

Towards Textual Studies' High Tide? Studying Research Epistemologies in Digital Humanities after the Machine Learning Turn

Dr. J. Berenike Herrmann (Basel)
Dr. des. Thomas Messerli (Basel)

Prof. Noah Bubenhofer (Winterthur)
Daniel Knuchel (Winterthur)

Digitization

- Vast amount of digital data & powerful text mining tools for Humanities scholars, e.g., sentiment analysis.
- Many tools were developed within engineering and commercial application.
- Machine Learning and Deep Learning can be effective even in the absence of any expert knowledge about the make-up of text.
- Need for reflection of Digital Humanities practices.

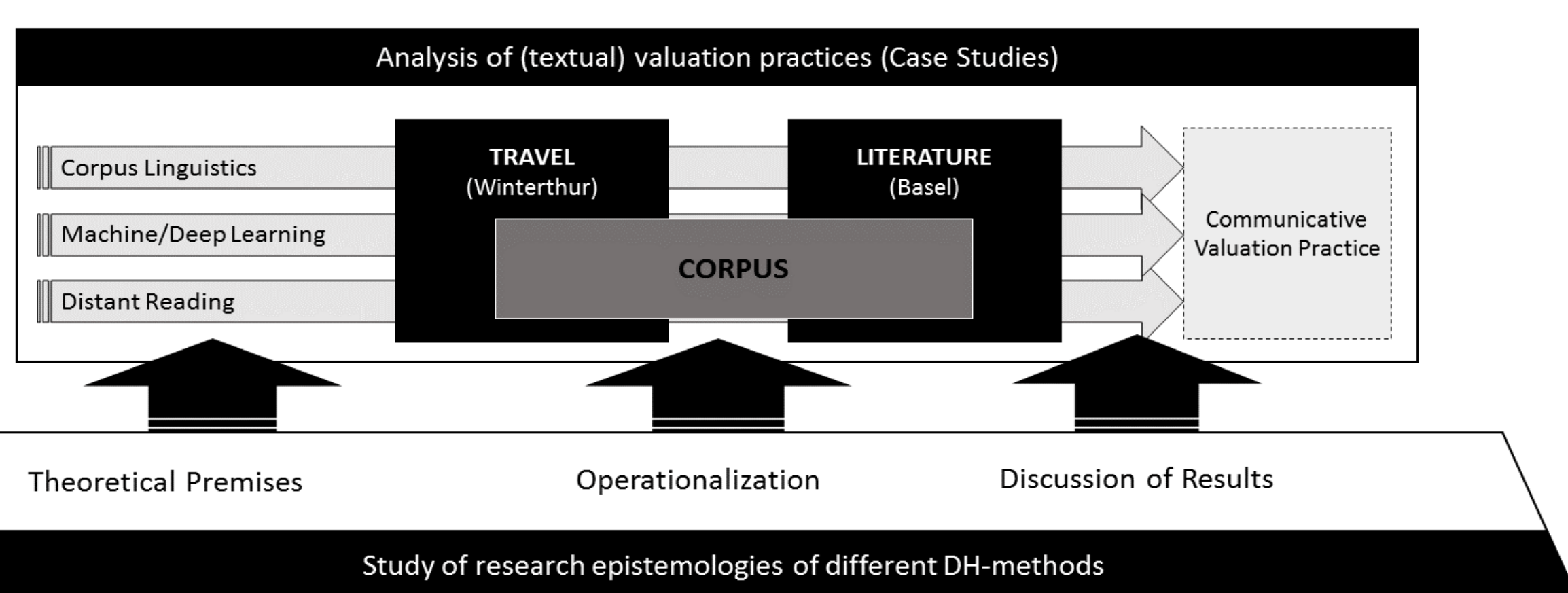
Overarching Research Questions

How is digitization transforming the humanities and social sciences?

- What is the application benefit of digital tools in the textual humanities?
- Do they answer pertinent questions and offer new perspectives?
- Do they aid –or possibly obstruct– a reflective Humanities scholarship?

Research Design

Epistemological research ↔ Two Web 2.0 Case Studies
Sentiment Analysis



Research epistemologies («Forschungslogiken»)

- Guiding schemata for scientific/scholarly action
- Mutual determination with regard to reality & humanity
- *Showing in research practice*
 - Types of research questions, argumentation, interpretation, methodology
 - Thought styles («Denkstile»; Fleck, 2012, first 1935)
 - Epistemic cultures («Wissenskulturen»; Knorr-Cetina, 1999) see also Bubenhofer & Dressen (2018)

