

Heike Neuroth, Andrea Rapp, Sibylle Söring (Hrsg.)

TextGrid: Von der Community – für die Community

Eine Virtuelle Forschungsumgebung
für die Geisteswissenschaften

```
<surface n="44" facs="D07_048.jpg">
```

```
<zone>
```

```
<line style="margin-left:1cm"><fw type="pageNum"><hi><seg style="text-decoration  
underline" rend="underline style="semicircle"/>88</seg></hi></fw></line>
```

```
<line style="margin-left:1.1cm">hier wohl alle Sorten,</line>
```

```
<line style="margin-left:1.0cm">namentlich auch Gardefüsilire.</line>
```

```
<line style="margin-left:1.0cm">man liest <hi>Sch  
<seg style="text-decoration underline">loßfreiheit</seg>
```

```
</hi></line>
```

```
<line style="margin-left:1.0cm"><hi>Bu<seg style="text-decoration underline">  
Wenigs</seg></hi><handShift script="Latn"/> etc <handShift script="Latn"/>  
>Wenige</line>
```

```
<line style="margin-left:0.9cm">Lokalitäten sind <hi>  
<seg style="text-decoration underline">lo</seg></hi> zer-
```

```
</line>
```

```
<line style="margin-left:1.1cm">flor worden, wie dieses.</line>
```

```
<line style="margin-left:0.9cm">schloßartige Haus </line>
```

```
<handShift script="Latn"/>
```

```
<line style="margin-left:1.2cm"><hi>
```

```
<seg style="text-decoration underline"
```

```
>LeBourget</seg></hi> <handShift script="Latn"/>
```

```
<line style="margin-left:1.1cm">der von Soiff<mod type="
```

```
><del rend="overwritten"><!-- unentz --><del>add</del></del>
```



Neuroth/Rapp/Söring (Hrsg.) ·
TextGrid: Von der Community – für die Community

**Heike Neuroth, Andrea Rapp,
Sibylle Söring (Hrsg.)**

**TextGrid:
Von der Community –
für die Community**

**Eine Virtuelle Forschungsumgebung
für die Geisteswissenschaften**

vwh

Verlag Werner Hülsbusch
Fachverlag für Medientechnik und -wirtschaft

TextGrid: Von der Community – für die Community

hrsg. von Dr. Heike Neuroth, Prof. Dr. Andrea Rapp und Sibylle Söring

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet unter <http://d-nb.de> abrufbar.

Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung 4.0 International Lizenz – <http://creativecommons.org/licenses/by/4.0/deed.de>



Das Buch ist auch als freie Onlineversion über die Homepage des Universitätsverlages <http://www.univerlag.uni-goettingen.de> sowie über den Göttinger Universitätskatalog (GUK) bei der Niedersächsischen Staats- und Universitätsbibliothek Göttingen <http://www.sub.uni-goettingen.de> erreichbar. Es gelten die Lizenzbestimmungen der Onlineversion.
DOI: <http://dx.doi.org/10.3249/webdoc-3947>

Einfache Nutzungsrechte liegen beim Verlag Werner Hülsbusch.

© Verlag Werner Hülsbusch, Glückstadt, 2015

vwh Verlag Werner Hülsbusch
Fachverlag für Medientechnik und -wirtschaft

www.vwh-verlag.de

In Kooperation mit dem Universitätsverlag Göttingen.

Markenerklärung: Die in diesem Werk wiedergegebenen Gebrauchsnamen, Handelsnamen, Warenzeichen usw. können auch ohne besondere Kennzeichnung geschützte Marken sein und als solche den gesetzlichen Bestimmungen unterliegen.

Korrektur und Satz: Werner Hülsbusch

Redaktionelle Betreuung: Sibylle Söring

Umschlaggestaltung: Sibylle Söring, Max Andree unter Verwendung eines Filmstills aus dem Film „Virtuelle Forschungswelten: Neue Technologien für die Geisteswissenschaften“ (<http://bit.ly/1BU9MDK>) sowie eines Digitalisats aus dem Notizbuch D07 von Theodor Fontane, Blatt 44v/45r, Staatsbibliothek zu Berlin, Preußischer Kulturbesitz, Handschriftenabteilung. Digitalisat und Code hat das Projekt „Genetisch-kritische und kommentierte Hybrid-Edition von Theodor Fontanes Notizbüchern“ der Universität Göttingen zur Verfügung gestellt.

Druck und Bindung: SOWA Sp. z o. o., Piaseczno

Printed in Poland · ISBN: 978-3-86488-077-3

Inhaltsverzeichnis

| | |
|---------------------------------|----------|
| Grußwort | 9 |
| <i>Cornelia Quennet-Thielen</i> | |

I. Tradition und Aufbruch

| | |
|---|-----------|
| Neuland Virtuelle Forschungswelten | 13 |
| Die Rolle der Bibliotheken als stabile Infrastruktur im wissenschaftlichen Forschungsprozess der Zukunft | |
| <i>Ein Gespräch mit Elmar Mittler</i> | |

| | |
|--|-----------|
| TextGrid – Eine Geschichte | 23 |
| <i>Werner Wegstein, Andrea Rapp und Fotis Jannidis</i> | |

II. Herausforderungen und erste Schritte: Einstieg in die VRE

| | |
|---|-----------|
| Materialautopsie | 39 |
| Überlegungen zu einer notwendigen Methode bei der Herstellung von digitalen Editionen am Beispiel der Genetisch-kritischen und kommentierten Hybrid-Edition von Theodor Fontanes Notizbüchern | |
| <i>Gabriele Radecke</i> | |

| | |
|---|-----------|
| Blumenbach digitalisieren | 57 |
| Einblicke in eine Digitalisierungswerkstatt | |
| <i>Martin Liebetruth</i> | |

| | |
|---|-----------|
| Usability in Forschungsinfrastrukturen für die Geisteswissenschaften | 63 |
| Erfahrungen und Einsichten aus TextGrid III | |
| <i>Ben Kaden und Simone Rieger</i> | |

| | |
|--|-----------|
| Authentifizierung und Autorisierung im Kontext Virtueller Forschungsumgebungen in den Geisteswissenschaften | 77 |
| <i>Peter Gietz</i> | |

**Metadaten, LOD und der Mehrwert standardisierter
und vernetzter Daten** **91**

Martin de la Iglesia, Nicolas Moretto und Maximilian Brodhun

**III. Kulturelles Erbe, Vielfalt der Disziplinen:
Digitale Forschung mit TextGrid**

**Das Textdatenbank- und Wörterbuchprojekt
des Klassischen Maya** **105**

Möglichkeiten und Herausforderungen digitaler Epigraphik

Christian M. Prager

Die *Bibliothek der Neologie* (BdN) **125**

Einblicke in ein aufklärungstheologisches Editionsprojekt

Bastian Lemitz

TextGrid lernt Noten **139**

Digitale Musikeditionen in der VRE

Joachim Veit

Bilder als Quelle in TextGrid **153**

Robert Casties und Dirk Wintergrün

Datenvisualisierung **165**

Geo-Browser und DigiVoy

Thomas Kollatz und Stefan Schmunk

**IV. *Store it, share it, use it*: Forschungsdaten
archivieren und zugänglich machen**

Hochverfügbarkeit und Performanz:

Beiträge zum langfristigen Betrieb von TextGrid **183**

Philipp Wieder

Vom Konzept zur Umsetzung –

Einblicke in die Entstehung des TextGrid Repository **191**

Stefan E. Funk und Wolfgang Pempe

| | |
|--|------------|
| Share It – Kollaboratives Arbeiten in TextGrid | 201 |
| <i>Andreas Aschenbrenner</i> | |
| Publizieren mit TextGrid | 211 |
| SADE und XML-Print | |
| <i>Ubbo Veentjer und Marc Wilhelm Küster</i> | |
| Ein virtuelles Bücherregal: Die <i>Digitale Bibliothek</i> im TextGrid Repository | 229 |
| <i>Katrin Betz</i> | |
| V. Von der Community für die Community: Potenziale und Perspektiven | |
| Das vielschichtige und komplexe Thema Nachhaltigkeit | 241 |
| <i>Ein Interview mit Heike Neuroth</i> | |
| Das TextGrid Laboratory: Zehn Jahre Software-Entwicklung | 251 |
| <i>Mathias Göbel</i> | |
| Grenzgänge | 259 |
| An der Schnittstelle von Geisteswissenschaft, Informatik und Informationswissenschaft | |
| <i>Sibylle Söring und Mirjam Blümm</i> | |
| TextGrid in Zahlen | 271 |
| NutzerInnen, MitarbeiterInnen, Finanzen | |
| <i>Nadja Grupe und Martina Kerzel</i> | |
| Der Weg in die TextGrid-Community | 287 |
| Schulungen, Nutzertreffen, Support | |
| <i>Oliver Schmid, Philipp Vanscheidt, Celia Krause, Andrea Rapp und Simon Rettelbach</i> | |
| Virtuelle Forschungsumgebungen nachhaltig organisieren – ein Erfahrungsbericht | 301 |
| <i>Andreas Witt, Ludwig Eichinger und Norman Fiedler</i> | |

Anhang

| | |
|---|------------|
| Verzeichnis der Autorinnen und Autoren | 317 |
| Projektpartner | 321 |
| Kooperationsprojekte | 323 |
| Meilensteine und Reports | 325 |
| Nutzertreffen und Summits 2011–2015 | 335 |
| Publikationen von und über TextGrid | 337 |
| TextGrid im Netz | 343 |

Editorische Notiz:

Soweit personenbezogene Bezeichnungen in diesem Band im Maskulinum Verwendung finden, wird diese Form aus Gründen der Lesefreundlichkeit verallgemeinernd verwendet und bezieht sich ausdrücklich auf beide Geschlechter.

Grußwort

des Bundesministeriums für Bildung und Forschung (BMBF)



Texte, Bilder und Töne zu digitalisieren bedeutet, sie weltweit verfügbar zu machen, neue Formen der Analyse wie der Zusammenarbeit zu ermöglichen und die Basis für Antworten auf neue Fragen zu schaffen. Die Entwicklung vollzieht sich in rasantem Tempo.

