Querying Prosopographic Datasets: a Case Study on **Ancient Near Eastern Persons in Cuneiform Sources**

Rossana Damiano^{1,*,†}, Claudio Mattutino^{1,†}, Claudia Finocchiaro^{3,†} and Elena Devecchi^{2,†}

¹Dipartimento di Informatica, Università di Torino ²Dipartimento di Studi Storici, Università di Torino

Abstract

The advent of Linked Data and of RDF-based representations in the Digital Humanities has brought unprecedented benefits to the collaborative, large scale collection and study of data. However, the models and formats that characterize Linked Data projects can hinder the easy of access to research data by domain experts. In this paper we describe the design and implementation of an interface for querying a prosopographic dataset encoded in a knowledge graph. The interface, designed in cooperation with the domain experts, has been developed according to a test-and-redesign approach to guarantee their capability to afford an easy to use, effective access to data.

Keywords

Knowledge graphs, Prosopography, Cuneiform documents, Access tools

1. Introduction

More than two decades ago, Semantic Web technologies [1] have brought to Digital Humanities an interlingua that has replaced project-specific database schemes with interoperable, publicly available computational ontologies that support scholars in their workflows [2, 3, 4]. The advent of Linked Data [5], then, has enabled the publication and sharing of the data as "Knowledge Graphs" (KG) where the entities of a domain are described through their interconnections in network-like representations that naturally lends themselves to the study and exploration of the relationships between the entities by human experts and machines [6, 7, 8].

In this paper, we describe the design and implementation of an interface for querying a prosopographic dataset sourced from cuneiform documents and encoded in a knowledge graph. According to Lawrence and Bodard "Prosopography, as distinct from biography, focuses not on the life of an individual but on a person as part of a collection with whom they may share, or not, a number of traits." [9]. Such traits define the social identities of individuals, such as names and titles, but also their network of relations within the socio-economic and cultural context they lived in.



Workshop on Advanced Visual Interfaces and Interactions in Cultural Heritage (AVICH 2024), June 4th, 2024, co-located with the 17th ACM Conference on Advanced Visual Interfaces, Arenzano (Genoa), Italy.

^{*}Corresponding author.

[†]These authors contributed equally.

rossana.damiano@unito.it (R. Damiano); claudio.mattutino@unito.it (C. Mattutino);

finocchiaroclaudia15@gmail.com (C. Finocchiaro); elena.devecchi@unito.it (E. Devecchi)

D 0000-0001-9866-2843 (R. Damiano); 0000-0002-0413-2436 (C. Mattutino)

^{© 2022} Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

The paradigm of Linked Data, mostly suitable to collaborative, distributed research initiatives, is particularly beneficial to prosopographic data, thanks to its capability of encoding relations over people, locations, and other domain entities (staples, offices, etc.) in a straightforward fashion. However, the Linked Data model and format cannot be directly exposed to the domain experts, despite the capability of the KG of being accessed through semantic queries that mirror very closely the research questions expressed by them. To bridge the gap between the KG and the information needs of the domain experts, we have designed and implemented a set of search forms for querying the KG in a structured way without knowing the query language for RDF, namely SPARQL¹. Initially developed from a initial set of research questions put forth by the domain experts [7], the forms have been subsequently refined through the feedback provided by the experts after using them to explore the KG, then redesigned to improve their usability.

This paper is structured as follows: after briefly reviewing the state of the art in Section 2, we describe the data modelling and access in Section 3. Section 4 describes the design considerations gathered after the testing of the interface by the domain experts and the actual design. Conclusion and Future Work end the paper.

