spatial@linkedscience – Exploring the Research Field of GIScience with Linked Data

Carsten Keßler[1], Krzysztof Janowicz[2], and Tomi Kauppinen[1]

[1] Institute for Geoinformatics, University of Münster, Germany
carsten.kessler|tomi.kauppinen@uni-muenster.de
[2] Department of Geography, University of California, Santa Barbara, USA
jano@geog.ucsb.edu

Abstract. Metadata for scientific publications contain various explicit and implicit spatio-temporal references. Data on conference locations as well as author and editor affiliations – both changing over time – enable insights into the geographic distribution of scientific fields and particular specializations. At the same time, these *byproducts* of scientific bibliographies offer a great opportunity to integrate data across different bibliographies to get a more complete picture of a domain. In this paper, we demonstrate how the Linked Data paradigm can assist in enriching and integrating such collections. Starting from the bibliographies of the GIScience, COSIT, ACM GIS, and AGILE conference series, we show how to convert the data to Linked Data and integrate the previously separate datasets. We focus on the spatio-temporal aspects and discuss how they help in matching and disambiguating entities such as authors or universities. We introduce a novel user interface to explore the integrated dataset, demonstrating the potential of Linked Data for innovative applications using spatio-temporal information, and discuss how more complex queries can be addressed. While we focus on bibliographies, the presented work is part of the broader vision of a Linked Science infrastructure for e-Science.

1 Introduction

Over the last decades several conference series and journals have been established to communicate and publish research on Geographic Information Systems and Science. However, without being a member of the research community and its different subfields, it is difficult to judge how the conferences and journals differ, which one should be selected for a specific publication, and where to meet scholars that share related research interests. While calls for papers and editorial boards can be used as indicators, they do not provide enough discriminatory power and overlap to a large degree. A single resource to learn about GIScience-related publications, events, researchers, their interconnections, and research affiliations is missing. While CiteSeer, Google Scholar, DBLP, SpringerLink, or Mendeley offer large amounts of metadata on scientific publications, the links between those datasets that would enable more complex queries are missing. Moreover, even within a specific bibliographic dataset it is difficult to track the identity of authors and their affiliations. For instance, several manual queries are required to find all spelling variants for a single author. Thus, even simple queries for co-authors that would help journal editors and program chairs to find reviewers without conflicts of interest are virtually impossible. This is especially striking as the information to answer such queries is usually part of the bibliographic data provided by publishers.

In this paper, we demonstrate that the Linked Data paradigm can provide us with the methods to interlink datasets and establish identity by using the spatio-temporal properties for matching and disambiguation. The term *Linked Data* refers to a set of principles to publish machine-readable and understandable data online that has been proposed by Tim Berners-Lee [1]. These principles make use of well-established Web standards for identifying and accessing data sources, along with lightweight semantics to create a global graph of data. This distributed and interlinked collection of datasets is also referred to as the Linked Data Cloud [2] and has been growing rapidly over the past years [3], with some geographic information sources such as GeoNames[1] acting as central hubs.

In this paper, we use an integration scenario for publications from different conference series, i.e., COSIT, GIScience, ACM GIS, and AGILE to illustrate the potential of spatio-temporal properties in Linked Data. As *non-information resources*, researchers or universities can only be in one place at a given time. Spatio-temporal information about these entities can act as identity criteria that allow us to match data that may be registered under different names in different datasets, and disambiguate different events that accidentally share a common (place) name. We discuss how to use the spatio-temporal properties in bibliographic data from conference series proceedings for identity reasoning. Besides the potential of spatio-temporal information to facilitate data integration, we demonstrate how the final product—an interlinked online collection of dereferenceable bibliographic resources, available through a standardized API—can act as the foundation for intuitive, exploratory user interfaces.

We show that by semantically annotating, integrating, and interlinking bibliographic datasets, we can answer complex queries. For instance, due to their history all conferences have their own areas of specialization and, therefore, attract a different audience. Which researchers act as bridges between these communities and conferences, i.e., have published in several conference series? Given a certain sub-field of GIScience, which event offers the highest probability to meet researchers who share the same interests? How do topics evolve over time, gain, and lose interest? We have made all presented datasets, APIs, and user interfaces freely available on the Web at `http://spatial.linkedscience.org` and are constantly enriching them. While we focus on bibliographic data here, the presented work has broader implications on e-Science [4] and is part of the vision of a Linked Science [5] infrastructure.

