Laidlaw 2020 Research

Introduction

My first introduction to research occurred at the start of my second year of college. Research has always been something I have been interested in trying and so, just before commencing my second year of Computer Science, I sent off an email to every professor I could possibly think of asking them if they knew of any on-going research projects that I could possibly join (voluntarily of course!). I was not very hopeful to secure anything as I was acutely aware that I had only just completed one year of university at that point. As it happens, I was incredibly lucky to have my current Laidlaw supervisor, Declan O'Sullivan, respond and, after a short interview with him, I was able to join his research team. Not long after, I received an email from my college advertising the Laidlaw Scholarship. I knew before even finishing the email that the Laidlaw program would be perfect for me. Up until that point, I had no prior leadership training in spite of it being something I utilised quite frequently in my extra-curricular activities (Class representative – leader to over 120 students, Computer Science Convenor – leader to over 700 students, S2S mentor – mentor of over 20 new international students). Thankfully, after much preparation both individually and alongside my supervisor, I managed to secure a place on the program! These two years have given me the most incredible opportunities to interact with some of the most interesting and intelligent people in the university, visit new places, attend countless workshops, and experience excellent mentorship.

Laidlaw 2020 Leadership

Leadership

Initial concept of leadership

Prior to Laidlaw, I did not have a very clear idea of what it meant to be a leader. If asked, I would have probably said something along the lines of "oh, the job a general in the military does, or perhaps a politician?". To me a leader would have been someone out alone, in front of everyone else, the ultimate decision maker. However, through my Laidlaw journey I have come to the realisation that for the most part, that idea is false.

New idea of leadership

My concept of leadership is still a work in progress. The various workshops and trips that I participated in through Laidlaw sparked an interest in me into the world of leadership. Following on from Laidlaw, during the summer, I helped lead a book club in which the book of choice was "Dare to Lead" by Bréne Brown. In her book, Brown delved deep into the psychology behind leadership and, in particular, the correlation between being vulnerable and a leader.

That being said, I do now have a much clearer idea of what it means to truly be leader. I believe that in order to be a great leader, one must be vulnerable. It seems contradictory to what my initial idea of a leader had been, i.e. strong, impenetrable. Now, I would argue the contrary, to be vulnerable is to be strong. Being vulnerable does not mean that you are opening yourself up to everyone, giving them glimpses of your deepest darker desires. Being vulnerable means that you are willing to try, to try and fail and try again.

Leadership Experiences

Puzzle Workshop

One of the more memorable Laidlaw workshops for me was one which involved jigsaw puzzles. As someone who has spent most of my life enjoying putting things together, I am naturally drawn to puzzles. However, even though I did love the act of putting the jigsaw puzzle together, the puzzle itself did have a purpose. The activity was as follows – my Laidlaw cohort was split up into five teams and each team was given a bag containing jigsaw puzzle pieces. We had to work together within our teams to put the jigsaw together. The catch was that we were not given the boxes the jigsaw puzzles came in and so we did not know what the pieces were supposed to make up.

Naturally, putting a jigsaw puzzle together when you do not know what the image is supposed to be is extremely challenging. As a team, we had to come up with a strategy to tackle the problem and were given a set amount of time to try put it together. After an iteration we had to analyse how we had performed and if the strategy needed to be refined. When finally we were given the box to look at, our progress improved. The analogy here was that without vision, while a team can still work together, it will not perform to its true capability.

Leadership Weekend

Engaging with people for a long length of time is something I have always struggled with. Therefore, the concept of going away for the weekend with a large group of people initially terrified me. Even though I knew that the people I would be going away with were some of the most incredible people I have ever met in my life, it still took me some time to convince myself to go. Wow, am I so glad that I went! The leadership workshops during the day were challenging but rewarding. In particular, one workshop that involved a leadership coach who brought up confronting and thought provoking topics. I found it interesting how he was able to divide and unite us and how even after the workshop was over we found ourselves continuing to build on conversations that had been started during his workshop. Through such conversations I continued to learn how to communicate with people from a vast array of different backgrounds on complicated ideas.

