Knowledge Engineering for Historians on the Example of the *Catalogus Professorum Lipsiensis*

Thomas Riechert[1], Ulf Morgenstern[2], Sören Auer[1],
Sebastian Tramp[1], and Michael Martin[1]

[1] AKSW, Institut für Informatik, Universität Leipzig, Pf 100920, 04009 Leipzig
`lastname@informatik.uni-leipzig.de`
`http://aksw.org`
[2] Historisches Seminar, Universität Leipzig, Pf 100920, 04009 Leipzig
`lastname@uni-leipzig.de`
`http://www.uni-leipzig.de/histsem`

Abstract. Although the Internet, as an ubiquitous medium for communication, publication and research, already significantly influenced the way historians work, the capabilities of the Web as a direct medium for collaboration in historic research are not much explored. We report about the application of an adaptive, semantics-based knowledge engineering approach for the development of a prosopographical knowledge base on the Web - the *Catalogus Professorum Lipsiensis*. In order to enable historians to collect, structure and publish prosopographical knowledge an ontology was developed and knowledge engineering facilities based on the semantic data wiki *OntoWiki* were implemented. The resulting knowledge base contains information about more than 14.000 entities and is tightly interlinked with the emerging Web of Data. For access and exploration by other historians a number of access interfaces were developed, such as a visual SPARQL query builder, a relationship finder and a Linked Data interface. The approach is transferable to other prosopographical research projects and historical research in general, thus improving the collaboration in historic research communities and facilitating the reusability of historic research results.

1 Introduction

The World Wide Web, as an ubiquitous medium for publication and exchange, already significantly influenced the way historians work: the online availability of catalogs and bibliographies allows to efficiently search for content relevant for a certain investigation; the increasing digitization of works from historical archives and libraries, in addition, enables historians to directly access historical sources remotely. The capabilities of the Web as a medium for collaboration, however, are only starting to be explored. Many, historical questions can only be answered by combining information from different sources, from different researchers and organizations. Also, after original sources are analyzed, the derived information is often much richer, than can be captured by simple keyword indexing. These factors pave the way for the successful application of knowledge engineering techniques in historical research communities.

P.F. Patel-Schneider et al.(Eds.): ISWC 2010, Part II, LNCS 6497, pp. 225–240, 2010.

In this article we report about the application of an adaptive, semantics-based knowledge engineering approach for the development of a prosopographical knowledge base. In prosopographical research, historians analyze common characteristics of historical groups by studying statistically relevant quantities of individual biographies. Untraceable periods of biographies can be determined on the basis of such accomplished analyses in combination with statistically examinations as well as patterns of relationships between individuals and their activities.

In our case, researchers from the historical seminar at Universität Leipzig aimed at creating a prosopographical knowledge base about the life and work of professors in the 600 years history of Universität Leipzig ranging from the year 1409 till 2009 - the *Catalogus Professorum Lipsiensis* (CPL). In order to enable historians to collect, structure and publish this prosopographical knowledge an ontological knowledge model was developed and incrementally refined over a period of three years. The community of historians working on the project was enabled to add information to the knowledge base using an adapted version of the semantic data wiki OntoWiki [1][1]. For the general public, a simplified user interface[2] is dynamically generated based on the content of the knowledge base. For access and exploration of the knowledge base by other historians a number of access interfaces was developed and deployed, such as a graphical SPARQL query builder, a relationship finder and plain RDF and Linked Data interfaces. As a result, a group of 10 historians supported by a much larger group of volunteers and external contributors collected information about 1,300 professors, 10,000 associated periods of life, 400 institutions and many more related entities.

The benefits of the developed knowledge engineering platform for historians are twofold: Firstly, the collaboration between the participating historians has significantly improved: The ontological structuring helped to quickly establish a common understanding of the domain. Collaborators within the project, peers in the historic community as well as the general public were enabled to directly observe the progress, thus facilitating peer-review, feedback and giving direct benefits to the contributors. Secondly, the ontological representation of the knowledge facilitated original historical investigations, such as historical social network analysis, professor appointment analysis (e.g. with regard to the influence of cousin-hood or political influence) or the relation between religion and university. The use of the developed model and knowledge engineering techniques is easily transferable to other prosopographical research projects and with adaptations to the ontology model to other historical research in general. In the long term, the use of collaborative knowledge engineering in historian research communities can facilitate the transition from largely individual-driven research (where one historian investigates a certain research question solitarily) to more community-oriented research (where many participants contribute pieces of information in order to enlighten a larger research question). Also, this will improve the reusability of the results of historic research, since knowledge represented in structured ways can be used for previously not anticipated research questions.

