KAD Milestone 1: Application Design

John Can Lokman

VU University Amsterdam

Abstract. This report documents the initial design of an application being made for the Knowledge and Data course of Vrije Universiteit Amsterdam. This application combines various data sources into linked data using an ontology, serializes this ontology and its instances as a triple store, and allows users to query it with a user-friendly interface. Below, a description of the application and its intended users were described and its initial design is summarized.

Date: 13.10.2017

1 Description of the Application and Users

1.1 Goal

The current project aims to create a linked-data interface that allows users to query a database of scientific publications in order to get detailed metadata concerning publication patterns . Through the interface, it is particularly aimed to allow a data-driven investigation of interdisciplinary collaboration patterns. Some research questions —among other possible exploratory ones— are as following:

- Are there any collaboration patterns that are biased towards a certain disciplines? For instance, when medical researchers and computer scientists collaborate on research projects, do they tend to publish their research on journals that belong to one of their respective disciplines (i.e., a medical journal versus a computer science journal).
- Does interdisciplinarity of a research project affect its impact (e.g., as measured with number of outgoing citations).
- Are there any publication patterns that can explain researcher career trajectories. For instance, do variables such as interdisciplinarity of an author's lifetime research, number of overall collaborations with other researchers (i.e., network size), and other similar variables affect career-related variables of researchers, such as influence (e.g., as measured by number of citation) or tenure attainment.

The project is part of the 10-month research project 'Knowledge Flows in Interdisciplinary Research' [10], and as the investigation progresses over the next few months, the current linked-data interface is expected to shed light to these research questions, and also motivate new ones.

1.2 Users

The intended initial user base for the linked data interface are the researchers involved in the Knowledge Flows in Interdisciplinary Research project: Dr. Ali Khalili, Dr. Sascha Friesike, Prof. Peter van den Besselaar; and Academy Assistants Frederik König and John Can Lokman. As this research trajectory continues in the following years, more researchers and students who are involved in proceeding —or similar— projects in both Vrije Universiteit Amsterdam and other universities can be expected to join the user base.
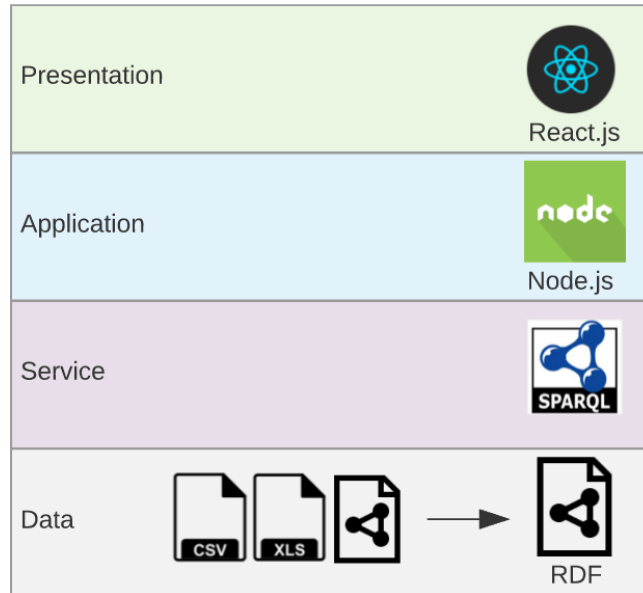
2 Design

Fig. 1. The technologies we are planning to use for the project, distributed to four layers of application design.

2.1 Data Layer

In the background, the application will be based on an ontology of scientific collaboration, which will model the landscape of scientific publications (i.e., scientific journals) and collaborations between authors and scientific fields. The ontology will be populated with instances by incorporating data from multiple sources

such as RISIS [8] and VU Research Portal of Vrije Universiteit Amsterdam—a service that is powered by Pure [3], Scopus [4], and Elsevier Fingerprint Engine [2]. More data sources that contain meta-scientific information could later be added as the project progresses. If such additional data sources do not come in RDF format, necessary transformations will be done using appropriate scripts or methods. (For a summary of application design, see Figure 1.)

2.2   Service Layer

The ontology and its instances will be serialized as a triple store and will be hosted online as linked open data.

2.3   Application Layer

Although this step is to be further specified, the application layer and SPARQL wrapper will likely be implemented using Node.js and Javascript, as part of Linked Data Reactor Framework [5].

