



IEG
Leibniz-Institut für
Europäische Geschichte

Mitglied der
Leibniz
Leibniz-Gemeinschaft

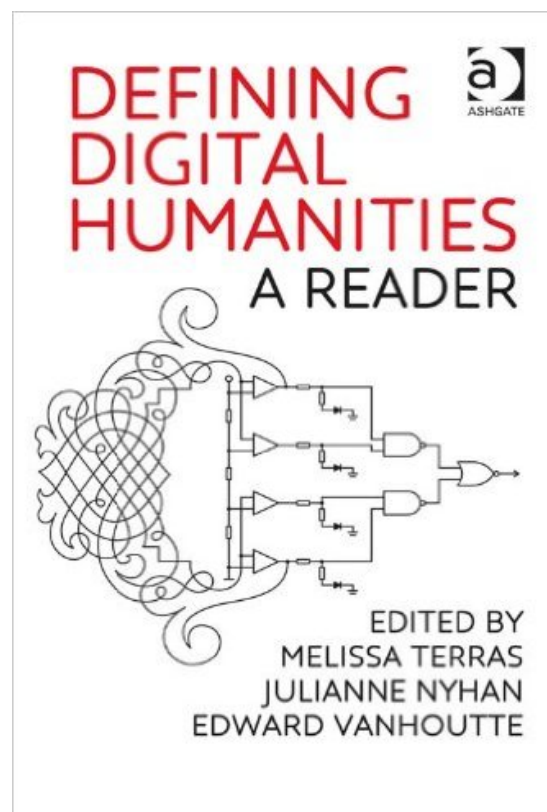
Digital Humanities, Computational Linguistics, and Natural Language Processing

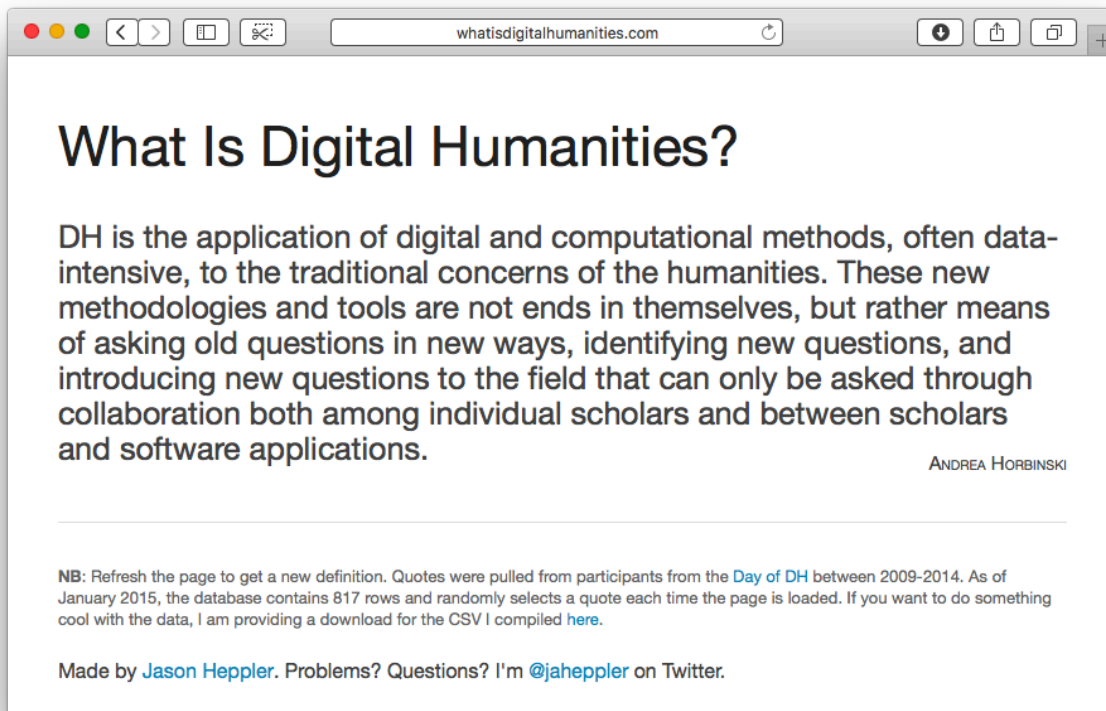
Dr.-Ing. Michael Piotrowski
Leibniz Institute of European History
<piotrowski@ieg-mainz.de>

Uppsala, March 4, 2016



Defining Digital Humanities





Do we really need a definition?