The article is structured as follows: we present the overall technical architecture of the knowledge engineering approach in Section 2. We describe how the collaboration in

[1] Online at: `http://ontowiki.net`
[2] Available at: `http://www.uni-leipzig.de/unigeschichte/professorenkatalog/`

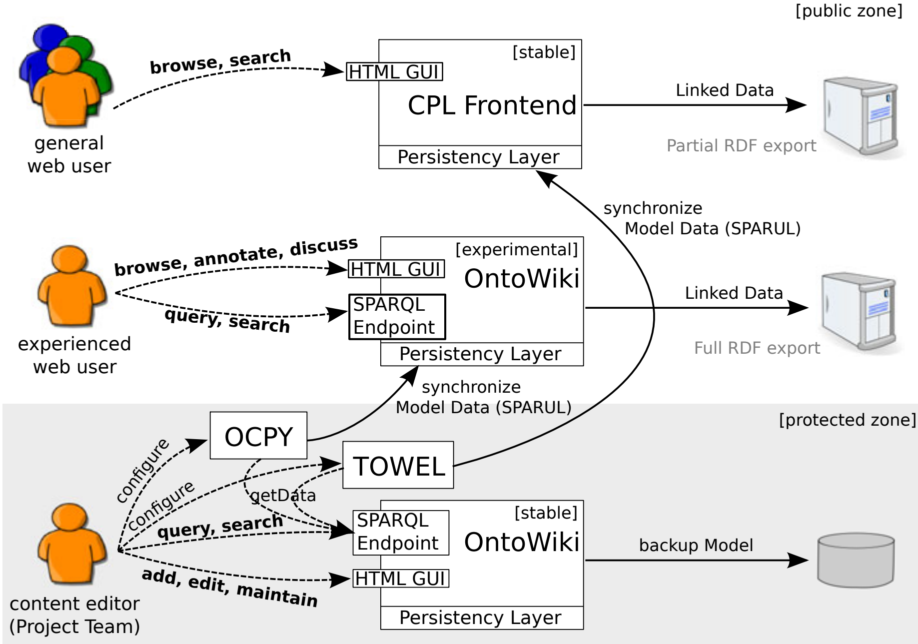

Fig. 1. Architectural overview about the project platforms

the historian community was facilitated by the semantic data wiki approach in Section 3. The underlying ontology is presented in Section 4. We elaborate on the knowledge engineering methodology in Section 5. Different exploration and access interfaces are showcased in Section 6. We present some prosopographical use cases in Section 7 and conclude in Section 8 with an outlook on future work.

2 Architectural Overview

The system architecture of CPL comprises a combination of different applications, which interact using standardized interfaces as illustrated in Figure 1. The project-set-up provides specialized human user interfaces for different user-groups according to the needs and tasks as well as generic access interfaces for machines including Linked Data and SPARQL endpoints. We divided the architecture into two separated zones (public and protected zone) due to technical constraints and in order to prevent security problems.

The semantic data wiki OntoWiki located in the protected layer [3] uses the *Catalogus Professorum Model* (CPM), which comprises several ontologies and vocabularies for structuring the prosopographical information (cf. Section 4). The project team, consisting of historians supported by knowledge engineers and semantic web experts, is working collaboratively and spatially distributed (e.g. in archives or libraries) to collect, structure and validate information about persons and institutions relevant to this

[3] `http://professoren.ontowiki.net` [restricted access]; OntoWiki-Version 0.85

knowledge domain. The resulting knowledge base is accessible only by the project team and is backed-up daily. Using the two configurable tools OCPY[4] (**O**ntology **C**o**PY**) and TOWEL[5] (**L**ight**w**eight **O**ntology **E**xport **T**ool) the knowledge base is exported in order to make it accessible for the public.

For general web users the catalog is integrated in the public website of the University of Leipzig[6]. Due to technical limitations and security considerations on the web server of the university, a simplified user interface consisting of plain HTML and Linked Data Information is generated nightly from the knowledge base, using TOWEL.

Domain experts, i.e. historians, are able to interact with CPL via an experimental version[7] of OntoWiki, which is presented in Section 3. The version of the catalog available there is synchronized using the tool OCPY, that exports data from the protected OntoWiki installation, transforms the exported data considering any linked knowledge bases and imports the changed data into this experimental installation. This experimental deployment in particular offers new functionality of OntoWiki for testing purposes. In addition to this complete catalog, smaller subsets are provided (e.g. a catalog containing only professors born in the city of Dresden[8]). Users are able to register at the platform, to use community features such as resource commenting and tagging, or sharing SPARQL queries with other users.

3 Collaboration Using a Semantic Data Wiki

The core of CPL is OntoWiki - a tool for browsing and collaboratively editing RDF knowledge bases. It differs from other Semantic Wikis insofar as OntoWiki uses RDF as its natural data model instead of Wiki texts. Information in OntoWiki is always represented according to the RDF statement paradigm and can be browsed and edited by means of views. These views are generated automatically by employing the ontology features such as class hierarchies or domain and range restrictions. OntoWiki adheres to the Wiki principles by striving to make the editing of information as simple as possible and by maintaining a comprehensive revision history. This history is also based on the RDF statement paradigm and allows to roll-back prior change-sets. OntoWiki has recently been extended to incorporate a number of Linked Data[9] features, such as exposing all information stored in OntoWiki as Linked Data as well as retrieving background information from the Linked Data Web [5]. Apart from providing a comprehensive user interface, OntoWiki also contains a number of components for the rapid development of Semantic Web applications, such as the RDF API Erfurt[10], methods for authentication, access control, caching and various visualization components.

OntoWikis main interface consist of two types of views as shown in Figure 2.

[4] http://catalogus-professorum.org/tools/ocpy/
[5] http://catalogus-professorum.org/tools/towel/
[6] http://www.uni-leipzig.de/unigeschichte/professorenkatalog/
[7] http://catalogus-professorum.org/
[8] http://catalogus-professorum.org/Dresden/
[9] http://linkeddata.org/
[10] http://aksw.org/Projects/Erfurt/

Fig. 2. OntoWiki views: (background) A tabular list view, which contains a filtered list of resources highlighting some specific properties of those resources and (foreground) a resource view which allows to tag and comment a specific resource as well as editing all property values.