My leadership style

Through Laidlaw I have become intensely aware of how I conduct myself in a group setting. I have learned that I am a good person for starting the conversation off but once everyone is then talking, I am happy to take a step back and observe the room. I have found that being a leader does not necessarily mean that you are the loudest in the room. By staying quiet for a while and instead listening, you are able to learn a lot more about the people you will be working with and overall be able to create a better dynamic. I look forward to continuing to develop my leadership style based upon the understanding I have gained from Laidlaw.

Final outcome

I think that the final outcome for me could be summed up in one word; *confidence*. While I would not necessarily say that prior to Laidlaw I was unconfident, I was perhaps confident in a different sense; outwardly confident but lacking inner belief. Through the many workshops and challenges

Laidlaw has provided, I have been able to prove to myself that no matter what the problem or situation, I have been able to find a way to successfully navigate through it.

Laidlaw 2020 Research

Research Space

The research space my project was centred in was *Linked Data*. Before getting into the exact details of what my research aimed to achieve, I will first go over some of the key terminology:

Linked Data – As the name suggests, Linked Data refers to data that is linked in some way, i.e. information that has a defined relationship to other information. For example, *I live in Dublin*. The two pieces of data, me (Claire) and Dublin (capital city of Ireland) are related via "lives": Claire –> lives –> Dublin. More information can be added in a similar way. For instance, *Claire owns a pet cat*. This can be added into the graph like so: cat <- owns <- Claire -> lives -> Dublin. The statements Claire -> owns -> cat and Claire -> lives -> Dublin are linked by me, Claire. These statements are called triples as they are made up of three distinct parts: *Claire* (Subject) *Lives* (Predicate) *Dublin* (Object).

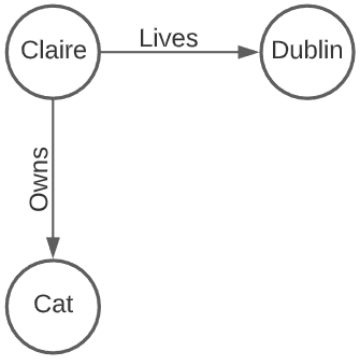

Figure 1: Graph view of the triples.

Resource Description Framework (RDF) – Having a number of statements about a single subject is called a resource. The data model that is used to model such data is called Resource Description Framework or RDF. RDF, as shown in figure 1, is a directed, labelled graph data format for representing information on the web[1]. This is the type of data format that I worked with during my research.

Why use Linked Data?

The world around us is not made up of static individual collections of information. Everything around us is interconnected. The specific area of Linked Data that my research was focused on was clinical Linked Data. In particular, clinical Linked Data that provided information about patients with a rare kidney disease. Unlike other instances of a person becoming ill, when a person with an autoimmune disease gets a flare up, it may not be immediately evident when looking at their patient records as to why it happened. This is because the patient record does not contain everything there is to know about that person. A person is made up of multiple collections of information. In the case of the Linked Dataset that I worked with, patients' records, weather records and pollution records were all

[1] https://www.w3.org/TR/rdf-sparql-query/

linked together to provide a knowledge graph that contained general information about a person's life. The links between the different sectors of information were location, time and a specific id that the patient had been assigned. The idea was that through exploration of the knowledge graph, patterns in the data may become evident that could help clinicians predict flare ups and hopefully understand the reason behind them occurring in the first place.

While health-related information is certainly very suited to being stored as Linked Data, the potential of Linked Data does not end there. DBPedia[2] is a well-established example of Linked Data in action. It provides Wikipedia style data as an open knowledge graph allowing the user to pose queries over the web such as retrieving *"all cities with low criminality, warm weather and open jobs"*[3].

The problem

As is evident from the previous section, Linked Data has enormous potential. One of the primary factors hindering the use of Linked Data is the gap incurred between the Linked Dataset and the domain expert, i.e. the person who understands what exactly the data means. Without extensive knowledge of Linked Data, it is difficult use Linked Data in its raw form. This is where my project comes in. My research aims to bridge the gap between the RDF dataset and the domain experts by providing a web application that allows domain expert researchers to explore and conduct preliminary analysis on Linked Data.

