

# Die Zukunft der Digital Humanities: Eher „Computational“ und mehr als nur „Humanities“?

**Burghardt, Manuel**

burghardt@informatik.uni-leipzig.de  
Computational Humanities, Universität Leipzig,  
Deutschland  
ORCID: 0000-0003-1354-9089

## Eine Geschichte zweier Wissenschaftskulturen ... und die Rolle der Informatik

In seinem vielzitierten Essay „The Two Cultures“ erzählt Charles Percy Snow (1959) die Geschichte zweier Wissenschaftskulturen, der Naturwissenschaften und der Geisteswissenschaften, die so grundverschieden sind, dass sie eine große Kluft trennt. Snow (1959) stellt jedoch auch fest, dass der Zusammenprall zweier Kulturen kreative Chancen birgt. Malte

Rehbein (2016, S. 14) hat argumentiert, dass die Digital Humanities (DH) eine solche kreative Chance darstellen: „Hier treffen geisteswissenschaftliche Fragestellungen auf neue, vor allem computerbasierte Forschungsmethoden, die aus der anderen, der mathematisch-technischen ‚Kultur‘ entlehnt werden“. Ich bin allerdings skeptisch bei der Einordnung der DH in dieses Szenario zweier Wissenschaftskulturen, denn der digitale Teil der DH – was immer das auch sein mag (vgl. Burghardt, 2020; Luhmann & Burghardt, 2022) – kann sicherlich nicht einfach mit den Naturwissenschaften gleichgesetzt werden. Auch wenn die Naturwissenschaften seit jeher durch ihre empirisch-nomothetische Ausrichtung eine stärkere Affinität zur Nutzung digitaler Methoden haben (vgl. Lauer, 2013), so sind diese in der Regel nicht genuin naturwissenschaftlich, sondern zumeist aus der Informatik entlehnt.

Über die DH wird nun auch für die andere Seite, die *Humanities* der Zugriff auf diese digitalen Methoden ermöglicht und die Informatik entsprechend integriert. Diese Nähe der DH zur Informatik wird noch deutlicher, wenn man an frühere Bezeichnungen, wie etwa *Humanities Computing* (McCarty, 2005), oder neuere Entwicklungen, wie die *Computational Humanities*<sup>1</sup>, denkt.

Wenn wir uns die Rolle der Informatik in diesem Zweikulturen-Szenario einmal näher betrachten, dann wird schnell deutlich, dass sie keiner der beiden Seiten ange-

hört, sondern vielmehr den Status einer Querschnittswissenschaft hat. Dieses Selbstverständnis der Informatik findet sich auch im Positionspapier „Was ist Informatik“ der Gesellschaft für Informatik (2006) wieder<sup>2</sup>. Weiterhin liest man dort, dass die Informatik mit ihren Werkzeugen und Methoden häufig auch regelrecht neue Disziplinen hervorbringt. Genannt seien etwa die sogenannten Bindestrichinformatiken wie Bioinformatik, Medieninformatik, Medizininformatik, Wirtschaftsinformatik oder eben auch die DH<sup>3</sup>. Das Vermittlungspotenzial der Informatik zwischen den zwei akademischen Kulturen liegt aus meiner Sicht vor allem in der systematischen Verbreitung informatischer Konzepte und Denkfiguren in naturwissenschaftlichen sowie auch geistes- und kulturwissenschaftlichen Disziplinen. Auf diese Weise erlangen sie eine gemeinsame, informatische Zweitsprache, die ihnen die grundlegende Verständigung jenseits klassischer disziplinärer Grenzen ermöglicht. Im Kontext der DH wurde diese informatische Zweitsprache zuletzt unter dem Stichwort der Algorithmizität (Geiger 2023, Burghardt 2023) diskutiert. Allgemeiner gesprochen könnte man sie aber wohl auch als *computational literacy* umreißen.