Yes, we do.

If you want to create a program of studies or devise a research agenda, you *must* commit yourself to *some* definition.

- ▶ However, most definitions focus on *methods* and say very little about *goals*.
- ▶ Related problem: Are the digital humanities a discipline of their own, an interdisciplinary field, a community of practice, or something else again?

Consensus

Relatively broad consensus, that the digital humanities bring together humanities and computer science; thus we have two aspects:

- Ⓐ Work on *humanities* research question using methods and tools from computer science
 - Ⓑ Work on *computer science* methods und tools for tackling research questions in the humanities
- **Term is inherently ambiguous**

Piotrowski 2012

The emerging field of digital humanities aims to exploit the possibilities offered by digital data for humanities research. The digital humanities combine traditional qualitative methods with quantitative, computer-based methods and tools, such as information retrieval, text analytics, data mining, visualization, and geographic information systems [GIS].

[Piotrowski 2012, p. 6]

Piotrowski 2013

In a narrow sense, “digital humanities” refers to the application of quantitative, computer-based methods for humanities research, usually complementing traditional qualitative methods [...]. The important point is that it is humanities research, i.e., you’re applying these methods to answer a humanities research question.

In a wider sense, it may also refer to the application of computer-based tools in humanities research (note that this definition does not require the use of quantitative methods). For example, creating a digital edition is not digital humanities in the narrow sense (because it does not use quantitative methods), but it is in the wider sense.

<http://nlphist.hypotheses.org/114>

Discussion

- + Relatively clearly delimited area of research
- + Uncontroversial, but not arbitrary
- Actually only a description of *practices*
- Nothing is said about motivations or goals of the digital humanities

Why Digital Humanities?

- ▶ Ultimate goal of all science and scholarship: gaining new insights by systematic research (“Erkenntnisgewinn”)
- ▶ What is the benefit of combining humanities and computer science *for the humanities*?
- ▶ Acceleration of research through digitization?
- ▶ Automatic analyses of large amounts of data?
- ▶ Attractive visualizations?
- Where is the advancement or “innovation”?

Piotrowski 2016

Definition (Digital humanities)

The digital humanities study the means and methods of constructing formal models in the humanities.

Definition (Digital history)

Digital history is concerned with the construction of formal models of historical circumstances and with the methodology of constructing such models.

Correspondingly: Digital literary studies, digital philosophy, etc.

These are *subfields of their respective disciplines*, characterized by the creation and use of formal models.

Formal models

- ▶ A model is a representation of a selected part of the world.
- ▶ Model \approx description \approx theory
- ▶ Слово “формальный” не означает ничего, кроме как “логически последовательный + однозначный + абсолютно явный”.

[Gladkij & Mel'čuk 1969, p. 9]

The word “formal” means nothing more than
logically coherent + unambiguous + explicit.

- ▶ There are different degrees of formalization; here we are primarily interested in a degree of formalization that allows models to be processed and manipulated by computers.

Formal models

- ▶ **All** scientific and scholarly research constructs models of their objects of research.
- ▶ In order to understand a complex object (phenomenon, situation, ...), you need to understand its parts and how they interrelate with each other.
- This is exactly what a model describes.
- ◎ In contrast to the natural sciences, models in the humanities are traditionally not formal and not directly accessible; narratives are *not* models, but *informal descriptions* of models.

Digital humanities as a metascience

Definition (Digital humanities)

The digital humanities study the means and methods of constructing formal models in the humanities.

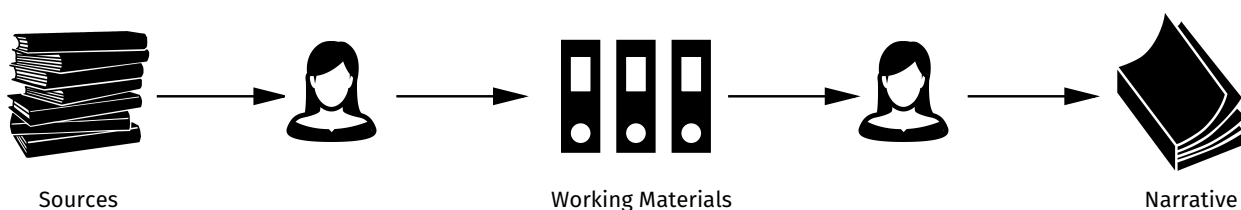
- The digital humanities are concerned with the “construction materials” for such formal models; thus: a **metascience**.

