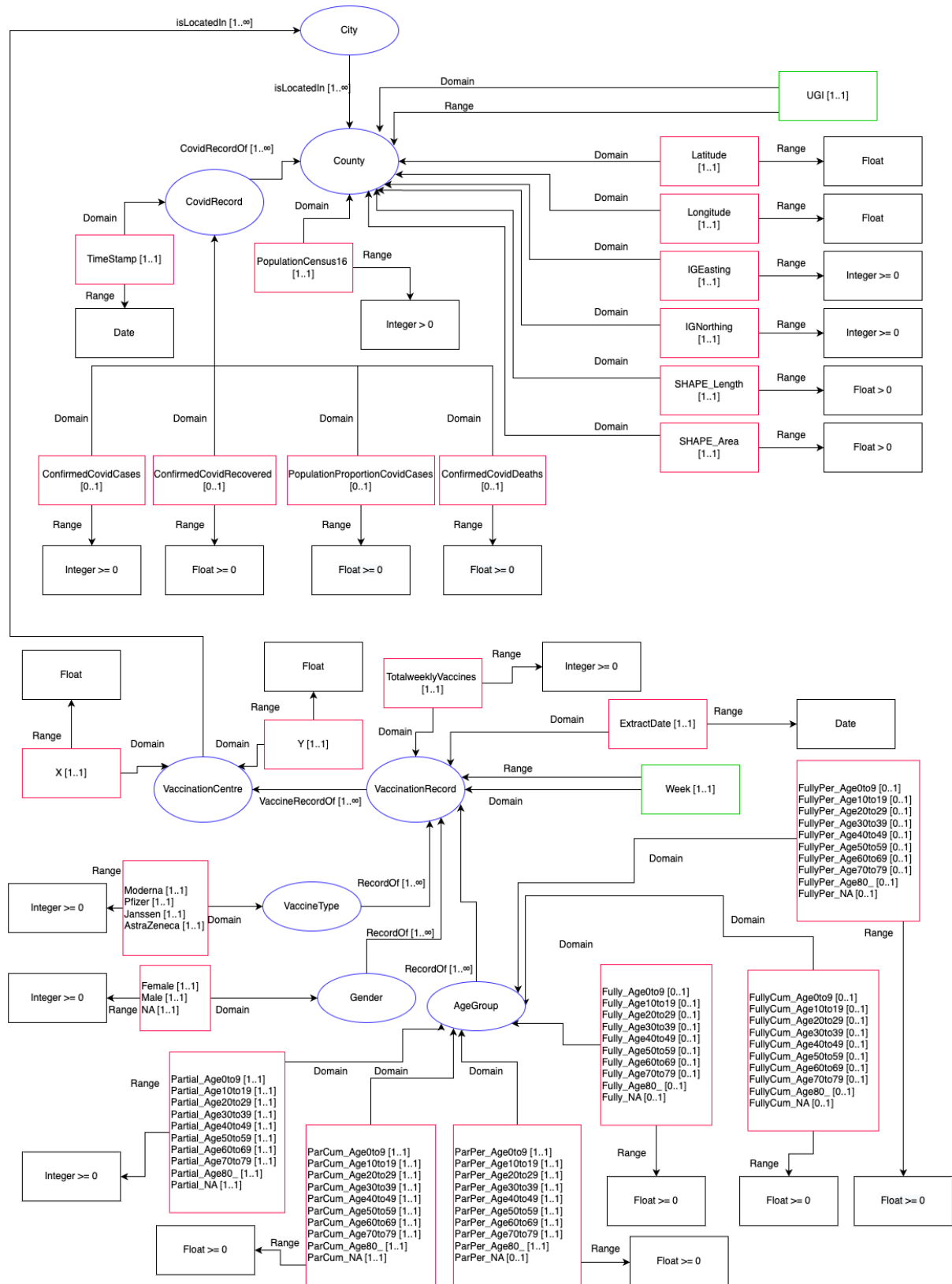


Figure 1. shows the complete ontology for both Task 1 and Task 4. Both the datasets are connected via County - City.

