From Research Objects to Research Networks: Combining Spatial and Semantic Search

Sara Lafia[1] and Lisa Staehli[2]

[1] Department of Geography, UCSB, Santa Barbara, CA, USA
[2] Institute of Cartography and Geoinformation, ETH Zurich, Zurich, Switzerland

Abstract. The spatial and semantic discovery of research objects extracted from sources available on the Web can be enabled with georeferenced and annotated metadata. Constraints on data retrieval are based on the types of queries and services that current repositories offer, which contribute to their limited usability. We address these constraints by illustrating a framework for a linked research network along with exemplary research questions, which demonstrate the added value that spatial and semantic annotation can contribute to information retrieval.

Keywords: annotation, data discovery, georeferencing, linked data

1 Introduction

The practice of publishing and sharing research data is broadly recognized as beneficial across diverse fields of study. Most current open data management systems either curate research institutional data across various domains, through a university library for instance, or manage domain-specific data, such as biomedical observations, across multiple institutions. As federated query capabilities across repositories are limited, search is either spatially restricted (e.g. to published UCSB research data) or thematically restricted (e.g. to dendritic cell research data). Initiatives promoting the open research data paradigm show the same trend. They evolve either by governmental inducement (e.g. Horizon2020[3]) or by research communities motivated to share knowledge and data (e.g. DC-Thera[4] for biology or PANGAEA[5] for environmental and earth sciences).

Spatial data infrasturctures, such as DataONE[6], span a wide range of datasets but do not enable the discovery of related publications. Keyword-based search engines, like Google Scholar, offer search for publications across repositories and disciplines, but do not provide access to associated datasets. Approaches that link publications by citations, such as the Citation Map [5], advance relations between research objects by exposing similar work being conducted in a particular research area, but they do not yet span disciplines.

---

[3] https://ec.europa.eu/programmes/horizon2020/
[4] http://dc-research.eu/
[5] https://www.pangaea.de/
[6] https://search.dataone.org/#data

2 Framework

A research network that enables both spatial and semantic discovery links heterogeneous research objects available on the Web. Adding spatial descriptors to research objects is a first step in the realization of integrated data networks. As location is an integrator of diverse contents [7], geo-semantic annotation improves information retrieval and enables data integration across domains [6].

Publications (hosted PDF files) and data (CSV files, geometry, imagery, etc.) are **research objects**, conceptualized as nodes in a research network. Geo-semantic annotations can be used to link research objects to a location and a spatial extent. Vocabularies like Dublin Core[7] provide a light-weight means for annotating and linking research objects, such as the term *coverage*, which can describe spatial scope. One publication may share network edges with one or more datasets, and conversely one dataset can be associated with multiple publications. Developing linked data **research networks** can enable broader discovery, horizontally across disciplines and vertically within topics.
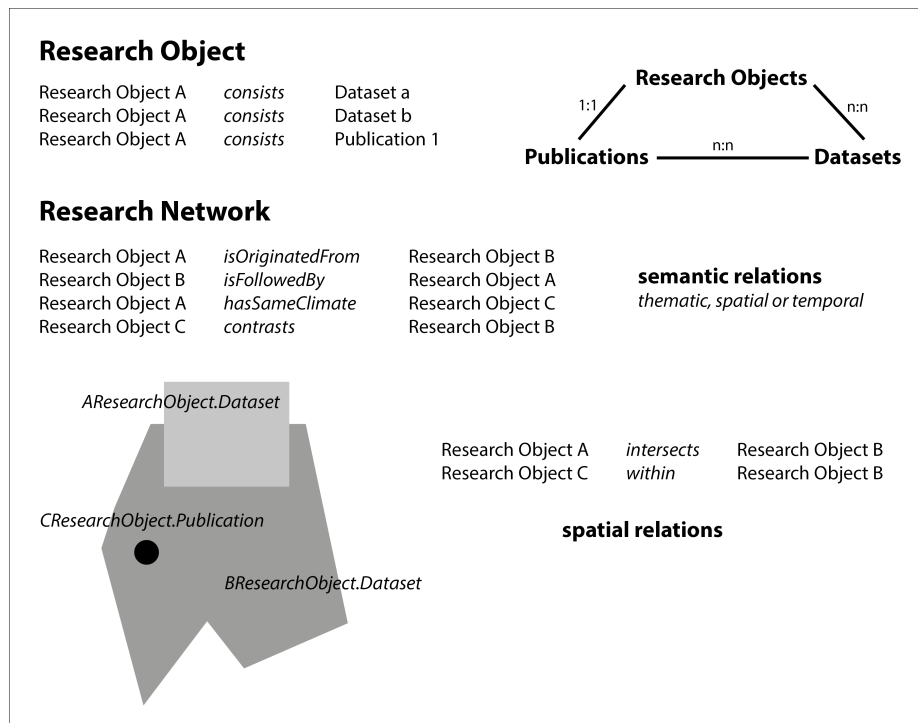