2. Related Work

The representation of prosopographic data with computational ontologies relies on the pioneering work carried out by Pasin and Bradley [6]. The Factoid-based Prosopographic Ontology (FPO) licensed by Pasin and Bradley revolves around the notion of *factoid*, namely a believedtobe-true event reported in some written source. This definition fits very precisely the data – inferred from text corpora – investigated by our case study. The Factoid model puts into play two basic entities: the Source where the factoid is asserted (here, a cuneiform text), and the Relation it describes (e.g., an administrative or legal relation), which refers to a given Person (Person Reference) and Location (Location Reference). The notion of factoid itself, in fact, matches well the attribution of a specific relationship to specific entities decoded from a source without blurring the distinction between the source and the relationship, intended as a conceptual entity inferred from it by the expert. The representation of factoids, individuals and sources in the case study relies on FPO, which has been extended to represent the peculiar aspects of the case study.

Following the paradigm of Linked Data, according to multiple models (and vocabularies) can be combine to integrate information of different type and source, a set of vocabularies are relevant to the representation of prosopographic data: the Friend Of A Friend vocabulary (FOAF) for the social identity of the individuals, the Bibliographic Ontology vocabulary (BIBO) for the relations with bibliographic sources, and the Dublin Core Metadata Element Set (DC-MES) for the provenance and description of data.

3. The Prosophographic Data Set

The KG and the access interface have been developed as part of a research project entitled "Networks of Power: Institutional Hierarchies and State Management in Late Bronze Age

¹https://www.w3.org/TR/rdf-sparql-query/

Western Asia (NePo)", funded by the Italian Ministry of Research. The project aims at a detailed analysis of court structures of Near Eastern kingdoms in Late Bronze Age (LBA) [10], their internal networks, and the economic systems they were controlling, on the basis of selected epigraphic and archaeological sources, with the ultimate goal of obtaining a comparative picture of the royal elites of LBA Western Asia through the analysis of multiple sources from the geographical areas under consideration.

3.1. Data Modelling

The design and testing of the form-based access interface concerns in particular in particular the data extracted form a corpus of cuneiform tablets from Babylonia dating to the Kassite period. Although Kassite material mostly comes from the city of Nippur, the case study focuses on a corpus of ca. 800 cuneiform tablets published by [11, 12] and dating to the 14th-13th centuries BC; their exact provenance is not certain, but they likely originated from an administrative center of the Nippur area. These documents currently represent the second largest, internally coherent set of Kassite-period sources after that of Nippur. They are administrative records mainly dealing with the income, storage, and redistribution of agricultural products (mostly cereals, but also sesame, pulses, and cress) and by-products (beer and flour), animal husbandry, and textile production; smaller groups of texts include legal documents and letters.

Starting from FPO, classes and properties have been specialized to adapt them to Kassite domain, yielding the Kassite FPO, or KFPO, by adding more specific classes and properties to the ones in the ontology, leaving the the core model unaffected (see [7] for details). In particular, Cuneiform texts and Seals have been introduced as specific source types; specific factoid types have been added to account for the information conveyed by the cuneiform texts in the domain, which concern Administrative, Kinship, and Professional relationships between the personages in the Kassite world; finally, new authoritative lists have been created to fit the representation of sources and personages, where dates refer to sovereigns, and people are described according to culture-specific terms for age and gender [13]. Due to the specificity of the Kassite domain, only a partial alignment has been accomplished so far with other resources: currently, the alignment is limited to a subset of the geographical entities in the Pleiades repository of ancient places ².

The prosopographic knowledge graph, stored into an installation of Apache Fuseki triple store ³, is a self-contained representation where each individual entity such as a specific person or text becomes a node connected to other entities through links labelled after the fields in the original entity description. For example, a given factoid will be *sourced from* a given source; a given person will be in a *co-worker relation* with another person. In some cases, an entity is directly linked to data instead of other entities: for example, a cuneiform text can be linked to a given year (a number) and king (a string), a person can be linked to its patronymic (a string) or title (a string). In this case, the knowledge graph will include the information about the data type (whether it is a number, string or a term from a controlled vocabulary).

