A Linked Data Driven Visual Interface for the Multi-Perspective Exploration of Data Across Repositories

Gengchen Mai[1], Krzysztof Janowicz[1], Yingjie Hu[2], Grant McKenzie[3]

[1] STKO Lab, University of California, Santa Barbara, USA
[2] University of Tennessee, Knoxville
[3] University of Maryland, USA

Abstract. As more data from heterogeneous sources become available, interfaces that support the federated exploration of these data are gaining importance to uncover relations between entities across multiple sources. Instead of explicit queries, visual interfaces enable a *follow-your-nose* style of exploration by which a user can seamlessly navigate between entities from different data sources. This requires an alignment of the ontologies used by said sources as well as the coreference resolution of entities across them. Together with Semantic Web technologies, the Linked Data paradigm provides the technological foundations to address these challenges. Nonetheless, the majority of work studies these components in isolation, focusing either on the alignment, coreference resolution, or visualization. Some interesting aspects, however, only arise when all puzzle pieces are in place. Two of these aspects are the seamless transitions between visualization and interaction paradigms as well as the combination of entity and type queries. In this work, we present a multi-perspective visual interface that enables the seamless exploration of major scientific geo-data sources that contain millions of RDF triples.

1 Introduction and Motivation

Linked Data as a paradigm describes how to break up data silos and support the publication, retrieval, reuse, and interlinkage of data on the Web. Together with other Semantic Web technologies, Linked Data shows promise to address many challenges that have affected semantic interoperability between repositories and services within and across domains that are highly heterogeneous in nature, e.g., the broader geosciences [9]. However, making use of the largely machine-oriented global graph of Linked Data also requires human-centric interfaces to query data or to explore it by following links across entities and even repositories. Unsurprisingly, user interfaces, vocabularies for their creation, and visual aids for the construction of SPARQL queries have been an active research area for many years [3,13,5,11,16]. The integration and deployment of such interfaces on top of heterogeneous and conflated sources, however, is still rare. In other words, research on topics such as ontology alignment, coreference resolution, visualization, querying, and so forth, often takes place in isolation. As a consequence, findings that only emerge once the full stack is implemented are frequently overlooked.

One such example is the fact that co-reference resolution without data conflation (fusion) hampers the reuse of data while one would intuitively assume that the opposite is true. The reason for this lies in the fact that data sources which contain data about the same entities share overlapping information. Consequently, to give a concrete example, establishing that two URIs identify the same entity without fusing the data about them leads to places having more than one (and different) population counts, geographic coordinates, names, and so forth, e.g., for Kobe, Japan in DBpedia and GeoNames.[1]

In this work, we are interested in aspects that arise from exploring Linked Data via graphical user interfaces and more specifically in three observations made when sharing scientific data from several major oceanographic repositories. **(I)** There is no *one size fits all* visualization and interaction paradigms. However, offering multiple perspectives on data works only if users can seamlessly change between these perspectives. **(II)** exploratory interfaces such as implemented by the popular Relfinder [10] benefit from query capabilities that enabled the user to select *entities* as well as *classes* as nodes. **(III)** With an increasing number of data sources (and triples), aspects that may seem like mere convenience function become essential features, e.g., the ability to expand and compress local nodes and edges in a graph view, support for multiple layers (from multiple data sources) in a map view, and so forth.

Graph	**#Triples**
bcodmo	592,467
combined *// (AGU+NSF)*	9,506,867
dataone	25,771,511
gebco	15,212
iodp	108,338
ngdb	5,817,710
r2r	692,873
sesar	2,445,348
wholib	113,977
Total count of triples	**45,064,303**

Table 1 Data repositories made available as Linked Data.

