

Datenbanken in der Alten Geschichte: Beobachtungen aus der Alten Welt

Andreas Hartmann

Zusammenfassung Behauptungen, dass der zunehmende Einsatz digitaler Werkzeuge in der Alten Geschichte mit einem methodischen Paradigmenwechsel korreliert, sind mit Skepsis zu betrachten. In vielen Fällen werden lediglich bewährte Herangehensweisen automatisiert. Aufgrund des ohnehin im Vergleich mit neueren Epochendisziplinen überschaubaren Quellenmaterials wird auch der Vorteil nahezu unbegrenzter Skalierbarkeit nicht im Sinne eines distant reading methodisch wirksam. Ungelöste methodische Probleme bleiben bislang die Sicherstellung textkritischer Standards in der digitalen Welt sowie die nachhaltige Bereitstellung von Forschungsdaten und digitalen Publikationen.

Abstract Allegations that the increasing use of digital tools in Ancient History correlates with a paradigm shift in methodology are to be viewed with skepticism. In many cases, only proven approaches are automated. Due to the source material, which is in any case manageable in comparison with recent epochal changes in other scholarly areas, the advantage of almost unlimited scalability do not become methodically effective in the sense of allowing for a distant reading. Unresolved methodological problems have so far included the securing of text-critical standards in the digital world as well as the sustainable provision of research data and digital publications.

Keywords methodology, paradigm shift

Die Diskussion über Potential und Gefahren elektronischer Verfahren in den Geisteswissenschaften ist so alt wie diese Verfahren selbst. Soweit sie die Grundsatzfrage nach dem Verhältnis qualitativer und quantitativer Forschungsmethoden berührt, reichen die vorgebrachten Argumente sogar noch weiter zurück.¹ Eine strikte Abwehrhaltung gegenüber allem Digitalen dürfte heute kaum noch anzutreffen sein und wäre praktisch auch kaum noch durchzuhalten. Dennoch schwelt die Debatte weiter, nicht zuletzt, weil mit dem Methodenstreit auch handfeste Verteilungskämpfe um Stellenzuweisungen und Mittel der Forschungsförderung verbunden sind.

1 Vgl. Röhle (2014), bes. 158.

Wer über das Potential elektronischer Datenbanken in der Alten Geschichte nachdenkt, tut daher gut daran, zunächst einmal den eigenen Standort zu klären, zum einen um selbst die Reichweite der eigenen Argumente besser einschätzen zu können, zum anderen um auch dem Leser eine kritische Kontextualisierung zu ermöglichen.

Die folgenden Überlegungen sind primär aus der Perspektive eines Althistorikers verfasst. Auf Bereiche wie die (Korpus-)Linguistik oder auch die Archäologie sind sie daher nicht übertragbar. Es liegt auf der Hand, dass informatischer Datenverarbeitung dort, wo es auf quantitative Auswertungen großer Textmengen mit dem Ziel einer Analyse der sprachlichen Gestalt (d.h. nicht der Erforschung der in den Texten referenzierten Sachverhalte) oder auf die Anwendung naturwissenschaftlicher Methoden der Materialanalyse und Bodenerkundung ankommt, eine elementare Bedeutung zukommt und Möglichkeiten des Erkenntnisgewinns entstehen, die es ohne sie nicht gäbe.

Meine Perspektive ist ferner die eines Historikers, der in seiner eigenen Forschung in hohem Maße, ja sogar vorzugsweise digitale Ressourcen verwendet und seit etlichen Jahren auch bemüht ist, solche Ressourcen für die wissenschaftliche Propädeutik zu erschließen.² Zudem bin ich in mehrere digitale Projekte involviert, insbesondere die Weiterentwicklung der *Gnomon Bibliographischen Datenbank*.³ Obwohl ich mich daher mit gutem Gewissen als einen digital-affinen early adopter bezeichnen würde, würde ich mich nicht als digital humanist oder Ähnliches einordnen: Schon von meiner Ausbildung in den Fächern Geschichte und Klassische Philologie her bin ich ein durchaus „konventioneller“ Historiker. Ein Hintergrund in der Informatik, der über das projektbezogen und autodidaktisch Angearbeitete hinausginge, fehlt mir. Das ist insofern wichtig, weil mein Interesse an Datenbanken immer von konkreten historischen Fragestellungen und den damit verbundenen Anwendungsszenarien herkommt. Die Begeisterung für das technisch Machbare um der reinen Machbarkeit willen teile ich folglich nur eingeschränkt, obwohl ich einen solchen Zugang aus einer anderen Perspektive heraus für völlig legitim halte. Die folgenden Ausführungen beziehen sich dieser persönlichen Orientierung entsprechend nicht auf den technisch-explorativen Aspekt der Digital Humanities als einer „Neuen Welt“, sondern fragen nach den Auswirkungen informatischer

2 Das bis heute fortgeführte *Tutorium Augustanum* (<http://www.historicum-estudies.net/etutorials/tutorial-alte-geschichte>, dieser Hyperlink und alle folgenden wurden zuletzt am 25.11.2019 aufgerufen) wurde zuerst 2003 als *Tutorium Quercopolitanum* publiziert.

3 Zu nennen sind diesbezüglich ein Kooperationsprojekt mit dem DAI im Rahmen von CLARIN-D (*Bibliographische Datenbanken als visualisiertes Wissensnetz*, 2013–2014) sowie die derzeit laufende Migration der Daten der *Gnomon-Datenbank* in den B3Kat verbunden mit einer engen Integration in den DFG-geförderten Fachinformationsdienst Altertumswissenschaften (Aktionsfeld 4: Nachhaltigkeit für exzellente Fachinformation, Arbeitspaket 16: Migration der Gnomon Bibliographischen Datenbank in die Verbunddatenbank (B3Kat), 2016–2018).

Arbeitsmittel auf die Bewohner der „Alten Welt“, d. h. den auf eine narrative Rekonstruktion vergangener Sachverhalte abzielenden Historiker.

1 Ein methodischer Paradigmenwechsel?

Niemand wird bestreiten, dass die Möglichkeiten der elektronischen Datenverarbeitung in den zurückliegenden vier Jahrzehnten seit der Einführung des PC die Arbeitswirklichkeit von Historikern auch und gerade im Bereich der Altertumswissenschaften grundlegend verändert haben.⁴ Bibliographische Datenbanken ermöglichen umfassende Literaturrecherchen in verhältnismäßig kurzer Zeit.⁵ Durch elektronische Textcorpora ist es möglich, fast die gesamte literarische Produktion der Antike sowie einen Großteil der edierten Inschriften und Papyri zu durchsuchen.⁶

4 Hohls (2007). Vgl. die diesbezüglichen Einführungen: Cohen et al. (2006); (2010); Schmale (2010); Galgano et al. (2013); Koller (2016); Berry et al. (2017). Während diese Publikationen im Wesentlichen von der Tätigkeit des Historikers ausgehen, bieten Kurz (2016) und Jannidis et al. (2017) eine Einführung in grundlegende Konzepte und Techniken der Informatik für Historiker. Spezifisch altertumswissenschaftliche Aspekte berühren alle aufgeführten Veröffentlichungen nicht oder nur am Rande.

Zeitlich noch weiter zurück reicht der Einsatz von Großrechnern für geisteswissenschaftliche Forschungsprojekte, doch blieb in solchen Einsatzszenarien die informatische Datenverarbeitung in Rechenzentren noch strikt von der alltäglichen Arbeitswelt des Historikers getrennt: Haber (2006) 169–171. Einen Überblick über die Entwicklung bis zur Mitte der 1970er Jahre geben Swierenga (1970) und Arnold (1974); vgl. auch Thaller (1990). Die wichtigsten frühen Projekte in Nordamerika im Feld der klassischen Altertumswissenschaften stellt der Sammelband Solomon (1993) vor. Instrukтив sind auch die Berichte über „Computers and the classics“ in den frühen Ausgaben der Zeitschrift *Computers and the Humanities*: McDonough, Jr. (1967); Delatte (1968); Waite (1968); Waite (1970); Waite (1971); Helgerson (1988).

5 Zu nennen wären im Gebiet der klassischen Altertumswissenschaften vornehmlich die mittlerweile von Brepols angebotene elektronische Variante der *Année philologique* (<http://apps.brepolis.net/LTool/Entrance.aspx?w=30>), die *Gnomon Bibliographische Datenbank* (<http://www.gnomon-online.de>) sowie *iDAI.bibliography/ZENON* (<https://zenon.dainst.org>) und *Dyabola* (<http://www.dyabola.de>). Hinzu treten Spezialbibliographien wie etwa *Nestor* (<https://classics.uc.edu/nestor>) oder die *Bibliographie Papyrologique* (<http://www.aere-egke.be/BP.htm>) und vereinfachte Version unter <http://papyri.info/bibliosearch>.

