

## *Digital Humanities zwischen Informatik und Geisteswissenschaften<sup>1</sup>*

Die methodischen Probleme dieser Gegenüberstellung werden derzeit oft thematisiert. Soll hier ein neues Fach etabliert werden, das bei der Verteilung der Ressourcen von Studium und Lehre in Konkurrenz tritt zu den anderen Fächern der Geisteswissenschaften oder sollen neue Methoden aus dem Information Retrieval wie etwa Text- und Datamining, oder Methoden der Visualisierung in die geisteswissenschaftliche Grundausbildung integriert werden? Der Trend ist noch nicht klar absehbar, denn es gibt noch nicht einmal eine deutlich charakterisierende Benennung des neuen Gebietes, wie an den verschiedenen Bezeichnungen wie Digital Humanities vs. eHumanities vs. Computational Humanities vs. Digitale Geisteswissenschaften gut ersichtlich ist. Hier wie im Folgenden wird aus Gründen der praktischen Darstellbarkeit der Begriff ‚Digital Humanities‘ verwendet, um zu beschreiben, daß es sich um eine Entwicklung handelt, die die Schnittstelle zwischen der ganz wesentlich von der Informatik beeinflussten Digitalisierung einerseits und den Geisteswissenschaften andererseits betrifft. Im Folgenden werden aus der weitläufigen Diskussion drei Aspekte herausgegriffen: eine kurz Skizze der Rahmenbedingungen (I.), die im Hinblick auf die institutionelle Entwicklung diskutiert werden und die Auswirkungen auf die disziplinäre Struktur (II.). Die größte Herausforderung dieser Diskussion um die Digital Humanities liegt jedoch im methodischen Bereich (III.): geht es hier um die Gegenüberstellung von hypothesengeleiteter vs. datengetriebener Forschung, gerichteter vs. ungerichteter Forschung bzw. die grundlegende methodische Frage, ob es eine ungerichtete Forschung in den Geisteswissenschaften überhaupt geben kann.

### *I. Rahmenbedingungen für die Entwicklung der Digital Humanities in Deutschland*

Die Empfehlungen des Wissenschaftsrates (Empfehlungen des Wissenschaftsrates zu Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften v. 28.1.2011) haben für die weitere Entwicklung und strukturelle Ausgestaltung der Digital Humanities ein Rahmengerüst formuliert, das sich sehr stark an dem Begriff der Forschungsinfrastruktur ausrichtet: Angestrebt wird auf der Basis von thematisch fokussierten Forschungsfragen der Geisteswissenschaften die Verbindung von Forschungsinfrastruktur mit Forschungsfragen: als Anstoß und Motor sollen methodische Innovationen dienen, die verändernde Prozesse („projektförmiger Beginn von Infrastrukturen“) einleiten.<sup>2</sup>

In den Richtlinien der DFG für Forschungsinfrastrukturen ist nun demgegenüber eine völlig andere Vorstellung von Forschungsinfrastrukturen erkennen. In den Leitlinien (Basic Requirements

---

<sup>1</sup> Eine ausführliche Fassung dieses Beitrags ist unter dem Titel: „Zauberlehrling und Meister: Digital Humanities zwischen Informatik und Geisteswissenschaften?“ erschienen in: J. Kobes / K. Ruffing / W. Spickermann, 20 Jahre Arbeitsgemeinschaft Geschichte und EDV (AAGE, 2), Gutenberg 2013. Den Herausgebern sei hier herzlich gedankt für die Erlaubnis, diese gekürzte Fassung publizieren zu dürfen.

<sup>2</sup> Empfehlungen des Wissenschaftsrates zu Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften v. 28.1.2011, 24.

for Research Infrastructures in Europe) zeigt sich eine deutliche Orientierung an Infrastrukturen.<sup>3</sup> Hiernach sollen in das von der DFG aufgebaute Informationsportal der Forschungsinfrastrukturen diejenigen Institutionen aufgenommen werden, die ein anerkanntes wissenschaftliches und *technologisches* (sic!) Angebot bieten, den Zugang über einen transparenten Auswahlprozeß auf der Basis von wissenschaftlicher Qualität und Machbarkeit des Projekts regeln sowie über ein nachhaltiges Management verfügen. Neben den Kriterien für Qualitätssicherung und Management sind vor allem die Leitlinien für den Zugang zu Daten und Datenrepositorien auf Standardisierung, Öffentlichkeit und Nachhaltigkeit ausgerichtet. Der sehr viel offenere Begriff von Forschungsinfrastruktur des Wissenschaftsrates ist demgegenüber mehr auf die konkreten Anforderungen des Forschungsprozesses selbst ausgerichtet und hält fest, insbesondere für die Geistes- und Sozialwissenschaften, daß neben „Großgeräten auch Wissensressourcen wie Sammlungen, Archive, digitale Datenbanken und Datensammlungen als Forschungsinfrastrukturen zu betrachten“ seien.<sup>4</sup>