Definition (Digital history)

Digital history is concerned with the construction of formal models of historical circumstances and with the methodology of constructing such models.

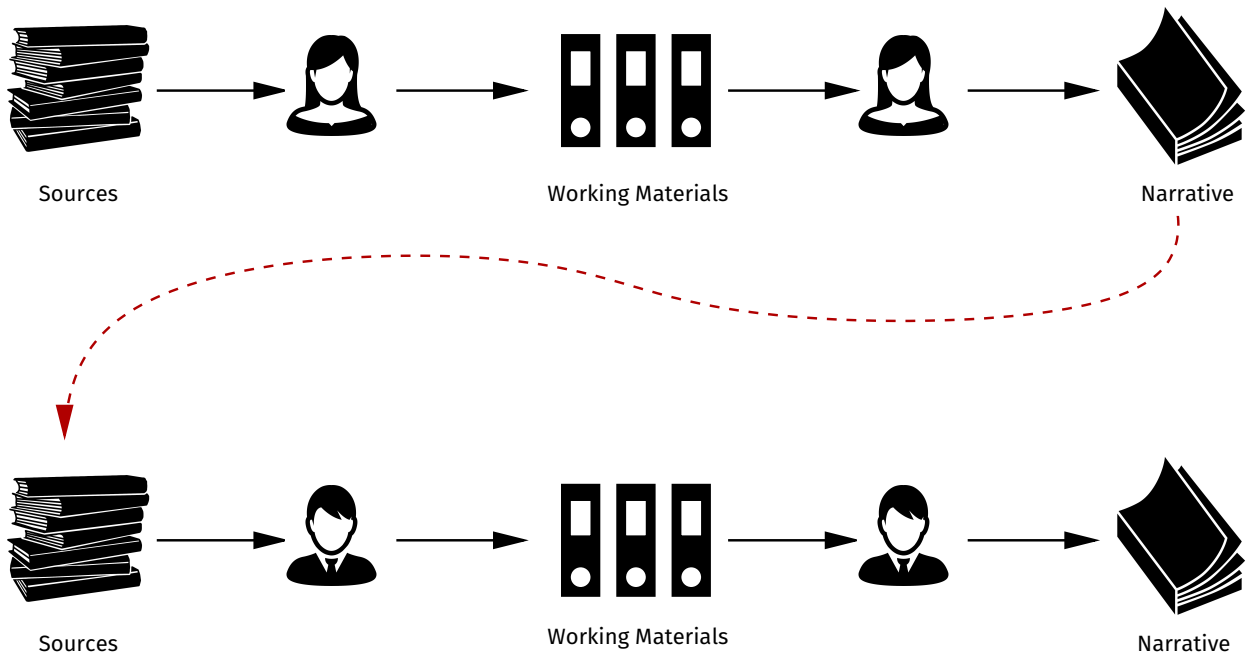
- Individual digital humanities subfields create *concrete* formal models of their research objects.
- ⊙ There is no strict boundary between digital humanities and individual digital humanities subfields.

Traditional research process

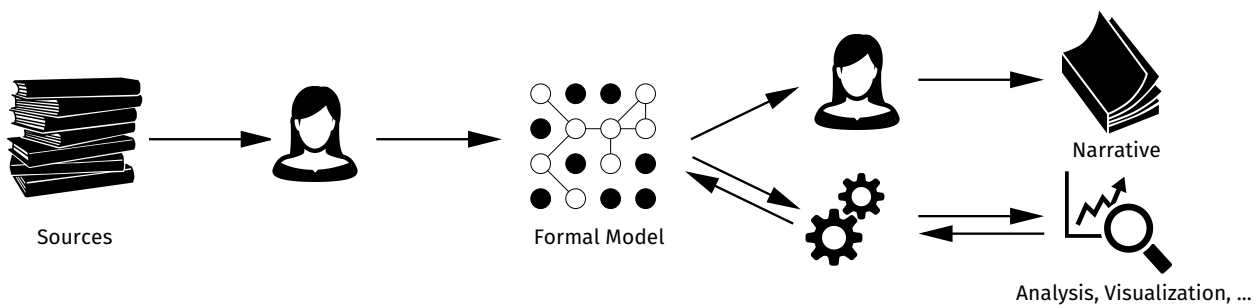


- ▶ Scholar reads and interprets primary and secondary **sources**
- ▶ Facts and insights are recorded as **working materials** in a variety of forms [on paper or electronically, as text, in spreadsheets, databases, etc.]
- ▶ Using the working materials, scholar constructs **mental model** to answer research question and describes the model in a **narrative**.