In this paper, we present an interface[2] that supports knowledge exploration across several federated geo-data sources by means of a modular collection of ontology design patterns[3], coreference resolution based on the owl:sameAs and skos:closeMatch predicates, and multiple perspectives including a tabular view (lens), a graph view, and a map view on the data. The used data sources include BCO-DMO, DataONE, IEDA, IODP, LTER, MBLWHOI Library, R2R, and a dataset of AGU abstracts and NSF award [8]; see table 1. Overall, the served data consists of more than 45 million triples about oceanographic (scientific) cruises, research vessels, instrumentation, researchers, research projects, undersea features such as seamounts, physical samples, organizations, and so forth. As the data stems from major repositories and is in use by

[1] Via *dbpedia:Kobe owl:SameAs geodata:Kobe. dbo:populationTotal 1536499. gn:Kobe gn:population1528478.*

[2] http://demo.geolink.org/

[3] http://schema.geolink.org/

the research community, a two-step process was taken for the coreference resolution. OWL:sameAs relations between entities within and across repositories are manually curated by domain experts. In addition, they are enriched with automatically learned skos:closeMatch relations. With respect to the graphical user interfaces, this means that sameAs links will be automatically explored, while an additional checkbox enables the integration of closeMatch results. For performance reasons, the involved repositories are regularly synchronized with a harvesting endpoint.[4]

Nonetheless, our work is not specific to any particular dataset. The multiple views are not merely different ways to represent the data visually but come with their own exploration styles. The tabular view supports classical follow-your-nose exploration. The map view supports a *layered* multi-source exploration of undersea features. Finally, the graph view implements a relation finder [10] access but extends it substantially by offering type-based queries and query compression on top. To the best of our knowledge, this is the first interface that supports layers from different sources and entity-to-type queries.

As a running example, we outline how the tabular view can assist users to get detail information about a researcher, how to relate this researcher to scientific cruises that he participated in, as well as the trajectories that scientific vessels took during these cruises. The initial view of the GeoLink interface is shown in figure 1.

2 Related Work

Visualizations have been widely applied in different research aspects of the Semantic Web. Visual analytic has been used in semi-automatic approached for ontology matching, e.g., AlignmentVis [1]. Visualization is also used for a comprehensive understanding of the evolution of ontologies or time-varying ontologies [4]. A user-oriented visual notation for OWL, VOWL [12], has been proposed to define a mapping from OWL language constructs to graph elements [6]. As for visual user interfaces for Linked Data exploration, lots of work have been proposed to facilitate users without any knowledge of Semantic Web technologies to construct SPARQL queries and explore Linked Data, a spatiotemporal example being the work by Scheider et al. [14].

CS AKTive Space [15] supports an overview of UK University research in Computer Science which includes topic similarity query and geographic representation. A tabular view of direct information of one entity and map representation of similar research topic are enabled. However, due to the lack of enough data, CS AKTive Space has been restricted to support a few services.

Linked Data Scientometrics [7] is another Linked Data-driven user interface to evaluate and analyze scientific works and explore the network of researchers. This interface serves as a middle layer to support users to query the dataset from different perspectives without requiring familiarity with SPARQL. It can be put on top of any Linked Dataset that uses the Bibo ontology.[5]

FedViz [17], a visual interface for SPARQL query and formulation, enables federated and non-federated SPARQL queries from distributed data sources from Life Sci-

[4] http://data.geolink.org/sparql

[5] http://bibliontology.com/

Fig. 1 The initial view of the GeoLink interface.

ence domains. However, FedViz can only support users to ask relatively simple questions which can be formalized as one federated/non-federated SPARQL query. Complex queries which require a combination of results of several different queries, like a path query over two given end nodes are not supported.

RelFinder [10] does support path queries over several nodes through a dataset. The nodes, however, are limited to entities which means it does not support entity-to-type path queries. RelFinder executes queries over only one dataset, e.g., DBpedia.

Based on analyzing several existing Linked Data-driven user interfaces supporting data exploration and querying, we present a novel user visual interface which enables the multi-perspective data exploration of different geo-data sources. A follow-your-nose exploration, map visualization, and path queries between entities as well as entity-to-type are supported.