Zusammengefasst kann man also sagen, dass der Informatik eine wesentliche Rolle beim transdisziplinären Brückenbauen zwischen den Wissenschaftskulturen zukommt. Über die DH haben die Geistes- und Kulturwissenschaften Kenntnis zentraler Konzepte der Informatik, etwa Algorithmen oder maschinelle Lernverfahren, und können nun auch mit vielen Bereichen der Natur- und Lebenswissenschaften auf einer grundlegenden methodischen Ebene kommunizieren. Genau hier liegt aus meiner Sicht erhebliches Potenzial für eine blühende Zukunft transdisziplinärer DH, die nicht länger ausschließlich auf ihren eigenen Bereich der „Humanities“ beschränkt sind, sondern über die gemeinsame Sprache der Informatik auch Fragestellungen, Methoden und Theorien der Naturwissenschaft in ihre Forschung einflechten können. Einige bestehende Forschungsbeispiele in diesem Dreiklang aus Humanities, Naturwissenschaft und Informatik stelle ich im zweiten Teil des Beitrags vor.

## Potenziele der Zusammenarbeit von Digital Humanities und Biologie

Um es nicht zu kompliziert zu machen, werde ich nachfolgend eine exemplarische Naturwissenschaft herausgreifen, die aus meiner Sicht besonders gut das Potenzial und die vielfältigen Möglichkeiten des Austauschs und der Zusammenarbeit mit den DH veranschaulicht. Es ist dies die Biologie und insbesondere auch die *Ökologie* – die sich speziell mit den Wechselwirkungen zwischen Organismen und ihrer Umwelt befasst.

### Environmental Humanities

Bevor ich einige konkrete Beispiele für ein Zusammenspiel von DH und Biologie vorstelle, möchte ich anmerken, dass insbesondere die Verbindung zwischen der Ökologie und den Geisteswissenschaften nicht völlig neu ist, sondern unter der Bezeichnung Environmental Humanities bereits eine längere Geschichte von mindestens 50 Jahren hat (vgl. Emmett & Nye, 2017). Eine wichtige Prämisse der EH ist, dass die Umweltkrise und damit zusammenhängende Probleme, wie z. B. die Klima- und Biodiversitätskrise, viel zu komplex sind, als dass sie von den Naturwissenschaften allein untersucht werden könnten. Wichtige Fragen, zu deren Beantwortung die Geisteswissenschaften beitragen können, sind: Wie sind wir in die aktuelle Krise gekommen und, was noch wichtiger ist, wie kommen wir da wieder heraus? Dementsprechend gibt es eine ganze Reihe von geisteswissenschaftlichen Ansätzen, die sich mit primär ökologischen Herausforderungen befassen, wie z. B. Literaturwissenschaftler\*innen, die sich für das Romangenre der Öko-Kritik und Öko-Romane interessieren, oder Philosoph\*innen und Sozialwissenschaftler\*innen, die sich für den Klimarassismus interessieren.

Und auch die Hinzufügung der Vorsilbe *digital* hin zu Digital Environmental Humanities (DEH) boomt seit einigen Jahren, wie eine Reihe von Aufsätzen und Sammelbänden belegt (vgl. etwa Jørgensen, 2014; Ryan & Hearn, 2023; Travis et al., 2023). Betrachtet man jedoch die digitalen Aspekte dieser bestehenden DEH-Ansätze, so sind diese meist in jenen Zweigen des Big Tent DH zu finden, in denen wir uns mit digitalen Medien, digitaler Kommunikation und anderen digitalen Phänomenen beschäftigen, die die Umwelt betreffen, oder auch mit Aspekten der Digitalisierung und der Erfassung und Archivierung von Umweltdaten über Crowdsourcing. Der Bereich der DH im Umweltkontext ist hingegen noch verhältnismäßig unterentwickelt, was meiner Meinung nach einige interessante Möglichkeiten birgt. Ich möchte nun im nächsten Teil des Beitrags den größeren konzeptionellen Rahmen der Environmental Humanities verlassen und auf einige spezifische Methoden aus der Bioinformatik eingehen, die seit vielen Jahren mit großem Erfolg in den DH genutzt werden.