Das BMBF hat sich in seiner jüngst veröffentlichten Roadmap ebenfalls geäußert: „Forschungsinfrastrukturen im Sinne dieser Roadmap sind umfangreiche Instrumente, Ressourcen oder Serviceeinrichtungen für die Forschung in allen Wissenschaftsgebieten, die sich durch eine mindestens nationale Bedeutung für das jeweilige Wissenschaftsgebiet auszeichnen sowie durch eine lange Lebensdauer (in der Regel über 10 Jahre). Eine Abgrenzung gegenüber kleineren Infrastrukturen und Geräten erfolgt für Natur-, Bio-, Umwelt und Technikwissenschaften sowie die Medizin über die Höhe der Investitionskosten, die mindestens 15 Mio. Euro betragen sollten. Für andere Wissenschaftsgebiete wie die Geistes-, Sozial- und Wirtschaftswissenschaften wird auf diese Grenze verzichtet.“<sup>5</sup> Interessant ist, daß das BMBF bei der Bewertung der Projekte und Einrichtungen, die als Forschungsinfrastrukturen in die Roadmap aufgenommen wurden, zwischen einem wissenschaftsgeleiteten und einen wirtschaftlichen Bewertungsprozeß unterscheidet.<sup>6</sup> Insofern wundert es nicht, daß die Roadmap, die insgesamt 48 Projekte umfaßt, lediglich drei sozialwissenschaftliche aufführt und nur zwei geisteswissenschaftliche.<sup>7</sup>

Den geisteswissenschaftlichen Forscherinnen und Wissenschaftlerinnen selbst kommen in den Konzepten offensichtlich sehr unterschiedliche Rollen zu und es bietet sich daher an, diese verschiedenen Vorstellungen auch graphisch zu visualisieren. Ausgehend von den beiden zentralen Polen der Forschung und der Infrastruktur lassen sich drei Modelle unterscheiden:

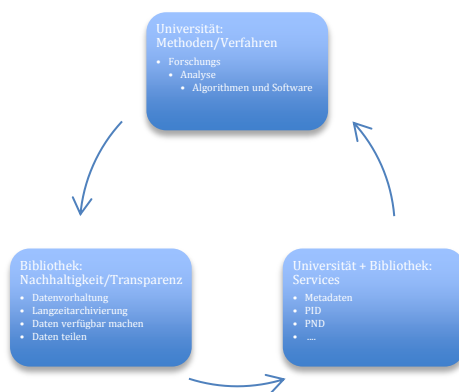
<sup>3</sup> <http://www.dfg.de/foerderung/programme/infrastruktur/wgi/forschungsinfrastrukturen/index.html> (22.4.2013)

<sup>4</sup> Empfehlungen des Wissenschaftsrats a.a.O. 5.

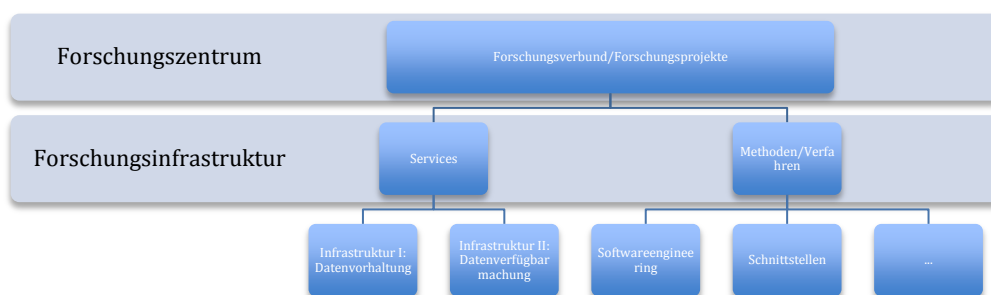
<sup>5</sup> Roadmap für Forschungsinfrastrukturen. Pilotprojekt des BMBF, 2013, 2. Presseerklärung am 29.04.2013. Abzurufen unter: <http://www.bundesforschungsministerium.de/de/12159.php> (2.5.2013).

<sup>6</sup> Roadmap für Forschungsinfrastrukturen. Pilotprojekt des BMBF, 3-4.