Task 4

In this task, we design an ontology with the dataset “COVID-19 HSE Weekly Vaccination Figures” that consists of a total of 73 columns as shown below:

```
'X', 'Y',  
'ExtractDate', 'Week', 'TotalweeklyVaccines',  
'Male', 'Female', 'NA',  
'Moderna', 'Pfizer', 'Janssen', 'AstraZeneca',  
'Partial_Age0to9', 'Partial_Age10to19', 'Partial_Age20to29',  
'Partial_Age30to39', 'Partial_Age40to49', 'Partial_Age50to59',  
'Partial_Age60to69', 'Partial_Age70to79', 'Partial_Age80_',  
'Partial_NA', 'ParCum_Age0to9', 'ParCum_Age10to19', 'ParCum_Age20to29',  
'ParCum_Age30to39', 'ParCum_Age40to49', 'ParCum_Age50to59',  
'ParCum_Age60to69', 'ParCum_Age70to79', 'ParCum_80_', 'ParCum_NA',  
'ParPer_Age0to9', 'ParPer_Age10to19', 'ParPer_Age20to29',  
'ParPer_Age30to39', 'ParPer_Age40to49', 'ParPer_Age50to59',  
'ParPer_Age60to69', 'ParPer_Age70to79', 'ParPer_80_', 'ParPer_NA',  
'Fully_Age0to9', 'Fully_Age10to19', 'Fully_Age20to29',  
'Fully_Age30to39', 'Fully_Age40to49', 'Fully_Age50to59',  
'Fully_Age60to69', 'Fully_Age70to79', 'Fully_Age80_', 'Fully_NA',  
'FullyCum_Age0to9', 'FullyCum_Age10to19', 'FullyCum_Age20to29',  
'FullyCum_Age30to39', 'FullyCum_Age40to49', 'FullyCum_Age50to59',  
'FullyCum_Age60to69', 'FullyCum_Age70to79', 'FullyCum_80_',  
'FullyCum_NA', 'FullyPer_Age0to9', 'FullyPer_Age10to19',  
'FullyPer_Age20to29', 'FullyPer_Age30to39', 'FullyPer_Age40to49',  
'FullyPer_Age50to59', 'FullyPer_Age60to69', 'FullyPer_Age70to79',  
'FullyPer_80_', 'FullyPer_NA', 'ObjectId'
```

The ObjectId is just to indicate the number of rows in the dataset, hence it should not be in the ontology. This ontology can be linked with the one created in Task 1 as VaccinationCenter can be linked as subclass of City.

To perform with triples, first, get the county of the vaccination centre located in with X, Y points from wikidata. However, it is useless to match the exact geometry points as there are always some non-standard deviation. Hence, the radius from 0.5 is calculated to return the nearest location from the geo-points. Eventually, rdfs:label is used to return the English name of the wikidata location. The county and city name returned are then the key connection to join both the datasets and to add the triples to our graph built in Task 3.

Task 6

Based on this dataset, a mobile application that could help everyone in Ireland to track the COVID-19 cases happening around them or nationally can be generated. Firstly, users are able to use their current location, and check if there's any COVID-19 cases happened in the current county in past 7 days with the knowledge graphs created. If number of cases has passed a certain threshold, government could send out warning for residence in the area. Other than that, users could also check the statistics of daily number of cases, the covid deaths and recovered value. It provides a handy tool to show what's happening around them. Besides daily statistics, data visualisations could be applied to show the overall proportion of covid cases happened, deaths and recovery rate across nation. In addition, users could also check the trend of the COVID-19 pandemic that started from 2020-02-27 until the latest date to see how did it go.

With the geometry points in the knowledge graphs, users could for the nearest vaccination centre around them. Furthermore, they are able to search for a specific county other than their current location to check the covid statistics. APIs and webhooks should be integrated in the application to take in input from users, then backend should use the input to generate SPARQL and query from the knowledge graphs and return the results to user.