6 Textdatenbanken: *Thesaurus Linguae Graecae* (<http://stephanus.tlg.uci.edu>), *Library of Latin Texts – Series A* (<http://apps.brepolis.net/LTool/Entrance.aspx?w=12>), *Bibliotheca Teubneriana Latina Online* (<https://www.degruyter.com/view/db/btl>), *PHI Latin Texts* (<http://latin.packhum.org>); Inschriften: *Searchable Greek inscriptions* (<http://epigraphy.packhum.org>), *Epigraphische Datenbank Clauss-Slaby* (<http://www.manfredclauss.de>), *Epigraphische Datenbank Heidelberg* (<http://edh-www.adw.uni-heidelberg.de/home>), *Ubi erat lupa* (<http://www.ubi-erat-lupa.org>), *Epigraphic Database Roma* (<http://www.edr-edr.it>), *Epigraphic*

Münzdatenbanken verbinden die Systematik der Standardcorpora mit Bildmaterial aus den bedeutendsten Münzsammlungen der Welt.⁷ Literaturverwaltungs- und Wissensorganisationsprogramme erleichtern das Verwalten der recherchierten Informationen und unterstützen das Herstellen von Querverbindungen.⁸ Für heutige Studierende ist ein Arbeiten ohne diese Hilfsmittel – von OPACs⁹ und Textverarbeitungsprogrammen ganz zu schweigen – kaum noch vorstellbar und für Wissenschaftler sind die dadurch erzielten Effizienzgewinne angesichts zunehmender bürokratischer Lasten eine wichtige Voraussetzung, um überhaupt noch effektiv forschen zu können.

Diese Umwälzungen werfen die Frage auf, inwieweit sie auch einen methodischen Paradigmenwechsel implizieren – gewissermaßen einen digital turn.¹⁰ Die Nutzung digitaler Ressourcen alleine kann freilich noch nicht als Indiz einer methodischen Umwälzung gelten,¹¹ denn sonst wäre bereits jeder Historiker, der E-Mail, Textverarbeitung oder OPAC benutzt, ein digital historian. Sinnvollerweise muss jedoch zwischen Methoden im engeren Sinne von heuristischen Prozeduren und allgemeineren unterstützenden Arbeitsprozessen wie Kommunikation, Materialrecherche, Wissensorganisation usw. unterschieden werden.¹² Gerne wird in diesem Zusammenhang auf das von dem italienischen Literaturwissenschaftler Franco Moretti entwickelte Konzept des distant reading verwiesen.¹³ Moretti ging aus von der Feststellung, dass die Vorarbeiten für eine wirkliche Literaturgeschichte, zumal eine Geschichte der Weltliteratur, die Lektürezeit eines einzelnen

Database Bari (<http://www.edb.uniba.it>), *HispaniaEpigraphica Online* (<http://eda-bea.es>), Metasuche im Rahmen des *europæana eagle project* (<https://www.eagle-network.eu>); Papyri: *Duke Databank of Documentary Papyri* (<http://papyri.info/browse/ddbdp>). Ich verzichte auf die Anführung der Datenbanken einzelner Sammlungen.

- 7 *Sylloge Nummorum Graecorum* (<http://www.sylloge-nummorum-graecorum.org>), *Pella* (<http://numismatics.org/pella>), *Coinage of the Roman Republic Online* (<http://numismatics.org/cirro>), *Online Coins of the Roman Empire* (<http://numismatics.org/ocre>). Ich verzichte auch hier auf die Anführung der Datenbanken einzelner Sammlungen. Das Material aus zahlreichen Auktionskatalogen erschließen *CoinArchives* (<https://www.coinarchives.com/a>) und *acsearch.info* (<https://www.acsearch.info>).
- 8 Citavi (<https://www.citavi.com>), Zotero (<https://www.zotero.org>), Lit-link (<http://www.litlink.ch>), Mendeley (<https://www.mendeley.com>), EndNote (<http://endnote.com>), RefWorks (<https://refworks.proquest.com>).
- 9 Heute ein selbstverständliches Basisangebot, das für Arnold (1974) 126 jedoch noch völlig außerhalb des technisch Denkbaren lag.
- 10 Der Begriff z. B. bei Hagner et al. (2013) 7. Vgl. das Schlagwort „computational turn“ bei Berry (2011). Eine differenzierte Diskussion des Problems bieten Rieder et al. (2012).
- 11 So jedoch tendenziell Zaagsma (2013).
- 12 Vgl. Rieder et al. (2012) 68–69.
- 13 Beispielfhaft Hayles (2012a) 46; Zaagsma (2013) 24. Mit einer gewissen kritischen Reserve verweisen Hagner et al. (2013), auf das Konzept. Vgl. für die Altertumswissenschaften die unter dem Titel „Close Reading and Distant Reading. Methoden der Altertumswissenschaften in der Gegenwart“ auf dem Historikertag 2014 in Göttingen durchgeführte Sektion.

Forschern bei weitem übersteigen würden.¹⁴ Ein solches Projekt sei daher mit den Verfahren des traditionellen close reading nicht zu bewältigen, sondern nur mithilfe maschinell-quantitativer Auswertungen, wofür er den Begriff distant reading prägte. Ob sich damit eine neue Methodik für die Alte Geschichte im Speziellen oder die Altertumswissenschaften im Allgemeinen beschreiben lässt, scheint mir jedoch aus mehreren Gründen zweifelhaft zu sein.

Zunächst ist anzumerken, dass Morettis Vorschlag das spezifisch literaturwissenschaftliche Problem der Kanonbildung adressiert.¹⁵ Da für den Historiker ein Text in der Regel jedoch kein Forschungsgegenstand um seiner selbst willen ist, sondern ein Mittel zum Zweck der Ergründung einer weiterführenden historischen Fragestellung, konnte das reine close reading einzelner Werke oder Passagen niemals das methodische Mittel der Wahl sein. Für ihn musste es stets um eine Kontextualisierung des Einzelzeugnisses im Rahmen der weiteren vorliegenden Überlieferung gehen. Moretti hat sich denn auch nicht von ungefähr auf einen Historiker, nämlich Fernand Braudel, als eines seiner Vorbilder bezogen.¹⁶

Speziell in den Altertumswissenschaften hat zudem das relativ schmale Corpus erhaltener antiker Texte von jeher den Anspruch auf möglichst Vollständigkeit der Kontextualisierung gefördert, also genau jenen Überblick, den Moretti mit dem distant reading zu erreichen sucht. Dies gilt a fortiori, wo die jeweilige Fragestellung eine weitere Eingrenzung der Quellenbasis erlaubt. Der von Moretti in den Blick genommene Fall, dass eine von menschlichen Interpreten nicht zu bewältigende Informationsmenge zu analysieren wäre,¹⁷ tritt in den Altertumswissenschaften praktisch kaum auf.¹⁸ Diese haben zudem Lexika und Indizes in großer Zahl erarbeitet, die als gedruckte Hilfsmittel die Materialdurchsicht erleichtern und damit genau das ermöglichen, was heute (für neuere Epochen sicher zu recht) als wesentlicher Fortschritt informatischer Datenverarbeitung bezeichnet wird. Dass solche Projekte mit Hilfe der EDV in schnellerer Zeit, mit weniger Personal, noch etwas genauer und unter Verzicht auf pragmatische Restriktionen durchgeführt werden können, steht auf einem anderen Blatt.¹⁹

14 Zur Position Morettis s. Moretti (2000); Moretti (2003); Moretti (2005). Vgl. auch Jockers (2013).

15 Vgl. auch Wilkens (2012).

16 Moretti (2000) 56–57; Moretti (2003) *passim*.

17 Vgl. auch Wettlaufer (2016), Kap. 7, der als spezifisches Potential der digital humanities den Fall ausmacht, dass diese „Möglichkeiten qualitativen Erkenntnisgewinns [schaffen], der ansonsten aufgrund der schieren Informationsmasse für menschliche Interpreten nicht zu bewältigen wäre“.

18 Der Plural ist wichtig, denn selbstverständlich soll nicht gesagt sein, dass ein *einzelner* Mensch beliebige Mengen an Material auswerten könne.

19 Ob aus diesem Mehr dann überhaupt noch relevante Erkenntnisfortschritte resultieren, allerdings auch.

Exemplarisch sei hier auf den Zettelkatalog des *Thesaurus Linguae Latinae* verwiesen:²⁰ Für diesen wurden die Texte bis etwa 200 n. Chr. vollständig verzettelt, während aus den jüngeren Texten nur besonders relevant erscheinende Stellen exzerpiert wurden. Bei den Inschriften beschränkte man sich für die vollständige Erfassung auf die republikanischen Monumente in *CIL I* sowie die Wortkonkordanz zu *CIL VI* (Inschriften der Stadt Rom), darüber hinaus wurden wiederum nur Exzerpte angefertigt. Derartige Einschränkungen wären heute nicht mehr notwendig, lassen sich doch die Belegstellen eines Wortes vergleichsweise leicht in den einschlägigen Textdatenbanken ermitteln.²¹ Andererseits darf die Beschränkung auf Exzerpte für einen Teil des Materials auch nicht ausschließlich als Defekt gesehen werden, denn sie stellt bereits einen wichtigen Schritt der Abstraktion und damit eine genuine Forschungsleistung dar. In einem Punkt ist das Zettelkastensystem des *Thesaurus* den heute verfügbaren Datenbanken sogar substantiell überlegen: es berücksichtigt auch Textvarianten. Darauf wird noch zurückzukommen sein.

Es ist auch nicht so, dass die Erstellung des Zettelkatalogs einen übermäßig langen Zeitraum benötigt hätte: Das Grundmaterial wurde von 1894 bis 1899 gesammelt, mithin in einer Zeitspanne, die selbst in den engen Zeithorizonten gegenwärtiger Forschungsförderung gerade noch vermittelbar wäre. Das Beispiel zeigt im Übrigen auch, dass die Materialsammlung allein das geringere Problem ist – die Publikation des *Thesaurus* ist auch nach über hundert Jahren noch nicht abgeschlossen. Die gesammelten Daten müssen eben auch durchgearbeitet werden – und dabei bleibt der Mensch nach dem Stand der Technik immer noch auf sich gestellt, weil die Erschließung von Bedeutungen für den Computer viel schwieriger ist als das Indexieren und Auszählen.

