Digital Innovation in Medieval Arabic History

Applicants

*Eligible proposals must have at least **two applicants from Humanities, preferably from two different institutes.***

Supervisor Name	Institute
Prof Dr Petra M. Sijpesteijn/ Dr Edmund Hayes	Leiden University Institute for Area Studies/ EMCO project
Dr Paul Vierthaler	Leiden University Centre for Digital Humanities (LUCDH)

Project description

The flourishing literary culture of early Islam produced thousands of texts, large and small, which survive both in contemporary documentary form (papyri, manuscripts), and in later copies, which are increasingly being digitized. In spite of the huge number of texts that survive from the first 3 centuries of Islam, little has been done to analyse these texts on a large scale using computational techniques. Scholars of Arabic and Islamic history are only very recently beginning to apply digital methods to the analysis of the large corpora of texts. Last year's LUCIS fellow, Sarah Savant (AKU London, KITAB) and her collaborator, Maxim Romanov (Vienna, KITAB OpenITI) have been putting together a corpus of literary texts which has been prepared and made available for such analysis (http://iti-corpus.github.io/).

In the framework of Petra Sijpesteijn's ERC project "Embedding Conquest" (2017-2021) (EMCO) (https://www.universiteitleiden.nl/en/research/research-projects/humanities/embedding-conquest#tab-1), the research trainees will work with a team of PhD students and postdocs to analyse and compare the extant physical documents with the huge corpus of literary materials that has now been digitized. This work was not foreseen in the ERC project as it is based on recent knowledge exchange/cooperation between LUCDH, the EMCO team and OpenITI offering exciting new opportunities.

This research project will identify, quantify and periodize the emergence and usage of social group vocabulary from the Islamic conquests in the 7th century CE until the eclipse of the Abbasid dynasty in the 10th century. Trainees will compare vocabulary in two sets of sources: physical documents (papyri, manuscripts, inscriptions); and literary sources (in particular biographical dictionaries). The project will involve two major tasks: (1) inputting and tagging Arabic texts, and (2) applying simple digital text-analysis techniques to the growing corpus of texts gathered and analysed. The research trainees will participate in producing a collaborative publication with EMCO members based on the results of the research. The project will provide trainees with concrete skills in coding, the digital analysis of historical texts which will be adaptable to many different research contexts in the future, as well as more traditional skills of writing academic papers and making presentations.

Research Trainee Profile

BA3, MA and ResMA students can apply to the traineeship. Research Trainees should have a demonstrable knowledge of Modern Standard Arabic at an advanced level (B2 of the Common European Framework of Reference for Languages2). They should also be familiar with the medieval history of the Near East.

The Research Trainees are expected to take the course 'Culture and Society in the Medieval Muslim World' (MA Middle Eastern Studies) or to have successfully participated in this course in the academic year 2014-2015 or 2015-2016. No specific technical skills are necessary, beyond a willingness to learn new skills, and an aptitude for problem solving.

General tasks of each of the Research Trainees:
- Taking a short online course on basic coding in Python (e.g. as provided by codeacademy.com), in order to learn how to computationally analyze the corpus under supervision
- Helping to build the corpus of letters for the team to work on
- Marking up letters with machine-readable tags according to social categories developed by the team
- Helping build and work through bibliographies of secondary literature to enrich the list of social group names and categories used by the team
- Participate in EMCO team meetings to embed them in the research environment
- Present findings at a public workshop or conference
- Contributing to a joint publication on social categories in early Islamic sources

Collaboration

By partnering between LIAS and LUCDH, digital humanistic methods developed in other fields can be applied to the analysis of Arabic corpora. In the context of Leiden University, there is a strong presence of sinologists working in digital humanities, reflecting the greater progress made in applying digital methods to Chinese literature than has yet been made in the field of Arabic and Islamic studies. This project and the publication that results will, therefore, contribute a pioneering step in the field of Arabic digital humanities, providing direction for the field as a whole. Within Leiden, the collaboration will represent a cross-pollination between scholars thinking systematically about ways of devising sets of objective categories to computationally analyse very different texts and societies. In addition collaboration will incorporate a wider network of digital humanists at other institutions, through collaboration with the OpenITI team.

Deliverables
- Publication of a collaborative article on social categories in early Islam between documentary and literary sources
- Building a corpus of tagged Arabic letters for analysis by the "Embedding Conquest" team. The corpus will also be put into the public domain by being uploaded to the OpenITI corpus to be used by other scholars in the future. (http://iti-corpus.github.io/)
- Research trainees will present findings and methodology at a public workshop including participation from the LIAS/LUCDH and the digital humanists from London and/or Vienna.
- Contributing to the young field of Arabic digital humanities.

Planning
- **Month 1: Orientation and skills building.** The trainees will take an online course on the Python programming language (www.codeacademy.com). At the same time, trainees will be introduced to the categories being used in the EmCo team, and start tagging texts under close supervision.
- **Months 2-5: Tagging letters corpus and mining secondary literature**. The trainees will work to tag texts a little more autonomously, though still under supervision to check for accuracy and to think about future analysis. Trainees will be integrated into ongoing team discussions about the validity of the categories used, which will flow back into refinement of the research goals. Team members will experiment with different types of quantitative analysis and visualization of results. Trainees will also read from an assigned bibliography to mine existing literature for social group names and categories which have been discovered by other scholars
- **Months 5-8: Analysis and visualization and presentation.** Trainees will carry out searches of the growing corpus, and participate in the team's analysis and visualization of the material. Trainees

will also be encouraged to run their own searches and provide their own analysis of the corpus, to the extent possible, in order to think about future research directions for themselves.

- **Month 8:** Presentation. Trainees will make a presentation of their findings and methodology at a public workshop to expose the "work in progress" of the "Embedding Conquest" project. Results will be incorporated into a collaborative publication authored by members of the "Embedding Conquest" team.

Student Application

To apply, students should send an email with resume and motivation letter to Edmund Hayes at e.p.hayes@hum.leidenuniv.nl