

Workshop Summary: New Ways of Transcribing, Visualizing, Publishing, and Providing Access to Data on Epidemics and Contagious Diseases

Date: 18 June 2025

Host: Institute of Evolutionary Medicine, University of Zurich

Organized by: Prof. Dr. Kaspar Staub

Affiliation: [IUSSP Scientific Panel on 'Epidemics and Contagious Diseases: The Legacy of the Past'](#)

This interdisciplinary online workshop brought together a global cohort of 48 researchers, data scientists, historians, and public health professionals to explore cutting-edge methods in digitizing, visualizing, and sharing historical data on epidemics. The event aimed to foster collaboration and share innovations in how we manage, interpret, and make accessible both structured and unstructured data from handwritten and printed historical sources.

Session Highlights

Opening Remarks

Prof. Kaspar Staub welcomed participants, emphasizing the workshop's mission to promote methodological exchange and community building among researchers working with historical and modern epidemic data.

Session I: Innovations in Transcription and Visualization

- **Vijay Kumar (India)** opened with an overview of digital innovations that enable more accurate geospatial transcription and visualization of epidemic data. He highlighted the value of open data platforms and AI tools for public health planning and research, particularly in the context of Indian administrative records.
- **Moana Rarere (New Zealand)** presented a new demographic framework for 19th-century Māori populations. Her work challenges colonial-era assumptions and integrates oral histories with written records to reshape historical demographic understandings.
- **Gaurav Raj (India)** demonstrated how Python-based approaches can unlock patterns from historical epidemic records, including automating data extraction from scanned sources and integrating diverse datasets for analytical modeling.

Session II: Standardization, Surveillance, and Public Health Integration

- **Gabi Wuethrich (Switzerland)** discussed the application of the Text Encoding Initiative (TEI) Guidelines to digitized historical statistics. Her talk focused on the challenges of rendering tabular data into FAIR-compliant formats, especially from handwritten or scanned sources.
- **Grace Kim (USA)** shared insights from projects linking climate and health data to improve early warning systems. Drawing from malaria and climate surveillance initiatives, she emphasized the role of interdisciplinary integration for real-time public health response.

- **Pamela R. Chacon Uscamaita (USA)** introduced a malaria surveillance platform in the Amazon Basin. Her team's work combines predictive modeling, remote sensing, and interactive dashboards to track disease spread and inform interventions.

Session III: Transcription at Scale and the Role of Automation

- **Tobias Hodel & Jan Blarer (Switzerland)** detailed best practices for large-scale processing of handwritten death registers. Their talk emphasized human-in-the-loop workflows, which combine machine learning with archival expertise for quality control.
- **Mads Perner (Denmark)** explored the epidemiological transition of Copenhagen via spatial analysis of causes of death and disease notifications from the 19th to early 20th century, highlighting the value of integrated mapping and registry systems.
- **Joana Maria Pujadas Mora & Adrià Molina (Spain)** presented automated transcription tools for processing demographic records. They emphasized accessibility and reproducibility, showing how machine learning can democratize the processing of historical population data.
- **Christian Møller Dahl (Denmark) & Rick Mourits (Netherlands)** concluded the session with a summary of a recent workshop on automatic transcription. They critically examined the strengths and limitations of automation when handling vast individual-level mortality data, calling for improved evaluation metrics and collaborative benchmarking.

Closing Remarks

Prof. Staub wrapped up the workshop by encouraging participants to pursue ongoing collaboration, and announced plans for further community engagement through follow-up meetings and joint projects.

This workshop underscored the dynamic evolution of how we work with epidemic data—from manual transcription to automated pipelines, and from isolated archives to open, interoperable data ecosystems. The contributions highlighted both the promise and the complexity of bringing historical health data into the digital age.