Ein weiterer problematischer Aspekt des distant reading ergibt sich aus der Tatsache, dass quantitative Auswertungen gerade im Bereich der aus der Antike erhaltenen literarischen Quellen aufgrund der spezifischen chronologischen und lokalen Schwerpunktbildung der Überlieferung durchaus problematisch sind.²² Lassen sich etwa auf den spärlichen Fragmenten der republikanischen Annalistik überhaupt noch statistisch signifikante Analysen gründen? Welche Reichweiten haben Auswertungen, die sich ganz überwiegend auf Texte von in Athen wirkenden Autoren stützen, die sich wiederum auf das 5. und 4. Jh. v. Chr. konzentrieren?

20 Die folgenden Angaben nach der Projekthomepage: <https://www.thesaurus.badw.de/ueber-den-tll/zettelmaterial.html>.

21 Dies gilt besonders dort, wo die in einer Datenbank enthaltenen Texte einer morphologischen Analyse unterzogen wurden.

22 Die Verzerrung der Materialbasis ist im Bereich der Altertumswissenschaften vor allem den Prozessen der Textüberlieferung bzw. dem archäologischen Forschungsstand geschuldet. Für neuere historische Epochen führt die wissenschafts- und geschichtspolitisch gesteuerte Schwerpunktsetzung der Digitalisierungsprojekte jedoch zu durchaus ähnlichen Problemen: Zaagsma (2013) 19–21.

Das sind offensichtlich keine neuen methodischen Probleme, aber der Versuch, vermittels *distant reading* langfristige Veränderungen und regionale Spezifika im größeren Kontext nachzuweisen, stößt dort an Grenzen, wo wir nur für einzelne Regionen und einzelne Epochen überhaupt eine Materialmenge besitzen, die statistisch aussagekräftige Auswertungen erlaubt.

Ferner konfrontiert uns der antike Textbestand immer wieder mit erheblichen Schwierigkeiten bezüglich Autorzuweisungen und Datierungsfragen, die aber auf das Ergebnis statistischer Auswertungen (und nichts anderes kann ein Computer ja vornehmen) signifikante Auswirkungen haben.²³ Es ist klar, dass die lückenhaften Datensets, die sich aus antikem Material zusammenstellen lassen, mit ganz anderer Vorsicht interpretiert werden müssen, als solche zu neueren Epochen, die tatsächlich Anspruch auf annähernde Vollständigkeit erheben können oder doch jedenfalls so breit angelegt sind, dass statistische Repräsentativität postuliert werden kann.²⁴

Diese spezifische Überlieferungslage bringt es auch mit sich, dass elektronische Verfahren in den Altertumswissenschaften gerade das, was nach dem Konzept des *distant reading* von ihnen erwartet wird, nur sehr eingeschränkt leisten können, nämlich ein Überschreiten des traditionellen Textkanons. Wir können nur die (literarischen) Texte überhaupt einer elektronischen Auswertung zuführen, die den Selektionsprozess der Überlieferung überstanden haben. Hinter diese Art der Kanonbildung führt auch mit informatischen Mitteln kein Weg zurück. Die hinter dem Konzept des *distant reading* letztlich liegende Hoffnung, den Kanon mit all seinen impliziten Blindstellen und Wertungen zu überwinden, lässt sich eben nicht realisieren, wenn das überhaupt erhaltene Material im Wesentlichen mit dem Kanon koinzidiert.

Von der Debatte um das *distant reading* einmal abgesehen darf die emphatische Neuheitsrhetorik, die heute oft mit allem Digitalen verbunden ist, nicht den Blick auf die Kontinuitäten verstellen:²⁵ Datenbanken im Sinne einer strukturierten Wissensablage sind kein Phänomen, das erst durch den Siegeszug des Computers entstand. Auch ein Zettelkasten ist eine (analoge) Datenbank, wie am bereits erwähnten Beispiel des *Thesaurus Linguae Latinae* klar erkennbar wird.²⁶ Doch auch Lexika

23 Die offensichtlichen Datierungsfehler in Abfragen des *Google Ngram Viewer* (Hodel (2013) 109–110) sind hingegen in der Sache meist leicht zu klären. Hier liegt das Problem „nur“ auf der Ebene der Datenqualität.

24 Auch hier ist freilich Vorsicht geboten: Nach Schöch (2013) 7–8 sind etwa von 20–30.000 im England des 19. Jh. publizierten Romanen überhaupt nur ca. 6.000 erhalten und von diesen wiederum nur gut die Hälfte mit OCR in ausreichender Qualität in maschinenlesbare Form überführt, um eine elektronische Auswertung zu ermöglichen.

25 Vgl. auch Ore (2012) 108–109, der die kritischen Zeichen antiker Homerkommentatoren und das Leidener Klammersystem als „Markup“-Sprachen behandelt.

26 Sichtbar wird diese analog-digitale Tradition zudem anhand der Versuche, das Zettelkastensystem von Niklas Luhmann in eine elektronische Form zu überführen: *Synapsen*

organisieren Wissen und machen es anhand bestimmter Selektoren – der Lemmata – leichter auffindbar. Unabhängig vom Forschungswert der jeweiligen Artikel liegt gerade in dieser Funktion der bis heute bleibende Wert der *Realencyclopädie*, dass sie als umfänglicher Index zur antiken Überlieferung genutzt werden kann.

Mit derartigen Formen der Wissenserschließung begann – lange vor dem digitalen Zeitalter – auch die oft beklagte Dekontextualisierung von Quellenstellen, die im Extremfall zur Nutzung der Ursprungstexte als bloßer Steinbruch für einzelne Informationsbrocken führt. Diese Art der Quellennutzung wurde nicht zuletzt auch dadurch gefördert (ja geradezu gefordert), dass mit zunehmender Dauer der Forschungstradition und dem starken Anwachsen der wissenschaftlichen Publikationstätigkeit im Laufe des 20. Jh. die Lektüre der modernen Forschungsliteratur einen zunehmenden Anteil der Lesezeit eines Forschers in Anspruch nahm. Gleichzeitig führte die Quantifizierung wissenschaftlicher Leistung und die stetig wachsende Bedeutung projektbezogener Forschungsförderung dazu, dass diese Lesezeit nicht in entsprechendem Maße ausgeweitet werden konnte.²⁷ Zweckfreie, d. h. nicht unmittelbar projektorientierte, Quellenlektüre findet im Zeitbudget eines professionellen Forschers heute nur noch mühsam einen Platz. Das dekontextualisierende distant reading ist daher seit langem in Ausbreitung begriffen und die Möglichkeiten elektronischer Volltextsuche sind nur der logische Endpunkt dieser Entwicklung.²⁸

Dabei soll gar nicht gesagt sein, dass derartige Suchmöglichkeiten keinen sinnvollen Weg der Materialerschließung darstellen (zumal sich andere Auswege aus dem skizzierten Dilemma nicht anbieten), nur muss man sich im Klaren sein, was ein solcher Zugriff – egal ob mit analogen oder digitalen Hilfsmitteln – nicht zu leisten vermag: Abgesehen von der Schwierigkeit, isolierte Passagen ohne ausreichende Kenntnis ihres argumentativen Zusammenhangs korrekt einzuordnen, entstehen auf diesem Weg kaum Lesefrüchte. Gerade diese Zufallsfunde führen in den Geisteswissenschaften jedoch oft zu den spannendsten Einsichten – wenn auch nicht unbedingt für das Projekt bzw. die Fragestellung, welche die Lektüre ursprünglich auslösten.

Ferner können weder die Schaffung eines externalisierten Gedächtnisses noch der Austausch zwischen Wissenschaftlern als Neuheiten des digitalen Zeitalters

(<http://www.verzetteln.de/synapsen>), *Zettelkasten* (<http://zettelkasten.danielluedecke.de>), jeweils mit explizitem Verweis auf Luhmann. Frühe Anwendungsfälle elektronischer Datenverarbeitung in den Geistes- und Sozialwissenschaften zielten oft darauf ab, bereits etablierte quantitative Auswertungsverfahren zu beschleunigen: Thaller (1990) 140–146; Rieder et al. (2012) 69–70; Röhle (2014).

27 Zum Problem der Überproduktion geisteswissenschaftlicher Literatur im Verhältnis zur Lesezeit vgl. Theisohn (2013) 15–16.

28 Vgl. Zaagsma (2013) 26–27.

gelten.²⁹ Bücher (insbesondere Nachschlagewerke und Handbücher) sind ja nichts anderes als ein solches externalisiertes Gedächtnis und Gelehrtenkorrespondenz funktionierte auch ohne social media.³⁰ Auch hier gilt wieder, dass die neuen digitalen Werkzeuge vieles erleichtern, doch bei nüchterner Betrachtung zeichnet sich ein breiter Wandel nicht ab.

