

Exploring research questions through browsing: ResearchSpace for MMM

Graham Klyne (with David Lewis), December 2020

Abstract

Background

MMM and the Research Questions

ResearchSpace

ResearchSpace Prototype

Comments on the implemented interface

Survey of research questions

Worksets and composition

This document summarizes an approach to exploring MMM research questions using ResearchSpace, and reflects the end point of some exploratory work. More technical details about the actual work can be found in the [GitHub project README file](#).

Abstract

This document describes an experiment using ResearchSpace to realise research questions devised for the Mapping Manuscript Migrations project (MMM) over project data. I also survey the questions to assess how well this approach would work, and briefly consider using *worksets*, named groupings of entities. The initial findings are encouraging, but further work will be needed to validate them in a real research context.

A research question may not initially present itself as a single, clear question. It is often the expression of the end-point of a process – a line of enquiry. What we propose is approximating such a line of enquiry using a hyperlinked series of dynamically-generated pages, allowing a user to explore and develop various research questions without having to frame a formal query at any point. Each query element would be handled by a single application page (like the "who collects" page described below), in such a way that different queries can be composed by following links (combined with filter and sorting functions built into results pages, as in our ResearchSpace prototype).

Background

MMM and the Research Questions

The Mapping Manuscript Migrations (MMM) project involved exporting data from three medieval manuscript catalogues (Bibale, the Bodleian catalogue, and the Schoenberg Database of Manuscripts) as Linked Data and aligning and unifying them in a single triplestore, using a model drawing heavily from CIDOC-CRM ontologies. The MMM project was defined and tracked in part based on predefined [research questions](#).

ResearchSpace

[ResearchSpace](#) is a web application created by Dominic Oldman and his team at the British Library, and [metaphacts](#). Given a SPARQL endpoint, it allows developers to create interfaces for exploring data.

ResearchSpace is template-based, and can dynamically select templates to use based on the type of data object being investigated (a manuscript would have a different template to a person, for example). Although it can work with any linked data, template presets exist for a range of CIDOC-CRM object types.

Creation and editing of new ResearchSpace pages is through a web interface integrated into the main application, facilitating rapid prototyping of the kind described here.

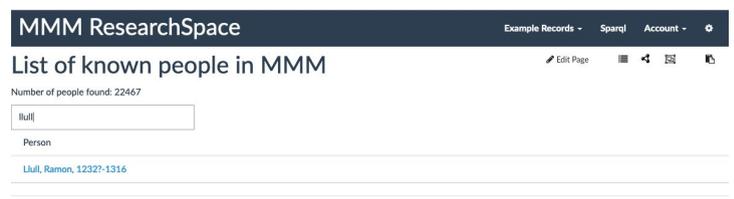
ResearchSpace Prototype

Taking the MMM research question labelled [B3](#) (see *Survey of research questions* below): “Who collects manuscripts with texts by Ramon Llull?”, we can construct a series of pages and navigation steps using the prototype:

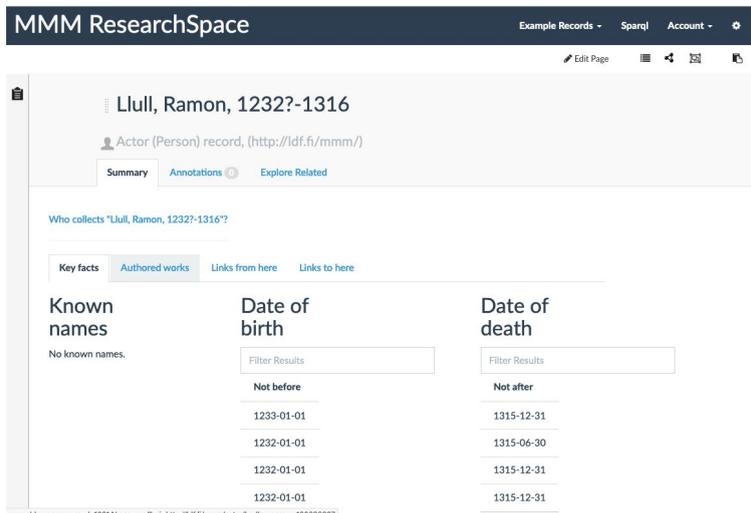
1. Go to the landing page



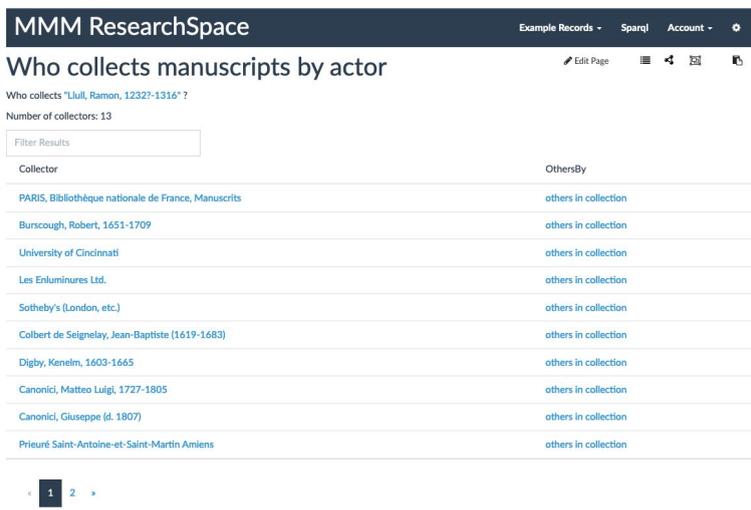
2. Click ‘List known people’ and filter for the author’s name



3. Click to view more details about Llull's record



4. Click to see "Who collects" manuscripts with his works



This gives an answer to the question, but frames it as part of an investigation. The investigation can be continued, for example by clicking on 'others in collection', which shows, for that collector, which manuscripts, and which works in those manuscripts, were owned by that collector.

There is also a [video](#) that illustrates use of the interface covered by the screenshots above.

Comments on the implemented interface

The "who collects" query is instigated by a direct link in the Actor Summary page. This is likely to become cumbersome as more query directions are added. I propose a new "Explore author" page to which the embedded logic that generates the parameterized link is moved, and which acts as a hub for explorations of authors. Similar exploration pages would be possible for works, manuscripts, places, etc. I'm not sure how to handle composites - I propose to ignore that concern until a real case arises from examination of the research questions.

Currently, the exploration page is parameterized by a single Actor URI: maybe this should be extended to allow multiple actors (queries effectively becoming list monads?). I'm not sure

how do-able this is (i.e. can we collect all selected matches and pass them forward?). For now, stick with the simpler single-instance case. (See also the section below "Worksets and composition" for some further thoughts about this.)

Survey of research questions

(Derived from

https://docs.google.com/spreadsheets/d/1Tdt3dNGkq5aEpC-IXCpxXZeEptnOhl60rVeT_m_mjw4)

For each question, I have tried to identify query elements that would need to be composed to achieve the desired results:

A1:

- (a) manuscripts by (production event) date
- (b) manuscripts by (production event) place (Europe)

A2:

- (a) manuscripts by surviving?
- (b) manuscripts by language content
- (c) manuscripts by place (Castile)
- (d) manuscripts by "produced for abbey or convent"
- (e) manuscripts by collector
- (f) collectors by nationality
- (g) collectors by is-private (or is-person?)
- (h) manuscripts by current owner (collector?)
- (i) collector by is-institution

A3:

- (a) collectors by nationality
- (b) manuscripts by acquisition date by collector (composite key)
- (c) current location by manuscript

B1:

- (a) manuscripts by author of work
- (b) manuscripts by purchase (acquisition?) date

B2: (there is no B2)

B3:

- (a) manuscript by author of work
- (b) collector by manuscript

B4:

- (a) manuscripts by author of work
- (This case challenges the composition capabilities: given two authors, find manuscripts that contain works by both authors.)

B5:

- (a) manuscript of works by "a medieval author"

- (b) work by manuscript
- (This case additionally requires finding the maximum count of any result value.)

