

# DH is the Study of dead Dudes

**Hall, Mark**

mark.hall@informatik.uni-halle.de

Martin-Luther-Universität Halle-Wittenberg, Deutschland

## DH is the Study of dead Dudes

Die Digital Humanities (DH) werden oft als Möglichkeit gesehen neue Methoden und neue Datenmengen zu nutzen um neue Ansätze auf Fragen der Geisteswissenschaften zu entwickeln (Barry 2011). Als Teil des DH Forschungsprozesses muss in vielen Fällen ein Datensatz oder eine Person als Fokus der Studie gewählt werden. Dass dabei existierende soziale Trends verstärkt werden, besonders was den Fokus auf den Kanon betrifft ist bekannt und in verschiedenen Hinsichten problematisiert worden (Liu 2012; McPherson 2012; Wernimont 2013; Fiormonte 2016; Gallon 2016; Rhody 2016). Es basiert aber keine dieser Kritiken auf einer expliziten Analyse der im DH Bereich unternommenen Forschung. Dieser Beitrag unternimmt einen ersten Schritt um dies zu korrigieren und präsentiert einen quantitativen, distant-reading Ansatz um der Frage nach Gender, Sprache, und Herkunftsland von in DH untersuchten Personen nachzugehen.

## Methode

Die hier präsentierte Analyse basiert auf den Konferenzabstracts der DHd Konferenzen 2016, 2017, und 2018. Aus den Abstracts wurden alle erwähnten Personen manuell identifiziert. Auf eine automatisierte Identifikation wurde verzichtet, da dies nur eine neue Fehlerquelle in die Analyse einführen würde, ohne inhaltlich irgendeinen Unterschied zu machen. Da die Abstracttexte selbst nicht im Detail gelesen wurden, ist der Ansatz als distant-reading zu verstehen und berücksichtigt auch keine Abstracts, die sich mit diesem Thema beschäftigen, aber keine Personen expliziten erwähnt.

Für die Identifikation der Personen wurden die Inklusions- und Exklusionsregeln in Tabelle 1 angewandt. Das zentrale Prinzip für das Zählen einer Person ist, dass sie explizit ausgewählt worden sein muss. Das heißt, Personen werden nur gezählt, wenn sie oder ihr Werk das primäre Untersuchungsobjekt sind (#1) oder weil sie als exemplarisches Beispiel eines Themas (#2) oder als Sample eines Datensatzes (#3) präsentiert werden. Personen, die sowohl als Studiensubjekte auftreten, aber auch selbst akademische Theorien entwickelt haben, werden nur gezählt wenn ihre Arbeiten das Ziel der Analyse sind, nicht aber wenn sie Teil des methodischen Ansatzes sind (#4). Zum Beispiel, wenn die Schriften

Adornos automatisiert analysiert werden, dann wird er gezählt. Wenn aber die Theorien Adornos zur Interpretation anderer Daten genutzt werden, dann wird er nicht gezählt. Letztens werden Personen nicht gezählt, wenn sie in einer vollautomatisch erstellten Liste auftreten, bei der kein manueller Eingriff in der Erstellung auftrat (#5). Zum Beispiel wenn bei einer Häufigkeitsanalyse die 20 häufigsten Namen gelistet werden, dann werden diese nicht gezählt, da die Studie Verzerrungen in der Auswahl von Studiensubjekten betrachtet und nicht Verzerrungen in den Datensätzen.

| # | Regel  | Aktion    |
|---|--|-----------|
| 1 | Person als primäres Untersuchungsobjekt          | Inklusion |
| 2 | Person als Beispiel für das untersuchte Thema    | Inklusion |
| 3 | Person als Sample aus dem untersuchten Datensatz | Inklusion |
| 4 | Person erwähnt als Theoriegeber_in               | Exklusion |
| 5 | Person erwähnt als vollautomatisiertes Ergebnis  | Exklusion |

Tabelle 1: Inklusions- und Exklusionsregeln für die Erstellung des Personendatensatzes.