The remainder of the paper is structured as follows. In the next section, we discuss relevant related work and the used technologies. Section 3 describes the data conversion process and data sources in detail. Section 4 presents the data

[1] See `http://geonames.org`.

integration and mapping methodology and evaluates it with sample queries. Finally, we present the system architecture and a user interface that makes our inter-linked and enriched dataset available to the GIScience community. We conclude our work by pointing out limitations and directions for future work.

2 Related Work

Linked Data was proposed by Tim Berners-Lee as a practical, data-driven approach towards the vision of the Semantic Web. The approach consists of four principles [1]: (1) use of URIs as names for things; (2) use of HTTP URIs to enable look-up; (3) use of the Resource Description Framework (RDF) and SPARQL Protocol And RDF Query Language (SPARQL) standards for data encoding and querying; and (4) interlinking of datasets to enable discovery. As these principles build on established standards that have proven useful in the "Document Web", the approach was quickly adopted by the community [3]. The Linked Data Cloud [2] that was built on these principles is constantly growing. As of September 2009, it consists of more than 31 billion RDF triples, which are the basic building blocks of Linked Data [6].

The bibliographic domain has been one of the first fields to embrace Linked Data as a new way to publish bibliographic records online in a machine-readable way. Several hubs in the Linked Data Cloud provide bibliographic collections, especially with an academic focus, such as the ACM library,[2] DBLP,[3] or CiteSeer.[4] Major libraries such as the British Library or the German National Library already offer Linked Data services. In order to semantically annotate the published datasets, the community has been working on a number of vocabularies to describe their datasets, of which the bibliographic ontology BIBO is most widely used [7]. The Semantic Publishing and Referencing (SPAR) ontologies are a more detailed attempt towards semantic publishing that goes beyond classical metadata [8]. ArnetMiner[5] is an ongoing research effort based on bibliographic data, building a dataset for exploring aspects such as advisor-advisee relationships [9] or social network mining [10].

Unfortunately, GIScience is under-represented in systems such as ArnetMiner. Many relevant conference series and journals are not listed. Analyzing GIScience as research field is not new; for instance, Skupin has applied document spatialization approaches to visualize the domain [11]. Agarwal et al. [12] have used a social graph to study the interconnectedness of the community, while Grossner and Adams [13] used Latent Dirichlet allocation to study research trends by analyzing the full papers of the last ten COSIT conferences.

Linked Data has also recently been explored by several researchers as a new way of publishing spatio-temporal data. Examples include a Linked Data ver-

[2] See http://thedatahub.org/dataset/rkb-explorer-acm.
[3] See http://thedatahub.org/dataset/fu-berlin-dblp.
[4] See http://thedatahub.org/dataset/rkb-explorer-citeseer.
[5] See http://arnetminer.org/.

sion of OpenStreetMap [14], Ordnance Survey Linked data,[6] and data from the Sensor Web [15]. The main motivations are twofold: First, Linked Data enables adding light-weight semantic annotations to the data—a task which has bothered geographic information scientists for more than a decade [16]. Second, Linked Data offers a way to expose geographic information to a wider audience that does not know anything about the standards used in spatial information infrastructures. Recent developments such as the GeoSPARQL [17] query language show a strong tendency in the field to open up and make use of generic standards for data exchange that do not only work for geographic information.

In this paper, we merge the efforts from these two domains. We document how we approached this problem and show how the result enables novel user interfaces and interaction paradigms. As such, this paper is one step in realizing the vision of *Linked Science*[7] [5] as an infrastructure for e-Science, an approach to semantically annotate and interconnect scientific resources such as models, data, methods and evaluation metrics [4].

3 Conversion Process

This section describes the conversion process of bibliographic metadata in BibTex format into an RDF representation.

3.1 Input Data

The input data for our process consist of metadata for the conference series *International Conference on Geographic Information Science* (GIScience), *Conference On Spatial Information Theory* (COSIT), *ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (ACM GIS), and *AGILE International Conference on Geographic Information Science* (AGILE). For each series, the metadata for each paper in BibTex format were used, containing information about authors, year, title, proceedings title, editors, digital object identifier (DOI), pages, publisher, and author affiliations. The metadata sum up to 1256 papers (110 for GIScience, 331 for COSIT, 699 for ACM GIS, and 116 for AGILE) from a total of 36 proceedings volumes (5 for GIScience, 11 for COSIT, 15 for ACM GIS, 5 for AGILE).[8] Unfortunately, the metadata for the proceedings of some early conferences of the GIScience, ACM GIS and AGILE series were not available. We are constantly adding and enriching new data, e.g., to integrate the SDH, AutoCarto, and Spatial Cognition conference series.

