



Mapping the Science of Interwar Czechoslovakia

Digital Reconstruction of Scientific Networks and Methodological Challenges

Filip Kersch^{1,2}

¹ Library of the Czech Academy of Sciences

² Institute of Information Science and Librarianship, Faculty of Arts, Charles University

Introduction

WHAT

This poster presents a project reconstructing **the information network** of scientific publishing in Czechoslovakia (1910–1939) to map the evolution of science in the newly established state, using **full-text data** from scholarly society journals.

WHY

While traditional historiography relies on archival records often overemphasizing prominent figures and isolated institutions, the historical scientific community actually relied on inter-institutional journal exchange.

Approaching science as the **circulation and externalisation of knowledge** via Digital Humanities and Information Science overcomes this archival bias.

Objectives

- **Map** dominant topics and track their transformations over time.
- **Link** authors, locations, institutions, and topics into a complex network of scientific communication.
- **Analyze** the impact of external factors (political shifts, state-building, economic crisis) on this information ecosystem.

Methodology

As the project is in its early stages, this poster discusses the design of the processing pipeline and outlines candidate digital methods for handling a large historical text corpus.

Pipeline:

- 1 Data Retrieval**
Harvesting images and full-text data from Czech digital libraries.
- 2 OCR Quality Control**
Validating existing text and ALTO layout metadata, re-running OCR for missing or low-quality layers.
- 3 Automated Article Segmentation**
Splitting unsegmented volumes using layout/typography heuristics (headers, author names) and semantic boundaries (*sentence-transformers*).
- 4 Named Entity Recognition (NER)**
Testing and comparing tools (NameTag3 vs. LLMs) to extract persons, locations, and institutions.
- 5 Topic Modeling**
Exploring traditional algorithms (LDA) with neural approaches (Top2Vec, BERTopic).
- 6 Graph Database**
Modeling authors, publications, periodicals, institutions, places, and topics as interconnected nodes.

NETWORK GRAPH PREVIEW
(IMG/NETWORK-GRAPH.PNG)

Interactive Presentation

Presenting current results from a pilot data sample. Scan the QR code to explore the interactive network model.

QR
(IMG/QR-CODE.PNG)