List views. deal with the filtering and managing of resource lists. The user creates a list view by submitting a search keyword or selecting a class in the navigation module[11]. Subsequently, the user can apply multiple filter conditions to the list, which modify the underlying SPARQL query operating on the triple store.

Resource views. combine all information stored in OntoWiki about a specific resource. They are rendered by selecting a resource from a list view, requesting a resource directly via bookmark or link from an external page or selecting a resource in any other OntoWiki module. Resource views also allow the user to manipulate the selected resource. Starting from a resource view one can easily add and delete statements, as well as tags or comments.

4 The Catalogus Professorum Model

In this section we give an overview on the *Catalogus Professorum Model* (CPM), which is used to structure the prosopographical knowledge base. Although, the conceptual

[11] This module is not restricted to display class hierarchies, but allows to navigate through all types of hierarchies (e.g. group, geo-spatial or taxonomic hierarchies).

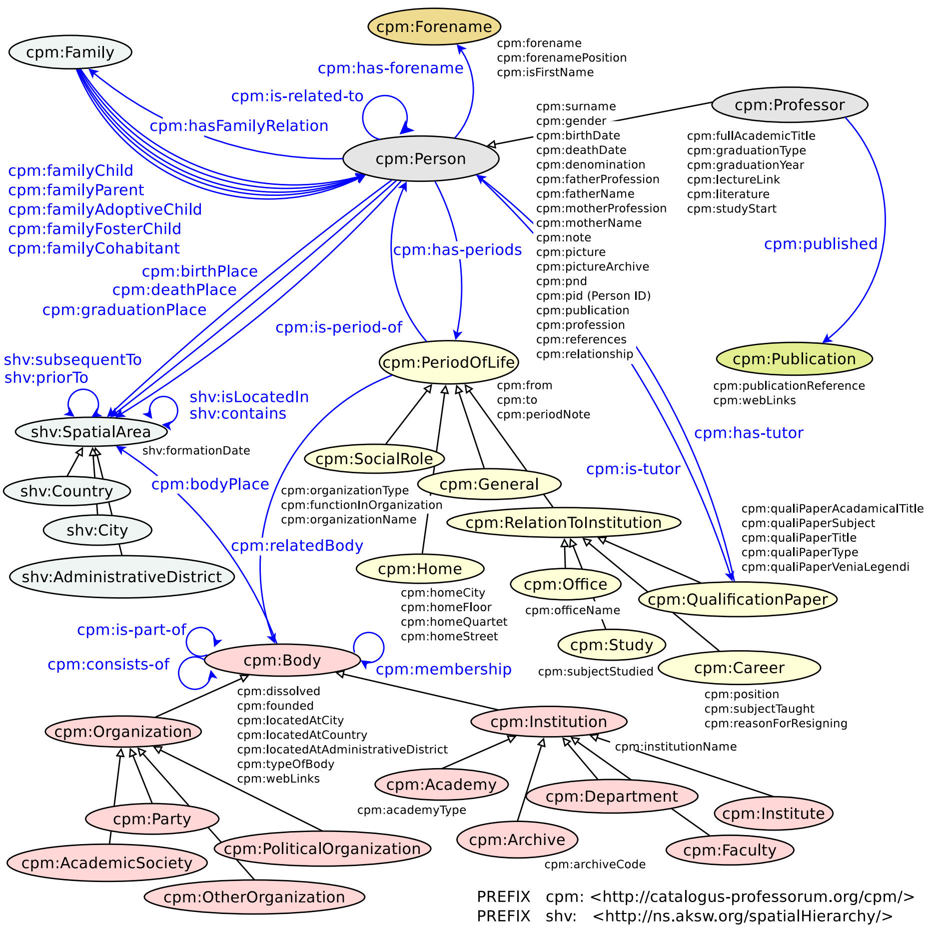

Fig. 3. The Catalogus Professorum Model (CPM)

model is initially only used for the Catalogus Professorum Lipsiensis the model was designed to be used for similar prosopographical knowledge bases at other universities. To facilitate reusability, CPM resides in its own namespace[12] and it uses identifiers with localized labels (currently English and German). This will in particular allow a simple integration of information from different sources at a later stage.

In its current version CPM contains 30 classes and 104 properties, 25 of which are object and 79 data properties. An birds-eye view of CPM is depicted in Figure 3. The central concepts in CPM are `cpm:Person` (with the subclass `cpm:Professor`), `cpm:FamilyRelation`, `cpm:Body` (with subclasses `cpm:Organization` and `cpm:Insitution`), `cpm:Publication`, `cpm:PeriodOfLife` and `cpm:SpatialArea` (with subclasses `cpm:Country`, `cpm:City` and `cpm:AdministrativeDistric`).