[6] See http://data.ordnancesurvey.co.uk/.

[7] See http://linkedscience.org.

[8] Note that these numbers only describe our input data; they are not intended to analyze the domain or make any statements about the importance of the different conference series.

3.2 From BibTex to RDF

The first step towards a Linked Data version of the series proceedings is the conversion to RDF. For this purpose, a Java converter has been developed that iterates through the collection of BibTex files and generates RDF statements. As mentioned in Section 2, URIs serve as identifiers for Linked Data resources. It is therefore crucial to develop URI conventions as a first step that defines how the URIs for different kinds of resources will be structured. These URI patterns are then filled by certain properties from the input data:

- Paper:
  `http://spatial.linkedscience.org/context/paper/doi`*DOI*
  Example:
  http://spatial.linkedscience.org/context/acmgis/paper/doi10.1145/1653771.1653787
- Person:
  `http://spatial.linkedscience.org/context/person/person`*MD5-Hash*
  Example:
  http://spatial.linkedscience.org/page/context/person/person4a54e293d2c33b74e2aab49bb5c182b6
- Affiliation:
  `http://spatial.linkedscience.org/context/affiliation/affiliation`*MD5*
  Example:
  http://spatial.linkedscience.org/page/context/affiliation/affiliationc429ca266aa3fce217af9c8ef1524f9a

We followed the strategy to re-use any unique identifiers that were already present in the data, such as DOIs for the single papers, and only created our own identifiers when necessary. In case of the author names and affiliations, we created MD5 hashes from the input strings in order to prevent overly long URIs. At the same time, the hashing also removes any special characters such as umlauts from the URIs. The set of resources that is created by following these URI patterns then needs to be interlinked using properties (also referred to as predicates) from RDF vocabularies. The fields in the BibTex files can be mapped to RDF properties defined in existing and widely used vocabularies. Hence, there was no need to create new vocabularies. We have used properties from six existing vocabularies: Dublin Core (namespace `dc`), Friend Of A Friend (`foaf`), the Bibliographic Ontology (`bibo`), the Ontology for vCards (`vcard`), and the W3C Basic Geo Ontology (`geo`).[9] Figure 1 gives an overview of how the different resource types are interlinked and annotated with the properties from those vocabularies.

The `geo:lat`, `geo:lon` and `vcard:ADR` properties shown in the figure cannot be created directly from the BibTex input. They serve to encode the georeferences for all affiliations. To generate them, we queried the affiliation string from

[9] See http://dublincore.org/documents/dcmi-terms/, http://xmlns.com/foaf/spec/, http://bibliontology.com/specification, www.w3.org/2006/vcard/ns, and http://www.w3.org/2003/01/geo/ for the respective specifications.