Initial idea

My initial idea centred around providing the user with a visual representation of what was in the dataset. This was achieved using classic tables and charts. However, before I could get to that stage, there were two key things I had to learn. Firstly, I had to learn how to query the data and secondly, I had to learn how to build a web application. Both were essential in bringing my initial idea to life.

1. *Querying the data.*

In order for the dataset to provide the answers I was looking for, I needed to learn how to format the questions in a way that the dataset would understand. The specific language that I used to do this is called SPARQL. SPARQL (SPARQL Protocol And RDF Query Language) is a query language used to query over RDF which relies on looking for patterns within the data. An example SPARQL query is shown in figure 2. Applied to the RDF data in figure 3 this query returns two records, those shown in figure 4. Both Claire and Mary's records are returned as their records are the ones that match the pattern: aged 22 and living in Dublin.

[2] https://wiki.dbpedia.org/
[3] https://wiki.dbpedia.org/about

Query:

```
PREFIX studentNo: <http://trinity.ie/studentNumber/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

SELECT ?studentNumber ?name
WHERE {
    ?studentNumber student:hasAge "22" .
    ?studentNumber student:livesIn "Dublin" .
    ?studentNumber student:hasName ?name .
} ORDER BY ASC(STR(?name))
```

Figure 2: sample query for the student number and name of students aged 22 living in Dublin.

Dataset:

```
@prefix student: <http://trinity.ie/student/> .
@prefix studentNo: <http://trinity.ie/studentNumber/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

studentNo:17332568 a student:information ;
    student:hasName "Claire McNamara" ;
    student:hasAge "22" ;
    student:livesIn "Dublin" .

studentNo:17332569 a student:information ;
    student:hasName "Jane Doe" ;
    student:hasAge "23" ;
    student:livesIn "Dublin" .

studentNo:17332570 a student:information ;
    student:hasName "John Doe" ;
    student:hasAge "22" ;
    student:livesIn "Cork" .

studentNo:17332571 a student:information ;
    student:hasName "Mary Smith" ;
    student:hasAge "22" ;
    student:livesIn "Dublin" .
```

Figure 3: sample RDF dataset

studentNo:17332568 “Claire McNamara”

studentNo:17332571 “Mary Smith”

Figure 4: results of the query in figure 2 applied to the data shown in figure 3

2. *Building a web application.*

Prior to starting my first research period with Laidlaw, web development was not an area I had explored. I therefore had to spend the first week or so learning the various components that go into building a web application. This involved learning four new coding languages (HTML, CSS, Javascript

and Python), a templating language (Jinja2) and a web framework (Flask). I used Python for the backend or the "brains" of the application; the part that would query the Linked Dataset, perform calculations on the retrieved results, and package up the information nicely to be displayed to the user. I found that I particularly enjoyed learning and coding in Python and it has since proved to be an extremely useful language to know both as an option to complete coding coursework in and also in extracurricular activities.

End of the first research period

By the end of the first summer, I had built a working web application prototype for the clinical dataset that I was using. This prototype gave the clinical researcher the ability to search by individual patient ID or get a list of patients over a specific timeline. (Figure 5)

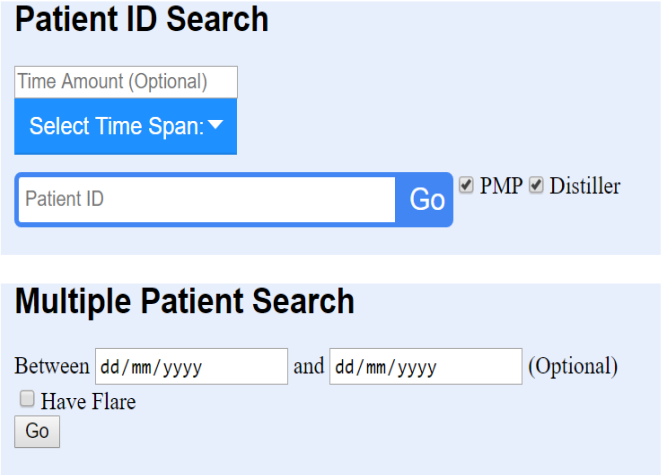

Figure 5: Individual and multiple patient search.