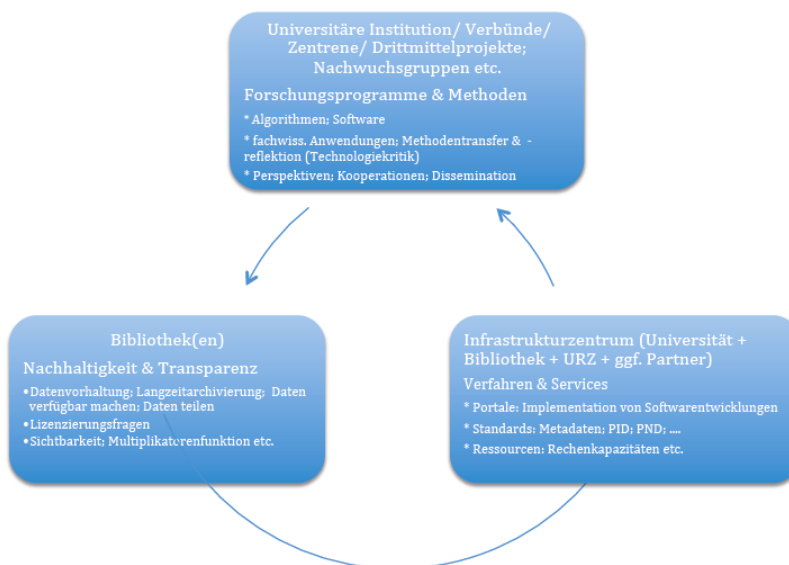
<sup>7</sup> Die sozialwissenschaftlichen sind: CESSDA: Council of European Social Science Data Archives (Verbund europ. Sozialwissenschaftlicher Datenarchive), ESS: European Social Survey (Querschnitterhebung zu Werten und Einstellungen und SHARE: Survey of Health, Ageing and Retirement in Europe (Längsschnitterhebung zur gesundheitlichen, ökonomischen und sozialen Situation älterer Bürger), die geisteswissenschaftlichen sind: CLARIN: Common Language Resources and Technology Infrastructure und DARIAH: Digital Research Infrastructure.



Modell 1: Forschungsinfrastruktur als Zyklus des Austauschs und der Kooperation zwischen Universitäten und Bibliotheken



Modell 2: Forschungsinfrastruktur als eigener „Anbieter“ von Services, Methoden und Verfahren für einzelne Forschungsverbände, Projekte und andere „Abnehmer“



Modell 3: Forschungsinfrastruktur als dynamischer Prozeß, der auf dem Zyklus des Austauschs der vorhandenen Infrastrukturen (Bibliotheken, Rechenzentren) aufsetzt und forschungsgetrieben die Methodenentwicklung vorantreibt

Das Anbieter-Abnehmer-Verhältnis in Modell 2 stellt den Aufbau und die nachhaltige Sicherung der Infrastruktur in den Vordergrund. In Modell 1 und Modell 3 hingegen sind die Wissenschaftlerinnen in einen dynamischen Austausch eingebunden, der den Forschungsprozeß mit seinen differenzierten, methodischen Verfahrensweisen und seine Anforderungen als Motor sieht, aber auch die Nachhaltigkeit der Ergebnissicherung und Datenzugänglichkeit einbindet und insbesondere auch der Vielfalt der geistes- und sozialwissenschaftlichen Fachkulturen gerecht wird. Der entscheidende Unterschied zwischen diesen beiden Richtungen ist die Zielvorstellung: Soll die Infrastruktur für sich stehen, mehr oder weniger unabhängig von den – sich durchaus in manchmal schnellen Zyklen wandelnden – Anforderungen der Fachkulturen oder soll die Infrastruktur – eben als echte Forschungsinfrastruktur – Ausdruck eines Prozesses sein, in dem die Anforderungen der Fachkulturen, hier also der Geisteswissenschaften, den leitenden Rahmen vorgeben?

Welche dieser Richtungen sich langfristig durchsetzen wird, ist noch keineswegs ausgemacht und wird einerseits davon abhängen, wie sich die Digitalisierung in den einzelnen Disziplinen entwickeln wird, andererseits aber auch davon, wie sehr die Wissenschaftlerinnen bereit sind, sich in diesem strukturpolitischen Feld für den Vorrang der geisteswissenschaftlichen Art der Forschung mit ihren spezifisch Leitkategorien einzusetzen.<sup>8</sup>

## *II. Auswirkungen und Entwicklung in der disziplinären Praxis*

Derzeit sind Bestrebungen zu beobachten, Digital Humanities als neues Fach mit einem eigenen Curriculum zu etablieren.<sup>9</sup> Im Vergleich zu den traditionellen Fächern, die z. T. auf eine jahrhundertalte Fachtradition zurückblicken, stellt sich für die Digital Humanities, wenn sie denn als eigenes, universitäres Fach begründet werden sollen, vorrangig die Frage nach den Inhalten und Methoden. So können die Inhalte der verschiedenen, bereits an einigen Universitäten angebotenen Digital Humanities-Studiengänge aus praktisch jedem Fach der Geistes- und Sozialwissenschaften stammen und darauf aufbauend soll vermittelt werden, wie Daten modelliert, Fragestellungen formalisiert und operationalisiert werden können.<sup>10</sup>

<sup>8</sup> European Science Foundation: Research Infrastructures in the Digital Humanities. Science Policy Briefing 42, Straßburg 2011, URL:

[http://www.esf.org/fileadmin/Public\\_documents/Publications/spb42\\_RI\\_DigitalHumanities.pdf](http://www.esf.org/fileadmin/Public_documents/Publications/spb42_RI_DigitalHumanities.pdf).

