

Beitrag aus:

Sonderband 2 der ZfdG: Digitale Metamorphose: Digital Humanities und Editionswissenschaft. Hg. von Roland S. Kamzelak und Timo Steyer. 2018. DOI: [10.17175/sb002](https://doi.org/10.17175/sb002)

Titel:

Von der Raupe zum Schmetterling oder Wie fliegen lernen – Editionsphilologie zwischen Infrastruktur und Semantic Web

Autor/in:

Roland S. Kamzelak

Kontakt:

roland.kamzelak@dla-marbach.de

Institution:

Deutsches Literaturarchiv Marbach

GND:

[173262678](https://nbn-resolving.org/urn:nbn:de:hbz:5:1-73262678)

ORCID:

[0000-0003-4512-2047](https://orcid.org/0000-0003-4512-2047)

DOI des Artikels:

[10.17175/sb002_004](https://doi.org/10.17175/sb002_004)

Nachweis im OPAC der Herzog August Bibliothek:

[1012353206](https://nbn-resolving.org/urn:nbn:de:hbz:5:1-73262678)

Erstveröffentlichung:

07.03.2018

Lizenz:

Sofern nicht anders angegeben 

Medienlizenzen:

Medienrechte liegen bei den Autoren

Letzte Überprüfung aller Verweise:

07.03.2018

GND-Verschlagwortung:

[Edition](#) | [Semantic Web](#) |

Zitierweise:

Roland S. Kamzelak: Von der Raupe zum Schmetterling oder Wie fliegen lernen – Editionsphilologie zwischen Infrastruktur und Semantic Web. In: Digitale Metamorphose: Digital Humanities und Editionswissenschaft. Hg. von Roland S. Kamzelak /Timo Steyer. 2018 (= Sonderband der Zeitschrift für digitale Geisteswissenschaften, 2). PDF Format ohne Paginierung. Als text/html abrufbar unter DOI: [10.17175/sb002_004](https://doi.org/10.17175/sb002_004).

Roland Kamzelak

Von der Raupe zum Schmetterling oder Wie fliegen lernen – Editionsphilologie zwischen Infrastruktur und Semantic Web

Abstracts

Die Editionsphilologie gehört zu den ersten Motoren für die Einführung digitaler Methoden in den Geistes- und Sozialwissenschaften. Für Editionen werden Infrastrukturen aufgebaut, die zu einem neuen Typus von Edition führen können, wenn die Daten jeweils so aufbereitet werden, dass sie editionsübergreifend nutzbar sind. Der Nachnutzung der Daten kommt eine immer größere Bedeutung zu, doch es ergeben sich daraus auch neue Probleme. Beispielhaft werden im Beitrag Chancen und Probleme von übergreifenden Editionen anhand der Infrastruktur AMIE dargestellt.

Scholarly editions are one of the first engines for introducing digital methods into the humanities. Digital infrastructures are being developed for single editions, which could lead to a new type of edition by making data available across different editions. Reuse of data becomes increasingly a subject, but there are also new problems involved. In this essay chances and risks of cross-reference editions are discussed by describing the infrastructure AMIE.

1. Virtuelle Forschungsinfrastruktur

Seit der Wissenschaftsrat 2011¹ aufgerufen hat, virtuelle Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften auszubauen, floriert der Begriff Virtuelle Forschungsumgebung (VFU)² in Anträgen und in Projektbeschreibungen. Dabei sind seit ersten Forderungen aus den Fachdisziplinen³ bis heute kaum funktionierende VFU im produktiven Einsatz. Einzig TextGrid⁴ kann als mustergültige Lösung angesehen werden, da von der Datenerhebung über die Werkzeuge bis hin zur Langzeitarchivierung⁵ alle notwendigen Funktionen von VFU abgedeckt scheinen. Eine Untersuchung, warum Projekte sich dieses Instrumentes nicht häufiger bedienen, steht noch aus. Vielmehr werden vielfach hauseigene Systeme entwickelt⁶ oder an verschiedenen Orten einzelne Module bereitgestellt.⁷ Generische Systeme, die zu einer VFU zusammenwachsen, scheint es trotz der Initiativen DARIAH⁸ und CLARIN-D⁹ in den Geistes- und Sozialwissenschaften nicht zu geben.

¹ Empfehlungen zu Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften 2011; siehe auch Forschungsinfrastruktur für die Geistes- und Sozialwissenschaften 2013.

² Für eine erste Definition siehe Wikipedia: Virtuelle Forschungsumgebung 2016.

³ Vgl. etwa Baßler / Karczewski 2009, S. 27–34.

⁴ Siehe TextGrid und Rapp 2006, S. 61–67 und Kerzel et al. 2009, S. 36–39.

⁵ Für eine knappe Darstellung von Langzeitarchivierung siehe Edlex: Langzeitarchivierung 2017.

⁶ Als ein Beispiel soll hier FuD, Universität Trier, genannt werden. Siehe FuD.

⁷ Vgl. Unsworth 2000, der dies als den richtigen Weg ansieht.

⁸ Siehe DARIAH-DE. Digitale Forschungsinfrastruktur für die Geistes- und Kulturwissenschaften.