Sie bietet gerade auch für die Geisteswissenschaften neue Chancen. Der Forschungsverbund TextGrid ist führend darin, sie zu erkennen und zu nutzen. Seit 2005 verfolgt er das Ziel, eine virtuelle Forschungsumgebung für

Geistes- und Kulturwissenschaftler zu schaffen, stetig zu verbessern und den Bedürfnissen eines wachsenden Benutzerkreises anzupassen. Dabei verbindet TextGrid ein fachwissenschaftliches Archiv, in dem digitale Forschungsdaten archiviert werden können, mit einer Software, in der verschiedene Dienste und Werkzeuge für die Zusammenarbeit in Forschergruppen genutzt werden können. Dies geschieht bereits bei einer Reihe von Forschungsprojekten:

- Im Vorhaben „Johann Friedrich Blumenbach – online“ der Akademie der Wissenschaften zu Göttingen wird die Sammlung naturhistorischer Objekte von Johann Friedrich Blumenbach (1752–1840) komplett digital rekonstruiert, verbunden mit einer digitalen Neuausgabe seiner Originaltexte, deren Übersetzungen und Folgeauflagen sowie die Erschließung seiner Korrespondenz. TextGrid verknüpft Schriften und Sammlungstücke und ermöglicht so, das Sammeln und Arbeiten eines Begründers der Zoologie und der Anthropologie im Zusammenhang zu erschließen.
- Das Projekt „Freischütz digital“ unter der Federführung des Musikwissenschaftlichen Seminars der Universität Paderborn codiert sowohl die Musik wie den Text. Es ermöglicht eine grafische und akustische Aufbereitung des „Freischütz“ von Carl Maria von Weber, einer der meistgespielten Opern der Romantik. Dabei wird zum ersten Mal der Musik-Codierungsstandard MEI für ein umfangreiches musikalisches Werk angewendet. Mit beispielhaften Untersuchungen zu verschiedenen

überlieferten Fassungen des „Freischütz“ befördert das Projekt auch die Diskussion über künftige Editionsformen.

- Die Datenbank epidat (Salomon Ludwig Steinheim-Institut, Essen) dient der Inventarisierung, Dokumentation, Edition und Präsentation von Inschriften jüdischer Gräber. In digitaler Edition online zugänglich sind gegenwärtig 145 digitale Editionen mit 28.963 Grabinschriften (62.615 Bilddateien). Die Benutzung von TextGrid ermöglicht es dabei mehreren Personen gleichzeitig, zeit- und ortsunabhängig an einem Objekt bzw. einer historischen Quelle zu arbeiten.

Diese Beispiele belegen eindrucksvoll, dass TextGrid Geisteswissenschaftlerinnen und Geisteswissenschaftler konkret dabei unterstützt, neue Forschungsansätze kooperativ umzusetzen.

TextGrid ist ein Vorreiter für die digitale Vernetzung der Geisteswissenschaften und gehört zu einer Reihe von Projekten und Forschungsinfrastrukturkomponenten der „E-Humanities“, die das Bundesministerium für Bildung und Forschung bisher mit über 60 Mio. € fördert.

Die zehnjährige Erfahrung von TextGrid ist außerordentlich hilfreich für die vielen in den letzten Jahren begonnenen Aktivitäten im Bereich virtueller Forschungsumgebungen über die Geisteswissenschaften hinaus.

Ich gratuliere TextGrid zum zehnjährigen Bestehen sehr herzlich und wünsche weiterhin viel Erfolg dabei, die Digitalisierung in den Geistes- und Kulturwissenschaften voranzutreiben.



Cornelia Quennet-Thielen

Staatssekretärin im Bundesministerium für Bildung und Forschung

I. Tradition und Aufbruch

Neuland Virtuelle Forschungswelten

Die Rolle der Bibliotheken als stabile Infrastruktur im wissenschaftlichen Forschungsprozess der Zukunft

Ein Gespräch mit Prof. Dr. Elmar Mittler

Herr Mittler, die Forschungslandschaft in den Geistes- und Kulturwissenschaften auf nationaler und internationaler Ebene hat sich im Zuge der technologischen Entwicklungen der letzten Jahrzehnte stark verändert. Wie sah die Forschungslandschaft in diesen Disziplinen aus, als Sie 1990 die SUB Göttingen als Direktor übernommen haben?



Elmar Mittler: Damals war man noch ganz analog ausgerichtet. Größter Wunsch der Geisteswissenschaftler war es, für ihre neuen Forschungsansätze unmittelbaren Zugang zur Literatur zu bekommen. Das habe ich mit der Öffnung der Historischen Bestände in Göttingen für ein gutes Jahrzehnt realisieren können. Ich habe aber schon damals die Vision gehabt, dass Bibliotheken und geisteswissenschaftliche Forschung im Rahmen der Digitalisierung

in einem bis dahin nicht vorhandenen Maße zusammenwachsen könnten – knapp gesagt nach dem Motto „Bibliothek wird Forschung, Forschung wird Bibliothek“. Dafür habe ich mich insbesondere im Rahmen meiner Mitgliedschaft im (damaligen) Bibliotheksausschuss der Deutschen Forschungsgemeinschaft eingesetzt. Diese Ansätze zu realisieren, habe ich auch als meine Aufgabe als Leiter der SUB Göttingen gesehen, die schon im 18. Jahrhundert die erste Forschungsbibliothek der Welt war – ein ideales Forschungsinstrument ihrer Zeit, wie Bernhard Fabian sie charakterisiert hat.

Welche Stimmung herrschte zur damaligen Zeit in der geisteswissenschaftlichen Forschungs- und Bibliothekslandschaft? Wie wurde diese Vision aufgenommen? Mussten Sie Widerstände überwinden?

Elmar Mittler: Die Forschung war zu dieser Zeit ganz klar am physischen Objekt und dem Zugriff darauf interessiert; sie hat die Möglichkeiten, die die Digitalisierung bietet, nur am Rande wahrgenommen. Gerade bei den ersten Projekten, die sich damit beschäftigten, wie die Digitalisierung in Deutschland aussehen könnte, konnte man erkennen, dass es Ängste vor diesem Prozess gab. Man hatte Sorge, dass die Digitalisierung nicht auf breiter Basis vorgenommen würde, sondern neue Barrieren aufgerichtet würden – dadurch, dass Bücher nur zu einem ganz kleinen Teil digital zugänglich gemacht werden können und diese Literatur hierdurch einen neuen Kanoncharakter und Ausschließlichkeitscharakter erhalten würde. Trotz – und vielleicht auch wegen – dieser Vorstellungen war es damals notwendig, sehr begrenzte Digitalisierungsprojekte bei der DFG zu beantragen. Wir haben uns beispielsweise sehr viel Mühe gegeben, dass die Bestände der Göttinger Wissenschaftsgeschichte digital erschlossen werden konnten. Wir brauchten ein hochkarätig zusammengesetztes Gremium, um diese Digitalisierung durchzusetzen. Es wäre sicher effizienter gewesen, damals den urheberrechtsfreien Bestand der Universitätsbibliothek Göttingen vollständig zu digitalisieren und so eine digitale Referenzbibliothek der wissenschaftlich relevanten Literatur des 18. Jahrhunderts zu schaffen.

Mit Ihrem Einsatz für die Digitalisierung haben Sie Neuland betreten. Wie kann man die Rolle der Bibliotheken im wissenschaftlichen Forschungsprozess in jenen Jahren beschreiben?

Elmar Mittler: Um die Jahrtausendwende herum lag der Akzent der Bibliotheken nicht mehr auf der Bewahrung des physischen Materials, wie es mit dem klassischen Bibliotheksverständnis verbunden wird. Der Akzent lag eher darauf, das Material zugänglich zu machen – was gerade auch für die Wissenschaftler ein zentraler Aspekt war und ist. Gerade auch für die historisch arbeitenden Wissenschaften spielt der Kontext, in dem die Bücher stehen, nicht nur der „content“, eine wesentliche Rolle – ein Grund, die historischen Bestände frei zugänglich zu machen, damit der Forscher nicht nur das Buch findet, das er sucht, sondern auch das, was sachlich dazu gehört und vielleicht neben dem von ihm gesuchten Titel steht. Kontext mit elektronischen Mitteln zu bilden, war auch das Ziel unserer ersten – aus heutiger Sicht ganz

kleinen – Projekte zum semantischen Erschließen in den 90er-Jahren, mit denen wir schon damals auf europäischer Ebene versucht haben, konkrete Anwendungen mit einzelnen Fachbereichen zu entwickeln. Heute sehe ich eine der wichtigsten Aufgaben der Bibliotheken darin, dass sie die Chancen, die *Linked Open Data* bieten, erkennen und realisieren. Für die Forschung ist diese semantische Erschließung durch Bibliotheken so wichtig, weil diese nicht nur die Informationen über Dokumente, sondern auch diese selbst bereitstellen können. Man kommt nur über die Bibliotheken an viele Materialien überhaupt erst heran. In der Vergangenheit bestand eine der Hauptaufgaben von Bibliotheken in dem, was man „Dekommodifizieren“ nennt: „Kommodifizieren“ bedeutet, dass man Güter in den Vermarktungsprozess bringt. Während die öffentlich geförderte Forschung selbst außerhalb der kommerziellen Verwertung steht, werden ihre Ergebnisse durch Verlage in den wirtschaftlichen Warenkreislauf eingebracht. Die Bibliotheken sind dafür da, diese Produkte wieder zu dekommodifizieren und sie so für neue Forschung frei zugänglich zu machen. Diese Aufgabe bleibt auch – oft in modifizierter Form – im digitalen Zeitalter bestehen.

Welche Rolle spielt in diesem Zusammenhang die langfristige Zugänglichkeit von Forschungsdaten, z. B. via Open Access? Inwiefern sehen Sie hier die wissenschaftlichen Bibliotheken in der Pflicht?

Elmar Mittler: Das ist ein wichtiger Punkt, an dem das moderne Bibliothekswesen ansetzen muss. Das heißt auf der einen Seite: *Open Access*, so viel man irgendwie bekommen kann – was bedeutet, nicht nur dafür zu sorgen, dass es *Open Access*-Repositorien bei den Bibliotheken gibt, sondern auch, dass durch neue Fonds kommerzielle *Open Access*-Publikationen finanziert werden können. Der nächste Schritt ist, das Material, das man zugänglich gemacht hat, dauerhaft verfügbar zu halten. Die Erfüllung dieser Aufgabe ist derzeit besonders gefährdet. Dies sieht man am Beispiel des Übergangs von Sondersammelgebieten zu Fachinformationsdiensten, der auf der einen Seite viele positive Effekte hat, aber auf der anderen Seite das systematische Dekommodifizieren und die dauerhafte Bereitstellung in einem möglichst umfassenden Rahmen durch Bibliotheken in Deutschland nicht mehr als Ziel hat. Dieses Netz des Wissens, das durch die Ergänzung mit langfristig gespeicherten Forschungsdaten in einem nie dagewesenen Umfang erweitert werden kann, darf nicht reißen, sondern muss fester und dauerhaft geknüpft werden. Die Bibliotheken sind die Einzigen, die in der Lage

sind, eine derartig komplexe Aufgabe überhaupt anzupacken und nachhaltig zu realisieren, da sie es gewohnt sind, Standards einzuhalten, und weil sie schon jetzt durch ihre internationale Verflechtung de facto ein weltweites Netz bilden.

