

September 10, 2026
Austrian Academy of Sciences
Georg-Coch-Platz 2, 1010 Vienna

Zoom: [https://ceu-edu.zoom.us/j/69174842543?](https://ceu-edu.zoom.us/j/69174842543?pwd=nrHVlrnn48nPv1VCrLx3aNNiKA5HmK.1)
[pwd=nrHVlrnn48nPv1VCrLx3aNNiKA5HmK.1](https://ceu-edu.zoom.us/j/69174842543?pwd=nrHVlrnn48nPv1VCrLx3aNNiKA5HmK.1)

Program

OCR/HTR Workshop for Under-resourced and Under- represented Languages Redux

Funded by the FWF Cluster of Excellence
EurAsian Transformations

9:00-9:15 (CET): Welcome

Alíz Horváth (Central European University): The Way We Work Now – Whither Text Recognition?

9:15-10:35: Session 1: Persistent challenges for non-Latin scripts

Alexander O'Neill (Musashino University): Reading the Unreadable: Pracalit versus Bhujimol and the Challenges of Under-Resourced Buddhist Manuscript Traditions

Clemens Neudecker (Berlin State Library): Document Layout Analysis and Reading Order Detection for Non-European Languages and Scripts

Matthias Arnold (Heidelberg University): 'I'm Sure That Someone Has Already Done This in China', Or How to Cope with Reviewer#2

10:35-10:50: Coffee break

10:50-12:10: Session 2: Workflows & Systems

Rachael Griffiths (École Pratique des Hautes Études): Building Ground Truth: Multi-stage Annotation for Tibetan HTR

Michal Racyn (Masaryk University): Modular Open-source Pipeline for Archival Research: From On-site Scanning to Data Visualization

Sinai Rusinek (Haifa University, Open University DH-Dev): Winding Workflows: Editors, Artificial Agents and Specialist ATR Models

12:10-13:10: Lunch

13:10-14:30: Session 3: Text recognition & co.

Seth Kulick (University of Pennsylvania): Using ATR of Yiddish to Study Language Change

Olaf Berg (Ruhr University Bochum): Processing Ottoman Documents: Text Recognition with Kraken/eScriptorium and Translation into Latin Script with LLM

Jan Odstrčilík (ÖAW) & Martin Roček (Charles University): Character Error Rate Diagnostics with CERatosaurus: Discover Hidden Issues

14:30-14:45: Coffee break

14:45-16:05: Session 4: AI for text recognition

Colin Brisson (École Pratique des Hautes Études): AnandaSky: A Vision-Language Model for Line-Level Transcription of Historical Sinographic Documents

Wayne de Fremery (Dominican University of California): Local-First AI for Historical Korean OCR

Andy Stauder (Transkribus/READ COOP): Is the Long Tail of Document Diversity a Glass Ceiling for General-purpose AI? An R&D Report

16:05-16:45: Catharsis: Mini writing sprint to revise/update the zine "10 annoying things about digitizing under-resourced and under-represented languages – And what might help fixing them"

16:45-17:30: Closing remarks and Special tour to explore the historic Postsparkasse Building with Jan Odstrčilík