The list view of patients gives the user the ability to filter by age and gender, and sort by gender and date of birth.

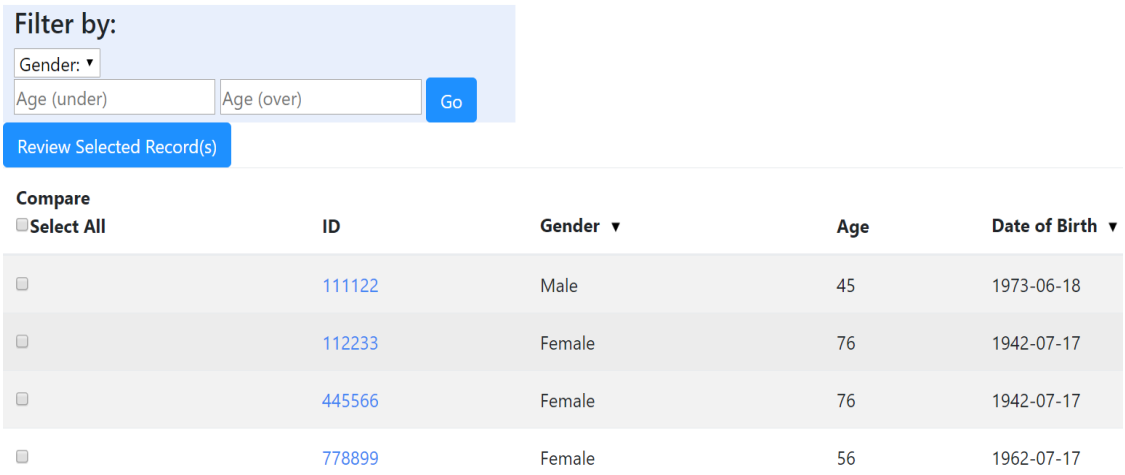

Figure 6: List of patients over a specific timeline.

The application also provides the user with the ability to select a number of medical fields in order to visualise how the fields change over time.