2.4   Presentation Layer

Presentation layer will be implemented using mainly HTML and React.js as part of Linked Data Reactor Framework [5]. As the current application is primarily aimed for research purposes, it will be mostly implemented as a an accessible linked data browser designed with 'What you see is what you query' principle [6] (see Figure 3). A short walkthrough to the presentation layer is provided below.

Introduction and Step 1: Datasets Users will be greeted with an introduction page explaining the purpose of the application, and then will proceed to a page that shows them the datasets that will be used for their queries in the next page. This page will be visually similar to the one in Figure 2, and the users will likely be given the option to include or exclude databases with checkboxes.

Step 2: Linked Data Browser After selecting datasets to query, users will proceed to a linked data browser that will be similar to Figure 3. In this interface, users will be allowed to explore the data in a visual and accessible way.

Although the time constraints on the current course will likely will not allow incorporation of additional interfaces, a few possible ideas that may be realized after the course are summarized below.
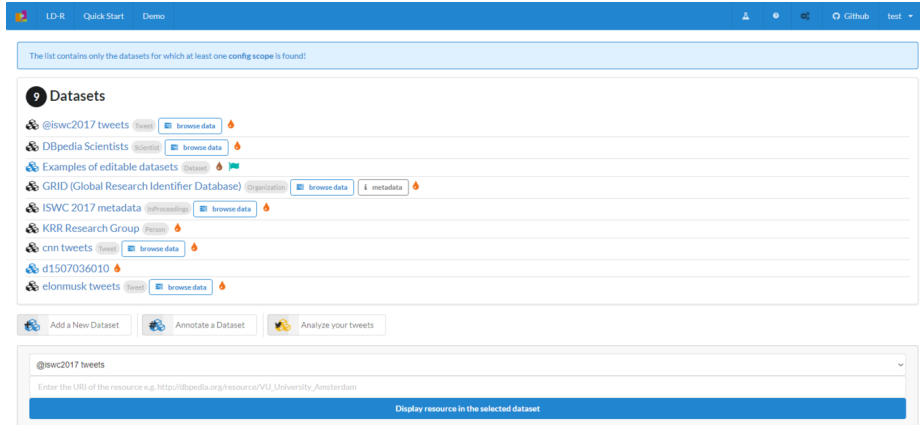
Fig. 2. An example page that lists available datasets. Screenshot taken from Linked Data Reactor [5]
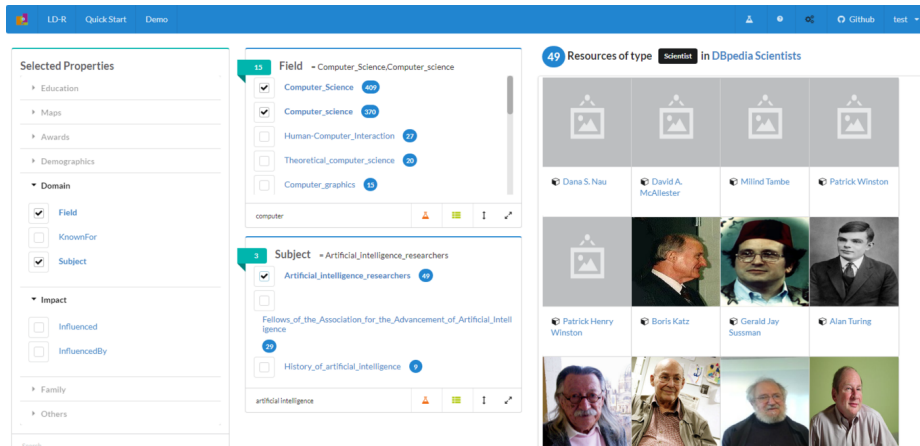

Fig. 3. A screen that closely resembles the envisioned linked data browser for the current project. Screenshot taken from Linked Data Reactor [5].