Nun ist es ja so, dass neben der Bereitschaft der wissenschaftlichen Bibliotheken, diese Rolle anzunehmen und auszufüllen, die Wissenschaftler die angebotenen Dienste – Virtuelle Forschungsumgebungen, Forschungsdatenmanagement, Digitales Publizieren – auch annehmen müssen. Schauen wir uns die Virtuelle Forschungsumgebung TextGrid an, deren Etablierung Sie ja auch von Anfang an befördert haben – entstanden aus der Idee, Geisteswissenschaftlern a) im digitalen Medium b) kooperatives Arbeiten zu ermöglichen; beides damals noch völliges Neuland. Wie wurde TextGrid zu Beginn aufgenommen? Muss man von einem „Clash of Cultures“ sprechen – hier die tradierte, im weitesten Sinne der „Gutenberg-Galaxis“ verpflichtete geisteswissenschaftliche Methodik, dort die „Agenten der Transformation“, die die neuen Möglichkeiten, die das digitale Forschen und Publizieren eröffneten, nutzen und mit gestalten wollten?

Elmar Mittler: Vielleicht können wir hier erst mal einen Schritt zurückgehen. Bereits das Digitalisieren an sich wurde sehr negativ aufgenommen. Der damalige Präsident der DFG, Prof. Frühwald, und ich sind öffentlich (sogar in der FAZ) scharf für das Vorantreiben der Digitalisierung kritisiert worden – bis hin zu Vorwürfen, wir wollten das Buch töten. Das kann man sich heute Gott sei Dank nicht mehr vorstellen.

Sie sprechen jetzt von vor 15, 20 Jahren.

Elmar Mittler: Ja, das wird so 1996/1997 gewesen sein.

Die Geschichte des Buchdrucks, der die geisteswissenschaftliche Forschungsmethodik und -publikation essenziell geprägt hat, blickt auf rund 500 Jahre Praxis zurück. Kaum zwei Dezennien sind da keine lange Zeit für einen so fundamentalen Transformationsprozess, wie Sie ihn beschreiben.

Elmar Mittler: Nein, überhaupt nicht. Bei der Beantwortung der Frage, ob die Wissenschaftler dazu bereit sind oder sein sollten, Angebote der Bibliotheken – wie Forschungsdatenmanagement, Digitales Publizieren und Virtuelle Forschungsumgebungen – anzunehmen, muss man auch die aktuelle Forschungsförderungskultur im Hinterkopf haben: Denn die Struktur der

Finanzierung von wissenschaftlichen Projekten ist tendenziell eher ein Gegen dauerhafter Services. Viele nur für kurze Fristen bewilligte Projekte fangen immer wieder am Nullpunkt an und verschwinden kurz nach Ende der Projektfinanzierung. Und das ist ein zusätzlicher Grund, warum wir E-Humanities und Infrastruktur-Angebote wie TextGrid brauchen. Bei den Förderungseinrichtungen müsste es selbstverständlich werden, dass es mindestens als Pluspunkt, wenn nicht sogar als eine Bewilligungsvoraussetzung gesehen wird, dass ein Projekt an bereits vorhandene und ausgereifte Tools anschließt. Ein Vorbild hierfür ist JISC¹ in England.

Welche Herausforderungen gibt es hier über die Förderstrukturen hinaus? Wissen wir, was die Forscher sich wünschen?

Elmar Mittler: Ein weiteres Problem ist, dass die bisherigen Angebote oft für die Geisteswissenschaftler viel zu komplex und zu wenig nutzerorientiert sind. Will man erfolgreiche wissenschaftliche Infrastrukturen aufbauen, ist es vor allem wichtig, die „Faulheit“ der Nutzer zu unterstützen. Ein Wissenschaftler will arbeiten, ein Wissenschaftler muss Ergebnisse bringen. Er hat keine Zeit, sich mit großem Aufwand in ein Tool einzuarbeiten. Bei der Nutzung von Tools gibt es Hürden, die teilweise technisch notwendig sind, teilweise aber auch dem Entstehungsprozess und den zu geringen Mitteln geschuldet sind, die zur Verfügung gestellt werden; das ist auch ein Grund, weshalb viele Tools gegenüber den Services kommerzieller Anbieter nicht konkurrenzfähig sind. Erst wenn es uns gelingt, einem Wissenschaftler deutlich zu machen, dass er durch die Nutzung unserer Services schneller, effizienter und dauerhafter arbeiten kann, dann haben wir eine Chance, etwas zu erreichen. Das ist natürlich unter den Rahmenbedingungen Kürze der Laufzeiten und Unterfinanzierung, unter denen die meisten Projekte laufen, keine einfache Sache.

Noch einmal zur Förderlandschaft. In Deutschland etwa wird derzeit vergleichsweise viel Geld für die Digital Humanities ausgegeben. Ist das nicht doch ein Hinweis darauf, dass auch die Förderer hier einen Bedarf erkannt haben? Die DFG beispielsweise empfiehlt inzwischen, sich für wissenschaftliche digitale Editionen Partner ins Boot zu holen, die bereits Erfahrungen auf dem Gebiet der texttechnologischen Erschließung, der Standardisierung von Metadaten, der Informationswissenschaft und auch der Langzeitarchi-

¹ <http://www.jisc.ac.uk>

vierung besitzen, sodass sie – ganz wie Sie es fordern – die entsprechenden Kompetenzen und Expertisen nicht selbst und jedes Mal aufs Neue entwickeln müssen.

Elmar Mittler: So etwas gibt es erfreulicherweise für Teilbereiche, aber wenn man sich die Breite der Landschaft anschaut, ist das noch lange nicht flächendeckend durchgesetzt. Eine wichtige Frage hierbei ist auch, wie man entwickelten Services Dauerhaftigkeit geben kann. Die Bibliotheken sind die am besten geeigneten Einrichtungen für solche Aktivitäten – wenn man ihren Wandlungsprozess unterstützt und es ihnen ermöglicht, neue Services dauerhaft anzubieten. Das ist allerdings nur der Fall, wenn die Universitäten akzeptieren, dass die Personalstellen den Bibliotheken beim Ausscheiden von Mitarbeitern erhalten bleiben. Die Gefahr der Stellenkürzung ist hier sehr groß, da die Bibliotheken dem Anschein nach nicht mehr so viele Aufgaben haben. Das ist nur sehr eingeschränkt der Fall; im Gegenteil bringen die digitalen Medien viele Zusatzbelastungen. Und trotzdem besteht die Chance, durch Rationalisierungsmaßnahmen bei den traditionellen Aufgaben Freiraum für den Aufbau neuer Services zu schaffen. Ich habe das systematisch betrieben: Mit Drittmittelprojekten habe ich junge Menschen gewonnen, die neue Services entwickelt haben. Die besten Kräfte habe ich dann auf Dauerstellen gebracht. Mit einer derartigen Strategie könnten auch Grundlagen dafür geschaffen werden, Großprojekte wie TextGrid und DARIAH-DE mit Dauerhaftigkeit zu versehen. Wo es gelingt, dafür einen Konsens mit der Universitätsleitung herzustellen, sehe ich die Chance, dass die Bibliothek als dauerhafte Informationseinrichtung eine wachsende Rolle spielen kann.

Welchen Gewinn hat eine Universität als Dachstruktur davon?

Elmar Mittler: Der Universität bietet sich damit eine Chance, auf Dauer zusätzliche überregionale Bedeutung und internationales Renommee zu gewinnen. Man muss sich ja auch darüber im Klaren sein, dass Förderer zwar bereit sind, für eine gewisse Zeit über Projekte Infrastrukturen anzustoßen. Aber schon das ist – wie die Erfahrung zeigt – von personellen und politischen Konstellationen abhängig, die schnell wechseln können. In jedem Fall ist die Bereitschaft, eine kontinuierliche Unterstützung zu geben und völlig neue Infrastrukturen zu schaffen, sehr beschränkt. Auch das spricht dafür, bewährte Infrastruktureinrichtungen wie die Bibliotheken innovativ weiterzuentwickeln und auszubauen.

Kommen wir nach der Forschung nun zur Lehre. Sie stellen TextGrid seit acht Jahren in Ihren buchwissenschaftlichen Seminaren vor, an denen Studierende ganz unterschiedlicher Fachrichtungen teilnehmen. Wie nehmen diese jungen Nachwuchs-Geisteswissenschaftler, die zur Generation der Digital Natives gehören, das Angebot einer Virtuellen Forschungsumgebung für ihre eigene Arbeit auf?

Elmar Mittler: Was die Studierenden betrifft: Man muss sich darüber im Klaren sein, dass die sogenannten *Digital Natives* auch jetzt noch zu guten Teilen alles andere als begeisterte Anhänger digitalen Forschens und Publizierens sind. Viele sind noch weit davon entfernt, digitale Methoden und Tools in ihren Alltag zu integrieren. Allerdings bemerke ich, dass beispielsweise unter den Historikern die Werkzeuge und Angebote aus dem Bereich der Digital Humanities inzwischen sehr viel besser bekannt sind, als das noch vor wenigen Jahren der Fall war. Wenn ich an die Plattform hypotheses.org² denke, die moderne Kommunikations- und Publikationskanäle wie das Bloggen für die Wissenschaft zu öffnen sucht, stelle ich fest, dass es noch vor etwa drei Jahren massive Widerstände gab und der Einsatz dieser Kommunikationsform in der Forschung für Unsinn gehalten wurde. Inzwischen ist das Bloggen auch in der Wissenschaft für viele eine Selbstverständlichkeit geworden. Die Studenten haben teilweise schon konkrete Erfahrungen mit einem Seminarblog. Man kann also feststellen, dass sich die neuen Methoden subkutan durchsetzen.

Wie gestalten Sie solche Seminare?