²https://pleiades.stoa.org/

³https://jena.apache.org/documentation/fuseki2/

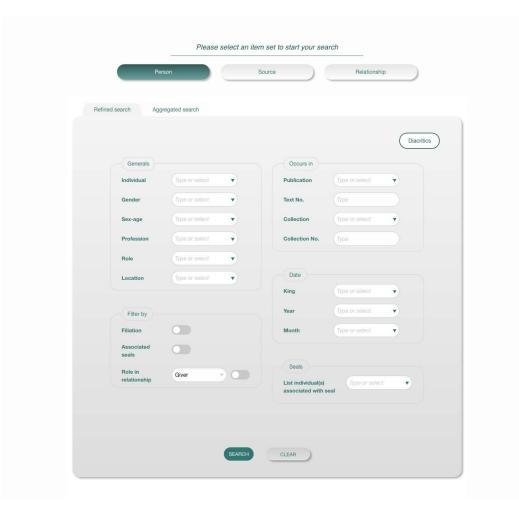


Figure 1: The Person search form.

3.2. Querying the data

Differently from the query languages for relational databases, SPARQL queries are expressed as patterns on data, making them very similar to the expression of the corresponding questions in natural language. For example, relationships can be queried through a pattern of the type "Find the persons who are in a certain type of relation with a given person".

To bypass the limitations posed by the KG model and format, for each entity type (Person, Source and Relationship) a set of research questions were collected and factorized into an exhaustive set of search parameters. Each parameter was translated in SPARQL to enable the on-the-fly creation of complete SPARQL queries that match the search parameters selected by the expert through the interface without requiring any knowledge of the language. Each query is then executed on the knowledge graph on the Fuseki server. Specific query forms have been created for the other relevant entities of the domains, such as the sources and the relations between the persons (Fig. 1 shows the form for searching persons).

Person	Source		Relationship
Refined search Aggregate	d search		
			Diacritics
Generals			
	Individual	Type or select	
	Kind of relationship	Administrative	
	Туре	ina qāt (from [the hanc 🔻	
	Object of transaction	Barley (ŠE, ŠE.BAR)	
Paired with			
	Relationships of type "paired with" arranged according to line number that occurs in text	BE 14, 11	
	SEARCH	CLEAR	
	SEARCH	CLEAR	
Showing 60 result(s)	SEARCH	CLEAR	
Showing 60 result(s) nameFactoid	SEARCH		objectOfTransaction
			objectOffransaction Barley (ŠE, ŠE.BAR)
nameFactoid	ni 1		

Figure 2: The Refined Search tab of the relationship search form.

PREFIX dcterms: <http: dc="" purl.org="" terms=""></http:>
PREFIX xsd: <http: 2001="" www.w3.org="" xmlschema#=""></http:>
PREFIX o: http://omeka.org/s/vocabs/o#>
select distinct ?IDFactoid ?nameFactoid ?objectOfTransaction
where{
?linkName a <kfpo:kassiteperson>.</kfpo:kassiteperson>
bind(?name as ?name)
?linkName dcterms:title ?name.
?linkName ?relation ?linkFactoid.
values ?relation { <kfpo:hasadministrativereference>}.</kfpo:hasadministrativereference>
?linkFactoid dcterms:title ?nameFactoid.
?linkFactoid o:id ?IDFactoid.
bind("ina āqt (from [the hand of])" as ?type)
?linkFactoid <kfpo:administrativerelationtype> ?type.</kfpo:administrativerelationtype>
bind("Barley Š(E, ŠE.BAR)" as ?objectOfTransaction)
<pre>?linkFactoid <kfpo:objectoftransaction> ?objectOfTransaction.}</kfpo:objectoftransaction></pre>

Listing 1: The SPARQL query matching the parameters in Fig. 2

For example, Figure 2 shows the results (for brevity, only a few lines of the 60 results are

listed) of a query on Administrative relationships that have barley as object and belong to a specific type of transaction acknowledged by the Kassite administrative system ("ina qāt", which means "from [the hand of]"). For each relationship found, the link to the record describing the corresponding transaction is provided. As shown in Listing 1, each selected parameter is translated into a specific element of the SPARQL query, using the terms from the KFPO ontology: the predicate kfpo:hasAdministrativeReference refers to administrative relations, the "Barley (ŠE, ŠE.BAR)" literal refers to the object of the transaction (a term from the controlled vocabulary for describing the objects of the transactions), the "ina qāt (from [the hand of])" literal refers to the specific type of administrative relation). Figure 3 shows the record corresponding to the second the line of the results, namely the factoid termed as "Sîn-balața-īriš receives from Mudammiq-Adad L", on the website where the dataset is published in textual form.