Possible Future Interface/Component - Search Engine In an alternative interface to that of Linked Data Reactor, a Google-like search engine built to search Meta-science could be used to greet users (see Figure 4). This search engine could allow searching for multiple research fields, researchers, and research projects. Alternatively, such a search box could also be added to the regular Linked Data Reactor browser in Figure 3.
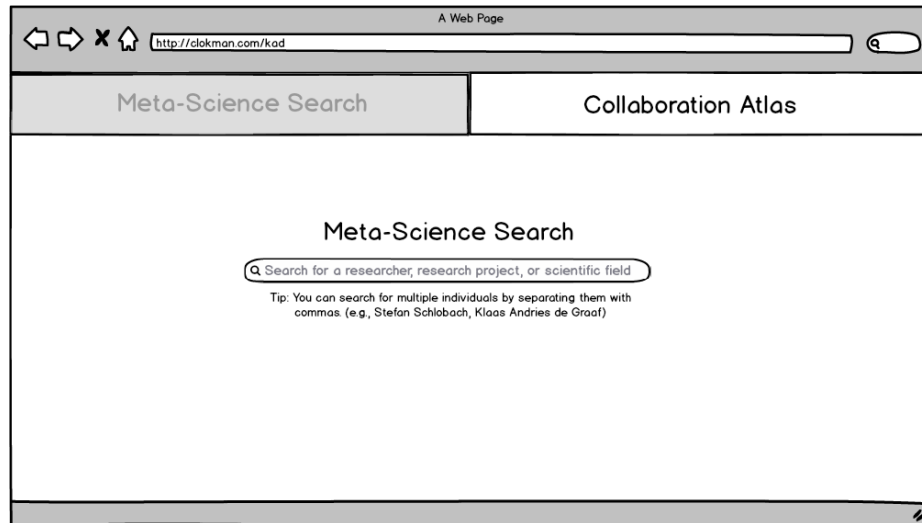
Fig. 4. A concept search engine to query a meta-scientific ontology and the instances it contains.

Possible Future Interface/Component - Linked-data-driven Visualizations The results of the search query could be returned as various interactive visualizations depending on the type of search queries entered (see Figures 5, 6, and 7). Such visualizations could be built using libraries such as D3 (for graphics) and React.js (for other interface elements and operations, such as removing keywords from search query on-the-go—i.e., without requiring users to go back to the initial search page).
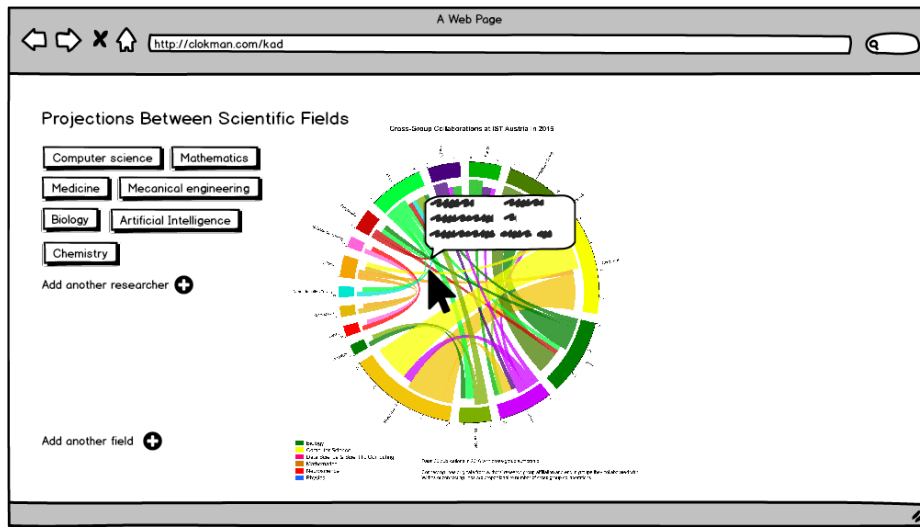
Fig. 5. A concept interactive visualization that could be generated as a response to a search query that relates to scientific fields or topics.
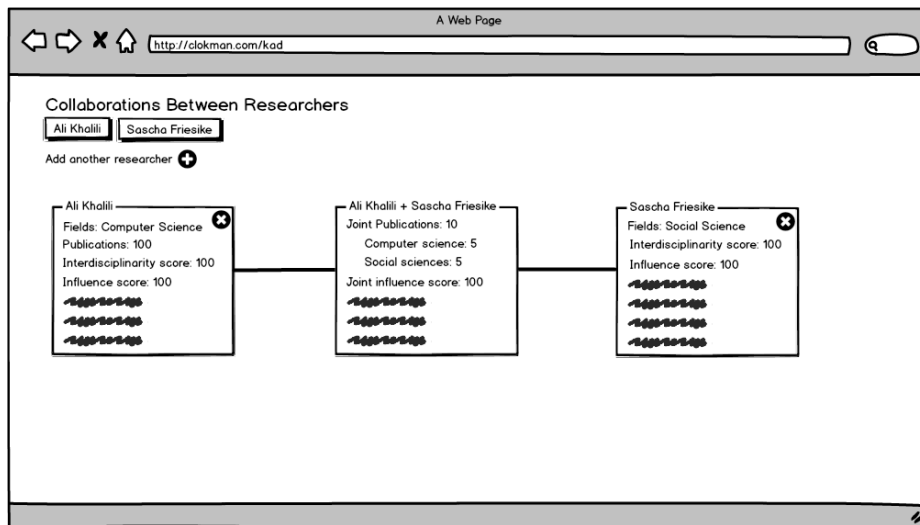