Elmar Mittler: In meinen Veranstaltungen versuche ich, durch Besuche in anderen, auf diesem Gebiet aktiven wissenschaftlichen Bibliotheken wie der HAB Wolfenbüttel oder durch das Einladen von Kolleginnen und Kollegen aus Göttinger Digital-Humanities-Projekten ins Seminar Studierende dauerhaft an die Digitalen Geisteswissenschaften heranzuführen. Oft gehen sie mit Begeisterung für die Digital Humanities aus der Veranstaltung und nehmen diese für ihr Studium und vielleicht auch danach wichtiger als vorher. Besonders glücklich bin ich, dass die Studenten durch praktisches, konkretes Arbeiten mit TextGrid, das durch XML und TEI ja durchaus eine gewisse Einstiegshürde hat, Spaß an der Sache bekommen und wir in Göttingen nun an einer Edition des ‚Bellifortis‘ von Konrad Kyesser arbeiten. Man-

² <http://hypotheses.org>

che Studenten beteiligen sich sogar ohne zusätzlichen Zertifikats-Erwerb weiter an der Edition mit TextGrid. Wir haben für sie eine Art „TextGrid light“ entwickelt, um das Arbeiten mit der Umgebung einzuüben. Hier zeigt sich, wie wichtig es ist, den zukünftigen Wissenschaftlern einen niedrigschwelligen Zugang zu digitalen Tools und Methoden zu bieten.

Sie sprechen das Stichwort Usability an – ein essenzielles, mitunter noch zu wenig beachtetes Thema in den Digital Humanities.

Elmar Mittler: Richtig. Es ist wichtig, die Leute dort abzuholen, wo sie gerade stehen. Auf der anderen Seite habe ich immer wieder festgestellt, dass Wissenschaftler von sich aus zu uns gekommen sind und bei uns Hilfe gesucht haben, wenn sie ein Interesse hatten, einem E-Humanities- oder E-Science-Projekt Dauerhaftigkeit zu geben. Dieses Vertrauen in die Bibliotheken auszubauen, ist für die Zukunft zentral. Wenn wir den Wissenschaftlern einen Service anbieten können, der für sie bequemer ist, als wenn sie vieles selber machen müssten, der einen effizienteren Einsatz von Hilfskräften sichert, der ihnen die Arbeit für sich allein oder mit anderen erleichtert und der ihnen zudem garantiert, dass ihre Ergebnisse leichter publiziert werden können und dauerhaft zur Verfügung stehen, dann ist ein Durchbruch für die Anwendung von E-Humanities in der Forschung möglich. Ein derartiges „Rundum-sorglos-Paket“ aber können dauerhaft nur die wissenschaftlichen Bibliotheken (selbstverständlich auch nur in Kooperation mit den Rechenzentren) anbieten.

Last, but not least: Sind die Digital Humanities in fünf oder zehn Jahren noch ein Thema? Werden uns die heutigen Diskussionen unverständlich sein, oder spricht man 2025 nur noch von den ‚Humanities‘, weil das ‚digital‘ für uns selbstverständlich geworden sein wird?

Elmar Mittler: Ich denke, Letzteres wird ganz sicher der Fall sein. Die Entwicklung geht aber nicht von allein weiter. Man muss dafür sorgen, dass die Wissenschaft mit systematisch ausgebildetem Personal unterstützt wird. Wir brauchen die Digital Humanities als universitäres Fach; sie müssen aber sehr praxisorientiert sein, so wie es auch die Bibliothekswissenschaft sein muss. Vielleicht werden die Digital Humanities „nur“ eine Hilfswissenschaft; sie müssen auf jeden Fall serviceorientiert sein. Wir brauchen die E-Humanities-Lehrstühle und -Institutionen, damit systematisch, kontinuierlich und mit dem Ehrgeiz, an der Spitze der Entwicklung zu stehen, neue Methoden

und Tools für die Geistes- und Kulturwissenschaften entwickelt werden, die dann von erfolgreichen Studierenden in die Bibliotheken, Archive und Museen, vor allem aber auch in die Forschung, eingebracht werden. Ich bin optimistisch, dass wir dafür in zehn Jahren eine etablierte Digital-Humanities-Forschungs- und -Infrastrukturlandschaft schaffen können.

Das Gespräch führten Sibylle Söring und Michelle Rodzis.

TextGrid – Eine Geschichte

von Werner Wegstein, Andrea Rapp und Fotis Jannidis

Vorbemerkung

In guter philologischer Tradition gliedern wir ‚TextGrid – Eine Geschichte‘ in ‚Die Vorgeschichte‘, ‚Das Projekt und seine Phasen‘ und den visionären Ausblick ‚Perspektiven‘. Mit der Dreiteilung ergibt sich die Verantwortung für die verschiedenen Teile von selbst. Da einige der frühen Erfahrungen ihren Ursprung an der Universität Würzburg haben, fällt dieser Teil auf mich, das ‚Projekt und seine Phasen‘ auf Andrea Rapp, weil sie schon an den Anfängen in Göttingen beteiligt war und anschließend alle Projektphasen von TextGrid mitgestaltet hat. Fotis Jannidis ist für den visionären Ausblick prädestiniert, denn er hat im Jahr 2007, auf der TEI-Jahrestagung *TEI@20: 20 Years of Supporting the Digital Humanities* in College Park, Maryland, U.S., schon einmal in die Glaskugel geblickt, Visionäres gesehen und in Worte gefasst, die man nicht so schnell vergisst, z. B. die verlassenen XML-Bäume in den TEI-Kodierungswäldern.¹ Eine Vision, wie es mit TextGrid nach zehn Jahren als Virtueller Forschungsumgebung weitergeht, wird ihm daher leichtfallen.

Werner Wegstein

*Ein Mann hat eine Erfahrung gemacht,
jetzt sucht er die Geschichte seiner Erfahrung ...*

Max Frisch²

Die Vorgeschichte

Der Anstoß geht von Prof. Dr. Wolfgang Riedel aus, seinerzeit Dekan der damaligen Philosophischen Fakultät II der Universität Würzburg. Im April

¹ Keynote presentations 1.3 *TEI in a crystal ball* (<http://www.tei-c.org/Vault/Members-Meetings/2007/program.html#jannidis>)

² Das aufgegebene Thema erinnert mich an Gedanken des Erzählers im Anfangsbereich des Romans ‚Mein Name sei Gantenbein‘ von Max Frisch (1964), S. 8 und S. 11 – daher mein Motto.

2003 lädt er für einen Workshop am 11. Juli 2003 nach Würzburg ein zum Thema: ‚Zukunft von TUSTEP und Perspektiven philologischer EDV‘.

Anlass ist die ungewisse Zukunft von TUSTEP. Wird es (in Tübingen oder woanders) weiterentwickelt? Wird es auf dem bis dato erreichten Entwicklungsstand ‚eingefroren‘ und in der gegebenen Form (und wie lange dann) weiterverwendet? Gibt es Alternativen zu TUSTEP? Die Würzburger Germanistik, die seit langem und intensiv mit TUSTEP arbeitet, betrachtet diese Ungewissheiten mit einiger Sorge. Es erscheint uns ratsam, die einschlägig kompetenten und engagierten Kollegen aus dem Fach einmal zu einem Meinungsaustausch und Brainstorming zusammenzurufen.³

Es folgt eine Liste von Fragen, über die diskutiert werden sollte, sowie die Namen bereits Eingeladener und ein Dank vorab für weitere Einladungsvorschläge.

Der Workshop am 11. Juli 2003 ist ein Erfolg. 33 Wissenschaftler (v. a. aus Würzburg, Tübingen, Trier) und der Schweiz (Zürich) nehmen teil, und diskutieren die skizzierten Fragen ausführlich und weitestgehend einmütig. Als Ergebnis ist festzuhalten:

1. Die Verdienste von TUSTEP sind unstrittig: die stolzen Bilanzen der Würzburger Forschergruppen *Prosa des deutschen Mittelalters* und *Das Bild des Krieges im Wandel vom späten Mittelalter zur frühen Neuzeit* sowie des Sonderforschungsbereichs 226 *Wissensorganisierende und wissensvermittelnde Literatur im Mittelalter* wären ohne TUSTEP nicht möglich gewesen. Die Erfahrungen daraus sind von Anfang an auch in die Lehre eingeflossen und haben die beiden Aufbaustudiengänge in Würzburg, ‚Linguistische Informations- und Textverarbeitung‘ und ‚EDV-Philologie‘ beeinflusst. Für die Organisation der Weiterentwicklung in Form einer Evolution wird ein Netzwerk empfohlen. Die leitende Frage bleibt: Wie wird philologische Software weiterentwickelt?
2. Die Diskussion am 11.07.03 zeigt auch, dass bislang nirgends erwogen worden ist, organisatorisch und technisch eine Lösung zu suchen. Deshalb erfährt der Vorschlag, an der Universität Würzburg ein Kompetenzzentrum ‚EDV-Philologie‘ einzurichten, Zustimmung von allen TeilnehmerInnen des Workshops. In Würzburg sind sowohl das Know-how in Forschung und Lehre als auch das Bedürfnis nach der Weiterentwicklung philologischer Software vorhanden.

³ Der Abdruck des Briefftextes erfolgt mit Zustimmung des Verfassers.

Als Konsequenz des Arbeitstreffens beantragt das Institut für deutsche Philologie im Juli 2003 die Einrichtung eines Kompetenzzentrums für EDV-Philologie am Institut für deutsche Philologie in Verbindung mit einer entsprechenden Professur.

An dem Wochenende um den 7./8. November 2003 kann das TEI Members Meeting⁴ in dem nicht allzu weit entfernten Nancy genutzt werden, seit 1990 bestehende Kontakte zur TEI-Community zu vertiefen und in Nancy weitere Ansichten zu den in Würzburg diskutierten Fragen einzuholen.

Zum 1. Dezember 2003 richtet die Universität Würzburg das ‚Kompetenzzentrum für EDV-Philologie‘ am Institut für deutsche Philologie ein und beauftragt mich mit der Vertretung einer Professur für EDV-Philologie.

Zum 19. März 2004 lädt nunmehr das Kompetenzzentrum für EDV-Philologie zum nächsten <philtag>-Arbeitstreffen ein. Nach dem Erfolg des Workshops im Juli 2003 bekommt dieses Arbeitstreffen damit die Nummer 2. (Die Nomenklatur schwankt eine Weile je nach Sprache der Veranstalter und Teilnehmer zwischen Workshop und Arbeitstreffen.) Es melden sich 20 Teilnehmende an.⁵ Zwölf TeilnehmerInnen registrieren sich für die geplante Projektgruppe <philtag>.