[12] CPM namespace: `http://catalogus-professorum.org/cpm/`

From a knowledge representation point-of-view, CPM currently uses rather shallow description logic expressivity. The used features are currently limited to subclass-superclass and subproperty-superproperty relationships, domain, range and simple cardinality restrictions, inverse properties and disjointness axioms. Most of the defined properties are defined as either object or datatype properties. CPM does not contain any deep class hierarchies. In the following paragraphs we describe some of the core concepts of CPM in more detail:

Historic persons and professors. The core information about persons in a prosopographical catalog comprises their name, information about birth and death, graduation and many more (cf. Figure 3). We explain in Section 5 why existing vocabularies such as FOAF [13], ULAN[14] or VIAF[15] were not reused and how a mapping from properties such as `cpm:birthDate` to these vocabularies will be achieved. A crucial resource for the interlinking of person data in the German speaking region is the Personen Namen Datei (PND) of the German National Library. Projects such as PND/BAECON[16] or LinkedHistory[17] enable the interlinking of existing databases on the basis of this identifier.

Periods of Life. For prosopographical research it is paramount to have a fine-grained representation of different periods within the life of a certain person. In order to capture such knowledge we introduced the concept `cpm:PeriodOfLife`, which is associated with a `Person` through the properties `cpm:has-period` and its inverse `cpm:is-period-of`. Different types of periods are distinguished by the subclasses `cpm:SocialRole`, `cpm:Home`, `cpm:RelationToInstitution`, `cpm:Office` (e.g. dean, rector), `cpm:Study`, `cpm:Career`, `cpm:QualificationPaper` (e.g. dissertation, habilitation) or `cpm:General`. Each of these types of periods of life is the domain of a number of properties, which are used to describe a particular instance in more detail, however, all inherit the delimiting properties `cpm:from` and `cpm:to`. Different periods of life of the same person can overlap, e.g. the `cpm:Home` usually overlaps with other periods.

Bodies. The bodies class is used to describe organizations (e.g. parties) and institutions (e.g. academies). Persons can be members of more than one body.

Historic specifics of spatial areas. CPM uses the independently developed *spatial hierarchy vocabulary* (SHV)[18] to represent spatial information. The core of SHV is the class `shv:SpatialArea` with subclasses `shv:City`, `shv:Country`, `shv:AdminstrativeDistrict`. Using the properties `shv:contains`

[13] Friend Of A Friend, `http://www.foaf-project.org/`
[14] Union List of Artist Names Online,
`http://www.getty.edu/research/conducting_research/vocabularies/ulan/`
[15] Virtual International Authority File, `http://viaf.org/`
[16] `http://meta.wikimedia.org/wiki/BEACON`
[17] `http.//linkedhistory.aksw.org/pnd/`
[18] `http://ns.aks.org/spatialHierarchy/`

and `shv:isLocatedIn` instances belonging to these classes can be arranged in spatial hierarchies of arbitrary depth. SHV is aligned with the World Geodetic System vocabulary WGS84[19]. In order to support the representation of historic administrative divisions, each spatial area can be equipped with properties `shv:formationDate`, `shv:dissolutionDate` as well as `shv:priorTo` and `shv:subsequentTo`. Thus, the historic administrative evolution can be easily represented. Königsberg, the early capital of Prussia, for example, is located in the following historic administrative districts (respective countries): Duchy of Prussia (from 10 April, 1525), the united Duchy of Prussia and Brandenburg (from August 27, 1618), Kingdom of Prussia (from 18 January, 1701), the Free State of Prussia (from 9 November, 1918) until its abolition at 30 January, 1934.

Representing family relationships. Family relationships are represented in CPM using the `cpm:Family` class. Instances of the class `cpm:Person` are then related to an instance of the `cpm:Family` class using one of the following properties: `cpm:familyChild`, `cpm:familyAdoptiveChild`, `cpm:familyFosterChild`, `cpm:familyParent`, `cpm:familyCohabitant`. Genealogy is a subfield of prosopographical research. Within this popular area of research some large database we already developed. A popular vocabulary for representing genealogical information is GEDCOM[20], which is also based on the family concept as central information asset. For evolutionary reasons, we still also included the properties `is-related-to` and `cpm:relationship`, which allow to model a family relationship more directly.

5 Knowledge Engineering Methodology

In order to describe the knowledge engineering methodology followed in CPL we statistically analyzed a number of usage indicators (cf. Figure 4). Editing activities using OntoWiki in the restricted zone was logged since September 2008. Statistics about accessing data via the user interface of the CPL frontend was logged since its launch in April 2009. This statistics does, however, not include any access information from the linked data interface.

It can be clearly seen, that most of the editing activity were additions of statements. In the learning phase (i.e. first four months of the project from September 2008 till January 2009), there was still notable but decreasing number of statement deletions and property changes, which indicate corrections and an increasing familiarity of the domain experts with the system. Till June 2009 - the month of the official public announcement of CPL - the activities intensified, with regard to added statements, professors and changed properties. Due to feedback from other historians working in the field and the general public, the editing activity remained relatively high after the announcement, but decreased slowly. The number of editors (i.e. historian domain experts) ranged between 5 and 10.

For the development of the CPM we have chosen a very pragmatic approach. The development of a first version of the ontology was simplified due to the availability of

[19] `http://www.w3.org/2003/01/geo/wgs84_pos#`

[20] Genealogical Data Communication, The GEDCOM standard release 5.5, 1996.

