Thoughts for using Solid-based working sets

Graham Klyne, *Mapping Manuscript Migrations* Project, Summer 2020

We conducted an experiment to explore MMM research questions using ResearchSpace, and some ideas were raised by our exploratory work. The ideas here are only partially tested, and these notes are intended as possible input to future work in this area. More technical details about the exploratory work can be found in the [GitHub project README file](#).

See also: [Reflections on addressing research questions](#)

In our discussions about navigating MMM data (which could apply to many similar datasets), we discussed the idea of research questions that are addressed through a composition of data-refining queries. A question that arises is: how can we make such composition easy for non-technical researchers to use?

The work so far shows how ResearchSpace allows for application-specific queries to be embedded in web pages, and connected by web hyperlinks. The demo queries are very specific to certain questions that one might ask of the data. But an analysis of the proposed research questions suggests that there may be a relatively small number of query elements that are combined in different ways when addressing different questions: many of these elements could be written once, and re-used, if there were a way to compose the results.

One way to approach this would be to represent a research exploration as a connected graph of intermediate results, where graph connections are represented by hyperlinks. The starting point for such a graph would be the entire set of data and a landing page. As new queries are applied, the current "working set" of data would be updated, and the derivation of any intermediate stage could be represented as a provenance graph from the starting point.

One thing required to implement this would be a way to capture and record the current working set at any point. For this, I propose to use a Solid Pod [1][2].
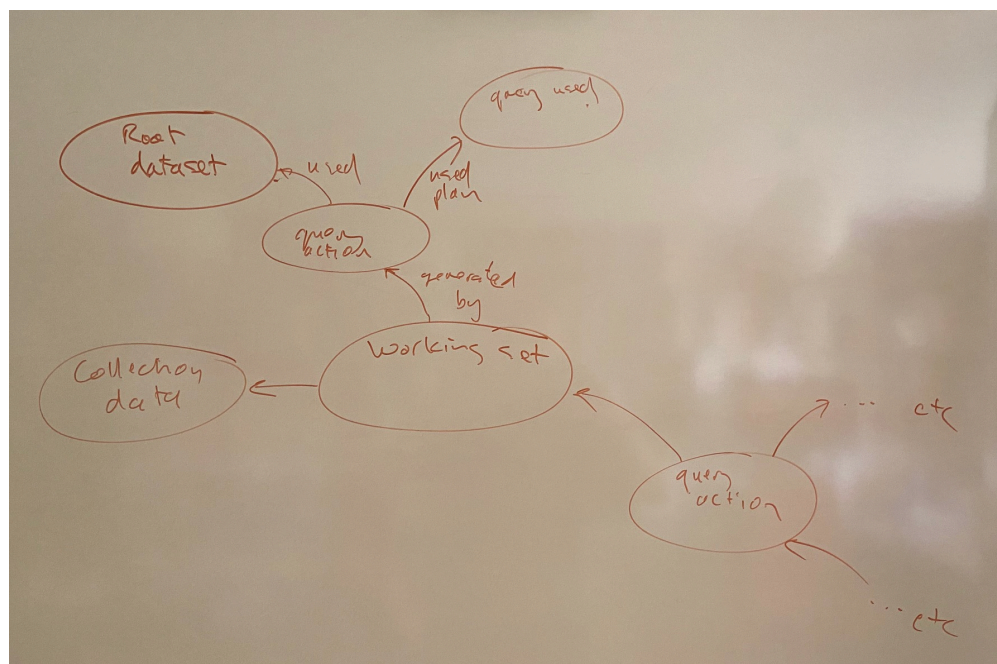
[1] https://solidproject.org/

[2] https://solidproject.org/use-solid/

In the pod would be a container, identified by a URI, that might (for example) correspond to a research investigation. Within this container would be worksets consisting primarily of a collection of entity data (e.g. for people, or works, or manuscripts, etc), a reference to another working set from which it was derived, and a query mechanism by which it was derived. The query used would correspond to a hyperlink on a displayed web page that forms part of the research exploration.

This calls for a set of conventions for composing queries and working sets. I imagine each query corresponds to a webpage that displays the result of that query. The query would be part of the web page, and used to dynamically construct the display from a supplied working set and additional SPARQL endpoint(s) to which the query is directed. The web page would be parameterized with the URI of a working set from which it is derived; e.g.

```
http://query.example.org/works-in-manuscript/?ws=http://pod.example.org/workset/1234
```

Thus, a single working set URI would allow access to a complete history of how that working set was created. And an exploration could be repeated against updated background data by repeating the derivation chain (NOTE: would need to recognize and substitute intermediate working set URIs when replaying a derivation).

Q: how to represent a working set.
A: (At least) Two options:

1. a set of URIs, so a set of resources - this was my initial thought
2. an RDF graph - I'm leaning toward this, because it means a single SPARQL defines the composition. But it does mean that the pages and the workset data might be more tightly coupled. Maybe use Solid/LDP vocabulary for the simple case of representing (1), leaving the way clear for more subtle com,position options?

Q: derivation from multiple working sets?
A: possibly, but this would be harder to implement using a simple hyperlink

Q: SPARQL endpoint(s) used defined as fixed within each composing web page, as part of a global environment, or as parameters to the composing web pages?
A: ... start simple (fixed in page?) and see what requirements emerge?

Q: Would OWL-QL be a reasonable way to abstractly represent a working set derivation, which can be translated into SPARQL queries? Maybe worth talking to Ian Horrocks' team in CS?
A: ...

Q: what about derivation steps using manual inspection. That rather breaks the repeatability story.
A: ...

Q: What about adding new query elements?
A: This would probably still need technical expertise to create a new composable query element page, but the amount of work required should be limited, and the resulting page available for any subsequent investigations.