Das Programm besteht aus sechs Vortragsblöcken und einer Diskussionsrunde. Eingebunden sind Informatik-Studierende aus Darmstadt (Jannidis) und Würzburg (Seipel), sowie Informatikstudierende aus Würzburg mit Nebenfach Linguistik (Wegstein).

4 <http://www.tei-c.org/Membership/Meetings/2003/mm22.xml>

5 Neben den elf Referenten und ihren Betreuern sind dies: Hans-Werner Bartz, Deutsches Wörterbuch Projekt, Univ. Trier; Dr. Thomas Burch, Kompetenzzentrum Elektronisches Publizieren, Univ. Trier; Conrad Drescher, Informatik, Univ. Würzburg; Dipl.-Inf. Marbod Hopfner, Informatik, Univ. Tübingen; Dr. Stephan Moser, Kompetenzzentrum EDV-Philologie, Univ. Würzburg; Martin Omasreiter, Informatik, Univ. Würzburg; Dr. Klaus Prätor, Berlin-Brandenburgische Akademie der Wissenschaften, Berlin; Dr. Andrea Rapp, SUB Göttingen; Dr. Peter Stahl, Kompetenzzentrum EDV-Philologie, Univ. Würzburg.

Die Vorträge:

1. Michael Stolz (Basel): Computergestütztes Kollationieren. Ein Werkstattbericht aus dem Basler Parzival-Projekt
2. Dietmar Seipel / Michael Wenger (Informatik, Würzburg): Textvergleich
3. Fotis Jannidis/Alexander Dotor/Thomas Ries (TU Darmstadt): Automatisches Tagging von Strukturen (Inhaltsverzeichnisse, Register, ...)
4. Werner Wegstein / Thomas Hemmerich / Florian Thienel (Universität Würzburg):
 - a. Archivierung von TUSTEP-Dateien im XML-Format (Hemmerich)
 - b. Editor's workbench: Kollationierungsarbeitsplatz mit XMetal und Grafik (Thienel)
5. Hans-Walter Gabler (Universität München): Werkzeuge für die Edition von Virginia Woolfs Roman ‚To the Lighthouse‘
6. Michael Nedo (Wittgenstein-Archiv, Cambridge): EDV-Werkzeuge für die Wittgenstein-Edition
7. Diskussionsrunde über Procedere und Präzisierung der Rahmenbedingungen
 1. Java +/- Eclipse
 2. Votum für Plattformunabhängigkeit und Open Source. Diskussion über die Nachteile von Eclipse; angeregt wird eine Untersuchung der Abhängigkeit und die Suche nach Alternativen zu Eclipse.
 3. XML-Dateiformat
 4. Kodierungsminimalia
 5. Einbindung anderer Module
 6. Integration weiterer Programmier- und Skriptsprachen (Prolog soll geprüft werden.)
 7. Test-Rahmen
 8. Dokumentationsstandards
 9. Stand der Arbeit in Würzburg

Zum Abschluss wird festgehalten, dass Termine im März aus verschiedensten Gründen ungünstig sind, und es wird angeregt, die Arbeitstreffen in den Herbst zu verlegen.

7.10.2004: TextGrid-Besprechung in Würzburg Zimmer 4E13:

Wenn ich mich recht entsinne, haben wir in Würzburg damals über die Zusammensetzung der Antragsgruppe diskutiert und versucht, eine gute Balance zu finden. Die Informatik im Kernbereich ist mit DAASI International GmbH (z. B. zur Anbindung an die Middleware) und mit der Saphor GmbH, die noch etwas später dazukommt, vertreten. Aber auch die FachwissenschaftlerInnen kommen für bestimmte Aufgaben nicht ohne InformatikerInnen in ihrem Teilprojekt aus. Die richtigen dafür zu finden, wird nicht einfach. Die Fachwissenschaft (Germanistik) selbst ist in guter Breite vertreten; die Mediävistik mit Andrea Rapp, die Neuere Deutsche Literaturwissenschaft mit Fotis Jannidis, die Sprachwissenschaft mit jemandem wie mir, der auch ein verkappter Mediävist sein könnte – wie viele Sprachwissenschaftler mit einem Schwerpunkt in der Sprachgeschichte. Rückblickend zeigt sich aber: Keiner von uns war von Dauer und ist heute noch dort, wo er war, als wir begannen. Unserem gemeinsam gestarteten Projekt hat das aus meiner Sicht nicht geschadet. Immerhin ist die SUB immer noch in Göttingen und das IDS in Mannheim, *stabilitas loci* – fast schon etwas Benediktinisches in all dem Wandel, aber für Langzeitprojekte aus meiner Sicht eine wichtige Grundlage für den administrativen Kern.

Außerdem bringen unsere Institutionen, die auf Dauer angelegt sind, Spezielles ein: die SUB natürlich die Metadaten, Katalogstrukturen und alles, was mit Buch im weitesten Sinn zu tun hat, und das IDS, die Institution, die auf Sprache (vor allem, aber nicht nur auf Gegenwartssprache) spezialisiert ist, bringt z. B. auch die Korpuslinguistik, die Lexikografie und weitere interessante Bereiche mit ein. Außerdem wurden 1971 und 1973 am IDS die ersten beiden Tagungen mit dem Titel ‚Maschinelle Verarbeitung altdeutscher Texte‘ veranstaltet, in Deutschland der Einstieg in die Textanalyse von Sprache mithilfe der Datenverarbeitung (vgl. Lenders/Moser 1978a; 1978b).

Das Einzige, dessen ich mir in meiner Erinnerung ganz sicher bin, ist die Tatsache, dass ich nach dieser Besprechung Heike Neuroth und Joachim Migl vom Hubland-Campus der Universität in Würzburg unter Nutzung aller Schleichwege und Abkürzungen so schnell (und ohne Strafzettel) an den Würzburger Bahnhof bringen konnte, dass sie den Zug nach Göttingen noch erreicht haben, obwohl wir bis zur letzten Minute zu diskutieren hatten.

Die Verlegung der <philtag>-workshops in den Herbst wird noch 2004 vollzogen. Als Frucht der TEI-Tagung in Nancy können Lou Burnard und Laurent Romary als Referenten gewonnen werden.

8./9.10.2004: philtag 3 in Würzburg TEI Workshop, 8.–9. Oktober 2004 mit Lou Burnard und Laurent Romary:⁶

Alle Details sind auf der Website festgehalten: die Workshop-Ziele, eine Leseliste einschlägiger Literatur, mit der man sich zuvor beschäftigen sollte, das Programm und die TeilnehmerInnen. Interessant erscheint im Rückblick die dichte Kreuzung von Aktivitäten: Donnerstag trifft sich die Würzburger philtag-Initiative nach dem Mittagessen und vor Lou Burnards Vortrag über XAIRA im Toscana-Saal der Würzburger Residenz und am Samstag wird der Spätnachmittag genutzt für ein Treffen des Technical Committee of the German Working Group on Philosophical Editions. Im Rundbrief Nr. 20 & newsletter 12 der Arbeitsgemeinschaft philosophischer Editionen der Deutschen Gesellschaft für Philosophie e.V. vom Dezember 2004, S. 3–4, findet sich dazu ein ausführlicher Bericht von Klaus Prätor.

Am 21.10.2004 wurde der Antrag beim BMBF eingereicht. Im Sommer 2005 waren noch Rückfragen von Gutachtern zu beantworten.

14./15.10.2005: philtag 4 in Würzburg mit Lou Burnard und Sebastian Rahtz (Oxford University):⁷

Alle Details sind auf der Website festgehalten: die Workshop-Ziele, eine Leseliste einschlägiger Literatur, mit der man sich zuvor beschäftigen sollte und das Programm. Die TeilnehmerInnenliste ist leider (schon in Würzburg) verloren gegangen. Einer der Gründe dafür mag der technische GAU gewesen sein, der in Würzburg am 15.10.2005 zuschlug. Im Laufe des Vormittags brachen im Abstand von 30 Minuten nacheinander beide Beamer, die zur Verfügung standen, zusammen. Die Glühbirnen versagten. Zusätzliche Reservebirnen hatten wir damals nicht.

25.10.2005: Das Bundesministerium für Bildung und Forschung bewilligt unseren Antrag. Freude macht sich breit. Jetzt kann es losgehen.

Werner Wegstein

6 <http://www.tei-c.org/Talks/2004/Wuerzburg/programme.html>

7 <http://projects.oucs.ox.ac.uk/teiweb/Talks/2005/Wuerzburg/>

Das Projekt und seine Phasen

„Die Geschichte, plur. ut nom. sing. von dem Zeitworte geschehen.

- 1) Was geschehen ist, eine geschehene Sache, so wohl in weiterer Bedeutung, eine jede, so wohl thätige als leidentliche Veränderung, welche einem Dinge widerfähret, als auch in engerer und gewöhnlicherer, von verschiedenen mit einander verbundenen Veränderungen, welche zusammen genommen ein gewisses Ganze[s!] ausmachen (...). Eine wahre Geschichte, im Gegensatze der erdichteten. In engerer Bedeutung führet nur die erstere den Nahmen der Geschichte. (...)
- 2) Die Erzählung solcher Geschichte oder geschehenen Begebenheiten; die Historie. (...)⁸

So wird ‚Geschichte‘ in Adelungs Wörterbuch erläutert, das hier nach der Version im *Trierer Wörterbuchnetz* zitiert ist, die auf den vom Bundesministerium für Bildung und Forschung gekauften und seitdem frei verfügbaren Daten der *Digitalen Bibliothek* von www.zeno.org basiert. Sowohl das Wörterbuchnetz als auch die *Digitale Bibliothek* sind Teil der mehr als zehn Jahre umfassenden TextGrid-Geschichte, die hier erzählt wird. Zehn Jahre bedeuten nicht nur in der digitalen Welt eine Zeit intensiven Wandels; dies spiegelt der vorliegende Band sehr deutlich. Daher ist es nicht die Aufgabe unseres Beitrags, die TextGrid-Geschichte so zu erzählen, dass ein „gewisses Ganzes“ entsteht; dies leisten vielmehr neben den zahlreichen TextGrid-Publikationen und -Berichten vor allem die TextGrid-Produkte und die Erfahrungen der EntwicklerInnen NutzerInnen. Unser Blick auf die Geschichte soll vielmehr ein ganz persönlicher und auf die „nicht herausgegebenen“ Rahmenbedingungen, Meilensteine, Schlüsselmomente und Veränderungen sein, die die Geschichte von TextGrid ausmachen.