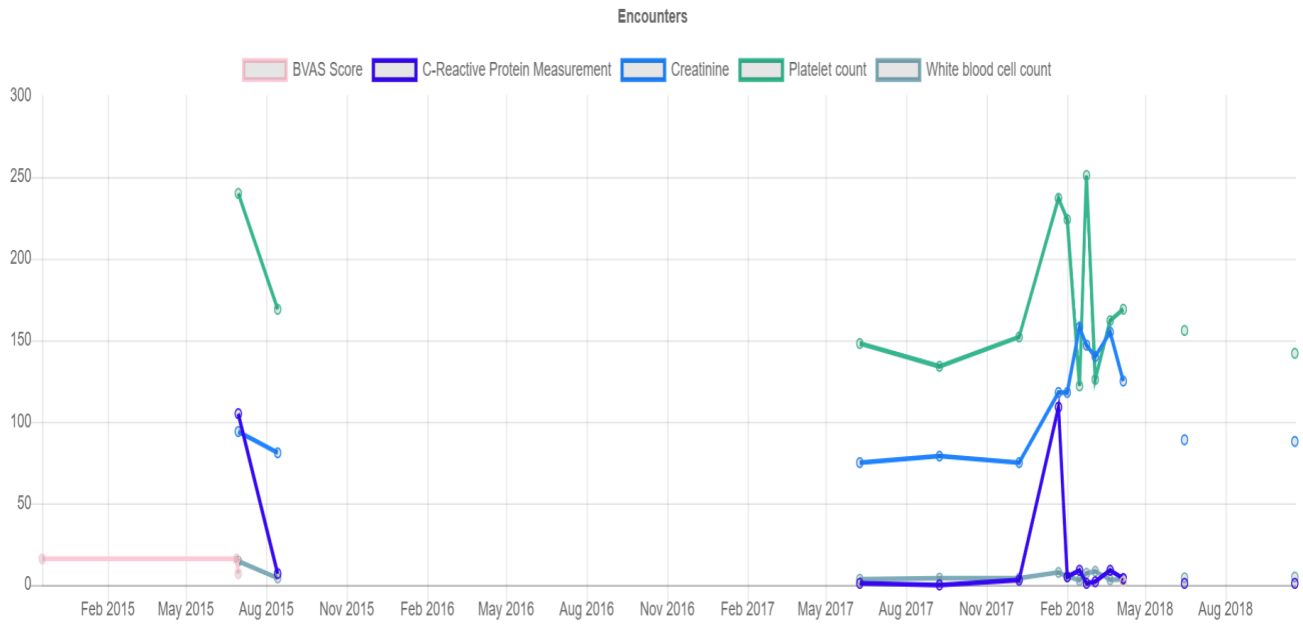

Figure 7: Graph view of a specific patient's medical data over time.

Problems with the original idea

Problem 1.

While the first prototype is functional and does provide some insight into the data, the application is not exploratory in nature. It does not provide a user who has never encountered the dataset before the opportunity to see what is in the dataset. Instead, it merely provides my idea of what a user would be interested in viewing when interacting with the dataset.

Problem 2.

The application built is very specific to the dataset I had at the time of building it. Any change to that dataset would mean that multiple changes would have to be made to the code. In essence, the prototype is very difficult to generalize.

Revised idea

Towards the end of my second research period I came up with a novel way to tie in exploration with the more traditional visualization provided by my original prototype. The idea is to use node by node exploration of fields within the dataset to build up a query. Then, once the query has been created, to use that query to provide back a list of patients that match the criteria specified - thereby connecting the work done last year with the new exploration layer.

The difference between my approach and the approach taken by Linked Data visualization tools in the past is that my idea is not to explore the data itself, but more so to explore the structure of the data instead.

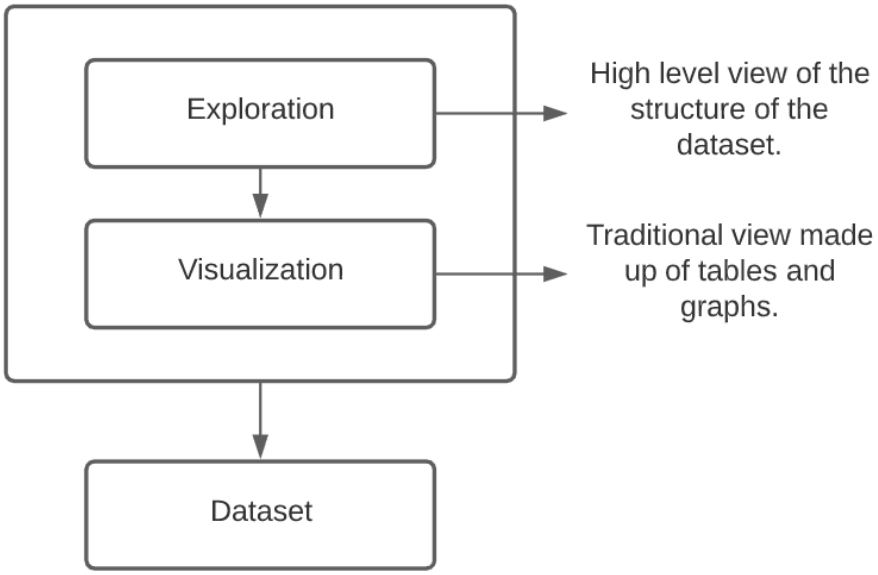

Figure 8: Layering Digram.

On the left hand side of the application there is a panel which provides a list of the main categories present in the patient record. When the user clicks on a category, a node is created on the screen for that category. Next, when the user clicks on the newly generated node, new nodes representing the sub-categories of that category are generated, all connecting to the central category node. If applicable, when the user hovers their mouse over a subcategory node, the potential values of that node are displayed.