⁹ Siehe Clarin-D.

Die Arbeitsgruppe ›Virtuelle Forschungsumgebungen‹ der Allianz der deutschen Wissenschaftsorganisationen hat 2011 definiert:

»Eine Virtuelle Forschungsumgebung (Virtual Research Environment - VRE) ist eine Arbeitsplattform, die eine kooperative Forschungstätigkeit durch mehrere Wissenschaftlerinnen und Wissenschaftler an unterschiedlichen Orten zu gleicher Zeit ohne Einschränkungen ermöglicht. Inhaltlich unterstützt sie potentiell den gesamten Forschungsprozess – von der Erhebung, der Diskussion und weiteren Bearbeitung der Daten bis zur Publikation der Ergebnisse – während sie technologisch vor allem auf Softwarediensten und Kommunikationsnetzwerken basiert. Virtuelle Forschungsumgebungen sind wesentliche Komponenten moderner Forschungsinfrastrukturen und spielen eine entscheidende Rolle für die Produktivität und Wettbewerbsfähigkeit der Forschung.«¹⁰

Der Wissenschaftsrat erhoffte sich durch »gänzlich neue Fragestellungen« einen »wesentlichen Beitrag zum Erkenntnisgewinn« besonders in den Geistes- und Sozialwissenschaften. Dabei ist nicht nur der Zweig Digital Humanities gemeint, sondern die ganze Breite der Geistes- und Sozialwissenschaften. Schaut man sich aber beispielsweise die Forderungen von Baßler / Karczewski von 2009 und das Resumé von Stäcker aus dem Jahr 2013¹¹ an, so scheint das Fach Germanistik noch sehr weit von einer qualitativen Verbesserung der Forschung entfernt zu sein: Ausgehend vom cultural turn der 80er Jahre wünschen sich Baßler / Karczewski ein digitales Archiv aller relevanten Texte, wobei auf jegliches Tagging verzichtet werden soll, da Tagging und Indices ihrer Meinung nach veralten. Stattdessen soll die Suche im Archiv im Vordergrund stehen, die Trefferlisten generiert. Die relevanten Suchanfragen sollen aber gespeichert werden können und erweiterbar sein. Intuitive Handhabung gehört ebenfalls zu den Forderungen, um die Lust an der Arbeit nicht zu verlieren. Die Frage, die sich dabei jedoch stellt ist, ob eine gespeicherte Suche nicht dem Tagging eines Textes zumindest sehr ähnlich ist und ebenfalls veraltet. Tagging hat dabei den Vorteil, dass Forschungsergebnisse direkt in den Text einfließen und allen sofort zur Verfügung stehen, während gespeicherte Suchanfragen ohne Kontext zunächst einmal unverständlich sind und auch ohne Ergebnis bleiben können. Auch, wie kann in nicht ausgezeichneten Texten etwa nach uneigentlichen Namen (beispielsweise findet sich in Texten schlicht ›G.‹ für Goethe oder ›meine Frau‹ für etwa Helene von Nostitz) gesucht werden? Normalisierung, die Auflösung von Abkürzungen und die Einbeziehung und Verarbeitung von fremdsprachlichen Texten sind für die gewünschte Bearbeitung absolut notwendige Grundvoraussetzungen. Mit dem Text von Baßler / Karczewski scheint ein Indiz vorzuliegen, dass rein analog arbeitende Forscher die Mechanismen elektronischer Texte nicht nutzen wollen oder können, weil die Verfahren nicht genügend bekannt und allgemein benutzbar sind.¹² Deshalb auch die Forderung nach möglichst intuitiver Bedienung elektronischer Ressourcen. Dabei sind Institutionen wie Archive und Bibliotheken auch nicht intuitiv benutzbar. Vielmehr ergeben sich erst mit einiger Archiverfahrung geeignete Mechanismen, um relevantes Material für seine Arbeit zu finden. Das ist in diesem Sinne mühsamer als reine Textarbeit. Der Forscher muss zu den

¹⁰Vgl. Schwerpunktinitiative "Digitale Information" der Allianz der deutschen Wissenschaftsorganisationen. Leitbild 2013-2017, S. 9-11.

¹¹Vgl. Baßler / Karczewski 2009 und Stäcker 2013.

¹²Es ist eine traurige Tatsache, dass die Dokumentation von Software, auch das intuitive Nutzerinterface, bei der Entwicklung in engen Projekt- und Finanzierungszeiträumen meist nicht mehr umgesetzt werden.

Archiven reisen, diese Reisen organisieren und finanzieren, und dabei ganz unterschiedliche Öffnungszeiten beachten, schwer lesbare Handschriften in verschiedenen Zeichensystemen entziffern und meist zuerst auch ganz unterschiedliche archivalische Ordnungssysteme verstehen, um überhaupt relevantes Material in den Lesesaal zu bekommen. Auch hier scheint das Indiz virulent, dass digitale Techniken noch nicht ins Fach eingedrungen sind, denn die Mühen der analogen Recherche fallen bei der Beurteilung nicht ins Gewicht, während im Digitalen alles intuitiv und einfach sein soll.