4. Interface design and evaluation

To allow scholars to test the advanced search interface, a preliminary form was created, shown in Figure 4. The form had a tripartite structure, with tabs for persons, sources and relationships. However, this preliminary interface, used by the domain experts for a period of six months, suffered from several usability issues, ranging from an excessively large number of parameters to the misplacement of the search buttons. Moreover, each form contained both the retrieval and the counting functions (implemented by using the SPARQL COUNT construct), needed to estimate the quantitative impact of the different entity features, resulting in the difficulty by the domain experts to select the appropriate function in a simple way. To overcome these issues, the interface was redesigned as follows:

• For each entity type, the retrieval and search functions were split into two different tabs ("Refined search" and "Aggregated search"), so that it is always clear which function is selected. Figure 5 shows the Aggregated search interface for persons.

Sîn-balața-īriš receives from Mudammiq-Adad L 📼



Figure 3: A record exemplifying the textual representation of a relationship factoid (second line of the search results in Fig. 2).

Kass	ite Prosopographic Records		SOURCE	RELATIONSHIP
rm + Diacritics				
Name:	8	Total number of individuals grouped by gender:		Diacritics
Gender:	0	Total number of individuals who have a specific sex-age designation:		ĀāÂâ
Sex-age:	8	Total number of individuals practicing a specific profession:		ĒēÊê
Profession:	8	Total number of individuals playing a specific role:		ΙτÎî
Role:	8	Total number of individuals associated with a specific location:		Ũ ū Û û
Location:	•	Total number of individuals who act as recipients:		ŠšŞş
Patronymic:				T t H b
Ethnonym:				
List associated seals:				
Person associated with:				
	Publication Text No. Collection Collection No.			
Occurs in:				
	King Year Month			
Date:				

Figure 4: The preliminary interface. The screenshot shows the menus for selecting the search parameters for persons such as Name, Gender or Role (left column), and the checkboxes for counting the number of occurrences found (right column) according to different grouping criteria.

- To help the user find their way around the number of parameters, the latter were grouped into subgroups, as exemplified in Figure 6 where search parameters are grouped into meaningful subgroups such as "Text", "Genre", "Seals", "Earliest date" and "Latest Date".
- To free up space in the interface, the "Diacritics" box, needed to write the characters that appear in the transliterated cuneiform texts, was removed and replaced by a button that makes it visible upon request.

Refined search	Aggregated search	
	Total number of individuals grouped by gender	
	Total number of individuals who have a specific sex-age designation	
	Total number of individuals practicing a specific profession	
	Total number of individuals playing a specific role	
	Total number of individuals associated with a specific location	
	SEARCH CLEAR	

Figure 5: The aggregated person search form, with radio buttons for specific grouping criteria.

Text			Earliest date		
Publication	Type or select		King	Type or select	•
Text No.	Туре		Year	Type or select	
Collection	Type or select	•	Month	Type or select	
Collection No.	Туре				
Provenience	Type or select	•	Latest date (2)		
Findspot	Туре		King	Type or select	•
			Year	Type or select	•
Genre			Month	Type or select	•
Genre	Type or select	•			
Subgenre	Type or select	•	Seals		
Subject matter	Type or select	•	Sealed by	Type or select	•
			Sealed with	Type or select	•
			Type of sealing	Type or select	•

Figure 6: The source search form. To reduce the number of high-level elements displayed to the user, related search parameters are grouped into boxes and labeled according to their type ("Text", "Genre", "Seals", "Earliest date" and "Latest Date").