Fig. 6. An interactive visualization that shows collaboration between two researchers as well as other related information for each researcher individually.
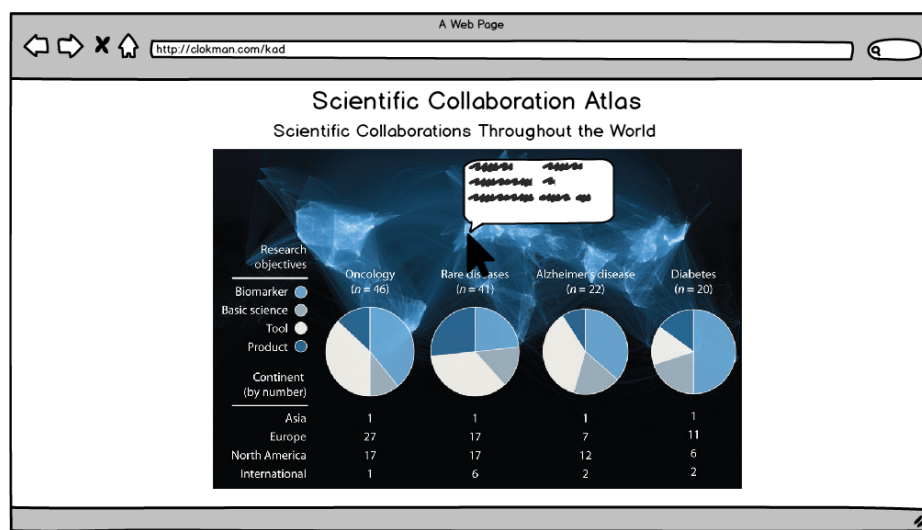
Fig. 7. A concept interactive visualization that maps collaborations based on author locations and research topic. Image taken from [7], and background visualization from [1].

References

1. Beauchesne, O.H.: Map of scientific collaboration between researchers. http://olihb.com/2011/01/23/map-of-scientific-collaboration-between-researchers/ (2011), `http://olihb.com/2011/01/23/map-of-scientific-collaboration-between-researchers/`, accessed on Fri, October 13, 2017
2. Elsevier: Elsevier Fingerprint Engine. https://www.elsevier.com/solutions/elsevier-fingerprint-engine (2017), `https://www.elsevier.com/solutions/elsevier-fingerprint-engine`, accessed on Fri, October 13, 2017
3. Elsevier: Pure. https://www.elsevier.com/solutions/pure (2017), `https://www.elsevier.com/solutions/pure`, accessed on Fri, October 13, 2017
4. Elsevier: Scopus. https://www.scopus.com/home.uri (2017), `https://www.scopus.com/home.uri`, accessed on Fri, October 13, 2017
5. Khalili, A., Loizou, A., van Harmelen, F., Andries de Graaf, K., Albert Merono-Penuela, Pek van Andel, P.v.: Linked Data Reactor. http://ld-r.org (2017), `http://ld-r.org`, accessed on Sat, October 14, 2017
6. Khalili, A., Merono-Penuela, A.: WYSIWYQ – What You See Is What You Query (2017), `http://research.ld-r.org/papers/wysiwyq.pdf`, accessed on Sat, October 14, 2017
7. Lim, M.D.: Consortium Sandbox: Building and Sharing Resources. Science Translational Medicine 6(242), 242cm6–242cm6 (jun 2014), `https://doi.org/10.1126%2Fscitranslmed.3009024`
8. RISIS Consortium: RISIS: Research Infrastructure for Science and Innovation Studies. http://risis.eu (2017), `http://risis.eu`, accessed on Fri, October 13, 2017
9. Stardog Union: Stardog: the Enterprise Knowledge Graph. http://www.stardog.com (2017), `http://www.stardog.com`, accessed on Fri, October 13, 2017
10. VU Network Institute: Academy Assistants and Projects. http://www.networkinstitute.org/academy-assistants/academy-projects-17/ (2017), `http://www.networkinstitute.org/academy-assistants/academy-projects-17/`, accessed on Fri, October 13, 2017