

The Cluster of Excellence  
**Understanding Written Artefacts**  
cordially invites you to the workshop

## **Data Linking Workshop 2023: Computer Vision and Natural Language Processing – Challenges in the Humanities**

Tuesday, 27 June 2023, 10:00 am – 9:00 pm CEST  
Wednesday, 28 June 2023, 9:30 am – 12:15 pm CEST

Warburgstraße 26, 20354 Hamburg, and online

Organised by Sylvia Melzer (Universität Hamburg,  
Universität zu Lübeck) and Haiyan Hu-von Hinüber  
(Max-Weber-Kolleg, Shandong-Universität)

Registration:

<https://www.csmc.uni-hamburg.de/register-workshop34>

Historians are increasingly using technologies to evaluate digitised texts in a machine-readable way, as well as techniques from the field of natural language processing (NLP) to analyse the content and context of language in written artefacts. These techniques can be used to analyse large corpora and identify patterns. In general, however, these methods often use training data from current rather than historical data. The use of these methods can lead to biases in the historical record, incurring the risk of false inferences about history. Therefore, the methods used should be fully investigated to account for any biases. In this Data linking workshop, the challenges of applying computer vision and NLP techniques in the humanities, and first solutions to them, will be presented.

## Programme

---

### Tuesday, 27 June, 10:00 am – 9:00 pm

- Chair: Haiyan Hu-von Hinüber (Max-Weber-Kolleg, Shandong-Universität)
- 10:00 – 10:20 Konrad Hirschler (Universität Hamburg), Sylvia Melzer (Universität Hamburg, Universität zu Lübeck)  
*Welcome*
- 10:20 – 10:50 Eva Wilden, AI: Charles Li (Universität Hamburg)  
*Tamilex -- Digital Lexicography*
- 10:50 – 11:10 Stefan Baums (Ludwig-Maximilians-Universität München), AI: Stephen White (Venice)  
*Computer Vision and Kharoṣṭhī Paleography*
- 11:10 – 11:30 Coffee Break
- Chair: Oliver Hellwig (Universität Düsseldorf)
- 11:30 – 11:50 Oskar von Hinüber (Universität Freiburg, Mainz Akademie), Haiyan Hu-von Hinüber (Max-Weber-Kolleg, Shandong-Universität), AI: Sylvia Melzer (Universität Hamburg, Universität zu Lübeck)  
*What the Buddhological Epigraphy can expect from the AI: The Information System 'Buddhist Bronzes Inscriptions'*
- 11:50 – 12:10 Kathrin Holz (Universität Würzburg), AI: Olga Serbaeva (University of Zurich)  
*The Proto-Śāradā Project: Towards the edition of a new collection of administrative letters and documents from pre-modern South Asia*
- 12:10 – 12:30 I. Konczak-Nagel (Sächsische Akademie der Wissenschaften zu Leipzig), AI: Erik Radisch (Sächsische Akademie der Wissenschaften zu Leipzig)  
*The Kucha Mural Information System: Taxonomy and Semi-Automated Image Recognition*
- 12:30 – 2:30 Lunch Break (in Pavilion for presenters)

- Chair: Hussein Mohammed (Universität Hamburg)
- 2:30 – 3:30 Ralf Möller (Universität zu Lübeck)  
*Aligned AI and the role of the humanities: Training AI systems using human feedback*
- 3:30 – 3:50 Isabelle Marthot-Santaniello (University of Basel)  
*The application of NLP in combination with Computer Vision for analysing ancient Greek handwritings on papyri*
- 3:50 – 4:10 Olga Serbaeva (University of Zurich)  
*Some features of the 17th century Newārī script: READ-based statistical approach to palaeography*
- 4:10 – 4:30 Lena Hinrichsen (Herzog August Bibliothek)  
*OCR technologies in research practice*
- 4:30 – 5:00 Discussion Round
- 7:00 – 9:00 Dinner (invited workshop participants)

### Wednesday, 28 June, 9:30 am – 12:15 pm

- Chair: Sylvia Melzer (Universität Hamburg, Universität zu Lübeck)
- 9:30 – 10:00 Oliver Hellwig (University of Düsseldorf)  
*Web-based information systems for Indian scripts and texts*
- 10:00 – 10:20 Simon Schiff (Universität zu Lübeck), Ralf Möller (Universität zu Lübeck)  
*Persistent Data, Sustainable Information*
- 10:20 – 10:40 Hamid Reza Hakimi (Universität Hamburg), Lisa Mischer (Universität Hamburg), Tariq Yousef (Universität Hamburg), Maxim Romanov (Universität Hamburg)  
*Finding and Linking Information in Arabic Historical Texts*
- 10:40 – 11:00 Coffee Break
- Chair: Simon Schiff (Universität zu Lübeck)
- 11:00 – 11:20 Sylvia Melzer (Universität Hamburg, Universität zu Lübeck)  
*Building Information Systems on Demand with ChatGPT?*

- 11:20 – 11:40 Martin Braun (University of Vienna), Hannes Fellner, Bernhard Koller (University of Vienna)  
*A Digital Paleography of Tarim Brahmi*
- 11:40 – 12:00 Hussein Mohammed (Universität Hamburg)  
*Computer vision beyond OCR: potentials and challenges for the study of written artefacts*
- 12:00 – 12:15 Concluding Remarks