Managing and Mining Historical Research Data over Time
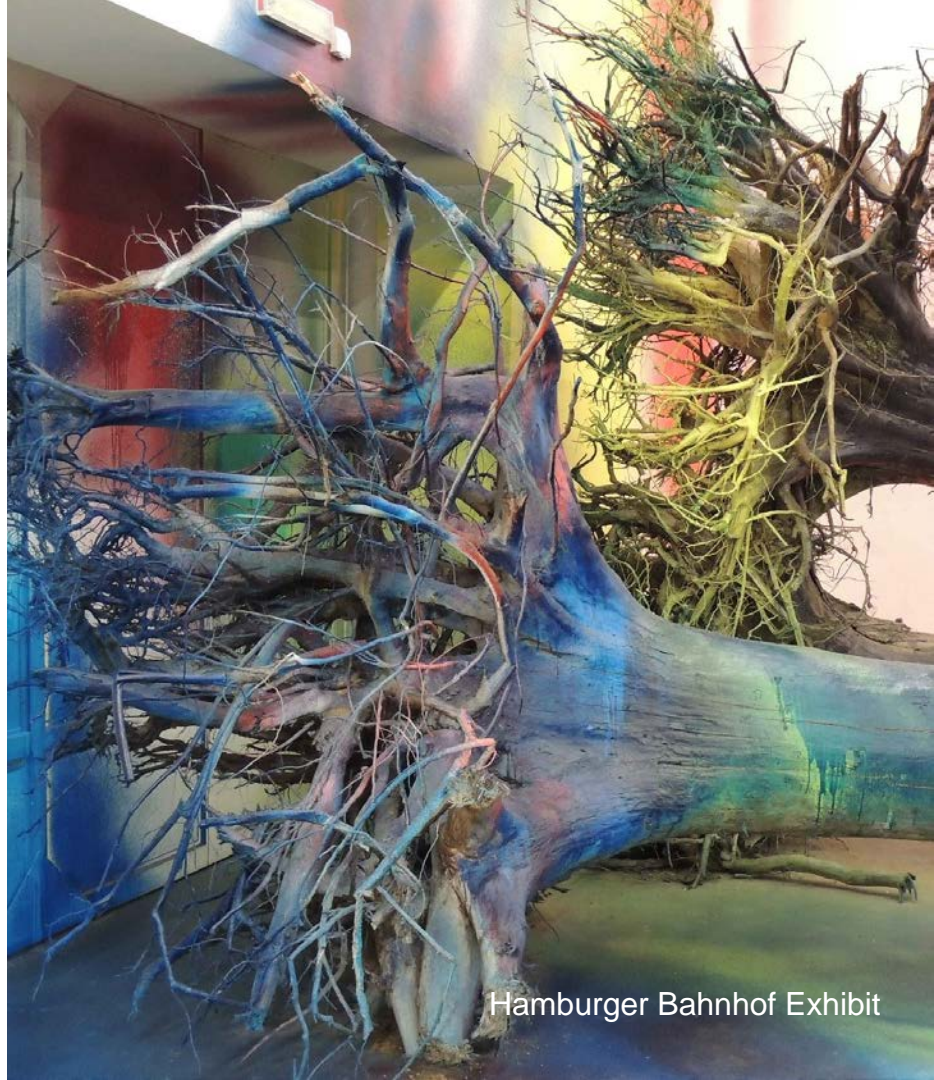
Michael Seadle

Berlin School of Library & Information Science

Humboldt-Universität zu Berlin

November 2014, Antalya, Turkey

Overview

- Reading and granularity
- Management & mining humanities research data
- Humanities data mining today & tomorrow
- Curriculum issues
- Archiving issues

Hamburger Bahnhof Exhibit

Reading

- Vocabulary
- Speed
- Comprehension
- Retention

Foujita, Lyon Museum

Granularity

- Reading books
- Reading chapters or articles
- Reading paragraphs
- Reading sentences
- Reading data

Foujita, Lyon Museum

Content Mining

- Reading speed
- Reviews
- Footnotes as links
- Bibliographies as links
- Human interaction as content mining

Foujita, Lyon Museum

Humanities Data

- Words, sentences, paragraphs
- Tables and Diagrams
- Images and photographs
- Voice and video recordings
- Ethnographic observation

Foujita, Lyon Museum

Managing historical data

Search-friendly environment:
- structured content
- (openly?) accessible content
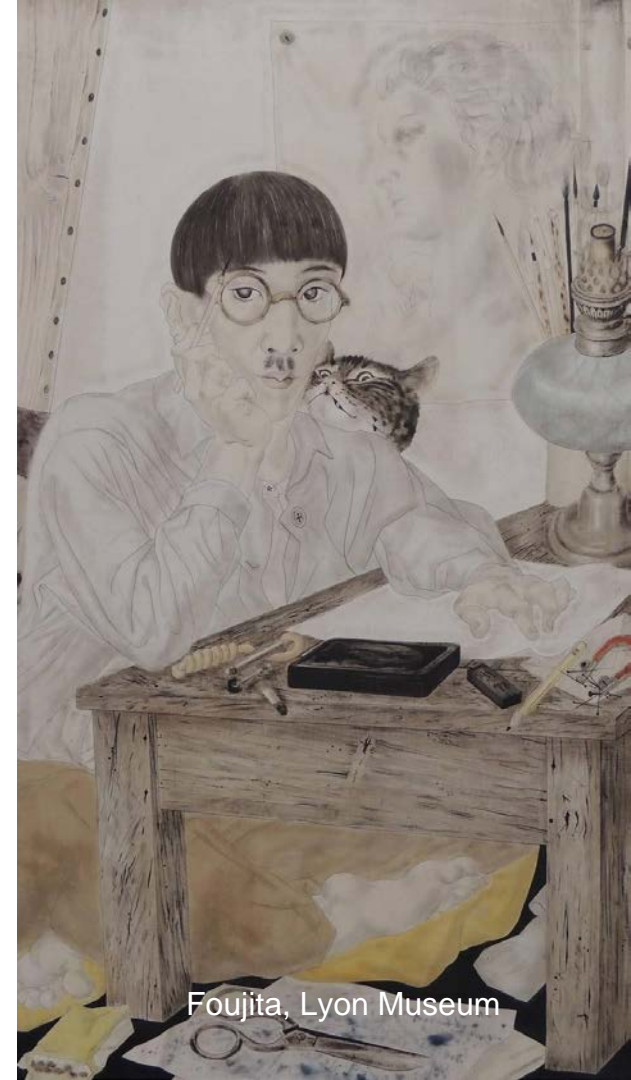- authentic content
- context information