<sup>9</sup> <http://www.dig-hum.de/aktuelles/digital-humanities-curricularentwicklungen> (1.5.2013).

<sup>10</sup> Die inhaltliche Beschreibung in der Broschüre „Digitale Geisteswissenschaften“, hrsg. vom CCeH, Universität zu Köln (2011), erstellt im Rahmen des Projekts DARIAH-DE, gefördert vom BMBF, zeigt, daß die fachspezifischen Inhalte an jedem Studienort anders sind, man könnte fast sagen: nahe an der Beliebigkeit. Auch die Struktur ist höchst unterschiedlich: So bietet die Universität zu Köln ein DH-Studienangebot BA/MA „Informationsverarbeitung“ an, der sowohl eine Schwerpunktsetzung im Bereich Sprachtechnologien als auch eine bei den Kultur- und Medientechnologien erlaubt. Die Universität Würzburg bietet einen BA-Studiengang Digital Humanities als Teil des Lehrangebots der Philosophischen Fakultät an. Im Studiengang sind Lehrende aus mehreren Fächern der Fakultät (Germanistik, Alte Geschichte, Landesgeschichte, Klassische Philologie) tätig. Organisatorisch betreut wird der Studiengang vom Lehrstuhl für Computerphilologie und Neuere Deutsche Literaturgeschichte. An der Universität Bamberg sind die Digital Humanities Teil des BA Angewandte Informatik. An an-

Damit definieren sich die Befürworter einer Fachrichtung ‚Digital Humanities‘ als Anwender von Methoden, die aus dem informationswissenschaftlichen Bereich in die Geistes- und Sozialwissenschaften transferiert werden. Bei den Digital Humanities handelt es sich also um ein Anwendungsfach mit einem breiten Methodenbereich: Anwendung in einem außerordentlich großen Bereich der Geistes- und Sozialwissenschaften, jedoch ohne eigenes inhaltliches Profil. Daher läge es nach Ansicht der Verfasserin dieses Beitrags sehr viel näher, diese Anwendung der Methoden in die Ausbildung der ‚klassischen‘ Fächer der Geistes- und Sozialwissenschaften zu integrieren, bspw. in den Grundlagenbereich (in die sog. Propädeutik).

Als Antwort wird die Einrichtung von “Research Technology Centers” bzw. „Digitalen Forschungszentren“ empfohlen. Allerdings stellt sich hier die Frage, welche Rolle Zentren dieserart neben den Universitätsbibliotheken, Rechenzentren und Forschungszentren (mit fachspezifischen Forschungsprogramm) einnehmen sollten: diejenige eines weiteren Dienstleistungs- und Servicezentrums, eines Infrastrukturzentrums oder diejenige eines Forschungszentrums? Für den erstgenannten Bereich existieren an Universitäten mit den Bibliotheken und Rechenzentren bereits gut ausgeprägte Infrastrukturen, die auch aus gutem Grund auf Dauer gestellt sind. Für den zweitgenannten Bereich wiederum würde die Betonung auf Forschung die fachdisziplinäre (auch die fach-interdisziplinäre) Ausrichtung in den Vordergrund stellen (analog den Sonderforschungsbereichen oder DFG-Forschungszentren), die immer entscheidend von einer Forschungsfrage bzw. einem größeren Forschungsprogramm geleitet ist und somit naturgemäß zeitlich begrenzt wird.

Einen bedenkenswerten Vorschlag, um genau diesen Hiat zu überwinden, haben P. Sahle und S. Kronenwett kürzlich unterbreitet, den sie am Beispiel des Data Center for the Humanities (DCH) der Universität zu Köln beschreiben:<sup>11</sup> Sie gehen von einem Modell funktionaler Schichten aus (Schicht 1: Datensicherung und Datenvorhaltung, Schicht 2: Schnittstellen, Dienste, Werkzeuge, Schicht 3: Präsentation und Oberflächen), dessen Funktionen sowohl die Daten und Ergebnisse einzelner Projekte aufnehmen können, als auch für Fachbereiche mit ihren Spezifika sowie zum zentralen Nachweis über alle Inhalte, Projekte und Datensammlungen des DCH durch ein zentrales Portal, das Gesamtportal des Datenzentrums, zur Verfügung stehen sollen. In dieser Konzeption sollen letztlich Archiv, Bibliothek, Museum und Werkstatt in einer Forschungsinfrastruktur zusammenfließen.

