Self-reflection report

Patrick Hughes

What I contributed - I attended all scheduled group meetings and contributed ideas to them. I chose to handle the weather datasets. Once we decided on the time scope of our data (yearly interval) and location scope of our data (by Republic of Ireland Counties) I began to scrape Met Eireann's online repository of historical weather data (https://www.met.ie/climate/available-data/historical-data). I manually downloaded the oldest data records per county per month (no yearly option available). I wrote a python script to format each csv correctly, average their monthly values by year, add county fields and combined these csvs to be uplifted to juma. I attended meetings on our ontology design and gave feedback on said design. I uplifted my data to juma and put my ttl file in our groups Google drive. I thought of questions our ontology/linked data could answer and wrote the corresponding SPARQL queries. I helped our group write and test our SPARQL queries at group meetings. I wrote up my designated part of our group technical report, talking about how our queries worked and the strengths and weaknesses associated with them.

Weaknesses: When choosing datasets it was a compromise between datasets that went back a long time and datasets that had a large number of weather categories e.g. sunshine duration, mean maximum temperature. I should have prioritized a broader range of weather categories as I feel this would have lead to more interesting queries.

Caoimhe Mooney

My contribution to the project involved cleaning and uplifting the homelessness dataset, writing some of the SPARQL queries and using WIDOCO for the documentation. The cleaning of the homelessness dataset took quite a bit of work as the information was not provided in an easy to handle format and because we wanted to query over counties. The homelessness data was split into regions rather than counties so I had to do many calculations to translate the values correctly from regions into individual

counties. Once the base data was cleaned it was easy to uplift the data from there. We wrote our SPARQL queries together as a group so we all had input and worked through issues together. Finally, I worked on the WIDOCO documentation, which as mentioned in our main report, brought along many of its own issues. I think as a group we worked very well together due to our communication. We did most of the work while all sitting in a meeting room which meant we could talk out any ideas and work through problems together. These group meetings also helped us to counteract any weaknesses that we may have had individually. I think a weakness I had personally going into this project was seeing the bigger picture of what we were trying to achieve. We spent many days discussing our ontology model and how each of our datasets should be able to link. I feel that at this initial part of the project I struggled to find those connections and understand the model we were trying to achieve. Once we had decided on the direction we wanted to go I felt far more comfortable with the overall picture and was able to achieve more in the project while writing the SPARQL queries and doing the documentation.

Arne Philipeit

My contributions include processing the csv files on housing prices with a simple Go program to prepare the data to be uplifted with Juma, including extrapolation of data for the remainder of 2019. I prepared a ttl file for housing prices, attended all but one group meetings, and strongly contributed to our discussion of the ontology we designed, especially how to find a sensible transitive property. I developed the user interface in Java and used the Apache Jena library for SPARQL query execution. This made it faster for the team to develop working queries. Based on common use cases and complaints during this process, I improved the user interface - limiting the output size and displaying helpful syntax errors from Jena. I helped with the development of SPARQL queries and included all queries in the final user interface code. I wrote parts of the technical report on transitive properties and described the user interface.
The initial approach we took for the housing data ontology was suboptimal, making two subclasses of Housing: Executions and Filings. I should have realized earlier that it would make the SPARQL queries and the uplifting process a lot easier to make

Execution or Filing a property of a shared Housing statistic class instead. We also did not end up using all the data provided by the housing data set in our SPARQL, only mean sales prices. Java's Swing library is deprecated and does not implement cut, copy, paste, undo, redo etc very well. I could have worked with JavaFX or another modern user interface library instead but chose to use Swing due to more familiarity and time constraints.

Alexandra Silva

My first contribution to the project involved cleaning the salary dataset. For salaries I was only interested in one value in the dataset which was Disposable Income per Person. I decided to remove the rest of the values as I felt we already had more than enough information from all of the datasets combined. The original dataset also had information on different regions of Ireland such as the south or the west as well as all of the counties in Ireland, I decided to remove these as the goal was to have every dataset relating by county. Another contribution of mine was helping with the ontology model. I attended all meetings to discuss the ontology and put forward my ideas during these meetings. I also downloaded protege to put the ontology together, this allowed me to learn how to use protege as I had never used it before this project. I believe that working on the ontology in protege was a strength of mine, as I was able to learn how to use protege very quickly and now would feel totally comfortable creating an ontology with protege in the future. I also uplifted my dataset with Juma. Another contribution of mine was writing some of the SPARQL queries. Finally the last of my contribution was writing parts of the technical report. I believe a strength in our project was how well we communicated and how on top of the work we all were. In my opinion we achieved this because we attended regular group meetings and in these meetings discussed and solved any challenges we were facing together and made sure that we were all on track with our work. One weakness I believe we had was in how we chose datasets with different timelines (salaries ranges from 2000 to 2016 while homelessness ranges from 2014 to 2019). This means that when querying the ontology you will only get results for both salaries and homelessness for 2014-2016, any years outside of this will return no results.

Caleb Teo

My contribution to the project involved data cleaning of the crime dataset, uplifting of the crime dataset and contributions to the designed ontology, SPARQL queries and Application Interface code. I attended all group meetings and worked with everyone on the design of the ontology. I helped discuss the ontology and how our datasets would be linked together. I was given the crime dataset to work on and I had to clean the dataset in preparation for JUMA. The crime dataset was split into quarters and I summed each quarter to get the total per year and county. I wrote a script that transformed the dataset into a table ready for JUMA. I then helped create the SPARQL queries from the questions we had made for the application. I also contributed to the application code by helping write the queries into application and formatting the questions on the interface. I wrote my contributions to the report for the overall project. The weakness of my contribution was my work with the crime dataset, as I was not careful in my uplifting which I then had to edit the script to transform the data and re-uplift the dataset. I have learnt that by being more careful in my approach would have mitigated this mistake which had an impact on the project. I felt my strength was working with the crime dataset as it was a very large dataset. I was able to uplift the data and complete it to a high quality. I think I was able to work with the team and we were able to be honest with each other on different design and approaches. This openness helped get all the different idea out and allow us to choose the best method.