Dies wird besonders am Beispiel des oft beschworenen kollaborativen Arbeitens deutlich: Zugrunde liegt diesem Konzept zum einen die aus naturwissenschaftlichen Forschungsdesigns abgeleitete Vorstellung, dass sich größere Fragestellungen stets logisch in kleinere unterteilen lassen, zum anderen die idealistische Annahme, dass Forscher primär zum Wohle eines höheren Zieles – der Wissenschaft oder der Wahrheit – arbeiten. Die erste Annahme ist auf geisteswissenschaftliche Fragestellungen nur bedingt übertragbar, in denen es um die Verknüpfung von Deutungen und Hypothesen geht, die in einem komplexen Wechselspiel ihre jeweilige Plausibilität beeinflussen. Die zweite Annahme stößt sich zumindest an den harten materiellen Rahmenbedingungen des Wissenschaftsbetriebes, wahrscheinlich aber auch am menschlichen Wesen generell: Zum ersten müssen Forscher auch und gerade in den Geisteswissenschaften auf die Erfordernisse universitärer Laufbahnen Rücksicht nehmen. Dies bedeutet, dass das Erreichen einer Stelle, aber auch die Einwerbung von Drittmitteln an den Nachweis individueller Forschungsleistungen geknüpft sind.³¹ Investitionen in kollaboratives Arbeiten zahlen sich dabei nicht unbedingt aus.³² Warum sollte beispielsweise der harte Kampf um Publikationsrechte für Inschriften nur deshalb enden, weil die Edition in digitaler Form erfolgt? Wo derartige materielle Zwänge keine Rolle spielen, greift die Kombination von individuellem Leistungsstolz mit Prestigedenken. Die edit wars auf *Wikipedia* zeugen eindrücklich von diesem weniger angenehmen Aspekt der *conditio humana*. Der von Egoismus und Eitelkeit völlig befreite Forscher, der zum Wohle des höheren Zieles der Erkenntnis anonym in das Glied eines kollaborativen Wissenskollektivs zurücktritt, wird wohl bis auf weiteres ein Phantasma bleiben. Entsprechend mager sind die Ergebnisse vieler mit hohen Zielen

29 In diesem Sinne jedoch Berry (2011) 10.

30 Vollends fatal wäre es, eine solche Kommunikation schon mit einer Publikation zu wechseln. Gegen solche Ideen Walkowski (2013) 47–48, der mit Recht darauf hinweist, dass „die Leistungen einer Publikation gerade aus der Unterbrechung des kommunikativen Flusses resultieren“.

31 Der reichlich optimistische Beitrag von Wampfler (2013) bemerkt zu diesem Problem immerhin, dass durch kollaborative Wissenschaftspraktiken „bewährte Verfahren zur Organisation und Verteilung von Reputation“ entfielen (S. 89), ohne aber zu erläutern, was an ihre Stelle treten soll.

32 Vgl. Greetham (2012). Hingegen dürfte eine Apologie digitaler Publikationsformen an sich, wie sie noch Raben (2008) zum Start des *Digital Humanities Quarterly* für notwendig hielt, heute überholt sein.

gestarteter kollaborativer Unternehmungen.³³ Erfolgversprechender könnte das Konzept von individuell zuweisbaren und zitierbaren Mikropublikationen sein, die etwa als Kommentare an Quelldatenbanken angedockt sein könnten. Freilich müssten dafür auch Strukturen geschaffen werden, wie solche Publikationen inhaltlich erschlossen und in den zentralen Fachinformationsdiensten indiziert werden können. Damit entstünde freilich auch die Notwendigkeit, eine gewisse Qualitätskontrolle zu gewährleisten. Ob die Datenbankbetreiber allerdings solche Herausgebertätigkeiten übernehmen möchten, bleibt abzuwarten.

Noch in einem weiteren Punkt scheinen mir letztlich die Kontinuitäten zwischen analoger und digitaler Welt zu überwiegen: die oft beklagte unkritische Nutzung verbreiteter Suchmaschinen zur wissenschaftlichen Informationsrecherche. Sicher ist richtig, dass Forscher etwa Google nur als eine Blackbox nutzen können, da weder transparent ist, welche Datenquellen indiziert werden, noch wie die Relevanz der Treffer geratet wird.³⁴ Freilich gilt dasselbe im Wesentlichen auch für Bibliographien, Archiv- und Bibliothekskataloge, deren inhaltliche Beschreibungen stets nur eine Möglichkeit unter mehreren darstellen.³⁵ Auch gegen die Nutzung von Lexika oder älterer wissenschaftlicher Literatur überhaupt ließe sich einwenden, dass in keiner Weise transparent ist, wie der jeweilige Autor die von ihm zitierten Quellen und Sekundärliteratur aufgefunden und nach welchen Prinzipien er sie gegebenenfalls ausgewählt hat. Letztlich ist sogar fraglich, ob das klassische Querlesen von größeren Textbeständen solch hohen Transparenzansprüchen genügt, denn auch das menschliche Gehirn benutzen wir im Wesentlichen als Blackbox. Schon in der analogen Wissenschaftswelt war der Prozess der Recherche als solcher nur ansatzweise nachvollziehbar, nämlich dort, wo der Autor ihn ausnahmsweise explizit verbalisierte. Ob eine vollständige Transparenz aller Irr- und

33 Exemplarisch sei auf das von Berry (2011) 18 Nr. 10 als „a good example of a wiki-based project“ angeführte <http://openprotagoras.wikidot.com> verwiesen: Tatsächlich beteiligte sich nach Ausweis der Änderungshistorie fast ausschließlich der Seitenbetreiber an dem Projekt einer „kollaborativen“ Übersetzung des platonischen Protagoras. Die letzten Änderungen scheinen auf das Jahr 2012 zurückzugehen. Die Übersetzung umfasst bislang die Stephanusseiten 309–322.

34 van Dijk (2010); Kemman et al. (2014). Eine kritische Beleuchtung des *Google Ngram Viewers* unternimmt Hodel (2013). Ähnliche Bedenken können freilich auch gegen viele wissenschaftliche Datenbanken vorgebracht werden: Rieder et al. (2012) 75–77. Folgende Probleme stehen einer Transparenz der verwendeten Suchalgorithmen entgegen: (1) Nichtverfügbarkeit des Quelltextes, (2) mangelnde Kenntnisse der Benutzer für eine kritische Evaluierung des Quelltextes, (3) Komplexität von Algorithmen im Bereich des maschinellen Lernens.

35 Selbst Kemman et al. (2014), Kap. 1 konzedieren, dass „[o]f course these issues were equally problematic in the analogue era“. Die anschließende Relativierung überzeugt nicht: Warum sollte bei weniger verfügbarem Material eine vorgeschaltete Selektion/Gewichtung weniger problematisch sein?

Abwege von der Themenfindung bis zur letzten Textstraffung für den Leser überhaupt wünschenswert wäre, sei einmal dahingestellt.

All dies soll nun keineswegs bedeuten, dass den Digital Humanities ihr Status als eigene Disziplin abgesprochen werden soll.³⁶ Die Frage, ob sich traditionelle Geschichtswissenschaft durch neue digitale Methoden ablösen ließe und hinfällig werde, möchte ich jedoch verneinen. Man mag sich in Erinnerung rufen, was Klaus Arnold in der Mitte der 1970er Jahre über elektronische Datenverarbeitung und Geschichtswissenschaft zu sagen hatte:

[...] vielmehr wollen wir die Anwendungsmöglichkeiten der EDV im Bereich der Geschichtswissenschaften als echte „Hilfs“-wissenschaft der Geschichte in ihrem Bemühen um die Erforschung von Handlungen und Verhältnissen des Menschen in der Vergangenheit aufzeigen. Die elektronische Datenverarbeitung wird nie mehr sein als ein Hilfsmittel auf diesem Weg. Intuition und Erfahrung des Historikers in Verstehen und Wertung des Gewesenen werden sich mit Sicherheit niemals durch eine noch so vollkommen und unter kaum vorstellbarem Aufwand an Zeit und Sorgfalt programmierte Großrechenanlage ersetzen lassen, auch wenn der Umgang mit der Datenverarbeitung den Geisteswissenschaftler zu einem Maß an Genauigkeit – etwa in seiner Terminologie – zwingt, über die er, gemessen an den „exakten“ Naturwissenschaften, ansonsten nicht verfügt.³⁷

Der Tonfall mag heute etwas harsch klingen, zumal auch die klassischen „Hilfswissenschaften“ begrifflich längst zu „Grundwissenschaften“ aufgewertet sind. In der Sache verdient das Argument aber nach wie vor Beachtung: Solange man Geschichtsschreibung nicht als bloße Datenakkumulation und -strukturierung begreift,³⁸ sondern (in gut antiker Tradition) als einen narrativen Prozess, der Verstehen (d.h. ursächliche Erklärung) und Werten einbezieht, kann jede elektronische Heuristik nur eine Vorarbeit für die „eigentliche“ Geschichtsschreibung darstellen. Mit der reduktionistischen Verkürzung von Geschichtsschreibung auf Datensammlung

36 Zum Stand dieser Diskussion vgl. Sahle (2015a).

37 Arnold (1974) 99.

38 Zu Forderungen in dieser Richtung s. Hayles (2012a) 51 („Some argue that the discovery of patterns is sufficient, without the necessity to link them to meaning.“); Walkowski (2013) 45 („Einige *Enhanced Publications* besitzen daher auch gar keine zentralen Textbestandteile mehr. Daten werden als Fakten verstanden, die es zu prozessieren, aber nicht zu interpretieren gilt. Die letzte Kritik am Text lautet daher, dass die relevanten wissenschaftlichen Ergebnisse gar keine textuelle Grundlage mehr besitzen und dieser auch nicht bedürfen.“). Vgl. dazu bereits die Warnung im Vorwort der ersten Ausgabe von *Computers and the Humanities*: „We need never be hypnotized by the computer’s capacity to count into thinking that once we have counted things we understand them“, (Prospect (1966) 2).