### **III. Methoden: Serendipity oder die Systematisierung des Zufalls**

Wenn wir akzeptieren, daß es ein von der Kultur unabhängiges menschliches Handeln nicht gibt, „alle menschlichen Handlungen kulturelle Akte [sind], durch die Menschen sich selbst begreifen

---

deren Universitäten werden Zertifikatskurse im Rahmen geisteswissenschaftlicher Studiengänge angeboten (in Leipzig z.B. Zertifikatskurse begleitend zum Masterstudiengang Klassische Antike).

<sup>11</sup> P. Sahle/ S. Kronenwett, „Jenseits der Daten: Überlegungen zu Datenzentren für die Geisteswissenschaften am Beispiel des Kölner ‚Data Center for the Humanities‘“. LIBREAS. Library Ideas, 23 (2013). <http://libreas.eu/ausgabe23/09sahle/> (8.12.2013).

und als kulturelle Wesen verorten“,<sup>12</sup> muß vor aller Anwendung von Methoden der Digital Humanities die Frage nach Strukturen, Objekten und Bedingungen des Wissens, nach Wahrnehmungsdispositiven und Textstrategien stehen.

Dies berührt im Grundsätzlichen die Frage nach den Daten und die Veränderung des Blicks auf die Daten. Unter dem Schlagwort ‚distant reading‘ sind neue Möglichkeiten diskutiert worden, aus der schier unermesslichen Menge von Daten durch die Anwendung algorithmenbasierter Auswertung mit den Methoden etwa des Textmining, des Clusterings oder des Topic Modeling neue Zusammenhänge aus sehr großen Text- und Datenmengen zu erkennen.<sup>13</sup> Komplexitätsreduktion, Visualisierung und exploratives Experimentieren haben aber auch auf ganz neue Fragen geführt, von denen nicht zuletzt die Qualität der zugrunde gelegten Daten eine wesentliche ist.<sup>14</sup>

Aber auch das, heute schon posthermeneutisch zu nennende Textverständnis ändert sich. Dies soll hier am Beispiel der Arbeiten, die im Rahmen des BMBF-geförderten Projekt eAQUA durchgeführt wurden,<sup>15</sup> exemplarisch erläutert werden, da in eAQUA am Beispiel der Adaption von Methoden aus dem Textmining auch methodische Überlegungen im Bereich der explorativen Suche entwickelt worden sind.<sup>16</sup> Ein methodisch besonders interessanter Weg hat sich aus der Kookkurrenzsuche ergeben, der jedoch auf anderen Vorüberlegungen basiert als die herkömmliche, informationswissenschaftliche Anwendung der explorativen Suche.<sup>17</sup> Die Kookkurrenzsuche kann semantische Zusammenhänge anzeigen, die weder offensichtlich und geläufig sind, noch durch andere, herkömmliche oder an-

<sup>12</sup> S. Kramer, Macht und Repräsentation. Die Autorität und Medialität interkultureller Diskurse, in: Furtwängler, Frank et. al. (Hg.): *Zwischen-Bilanz. Eine Internet-Festschrift zum 60. Geburtstag von Joachim Paech*. <http://kops.ub.uni-konstanz.de/static/paech/zdk/main.htm> (15.5. 2013) setzt die Medienanalyse in den interkulturellen Kontext und erläutert, wie Medien selbst als dynamischer und konstitutiver Teil von Kultur betrachtet werden und vor allem die „Bedeutung ... der technisch-apparativen audiovisuellen Medien im Kontext der europäischen Kulturen der Moderne, welche diese Medien erst hervorgebracht haben, hängt dagegen eng zusammen mit der photographischen Illusion des „Das, was man sieht, kann man auch glauben“. Vgl. auch G. Lauer, Die Vermessung der Kultur. Geisteswissenschaften als Digital Humanities. Geiselberger, H., & Moorstedt, T. (Eds.). *Big Data. Das neue Versprechen der Allwissenheit*. Berlin, 2013, 99-116, mit einem Überblick zur Entwicklung der Digital Humanities unter der Prämisse, daß „die digitalen Geisteswissenschaften die Kultur im buchstäblichen Sinn neu“ vermessen (a.a.O. 109).

<sup>13</sup> Die Formulierung ‚distant reading‘ stammt von F. Moretti, *Graphs, Maps, Trees*, London/New York 2007; zu der Methodendiskussion vgl. v.a. L. Manovich, *The language of new media*, MIT Press 2007 und G. Crane, *What Do You Do with a Million Books?* D-Lib Magazine, Vol. 12/3 March 2006 (<http://www.dlib.org/dlib/march06/crane/03crane.html>, 26.4. 2013).