Stäcker geht im Ansatz wesentlich weiter: Für ihn ist die Arbeit mit digitalen Ressourcen alltäglich und richtig. Er beklagt jedoch, dass zwischen In- und Output unterschieden wird. Um den Kreislauf der Wissenschaft zu schließen und die bestmögliche Nutzung von wissenschaftlichen Ergebnissen zu erreichen, muss auch elektronisch publiziert werden. Erst durch die direkte Aufbereitung von Ergebnissen mit Metadaten und der Verknüpfung mit den verwendeten Ressourcen schließt sich dieser Kreis. Dabei spielt Open Access für ihn eine große Rolle. Sein Beitrag zeigt sehr eindrücklich, dass der Weg zu einer digitalen Forschung in den Geistes- und Sozialwissenschaften noch sehr weit ist. Dazu gehört auch die Erforschung der Veränderung von Forschungsgegenständen selbst. Was macht Digitalität mit den Forschungsgegenständen? Wie verändert sie die Rezeptions- und Produktionsprozesse etc.?

Auf dem Weg zu einer digitalen Geistes- und Sozialwissenschaft muss man auch diesen Fragen nachgehen: Welche Chancen und Risiken birgt Open Access? Welchen wissenschaftlichen Nutzen bringt die Digitalisierungswelle¹³? Welche Nachteile birgt sie? Gibt es geeignete Konzepte für die Langzeitarchivierung? Gelingt die Langzeitverfügbarkeit? Welche Metadaten, welche Normdaten sind verlässlich? Wie kann man verlässlich ausgezeichnete Daten zu einem Semantic Web zusammenschließen? Mit welchen Suchmaschinen oder Visualisierungen kann man neue Forschungsfragen entwickeln? Eröffnet Crowdsourcing neue, auch qualitativ hochwertige Ergebnisse? Und schließlich: Wie generisch sind Projektergebnisse wirklich angelegt?

Hier soll der Blick fokussiert werden auf einen Bereich, der seit längerer Zeit bereits elektrifiziert und interdisziplinär ist: die Editionsphilologie. Editionen sind als Quellenmaterial Ausgangspunkt für die Forschung, Input für Normdatenpools und Lieferant von geeigneten Programmen (Suchmaschinen, Anzeigesoftware, Auszeichnungshilfen usw.). Zwar gibt es Ansätze in allen diesen Bereichen des Forschungsablaufs, doch amalgamieren sie nicht zu einem gemeinsam nutzbaren Forschungsraum.¹⁴ Auch ein intensiver Blick in die USA, dem Ursprungsland des wohl einzigen weltweiten de facto Standards für die Auszeichnung von elektronischen Texten, der TEI¹⁵, hat keine funktionierenden Virtual Research Environments

¹³Mit Digitalisierung ist nicht schlicht die Transformation von analogen Trägermedien in digitale als Abbild gemeint (Scan, Audio- oder Bilddatei), sondern die Anreicherung durch Metadaten, die Hinterlegung von normalisierten Volltexten, Transkriptionen usw. und etwa die semantische Aufbereitung durch Verknüpfungen.

¹⁴Aus diesem Grund habe ich für das [Verbundprojekt MWW](#) den Aufbau einer Virtuellen Forschungsinfrastruktur für die Arbeit mit literarischen Quellen vorgeschlagen, die von 2013-2018 aufgebaut wurde und in der zweiten Förderphase ausgebaut werden soll; Workshop in Frankfurt am Main mit dem Vortrag »Forschungsraum. Virtual Research Environment for Archives and Libraries«.

¹⁵Siehe [TEI Text Encoding Initiative](#).

erkennen lassen.¹⁶ Noch 2014 war es schwer, Vertreter der Digital Humanities an amerikanischen Universitäten und Bibliotheken zu finden.¹⁷ Inzwischen hat sich dieser Forschungszweig auch in den USA ausgebildet und ist zu einer festen Größe geworden.

2. Von der Edition des Tagebuches von Harry Graf Kessler zur Infrastruktur AMIE

Die Edition des Tagebuches von Harry Graf Kessler hat 1994 begonnen. Bei einem Umfang von 16.000 Seiten handschriftlichem, teilweise sehr schwer lesbarem Material und der Erwähnung von weit über 10.000 Personen, mehr als 4.000 Orten und 4.000 Werken der bildenden Kunst und Literatur, ist eine Laufzeit von über zwanzig Jahren nicht selten.¹⁸ Damit geht auch einher, dass bis heute einige Systemumstellungen notwendig waren. Das Tagebuch ist ein Kaleidoskop des kulturellen Lebens der Jahrhundertwende und das Personenregister liest sich wie ein Who is Who der Zeit zwischen 1900 und 1937.¹⁹ Da bereits 1994 an Auszeichnungen nach den Richtlinien der TEI gedacht wurde, bezogen sich die notwendigen technischen Umstellungen ausschließlich auf die Textverarbeitungsprogramme, nicht auf das verwendete Format, das seit dem Jahr 2000 direkt XML ist.²⁰

Ein wichtiges Element der Edition war von Anfang an das Register. Sehr früh war die Erkenntnis da, dass die Aufnahme von Registerbegriffen in einem möglichst frühen Stadium in ein zentrales Instrument für die Arbeit am Text hilfreich, wenn nicht sogar notwendig ist. So kann eine Nennung wie »Gee« ohne Probleme in eine Datenbank²¹ aufgenommen werden, auch wenn der Eintrag in der Edition kein Lemma im Register werden wird. Für die Referenzierung dient dann von Anfang an die Identnummer der Datenbank, obwohl der Haupteintrag sich durch tiefere Recherche und weiteren Erkenntnissen aus dem Text zu »Brion, Wilma Karoline Louise Alice, Marquise de« ändert. Für die interne Recherche bleibt die ursprüngliche Nennung »Gee« als Verweisung erhalten.