### Bestehenden Digital Humanities-Methoden aus der Bioinformatik: Alignment-Algorithmen und Phylogenetische Bäume

Einige Methoden aus der Bioinformatik sind längst auch als Standardverfahren in den DH adaptiert worden. Zum Beispiel bei der Anwendung von Bioinformatik-Algorithmen, wie sie für die Gensequenzierung verwendet werden, auf das Textmining. Insbesondere *alignment*-Algorithmen kommen hier häufig zum Einsatz, um ähnliche Textstellen in unterschiedlichen Korpora zu identifizieren (vgl. etwa Vierthaler & Gelein, 2019). Bei dem anderen Beispiel handelt es sich eher um einen Theorieimport aus der Biologie in die Geisteswissenschaften. Konkret wird hier die Evolutionstheorie, mit ihren Methoden, bspw. phylogenetischen Bäumen, adaptiert (vgl. etwa Da Silva & Tehrani, 2016).

Der Grundgedanke von *cultural evolution* ist dabei, dass die menschliche Kultur selbst ein evolutionärer Prozess ist, der die wichtigsten darwinistischen Mechanismen von Variation, Wettbewerb und Vererbung aufweist (Mesoudi, 2011). Zwei weitere Fallstudien, die das Innovationspotenzial von ökologie-inspirierten DH-Projekten illustrieren, folgen im nächsten Abschnitt.

### Neuere Ansätze für ökologie-inspirierte Methoden in den Digital Humanities

#### Beispiel 1: „Forgotten Books“

Im Artikel "Forgotten books: The application of unseen species models to the survival of culture" von Kestemont et al. (2022) geht es um die Anwendung eines ökologischen Modells für unentdeckte Arten (*unseen species*), um die Menge der europäischen, mittelalterlichen Bücher zu schätzen, die im Laufe der Zeit verloren gegangen sind, zum Beispiel durch Bibliotheksbrände oder mutwillige menschliche Zerstörung. Zum Einsatz kommt dabei das Chao1-Modell, dass im Wesentlichen wie folgt funktioniert: Stellen Sie sich vor, Sie erforschen einen Wald, um herauszufinden, wie viele verschiedene Arten von Tieren dort leben. Sie stellen einige Kameras auf, um sie zu fotografieren. Nach einer Woche gehen Sie Ihre Fotos durch und stellen fest, dass Sie mehrere verschiedene Tiere entdeckt haben. Zuerst zählt man, wie viele verschiedene Arten von Tieren tatsächlich gesehen wurden. Als Nächstes schaut man sich an, wie viele dieser Tiere nur einmal gesehen wurden (*singletons* = seltene Tiere). Außerdem schaut man sich an, wie viele Tierarten man genau zweimal gesehen hat (*doubletons* = eher seltene Tiere). Das Chao1-Modell verwendet die Anzahl der *singletons* und *doubletons*, um die Gesamtzahl der verschiedenen Tiere zu schätzen. Die Idee dahinter ist, dass, wenn man viele Tiere nur einmal und einige wenige Tiere zweimal gesehen hat, es wahrscheinlich noch mehr Tierarten gibt, die man überhaupt nicht gesehen hat. Überträgt man diese Methode auf den Bereich der mittelalterlichen Bücher, so ist die Analogie, dass Werke eine Spezies sind und verschiedene Kopien eines Werks den Sichtungen verschiedener Spezies entsprechen. Durch die Anwendung des Chao1-Modells ist es Kestemont et al. (2022) möglich zu schätzen, wie viele mittelalterliche Bücher im Laufe der Zeit verloren gegangen sind und heute nicht mehr „in freier Wildbahn“ (d.h. in Archiven oder Sammlungen) beobachtet werden können.