8 Johann Christoph Adelung: *Grammatisch-kritisches Wörterbuch der Hochdeutschen Mundart mit beständiger Vergleichung der übrigen Mundarten, besonders aber der oberdeutschen*. Zweyte, vermehrte und verbesserte Ausgabe. Leipzig 1793–1801, s. v. ‚Geschichte‘, Bd. 2, Sp. 605 f. http://woerterbuchnetz.de/cgi-bin/WBNetz/wb-gui_py?sigle=Adelung&lemid=DG01581; Imagedigitalisat: <http://www.zeno.org/Adelung-1793/K/adelung-1793-02-0605>; <http://www.zeno.org/Adelung-1793/K/adelung-1793-02-0606>.

Das TextGrid-Konsortium bildete eine außergewöhnliche, aber im Rückblick sehr gut funktionierende Mischung: Die beteiligten Philologinnen mit unterschiedlicher Schwerpunktsetzung (s. o.) vertrauten sich in der Leitung des Projekts einer erfahrenen Bibliothek an, hinzu kamen zwei Unternehmen mit technologischer, aber auch philologischer Expertise. In den folgenden Phasen wurde dieser Kreis erweitert durch weitere fachwissenschaftliche Ausdifferenzierung wie Klassische Philologie, Judaistik oder Musikwissenschaften sowie durch Rechenzentren als zentrale Infrastrukturpartner.⁹ Obwohl die Initiativgruppe vordergründig einen heterogenen Eindruck machte, gab es für einen großen Teil der Projektpartner einen gemeinsamen Hintergrund, sie hatten bereits eine gemeinsame Geschichte – nicht zuletzt im TUSTEP-Kontext, in der sie Textwissenschaften mit dem Einsatz und der Entwicklung digitaler Technologie verbunden hatten. Einer der Schlüsselmomente des Projektes war sicherlich die erste Antragsbesprechung des Konsortiums in Raum 4E13 in Würzburg am 7.10.2004 (s. o.), bei der die Kerngruppe – deren genaue Zusammensetzung bis dahin nicht allen bekannt war – zusammentraf und überrascht feststellte, dass alle sich in unterschiedlichen Konstellationen kannten und bereits in irgendeiner Form zusammengearbeitet hatten. Offenheit und ein aus Erfahrungen resultierendes Grundverständnis für die jeweils anderen Forschungskulturen waren also vorhanden.

Dennoch erforderte die Zusammenarbeit im Projekt eine Phase, in der eine gemeinsame, die Disziplinen überbrückende Sprache und eine spezifische gemeinsame Projektkultur entwickelt werden mussten. Auf dieser Basis gelang es, Kontinuität mit Flexibilität und Entwicklungsfähigkeit zu verbinden und über den für Projekte sehr langen Zeitraum von zehn Jahren vertrauensvoll zusammenzuarbeiten. Überhaupt scheint mir diese spezifische Projektkultur – nicht nur Aspekte aus den Geisteswissenschaften und der Informatik zusammenfügend, sondern ein neues Drittes bildend – durchaus kennzeichnend für die Digital Humanities zu sein und damit etwas Verbindendes zu stiften in diesem Disziplinbündel, das – kaum definiert – sich bereits entfaltet und ausdifferenziert. Auch aus diesem Blickwinkel scheint TextGrid ein Digital-Humanities-Modellprojekt zu sein, von dem mehr bleiben wird als eine Software-Entwicklung, da es ein wesentlicher Motor der Community-Entwicklung und Kulturbildung ist.

⁹ Vgl. dazu auch die Beiträge von Joachim Veit, Philipp Wieder und anderen in diesem Band.

Auch die Förderer im Bundesministerium für Bildung und Forschung schenkten der Gruppe Vertrauen und ließen sich auf das Experiment ein, den Geisteswissenschaften Zugang zu avanciertester Technologie zu ermöglichen. Die Bereitschaft zu referatsübergreifender Kommunikation und Zusammenarbeit kann wohl ebenfalls kaum hoch genug eingeschätzt werden.

Technologische Rahmenbedingungen wurden zunächst von der durch die Weitsicht Elmar Mittlers¹⁰ ermöglichten Einbindung in den D-Grid-Verbund gesetzt, was einerseits Folgen für technologische Grundlagen und Architekturentscheidungen, andererseits aber auch für die Kommunikation und für die Vermittlung der TextGrid-Idee in die geisteswissenschaftlichen Communitys hatte. Sie wurde häufig als „zu naturwissenschaftlich“ empfunden und war in den Geisteswissenschaften nicht immer einfach zu erläutern. Mittlerweile sind jedoch viele Aspekte der Grid-Philosophie in allen Wissenschaftszweigen angekommen: digitale Forschungsinfrastrukturen, netzbasiertes und kollaboratives Arbeiten, Datenoffenheit, Nachnutzbarkeit usw. Zugleich konnten die Mitglieder des TextGrid-Konsortiums vom Austausch mit den anderen natur- und ingenieurwissenschaftlichen Grid-Communitys profitieren, wie überhaupt der disziplinübergreifende Blick auf den Technologiebedarf bereichernd für alle Beteiligten war.¹¹ Ein weiterer Meilenstein für die TextGrid-Initiatoren war daher sicherlich die Zwischenbegutachtung des D-Grid-Verbunds am 1. und 2. März 2007 in Berlin im Konrad-Zuse-Zentrum, bei der wir den Eindruck gewonnen hatten, „angekommen zu sein“ – und das in doppelter Perspektive: Die geisteswissenschaftlichen Angebote, Fragestellungen und Bedarfe stießen nicht nur auf Verständnis und Interesse bei den übrigen Communitys, sondern wurden als Bereicherung erkannt, während wir unsererseits die naturwissenschaftlichen Infrastruktur-Perspektiven immer besser integrieren konnten.

Bereits in den ersten vorsichtigen Projektideen und Konzepten spielten frei verfügbare, vor allem qualitätvolle digitale Inhalte eine wesentliche Rolle. Qualität definierte sich nach den Kriterien von inhaltlicher Qualität (aktuelle und wissenschaftlich akzeptierte kritische Referenzangaben), Erfassungsqualität (Erfassungsgenauigkeit, Fehlerrate, Vollständigkeit, Zitierfähigkeit), Datenformat und Aufbereitungsqualität (Standards, XML/TEI, Metadaten) sowie Archivierbarkeit. Allerdings standen diese Dinge in der

10 Siehe dazu auch sein Interview in diesem Band.

11 Diese Aspekte wurden systematisch weiterverfolgt und dokumentiert im Projekt *WissGrid – Grid für die Wissenschaft*; s. www.wissgrid.de (26.01.2015).

ersten Projektphase nicht im Fokus, da diese Phase sich auf die Entwicklung einer Workbench für textwissenschaftliche (insbesondere editionswissenschaftliche) Arbeits- und Forschungsprozesse konzentrierte. Gleich zu Beginn wurden jedoch die Wörterbücher des Trierer Wörterbuchnetzes eingebunden.¹² Die Wörterbuchdaten wurden zu eng mit anderen Werkzeugen verknüpften Services (vgl. Büdenbender/Leuk 2009).

Die Inhalte rückten während der Projektlaufzeit immer stärker ins Blickfeld und es wurde deutlich, dass der zuverlässige und langfristige Zugang dazu ein entscheidendes Kriterium für die Weiterentwicklung und die Akzeptanz der Digital Humanities sein würde. Die Community artikulierte ganz dezidiert den entsprechenden Bedarf nach dem TextGrid Repository. Ein besonderer Glücksfall und wiederum wichtiger Meilenstein war daher der Kauf der Daten der *Digitalen Bibliothek*, die seitdem in Würzburg weiter aufbereitet werden und für Forschung und Lehre frei zur Verfügung stehen.¹³ Erst wenn genügend standardisierte und offene Daten, die Geisteswissenschaftlerinnen und Geisteswissenschaftler für ihre Forschung benötigen, zur Weiterverarbeitung und Nachnutzung zur Verfügung stehen, werden Digital-Humanities-Methoden in der Mitte der Fächer ankommen. Gerade auch in der Lehre ist ein solcher freier Zugang zu verlässlichen Daten, mit denen man Studierenden auch Qualitätsaspekte nahe bringen kann, von immenser Bedeutung.

Als glücklichen Umstand sehe ich schließlich, dass die Entwicklung von TextGrid parallel mit der Entwicklung und der Etablierung der Digital Humanities erfolgte und dass es uns ermöglicht wurde, TextGrid immer stärker als Forschungsprojekt zu profilieren, wenn auch der Infrastrukturaspekt nie aus den Augen geriet. Auf diese Weise kann – so hoffen die Initiatoren – der Anspruch verfolgt werden, als Forschungsinfrastruktur „wesentliche Beiträge zum wissenschaftlichen Erkenntnisgewinn“ zu liefern, zu „Inkubatoren für neue und innovative wissenschaftliche Fragestellungen“ zu werden und „ganz neuartige Möglichkeiten der forschenden Erschließung von Bibliotheks-, Archiv- und Sammlungsbeständen“ zu bieten sowie „eine Kultur des internationalen und zunehmend auch interdisziplinären Austauschs“ zu fördern.¹⁴

Andrea Rapp

12 www.woerterbuchnetz.de (26.01.2015).

13 Siehe hierzu den Beitrag von Katrin Betz im vorliegenden Band.

14 Empfehlungen des Wissenschaftsrates zu Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften, 28.1.2011, Drs. 10465-11, S. 7: <http://www.wissenschaftsrat.de/download/archiv/10465-11.pdf>.