Was aber meist über die Projektlaufzeit hinaus nicht erhalten bleibt, sind die vielen Recherchematerialien. Die mit viel Aufwand recherchierte und vor Ort eingesehene Geburtsurkunde, um das wahre Geburtsdatum einer Person zu verifizieren, fließt als blanke Jahreszahl ins Register ein. Im nächsten Projekt könnte man sich zwar auf die erste Edition

¹⁶Für diesen intensiven Blick hat mir die [Gerda Henkel Stiftung](#) 2014 ein Forschungsstipendium bewilligt, das ich am [Maryland Institute for Technology in the Humanities \(MITH\)](#) verbracht habe. Der Stiftung und den Mitarbeitern des MITH gilt mein ausdrücklicher Dank. Viele der Anregungen aus den Gesprächen dort sind in die Entwicklung von AMIE eingeflossen.

¹⁷Unter den frühen Vertretern des Faches, denen ich allen für wichtigen Input danke, sind zu nennen: Neil Freistat (College Park), Mathew Kirschenbaum (College Park), Tanya Clement (Texas), Mitch Fraas (Philadelphia) und Andrew Goldstein (Princeton / New Brunswick).

¹⁸Kamzelak / Ott 2004–2018. Band I, der letzte noch fehlende Band, ist in Vorbereitung und erscheint im Herbst 2018.

¹⁹Über die Bedeutung des Tagebuches für die Geschichte, Literatur- und Kunstgeschichte ist seit den 60er Jahren viel publiziert worden. Hier soll stellvertretend auf zwei neuere Publikationen verwiesen werden: Kamzelak et al. 2015 und Kamzelak 2015a.

²⁰Zunächst wurden in MS Word Druckformatvorlagen zur Textauszeichnung verwendet, die 2000 in XML/TEI automatisch durch Makrobefehle umgewandelt werden konnten.

²¹Das Projekt arbeitet zunächst mit der DOS-Datenbank Allegro-C, dann mit MS Access 2.0 und schließlich MS Access 97.

berufen, wird aber eher nochmals ins Geburtsregister schauen, um sicher zu sein. An dieser Stelle setzt die Idee zur Entwicklung einer Forschungsumgebung aus der ursprünglich nur für ein einziges Projekt gedachten Datenbank ein, denn für die Edition des Tagebuches wurden auch Kesslers Briefe eingesehen, teilweise transkribiert, da es natürlich viele Verbindungen zwischen Tagebuch und Briefen gibt. Über eine zentrale Datenbank lassen sich diese Verbindungen in der Recherche gegenseitig nutzen. Die Datenbank AMIE²² sammelt, neben den Lemmata, alle Rechercheergebnisse aller Projekte, so dass alle gegenseitig voneinander profitieren. Dabei stehen allen Nutzern nicht nur die in die Edition eingeflossenen Daten zur Verfügung, sondern alle Hintergrundinformationen dazu. So kann die Briefedition Harry Graf Kessler – Wilma de Brion dieselbe Identnummer für »Wilma« verwenden wie die Tagebuchedition. Somit ist ein Grundstein für die Verbindung der beiden Editionen gelegt. Hinterlegt man überall, wo dies möglich ist, auch noch Normdaten wie die GND-Nummer²³ für Personen, Werktitel, Körperschaften und Sachbegriffe, so lassen sich Verbindungen auch außerhalb der Editionen selbst legen. Der gegenseitige Nutzen ist evident: Je näher die eigene Edition zeitlich und thematisch an bereits weiter fortgeschrittenen Editionen ist, desto größer die Wahrscheinlichkeit, dass Recherchen bereits erledigt sind. Ggf. muss nur die neueste Forschung noch berücksichtigt werden. Verknüpfen genügt, um einen qualifizierten Registereintrag zu erzeugen. Dabei sind auch diejenigen Daten wichtig, die selbst nicht registerfähig sind, wie im oberen Beispiel »Gee«, oder Nennungen wie »Onkel Fritz« usw.