<sup>14</sup> Zu der Kritik an dem Google Ngram-Viewer: D. Sullivan <http://searchengineland.com/when-ocr-goes-bad-google-ngram-viewer-the-f-word-59181> (26.4.2013). Zu der für die Altertumswissenschaften nicht minder wichtigen Frage, welche Edition in der jeweils verwendeten Datenbank enthalten ist: Ch. Schubert, *Klassische Altertumswissenschaften, Digital Classics und das Feld des „new media encounter“* in: W. Fach, F. Kolovou, S. Rieckhoff, B. Streck, *Working Papers Contested Order No.3*, Sept 2011 ([http://www.uni-leipzig.de/~order/content/index.php?option=com\\_content&task=view&id=161&Itemid=151](http://www.uni-leipzig.de/~order/content/index.php?option=com_content&task=view&id=161&Itemid=151), 26.4.2013).

<sup>15</sup> S. dazu [www.eaqua.net](http://www.eaqua.net) und die Publikationen unter: [http://www.uni-leipzig.de/~order/content/index.php?option=com\\_content&task=view&id=161&Itemid=151](http://www.uni-leipzig.de/~order/content/index.php?option=com_content&task=view&id=161&Itemid=151) (26.4.2013).

<sup>16</sup> Grundsätzlich dazu G. Heyer, D. Keim, S. Teresniak, D. Oelke, *Interaktive explorative Suche in großen Dokumentbeständen* (Datenbank Spektrum (2011) 11:195–206; DOI 10.1007/s13222-011-0072-4). Zur Definition a.a.O. 196: „Mit dem Begriff der explorativen Suche werden in der Literatur alle Formen von Suchanwendungen bezeichnet, bei denen der Nutzer iterativ verschiedene Arten von Informationen zusammentragen und bewerten muss und seine Suchanfragen entsprechend seinem Informationsbedürfnis und den bisher erhaltenen Informationen ggf. neu formulieren kann.“ Die Suche erfolgt in zwei eingrenzenden Schritten: über die Eingrenzung von Zeiträumen und Texten und über die Auswahl aufgrund der Anzeige der Kontextvolatilität.

<sup>17</sup> Ausf. dazu Schubert 2011 (wie Anm.14), aus dem der folgende Abschnitt im Wesentlichen entnommen ist.

derweitig etablierte Suchstrategien bzw. –methoden erkannt werden können, wie sie sich aus der Verwendung von Lexika, Nachschlagewerken, Konkordanzen, Indizes, Suchprogrammen mit einfacher oder Bool'scher Wortsuche entwickelt haben. Kookkurrenzen, die im gesamten Corpus z.B. der antiken griechischen Literatur lediglich ein- bis dreimal auftreten, sind, wenn sie nicht gerade in den Werken der sehr bekannten, d.h. viel gelesenen Autoren, enthalten sind, ansonsten nur durch den Zufall – den glücklichen Zufall! – zu finden.

Das Phänomen des ‚glücklichen Zufalls‘, d.h. in diesem Fall einer wissenschaftlichen Entdeckung, die unbeabsichtigt gelingt, weil ursprünglich etwas ganz anderes gesucht wurde, ist wissenschaftshistorisch von Robert K. Merton als ‚serendipity‘ wiederentdeckt worden.<sup>18</sup> Ursprünglich von Horace Walpole am Beispiel der ‚Three Princes of Serendip‘ (Serendip als Bezeichnung für Sri Lanka) als Begriff entwickelt, ist er aber erst durch die wissenschaftshistorische Untersuchung von Merton konzeptionalisiert worden. Das mittlerweile breit rezipierte und mit vielen Untersuchungen unterlegte Konzept läßt sich systematisch als ‚serendipity pattern‘ im Anschluß an Merton in dreierlei Hinsicht konkretisieren:<sup>19</sup>

- als nicht vorhergesehenes Ereignis oder Ergebnis, etwa ein glückliches Nebenergebnis, eine unerwartete Beobachtung, die sich auch im Hinblick auf den ursprünglichen, wissenschaftlichen Erwartungshorizont auswirkt

- als überraschende Beobachtung, die deshalb überrascht, weil sie mit den bisherigen Theorien, Fakten und Meinungen nicht übereinstimmt und daher Neugier und weitere Untersuchungen stimuliert

- als unerwartete Beobachtung bzw. unerwartetes Ergebnis, das jedoch eine strategische Wirkung entfaltet, da es Auswirkungen auf allgemeine und generelle Theorien, Interpretationen und Meinungen hat.