Für die derart identifizierten Personen wurden dann Gender, Sprache, und Herkunftsland identifiziert, wobei als Quelle primär Wikipedia genutzt wurde. Bei der Klassifikation von Sprache und Herkunftsland wurde eine Mehrfachzuordnung durchgeführt, um Unterschiede zwischen heutigen und historischen Länder und Sprachgrenzen abzubilden und Mehrsprachigkeit der historischen Personen zu berücksichtigen. Aus den insgesamt 230 erwähnten Personen konnte nur für 4 keine Zuordnung durchgeführt werden. Basierend auf der Zuordnung wurde dann pro Abstract die Zahl der Erwähnungen ermittelt. Pro Abstract wurde eine Person die mehrmals erwähnt wurde nur einmal gezählt. Zugleich wurde eine Person die in verschiedenen Abstracts erwähnt wird, pro Abstract je einmal gezählt.

Die Abstracts wurden mit Publikationsjahr und Kategorie (Vortrag, Poster, Workshop, ...) annotiert, wobei hier nur die Vortrags- und Poster-abstracts analysiert werden.

## Ergebnisse

Insgesamt wurden 342 Abstracts analysiert (162 Vorträge, 180 Poster), in denen in 104 explizit Personen erwähnt wurden (60 in Vorträgen, 44 in Postern) (Tabelle 2). Bei den Vorträgen stellt dies einen Anteil zwischen 33% und 50% aller Abstracts da, bei den Postern liegt der Anteil

zwischen 19% und 28%. Für die Analyse wurden erstens die Erwähnungen pro Abstract aggregiert und dann die Zahl der Abstracts mit zumindest einer Erwähnung betrachtet. Zweitens wurden die Erwähnungen auch unabhängig von den Abstracts analysiert.

|         | 2016   | 2017         | 2018   |              |        |              |
|---------|--------|--------------|--------|--------------|--------|--------------|
|         | Gesamt | mit Personen | Gesamt | mit Personen | Gesamt | mit Personen |
| Vortrag | 60     | 20           | 36     | 18           | 66     | 22           |
| Poster  | 77     | 18           | 36     | 7            | 67     | 19           |

Tabelle 2: Zusammenfassung des analysierten Datensatzes.

## Gender

Tabellen 3 und 4 zeigen die Genderverteilung für die Vortrags- und Posterabstracts. In drei Jahren Vortragsabstracts gibt es keinen einzigen in dem nur weibliche Personen erwähnt werden. Bei den Posterabstracts ist die Situation marginal besser, mit je einem Abstract in 2016 und 2017. Insgesamt wurden über die drei Jahre hinweg nur in 15% aller Vortragsabstracts und in 7% aller Posterabstracts, in denen Personen erwähnt wurden, Frauen namentlich erwähnt.

|         | 2016 | 2017 | 2018 |    |   |      |   |    |      |   |   |
|---------|------|------|------|----|---|------|---|----|------|---|---|
|         | M    | F    | M&F? | M  | F | M&F? | M | F  | M&F? | M | F |
| Vortrag | 17   | 0    | 3    | 15 | 0 | 3    | 0 | 19 | 0    | 3 | 0 |
| Poster  | 16   | 1    | 0    | 1  | 6 | 1    | 0 | 18 | 0    | 1 | 0 |

Tabelle 3: Anzahl an Abstracts die entweder nur männliche (M), nur weibliche (W), sowohl männliche und weibliche (M&W), oder nur unbekannte (?) Personen erwähnen.

Bei den Einzelerwähnungen (Tabelle 4) ist die Situation ähnlich. Bei den Vortragsabstracts machen Frauen 9% der Erwähnungen aus, bei den Posterabstracts 5%.

|         | 2016 | 2017 | 2018 |    |   |   |    |   |   |    |   |
|---------|------|------|------|----|---|---|----|---|---|----|---|
|         | M    | F    | ?    | M  | F | ? | M  | F | ? | M  | F |
| Vortrag | 33   | 4    | 3    | 55 | 7 | 0 | 63 | 3 | 0 | 33 | 1 |
| Poster  | 19   | 1    | 1    | 6  | 1 | 0 | 33 | 1 | 0 | 33 | 1 |

Tabelle 4: Anzahl an Erwähnungen männlicher (M), weiblicher (W), oder unbekannter (?) Personen.

