Proposal Title

*Horizons of interest: the geographical scope of knowledge in early modern chronicles (1500-1850)*

Applicants

<table>
<thead>
<tr>
<th>Supervisor Name</th>
<th>Department/Group</th>
<th>Faculty</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Dr. Erika Kuijpers</td>
<td>History</td>
<td>FGW</td>
</tr>
<tr>
<td>2. Dr. Ronald Siebes</td>
<td>Computer Science</td>
<td>FEW</td>
</tr>
</tbody>
</table>

Project description

Scholars agree that cultural changes in early modern Europe (c. 1500-1800) were both accompanied and precipitated by an information revolution. The use of printed media filtered down into local chronicles. These are hand-written narratives produced usually by middle class authors, that recorded events and phenomena they considered important (local politics, upheavals, climate, prices, crime, deaths, but also supra local and international news). During the seventeenth and eighteenth century, these chronicles came to include news and topics deriving from a greater variety of information sources and, presumably, from an increasingly global information network. The question this project wishes to answer is in what way the geographical scope of the chroniclers’ access to information and news changed in the period 1500-1850. What places are mentioned by these authors in their accounts? How did the ratios between local, supra local or regional and international news develop over time?
In order to answer these questions and to study the spatial dimensions of the chroniclers' worlds of interest, this project will apply Linked Data technology to map and compare the places, people, and events mentioned in these chronicles over time. As a starting point, we will use the corpus of the project Chronicling Novelty, which consists of about 300 digitized handwritten chronicles in Dutch from the period 1500-1850. Volunteers manually annotated the named entities in these texts as well as the date of entry. We will apply existing semantic tools to enrich the mentioned locations with identifiers from the Linked Open Data Cloud, and then query and analyze the data to answer the research questions mentioned above. We will publish the geodata and their attributions as Open Linked Data with the World Historical Gazetteer, which has recently become a partner of CLARIAH.

Project Organization

The Academy Assistant in the History Department will be responsible for the correct conceptual representation of the historical background and the involved sources, and for the design of the evaluation and the interpretation of the results. The Academy Assistant in the User Centric Data Science (UCDS) group will be focusing on the Linked Data technology and the formal modeling aspects.

Requirements:
The student in History should be interested in early modern history of the Low Countries, and must have an excellent understanding of (early modern) Dutch and affinity with the use of digital research tools and resources for the media history of the Low Countries.
The computer science student needs to be familiar with text processing technology and Linked Data and have demonstrated expertise in (Web) programming, both on the ‘backend’ (database technology, APIs and data manipulation) and the ‘frontend’ (HTML + JavaScript + GeoMaps).

Both students need good communication and analytic skills and should be able to work reliably and independently. They should show willingness to work in close cooperation and learn from each other.

Supervision: Ronald Siebes is a senior researcher in the UCDS group with expertise in Linked Data technology applied in various European and national projects. Erika Kuijpers is assistant professor in Cultural History and one of the two project leaders of the interdisciplinary NWO project Chronicling Novelty.

The team will also be advised by Tobias Kuhn (FEW), Rombert Stapel (IISG/CLARIAH+), and Alie Lassche, who is one of the two PhD candidates within the project Chronicling Novelty and whose research is focused on the changing mediascapes of early modern chroniclers.

The team will have weekly meetings in order to guarantee continuity and quick responses to eventual problems. We will encourage the students to regularly work together on campus.

Collaboration

This collaboration will open new research possibilities for both groups and a new strong connection between them. The History group will learn about the advantages of publishing their enriched data as LOD and will learn to query and analyse them. The User Centric Data Science group investigates from a computer science angle different aspects of socio-technical systems. This collaboration will look at such systems from the perspective of the history discipline, to study users in the contexts of such historical data collection endeavors, and to focus on this particular kind of rich data.

The academic challenge for the students is to be involved in a truly interdisciplinary project supervised and carried out by both historians and computer scientists. They will learn how to understand each other’s terminology, skills and questions. The history student shall develop some basic skills needed for data curation and the querying of linked data (e.g. using SPARQL or Open Refine). The CS student will be introduced to early modern chronicles, their contents and contexts, and learn about hermeneutics, data criticism, and the importance of provenance in historical research. While the CS student will design the methodology and architecture for the automated processes, the data model should be the result of close collaboration. The approach will make use of existing technology developed in related projects (e.g. within CLARIAH).

Our aim in involving students is not only to give them a taste of ongoing research, but also to make them acquainted with the practice of interdisciplinary research at the crossroads of history and computer science. Collaboration and frequent contact will be essential. The students can count on the supervision and collaboration of a research team that consists of both historians, computer scientists and computational linguists.
Deliverables

Several intermediate deliverables are planned as well as a final co-authored research paper.

- Deliverable 1a: Summary of literature on the topics of extracting persons, time periods and places for the chronicle entities (month 3)
- Deliverable 1b: Summary of literature (Status Quaestionis) on the practice of chronicling and the spatial dimensions of early modern information infrastructures (month 5)
- Deliverable 2: Short report on the chosen selection of technologies to build the data processing pipeline, including extraction, transformation, linking, visualization, and analysis steps (month 5)
- Deliverable 3: Enriched dataset published as LOD (month 10)
- Deliverable 4: Research paper on the overall approach and its results, ready to be submitted to a journal (month 10)

These deliverables will demonstrate the feasibility and value of our approach and result in rich annotations of the available corpus that can be processed efficiently by humans and machines.

All deliverables will be produced jointly. If the students would like to, we will help them expand their work into a (R)MA thesis, or help them reflect on a PhD proposal based on either the corpus or the used methods.

(181 words)

Planning

<table>
<thead>
<tr>
<th>October-December 2021</th>
<th>Assistant History: data curation: normalization of dates, explore the variety in historical toponyms as well as resources for disambiguation. Assistant UCDS: background research on Linked Data representation/mining techniques for historical data. Both: familiarize with the data; develop data model.</th>
</tr>
</thead>
<tbody>
<tr>
<td>January-February 2022</td>
<td>Assistant History: Read into the history of Media and information networks, infrastructure of power, commerce and communication. Write a Status Quaestionis. Disambiguation of place names. Assistant UCDS: define and set up a pipeline to transform existing XML data to Linked Data</td>
</tr>
<tr>
<td>March-May 2022</td>
<td>Assistant History: Start analyze and contextualize the data. Assistant UCDS: Aggregate data; fetch additional information from LOD cloud. both: Study how this can answer the given historical questions, which visualizations/user cases will be developed for the final paper.</td>
</tr>
<tr>
<td>June-July 2022</td>
<td>Assistant UCDS: Publish the enriched dataset to the LOD and an accompanying article demonstrating various use-cases. both: Write final paper</td>
</tr>
</tbody>
</table>

(148 words)
Bibliography

https://doi.org/10.1093/pastj/gtw029.