3. Von der Infrastruktur zum Semantic Web

Die Datenbank, die unter der Editionsinfrastruktur AMIE steckt, im Detail zu beschreiben ist hier nicht der richtige Ort. Es ist eine Oracle Datenbank mit der Oberfläche APEX. APEX ist vielseitig und eingeschränkt zugleich, wie wahrscheinlich jede Applikation. Positiv zu bewerten ist der schnelle Einstieg in die Konfiguration und Präsentation. Es gibt positive Standardfeatures wie die gute kaskadierende Suche, einfache Downloads in verschiedenen Formaten und vielfache Anpassungsmöglichkeiten der Ansichten durch die Nutzer selbst. Eingeschränkt sind die Navigation durch interne Links und das Anzeigen aus verschiedenen Tabellen. Mit normalen Mitteln ist das Taggen und Verlinken innerhalb von Feldern nicht möglich, z. B. im Erläuterungsfeld. Dieser Nachteil wurde allerdings in AMIE inzwischen zu einer Stärke ausgebaut, durch die Verwendung von Relationen. Diese können leicht auch als RDF-Triple, eine Grundlage für das Semantic Web, ausgegeben werden. Letztlich ist es aber auch nicht so entscheidend, welches Instrument verwendet wird. Die Chancen und Risiken messen sich anders:

So gibt es bei der mehrere Editionen erfassenden Arbeit qualitativ neue Probleme mit der Ansetzung: In AMIE wird die in Deutschland gebräuchliche Form angesetzt (wie es bibliothekarisch nach RAK²⁴ Regel ist). So ist aber der Originaltitel manchmal der Haupteintrag, manchmal der Nebeneintrag. Dieses Problem ist bei analogen Editionen zwar ebenfalls

²²AMIE steht für administro editiones, eine Oracle APEX-Applikation am Deutschen Literaturarchiv Marbach, entwickelt von Roland S. Kamzelak unter Mitarbeit von Angelika Kreh.

²³In AMIE sind die GND-Nummer und Geokoordinaten Pflicht. Es können (und sollen) aber alle Normdaten hinterlegt werden, die bei den Recherchen gefunden werden.

²⁴Vgl. Regeln für die alphabetische Katalogisierung in wissenschaftlichen Bibliotheken (RAK-WB) 2002.

vorhanden, potenziert sich durch die Menge jedoch bei übergreifenden digitalen Editionen. Auch ist die gebräuchliche Form zeitabhängig: »Schuld und Sühne« oder »Raskolnikow« wie bei Kessler genannt? Oder ganz neu »Verbrechen und Strafe«. Freilich kann von dem einen auf die anderen Titel verwiesen werden, aber irgendwo müssen die Daten und Erläuterungen angehängt werden und verleihen dem Eintrag dadurch Gewicht. Konsistenz ist über die Zeit so kaum zu erreichen. Ändert man seine Meinung und setzt an, was in einer anderen Zeit gebräuchlich ist, dann ergeben sich ggf. Änderungen an anderen Editionen unter der Hand. Da das nicht sein darf, erschließt man nach Edierzeit, also »was zum Zeitpunkt der Aufnahme gebräuchlich« war, was wiederum schnell zu Schiefklagen führen kann, weil für eine Edition nicht alles neu aufgenommen wird, sondern vorhandene Sätze verwendet, und damit Einträge vor der Edierzeit.

Die digitale Welt kennt zwar keine Grenzen, doch wie kann man international werden, wenn man nicht alle Sprachen beherrscht und auch nicht überall hin mappen kann, weil es nicht überall nationale Normdaten gibt?²⁵ Soll man hier einfach auf die Zukunft vertrauen? Soll man strikt nationale Ausgaben herstellen und auf künftige Konkordanz über nationale Normdaten vertrauen? Diese Fragen stellen sich erst bei übergreifenden Editionen und sind bislang nicht diskutiert worden.

Ein zweiter, ähnlicher Problemkomplex bezieht sich auf die Entscheidung, wann welche Namensform verwendet wird. »Hardenberg« oder »Novalis«? Gebräuchlich ist Novalis, also Haupteintrag bei der Tagebuchedition. In Briefen wird der Autor jedoch auch als Hardenberg angesprochen. In AMIE zeigt Hardenberg wegen der früher bearbeiteten Tagebuchedition jedoch immer auf Novalis, das kann in einem Briefwechsel seltsam wirken. Auch wäre es interessant zu wissen, wo Hardenberg, wo Novalis verwendet wird. Das aber stellt nochmals hohe Ansprüche an das Mapping innerhalb der Datenbank.

Bei zwei Alternativen mag das zu bewältigen sein. Wie ist es aber zum Beispiel bei Kesslers Schwester »Wilhelma« oder »Wilma«. Kessler nennt sie »Gee« bis zu seinem Tod. Als Junge aber auch »Bee« und »Baby«. Verschreiber wie »Be« oder »Bie« sind auch dabei. Müssen diese Varianten alle in die Datenbank aufgenommen und müssen sie direkt referenziert werden können? Das wird zwar sehr unübersichtlich, wäre aber eine gute Informationsquelle für die Nutzer. Vereinheitlichung war bei analogen Editionen stets eines der Ziele, ggf. auch zum Preis von »Datenverlust«, digitale Register können alle Nennungen aufnehmen, doch muss man sich beim Zusammenführen des Anzeigeprogramms auf eine adäquate und richtige Wiedergabe verlassen können. Und welcher Philologe kann den dahinterliegenden Algorithmus verlässlich prüfen?