und der Gleichsetzung von Daten mit Fakten geht bisweilen die Hoffnung einher, durch elektronische Methoden den Menschen als „irrationalen“ Störfaktor wissenschaftlicher Erkenntnis aus dem Wissenschaftsprozess überhaupt auszuschließen.³⁹ Sowohl die Programmierung der Suchalgorithmen selbst als auch die Selektion der untersuchten Materialbasis sowie die Festlegung der spezifischen Suchparameter werden jedoch nach wie vor von Menschen verantwortet.⁴⁰ Davon einmal abgesehen liegt solchen Vorstellungen ein Wissenschaftsbild zugrunde, das die Funktion von Wissenschaft als kultureller Sinnstiftungspraxis nicht angemessen berücksichtigt. Eine Wissenschaft, die sich in der quantitativen Vermessung ihres Materials erschöpft, hätte ihren Zeitgenossen wenig zu sagen und würde im Grunde nur ein überholtes positivistisches Wissenschaftsverständnis reaktivieren.⁴¹

Unbenommen sind freilich die zunehmende Bedeutung digitaler Werkzeuge und die Notwendigkeit, entsprechender Expertise einen institutionellen Ort an den Universitäten zu geben. Die Komplexität sowohl der Probleme als auch der erforderlichen Lösungen machen rein kooperative Modelle, in denen informatisch ungebildete Geisteswissenschaftlicher und geisteswissenschaftlich ahnungslose Informatiker zusammenkommen, wenig effektiv.⁴² Die Etablierung der Digital Humanities ist daher meines Erachtens die Konsequenz eines Spezialisierungsprozesses, und die Annahme, dass Digitalität in erster Linie neue Werkzeuge zur Verfügung stellt, keineswegs gleichzusetzen mit einem Aufruf zur Zerstörung der Digital Humanities als Disziplin.⁴³

Tatsächlich ist die Institutionalisierung von Wissenschaftsbereichen gar nicht in erster Linie vom Nachweis einer eigenständigen Methodik abhängig.⁴⁴

39 Hayles (2012a) 47–48.

40 Vgl. Rieder et al. (2012) 70: „What is too often forgotten, though, is that our digital helpers are full of ‘theory’ and ‘judgement’ already. As with any methodology, they rely on sets of assumptions, models, and strategies. Theory is already at work on the most basic level when it comes to defining units of analysis, algorithms, and visualisation procedures.“ Ein instruktives Beispiel aus dem Bereich der Altertumswissenschaft stellen die schwierigen konzeptionellen Entscheidungen dar, die bei der Erstellung des *TLG Canon* zu treffen waren und die die Suchmöglichkeiten in der Datenbank bis heute bestimmen: Berkowitz (1993).

41 Vgl. Rieder et al. (2012) 71–73 zur „lure of objectivity“; Walkowski (2013) 49–50. Bei Rieder auch eine Kontextualisierung im Rahmen der Geschichte der Naturwissenschaften, in denen technischen Werkzeugen als Mittel der Wissensproduktion seit jeher eine wichtige Rolle zukommt.

42 So mit Recht Sahle (2015a), Kap. 4. Dagegen jedoch Gladney (2012), der vehement gegen eine Abtrennung der digital humanities von der Informatik plädiert, da dies zu einem Arbeiten am Forschungsstand vorbeiführe.

43 So jedoch Sahle (2015a), Kap. 4.

44 Ansprüche auf eine neue Methodik sind jedoch im Rahmen der Begründungslogik von Anträgen auf Forschungsförderung bedeutsam. Die Debatte um spezifisch digitale Methoden

Bekanntlich vermochte Theodor Mommsen eine solche in seiner Rektoratsrede von 1874 auch in der Geschichtswissenschaft nicht auszumachen, die sich an deutschen Universitäten gleichwohl recht erfolgreich institutionalisierte und ausdifferenzierte.⁴⁵ Entscheidend ist vielmehr das gesellschaftliche bzw. wissenschaftspolitische Interesse an bestimmten Fragestellungen sowie Prozesse der Spezialisierung und Professionalisierung, die zur Verselbstständigung von Teil- oder Randbereichen führen. Eine grundsätzliche methodische Umorientierung der Geschichtswissenschaft als solcher ergibt sich daraus jedoch noch nicht. Allenfalls wird eine Verschiebung der Forschungsschwerpunkte dadurch ermöglicht, dass nun einzelne Forscher (etwa im Rahmen von Qualifikationsschriften) Fragestellungen bearbeiten können, die früher Großprojekten mit zahlreichen Mitarbeitern und Hilfskräften vorbehalten waren.

2 Probleme der Datenbasis

Wie frühere Medienbrüche und die Einführung neuer Erschließungssysteme bedeutet auch die Digitalisierung in einer Hinsicht eine entscheidende Herausforderung: Die Quellenzeugnisse, die den Sprung in die neue digitale Welt nicht schaffen, drohen letztlich vergessen zu werden. Die Erwartung eines Forschers, in Datenbanken letztlich „alles“ geboten zu bekommen ist kaum anders geartet als die verbreitete Praxis, Google-Suchergebnisse mit „dem“ Internet oder gar dem kompletten zu einem Thema verfügbaren Wissen gleichzusetzen.⁴⁶ Ein entscheidender Vorteil der Altertumswissenschaften ist in dieser Hinsicht, dass aufgrund der beschränkten Menge der erhaltenen Texte vollständige Corpora der literarischen Quellen tatsächlich mit überschaubarem Aufwand realisierbar sind.⁴⁷ Nicht zuletzt daraus erklärt sich die Vorreiterrolle der klassischen Altertumswissenschaften in diesem Bereich.⁴⁸ Etwas anders sieht es dort aus, wo regelmäßig Erweiterungen des Quellenbestandes durch Neufunde zu verzeichnen sind, nämlich bei Inschriften und Papyri. Zwar ist auch die Erfassung solcher Neufunde ein quantitativ überschaubares Problem, doch tut sich hier eine Daueraufgabe auf, für die es entsprechend

ist daher auch im Kontext der Mittelallokation für DH-Projekte zu betrachten: Zaagsma (2013) 16. Die vergleichsweise hohen Kosten digitaler Projekte (insbesondere auch für den langfristigen Unterhalt) betont auch Berry (2011) 11.

45 Mommsen (1905).

46 Dazu Haber (2013) mit den Bemerkungen von Sarasin (2013).

47 Wobei man sich freilich der Lückhaftigkeit der Überlieferung bewusst bleiben muss.

48 Die im Jahre 1985 veröffentlichte Version A des *Thesaurus Linguae Graecae* war die erste CD-ROM für Endanwender, die keine Musik enthielt: <http://stephanus.tlg.uci.edu/history.php>. Gängige Softwareanwendungen wurden damals noch auf Disketten ausgeliefert.

dauerhafter und zentralisierter Strukturen bedürfte. Diese haben aber im Rahmen der derzeit üblichen Projektförderung keinen Platz. Dem ließe sich allenfalls durch Selbsthilfe der community, sprich verteilte Datenerfassung abhelfen.⁴⁹ Dies würde aber zumindest voraussetzen, dass Editionsplattformen mit entsprechender Funktionalität zentral zur Verfügung gestellt werden.⁵⁰

Ein wesentlicher Kritikpunkt an den derzeit verfügbaren Corpora wurde bereits vor mehr als zehn Jahren von Uwe Walter in einer Rezension der damals auf CD-ROM veröffentlichten *FGrHist* benannt: dass nämlich die meisten Textdatenbanken lediglich einen einfachen Lesetext bieten und auf einen kritischen Apparat verzichten.⁵¹ Dies ist umso bedauerlicher, als sich gerade textkritische Informationen dafür anböten, die Potentiale hypertextueller und dynamischer Publikationsformen zu ergründen. Da nicht einmal immer klar ist, welcher gedruckten Textvorlage eine Datenbank folgt,⁵² kommt dies in der Tat einem Rückschritt in ein vorkritisches Zeitalter gleich, denn die Vorstellung, dass Volltextdatenbanken als reines Suchwerkzeug benutzt werden, dürfte angesichts veränderter Nutzungsgewohnheiten der viel beschworenen digital natives ein frommer Wunsch sein.

49 Jedenfalls, sofern man die Integration der Neufunde in die zentralen Datenbanken als Messlatte nimmt. Dass die *Année épigraphique* mittlerweile auch in digitaler Form erscheint (<https://www.cairn.info/revue-annee-epigraphique.htm>, ältere Bände auch unter <https://www.jstor.org/journal/anneepig>) genügt diesem Anspruch selbstverständlich nicht, da es sich um eine bloße digitale Repräsentation der Druckausgabe ohne eigentliche Datenbankfunktionalität handelt. Anders verhält es sich mit dem *Supplementum Epigraphicum Graecum Online*, das von Brill auf seiner Datenbankplattform angeboten wird (<http://referenceworks.brillonline.com/browse/supplementum-epigraphicum-graecum>). Die Suchwerkzeuge dieser Plattform sind allerdings für Inschriftentexte nicht angepasst. Der wesentliche Mehrwert ergibt sich daher aus der Verwendung der Indizes in elektronischer Form – eine sehr „analoge“ Herangehensweise an das Problem.