Diese Kombination aus Überraschung, nicht-beabsichtigtem Ergebnis und strategischer Bedeutung erweist ‚serendipity‘ als einen speziellen Zufall, der sich vor allem durch seine Wirkung und Anschlußfähigkeit auszeichnet, aber auch grundsätzlich auf einer völlig anderen Prämisse ruht. Das informationswissenschaftliche Verständnis der explorativen Suche geht demgegenüber von einer analytisch definierten Suchstrategie aus, die mit einer facettierten Suche kombiniert wird.<sup>20</sup> Darin kommt nun auch ein grundlegender Unterschied im Anspruch der Aussagefähigkeit einer Suchmethode zum Ausdruck: Zwar wird seit Jahren, im Grunde seit Jahrzehnten, in den Naturwissenschaften breit diskutiert, was Serendipity als ‚glücklicher Zufallsfund‘ eigentlich bedeutet: die Entdeckung des Penicillins durch Alexander Fleming ist der vielleicht berühmteste Fall einer solchen Entdeckung, deren Charak-

---

<sup>18</sup> R. K. Merton, *Social Theory and Social Structure*, New York 1968=1957; ders., E. Barber, *The Travels and Adventures of Serendipity*, PUP 2006.

<sup>19</sup> Merton 1968 (wie Anm.18) 157-162.

<sup>20</sup> Heyer et al. a.a.O. (wie Anm.16).

teristik darin liegt, daß etwas anderes als das eigentlich Gesuchte gefunden wird.<sup>21</sup> Doch daß Serendipity nicht unbedingt etwas Zufälliges sein muß,<sup>22</sup> ist auch längst gesehen worden.<sup>23</sup>

Die Kookkurrenzsuche kann, wie in der Arbeitsgruppe des Projektes eAQUA für mehrere Fälle gezeigt, auf solche unbeabsichtigten, weil nicht intendierten bzw. nicht zielgerichtet gesuchten ‚Zufallsfunde‘ führen.<sup>24</sup> Einer dieser ‚Zufallsfunde‘ war beispielsweise die Kookkurrenz von νομάδας (nomadas [Akk.]: = Nomaden) und Ἀθίδος (Atthidos [Gen.]: = Werk eines Autors über Athen): Diese Kookkurrenz verweist auf zwei überhaupt nicht miteinander in inhaltlicher Verbindung stehende Dinge, die in einem Kontext stehen – Nomaden und die Geschichte der Stadt Athen. Die Erwähnung beider zusammen findet sich ein einziges Mal in der antiken Literatur und ist später lediglich zweimal von anderen Autoren zitiert worden. Aufgrund dieses seltenen Vorkommens und des auf den ersten Blick so unwahrscheinlichen inhaltlichen Kontextes – denn was haben Nomaden in Athen zu suchen, in einer Stadt, die sich als hochkultureller Mittelpunkt der griechischen Welt sah? – ist der Zusammenhang aus dem Originalzitat von dem Herausgeber der Fragmente der griechischen Historiker, Felix Jacoby, sogar athetiert worden, also als vermeintlich falsch ‚gelöscht‘ worden.<sup>25</sup>

Die Übersicht der Kookkurrenzen läßt jedoch aus der Überraschung, die das Auftauchen eines völlig unerwarteten Zusammenhangs andeutet, zusammen mit dem semantischen Kontext, wie ihn die graphische Visualisierung der Kookkurrenzergebnisse anzeigt, die berechtigte Vermutung werden, daß hier eine inhaltlich plausible Argumentation vorliegt, die der antike Autor in einen größeren Zusammenhang der attischen Geschichte eingebunden hat. Im Ergebnis hat sich dann aus der detaillierten Analyse gezeigt, daß sich dieser Gedanke bei einer ganzen Gruppe von antiken Historikern, den sog. Atthidographen, in ein ganz anderes als bisher angenommenes Geschichtsbild integriert hat.<sup>26</sup>

Die Suche nach den seltenen Kookkurrenzen hat sich mittlerweile bewährt und gezeigt, daß die ‚seltenen Ereignisse‘ die vielversprechenderen Kandidaten für das Auffinden neuer, interessanter Zusammenhänge sind im Vergleich zu den statistisch häufigeren.<sup>27</sup> Denn versierte Fachwissenschaftlerinnen kennen ihre Quellentexte in der Regel recht gut, werden also von den aufgrund der statistischen Häufigkeit angezeigten Zusammenhängen kaum überrascht sein. Dieser Befund zeigt eine inte-

<sup>21</sup> R. Campa, Making Science by Serendipity. A review of Robert K. Merton and Elinor Barber’s *The Travels and Adventures of Serendipity*, *Journal of Evolution and Technology*, 17/1, 2008, 75-83, <http://jetpress.org/v17/campa.htm> (27.5.2013).