The graphics was redesigned using the Foundation framework employed to manage the style at the site level⁴. The new search interface has been made available to a group of domain experts for validation. At the end of this phase we plan a new iteration of minor modifications.

5. Conclusion and Future Work

The advent of Linked Data and of RDF-based representations in the Digital Humanities has brought undoubted benefits to the collection and study of data. However, the tools for accessing RDF representations introduce an interpretation gap between data and experts, setting the need for access tools that lift the latter from the need to write complex queries in formal languages. In this paper, we presented a set of access forms for a prosopographic data set aimed at bypassing the complexity of writing queries but maintaining the flexibility of semantic queries. The forms, designed in cooperation with the domain experts, have been developed according to a test-and-redesign approach to guarantee their capability to afford an easy and effective access to research data.

As future work, we intend to make the search interface publicly available to receive feedback by a larger community of users and study the most common searches performed by the users for optimization purposes. Moreover, we plan to adapt the forms to similar data sets.

⁴https://get.foundation/

References

- O. Lassila, J. Hendler, T. Berners-Lee, The semantic web, Scientific American 284 (2001) 34–43.
- [2] G. Bruseker, N. Carboni, A. Guillem, Cultural heritage data management: the role of formal ontology and cidoc crm, Heritage and archaeology in the digital age: acquisition, curation, and dissemination of spatial cultural heritage data (2017) 93–131.
- [3] V. A. Carriero, A. Gangemi, M. L. Mancinelli, L. Marinucci, A. G. Nuzzolese, V. Presutti, C. Veninata, Arco: The italian cultural heritage knowledge graph, in: The Semantic Web–ISWC 2019: 18th International Semantic Web Conference, Auckland, New Zealand, October 26–30, 2019, Proceedings, Part II 18, Springer, 2019, pp. 36–52.
- [4] V. Lombardo, R. Damiano, T. Karatas, C. Mattutino, Linking ontological classes and archaeological forms, in: International Semantic Web Conference, Springer, 2020, pp. 700–715.
- [5] C. Bizer, T. Heath, T. Berners-Lee, Linked data: Principles and state of the art, in: World wide web conference, volume 1, Citeseer, 2008, p. 40.
- [6] M. Pasin, J. Bradley, Factoid-based prosopography and computer ontologies: towards an integrated approach, Digital Scholarship in the Humanities 30 (2015) 86–97.
- [7] A. Urbinati, E. Burdisso, C. Mattutino, S. Vilella, A. Semeraro, G. Ruffo, C. Corti, S. DE MAR-TINO, E. Devecchi, E. Scarpa, et al., Bridging representation and visualization in prosopographic research: A case study, in: CEUR WORKSHOP PROCEEDINGS, Sun SITE Central Europe, Technical University of Aachen, 2022, pp. 80–92.
- [8] C. Waerzeggers, Social network analysis of cuneiform archives: A new approach, Documentary Sources in Ancient Near Eastern and Greco-Roman Economic History: Methodology and Practice (2014) 207–233.
- [9] K. F. Lawrence, G. Bodard, Prosopography is greek for facebook: The snap: Drgn project, in: Proceedings of the ACM Web Science Conference, 2015, pp. 1–2.
- [10] E. Devecchi, E. Scarpa, Società, amministrazione ed economia nella babilonia di età cassita: un progetto di digital prosopography, Scienze dell'Antichità 29 (2023) 107–120.
- [11] E. Devecchi, Middle Babylonian Texts in the Cornell Collections, Part 2, Penn State University Press, 2020.
- [12] W. H. Van Soldt, Middle Babylonian Texts in the Cornell University Collections. Part I, CDL Press, 2015.
- [13] R. Damiano, Design and development of knowledge graphs for hittite and kassite prosopographic data, Mesopotamia 58 (2023) 87–91.