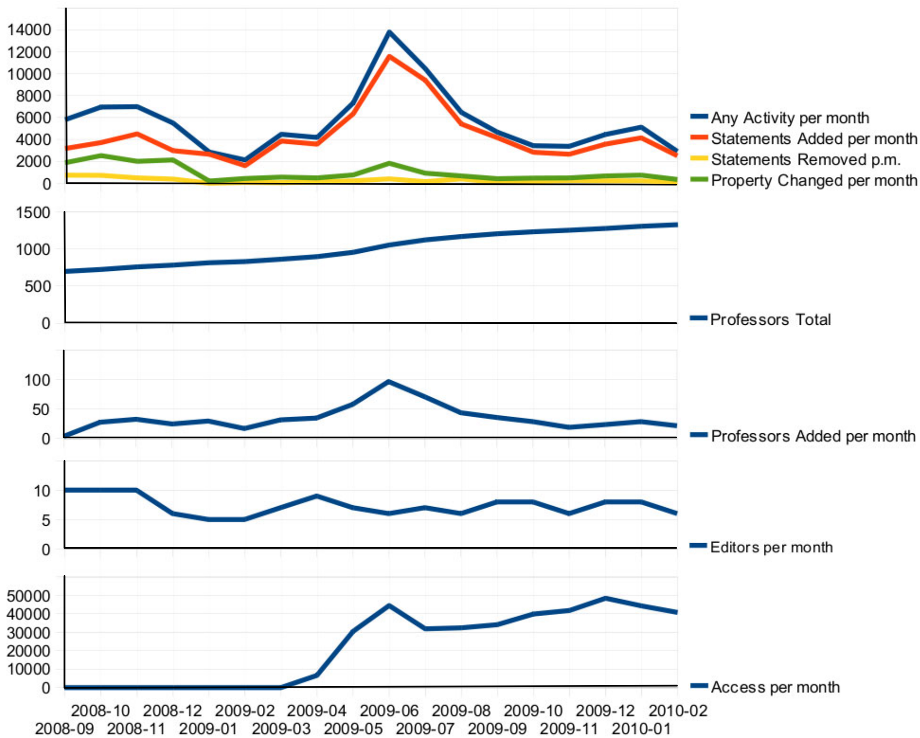

Fig. 4. Statistics about various CPL usage indicators

legacy data. The idea of using the FOAF and other existing vocabularies turned out inappropriate. The main reason for this being the absence of a precise understanding of the semantics to be represented within the catalog. Discussions between ontology engineers and historians about the overall ontology structure turned out to be very intricate due to the lack of understanding of the respective field of the other party. A solution to this dilemma was not to focus on the general ontology structure, but rather on small concrete representation issues and facilitate the evolution of the CPM and the CPL in an ontology/application co-design process. Overall, the engineering co-design methodology of the CPL can be characterized so far by 6 phases, which we describe in the sequel.

Before CPL. In 2006 the project was started using an single data table. This table was edited asynchronously by the project team. Since the database table become more complex and the number of needed columns reaches the column limit of 255 the project team requested technical support from database experts.

(1) Information analysis. Based on the existing database table, an analysis and remodeling of the database had been done. The resulting entity-relationship (ER) model was discussed and improved iteratively by the team of historians and knowledge

engineers for about half a year. Modelling this database structure became increasingly complex and discussions about entities and relationships were soaring. This has been the major argument to introduce the agile knowledge engineering method RapidOWL [2] based on wiki technologies and semantic knowledge representation. The existing ER schema was transformed into an preliminary RDF/RDF-Schema/OWL ontology.

(2) Initialization. In September 2008, OntoWiki has been deployed for the project and initialized with the preliminary version of the CPM. The existing data was transformed and imported from the database table into OntoWiki. The CPL starts with information about 700 professors.

(3) Wiki-based knowledge acquisition and ontology refinement. During the acquisition of new instance data with the now OntoWiki based CPL the project team detect limitations of the initial CPM. These were discussed within the project team. Advantages and disadvantages of different modeling approaches were presented by the Semantic Web experts to the historians and resulted in an substantially extended and restructured ontology.

(4) Publication of the catalog. As shown in the diagrams in Figure 4 the publication of the catalog data also results intensified knowledge acquisition and engineering activities. This is caused mainly by the feedback of the historian community and web users, but also result in a motivation boost for the historians working on CPL. The possibilities of accessing data has been permanently improved, e.g. by launching the experimental OntoWiki server (cf. Figure 1).

(5) Interlinking other datasets. The last phase of CPL so far was is the interlinking of the catalog with other databases. This will enhance working with the wiki and querying the data even more. Interlinking was performed with DBpedia [6], the German National Library using the unique PND [8] identifier and with the catalog of lecture directories[21] of the University of Leipzig.

(6) Alignment to other ontologies. As a result of the phases 4 and 5 the CPL gained quite much attention within the research community. As a consequence the interlinking with other knowledge bases became more important and there was a strong feedback with regard to aligning the CPM with other prosopographical knowledge bases.

6 Exploration and Access Interfaces

In order to facilitate the interaction of domain experts and interested people, CPL is accompanied with a number interfaces for accessing and exploring the information. In this section we briefly showcase the generic access interfaces Linked Data, Visual Query Builder and Relationship Finder, as well as the specifically developed public CPL website.