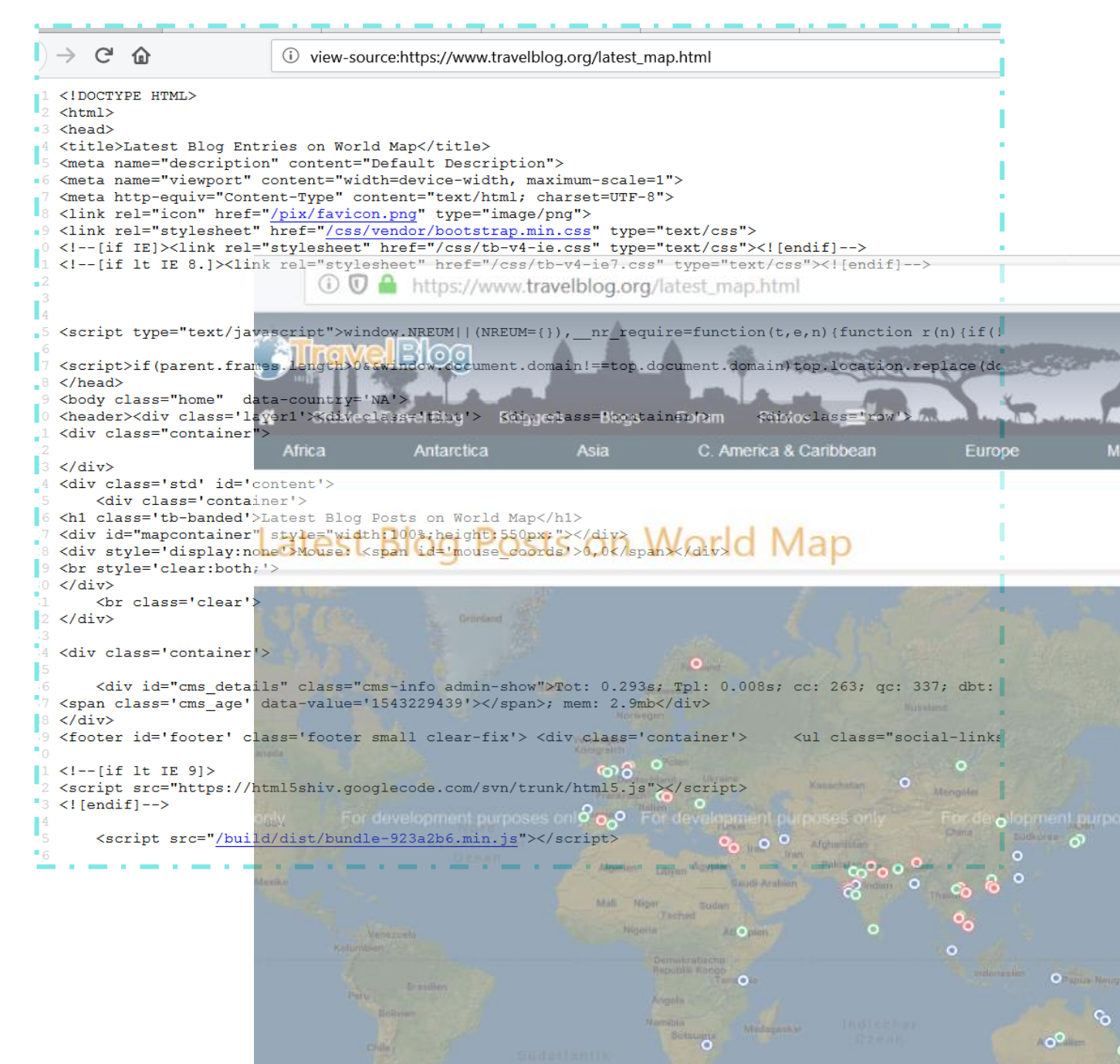
➤ **Theoretical Premises:** What understanding of language and communication motivate the different approaches? What assumptions are made about the world, reality and humanity?

➤ **Operationalization:** What, if any, linguistic and literary categories and theories are being operationalized? What covert and overt decisions inform existing analytical practices?