3 Follow-your-nose Tabular Exploration

The first view of our combined interface supports classical follow-your-nose exploration which is the most common interaction with Linked Data (aside of direct SPARQL queries). In a first step, the user can select a type of entity, e.g., *Cruise* or *Researcher*, and then use search-while-you-type to select a particular entity of said type. Fig. 2 shows results for the oceanographer Peter Wiebe. The search spans multiple repositories and as long as coreference resolution links (here owl:sameAs) exist, the data will be grouped together in predicate-object style. The user can click on the objects to trigger another query that will select all predicate-object pairs for the newly selected subject, e.g., a specific cruise, thereby revealing information about cruises with Peter Wiebe as a participant. Consequently, further exploring the data will yield results such as the types of instruments used on a cruise in which Peter Wiebe participated. So far, nine core entity types are supported: datasets, cruises, vessels, instruments, physical samples, gazetteer features, researchers, organizations, and awards. Each of them offers different predicates to be explored, e.g., roles played on cruises, affiliations to institutions, trajectories taken by vessels during their cruises, and so forth. Finally, during any stage of the exploration, the user can click on the graph (or map) view icons to *seamlessly* switch to anther perspectives.

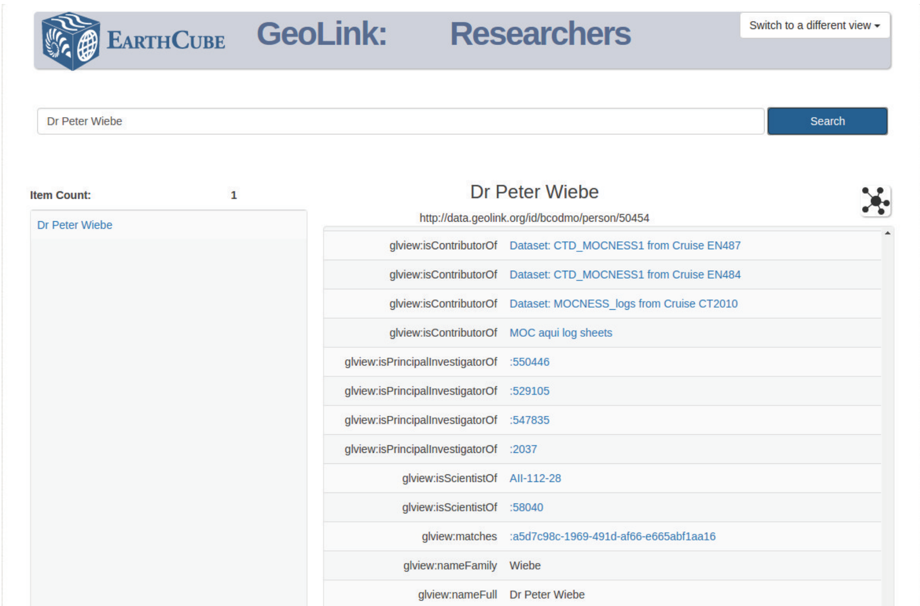

Fig. 2 The tabular view showing detail information of Peter Wiebe.

4 RelFinder Exploration Including Entity-to-type Queries

The second view builds up on the RelFinder system [10] and extends it with various features such as compressing and expanding a path, range queries around nodes, and mixed entity-to-type queries. In contrast to the first view, the user does not navigate step by step through the data but selects a source node, here Peter Wiebe, and a target node that can either be another entity, e.g., a specific vessel or researcher, or a type of entity such as *Cruise*. Our interface then performs n-degree path queries to uncover all subjects, predicates, and objects that are along the path from source to target. Fig. 3 shows a query from Peter Wiebe to the Cruise type. Depending on the maximum path distance (set to 4 here) the results will contain s-p-o chains such as scientific datasets to which Peter contributed and which were collected during certain cruises. To keep the interface responsive and clean, the user can request more paths (beyond the 10 set as default) and also compress or expand certain paths. Fig. 3 shows some expanded paths while others remain compressed.