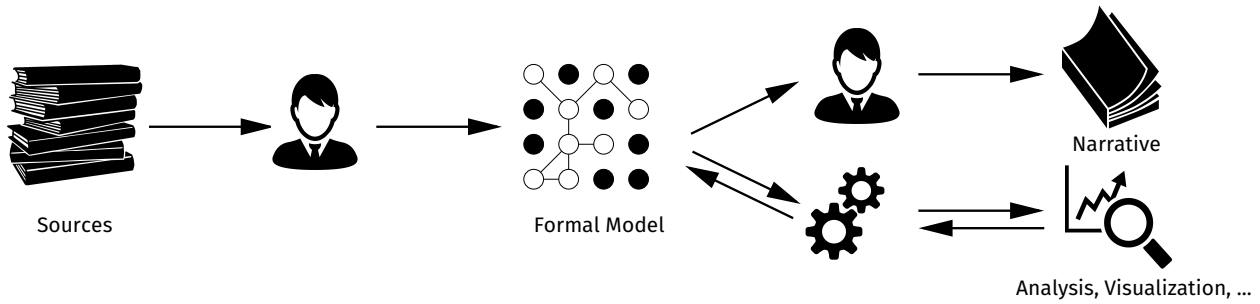
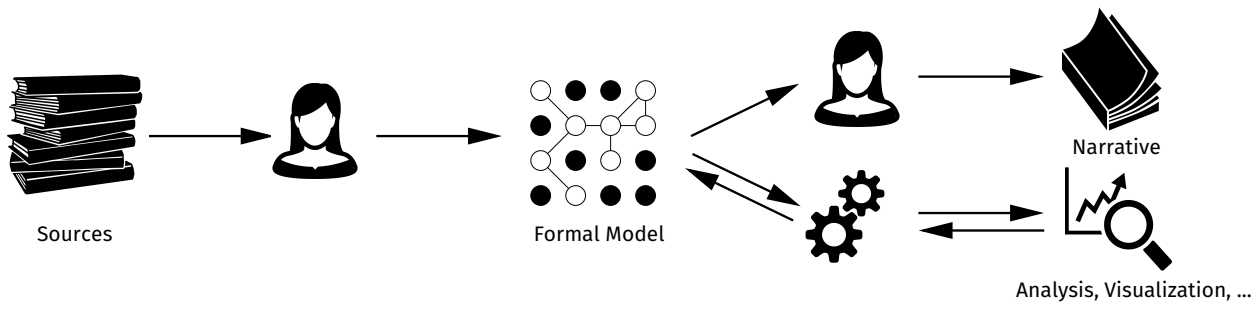
Building on the work of others (traditional process)