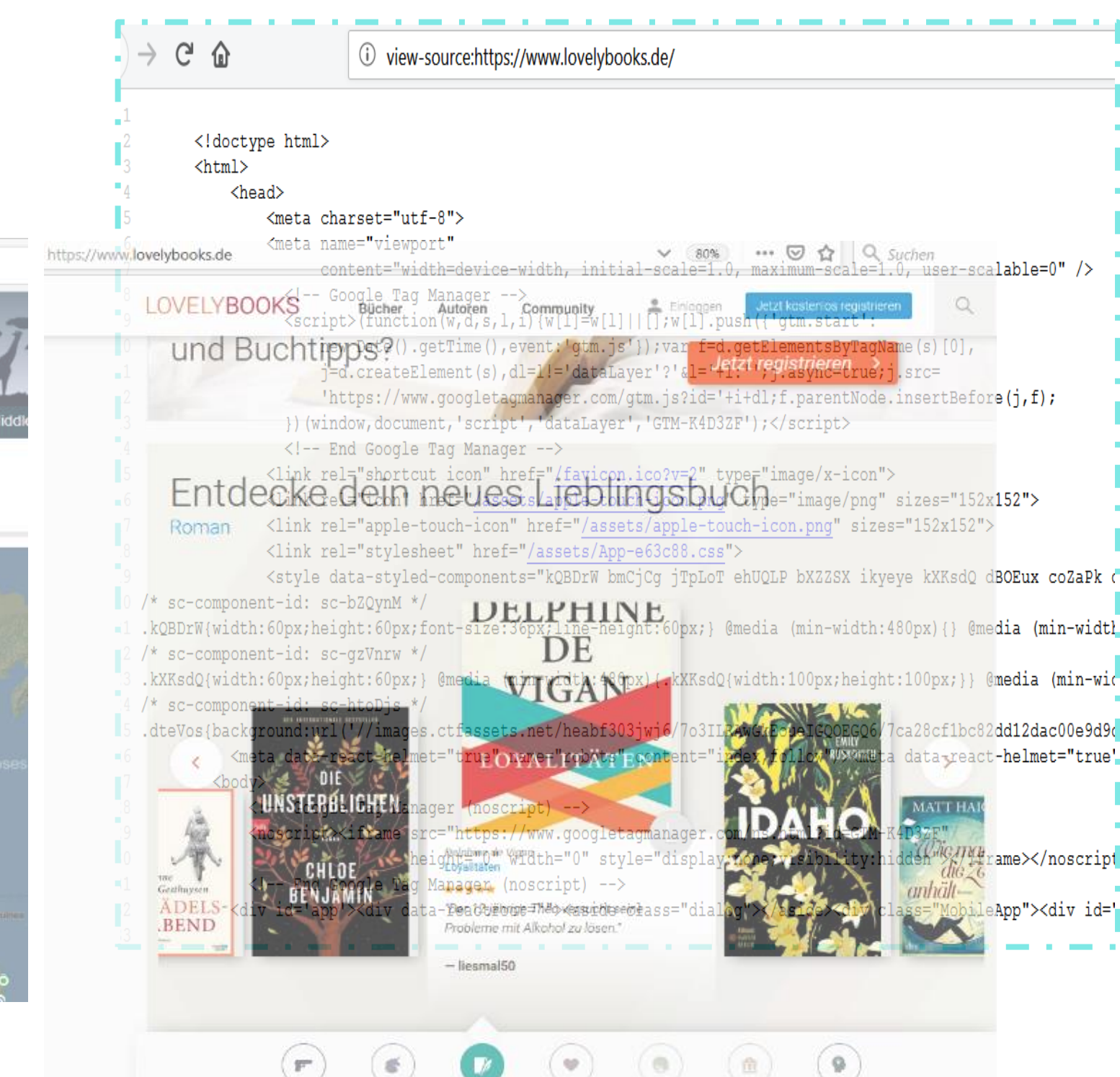
➤ **Discussion:** What rationale is used to discuss and interpret results? How do interpretations of results tie in with the respective theoretical frameworks? What open questions are foregrounded? What role and importance are given to the interpretation of data within the overall research projects? How is subjectivity discussed?

➤ **Research Practice:** What explicit and implicit rationale motivates the employment of particular research processes?

Travelblog.org
ca. 10,000 travel blog entries



Lovelybooks.de
ca. 1 Mio. book reviews



Social Web Discourses

(E)valuation Practices

- NLP and Psychological Sentiment Analysis (e.g., Buechel et al., 2017; Egloff et al., 2018; Jakobs, 2018; Liu, 2012)
- (Digital) Linguistics (e.g., Benamara et al., 2012; Biber & Zhang, 2008; Hunston & Thompson, 2000; Sandig 1979)
- Aesthetic judgement (Bourdieu, 1984; Cassirer, 1932; Kant, 1790; Wolff, 1732)
- Literary Criticism (e.g., Heydebrand & Winko, 1996; Rehfeldt, 2017)
- Distant Reading (e.g., Archer & Jockers, 2016; Piper & Portelance, 2016)
- Web2.0 Digital Lives: Writer-Reader ("prosumer") (e.g., Kaplan & Haenlein, 2010); Social Digital Reading (e.g., Piper, 2012; Stein, 2010)

RESEARCH PLAN	PREPARATION	Literature (Basel)	Travel (Winterthur)	
		Data Acquisition	Web data processing, adapting previously developed tools (e.g., see Bubenhofer, 2018)	
ANALYTICAL STEPS	Data Processing	Processing of raw data (corpus-linguistic standard)		
	Method	Machine Learning	Distant Reading	Corpus Linguistics
	Lead	joint	Basel (Herrmann)	Winterthur (Bubenhofer)
	Method Development	Discussion/Definition of approach (with project partners Cieliebak & Roth)	Review of research in literary studies, incl. Distant Reading	Review of (corpus)linguistic research
	Fine-tuning of Method	Coding of training data Development & Adaption of a fitting model	Devising and testing different procedures Discussion of developed methods in research meetings	
Method Application	Analysis of both corpora Interpretation of results			

Methods

- Harvesting and preprocessing of two corpora
- Literary and linguistic methods: Corpus linguistics, distant reading, hermeneutic
- NLP: Sentiment and opinion mining tools
- Machine learning: Statistical and neuronal
- Cultural anthropology: Participatory observation, Meta-annotation (e.g., Gius & Jacke, 2017), logbooks (Glaser, 2008)

Perspectives & Potential

Applied reflection of Digital Humanities practices, with new insights on:

- Research Epistemologies in two communities of textual DH
- Sentiment Analysis across textual DH
- The Role of Machine Learning in textual DH
- (E)valuative practices of Web 2.0 users
- Digital transformation of Humanities: Perils and potentials