#### Beispiel 2: „Biodiversity in Literature“

Während das vorherige Beispiel Methoden der Ökologie für die Geisteswissenschaften fruchtbar macht, so möchte ich nun noch ein Projekt präsentieren, welches DH-Methoden mit Biodiversitätsmetriken kombiniert, um insbesondere neue Erkenntnisse in der Biodiversitätsforschung zu erlangen (Langer et al., 2021). Bei der Erforschung der

biologischen Vielfalt gibt es ein großes Problem: Die Forscher haben in der Regel keinen Zugang zu systematischen Aufzeichnungen über die Situation der biologischen Vielfalt in der Vergangenheit. Da sie nicht in der Lage sind, in die Vergangenheit zu reisen und ihre ausgefeilten Biodiversitäts-Feldexperimente durchzuführen, brauchen sie einen *Proxy*, der verwendet werden kann, um die historische Biodiversität anzunähern. Die Idee dieses zweiten Beispiels ist es entsprechend, literarische Werke als Spiegel für die historische Biodiversität heranzuziehen. Dabei wird angenommen, dass die literarische Verwendung von biologischen Taxonbezeichnungen – das sind Pflanzen- und Tiernamen – mit dem gesellschaftlichen Bewusstsein für Biodiversität der jeweiligen Zeit korreliert ist. Als Korpus wurden hier ca. 16.000 Werken von rund 3.800 Autoren aus dem standardisierten Projekt Gutenberg (Gerlach & Font-Clos, 2020) verwendet. Alle Texte sind in englischer Sprache und wurden aus dem Zeitraum 1705 – 1969 gesammelt. Der Ansatz für die Taxonererkennung basiert im Wesentlichen auf einem großen Begriffslexikon, das über Datenbanken wie *WikiData* und *WikiSpecies* erstellt wurde. Nach dem automatischen Entfernen von Rechtschreibfehlern und Duplikaten enthält dieses Lexikon etwa 100.000 Einträge zu Tier- und Pflanzennamen.

Im Analyseteil kommen dann ökologische Metriken zum Einsatz. Zwei grundlegende Maße sind hier die *richness* und die *abundance*, die dem *type/token*-Konzept des *Natural Language Processing* sehr ähnlich sind. Die *richness* ist also die Anzahl der eindeutigen Taxonbezeichnungen, während die *abundance* die Gesamtzahl aller Taxonbezeichnungen ist, einschließlich der wiederholten Erwähnungen. Weiterhin gibt es verschiedene Diversitätsmaße wie etwa die *Shannon diversity*, die *richness* und *abundance* in Beziehung setzen. Eines der wichtigsten Ergebnisse dieser Studie ist, dass tatsächlich ein Anstieg der Biodiversität bis in die Mitte der 1830er Jahre und ein Rückgang danach gemessen werden kann. Langer et al. gehen davon aus, dass dieser Wechsellpunkt stark durch die beginnende Industrialisierung beeinflusst wurde, die zu einer enormen Entkopplung zwischen Natur und Mensch führte und uns u.a. in die aktuelle Biodiversitäts-Krise geführt haben könnte, da sich viele Menschen des potenziellen Artenreichtums der Natur gar nicht bewusst sind.

## Fazit

In diesem Artikel habe ich dargelegt, wie die Geisteswissenschaften über die Sprache der Informatik in gewinnbringender Weise mit den Naturwissenschaften kommunizieren können. Im Ergebnis können so einzelne bioinformatische Methoden wie etwa *alignments* oder phylogenetische Bäume für die Analyse von geisteswissenschaftlichen Textkorpora adaptiert werden. Es bleibt abzuwarten, welche weiteren Methoden, Modelle und Theorien aus naturwissenschaftlichen Disziplinen importiert werden können. Die beiden obigen Beispiele für neuere Ansätze ökologieinspirierter Methoden in den DH geben hier erste Hinweise.

Ich bin gespannt, welche Synergien sich künftig mit weiteren Naturwissenschaften, wie etwa der Geologie, der Meteorologie oder der Chemie, ergeben. Für Letztere gibt es bspw. bereits interessante Vorstöße, bei denen DH-Methoden genutzt werden um aus alchemischen Büchern historische Verfahren und Laborsettings der Chemie zu rekonstruieren (Lang et al., 2023).