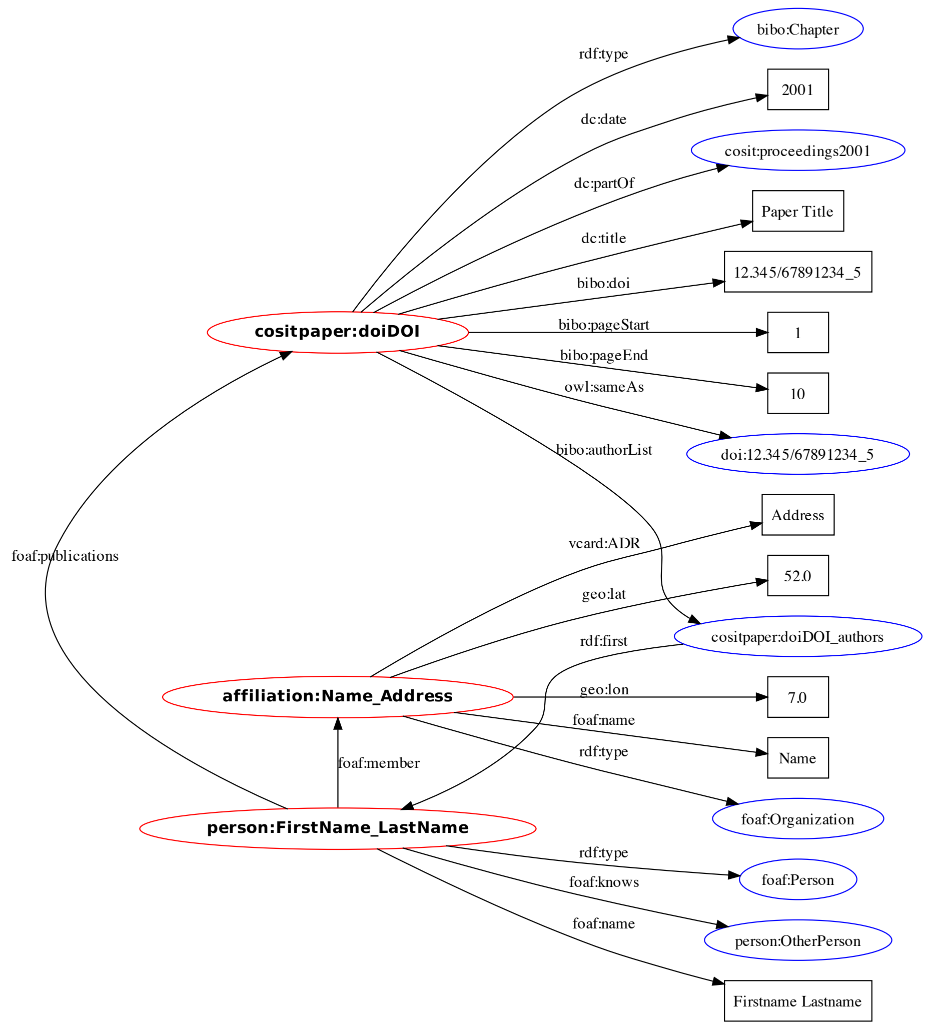

Fig. 1. Overview of interlinking schema. Resources are depicted as ellipses, with the three main types highlighted in bold font. Literals are depicted as boxes.

each BibTex file against the Google Geocoding API,[10] which returns both the latitude and longitude for the given location, along with a properly formatted version of the address. The affiliation strings often include very specific information that is not required for the geocoding process, such as working group or department information. This information is not only superfluous for the geocoder, it often even prevents finding a result. To cover these cases, we iteratively removed words from the beginning of the affiliation string until a result was returned, following the rationale that the strings usually start with more specific information and get generic towards the end. With this approach, we could successfully georeference all but 3 of the 1021 distinct affiliations. Note that the total number of affiliations (1021 for 1256 papers) is so high because we still have separate affiliation resources for every name and spelling variant at this point. We will tackle this issue in the following section.

The author lists and affiliations required a more complex transformation in order to fully represent the original information in RDF. In case of the author lists, it is not sufficient to link a paper to its authors via the `foaf:publications` property. This approach drops the order of authors, as any statement about a resource is treated equally. Therefore, when multiple statements of the same kind are available, they will appear in an arbitrary order in the output, thus losing the information who is first author, second author, and so on. The `author` entry from the BibTex file is therefore transformed into an RDF list that is linked to the paper resource via the `bibo:authorList` property, as shown in Figure 2.

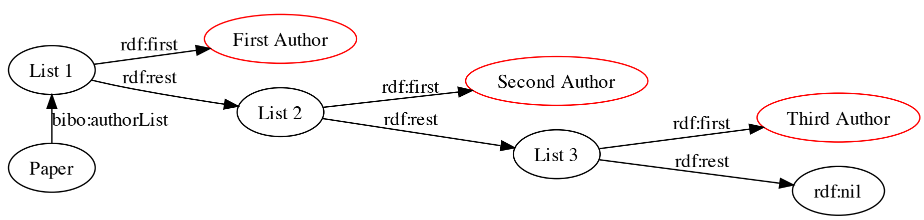

Fig. 2. Schematic view of nested author lists. `rdf:nil` marks the end of the list.