## Sprache

Bei den Sprachen der erwähnten Personen stellt sich die Situation anders da. Bei den Vortragsabstracts bewegt sich der Anteil an Abstracts mit nicht-deutschsprachigen

Personen zwischen 29% und 58% (Tabelle 5) und der Anteil an erwähnten nicht-deutschsprachigen Personen zwischen 46% und 60% (Tabelle 6). Bei den Posterabstracts sind die jeweiligen Zahlen etwas niedriger, mit Anteilen zwischen 20% und 60% (Tabellen 7 und 8).

|      | Deutsch | Englisch | Französisch | Portugiesisch | Spanisch | Andere |
|------|---------|----------|-------------|---------------|----------|--------|
| 2016 | 14      | 3        | 2           | 2             | 0        | 3      |
| 2017 | 14      | 3        | 4           | 0             | 0        | 2      |
| 2018 | 13      | 7        | 3           | 1             | 2        | 5      |

Tabelle 5: Anzahl an Vortragsabstracts die mindestens eine Person mit dieser Sprache erwähnen. Sprachen die nur einmal vorkommen sind in die Kategorie "Andere" aggregiert.

|      | DE | EN | FR | PT | ES | NL | SE | IT | Andere |
|------|----|----|----|----|----|----|----|----|--------|
| 2016 | 19 | 4  | 2  | 3  | 7  | 0  | 0  | 0  | 2      |
| 2017 | 31 | 16 | 7  | 0  | 0  | 2  | 2  | 0  | 0      |
| 2018 | 24 | 13 | 11 | 0  | 4  | 0  | 0  | 3  | 5      |

Tabelle 6: Anzahl an Personen pro Sprache in den Vortragsabstracts.

|      | Deutsch | Englisch | Andere |
|------|---------|----------|--------|
| 2016 | 14      | 2        | 2      |
| 2017 | 2       | 3        | 0      |
| 2018 | 11      | 3        | 5      |

Tabelle 7: Anzahl an Posterabstracts die mindestens eine Person mit dieser Sprache erwähnen.

|      | DE | EN | FR | IT | RU | Andere |
|------|----|----|----|----|----|--------|
| 2016 | 16 | 2  | 0  | 0  | 0  | 2      |
| 2017 | 2  | 3  | 0  | 0  | 0  | 0      |
| 2018 | 16 | 3  | 2  | 2  | 2  | 2      |

Tabelle 8: Anzahl an Personen pro Sprache in den Posterabstracts.

## Herkunftsland

Die Ergebnisse für das Herkunftsland sind ähnlich (Tabellen 9, 10, 11 und 12), wobei durch die Unterteilung der deutschsprachigen Länder (Deutschland, Österreich, Schweiz) die Anteile nicht-deutscher Personen etwas höher sind (37% bis 71%).

|      | DE | AT | CH | FR | GB | US | ES | IT | Andere |
|------|----|----|----|----|----|----|----|----|--------|
| 2016 | 11 | 3  | 0  | 2  | 0  | 2  | 0  | 0  | 8      |
| 2017 | 10 | 1  | 3  | 4  | 3  | 2  | 0  | 0  | 4      |
| 2018 | 12 | 2  | 0  | 3  | 6  | 2  | 2  | 2  | 7      |

Tabelle 9: Anzahl an Vortragsabstracts die mindestens eine Person mit diesem Herkunftsland. erwähnen.

|      | DE | AT | CH | FR | GB | US | ES | IT | NL | SE | AR | Andere |
|------|----|----|----|----|----|----|----|----|----|----|----|--------|
| 2016 | 15 | 4  | 0  | 2  | 0  | 3  | 7  | 0  | 0  | 0  | 2  | 6      |
| 2017 | 27 | 0  | 3  | 7  | 3  | 15 | 0  | 0  | 2  | 2  | 0  | 2      |
| 2018 | 19 | 6  | 0  | 11 | 9  | 7  | 4  | 3  | 0  | 0  | 0  | 7      |

Tabelle 10: Anzahl an Personen pro Herkunftsland in den Vortragsabstracts.