50 Ein aktuelles Beispiel für verteilte Datenerfassung liefert im Bereich der Numismatik *NUMiD* (<http://www.numid-verbund.de>). Im Rahmen dieses Projektes werden auf einer gemeinsamen technischen Plattform (vgl. dazu <https://www.kenom.de/projektseite>) universitäre Münzsammlungen in Deutschland digitalisiert und erschlossen. Die generierten Daten stehen als Linked Open Data zur Verfügung. In diesem Falle ergibt sich die verteilte Erfassung gewissermaßen natürlich aus der Zusammenarbeit von einzelnen Sammlungen. Vgl. für Papyri: <http://papyri.info/editor> und zukünftig für Inschriften <http://epigraphy.info>.

51 <https://www.hsozkult.de/publicationreview/id/rezbuecher-6319>. Die genannte CD-ROM stellt insofern die löbliche Ausnahme von der Regel dar, als auf ihr auch die kritischen Apparate aus Jacobys Edition vollständig übernommen wurden. Dass die CD-ROM nach gerade einem Jahrzehnt auf aktuellen Betriebssystemen nur noch mit einigen Umwegen genutzt werden kann, weist auf das im Folgenden noch vertiefte Problem digitaler Nachhaltigkeit hin.

52 Dies trifft etwa auf die *Epigraphische Datenbank Claus-Slaby* zu. Einen kritischen Apparat bieten z. B. die *Duke Databank of Documentary Papyri* oder die *Roman Inscriptions of Britain*: <https://romaninscriptionsofbritain.org>.

Davon abgesehen können für Projekte der Digital Humanities die aktuellen wissenschaftlichen Textausgaben aus lizenzrechtlichen Gründen oft nicht genutzt werden.⁵³ Man mag argumentieren, dass neue digitale Werkzeuge auch an älteren Textausgaben entwickelt werden können. Aus fachwissenschaftlicher Perspektive stimmt es freilich bedenklich, wenn (ver-)alte(te) Editionen in elektronischer Umsetzung neue Prominenz gewinnen.⁵⁴ Es wäre daher dringend angezeigt, die Entstehung von kritischen Texteditionen, die von vornherein auf eine elektronische Publikation und Weiternutzbarkeit ausgelegt sind und im Open Access zur Verfügung stehen, aktiv zu fördern und dafür auch entsprechende standardisierte Werkzeuge anzubieten.⁵⁵ Dies ist derzeit am ehesten im Bereich der Papyrologie der Fall, wo die Plattform *papyri.info* mit *Son of Suda On Line* unter Sicherstellung wissenschaftlicher Qualitätsstandards auch einen Webeditor für kollaborative Beiträge anbietet.⁵⁶

Tobias Hodel hat die gegenwärtige Blüte der Materialvorlage im Rahmen von Digitalisierungsprojekten mit den großen Editionsunternehmungen verglichen, die im 19. Jh. gestartet wurden.⁵⁷ In der Tat wäre Theodor Mommsen wohl ziemlich begeistert gewesen von den heutigen Möglichkeiten, die Archive der Vergangenheit zu ordnen. Die aktuellen Digitalisierungsbemühungen bleiben jedoch im kritischen Anspruch zumeist deutlich hinter den Standards des 19. Jh.s zurück. Stattdessen beleben sie eine Sammlungspraxis wieder, die für die frühneuzeitlichen Inschriftencorpora kennzeichnend war: die weitgehend unkritische Aggregation von Daten aus verschiedensten Quellen.

Hochwertige Daten, die eine inhaltliche Erschließung leisten und auch die textkritische Tradition aufarbeiten, entstehen jedoch nicht nebenher. Im gegenwärtigen Förderprogramm der DFG besteht hier allerdings eine Lücke: Einerseits

53 Vgl. Wettlaufer (2016), Kap. 3.

54 Als Beispiel sei auf die *Digital Fragmenta Historicorum Graecorum* (www.dfhg-project.org) verwiesen, welche die von Karl Müller zwischen 1841 und 1872 publizierte Fragmentensammlung in eine digitale Form überführen. Dazu heißt es in der Projektbeschreibung: „In particular, it is very suitable for providing rapid, broad coverage and an extensive foundation upon which a new generation of born digital editions of fragmentary texts can build“, (<http://www.dh.uni-leipzig.de/wo/dfhg>). Dies klingt, als sollten sowohl Jacobys *FGrHist* als auch *Brill's New Jacoby* auf dem Weg in die Zukunft gewissermaßen überbrückt werden.

55 Solche nativ digitalen Editionen würden es auch ermöglichen, Digitalisate der zugrundeliegenden Manuskripte einzubinden. Was hier möglich wäre, zeigt in Ansätzen der *Classical Text Editor* (<http://cte.oeaw.ac.at/?id0=main>, besonders <http://cte.oeaw.ac.at/?id0=epub>), der jedoch noch primär auf die Vorbereitung von Druckpublikationen ausgerichtet ist. Die Entwicklung von digitalen Editionswerkzeugen ist Gegenstand des DFG-Förderprogramms „Infrastruktur für elektronische Publikationen und digitale Wissenschaftskommunikation“: http://www.dfg.de/formulare/12_11/12_11_de.pdf.

56 <http://papyri.info/editor>.

57 Hodel (2013) 105.

fällt „reine“ Dateneingabe nicht in den Bereich der Forschungsförderung im engeren Sinne, andererseits sind die Ausschreibungen im Bereich Wissenschaftliche Literaturversorgungs- und Informationssysteme stark auf die Entwicklung von Werkzeugen fokussiert, die möglichst generisch und nachnutzbar sein sollen (wie weit das auch immer über die spezifischen Anforderungen bestimmter Fächer hinaus möglich ist). Dies mag auf den ersten Blick effizienter scheinen als eine fach- oder gar problemspezifische Datengenerierung. Dieser Ansatz übersieht jedoch, dass Auswertungen stets nur so gut sein können, wie die zugrundeliegenden Daten: garbage in, garbage out. Dateneingabe, Datenverbesserung und Datenaufbereitung sind daher eminent förderungswürdige Ziele, wobei ohnehin in jedem Einzelfall dargelegt werden müsste, worin der Mehrwert einer vorgeschlagenen Maßnahme besteht.⁵⁸

Wünschenswert wäre auch, dass die Themen „Open Access“ und „informatische Nachnutzbarkeit“ auch und gerade im Hinblick auf Editionen und Materialvorlagen stärker von den einzelnen Universitäten sowie den Fachgesellschaften aufgenommen werden. Zwar können diese Institutionen den einzelnen Forschern keine verbindlichen Vorgaben machen, aber die Formulierung einer Open-Access-Policy sowie die Benennung von kompetenten Ansprechpartnern auf der lokalen Ebene könnte mittelfristig einen Bewusstseinswandel befördern.⁵⁹

3 Probleme der digitalen Nachhaltigkeit

Ein ganz praktisches Problem bei der Nutzung von Datenbanken besteht in der Frage, ob und wie gegebenenfalls sehr umfangreiche Suchergebnisse, die sich einer Auflistung in Fußnoten entziehen, zitierbar gemacht werden können. Die naheliegende Idee, dass eine solche Funktionalität von den Datenanbietern selbst vorgehalten wird – etwa in Form persistenter Links zu bestimmten Suchanfragen –, scheint freilich wenig realistisch. Es genügt nämlich im Sinne einer nachvollziehbaren Zitation keinesfalls, eine identische Suchanfrage in der Datenbank auszulösen, sondern die gesamte Datenbank müsste vollständig einer Versionierung unterworfen werden, die es ermöglicht, das Ergebnis einer bestimmten Suchanfrage zu jedem beliebigen

58 Crane (2004) 48 referiert eine mündliche Stellungnahme David Packards aus dem Jahr 1983: „He observed that software and systems were ephemeral but that primary sources such as well structured, cleanly entered source texts were objects of enduring value.“ Vgl. auch das Plädoyer von Schöch (2013) 10 für „bigger smart data or smarter big data“.

59 Über solche Policies verfügen beispielsweise die Universitäten Heidelberg (<https://www.uni-heidelberg.de/universitaet/profil/openaccess>) und Regensburg (<http://www.uni-regensburg.de/publikationen/medien/open-access-policy.pdf>).

Zeitpunkt in der Vergangenheit zu replizieren. Dieses Problem wäre technisch gewiss lösbar, dies würde aber den Aufwand für die Anbieter ganz erheblich erhöhen.

Die viel grundsätzlichere Schwierigkeit, wie die dauerhafte Vorhaltung einer Datenbank überhaupt sichergestellt werden kann, ist davon aber noch gar nicht berührt.⁶⁰ Angesichts der Fixierung der Wissenschaftsförderung auf kurzfristige Projektförderung sind selbst langjährig etablierte Angebote stets vom Aus bedroht.⁶¹ Zwar wird von der DFG zumindest bei Anträgen im Rahmen der LIS-Förderlinie die Beteiligung einer Infrastruktureinrichtung gefordert, doch muss bei nüchterner Betrachtung klar sein, dass auch diese nur eine begrenzte Zahl an verschiedenen technischen Plattformen betreuen können. Anders als ein Buch ist ein elektronisches Angebot nie wartungsfrei abgeschlossen, selbst wenn der Datenbestand als solcher nicht mehr verändert wird.⁶² Der Serverunterbau muss schon aus Sicherheitsgründen immer wieder auf einen aktuellen Versionsstand gebracht werden, was dann erfahrungsgemäß weitere Anpassung notwendig macht.⁶³ Hinzu kommt, dass selbst noch vor kurzem unverzichtbar und zukunftsicher erscheinende Basistechnologien ersetzt werden können, wie der weit vorangeschrittene Wechsel von Flash zu HTML5 zeigt.⁶⁴ Natürlich könnte man die jeweils zugrundeliegenden Daten in ein möglichst einfach lesbares Standardformat exportieren, doch würde dies bedeuten, dass die eigentlich gewünschte Abfragemöglichkeit bzw. die Auswertungen den Nutzern nicht mehr unmittelbar zur Verfügung stehen.⁶⁵ Form und Inhalt können in diesen Fällen nicht ohne Weiteres getrennt werden.⁶⁶ Erforderlich ist demnach

60 Zu Problemen der Langzeitarchivierung vgl. Thaller (2012) 18–19; Buddenbohm/Engelhardt/Wuttke (2016). Von diesem Problembereich naturgemäß besonders stark berührt sind elektronische Publikationsformen, die vornehmlich auf andere elektronische Angebote verweisen und diese aggregieren: Walkowski (2013) 47.