Auch bei Ortsnamen treten Probleme auf: Durch die Geodaten kann ich Kilchberg (Schweiz) von Kilchberg (Deutschland) unterscheiden, doch »Kilchberg«, als Thomas Mann von dort aus schrieb, lag »bei Zürich«, heute gehört es zu Zürich. Blankenese bei Hamburg ist heute Teil von Hamburg. Hollywood ist offiziell Teil von Los Angeles. Sammelt man die Referenzen beim

²⁵Eine Übersicht verfügbarer nationaler Normdaten findet sich bei [VIAF \(Virtual International Authority File\)](#).

Haupteintrag, also Los Angeles, so wirkte das gerade bei der Filmstadt Hollywood befremdet oder zu technisch, auch wenn es korrekt wäre.

Ein weiterer Problemkomplex betrifft die Erläuterungstiefe bei Relationen: Bei Personen, die im Text genannt werden, kann man bei der Verzeichnung von Vater und Mutter aufhören (also nicht noch vom Vater und der Mutter jeweils auch deren Vater und Mutter verzeichnen), weil damit genügend Informationen generiert werden, um Geschwister, Cousins, Onkel usw. zu errechnen. Bei Orten zeigen sich weitere Probleme mit Hierarchien: Kilchberg ist Teilort von Zürich. Zürich ist Teilort der Schweiz. Die Schweiz ist Teilort von Europa. Europa ist Teilort der Welt. Auch wenn die Kette unmittelbar einleuchtet, so fehlen doch z. B. die Kantone. In den USA die Countries. In Deutschland die Länder, Landkreise, Regierungsbezirke. Auch Gaue, die im Dritten Reich politische Einheiten waren, sind heute wieder Einheiten im Bereich des Sports. Orte haben eine historisch-politische Dimension, die bei einer übergreifenden Registererfassung erst sichtbar zu einem Problem wird. Bislang haben Kontaktaufnahmen zu Historikern und Geografen dieses Problem noch nicht gelöst.²⁶ Für die Visualisierung von Briefen etwa sind historische Karten ein Desiderat; generell auch Geodaten für Flächen. Und alle mit Geodaten versehenen Einträge müssen mit einem Zeitstempel kombiniert sein, denn beispielsweise die Altstadt von Jerusalem ist eben heute nicht am selben Ort wie vor 2.000 Jahren.

Ohne die Metapher »digitale Metamorphosen« zu sehr belasten zu wollen – schließlich wären die »analogen« Editionen im Bild die Raupe, langsam, gefräßig, grau und unansehnlich, die digitalen Editionen dagegen die schönen, anmutigen, flinken Wesen – geht es darum, was sich beim Editionsprozess im Digitalen und mit digitalen Infrastrukturen verändert. Nicht alles ist automatisch positiv. Beim Text selbst gilt oder sollte gelten, was Bodo Plachta zu den Standards ausgeführt hat²⁷. Altmodischer aber wichtigster Begriff ist dabei die »Sorgfalt«. Sie muss Grundlage jedes editorischen Schrittes sein, ob im »analogen« oder im digitalen Editionsprozess. Bei einer digitalen Edition muss Sorgfalt aber auch im Bereich der Nachnutzung und des Einbindens in neue Zusammenhänge walten, sind die Möglichkeit zur Nachnutzung und zur weiteren Kontextualisierung doch wichtige Bestandteile einer modernen Edition.

Die Entwicklung von der Kessler Tagebuchedition hin zur Infrastruktur AMIE steht beispielhaft für alle virtuellen Forschungsumgebungen. Sie ist eine Bewegung vom Einzelnen zum Allgemeinen oder Übergreifenden. Bei einer analogen Text-Edition hat der Text alles regiert. Die Editionsrichtlinien für diesen Text haben die Ansetzung, die Register, die Kommentare, die Erläuterungstiefe bestimmt. Bei übergreifenden Editionen regiert nun in großen Teilen der eigentlich aparte Text einer Edition: das Register und die Erläuterungen, weil diese eben nicht nur einem einzigen Text dienen, sondern vielen. Wo früher der einzelne Editor für die Bearbeitung oder die Herausgabe eines Textes zeichnete, arbeiten heute viele kollaborativ am Werden einer Edition. Die Strukturen ändern sich im Digitalen zum

²⁶ Dieses Thema wäre einen eigenen Workshop wert.

²⁷ Plachta 2018.

moderierten Crowdsourcing. Neil Freistat nennt dies den *participatory turn*.²⁸ Das geht so weit, dass der Editor auf Erläuterungen zu Gunsten eines persistenten Links auf qualifizierte Inhalte verzichten kann. Damit steht die Edition in einer neuen Ordnung und in neuen Strukturen. Dies mag noch nicht so sichtbar sein, weil die *Synapsen* (qualifizierte Links oder APIs) noch nicht alle ausgebildet sind. Aber in der Tendenz ist das bereits der Fall. Sobald die Datenbank AMIE als RDF-Beacon freigegeben wird, werden sich neue Strukturen ergeben können.²⁹ Die Daten werden Teil des Semantic Web. Früher haben Editoren – man verzeihe mir den harten Ausdruck – viel abgeschrieben, jetzt übernehmen sie direkt durch Verlinken zu verlässlichen Quellen. Das beschleunigt den Editionsprozess an vielen Stellen, auch wenn an anderen die Arbeit vermehrt wird. Bei Editionen muss dabei teilweise für ein Lemma viel tiefer recherchiert werden, als das im Textzusammenhang notwendig wäre, nur weil die Richtlinie von AMIE dies zu Gunsten anderer Editionen oder Zwecke vorschreibt. Potentiell jedoch ergibt sich insgesamt eine Effektivitätssteigerung durch gute Infrastrukturen, da man viele Informationen direkt übernehmen kann.