A propos Laborsettings: der von Pawlicka-Deger (2020) ausgerufene *laboratory turn* in den DH schließt nahtlos an die Argumentation einer stärkeren Annäherung unterschiedlicher Wissenschaftskulturen an. Wie Hannaway (1986: 585) feststellt, besteht der Hauptzweck des naturwissenschaftlichen Labors in der Beobachtung und Manipulation der *Natur* mit Hilfe spezieller Instrumente, Techniken und Apparate, die sowohl manuelle Fähigkeiten als auch konzeptionelles Wissen über ihre Konstruktion und ihren Einsatz erfordern (Hannaway, 1986: 585). Ersetzt man den Begriff *Natur* im vorhergehenden Satz durch (*sozio-)*kulturelle Artefakte, so ergibt sich sofort eine grundlegende Agenda für DH-Labs: Hier werden Daten aus den Geistes- und Sozialwissenschaften so aufbereitet, transformiert und kodiert, dass sie mit speziellen Instrumenten und Verfahren – also Werkzeugen und Methoden aus den DH (statistische Modelle, Deep-Learning-Workflows, Algorithmen usw.) – einer Untersuchung zugänglich gemacht werden können.

Liegt die Zukunft unseres Faches also in einer noch stärkeren Betonung des Informatischen und einer gleichzeitigen Öffnung der Humanities, um synergetisch auch naturwissenschaftlichen Methoden und Fragestellungen einzubeziehen? Oder besteht durch solch eine Öffnung die Gefahr der Verwässerung – Identitätskrisen sind in den DH ja durchaus bekannt (vgl. etwa Terras, Nyhan & VanHoutte, 2016) –, hin zu generischen Konstrukten wie „Interdisciplinary Digital Sciences“<sup>4</sup>? Ich hoffe mit diesem Beitrag auf eine Fortsetzung der Diskussion zur Frage „Quo Vadis DH?“ (Motto der DHd 2024), befindet sich doch die Entwicklung des Faches aus meiner Sicht weiterhin stark *under construction* (Motto der DHd 2025).

## Fußnoten

1. Vgl. etwa die seit 2020 bestehende Konferenz *Computational Humanities Research* (CHR): <https://2024.computational-humanities-research.org/>
2. Diese Bezugnahme auf das GI-Positionspapier findet sich bereits bei Piotrowski (2014) und seinen Überlegungen zum Status der Informatik in den Digital Humanities.
3. Vgl. auch die GI-Fachgruppe „Informatik und Digital Humanities“, die sich dieser besonderen Schnittstelle widmet: <https://fg-inf.dh.gi.de/>
4. Dies ist der Titel eines entsprechenden Forschungszentrums an der TU Dresden: <https://tu-dresden.de/cids>.