Likewise, it is not sufficient to only attach the affiliations directly to the authors via `affiliation:X foaf:member person:Y`, as this approach would lose the information *when* the author Y was affiliated with the institution X. We therefore reify the affiliation relationship, i.e., we turn the membership property into a resource, that has subject, predicate, and object attached as properties. This allows us to attach additional metadata to the whole statement—in this case, we attach the year of publication to retain the information when this relationship was valid:

[10] See http://code.google.com/apis/maps/documentation/geocoding/.

```
membership123  rdf:type        rdfs:Statement ;
               rdfs:subject    affiliation:X ;
               rdfs:predicate  foaf:member ;
               rdfs:object     person:Y ;
               dc:date         "2005" .
```

In order to include the social network aspect in the dataset, we also add a `foaf:knows` link between two persons if they have published a paper together.

4 Data Integration and Mapping

This section describes the mapping approach we applied to integrate and enrich the data. In order to keep track of data provenance, the RDF data for each conference is stored in a separate named graph (e.g. http://spatial.linkedscience.org/context/giscience). This allows us to keep track of where a specific triple comes from. In the following, we describe how we consolidated multiple URIs for the same person or affiliation and how we interlinked the data across the four named graphs.

4.1 Mapping Approach

The high number of affiliation resources generated by our initial conversion step – 1021 affiliations from 1256 papers – shows that people use different spelling variations of their affiliation in different papers. The same applies to author names, where some papers include first name and middle initials, others include only the first name (in potential spelling variants, e.g. *James* vs. *Jim*), or just initials. While it is fairly easy to see for a human reader that the author names *Michael F. Goodchild*, *M.F. Goodchild* and *Mike Goodchild* probably refer to the same person, these heterogeneities in the input data create challenges for the RDF generation, where we strive for a single URI that identifies a person or affiliation.

We approached this problem by combining spatial distance measures with string similarity measures. This mapping was carried out with the Silk link discovery framework [18]. In order to consolidate the high number of URIs for affiliations, we have compared the names of all organizations that were not located more than 10km apart. We had to pick such a comparably high range for the spatial search because the output of the geocoding process varied widely in terms of the levels of the location that were recognized. Depending on the input string, the results range from locations of exact street addresses or intersections to cases where only the state or even only the country was recognized. The 10km radius was selected as a rule of thumb in order to compare all places that have been automatically georeferenced into the same city. Within this range, all organization names were compared based on the Jaccard similarity coefficient [19]. It provides a normalized measure of token-based string similarity, calculating the size of the intersection divided by the size of the union of two sets of words.

It ignores the order of the words in the affiliation title, so that e.g., the two strings *'University of California NCGIA and Geography Department Santa Barbara'* and *'Department of Geography University of California Santa Barbara'* still yield a high similarity (0.78). The application of the Jaccard coefficient prevents matching of different organizations that are located near each other.

In case of the author names, we have applied a Levenshtein distance measure as a first step, which counts the number of editing steps between two strings [20]. Two names with an overall edit distance below 4 *and* exactly matching last names were used as candidates for consolidation. To prevent merging two different persons that happen to have very similar names, their affiliations were taken into account: if no common affiliations could be found, these persons were kept separate.

For both the affiliations and authors, the mappings were stored into a separate named graph at http://spatial.linkedscience.org/context/sameas. Each triple in this graph links two resources that have been identified as representing the same organization or person using the owl:sameAs property. As this property formally assigns *all* information about one of the linked resources to both of them, special care must be taken not to map resources if there is any doubt that they really refer to the same person or organization [21].

4.2 Linking Out

The fundamental idea of Linked Data is to interlink resources across different datasets. This process enriches local datasets with external data and embeds a dataset into the global graph. Our conference dataset contains external links to the Semantic Web Service of the GeoNames gazetteer.[11] For each georeferenced affiliation, we have created an outgoing link to the closest entry in GeoNames through its findNearby API, such as <http://spatial.linkedscience.org/context/affiliation/affiliationdb445b559df12b0d28fe03b432be04a0> foaf:basedNear <http://sws.geonames.org/2867543/> .

Adding a pointer to GeoNames to the dataset may seem redundant, since the affiliations are already georeferenced. Using the external data provided by GeoNames, however, enables new spatial queries, especially concerning administrative hierarchies. As every resource in GeoNames is part of a hierarchy tree, these outgoing links enable queries by country or continent, for example. The outgoing links to GeoNames are stored in a separate graph at http://spatial.linkedscience.org/context/geonames.