While entity-to-entity (e.g., between researchers Wiebe and Chandler) queries will yields results within a reasonable time even for 6-degree queries, these will likely time out for most entity-to-type queries as all entities of the given target type have to be taken into account. Typical use cases for the relfinder-style view include finding all researchers that are using the same instruments as a particular researcher or that went on the same cruises. Right-clicking on any nodes allows the user to switch seamlessly to the table or map view, to visualize the immediate (1-degree) neighborhood of said node, or to set this node as source or target node for further exploration. For compressed paths, their path lengths are shown as numbers. Figure 3 also show owl:SameAs relations between researcher URIs and between cruise URIs.

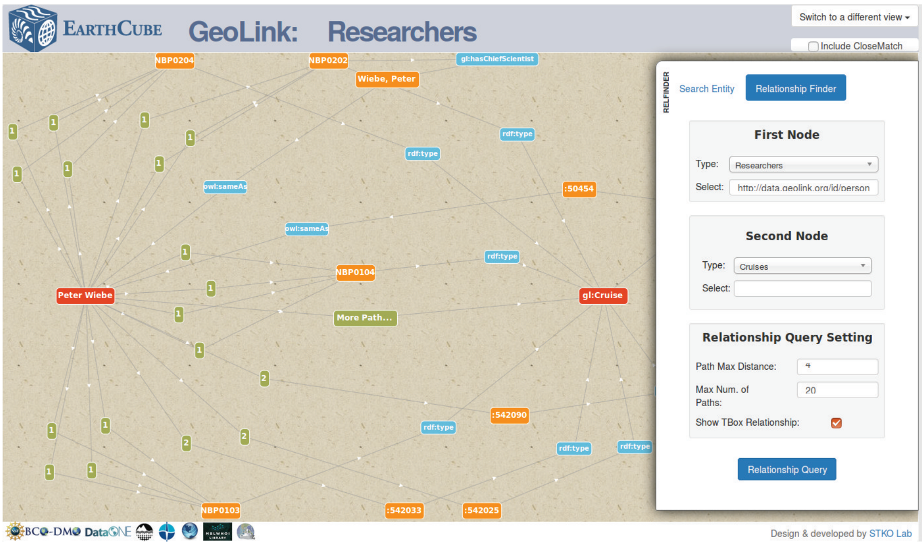

Fig. 3 Graph view showing cruises and datsets related to Peter Wiebe.

5 Multi-Layer Map Exploration

For a geospatial entity like cruise *AL9508*, it may be more appropriate to map out its geometry when a user is exploring the data repositories. For instance, after a user retrieved all cruises related to Peter Wiebe, (s)he can map out the geometries of any cruise by selecting the *Go to Map Visualization* option in the context menu, see Figure 4. Note that only entities which have a GeoSPARQL-conform WKT geometry will display this option in their context menu. The map layer container enables users to organize the geographical data in a map. The user can also map out any other geographical entities of the currently selected entity type using the search bar and the *Map Result* button. This functionality, for instance, can be used to retrieve, the trajectory of all cruises in which a certain researcher took part and then load in oceanographic gazetteer features to determine which of them may have been visited. By selecting such a feature, e.g., the Bahama Escarpment, the user can switch back to the tabular view or the graph view. Multiple layers can be added and enabled/disabled by a checkbox. These data can originate from different repositories and be of different types, e.g., undersea features, buoys, cruise trajectories, and so forth.