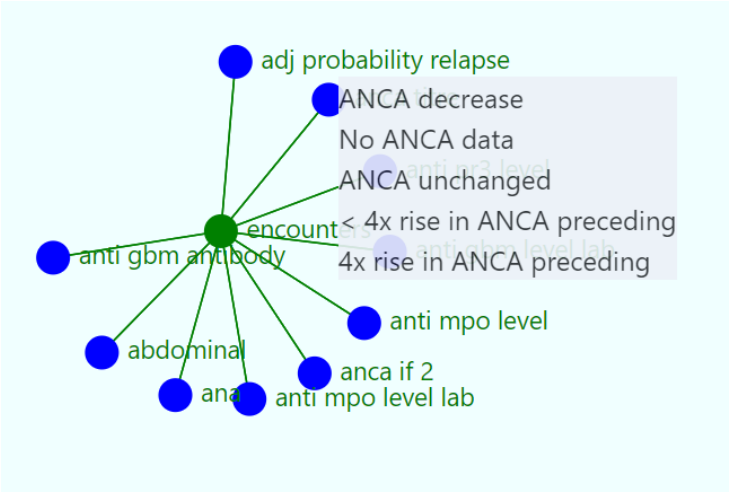

Figure 9: Central node with sub-categories.

When one of the subcategory nodes is double clicked, that subcategory is appended onto the query. If the subcategory has potential values, the user can use a drop down menu to select the value that they are looking for.

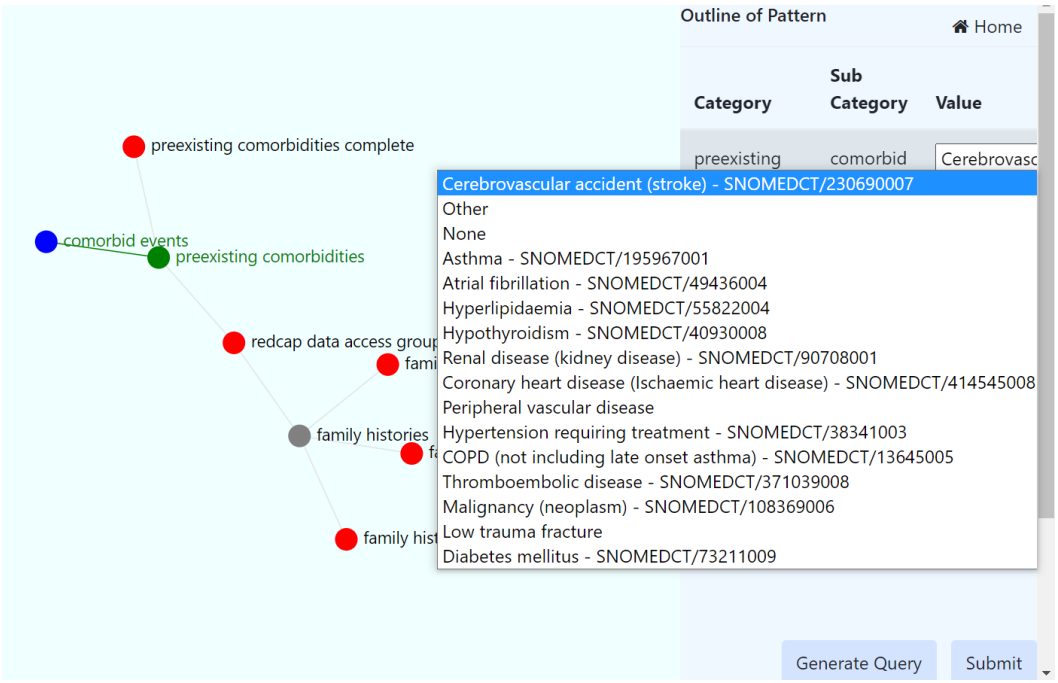

Figure 10: Demonstration of adding a subcategory to the query.

If two categories are connected by a subcategory, the subcategory is displayed with two links coming out of it – one to each category.