<sup>22</sup> M. de Rond, *The Structure of Serendipity*, Judge Business School, University of Cambridge, Working Paper No.7, 2005, [https://www.jbs.cam.ac.uk/research/working\\_papers/2005/wp0507.pdf](https://www.jbs.cam.ac.uk/research/working_papers/2005/wp0507.pdf) (27.5.2013), weist daraufhin, daß auch dem von T. Kuhn geprägten Modell des ‚Paradigmenwechsels‘ genau dieses Phänomen der Serendipity zugrunde liegt.

<sup>23</sup> De Rond 9 (wie Anm. 22).

<sup>24</sup> Schubert, *Neue Methoden der geisteswissenschaftlichen Forschung. Eine Einführung in das Portal eAQUA* (zus. mit G. Heyer), 2010 und dies., Detailed description of eAQUA search portal, in: Working Papers Contested Order No.3, Sept 2011 beide abzurufen unter: ([http://www.uni-leipzig.de/~order/content/index.php?option=com\\_content&task=view&id=161&Itemid=151](http://www.uni-leipzig.de/~order/content/index.php?option=com_content&task=view&id=161&Itemid=151)) sowie: dies., Perikles und Thukydides, Sohn des Melesias: Der Kampf um die politische Vorherrschaft als Ausdruck konkurrierender Konzepte, in: *Museum Helveticum* 2008, 129-152

<sup>25</sup> Schubert, Detailed Description (wie Anm.24).

<sup>26</sup> Schubert, *Formen der griechischen Historiographie: Die Atthidographen als Historiker Athens*, in: *Hermes* 2010, 259-275.

<sup>27</sup> Schubert, Detailed Description (wie Anm.24).



ressante Parallele zu einem von Jaspers hervorgehobenen Phänomen, das er folgendermaßen beschrieb:<sup>28</sup> „Das Höchstdifferenzierte ist das Seltene. Insofern ist das Seltene, nicht die Kuriosität, sondern als klassischer, als extremer, als völlig entwickelter Fall, geradezu der Orientierungspunkt für die Erkenntnis.“

Das bedeutet, daß man den Faktor ‚Zufall‘ hier durchaus nicht für trivial oder selbstevident halten kann, oder wie etwa Popper, der den ‚Kontext der Entdeckung‘ für völlig uninteressant und unbedeutend hielt (Popper 1968, 31-32):<sup>29</sup>

Die Suche nach den ‚seltenen Ereignissen‘ ist demnach hier durchaus als Fall der Fähigkeit „matching pairs“ zu sehen oder, wie es vielleicht exakter zu formulieren wäre: einer systematisierten ‚serendipity‘. Doch auch die Einschränkung von Van Andel (1994, 635) gilt noch immer:<sup>30</sup> „For it obviously requires a theoretically sensitized observer to detect the universal in the particular.“ Für den Einsatz der Kookkurrenzsuche läßt sich hieraus als methodisches Prinzip ableiten, in der speziellen Form der explorativen Suche die seltenen Kookkurrenzen zu betrachten, um genau solche ungewöhnlichen, seltenen und bisher von der Forschung nicht gesehenen oder für abwegig gehaltenen Zusammenhänge aufzudecken.

#### **IV. Fazit**

Die Digitalisierung der Geisteswissenschaften wird zu neuen Forschungsinfrastrukturen führen, wahrscheinlich auch – je nach Positionierung der Digital Humanities – zu erheblichen Auswirkungen auf die disziplinären Strukturen. Methoden und Theorien können davon profitieren, und dies durchaus sowohl auf der Seite der Geisteswissenschaften wie der der Informatik. Die mathematische Handlungsvorschriften, die den Suchalgorithmen, Filtertechnologien, Sortiermöglichkeiten zugrunde liegen, verweisen darauf, daß es sich bei der Anwendung dieser Handlungsvorschriften gar nicht um datengetriebene Forschung handeln kann, sondern immer um hypothesengetriebene Forschung und/oder zweckgerichtete Anwendung, die aus Daten erst Information werden läßt. Diesen Zusammenhang von beiden Seiten aus anzugehen, könnte der Informatik die Perspektive einer Kulturalisierung und den Geisteswissenschaften über die Digital Humanities diejenige einer methodischen Erweiterung geben.

Charlotte Schubert, Leipzig

---

<sup>28</sup> K. Jaspers, *Allgemeine Psychopathologie*, Springer-Verlag Berlin/Heidelberg/New York (9.unveränd.Aufl.) 1973, 11.

<sup>29</sup> K. Popper, *The Logic of Scientific Discovery*, New York 1968, 31-32.

<sup>30</sup> P. Van Andel *Anatomy of the Unsought Finding. Serendipity: Origin, History, Domains, Traditions, Appearances, Patterns and Programmability*, in: *The British Journal for the Philosophy of Science* 45, 1994, 631-648, hier: 635.