[21] `http://histvv.uni-leipzig.de`

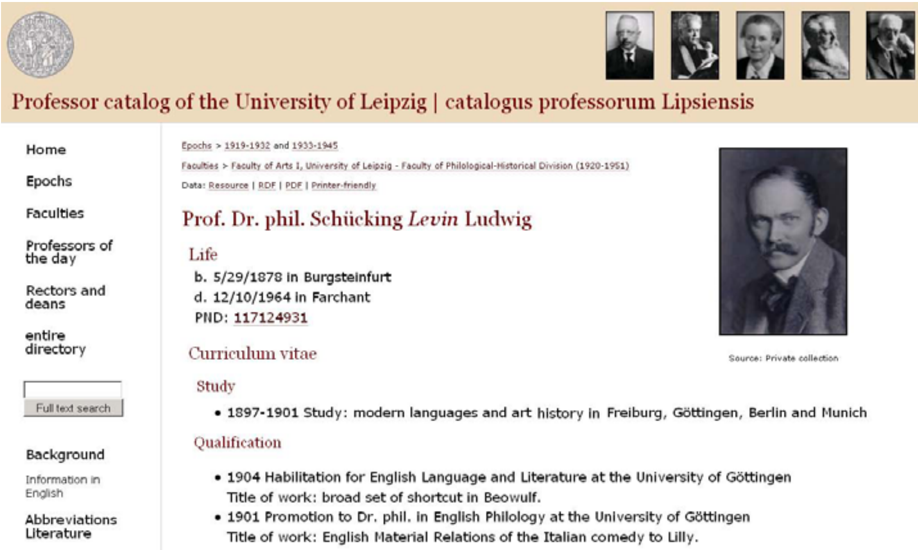

Fig. 5. Visualization of the resource representing "Prof. Dr. phil. Levin Ludwig Schücking"' on the public CPM website

Public Website. CPL is not just a tool for historians, but aims to showcase the results of historic research to the wider general public. For that purpose a special public website was created (as shown in Figure 5). The user interface of the public website is geared towards simplicity. The knowledge base can be explored by epochs, faculties, functions of professors (i.e. rector or dean) or alphabetically. Professors of the day are automatically selected based on important days in the life of a professor (i.e. birth or death). Furthermore, the public website comprises a full-text search, which searches within all literals stored in the CPL knowledge base.

Linked Data. The term Linked Data here refers to a set of best practices for publishing and connecting structured data on the Web[22]. These best practices have been adopted by an increasing number of data providers over the past three years, leading to the creation of a global data space that contains many billions of assertions Using OntoWiki's build-in endpoint functionality CPL is immediately available as Linked Data. Linked Data information is easy accessible e.g. using the Tabulator tool[23] [3]. Within the Linking Open Data effort, hundreds of data sets have already been connected to each other via `owl:sameAs` links. By interlinking CPL with other related datasets we aim at establishing CPL as a linked data crystallization point for academic prosopograpical knowledge.

---

[22] `http://www.w3.org/DesignIssues/LinkedData.html`

[23] Resource "Schücking" in Tabulator is available at: `http://dig.csail.mit.edu/2005/ajar/release/tabulator/0.8/tab.html?uri=http://catalogus-professorum.org/lipsiensis/Schuecking_144`

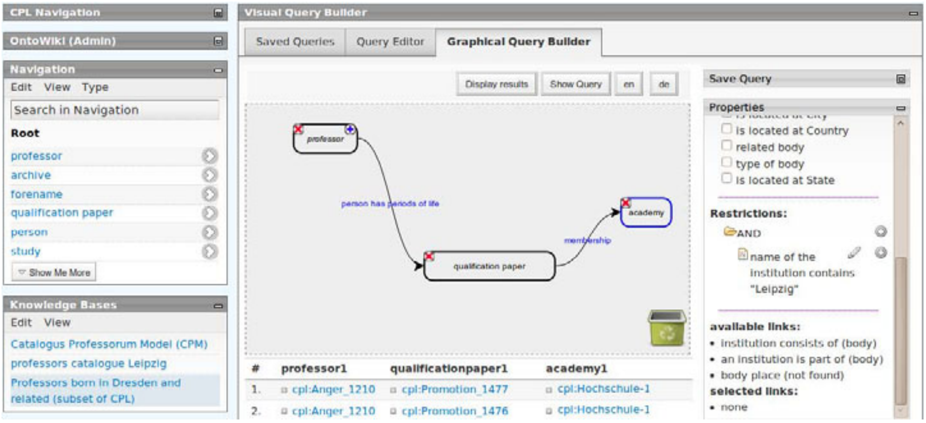

Fig. 6. *Visual Query Builder*

Visual Query Builder. OntoWiki also serves as a SPARQL endpoint, however, it quickly turned out that formulating SPARQL queries is too tedious for the historian domain experts. In order to simplify the creation of queries for the historians, we developed the *Visual Query Builder*[24] (VQB) as an OntoWiki extension, which is implemented in JavaScript and communicates with the triple store using the SPARQL language and protocol. VQB allows to visually create queries to the stored knowledge base and supports domain experts with an intuitive visual representation of query and data. Developed queries can be stored and added via drag-and-drop to the current query. This enables the reuse of existing queries as building blocks for more complex ones. VQB also supports the *set-based browsing* paradigm by visualizing different connectives, such as join, union, intersection, difference between queries. The incremental query building is facilitated by displaying results already during query creation. The VQB user interface is visualized in Figure 6. It consists of 5 panels, which visualize different aspects of the query creation:

– *Center panel:* the main workspace, where the query is visualized as a graph. Individual elements of the graph can be selected, deleted or moved on the canvas.
– *East panel:* displays information about the currently selected element. If the selected element is, for example, a class this panel contains the properties, which are used with instances of the class. The panel also contains controls for changing the query, such as via filter conditions.
– *West panel:* contains available classes in a tree display. Queries are grouped with the classes contained in the query.
– *South panel:* visualizes the result of the query.
– *North panel:* shows notifications such as usage hints, error or event descriptions and the generated SPARQL query.