Fig. 1: Research object and research network conceptualization

A research object consists of a spatial and a semantic component. Links between research objects are the edges of the research network, representing either

a spatial or a semantic relation. Figure 1 shows the distinctions between relations. Spatial relations between research objects occur whenever the nodes share location, such as a study area extent or an institution. Semantic relations can also include spatiality, (e.g., the same conference to which several publications have been submitted). Moreover, semantic relations consist of a thematic and a temporal dimension. Once relations between objects have been formalized, they can be made explicit through annotation [6] following the linked data concept by storing triples that link research objects. Linked research objects can be explored within a spatial (similar location) or semantic (similar topic) neighborhood.

3   Use Cases

Simple extension of metadata enables the combination of spatial and semantic discovery for research objects. Sections 3.1 and 3.2 show exemplary research questions that can be answered by geo-semantic annotation of research objects along with existing search interfaces. Section 3.3 demonstrates queries that necessitate combined spatial and semantic search across a research network.

3.1   Local and Regional Scale

On a local or regional scale, exemplary questions raised from the research community focus on interdisciplinary data integration and reuse of existing datasets.
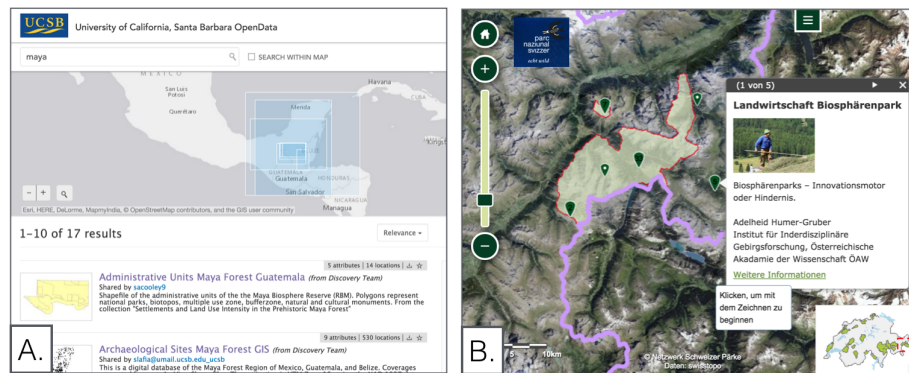

Fig. 2: Local scale search interfaces (A. UCSB Open Data[8]; B. Swiss National Park [9])

- What research has already been done with **similar data** in a location?
  Smart Cities: air quality, traffic, housing conditions, social media, etc.
- What kind of research has already been done in the **same study area**?
  Natural Reserves: Species distribution, soil and water samples, etc.

---

[8] http://discovery.ucsb.opendata.arcgis.com/
[9] http://www.nationalpark.ch/de/forschung/aktuelle-forschungsprojekte/

3.2 Global Scale

On a global scale, researchers are interested in reproducibility and comparison of phenomena across cultural or social contexts at multiple granularities.
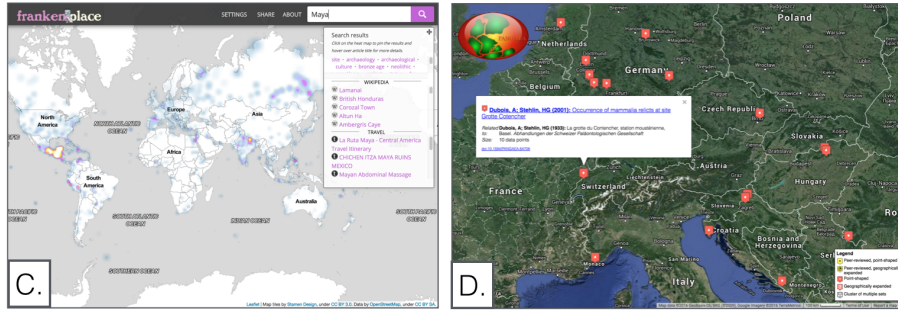