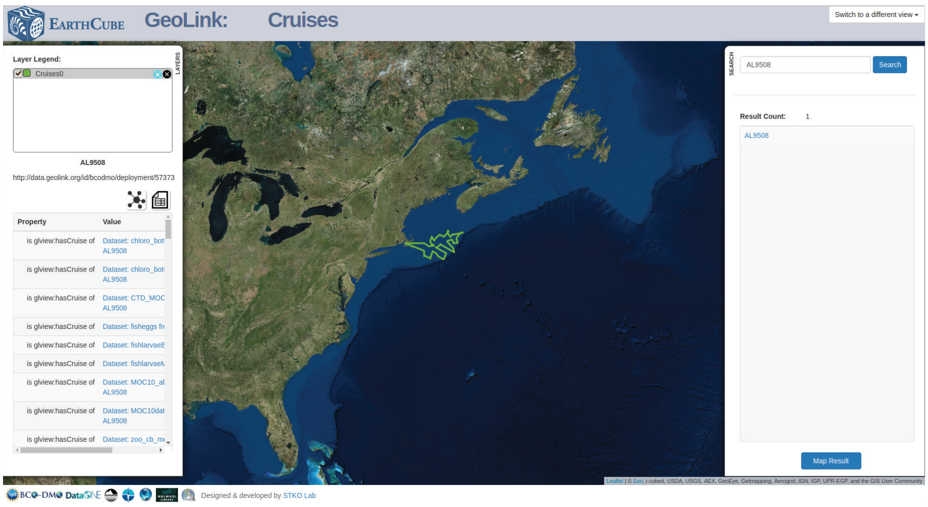

Fig. 4 Map view showing cruise ‘AL9508’ related to Peter Wiebe.

6 Conclusions

In this work, we introduced a Linked Data driven, multi-perspective interface that allows users to discover data across different repositories from three seamless perspectives, a tabular view, a graph view, and a map view. These perspectives enable users to discover detailed information about an entity, relationships between entities and between entity types, as well as the spatial distribution of entities. Our work thereby contributes to research on knowledge exploration across repositories. The data stems from 9 (major) oceanographic data sources and includes diverse data about researchers, institutes, research vessels, cruises, physical samples, instruments, datasets, undersea features, and so forth. While the data stems from different repositories, semantic interoperability is enabled via a set of ontology design patterns [8] together with manually curated owl:sameAs links and automatically mined skos:closeMatch relations for coreference resolution. The key challenge for a useful querying of Linked Data by domain experts lies in the realization of features that only become obvious when all the aforementioned components are in places.

Here we focused on three of them, namely the need for seamless changes between multiple perspectives on the data, relation-based exploration queries over entities and types, and convenience functions. For example, a user has to be able to switch from the tabular view about a specific cruise to a graph and a map view without losing focus, i.e., without having to enter the URI or the ID of the cruise again. While such a tabular perspective enables a user to follow his/her nose and explore new data step by step, other paradigms enable the user to explore the relations between two nodes or to map multiple geographic features at the same time. With respect to relation exploration, one interesting finding is that entity-to-type queries are often more useful than entity-to-entity queries. While features such as allowing for multiple layers, local range queries, collapsing property chains, and so forth, seem like mere convenience functions when regarded in isolation or toy examples, they rapidly gain importance for scientific application and when multiple sources are involved. In the future, we plan to add additional data sources and interaction possibilities to further strengthen the interface. A key issue that will define the success of exploratory interfaces is the quality and extent of coreference resolution which is currently ongoing. Finally, we also plan to test the interface by means of a user study.

On a side note, with respect to the underlying data, our work resonates with other current findings of the need for centralization [2] to achieve acceptable query performance and uptime. We believe that this is an issue that needs more attention and an open discussion within the Semantic Web community.

Acknowledgements.

The presented work is partially funded by the NSF award 1440202 *EarthCube Building Blocks: Collaborative Proposal: GeoLink – Leveraging Semantics and Linked Data for Data Sharing and Discovery in the Geosciences.*

References

1. Aurisano, J., Nanavaty, A., Cruz, I.F.: Visual analytics for ontology matching using multi-linked views. In: ISWC International Workshop on Visualizations and User Interfaces for Ontologies and Linked Data (Voila (2015)