The Internet Archive is working on this.

Foujita, Lyon Museum

Mining humanities data today

- Abstracting services
- Full text search
- Google [Ngram Viewer](#)
- Programs for mining available digital text
- Distant Reading / text mining

Foujita, Lyon Museum

Mining humanities data tomorrow

Desires and illusions:
- Cross-platform searching
- Format standardization
- Built-in context awareness
- Useful metadata

Packages vs. programming

Foujita, Lyon Museum

Curriculum issues

- Programming languages
- Descriptive and inferential statistics
- Context (ethnography)
- Context (history)

Eatily, NYC

Archiving issues

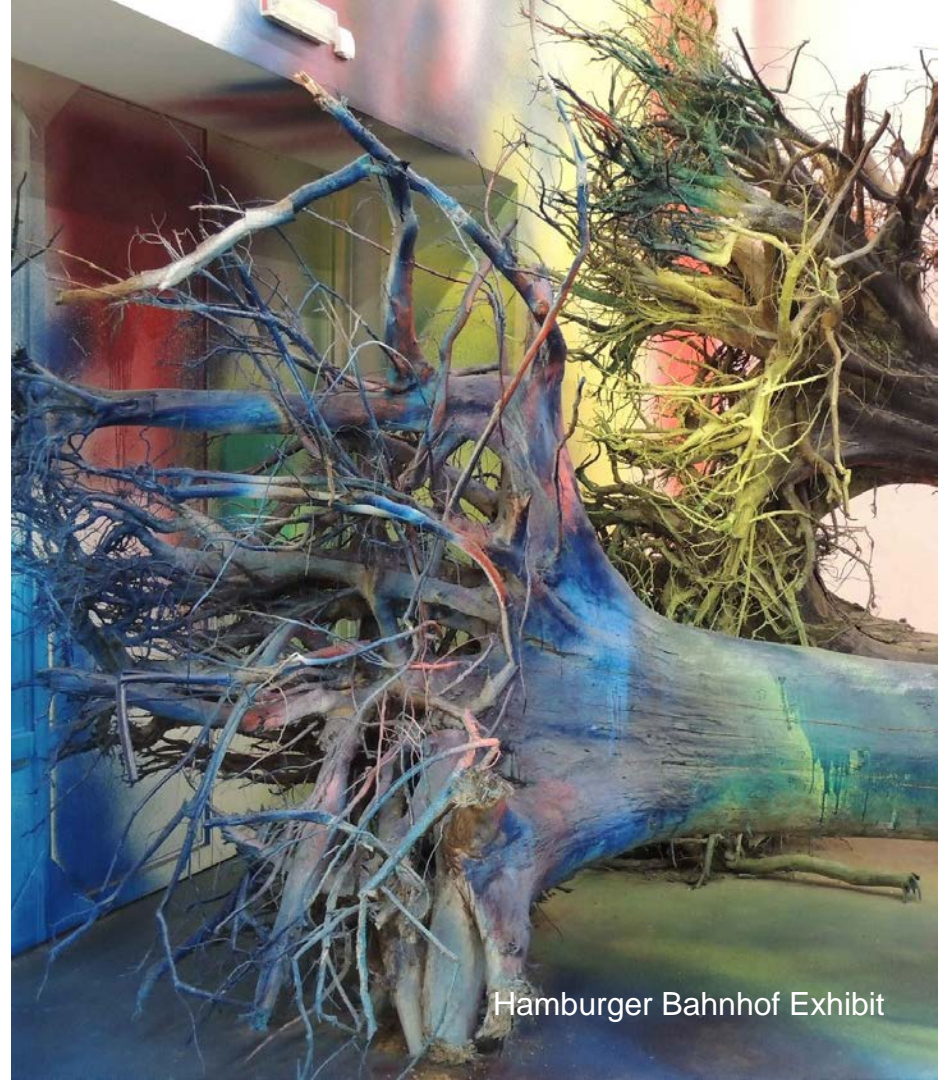Secondary sources (of course)
Primary sources:
- text (e.g. email)
- drafts
- images
- digital environments

Eatily, NYC

Conclusion

- Packages & skills
- Text & data
- Technology & history
- Focus & context

Hamburger Bahnhof Exhibit

Questions

Contact information

Prof. Dr. Michael Seadle

Room 03 in the
Sekretariat
Dorotheenstr 26, Berlin

seadle (at) ibi.hu-berlin.de

+49 30 2093 4249