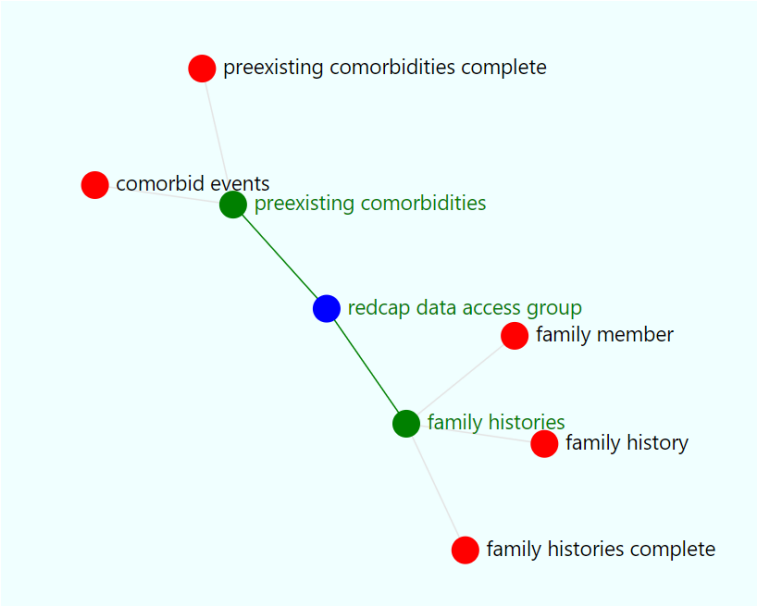

Figure 11: Linking two categories together by a shared sub category.

Results

From research period 1:

- Working version of visualization.

From research period 2:

- Working version of exploration.
- Proof of concept of layering exploration tool over visualization tool.

Future work

1. Generalise the backend

As has been mentioned previously, the final tool that I have built has thus far been designed specifically for the clinical dataset that I had access to. This means that while the tool itself has proven to be a very successful means of exploring a Linked Dataset, if a different dataset was provided to the tool, the tool would simply fail to work correctly.

An analogy would be a pair of glasses, *x*, with lenses prescribed for a specific person, *a*. In this example, the person is the clinical Linked Dataset that I was working with and the pair of glasses is my exploratory application. As *x* was built for *a*, *x* works as intended for *a* allowing them to view the world (explore). If another person, *b*, comes along and tries to use *x*, it is extremely likely that either 1. The effect on person *b* is that their vision will be dramatically decreased or 2. *b*'s vision will improve but only mildly and not enough to be of any real use. The problem is that, even though *a* and *b* could both benefit from utilizing glasses such as *x*, *x* was made specifically for *a* and not *b* and therefore *b* will not truly benefit from the full potential of *x*. Going forward, my research aims to create a tool that will provide benefit to both *a* and *b*, i.e. creating a generic tool that can be applied to multiple different types of Linked Datasets. Such datasets could include historical Linked Data, weather Linked Data, financial Linked Data and so on.

Generalizing the application will require further investigation in order to work out the best way to get and store the structure of the linked dataset (schema). It could be done either by storing schema as the dataset is being created, or, by increasing the functionality of the application to automatically generate the schema when given a linked dataset. The creation of the schema itself will also need to be researched more in order to figure out the best way to represent it in storage.

2. Update the UI/UX.

While the application I designed is user-facing, my primary interest was in the backend work described above. I therefore have not thoroughly tested the current user interface (UI) with the intended user, a clinical researcher. In order to conduct research on users I would need organize sample groups, come up with UI concepts and iterate through them, ultimately attempting to work out the best possible layout, colour and work flow for the researcher. As there were a limited number of clinical researchers who were a part of the research group I got my Linked Dataset from, and those individuals were extremely busy, it would have not been possible to direct my research in this area this time around anyway.

That being said, as the tool becomes more generalized, it will become more and more necessary to conduct focused user interface and user experience (UX) research that concentrates specifically on

the user (a domain expert) and their interaction with the application. There is no point ending up with a powerful, generic application that is unusable by its target audience.

I would suggest going forward that further work is put into designing such studies based upon the exploratory proof of concept that has been the result of my Laidlaw research.