61 Vgl. etwa Feraudi-Gruénais/Grieshaber (2016). Ähnlich prekär ist die Situation im Bereich der Fachbibliographien.

62 Ob rein digitale Publikationsformen langfristig tatsächlich kostengünstiger sind (so z.B. Wampfler (2013) 90), wäre erst einmal empirisch nachzuweisen. Das Ergebnis wird sicher wesentlich von der Standardisierung digitaler Publikationsformen unter dem Aspekt effizienter Archivierbarkeit abhängen.

63 Vgl. Sahle / Kronenwett (2013) 89–90 zur Praxis am Kölner Data Center for the Humanities.

64 Beispiele für Opfer solcher Umstellungen lassen sich leicht finden: Das an der Universität Wien entwickelte Hypertext-Projekt *Past Perfect* zur Geschichte der Frühen Neuzeit ist nicht mehr zugänglich, die auf der Homepage für Herbst 2013 angekündigte Wiedereröffnung hat offensichtlich nicht stattgefunden. Das Projekt war vom Jubiläumsfonds der Österreichischen Nationalbank mit insgesamt 122.000 Euro gefördert worden.

65 Vgl. die Unterscheidung von Aufbewahrung und Nutzung bei Sahle/Kronenwett (2013) 83.

66 Vgl. Buddenbohm/Engelhardt/Wuttke (2016), Kap. 2. Dies ist umso problematischer, als derzeit nicht abzusehen ist, „ob sich in den Geisteswissenschaften überhaupt eine Kultur der Zweitverwertung einmal erhobener und erschlossener Daten auf der Ebene der Daten und nicht auf der Ebene der Präsentations- und Publikationsformen etablieren wird“

für elektronische Ressourcen ein System der funktionalen Langzeitarchivierung.⁶⁷ Angesichts objektiv limitierter Ressourcen wird dies aber auch bedeuten, dass eine stärkere Steuerung weg von jeweils individuell erstellten Systemen hin zu generischen Werkzeugen sowie Daten- und Schnittstellenstandards erfolgen muss.⁶⁸

Eine solche Entwicklung wäre auch insofern sinnvoll, weil die meisten Fachwissenschaftler als potentielle Inhaltslieferanten weder über die notwendigen technischen Kompetenzen noch über die materiellen Ressourcen verfügen, die für ihr Projekt notwendigen Plattformen aufzubauen (und vor allem dauerhaft zu pflegen).⁶⁹ Unter generischen Werkzeugen möchte ich dabei nicht die bloße Nachnutzbarkeit einzelner Softwarekomponenten verstehen, sondern die Zurverfügungstellung eines unmittelbar nutzbaren Dienstes. Hier sollten unter Einbeziehung der community bestimmte Musterfälle identifiziert werden, die im Rahmen solcher Angebote abzudecken wären (z.B. philologische Edition (ggf. mit Übersetzung), papyrologisch/epigraphische Edition, Fragmentsammlung, thematische Quellsammlung).

Das ist freilich noch weitgehend Zukunftsmusik. Solange die genannten Probleme aber nicht gelöst sind, bleibt der Verweis auf Datenbankabfragen zu Belegzwecken in wissenschaftlichen Werken unbefriedigend, da die exakte Reproduzierbarkeit in Frage steht. Vorläufig bleibt die Aufnahme der Daten auf dem konsultierten Stand in eine Fußnote oder einen Anhang die beste Lösung. Freilich können solche Anhänge mit Forschungsdaten einem Buch durchaus in elektronischer Form beigegeben werden.⁷⁰ Dies würde jedoch wesentlich erleichtert, wenn die Datenbanken durchgängig entsprechende Exportoptionen bieten würden.⁷¹

(Sahle/Kronenwett (2013) 82). Mit einer Archivierung reiner Rohdaten wäre folglich für den Großteil der community-Mitglieder wohl wenig gewonnen.

67 Damit ist eine Archivierung gemeint, die über das bloße Abspeichern von Rohdaten hinausgeht. Zu diesem letzteren Zweck befindet sich das Portal *IANUS* im Aufbau (<https://www.ianus-fdz.de>). Zu Problemen der Langzeitarchivierung vgl. Neuroth et al. (2012) mit Beiträgen zu den Geisteswissenschaften allgemein (Pempe (2012)) und den Altertumswissenschaften (Dally et al. (2012)), worunter hier allerdings ganz vorrangig die Archäologie bzw. Archäoinformatik verstanden werden.

68 Vgl. Buddenbohm/Engelhardt/Wuttke (2016), Kap. 3.3.

69 Dies gilt insbesondere für die kleineren und mittleren Universitäten. Vgl. Sahle/Kronenwett (2013) 84.

70 Exemplarisch sei auf den Aufsatz Feraudi-Gruénais 2017 verwiesen. Ergänzend erschien auf der Publikationsplattform des Fachinformationsdienstes *Propylaeum* ein Repertorium der Schiffsdarstellungen auf Grabdenkmälern der hellenistischen und römischen Zeit unter besonderer Berücksichtigung der Inschriften: <http://dx.doi.org/10.11588/propylaeum.77.71>. Die zugehörigen Forschungsdaten wurden unter <https://doi.org/10.11588/data/OOC0ZI> hinterlegt.

71 Einen Export von Treffermengen als CSV-Datei ermöglichen z.B. die *Epigraphische Datenbank Heidelberg* und die Münzdatenbanken unter <http://numismatics.org> (s. Anm. 7).

4 Kompetenzprobleme

Der Literaturwissenschaftler Philipp Theisoht malte vor einigen Jahren den Geisteswissenschaften eine düstere Zukunft aus, in der die Inhalte leicht für jedermann verfügbar im Internet zur Verfügung stünden und dem Wissenschaftler nur noch die narrativ verdichtende Paraphrase bliebe.⁷² Einmal abgesehen davon, dass mir der Begriff der Aneignung besser geeignet schiene als der der Paraphrase, um einen solchen Prozess zu umschreiben – die Sorge ist schon deshalb unbegründet, weil auch digitale Hilfsmittel nicht voraussetzungslos zu benutzen sind. Vor allem müssen die aufgefundenen Informationsbrocken verstanden, kritisch geprüft und kontextualisiert werden. In den Altertumswissenschaften beginnt das schon mit der Beherrschung der Quellsprachen. Ferner hängt die Qualität und Aussagekraft von Suchergebnissen wesentlich von den gewählten Suchbegriffen ab. Wie jeder weiß, der einmal Studienanfänger in die Benutzung bibliographischer Informationssysteme eingeführt hat, führt dies zu dem heuristischen Paradox, dass die Formulierung optimaler Suchbegriffe bereits ein hohes Maß an Sachkenntnis voraussetzt.⁷³ Für zielgerichtete Recherchen braucht es Kenntnisse des Fachvokabulars und der von der Forschung verwendeten Kategorien. Keineswegs lässt ein Internetanschluss allein bereits alle Informationsprobleme verschwinden.⁷⁴ Vielmehr stellen die digitalen Hilfsmittel eine Herausforderung dar, die erst einmal bewältigt werden will.

Auf einer höheren Ebene stellt sich jedoch die Frage, welches Maß an IT-Kompetenzen Historiker zukünftig benötigen werden. Im Bereich der Digital Humanities im engeren Sinne scheint mir ein Professionalisierungsprozess im Gange zu sein, der Fachwissenschaftler ohne informatische Grundlagenausbildung mittelfristig ausschließen wird, jedenfalls sofern sie nicht die Möglichkeit haben, eng mit Kooperationspartnern aus der Informatik zusammenzuarbeiten. Davon abgesehen werden Historiker generell ein Grundverständnis digitaler Ressourcen benötigen. Eine Arbeitsgruppe im Rahmen des Projektes *IANUS* hat jüngst Vorschläge dazu vorgelegt.⁷⁵ Demzufolge sollten Altertumswissenschaftler zukünftig grundlegende Kenntnisse unter anderem in Forschungsdatenmanagement, Urheberrecht und Lizenzen, Datenschutzrecht, Internetprotokollen, Metadatenstandards, Normdatenvokabularen und Dateninteroperabilität, Desktop-Publishing und

72 Theisoht (2013), bes. 30–32. Das von Theisoht ausgemachte Problem der Repetitivität geisteswissenschaftlicher Forschung hat in Wahrheit weniger mit spezifischen Auswirkungen der Digitalität zu tun, sondern ist eine Konsequenz daraus, dass (1) die Zahl der Geisteswissenschaftler im Zuge der Bildungsexpansion stark zugenommen hat und (2) zu zentralen Texten/Quellen nach einer teilweise mehr als zweihundertjährigen Forschungsgeschichte kaum noch Neues gesagt werden kann.