Fig. 3: Global scale search interfaces (C. Frankenplace[10]; D. Pangaea[11])

- How can my research question be applied to other datasets with a different cultural or social **context**?
  Linguistic reasoning, navigation studies, social science, brain images, etc.
- Can I **reproduce** my results or are there any significant differences if I use data from another spatial context? Where can I find more contrasting datasets and observations to **extend** my models?
  Image recognition, machine learning, social media data, climate models, etc.

3.3 Networks

The following queries can only be addressed by a research network that contains relations of a spatial, thematic, or temporal nature.

- *hasSameClimate*: Where can I find data that has been collected in the same climate? What other research has been done under the same conditions?
- *contrasts, supports*: Is there any research that shows a counterexample? Is there a publication that supports my research question or results?
- *isFollowedBy, isOriginatedFrom*: What research follows from this publication/dataset? Which publication mentioned the research interest first?
- *hasSimilarRegion, hasTemporalFrequency*: Where are mountain regions in which researchers monitor glaciers at least once every three years?
- *isNear*: Where are anthropology research clusters located?

Extracting georeferences from published research object metadata poses challenges to this vision [6]. However, extending terms in ontologies, such as *Prov:Location*

---

in PROV-O, is a first step toward better spatial descriptions of research data. Future challenges also include the use of spatio-temporal reasoning to parse, disambiguate, and reason on queries, necessitating a combination of spatial discovery and semantic analysis. The vision for a research network is best understood as a compiler placed between a back-end database and front-end web applications.

4 Prospects

Place-based search for research data is not a new notion [4] and is expressly supported by myriad Spatial Data Infrastructure initiatives, but the idea of integrating semantic discovery of associated research objects, such as publications, is novel. Some data management systems allow users to preview geographical data and perform place-based search for publications, combining spatial search with faceted browsing and keyword query. Publishing interfaces, such as DataONE, must continue to advance this trend by enabling researchers to easily annotate their observation metadata spatially, with footprints or coordinates, and semantically, with controlled vocabulary keywords.

On the front-end, interfaces for federated exploration across research networks already exist, but require back-end development to achieve integration. Additionally, front-end interfaces need to support spatial and semantic search functionality for users without any prior knowledge nor experience with geographic information systems. Data deposit and search tools can be made more accessible for researchers from diverse disciplines, enabling them to connect with an increasingly interdisciplinary user-base.

5 Acknowledgments

References

1. Adams, J. (2012). Collaborations: The rise of research networks. Nature, 490(7420), 335-336.
2. Bechhofer, S., Buchan, I., De Roure, D., Missier, P., Ainsworth, J., Bhagat, J., Goble, C. (2013). Why linked data is not enough for scientists. Future Generation Computer Systems, 29(2), 599611.
3. Field, D. (2008). Working together to put molecules on the map. Nature, 453(7198), 978.
4. Goodchild M. F., Anselin L., Appelbaum R. P., Harthorn B. H. (2000): Toward Spatially Integrated Social Science. International Regional Science Review, 23:2, 139-159.
5. Hu, Y., McKenzie, G., Yang, J.A., Gao, S., Abdalla, A. and Janowicz, K. (2014). A Linked-Data-Driven Web Portal for Learning Analytics: Data Enrichment, Interactive Visualization, and Knowledge Discovery. In LAK Workshops.

6. Janowicz, K., Scheider, S., and Adams, B. (2013). A geo-semantics flyby. In Reasoning web. Semantic technologies for intelligent data access (pp. 230-250). Springer Berlin Heidelberg.
7. Kuhn, W. (2012). Core concepts of spatial information for transdisciplinary research. International Journal of Geographical Information Science, 26(12), 2267-2276.