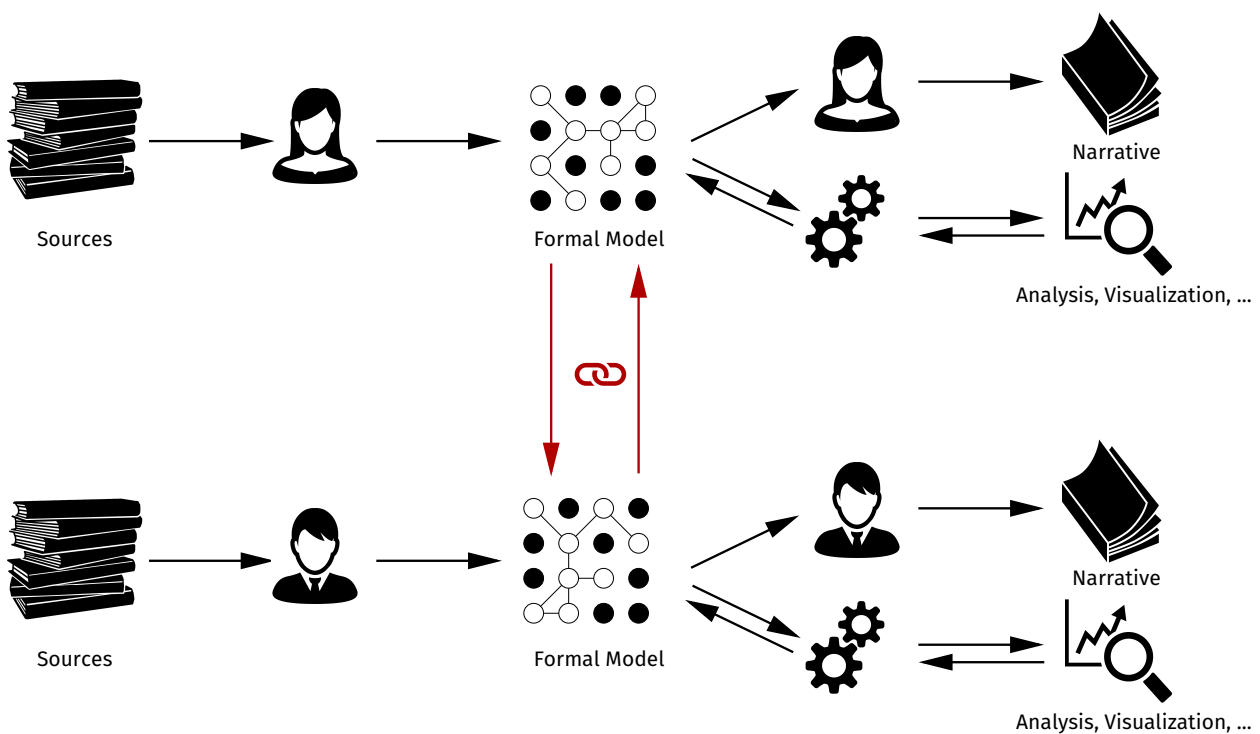
Where do formal models come into play?



Collaboration on a higher level



Collaboration on a higher level



What do we need?

- ▶ Humanities research questions and results are primarily qualitative.
- ▶ Digital humanities are primarily qualitative.
- **Knowledge representation** is central for the creation of formal models in the humanities.

DH, CL, and NLP

- ▶ Linguistics has a “vantage point” for observing the digital humanities, because it has essentially completed the transformation from “armchair linguistics” to an empirical science using formal models
- ▶ The role of computational linguistics corresponds to that of digital humanities
- ▶ The role of corpus linguistics corresponds to that of the digital humanities subfields (such as digital history)
- ▶ Where is the place of NLP?
 - ▶ Applied computational linguistics?
 - ▶ Engineers’ take on linguistics?
 - ▶ Computer science?
 - ▶ Toolsmiths?
- ▶ What is the role of NLP in digital humanities?

NLP and DH

- ▶ If the humanities seriously want to base their research on large quantities of text (and quantitative methods), they will need NLP as basis for all higher-level analyses
- ▶ For digital historical scholarship, NLP must then be regarded as an auxiliary science of history, similar to diplomatics, codicology, paleography, numismatics etc., which are indispensable for evaluating and using historical sources

Il n'est pas indispensable que le philologue établisse lui-même le programme, encore que ce soit infiniment souhaitable ; il devrait au moins connaître assez le langage de programmation pour contrôler le travail du technicien ; en effet, l'expérience m'a appris qu'il ne faut pas s'en remettre les yeux fermés aux électroniciens, mal préparés par leur formation mathématique à se faire une idée juste de problèmes concrets qui se posent dans la domaine de la philologie.

[Jacques Froger, 1970]

Summary

- ▶ The digital humanities do not merely aim to accelerate research or to analyze larger amounts of data.
- ▶ The key is **formal modeling** of scholarly knowledge and insights in machine-processable form.
- ▶ Formal models increase coherence, precision, and explicitness, encourage cooperation and sharing, and help researchers to directly build upon each other's work.
- ▶ **Knowledge representation** techniques are thus the foremost tools for creating formal models in the humanities.
- ▶ The “digital humanities discussion” can benefit from studying the development of linguistics.
- ▶ Digital humanities subfields can learn from corpus linguistics.
- ▶ NLP should be considered an *auxiliary science*—as such, DH researchers have to get acquainted with its methods and tools.



IEG

Leibniz-Institut für
Europäische Geschichte

Mitglied der
Leibniz
Leibniz-Gemeinschaft

Digital Humanities, Computational Linguistics, and Natural Language Processing

Dr.-Ing. Michael Piotrowski
Leibniz Institute of European History
<piotrowski@ieg-mainz.de>

Uppsala, March 4, 2016