C1: (none)

C2:

- (a) manuscripts by collector (or "associated with?")
- (b) manuscripts by "thirteenth-century bible"
- (c) manuscripts by "historiated initials"

F1:

- (a) manuscripts by collection (or by owner?)
- (b) manuscripts by "specific physical feature" (e.g. enluminés)

F3:

- (a) actors (and their roles) by (association with) collection

F4:

- (a) manuscripts by collection (or by owner?)
- (b) manuscripts by work
- (c) works by subject category

F5:

- (a) manuscripts by collection (or by owner?)
- (b) places by (production of) manuscript
- (c) dates by (production of) manuscript

F7:

- (a) manuscripts by collection (or by owner?)
- (b) events (provenance and production) by manuscript

F8:

- (a) manuscripts by (no) current location; last observed/known location as a kind of proxy?

F9: ?

G1:

- (a) manuscripts by work
- (b) manuscripts by "specific physical feature" (e.g. enluminés)

G2:

- (a) all works
- (b) manuscripts by work
- (plus some composition filtering to exclude works with >1 manuscript)

G4:

- (a) People by life-dates
- (b) Works by person
- (c) Works by language
- (d) People by association with work
- (e) Life-dates by person

G5:

- (a) manuscripts by work

- (b) production date+place by manuscript
- ??:
- (a) manuscripts by place of production event
- ??:
- (a) manuscripts by place of production event and date range (composite key)
- ??:
- (a) manuscripts by (previous/sometime) owner
 - (b) manuscripts by current location/owner
- ??:
- (a) manuscripts by work
 - (b) filter manuscripts by text in title
 - (c) number of folios by manuscript
 - (plus some counting/computation)
- ??:
- (a) manuscripts by seller
 - (b) owners (actors) by manuscript
 - (c) location by actor
 - (d) "kind of manuscript and earlier history" by manuscript
- ??:
- (a) works by language
 - (b) works by date
 - (c) manuscript by production data
 - (d) works by manuscript
- ??:
- (a) manuscripts by (sometime) owner
 - (b) purchase/sale events by manuscript and owner (composite key)
 - (c) advertisements placed by manuscript and owner (composite key)
 - (d) date by advertisement placed

The approach here is not to reduce the questions to a single SPARQL expression. Rather, it is to allow successive refinements/explorations that build up a set of results using multiple, simpler queries. For example, starting with an author of interest, a hyperlink presents collections containing works by that author, and from there further links to enumerate those works. Some steps may involve manual inspection and selection that draws on the investigating researcher's domain knowledge or interests (such as selecting a subset from a list of results presented, for example manuscripts owned by a subject, explored by places of production and distribution of dates [research question F4]).

The next technical step on this path might be to list the sub-queries indicated above and remove duplicates. The result would be a list of query templates and pages to be coded in order to build a minimal version of the proposed hypertext navigation system on ResearchSpace. Such analysis could help to size any further development proposal along these lines. For example, several of the research questions involve locating manuscripts by

author of work - these could be handled by a shared, pre-defined query element that is available (as a hyperlink) for exploring data connected with a person, or when viewing some other work by the same author.

Future work: it might be that OWL-QL would provide a way to describe the compositions, in a way that can be mapped to SPARQL queries. Could that reduction be done using JS in the browser?

Worksets and composition

A workset is a named collection of entities that can be saved, pointed to, and used as the basis for further investigations. If users could store the results of queries – or arbitrary lists of entities they are interested in – as worksets, then these can be fed into the querying process, provided the user interface can accommodate them. A workset used in this way could capture something of the context of a line of enquiry, which can be carried forward into subsequent queries. For example, when presented with a list of works by an author, to focus on those about a particular subject matter for further investigation. Such a subject-based list of the author's works is not directly available by built-in browsing, but once it has been placed in a named workset, it can be used as the basis for further browsing-based exploration. This could be used to answer questions such as whether books on medicine or music were sought by different collectors.

This use of worksets is a very nascent idea, and would need further investigation to fully assess and articulate its value to a researcher.

Future work: Thinking about worksets... maintain a local (or parallel) Solid Pod for storing workset lists of references. Use URI of workset collection to pass workset context forward when composing queries. When a new selection is made, use Javascript in the query web page to create a new workset entry, and pass its URI forward? This idea is explored further in [Thoughts for using Solid-based working sets](#).

This document was prepared for the Mapping Manuscript Migrations project, and is made available under a CC-BY-NC licence.

For further information, please contact:

Graham Klyne, Oxford e-Research Centre - graham.klyne@oerc.oc.ac.uk

Mapping Manuscripts Migrations: <https://mappingmanuscriptmigrations.org/>