

# An Introduction to Digital Humanities for EAS

Monday Mar. 3, 2025  
11:00-2:00 EST

East Asian Studies, University of Toronto  
Brought to you by:  
Rose Ting-Yi, PhD Candidate

*Guest Speakers*

**Dr. Paul Vierthaler**  
East Asian Studies, Princeton University  
**Peichao Qin**  
East Asian and Middle Eastern Studies,  
Cambridge University

# Workshop Outline

11:00-11:30 Rose's overview

What is Digital Humanities

Data, Methods, Tools

11:30-2:00 Guest Lectures  
(with Q&A after each talk)

\*Workshop is recorded

# What is Digital Humanities

Data, Methods,  
Tools

Future sessions

Guest Lectures

# Digital Humanities *The Conceptuals*

- Humanities research on issues regarding the digital age
  - Digital humanities research ethics

or

- **Digital humanities as tools and methods**

What is Digital  
Humanities

Data, Methods,  
Tools

Future sessions

Guest Lectures

# Digital Humanities as method:

Using quantitative research methods and  
tech tools to investigate humanist questions  
and problems

What is Digital  
Humanities

Data, Methods,  
Tools

Future sessions

Guest Lectures

Digital Humanities apply to:

Any research question/materials fit for  
quantitative investigation

# Digital Humanities research relies on...

What is Digital  
Humanities

Data, Methods,  
Tools

Future sessions

Guest Lectures

## Digital tools

Eg. Markus, Gephi,  
Paraview, ArcGIS,  
DocuSky, Python, R, GenAI

## Digital archives

Eg. CBDB, CTEXT, 中國基本  
古籍庫...

# DH methods can offer us:

- New perspectives
- “Big Pictures”
- A different kind of evidence
- Possibly: different conclusions and new insights

# But, DH is not omnipotent

- “Distant”
- “Generalizations”
- 
- DH = a set of analytic tools, a way of thinking
- 
- What really matters:
  - *You*: Research Questions, Interpretations

# Digital Humanities as Method

What is Digital  
Humanities

Data, Methods,  
Tools

Future sessions

Guest Lectures

## *The Technicals*

# Digital Humanities as *Method*

What is Digital  
Humanities

Data, Methods,  
Tools

Future sessions

Guest Lectures

## Data collection

## Data analysis

# Types of Data

What is Digital  
Humanities

Data, Methods,  
Tools

Future sessions

Guest Lectures

- Quantitative →
  - Amount, measurement
- Qualitative →
  - Property, feature, attribute
    - **Qualitative data can be quantified (to a degree)**
- Numerical
  - Discrete (count)
  - Continuous
  - Interval scale
  - Ratio scale
- Categorical
  - Nominal
  - Ordinal (ranked)

# Types of Data Analysis

## *Coding and Analyzing Qualitative Data*

2 types of data analysis:

- **Descriptive statistics**
  - Simple summaries about basic features/patterns of the dataset
- **Inferential statistics**
  - Hypotheses and mathematical tests

What is Digital  
Humanities

Data, Methods,  
Tools

Future sessions

Guest Lectures

# Coding and analyzing qualitative data

## Typical Tools:

- **Coding surveys and interviews:** *NVivo*
- **Data collecting, organization, quantitative analysis:** *Excel, Google Sheet, PAST, SAS, R*
- **Network Analysis:** *Gephi, Paraview*
- **Text markup:** *Markus, Markdown*
- **Mapping:** *CHGIS, ArcGIS, Google Map, Geojson, Python*
- **Visualization:** *Excel, Python, R*
- **Lit. Review:** *CiteScope*
- **General assistance:** *GenAI\**

What is Digital  
Humanities

Data, Methods,  
Tools

Future sessions

Guest Lectures

# Today's GenAI Chatbots

- LLM - Large Language Model
  - Simulation of “neural networks”; input-output matrices
  - *Probability*-based output
- Current GenAI models on the market:
  - ChatGPT (Open AI), Claude (Anthropic), Gemini (Google), Copilot (MS), Deepseek, etc. etc.
  - **Still all just “weak” models - “monkeys in a room with a Chinese dictionary”**
    - Biases, inconsistencies, “hallucinations”, “laziness”, privacy issues, ethical concerns
- Already very useful!
- GenAI models/agents in the future: self-conscious, real intelligence?
- 
- GenAI for data analysis: *let machines do the legwork*