Wichtiger aber als eine Effektivitätssteigerung innerhalb einer Edition ist die Nachnutzung in anderer Hinsicht: Zwar wird auch auf den TEI-Tagungen selbst immer wieder bezweifelt, dass TEI-Dokumente tatsächlich so sehr nachgenutzt werden, ja nachnutzbar sind,³⁰ die Metadaten der Datenbank AMIE jedoch lassen sich heute schon vielfach nachnutzen. Kesslers Itinerar etwa lässt sich problemlos mit wenigen Klicks visualisieren, wenn man die Registerdaten als csv-Datei ausspielen kann.³¹ Personen aus der Edition lassen sich spielend nach verschiedenen Kriterien gruppieren, nach Geschlecht, nach Beruf, nach Konfession, nach Geburtsort, nach allem, was als Relation beigegeben ist. Durch die Änderung von Stellenkommentar zu strukturierter Relation – die innerhalb der Edition eine Kommentarfunktion übernimmt wie früher die Erläuterung selbst – lassen sich vielfältige Nachnutzungen realisieren. Kollaborativ ist in diesem Sinne nicht nur die arbeitsteilige Erstellung der Begriffe in der Datenbank, sondern auch die Nutzung dieser Daten außerhalb der Edition selbst.

Ein Problem bei allen online verlinkten Editionen oder sogar Texten allgemein ist die Grenzziehung. Wie das World Wide Web prinzipiell grenzenlos ist, so ist die digitale Edition grenzenlos. Wie kann man sie lesen? Wie kann man eine Idee, ein Konzept vermitteln, wenn ständig Grenzen überschritten werden können? Auch die Frage der Qualitätssicherung ist zu stellen, denn alle Inhalte sind im Fluss. Konsistente Adressen helfen da nur bedingt, wenn sich unter der Adresse Inhalte verändern können. Wie wird man sich schließlich durch die Versionen navigieren können und die Unterschiede verstehen? In einem frühen Text habe ich einmal den neuen Wissenschaftszweig Elektropaläographie³² prognostiziert, der heute als »digital forensics«³³ bekannt ist. Heute denke ich an einen neuen Zweig: die Digitale Wissenschaftsgeschichte, die sich zum Ziel setzt, das Verstehen zu einer bestimmten Zeit unter Rekonstruktion der sich wandelnden digitalen Quellen zu untersuchen.

²⁸Ich danke Neil für die Erlaubnis, seinen Begriff hier verwenden zu dürfen.

²⁹Dieses Vorhaben soll in der zweiten Phase des [Forschungsverbundes MWW](#) realisiert werden. S. [Fußnote 14](#).

³⁰Vgl. Burnard 2013.

³¹Siehe Kamzelak 2015b.

³²Siehe Kamzelak 1999, S. 125.

³³Vgl. Ries 2010.

Zunächst jedoch ist zu wünschen, dass digitale Infrastrukturen, vor allem Editionen, viele wissenschaftlich seriöse Schnittstellen schaffen und sich dieser mit etwas mehr Mut bedienen. Dazu bedarf es eines seit vielen Jahren schon geforderten Systems der akademischen Kreditierung für Webinhalte aller Art, auch kollaborativ erstellter. Und etwas mehr Mut, Daten freizugeben, sie in anderen Zusammenhängen zuzulassen.

Auch muss die Technik sich weiterentwickeln. Ein wirkliches semantisches Web wird es erst geben können, wenn Schnittstellen nicht mehr nur abgefragt werden, wenn sie bekannt sind (andocken), sondern wenn sie auf Fragen antworten, die einfach in den digitalen Raum hineingegeben werden, also auch unaufgefordert (senden). Der RDF-Beacon, der Sparql-Endpoint muss sich von der Schnittstelle zum Sender entwickeln, der Daten in einem universell verstehbaren Format aussendet, ähnlich der Technik beim Mobilfunk. Der Aufbereitung der eintreffenden Daten, der Priorisierung in der Visualisierung kommt dann eine große Bedeutung zu. Big data wird erst dann wirklich zu Big data und die Forschungsinfrastrukturen wachsen zu einem großen semantischen Netz zusammen.