5 Architecture and Interaction

This section describes the client-server architecture of the application built on top of the dataset. It shows the user interface, explains the interaction workflow, and shows how the SPARQL endpoint can be used for complex queries.

[11] See http://www.geonames.org/ontology/.

5.1 Architecture

Having an integrated and inter-linked data set of GIScience-related conference series allows us to develop novel applications on top of it. The Web application described in this section is available at `http://spatial.linkedscience.org`. It is based on a client-server architecture that uses asynchronous JavaScript requests (AJAX) for communication. Figure 3 shows an overview of the architecture: The server hosts a static HTML interface, served through an nginx HTTP server; and a Parliament triple store that hosts the data and offers a GeoSPARQL endpoint for querying the data.[12] With this setup, it is possible to deliver an empty visualization frame to the client, which then updates the shown data after every user interaction by fetching the corresponding data from the SPARQL endpoint, without interrupting the user workflow by reloading the whole page.

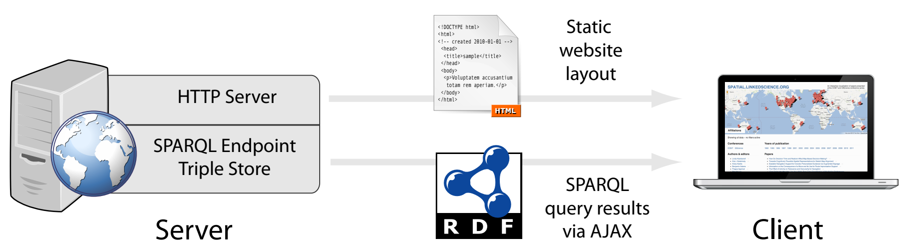

Fig. 3. Architecture overview.

5.2 User Interface and Interaction Workflow

The user interface shown in Figure 4 consists of five parts that show different facets of the dataset: the map for affiliations, a list of the conference series, a list of the years of publication, a list of all authors in the dataset, and a list of the titles of all papers. Each of the shown resources can be clicked to get further information. Clicking a marker on the map, for example, lists all authors affiliated with the corresponding organization, along with their papers, the years of publication, and conference series where they have published. Likewise, clicking a year will list all papers that have been published that year, along with their authors, affiliations, and the conferences that took place that year. With all data acting as thematic filters to the dataset, the user interface offers an *exploratory* interaction approach that allows the user to easily browse and navigate the collection. Figure 5 shows the map visualization that is triggered by clicking an author name. In this case, all affiliations of this author are selected from the dataset, ordered by date and connected by a polyline on the map. This visualization requires the reification discussed in Section 3.2, which allows us to annotate the `foaf:member` property with a timestamp.

[12] See `http://nginx.org/en/` and `http://parliament.semwebcentral.org/`. Both components are free and open source software.

Fig. 4. Screen shot of the user interface of http://spatial.linkedscience.org.

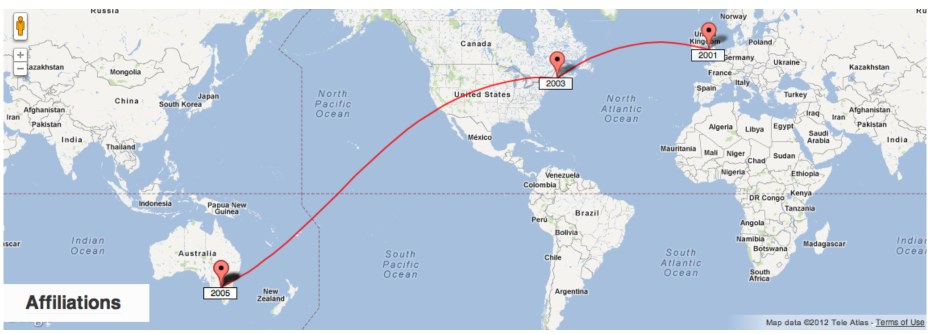

Fig. 5. Example of an author trace for Matt Duckham, showing the author's different affiliations including time stamps.

Any click on an element of the user interface causes an AJAX query to the server, which returns the corresponding results from the SPARQL endpoint in Javascript Object Notation (JSON) and visualizes them on the map. This asynchronous client-server interaction is handled by the JQuery framework.[13]

[13] See http://jquery.com/.