# GenAI for data analysis

What is Digital  
Humanities

Data, Methods,  
Tools

Future sessions

Guest Lectures

Eg. Configuring LLM for Google Sheet with  
OpenAI/Gemini API

<https://www.kwokleongtang.net/GenAI/use-with-ot-her-tools/google-sheet-with-llm-api/>

# DH platforms/resources for China Studies

What is Digital  
Humanities

Data, Methods,  
Tools

Future sessions

Guest Lectures

**Markus** - online Markup tool for texts

<https://dh.chinese-empires.eu/markus/>

**Docusky** - Chinese DH platform

<https://docusky.org.tw/docusky/home/v5/>

**CHGIS** - China Historical GIS

<https://chgis.fairbank.fas.harvard.edu/>

**CBDB** - Historical Biographies

<https://projects.iq.harvard.edu/cbdb>

**CTEXT** <https://ctext.org/>

# My Projects

## Mapping Han dynasty Jade Suits

- CHGIS; Google Map; Python

<http://rosework.space/research-jadesuits.html>

## Women Artists of Premodern China

- Text Markup for data collection
- Word Cloud
- Mapping Women Painters

● <http://rosework.space/research-women-artists.html>

- Next Step: Social Network of Women Artists

What is Digital  
Humanities

Data, Methods, Tools

Future sessions

Guest Lectures

# DH Interest Group



**DH Interest Group  
sign-up**

DH Asia on discord

<https://discord.gg/pYpHFrsZ>

Digital Orientalist on discord

<https://discord.gg/mBG5UPCu>

# What is Digital Humanities

Data, Methods, Tools

Future sessions

Guest Lectures

- Markup for textual analysis
- Gen-AI: utilizing AI for your research
- Data Visualization
- Mapping with Geojson, Google Map, Python, CHGIS
- Building a simple website

Rose Ting-Yi 琹怡

East Asian Studies, University of Toronto

[tingyi.liu@mail.utoronto.ca](mailto:tingyi.liu@mail.utoronto.ca)



Early China | Gender Archaeology | Feminist Art History | Digital Humanities



Interested in a hands-on session?  
Let me know!  
[tingyi.liu@mail.utoronto.ca](mailto:tingyi.liu@mail.utoronto.ca)

# Guest Lectures

(1 hour talk + 15 min Q&A) x 2

**\*Q&A** following each guest talk:

**In-person audience:** please use the mic

**Online audience:** please use the Chat or Q&A function, or raise your hand

# Guest Speaker: Dr. Paul VIERTHALER

<https://www.pvierth.com/>

## *“Machine learning and uncovering source material in the late Ming novel Jinpingmei”*

Paul Vierthaler is an assistant professor of late imperial Chinese literature and interdisciplinary data science at Princeton University. He has previously taught Chinese literature at William & Mary and the digital humanities at Leiden University (where he helped establish the Leiden University Centre for Digital Humanities). His first monograph project tracks how historical information in late imperial China was transmitted and deformed through fiction and unofficial histories (“quasi-histories”) using traditional critical analysis and computational analytics (natural language processing, corpus linguistics, machine learning, and other data analyses). He is also working on a variety of projects at the intersection of the humanities and machine learning. This includes a second monograph project focused on the late Ming novel Plum in the Golden Vase, and a set of interconnected projects studying Ming/Qing dynasty censorship, editorial practices, and their impact on the creation of large language models.

# Guest Speaker: Peichao QIN

<https://oracular.azurewebsites.net>

## *"From Image to Text: Automatic Information Extraction and Processing for 2D Artifact Images via Computer Vision"*

Peichao Qin is a fourth-year PhD student at the Faculty of Asian and Middle Eastern Studies (FAMES) at the University of Cambridge. He holds a B.A. degree from the Chinese University of Hong Kong (CUHK) and M.A. from The University of Edinburgh. His PhD dissertation, supervised by Prof. Roel Sterckx, focuses on the Shang oracle bone collection currently stored on site at Cambridge University Library, which was donated by Lionel Charles Hopkins (1854-1952) in the 50s and includes over 600 unique items. As part of his research to facilitate the study of the archive, Peichao has developed a comprehensive computer font that allows users to type out complex oracle bone characters and a website called "Jingyuan Digital Platform" for searching the glyphs and visualizing the geo and temporal distribution of the oracle bone artifacts. Peichao also has other DH research projects that are related to various tasks such as automatic artifact restoration, image upscale, handwritten OCR using image processing algorithms and Machine Learning models, some of which are launched or in active development.