Bibliographische Angaben

Forschungsinfrastrukturen für die Geistes- und Sozialwissenschaften. Hg. vom Bundesministerium für Bildung und Forschung (BMBF). Bonn 2013 PDF. [\[online\]](#) [\[Nachweis im OPAC\]](#)

Moritz Baßler / Rainer Karczewski: Computergestützte Literaturwissenschaft als Kulturwissenschaft. Eine Wunschliste. [\[online\]](#) In: Jahrbuch für Computerphilologie 9 (2009), S. 27–34. [\[online\]](#) [\[Nachweis im OPAC\]](#)

Lou Burnard: Interoperability of TEI Projects: Apotheosis or Chimera? In: Foxglove. A British Perspective on the Digital Humanities in France. Blogbeitrag vom 20.05.2013. [\[online\]](#)

Empfehlungen zu Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften. Hg. vom Wissenschaftsrat (WR). Berlin 2011. PDF. [\[online\]](#)

Kessler, der Osten und die Literatur. Mit dem Erstdruck des Dramenfragments Ivan Kaliáieff von Harry Graf Kessler. Hg. von Roland S. Kamzelak. Münster 2015a. [\[Nachweis im OPAC\]](#)

Roland S. Kamzelak: Was verrät Harry Graf Kesslers Itinerar über sein Europäertum? Analyse und Visualisierung eines statistischen Befundes. In: Grenzenlose Moderne. Die Begegnung der Kulturen im Tagebuch von Harry Graf Kessler. Hg. von Roland S. Kamzelak / Alexandre Kostka / Ulrich Ott / Luca Renzi. Münster 2015b, S. 123–133. [\[Nachweis im OPAC\]](#)

Roland S. Kamzelak: Hypermedia - Brauchen wir eine neue Editionswissenschaft? In: Computergestützte Text-Editon. Hg. von Roland S. Kamzelak. Tübingen 1999, S. 119–126. [\[Nachweis im OPAC\]](#)

Grenzenlose Moderne. Die Begegnung der Kulturen im Tagebuch von Harry Graf Kessler. Hg. von Roland S. Kamzelak / Alexandre Kostka / Ulrich Ott / Luca Renzi. Münster 2015. [\[Nachweis im OPAC\]](#)

Harry Graf Kessler: Das Tagebuch 1880–1937. Hg. von Roland S. Kamzelak / Ulrich Ott. 9 Bde. Stuttgart 2004–2018. (= Veröffentlichungen der Deutschen Schillergesellschaft, 50) [\[Nachweis im OPAC\]](#)

Martina Kerzel / Jens Middelbach / Thorsten Vitt: TextGrid. Virtuelle Arbeitsumgebung für die Geisteswissenschaften. PDF. [\[online\]](#) In: Künstliche Intelligenz 23 (2009), H. 4, S. 36–39. [\[online\]](#) [\[Nachweis im OPAC\]](#)

Langzeitarchivierung. In: Edlex. Das Editionslexikon. Lexikonartikel vom 11. Juni 2017. [\[online\]](#)

Bodo Plachta: Der »edierte Text: Grundpfeiler der Edition oder »Zugeständnis« an den Leser? DOI: [10.17175/sb002_002](#). In: Hg. von Roland S. Kamzelak / Timo Steyer. Wolfenbüttel 2018 (= Sonderband der Zeitschrift für digitale Geisteswissenschaften, 2). DOI: [10.17175/sb002](#)

Andrea Rapp: Das Projekt "TextGrid. Modulare Plattform für verteilte und kooperative wissenschaftliche Textdatenverarbeitung – ein Community-Grid für die Geisteswissenschaften". Chancen und Perspektiven für eine neue Wissenschaftskultur in den Geisteswissenschaften. In: Jahrbuch der historischen Forschung der Bundesrepublik Deutschland (2006), S. 61–68. Frankfurt/Main 2002. [\[Nachweis im OPAC\]](#)

Regeln für die alphabetische Katalogisierung in wissenschaftlichen Bibliotheken: RAK-WB. Hg. von der Deutschen Bibliothek. 2., überarbeitete Ausgabe, 4. Ergänzungslieferung. Frankfurt/Main 2002. [\[online\]](#) [\[Nachweis im OPAC\]](#)

Thorsten Ries: »die geräte klüger als ihre besitzer« Philologische Durchblicke hinter die Schreibszenen des Graphical User Interface. Überlegungen zur digitalen Quellenphilologie, mit einer textgenetischen Studie zu Michael Speisers 'ausfahrt st. nazaire'. In: Editio 24 (2010), S. 149–199. [\[Nachweis im OPAC\]](#)

Schwerpunktinitiative "Digitale Information" der Allianz der deutschen Wissenschaftsorganisationen. Leitbild 2013–2017. Hg. von der Allianz der deutschen Wissenschaftsorganisationen. Berlin 2012. PDF. [\[online\]](#)

Thomas Stäcker: Wie schreibt man Digital Humanities richtig? Überlegungen zum wissenschaftlichen Publizieren im digitalen Zeitalter. DOI: [10.1515/bd-2013-0005](#) In: Bibliotheksdienst 47 (2013), H. 1, S. 24–50. [\[online\]](#) [\[Nachweis im OPAC\]](#)

John Unsworth: Scholarly Primitives: what methods do humanities researchers have in common, and how might our tools reflect this? (Humanities Computing: formal methods, experimental practice, London, 13.05.2000) London 2000. [\[online\]](#)

Virtuelle Forschungsumgebung. In: Wikipedia. Die freie Enzyklopädie. Lexikonartikel vom 9. April 2016. [\[online\]](#)