5.3 Complex Queries

While the previous subsection discussed querying via the graphical user interface, more complex queries can be directed to the SPARQL endpoint. For instance, to study the differences between sub-communities and their preferred conferences, one may query for those researchers who have published at all major series and, thus, act as bridge builders. The following query selects authors that have papers at ACM GIS, COSIT, and GIScience.

```
prefix foaf: <http://xmlns.com/foaf/0.1/>
SELECT DISTINCT ?author ?name WHERE {
  GRAPH <http://spatial.linkedscience.org/context/acmgis> {
   ?author foaf:publications ?a ; foaf:name ?name .}
  GRAPH <http://spatial.linkedscience.org/context/cosit> {
   ?author foaf:publications ?c .}
  GRAPH <http://spatial.linkedscience.org/context/giscience> {
   ?author foaf:publications ?d .}
}
```

Out of the resulting 23 researchers, we have manually selected those that have full papers in all series; see Table 1. Note that during the last 20 years, some of the conferences have changed their paper categories. We therefore manually excluded extended abstracts from ACM GIS. To show how the list of researchers changes if we add another conference series, we have additionally filtered for researchers that have published full papers at AGILE. While this is an international conference series, it is organized by the Association of Geographic Information Laboratories for Europe and, thus, rather attracts researchers that are or have been based in Europe.[14]

1. Benjamin Adams	7. Mark Gahegan	13. Andrea Rodríguez
2. Christophe Claramunt ([a])	8. Krzysztof Janowicz ([a])	14. John Stell
3. Matt Duckham	9. Christopher B. Jones	15. Egemen Tanin
4. Max J. Egenhofer	10. Lars Kulik	16. Stephan Winter ([a])
5. Leila De Floriani	11. Kai-Florian Richter	17. Michael Worboys
6. Andrew U. Frank ([a])	12. Claus Rinner	

Table 1. Authors that published full papers at ACM GIS, GIScience, and COSIT. Authors marked by an ([a]) also published full papers at the AGILE series.

A majority of the authors listed in Table 1 have either a background in computer science or are working together with computer scientists. This is due to the strong focus of ACM GIS on '*algorithmic, geometric, and visual considerations*'[15] and a noticeable difference to the GIScience conference series. Adding AutoCarto and the Spatial Cognition series to the query would change this picture and highlight different researchers and associated topics.

[14] So far, our data only covers AGILE papers published by Springer (starting 2007).

[15] See http://www.sigspatial.org.

6 Conclusions

Collections of bibliographic metadata allow for a detailed analysis of a research field and the corresponding community. So far, publishers and libraries as operators of such collections have come short of tapping this potential. In this paper, we have described a conversion and enrichment process based on the Linked Data paradigm that demonstrates the potential of such collections for the field of Geographic Information Science. We have discussed how the conversion and interlinking processes have been implemented. This workflow transforms the input into RDF and makes use of different online APIs and existing Linked Data sources for data consolidation and enrichment. We have shown how the spatio-temporal properties in the data can be exploited for more efficient data integration and reconciliation of resources from different origins.

Since we use the common BibTex format for the input data, the collection can be easily extended with further proceedings and additional conference series, which will be the next step to make http://spatial.linkedscience.org a reference portal for the community. Moreover, we plan to add new functionality such as free-text search. Adding more data will also bring up new research challenges, as an increase in the number of publications and authors brings up new challenges for the organization of the data. In order to facilitate browsing the collection by topic, we plan to add further keywords to the publications. Previous research has shown that such keywords can be extracted using Latent Dirichlet allocation [13]. These annotations would facilitate new kinds of analysis on the contents, such as the development of a certain research topic over time. While more data is important, mining for more relations is of equal value. In the future, we plan to subclass our *knows* relation with a supervisor-student relations and add thematic roles such as reviewers or organizers. In order to generalize this approach and make it useful for other research fields, some analyses would have to be reconsidered. The convention of the first author being the main investigator on the paper, for example, is not consistent across all fields.

Acknowledgements

This research has been partly funded by the *Linked Open Data University of Münster* project (see http://lodum.de), the International Research Training Group *Semantic Integration of Geospatial Information* (DFG GRK 1498), and the 52° North semantics community.