|      | DE | AT | GB | US | Andere |
|------|----|----|----|----|--------|
| 2016 | 10 | 3  | 2  | 0  | 2      |
| 2017 | 2  | 0  | 0  | 0  | 3      |
| 2018 | 10 | 0  | 0  | 2  | 8      |

Tabelle 11: Anzahl an Posterabstracts die mindestens eine Person mit diesem Herkunftsland erwähnen.

|      | DE | AT | GB | RU | US | FR | IT | CH | Andere |
|------|----|----|----|----|----|----|----|----|--------|
| 2016 | 12 | 3  | 2  | 0  | 0  | 0  | 0  | 0  | 2      |
| 2017 | 2  | 0  | 1  | 0  | 1  | 0  | 0  | 1  | 0      |
| 2018 | 15 | 0  | 0  | 5  | 2  | 2  | 2  | 2  | 4      |

Tabelle 12: Anzahl an Personen pro Herkunftsland in den Posterabstracts.

## Diskussion

Die Ergebnisse der Länder- und Sprachanalyse zeichnen ein positives Bild des Engagements mit Inhalten und Personen außerhalb Deutschlands, mit über 15 verschiedenen Herkunftsländern. Dies ist positiv, da aufgrund der Datenquelle eine Tendenz zu deutschsprachigen Themen zu erwarten ist.

Zugleich sind die Ergebnisse im Genderbereich katastrophal. Über einen Zeitraum von drei Jahren werden in insgesamt 342 Abstracts in 100 (29%) Männer erwähnt, aber nur in 12 Frauen (3.5%), wobei nur in 2 Abstracts (0.5%) ausschließlich weibliche Personen das Ziel der Studie bzw. die erwähnten Beispiele sind. Zwar stellen Abstracts mit Personenerwähnungen nur einen Teil der DH Forschung da, aber trotzdem zeigt die Genderverteilung eine Schieflage, welche den historischen Blickwinkel, der Frauen aus dem öffentlichen Diskurs verdrängt, verstärkt. Der Zugang zu den Outputs von Frauen ist zwar aufgrund der historischen Sozialstrukturen schwieriger, aber die Wahrscheinlichkeit, dass diese Verteilung das Vorhandensein potentieller weiblicher Studiensubjekte auch nur ansatzweise korrekt abbildet ist unwahrscheinlich. Besonders da Studien wie Fischer und Jäschke (2018) quantitativ zeigen, dass es hier hinreichend weibliche Studiensubjekte gäbe. Es ist wesentlich wahrscheinlicher,

dass hier ein kognitiver Bias existiert, dem die DH als Disziplin aktiv entgegen treten muss.

## Bibliographie

**Berry, David M. (2011):** *"The Computational Turn: Thinking about the Digital Humanities"*, in: *Culture Machine* 12: 1–22.

**Fiormonte, Domenico (2016):** *"Toward a Cultural Critique of Digital Humanities"*, in **Gold, Matthew K / Klein, L. (ed.):** *Debates in the Digital Humanities 2016*, University of Minnesota Press 438–458.

**Fischer, Frank / Jäschke, Robert (2018):** *"Liebe und Tod in der Deutschen Nationalbibliothek"*. in *DHd2018: Kritik der digitalen Vernunft* 261–266.

**Gallon, Kim (2016):** *"Making a Case for the Black Digital Humanities"*. in **Gold, Matthew K / Klein, L. (ed.):** *Debates in Digital Humanities 2016* 42–49.

**Liu, Alan (2012):** *"Where is cultural criticism in the digital humanities?"* in **Gold, Matthew K (ed.):** *Debates in the Digital Humanities*, University of Minnesota Press 490–509.

**McPherson, Tara (2012):** *"Why are the digital humanities so white? Or thinking the histories of race and computation"*. in **Gold, Matthew K (ed.):** *Debates in the Digital Humanities*, 139–160.

**Rhody, Lisa M. (2016):** *"Why I dig: Feminist approaches to text analysis"*. in **Gold, Matthew K / Klein, L. (ed.):** *Debates in Digital Humanities 2016* 536–539.

**Wernimont, Jacqueline (2013):** *Whence Feminism? Assessing Feminist Interventions in Digital Literary Archives*. in *DHQ: Digital Humanities Quarterly*, 7(1).