Die Zukunft TextGrids

Das Projekt TextGrid wird 2015 nach neunjähriger Förderung abgeschlossen sein. Wie steht es um seine Zukunft? Auf diese Frage ist keine einfache Antwort möglich. Anders als bei vielen anderen Projekten haben sich die Verantwortlichen engagiert, ihr Projekt über dessen Förderlaufzeit hinaus weiter zu betreiben. Dazu haben sie sich in einem Verein zusammengeschlossen, dessen vornehmliche Aufgabe die „persistente Sicherstellung des fachwissenschaftlich nachhaltigen Gebrauchs der Angebote der TextGrid-VFU in einer heterogenen Community wissenschaftlicher Nutzerinnen und Nutzer“¹⁵ ist. Und gerade Institutionen wie das Institut für Deutsche Sprache und die Staats- und Universitätsbibliothek Göttingen, die plausibel Nachhaltigkeit versprechen können, sind dabei in führender Position engagiert. Zugleich ist TextGrid ein Modul in dem europäischen Infrastruktur-Projekt DARIAH-EU, das zurzeit als DARIAH-DE in Deutschland wiederum mit Förderung des BMBF umgesetzt wird.¹⁶

Das klingt alles sehr gut, aber es wird meines Erachtens nichts an der Tatsache ändern, dass TextGrid in einigen, wahrscheinlich wenigen Jahren verschwunden sein wird. Und das ist gut so. Lassen Sie mich erläutern, wie ich das meine. Schon seit einiger Zeit ist es klar geworden, dass DH-Institutionen wie TEI oder eben auch TextGrid aus mehreren Komponenten bestehen: zum einen aus der Software, und hier kann man im Fall von TextGrid noch zwischen dem Repository, in dem die Daten liegen und mit einem Persistent Identifier versehen sind, der API und den Clients unterscheiden. Zum anderen handelt es sich um eine Community von Geisteswissenschaftlern, Vertretern von Bibliotheken und Rechenzentren sowie Informatikern. Und schließlich handelt es sich um Standards und Best Practices, die im Laufe des Projekts entwickelt oder adaptiert wurden. Wenn ich nun sage, dass TextGrid in einigen Jahren verschwunden sein wird, dann bezieht sich das vor allem auf die Community und die Clients. Die Community wird sich, gerade weil TextGrid Teil einer umfassenderen Infrastruktur wird, ändern und daran adaptieren. Für die Clients muss ich ein wenig weiter ausholen.

Zu den wichtigsten Lehren, die zumindest ich aus meinen Erfahrungen mit der Entwicklung von TextGrid ziehen kann, gehört die Einsicht, dass

15 <http://www.textgrid-verein.de/>

16 DARIAH-DE (<https://de.dariah.eu/>) ist ein Teil der gesamteuropäischen Initiative (<https://www.dariah.eu/>).

digitale Infrastrukturen einem besonderen Entwicklungsdruck unterliegen – und das gilt natürlich umso mehr, wenn die angezielte Nutzergruppe selbst sich in einem so grundlegenden Wandel befindet, wie das in den Digitalen Geisteswissenschaften der Fall ist. Das Design einer Infrastruktur muss sich unter diesen Umständen auf die Extrapolation von Entwicklungstendenzen in eine doch weitgehend ungewisse Zukunft stützen. Und nicht nur das Verhalten der Anwender unterliegt Wandlungsprozessen, auch die Techniken zur Lösung von bestimmten Problemen ändern sich. In den Jahren, in denen wir an TextGrid gearbeitet haben, hat sich etwa der Browser als Arbeitsplattform für Anwendungen im Bereich des PC etabliert – und dann ist auch diese Technik wieder unwichtiger geworden, da im Rahmen der mobilen Geräte eine neue Kultur stark spezialisierter Programme Mode wurde.

Ein Grundproblem von Software im Bereich der Digitalen Geisteswissenschaften scheint gleich zu bleiben: Wie kann man die weniger technikaffinen Anwender komplexe Arbeiten ausführen und sich dabei ganz auf ihre fachwissenschaftliche Fragestellung konzentrieren lassen, ohne sie in ein Korsett zu zwingen, das der jeweiligen Fragestellung erst mühsam angepasst werden muss? Doch mögliche Antworten auf diese Frage sind von den größeren Paradigmen im Umgang mit Computern in all ihren Formen abhängig. Anders gesagt: Eine gelungene Lösung zu finden angesichts von sich ändernden Anwendererwartungen und -kompetenzen, von Techniken und Standardlösungen, die stets im Fluss sind, kann nur das Ergebnis eines längeren Forschungsprozesses sein, der nie endgültig abgeschlossen sein wird. Die Rolle der Fachwissenschaftler wird sich hierbei wohl vor allem darauf beschränken, die wissenschaftlichen Anforderungen zu formulieren, die Prototypen auf ihre Funktionalität zu testen und die Institutionen, die Dauer versprechen, dabei zu unterstützen, in die richtige Richtung weiterzuentwickeln.

Diese Teile von TextGrid werden sich also auflösen, werden teilweise eingehen in andere Strukturen, werden teilweise obsolet werden und werden evtl. zum Teil Grundlage für eigenständige neue Entwicklungen sein. Dass bei all der hierbei notwendigen Flexibilität und Beweglichkeit einige Dinge doch sicher und verlässlich bleiben – nicht zuletzt die Daten der Nutzer, die publizierten Daten und ihre URLs, kurzum das Repository –, dafür stehen die Betreiber von TextGrid.

Es ist sehr zu hoffen, dass TextGrid dabei auch zum Anlass wird, die Förder- und Unterhaltsstruktur in Deutschland zu überdenken. Zurzeit sehen sich alle einschlägigen Fördereinrichtungen in Deutschland nicht in der Lage, dauerhafte Infrastrukturen nach dem Auslaufen der Projektförderung zu fi-

nanzieren. Das Problem ist allen Beteiligten bekannt, aber die politischen und finanziellen Herausforderungen, die damit verbunden sind, erlauben bislang keine einfache, klare Lösung. Die Zeit, in der Fachwissenschaftler versuchen konnten, ihre Bedürfnisse im Rahmen von Projekten durch den Aufbau von ‚Infrastrukturprojekten‘ zu befriedigen, scheint auf jeden Fall vorbei zu sein. Die Notwendigkeit, solche Projekte nachhaltig zu gestalten, erfordert die Kooperation mit Institutionen, die – durchaus mit sehr guten Gründen – ganz anders arbeiten, als es der ungeduldige Fachwissenschaftler tut. Für diesen Prozess, der einerseits die Einsicht des Wissenschaftlers braucht, welche Werkzeuge fehlen, und der andererseits diesen übergreifende Stabilität abverlangt, gibt es einige Modelle, z. B. die Open-Source-Entwicklung oder die Webservice-Architekturen, aber hier ist der richtige Weg, Kreativität und stabile Organisation zu verbinden, wohl noch nicht gefunden.

Aber natürlich ist TextGrid nicht nur ein Forschungsprojekt in Sachen Virtuelle Forschungsinfrastruktur in den Geisteswissenschaften – TextGrid ist auch Infrastruktur, deren Nutzer in diesem Kontext das Arbeiten mit digitalen Werkzeugen lernen und weiterentwickeln. Ihre Anforderungen, die sich aus der Auseinandersetzung mit TextGrid und den Forschungsplänen für ihr Material ergeben, werden ihre Umgangsweise und Ansprüche an alle folgenden Werkzeuge prägen. Auch in dieser Form wird TextGrid also weiterleben.

Fotis Jannidis

Literaturverzeichnis

- Büdenbender, Stefan; Leuk, Michael (2009): Daten als Dienste: Wörterbücher als Erschließungsinstrumente in der virtuellen Arbeitsumgebung „TextGrid“. In: *it – Information Technology* (Themenheft „Informatik in den Geisteswissenschaften“) 51 (4): 191–196.
- Lenders, Wilfried; Moser, Hugo (Hrsg.) (1978 a): *Maschinelle Verarbeitung altdeutscher Texte. 1 Beiträge zum Symposium Mannheim, 11./12. Juni 1971*. Berlin: E. Schmidt.
- Lenders, Wilfried; Moser, Hugo (Hrsg.) (1978 b): *Maschinelle Verarbeitung altdeutscher Texte. 2 Beiträge zum Symposium Mannheim, 15./16. Juni 1973*. Berlin: E. Schmidt.

II.

**Herausforderungen und erste
Schritte: Einstieg in die VRE**

Materialautopsie

Überlegungen zu einer notwendigen Methode bei der Herstellung von digitalen Editionen am Beispiel der Genetisch-kritischen und kommentierten Hybrid-Edition von Theodor Fontanes Notizbüchern

von *Gabriele Radecke*

Einleitung

Virtuelle Forschungsumgebungen wie beispielsweise TextGrid sind wichtige Infrastrukturen für die Erarbeitung komplexer digitaler Editionen. Zum einen ermöglichen sie ein kollaboratives und ortsunabhängiges Forschen, zum anderen unterstützen sie durch ihre Dienste und Werkzeuge die digitale Aufbereitung des Ausgangsmaterials maßgeblich.¹ Die Arbeit an digitalen Editionen beginnt jedoch nicht erst in einer Virtuellen Forschungsumgebung, sondern setzt schon mit der Konzeption und „der Vorbereitung und Zusammenstellung des Forschungsgegenstandes“² ein. Die Materialautopsie – d. h. die systematische Sichtung, Beschreibung und Analyse von Handschriften, Typoskripten und Drucken – bildet dabei eine unabdingbare Grundlage; ihre Ergebnisse sind keine bloßen Vorarbeiten, sondern entscheidende Voraussetzungen für die mikrostrukturellen Editionsrichtlinien³ sowie für alle weiteren Arbeitsbereiche und Methoden in analogen und digitalen Editionsprozessen. Dieser Zusammenhang wurde bisher nur am Rande diskutiert,⁴ sodass es angebracht erscheint, sich dem Komplex der Materialautopsie sowohl in editionstheoretischer als auch -praktischer Hinsicht einmal anzunähern. Da-

1 Vgl. exemplarisch Neuroth et al. (2011).

2 Burch (2014), Abschnitt „Metadatenerfassung“

3 Im Unterschied zum makrostrukturellen Editionsprinzip, das das allgemeine Editions-konzept meint, sind mit den mikrostrukturellen Editionsrichtlinien die einzelnen editorischen Regeln, etwa zur Transkription, zur Textkonstitution und zur Kommentierung, gemeint.