2. Beek, W., Rietveld, L., Schlobach, S., van Harmelen, F.: Lod laundromat: Why the semantic web needs centralization (even if we don't like it). IEEE Internet Computing **20**(2), 78–81 (2016). DOI 10.1109/MIC.2016.43
3. Berners-Lee, T., Chen, Y., Chilton, L., Connolly, D., Dhanaraj, R., Hollenbach, J., Lerer, A., Sheets, D.: Tabulator: Exploring and analyzing linked data on the semantic web. In: Proceedings of the 3rd international semantic web user interaction workshop, vol. 2006. Athens, Georgia (2006)
4. Burch, M., Lohmann, S.: Visualizing the evolution of ontologies: a dynamic graph perspective. In: Proceedings of the International Workshop on Visualizations and User Interfaces for Ontologies and Linked Data (VOILA 2015). CEUR-WS, vol. 1456, pp. 69–76 (2015)
5. Dadzie, A.S., Rowe, M.: Approaches to visualising linked data: A survey. Semantic Web **2**(2), 89–124 (2011)
6. Haag, F., Lohmann, S., Siek, S., Ertl, T.: Visual querying of linked data with queryvowl. Joint Proceedings of SumPre pp. 2014–15 (2015)
7. Hu, Y., Janowicz, K., McKenzie, G., Sengupta, K., Hitzler, P.: A Linked-Data-Driven and Semantically-Enabled Journal Portal for Scientometrics. In: The Semantic Web–ISWC 2013, pp. 114–129. Springer (2013)
8. Krisnadhi, A., Hu, Y., Janowicz, K., Hitzler, P., Arko, R., Carbotte, S., Chandler, C., Cheatham, M., Fils, D., Finin, T., Ji, P., Jones, M., Karima, N., Lehnert, K., Mickle, A., Narock, T., O'Brien, M., Raymond, L., Shepherd, A., Schildhauer, M., Wiebe, P.: The geolink modular oceanography ontology. In: Proceedings of The 14th International Semantic Web Conference, Bethlehem, PA., pp. 301–309. Springer (2015)
9. Kuhn, W., Kauppinen, T., Janowicz, K.: Linked data-a paradigm shift for geographic information science. In: Proceedings of The Eighth International Conference on Geographic Information Science (GIScience2014), Berlin., pp. 173–186. Springer (2014)
10. Lohmann, S., Heim, P., Stegemann, T., Ziegler, J.: The relfinder user interface: interactive exploration of relationships between objects of interest. In: Proceedings of the 15th international conference on Intelligent user interfaces, New York, NY, USA, pp. 421–422. ACM (2010)
11. Mazumdar, S., Petrelli, D., Ciravegna, F.: Exploring user and system requirements of linked data visualization through a visual dashboard approach. Semantic Web **5**(3), 203–220 (2014)
12. Negru, S., Lohmann, S.: A visual notation for the integrated representation of owl ontologies. In: WEBIST, pp. 308–315 (2013)
13. Pietriga, E., Bizer, C., Karger, D., Lee, R.: Fresnel: A browser-independent presentation vocabulary for rdf. In: The semantic web-ISWC 2006, pp. 158–171. Springer (2006)
14. Scheider, S., Degbelo, A., Lemmens, R., van Elzakker, C., Zimmerhof, P., Kostic, N., Jones, J., Banhatti, G.: Exploratory querying of sparql endpoints in space and time. Semantic Web (Preprint), 1–22 (2015). DOI 10.3233/SW-150211
15. Shadbolt, N.R., Gibbins, N., Harris, S., Glaser, H., et al.: Cs aktive space: representing computer science in the semantic web. In: Proceedings of the 13th international conference on World Wide Web, pp. 384–392. ACM (2004)
16. Zainab, S., Hasnain, A., Saleem, M., Mehmood, Q., Zehra, D., Decker, S.: Fedviz: A visual interface for sparql queries formulation and execution. In: Proceedings of the International Workshop on Visualizations and User Interfaces for Ontologies and Linked Data co-located with 14th International Semantic Web Conference (ISWC 2015), vol. 1456, pp. 49–60 (2015)
17. e Zainab, S.S., Hasnain, A., Saleem, M., Mehmood, Q., Zehra, D., Decker, S.: Fedviz: A visual interface for sparql queries formulation and execution