The user interface can be adjusted by scaling or deactivating unused panels.

[24] `http://aksw.org/Projects/OntoWiki/Extension/VQB`

Relationship Finder. An important aspect of historical investigations is the search for relationships between different persons or entities of interest. An application supporting such investigations within RDF knowledge knowledge bases is *RelFinder*[25] [4]. With the help of *RelFinders* relationships between individual entities can be easily discovered and visualized. Figure 7, for example, visualizes the relationship between the entities "Schücking" and "München"[26]. In this example, three connections were found and visualized as paths through the knowledge base. *RelFinder* is a generic tool and can be used in conjunction with arbitrary SPARQL endpoints.

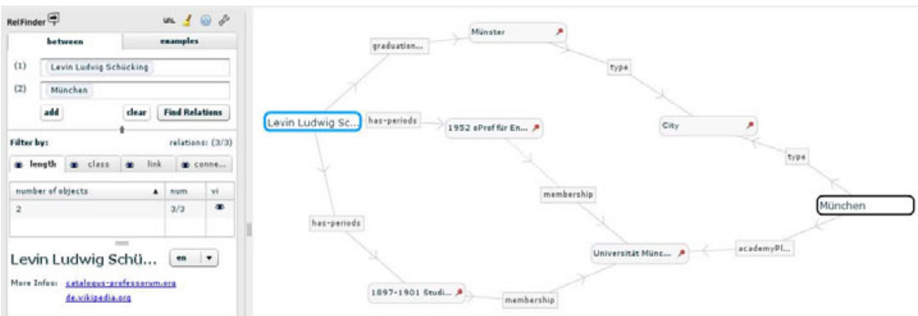

Fig. 7. Visualization of relationships in *RelFinder*

7 Prosopographical Use Cases

In this section we introduce some important prosopographical research use cases, which are facilitated by the ontological knowledge representation and the various exploration and access interfaces. These include in particular historical social network analysis, academic self-complementation analysis and the relationship between religion and university.

Historical Social Network Analysis. The analysis to what extend a certain professor influenced his students, colleagues and friends was previously only possible with a hardly justifiable manual effort. Although the CPL comprises primarily professors working in Leipzig, it reaches far beyond the limits of the Saxon state university, since all periods of life of a professor are included. Due to the semantic representation, it is easily possible to identify direct as well as indirect connections between individuals, such as, for example, an overlap in certain periods of life (such as a common school or university visit). Also, the detailed representation of qualification steps (such as doctoral and post-doctoral studies) facilitates the reconstruction of teacher-student relationships and thus the identification of certain schools of thought and on a more general level the establishment *academic genealogy*.

[25] Online at: `http://relfinder.semanticweb.org`

[26] More interesting relationships obtained from RelFinder are listed at: `http://catalogus-professorum.org/tools/relfinder`

Academic Self-Complementation Analysis. A crucial aspect of academic historic research is the analysis of the self-complementation functions of the different faculties. In particular cousin-hood was an important factor for chair appointments in German universities of the early modern period (i.e. from approx. AD 1500 to around AD 1800), which thus were heavily influenced by scholar dynasties. In the 19th century this practice changed dramatically, but still intellectual families aimed at preserving their social prestige by fostering the academic careers of their sons. CPL allows far reaching qualitative and quantitative research due to the fine-grained representation of family relationships. An interesting observation, for example, revealed by a query to CPL was that a common pattern of scholar biographies in the 19th century was the marrying of the daughter of ones academic advisor. Regarding academic self-complementation, CPL also allows to easily observe the popularity of faculty internal appointments during different periods of time: while very common in the early modern period it became much less popular after the Humboltanian reforms in the 1820s, but gained again popularity in socialist East Germany. Another interesting area of research is the political influence on appointments, which is facilitated, for example, by the inclusion of party memberships in CPL. The allows, in particular, to determine whether the membership in a certain political party statistically significantly affects ones career in a certain historic period.

Relationship between Religion and University. Religion was not only a founding factor of many European universities, but influenced academic live in one way or the other throughout the centuries. After the *Confessio Augustana* in the year 1543, Lutheranism was the obligatory confession at Universität Leipzig. Only in the 19th century it became possible to admit catholic members to the faculties and it is an interesting research question, facilitated by CPL, to investigate to what extend and in which fields this was actually the case. During the Third Reich period the destiny of Jewish professors is in the center of interest and later in, Eastern Germany, the non-confessionalism or reserves of religiosity in certain regime-distant fields is in the focus of interest.

8 Lessons Learned, Conclusions, Related and Future Work

Lessons Learned. A lesson we learned is, that such a project involving people with very different backgrounds and with very limited resources requires to establish a working *knowledge base / application co-design*, where both - knowledge bases and knowledge-based applications - are iteratively refined. Given the timely visibility of the knowledge base for a wider community, additional refinements are triggered by the interaction with the community. As we experienced with the historian domain expert team, the motivation boast due to the early public availability of the knowledge base project and the direct visibility of improvements and corrections can hardly be underestimated and by far outweighs initial maturity gains with longer development cycles. A growing added value for domain experts is the availability of background knowledge on the linked Data Web. CPL is one of the first prosopographical knowledge bases on the data web, but still the interlinking and fusing of information, for example, from DBpedia and Geonames is of great use for enhanced querying and exploration of the information. As more prosopographical knowledge bases become available this effect will be even more amplified, as we are currently already experiencing with applications of the CPL infrastructure for other prosopographical use cases.