## Bibliographie

- Burghardt, Manuel** (2020): Theorie und Digital Humanities – Eine Bestandsaufnahme. Digital Humanities Theorie. Abgerufen am 24. Juli 2024 von <https://dhtheorien.hypotheses.org/680>
- Burghardt, Manuel** (2023): Kritische Überlegungen zum Algorithmizitätsbegriff. Digital Humanities Theorie. Abgerufen am 24. Juli 2024 von <https://dhtheorien.hypotheses.org/1316>
- Da Silva, S. G., & Tehrani, J. J.** (2016). Comparative phylogenetic analyses uncover the ancient roots of Indo-European folktales. *Royal Society open science*, 3(1), 150645.
- Emmett, R. & Nye, D.** (2017). *The Environmental Humanities: A Critical Introduction*. MIT Press.
- Geiger, J.** (2023). „Algorithmizität“ – Begriffsanalyse und Plädoyer. Digital Humanities Theorie. Abgerufen am 24. Juli 2024 von <https://dhtheorien.hypotheses.org/1400>
- Gerlach, M., & Font-Clos, F.** (2020). A standardized Project Gutenberg corpus for statistical analysis of natural language and quantitative linguistics. *Entropy*, 22(1), 126.
- Gesellschaft für Informatik** (2006). Was ist Informatik? Positionspapier der Gesellschaft für Informatik. Online: <https://gi.de/fileadmin/GI/Hauptseite/Themen/was-ist-informatik-lang.pdf>
- Hannaway, O.** (1986). Laboratory Design and the Aim of Science: Andreas Libavius versus Tycho Brahe. In *Isis – Journal of the History of Science Society* 77(4), 585-610.
- Jørgensen, F. A.** (2014). The armchair traveler's guide to digital environmental humanities. *Environmental Humanities*, 4(1), 95-112.
- Kestemont, M., Karsdorp, F., de Bruijn, E., Driscoll, M., Kapitan, K. A., Ó Macháin, P., ... & Chao, A.** (2022). Forgotten books: The application of unseen species models to the survival of culture. *Science*, 375(6582), 765-769.
- Lang, S., Liebl, B., & Burghardt, M.** (2023). Toward a Computational Historiography of Alchemy: Challenges and Obstacles of Object Detection for Historical Illustrations of Mining, Metallurgy and Distillation in 16th-17th Century Print. In *CHR* (pp. 29-48).
- Langer, L., Burghardt, M., Borgards, R., Böhning-Gaese, K., Seppelt, R., & Wirth, C.** (2021). The rise and fall of biodiversity in literature: A comprehensive quantification of historical changes in the use of vernacular labels for biological taxa in Western creative literature. *People and nature*, 3(5), 1093-1109.
- Lauer, G.** (2013). Die Vermessung der Kultur. Geisteswissenschaften als Digital Humanities. In Heinrich Geiselberg & Tobias Moorstedt (Hrsg.): *Big Data. Das neue Versprechen der Allwissenheit*. Suhrkamp.
- Luhmann, J., & Burghardt, M.** (2022). Digital Humanities—A discipline in its own right? An analysis of the role and position of digital humanities in the academic landscape. *Journal of the Association for Information Science and Technology*, 73(2), 148-171.
- McCarty, W.** (2005). *Humanities Computing*. Basingstoke: Palgrave Macmillan.
- Mesoudi, A.** (2011). *Cultural Evolution. How Darwinian Theory Can Explain Human Culture and Synthesize the Social Sciences*. Chicago: The University of Chicago Press.
- Pawlicka-Deger, U.** (2020). The Laboratory Turn: Exploring Discourses, Landscapes, and Models of Humanities Labs. *DHQ: Digital Humanities Quarterly*, 14(3).
- Piotrowski, M.** (2014). Informatik – technischer Hilfsdienst der Digital Humanities? Book of Abstracts zum DHd-Workshop “Informatik und die Digital Humanities”, Leipzig.
- Rehbein, M.** (2016). Was sind Digital Humanities? *Akademie Aktuell* 01-2016, S. 13-17. Online: [https://badw.de/fileadmin/pub/akademieAktuell/2016/56/0116\\_04\\_b\\_Rehbein\\_V04.pdf](https://badw.de/fileadmin/pub/akademieAktuell/2016/56/0116_04_b_Rehbein_V04.pdf)
- Ryan, J., Hearn, L., & Arthur, P.** (2023). The digital environmental humanities (DEH) in the anthropocene: Challenges and opportunities in an era of ecological precarity. *Digital Humanities Quarterly* 17(3).
- Snow, C. P.** (1959). *The Two Cultures*. London: Cambridge University Press.
- Terras, M., Nyhan, J., & Vanhoutte, E. (Eds.).** (2016). *Defining Digital Humanities: A Reader*. Routledge.
- Travis, C., Dixon, D., Bergmann, L., Legg, R., & Crapsie, A.** (2023). *Routledge Handbook of the Digital Environmental Humanities*. Routledge.
- Vierthaler, P., & Gelein, M.** (2019). A BLAST-based, language-agnostic text reuse algorithm with a markus implementation and sequence alignment optimized for large Chinese corpora. In *Journal of Cultural Analytics*.