73 Vgl. Zaagsma (2013) 25.

74 Vgl. Haber (2013) 185.

75 <https://www.ianus-fdz.de/news/169>.

Archivierungsformaten, Zeichenkodierung, Auszeichnungssprachen, Grafikbearbeitung, Datenbankprogrammierung, Umgang mit Geo-Daten, Statistik, Netzwerkforschung mit Visualisierungstools und verschiedenen Programmiersprachen besitzen. Bei nüchterner Betrachtung ist ziemlich klar, dass dieser umfangreiche Katalog eher ein Kompetenzkatalog für einen Digital Humanities-Studiengang im Sinne der oben diagnostizierten Ausdifferenzierung ist, denn eine praktikable Blaupause für eine Ergänzung bestehender historischer Studiengänge.⁷⁶

Es ist nämlich trotz aller digitalen Herausforderungen unumgänglich, dass Historiker weiterhin auch Experten für historische, d. h. analoge, Wissenspraktiken sein müssen. Dies gilt jedenfalls mit Ausnahme der zeitgeschichtlichen Forschung, die eine zunehmend digital geprägte Gegenwart zukünftig nur noch mit digitalen Mitteln zweckmäßig erfassen können wird.⁷⁷ Gerade weil jedoch unsere Gegenwart immer mehr auf digitale Medien fokussiert ist, können im Umgang mit Überresten weiter zurückliegender Epochen erforderliche „analoge“ Basisqualifikationen nicht mehr einfach vorausgesetzt werden. Das gilt insbesondere dort, wo entweder ein erheblicher Teil der Quellenbestände (noch?) nicht digitalisiert ist oder textkritische Probleme eine Auseinandersetzung mit dem originalen Schriftträger oder dem originalen Objekt zwingend erfordern.⁷⁸

Im Rahmen einer vor allem von Mediävisten ausgelösten Debatte um den Wert der historischen Grundwissenschaften wurde auf die Paradoxie hingewiesen, dass in zahlreichen Datenbanken zwar viel Quellenmaterial digitalisiert und damit leichter zugänglich gemacht wird, gleichzeitig aber viele dieser digitalen Angebote ohne traditionelle Kompetenzen gar nicht benutzbar sind: Auch digitalisierte Manuskripte erfordern Paläographiekenntnisse.⁷⁹ Die großen altertumswissenschaftlichen Textcorpora kann nur verwenden, wer die entsprechenden Quellsprachen beherrscht. Die Anarbeitung dieser unverzichtbaren Sprach- und Lesekompetenzen

76 Jedenfalls, sofern nicht ein Hauptfach Geschichte mit einem Nebenfach Informatik oder Digital Humanities kombiniert wird.

77 Berry (2011) 4–5; Rieder et al. (2012) 67. Zu den Aufgaben und Problemen einer web history vgl. Brügger (2012).

78 Zum Verlust der Materialität durch Digitalisierung vgl. Zaagsma (2013) 25–26. Dem lässt sich entgegenhalten, dass dies nicht der Digitalisierung an sich, sondern einer defizienten Form von Digitalisierung zuzuschreiben ist: Turkel (2011). Die Diskussion kann freilich nur von gegenwärtig vorhandenen Technologien ausgehen. Voraussetzung für eine „vollständige“ Digitalisierung, die auch vom menschlichen Nutzer erfassbar sein soll, wäre eine Erweiterung der Computer-Interfaces mindestens um eine dreidimensionale und taktile Ausgabe. Eine so angelegte Digitalisierung wäre allerdings auch sehr viel aufwändiger als die gegenwärtig praktizierten Modelle.

79 <https://www.hsozkult.de/text/id/texte-2890>, besonders <https://www.hsozkult.de/debate/id/diskussionen-2866>. Dies gilt wenigstens solange, bis ausreichend treffsichere Verfahren computerisierter Handschriftenerkennung zur Verfügung stehen. Dazu vgl. Wettlaufer (2016) Kap. 4.

erfordert jedoch angesichts schwindender schulischer Vorkenntnisse eher mehr denn weniger Studienzeit. Ein Ausspielen analoger gegen digitale Kompetenzen scheint mir daher nicht zielführend. Vielmehr führt an Ausdifferenzierung und Institutionalisierung der Digital Humanities als eigenes Fach meines Erachtens kein Weg vorbei.⁸⁰ Nur so können letztlich auch passgenaue Studienangebote zur Verfügung gestellt werden, die in Form von Modulpaketen, Nebenfächern oder Zertifikatskursen konventionelle historische Studiengänge ergänzen können. Dabei muss aber bewusst bleiben, dass archäologische, literaturwissenschaftliche, philosophische oder rechtsgeschichtliche Perspektiven der historischen Erkenntnis genauso förderlich sein können wie informatische.

Dies bedeutet aber auch, dass die Entwicklung digitaler Werkzeuge stärker auf die Bedürfnisse ihrer potentiellen Nutzer in den Fachwissenschaften Rücksicht nehmen müssen. Andernfalls droht ein Zustand, in dem Datensammlungen und Auswertungswerkzeuge vorrangig deshalb realisiert werden, weil es eben technisch möglich ist, ohne dass auf der genuin geisteswissenschaftlichen Ebene tatsächlich ein Erkenntnisfortschritt einträte. Nutzerorientierung meint dabei in erster Linie, dass viel mehr Wert auf Nutzeroberflächen gelegt werden.⁸¹ Dies ist auch eine grundlegende Frage des Verhältnisses zwischen Mensch und Maschine. Gegenwärtig müssen Textbestände, die für menschliche Rezipienten produziert und optimiert wurden, mit großem Aufwand in digitale Form überführt und in verschiedener Weise durch Metadaten und/oder tagging maschinenlesbar erschlossen werden. Dies bedeutet aber nichts anderes als dass Menschen nun nicht mehr für andere Menschen, sondern für den Computer schreiben. Dies führt zu der unbefriedigenden Situation, dass nicht nur die Abfassung des „Textes“ für den Autor mit zusätzlichen Mühen verbunden ist, sondern dieser „Text“ auch vom Autor selbst nur mit Mühe gelesen werden kann. Das direkte tagging in XML-Dateien wird sich daher in weiten Kreisen der Altertumswissenschaften vermutlich genauso wenig auf breiter Front durchsetzen wie der Textsatz mit LaTeX außerhalb der Naturwissenschaften oder die direkte HTML-Programmierung von Institutshomepages. An sich fraglos sinnvolle Standards wie EpiDOC werden daher erst dann für die Editionspraxis breitere Relevanz gewinnen, wenn ein Editor zur Verfügung steht, der ein tagging der edierten Texte ohne genauere Kenntnisse des EpiDOC-Standards erlaubt.⁸² Wer glaubt, dies sei unnötig, muss nur auf die Geschichte des PC schauen,

80 Zur Lehre in den Digital Humanities vgl. Hirsch (2012).

81 Der Katalog der IANUS-Arbeitsgruppe fordert genau das Gegenteil, nämlich die Fähigkeit zur „Implementation von neuen Funktionalitäten, die in Programmen möglichen grafischen Oberflächen (GUI) hinausgehen (sic!), um neue Methoden, Forschungsansätze und Arbeitsabläufe implementieren zu können“.

82 Vgl. die Einschätzung des übergeordneten TEI-Standards als „sehr komplex“ bei Sahle/Kronenwett (2013) 82.

dessen Erfolg auf das engste mit dem Aufkommen graphischer Nutzeroberflächen verbunden ist. Ferner muss es auch für Forscher an kleineren Instituten, die keine Unterstützung durch schlagkräftige IT-Abteilungen erwarten können, möglich sein, mit Hilfe eines solchen Editors kreierte standardkonforme Daten auch wieder anhand der EpiDOC-XML-Dateien in einer für menschliche Leser optimierten Publikationsform auszugeben, was wiederum auf die Etablierung zentral (und dauerhaft) bereitgestellter Dienste hinausliefere.

5 Zusammenfassung

Insgesamt scheint mir eine grundlegend epistemische Neuorientierung im Bereich der Alten Geschichte durch digitale Methoden nicht greifbar. Wohl aber wird die digitale Transformation der Arbeitswelt des Historikers weiter voranschreiten: Gedruckte Publikationen und Erschließungsmittel dürften mit zunehmender Geschwindigkeit durch digitale Angebote abgelöst werden. Dabei sollte der Wert hochwertig kuratierter Daten stärker gewürdigt werden, insbesondere müssen textkritische Standards endlich auch für digitale Datenbanken konsequent umgesetzt werden. Die Nachhaltigkeit digitaler Ressourcen kann nur durch stärkere Standardisierung und Konzentration auf zentrale Plattformen gewährleistet werden, die sich als Dienstleister für möglichst viele externe Inhalteanbieter verstehen. Dies würde freilich voraussetzen, bestimmte Dienste als elementar für eine fachwissenschaftliche community zu definieren und für diese Daueraufgaben dann auch eine entsprechende langfristige Finanzierung bereitzustellen. Insofern wird die Digitalisierung nicht zuletzt eine Herausforderung für eine fast vollständig auf kurzfristige Projekte ausgerichtete Forschungsförderung darstellen.⁸³

83 Verwiesen sei darauf, dass zuletzt sogar die ehemaligen Sondersammelgebiete für die überregionale Versorgung mit Spezialliteratur durch die projektorientierten Fachinformationsdienste abgelöst wurden, deren Laufzeit drei Jahre beträgt: http://www.dfg.de/foerderung/programme/infrastruktur/lis/lis_foerderangebote/fachinformationsdienste_wissenschaft.