Conclusions. With CPL we demonstrated a successful application of semantic knowledge representation techniques and an agile collaboration methodology in social sciences. For historians the followed knowledge base approach resulted in completely new research opportunities, compared to the book/lexicon based methodologies prevalent in prosopographical research. The use of knowledge bases and agile, web-based collaboration has the potential to trigger a paradigm shift in historic research: from individual centered research aiming to solve a specific research task towards collaborative research, which's results can be re-purposed in order to answer unanticipated research questions.

Related Work. *SemanticWeb for History*[27] (SWHi) is a project which was carried out from 2006 to 2008 at the University of Groningen in the Netherlands had the aim to combine available vocabularies to be able to model the early American history. A semantic browser[28] for this data has been developed. The vocabulary[29] developed during this project is not published so far. The *Historical Event Markup and Linking Project*[30] (HEML) primarily aims at representing early Greek history. The developed vocabulary [9] is based on events and periods. Central elements of the vocabulary[31] are the classes `heml:Event`, `heml:Person`, `heml:Role` and `heml:Evidence`.

Beyond history, wiki-based knowledge engineering has been successfully applied to other knowledge domains before (e.g. Brede Wiki for Neuroscience data [7] or the Modelling wiKi MoKi [10]).

Future Work. One hotspot of future work will be the expansion of the usage of the Catalogus Professorum. Currently, we are planning to extend the catalog to include the universities Munich and Berlin, which for a long time represented the academic center together with Universität Leipzig. In addition there are cooperations with the universities Utrecht and Zürich and a number of other prosopographical databases are currently investigating how to interlink with the CPL and the Linked Data Web in general. In addition we aim to adopt the developed techniques in order to support other historic knowledge engineering projects beyond prosopographical databases. Also, as for the Data Web in general we aim to increase the coherence by tighter interlinking of CPL with related and complementary knowledge on the Data Web.

Acknowledgments

We would like to thank our colleague Christian Augustin and all of the young researchers, involved in archival work and interviews, for their helpful comments and

[27] http://americanimprints.ub.rug.nl/
[28] http://semweb.ub.rug.nl/
[29] http://semweb.ub.rug.nl/swhi#
[30] http://heml.mta.ca/
[31] http://www.heml.org/rdf/2003-09-17/heml/

inspiring discussions during the development of CPL. This work was supported by grants from the German Federal Ministry of Education and Research (BMBF), provided for the project LE4SW (project number 03WKP02A) and from the European Union's 7th Framework Programme provided for the project LOD2 (GA no. 257943).

References

1. Auer, S., Dietzold, S., Riechert, T.: OntoWiki – A Tool for Social, Semantic Collaboration. In: Cruz, I., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P., Uschold, M., Aroyo, L.M. (eds.) ISWC 2006. LNCS, vol. 4273, pp. 736–749. Springer, Heidelberg (2006)
2. Auer, S.: The RapidOWL Methodology–Towards agile knowledge engineering. In: Proceedings of the 15th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises, pp. 352–357. IEEE Computer Society, Los Alamitos (2006)
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: Procedings of the 3rd International Semantic Web User Interaction Workshop (SWUI 2006), p. 6 (2006)
4. Heim, P., Hellmann, S., Lehmann, J., Lohmann, S., Stegemann, T.: RelFinder: revealing relationships in RDF knowledge bases. In: Chua, T.-S., Kompatsiaris, Y., Mérialdo, B., Haas, W., Thallinger, G., Bailer, W. (eds.) SAMT 2009. LNCS, vol. 5887, pp. 182–187. Springer, Heidelberg (2009)
5. Heino, N., Dietzold, S., Martin, M., Auer, S.: Developing semantic web applications with the ontowiki framework. In: Networked Knowledge - Networked Media. SCI, vol. 221, pp. 61–77. Springer, Heidelberg (2009)
6. Lehmann, J., Bizer, C., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., Hellmann, S.: DBpedia - a crystallization point for the web of data. Journal of Web Semantics 7(3), 154–165 (2009)
7. Nielsen, F.Å.: Brede wiki: Neuroscience data structured in a wiki. In: 4th Semantic Wiki Workshop (SemWiki 2009) at ESWC 2009. Proceedings. CEUR WS, vol. 464 (2009)
8. Pfeifer, B., Senftleben, S.: Die Personennamendatei (PND). Leipziger Beiträge zur Informatik, vol. Band XXI, pp. 137–144. LIV (2010)
9. Robertson, B.: Exploring historical rdf with heml. Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure 3(1) (Winter 2009)
10. Rospocher, M., Ghidini, C., Pammer, V., Serafini, L., Lindstaedt, S.N.: Moki: the modelling wiki. In: 4th Semantic Wiki Workshop (SemWiki 2009) at the 6th European Semantic Web Conference (ESWC 2009), Hersonissos, Greece. CEUR Workshop Proceedings, vol. 464 (June 1, 2009)