References

1. Berners-Lee, T.: Linked Data – Design Issues. Online: http://www.w3.org/DesignIssues/LinkedData.html; last accessed 2011-12-29 (2009)
2. Cyganiak, R., Jentzsch, A.: Linking Open Data cloud diagram. Online: http://richard.cyganiak.de/2007/10/lod/; last accessed 2012-01-06 (2011)

3. Bizer, C., Heath, T., Berners-Lee, T.: Linked Data – The Story So Far. International Journal on Semantic Web and Information Systems **5**(3) (2009) 1–22
4. Brodaric, B., Gahegan, M.: Ontology use for semantic e-science. Semantic Web **1** (April 2010) 149–153
5. Kauppinen, T., Baglatzi, A., Keßler, C.: Linked Science: Interconnecting Scientific Assets. In Critchlow, T., Kleese-Van Dam, K., eds.: Data Intensive Science. CRC Press, USA (forthcoming 2012)
6. Bizer, C., Jentzsch, A., Cyganiak, R.: State of the LOD Cloud. Online: http://www4.wiwiss.fu-berlin.de/lodcloud/state/; last accessed 2011-12-29 (2011)
7. Giasson, F., D'Arcus, B.: Bibliographic Ontology Specification. Online: http://bibliontology.com/specification; last accessed 2011-12-29 (2009)
8. Shotton, D., Portwin, K., Klyne, G., Miles, A.: Adventures in Semantic Publishing: Exemplar Semantic Enhancements of a Research Article. PLoS Comput Biol **5**(4) (2009)
9. Wang, C., Han, J., Jia, Y., Tang, J., Zhang, D., Yu, Y., Guo, J.: Mining advisor-advisee relationships from research publication networks. In: Proceedings of the 16th ACM SIGKDD int. conference on Knowledge discovery and data mining, ACM (2010) 203–212
10. Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., Su, Z.: ArnetMiner: Extraction and mining of academic social networks. In: Proceedings of the 14th ACM SIGKDD int. conference on Knowledge discovery and data mining, ACM (2008) 990–998
11. Skupin, A., Liggio, R.: Label determination for document spatialization. In: International Conference on Geographic Information Science (GIScience 2006), September 21–23, 2006, Münster, Germany. (2006)
12. Agarwal, P., Béra, R., Claramunt, C.: A Social and Spatial Network Approach to the Investigation of Research Communities over the World Wide Web. In Raubal, M., Miller, H., Frank, A., Goodchild, M., eds.: Geographic Information Science. Volume 4197 of Lecture Notes in Computer Science. Springer (2006) 1–17
13. Grossner, K., Adams, B.: COSIT at 20: Measuring Research Trends and Interdisciplinarity. In: Conference on Spatial Information Theory (COSIT 2011). (2011)
14. Auer, S., Lehmann, J., Hellmann, S.: LinkedGeoData: Adding a Spatial Dimension to the Web of Data. In: ISWC '09: Proceedings of the 8th International Semantic Web Conference, Berlin, Heidelberg, Springer-Verlag (2009) 731–746
15. Keßler, C., Janowicz, K.: Linking Sensor Data – Why, to What, and How? In Taylor, K., Ayyagari, A., De Roure, D., eds.: Proceedings of the 3rd Int. workshop on Semantic Sensor Networks 2010 (SSN10) at 9th Int. Semantic Web Conference (ISWC 2010), Shanghai, China. CEUR-WS 668 (2010)
16. Bishr, Y.: Overcoming the Semantic and Other Barriers to GIS Interoperability. International Journal of Geographic Information Science **12**(4) (1998) 299–314
17. Open Geospatial Consortium: OGC GeoSPARQL – A Geographic Query Language for RDF Data. Request for comments avaliable from http://www.opengeospatial.org/standards/requests/80; last accessed 2011-12-29 (2011)
18. Volz, J., Bizer, C., Gaedke, M., Kobilarov, G.: Discovering and Maintaining Links on the Web of Data. In: International Semantic Web Conference (ISWC2009), Westfields, USA. (2009)
19. Tan, P.N., Steinbach, M., Kumar, V.: Introduction to Data Mining. Addison Wesley (June 2006)
20. Levenshtein, V.: Binary codes capable of correcting deletions, insertions, and reversals. Soviet Physics Doklady **10**(8) (1966) 707–10
21. Halpin, H., Hayes, P.J.: When owl:sameAs isn't the Same: An Analysis of Identity Links on the Semantic Web. In: Workshop on Linked Data on the Web. (2010)