4 Vgl. erstmals Radecke et al. (2013: 100).

bei spielen diejenigen Erfahrungen eine Rolle, die während der Arbeit an der Genetisch-kritischen und kommentierten Edition der Notizbücher Theodor Fontanes gemacht worden sind.⁵

Begriffsbildung und editionstheoretische Voraussetzungen

Im Rahmen des Fontane-Notizbuch-Projekts wurde ein projektspezifischer Workflow entwickelt, der die drei wesentlichen Abschnitte der interdisziplinären Zusammenarbeit umfasst: die Bereitstellung und Erschließung des Materials, die Aufbereitung des Materials in der Virtuellen Forschungsumgebung TextGrid sowie die verschiedenen digitalen und analogen Nutzungsmöglichkeiten im Portal, im Repository und als Buch (vgl. Abb. 1).⁶



Abb. 1 Workflow im Fontane-Notizbuch-Projekt; Grafik erstellt vom TextGrid-Team, ergänzt und bearbeitet von Gabriele Radecke, Martin de la Iglesia und Mathias Göbel

⁵ Das von der Deutschen Forschungsgemeinschaft von 2011 bis 2017 geförderte Editionsprojekt entsteht unter meiner Gesamtleitung an der Theodor Fontane-Arbeitsstelle der Universität Göttingen in enger Zusammenarbeit mit der SUB Göttingen und TextGrid unter der informationswissenschaftlichen und -technischen Leitung von Heike Neuroth an der Abteilung Forschung & Entwicklung und an der Gruppe Metadaten und Datenkonversion. Weitere Mitarbeiter sind: Judith Michaelis und Rahel Rami (editorische Assistenz), Martin de la Iglesia (Metadaten) und Mathias Göbel (IT, Visualisierung und Portalentwicklung).

⁶ Vgl. die ausführliche Darstellung in Radecke et al. (2013: 99–102).

Aufgrund der allgemein formulierten Arbeitsschritte lässt sich der Workflow auf weitere digitale Editionsprojekte übertragen, die ebenso wie die Notizbuch-Edition die Transkriptionen und kommentierten Edierten Texte auf der Grundlage von Handschriften und historischen Drucken veröffentlichen. Mit *digitalen Editionen* sind hier *born digital* erstellte Editionen gemeint,⁷ die von Anfang an auf philologischen und digitalen Methoden basieren und dem neugermanistischen Editionsverständnis verpflichtet sind, der Editionen als kommentierte, auf der Grundlage textkritischer Methoden erarbeitete digitale und analoge Textausgaben definiert.⁸

Zum Begriff der Materialautopsie

Die editorische Arbeit beginnt mit der Methode der Materialautopsie (vgl. Abb. 1, Nummer 1). Die Bezeichnung *Materialautopsie* wurde von mir schon während der Projektvorbereitungen 2009/2010 gewählt, weil man dabei nicht die digitalisierte Repräsentation des Originals, etwa Handschriftenscans, fokussiert, sondern vielmehr das zu bearbeitende Ausgangsmaterial. Der Begriff *Autopsie*, der ursprünglich aus der Gerichtsmedizin stammt, wird aufgrund seiner zweiten Bedeutung, der „Prüfung durch persönliche

7 Zum Unterschied zwischen einer *born digital*-Edition und einer retrodigitalisierten Edition, die lediglich das „Layout und Referenzsystem der Buchausgabe“ beibehält, vgl. Burch (2014), Einleitung; vgl. auch Radecke et al. (2015; im Druck).

8 Der weit ausgelegte editionswissenschaftliche Textbegriff umfasst neben den literarisch-ästhetischen und publizierten Texten auch nicht abgeschlossene und unveröffentlichte Entwürfe, Konzepte und Niederschriften sowie Rechts- und Gebrauchstexte, musikalische Texte sowie zeichnerische Skizzen und Kunstwerke (vgl. Plachta 2011: 66). Mit dem Aufkommen der vielfältigen digitalen Repräsentationsformen von Handschriften und Drucken – der Bereitstellung von Digitalisaten, digitalen Regesten und Nachlassdatenbanken sowie der Transkriptionen und textkritisch konstituierten und kommentierten Texte – wird der Editionsverständnis inzwischen häufig auf alle digitalen Repräsentationen angewendet. Diese Beliebigkeit ist allerdings insofern problematisch, weil somit die entscheidenden Akteurinnen und Akteure einer Edition, die Editoren und ihre textkritischen und philologischen Methoden, keine Rolle mehr spielen. Vgl. etwa Jutta Weber, die in ihrem Beitrag die „Erstellung und Langzeitarchivierung der Metadaten zum Gesamtnachlass sowie für [...] die Digitalisate“ nicht ganz zutreffend als „erste Edition“ bezeichnet, die von anderen Editionen, etwa von Historisch-kritischen Ausgaben, nachgenutzt werden (vgl. Weber 2013: 33f.).

Inaugenscheinnahme“⁹, inzwischen auch innerhalb der Bibliotheks- und Editions-wissenschaft verwendet – etwa, wenn es um die bibliografische Autopsie von Büchern bei der Titelaufnahme geht oder wenn der Stellenkommentar einer Edition zur Vermeidung von Fehlern nicht auf anderen Kommentaren beruht, sondern auf Quellenautopsie fußt (vgl. Radecke/Detering 2014: 896).

Die Reflexionen über die Methoden und Arbeitsschritte im Fontane-Notizbuch-Projekt verdeutlichen nun, dass der editionswissenschaftliche Begriff der Autopsie auch für die Beschreibung und Analyse des Editionsgegenstandes geeignet ist. Eine erste Publikation zum Thema Materialautopsie, die den Begriff aus editionspraktischer Perspektive eingeführt hat, wurde 2013 vorgelegt (vgl. Radecke et al. 2013: 100); ein weiterer Aufsatz auf der Basis pragmatischer Überlegungen ist im Druck.¹⁰

Dass der Begriff der Materialautopsie inzwischen auch auf andere (Editions-) Vorhaben übertragbar ist und somit einen universalen Anspruch erheben darf, zeigt die Workflow-Grafik des Projekts *Johann Friedrich Blumenbach – online*, das an der Göttinger Akademie der Wissenschaften in Kooperation mit TextGrid an der SUB Göttingen entsteht (vgl. Kerzel et al. 2013: 125). Auch hier beginnt die wissenschaftliche Arbeit mit einer Materialautopsie, die im Unterschied zum Fontane-Projekt jedoch nicht nur (Druck-) Schriften umfasst, sondern auch Objekte von Blumenbachs naturhistorischen Sammlungen. Der Autopsie der Überlieferung – im Fontane-Projekt sind es 67 Notizbücher, die zwischen dem Ende der 1850er- und dem Ende der 1880er-Jahre benutzt wurden –¹¹ kommt nun eine konstitutive Bedeutung zu, weil ihre Ergebnisse die Editionsrichtlinien und das Metadatenschema begründen.

Umgekehrt formuliert bedeutet das, dass sowohl die Transkriptionsregeln¹², ein wesentlicher Teil der Editionsrichtlinien, als auch das Codie-

9 Duden online; www.duden.de (Stichwort „Autopsie“)

10 Vgl. Radecke et al. (2015; im Druck), Kapitel II Materialautopsie.

11 Zu weiteren Informationen zu Benutzung, Funktionen und Notizbuchinhalten vgl. Radecke (2010: 98f.) und Radecke (2013: 156–163).

12 Wie aus der Workflow-Grafik (Abb. 1, Nr. 4 und 6) hervorgeht, wird im editorischen Arbeitsprozess zwischen der Transkription der Zeichen und der Codierung und Tiefenerschließung der Transkriptionen unterschieden. Aufgrund der schwer zu entziffernden Handschrift Fontanes sowie der anspruchsvollen Codierung wurde im Laufe der Projektzeit entschieden, die Entzifferung zunächst in Word vorzunehmen und nach der Konvertierung durch OxGarage mit der Tiefencodierung fortzufahren.

rungs-, Metadaten- und Visualisierungsschema vom Resultat der Materialautopsie abhängig sind. So können bei der Transkription, Codierung und Visualisierung beispielsweise nur das Trema (ë; ï; ÿ) oder auch das Schaft-s (l) berücksichtigt werden, wenn die Autorhandschrift diese Zeichen belegt und diese auch erkannt wurden. Allerdings werden in vielen, auch digitalen Editionen diese Zeichen stillschweigend z.B. durch ein „e“ und rundes „s“ ersetzt, obwohl mit der damit einhergehenden Normalisierung – ähnlich wie bei einer modernisierten Textgestalt – ein sprachhistorischer Bedeutungsverlust entsteht. Je genauer die Durchsicht des Materials und die darauf aufbauende Beschreibung und Analyse der Befunde erfolgt, desto präziser können die Editionsrichtlinien erfasst und desto genauer kann die Transkription der Zeichen und die Tiefenerschließung durch die XML/TEI-Codierung erfolgen.¹³ Für eine exakte Beschreibung und Auswertung des Ausgangsmaterials sind editionsphilologische Fähigkeiten erforderlich, die es ermöglichen, die Relevanz der Materialbefunde für eine Edition zu erkennen, adäquat zu benennen, zu deuten und editorisch aufzubereiten. Editionsphilologische Methoden tragen schließlich entscheidend dazu bei, dass die überlieferungsbedingten Befunde in Editionen nicht verwischt, sondern selbst durch den Medienwechsel von der Handschrift zum Buch/zum Portal bewahrt und somit auch rezipiert werden können.

Die differenzierte Analyse insbesondere der Handschriften durch Materialautopsie hat jedoch nicht allein editionspragmatische und -praktische Gründe; vielmehr steht die Notwendigkeit der exakten Beschreibung und darauf basierender Analysen im Zusammenhang mit editionstheoretischen Überlegungen, die Hans Zeller schon 1971 in seinem Aufsatz „Befund und Deutung“ formulierte und die sich bis heute bewährt haben. Demnach ist es für eine wissenschaftliche Edition, die den Kriterien der Objektivität verpflichtet ist, erforderlich, zwischen dem Befund der Überlieferung und seiner editorischen Deutung zu unterscheiden (vgl. Zeller 1971: 77–86).¹⁴ Diese

13 Zum Konzept des Metadatenschemas und zur XML/TEI-Codierung vgl. Radecke et al. (2015; im Druck), Kapitel III Metadaten und TEI.

14 Aufgrund der räumlichen Beschränkung muss auf eine ausführliche Darstellung des editionswissenschaftlichen Diskurses von *Befund* und *Deutung* sowie auf Zellers kritische Auseinandersetzung mit der Objektivität von Editionen an dieser Stelle verzichtet werden. Die Trennung von Befund und Deutung, die freilich nicht immer exakt zu ziehen ist, kann auf unterschiedliche Weise geschehen: Im Editionsbericht werden alle editorischen Entscheidungen diskutiert; im textkritischen Apparat sollten diese allge-

Prämisse trifft einmal mehr zu bei Editionen wie beispielsweise der Fontane-Notizbuch-Edition, die auf einer heterogenen handschriftlichen Überlieferung aufbaut und dem editionstheoretischen Konzept der Authentizität verpflichtet ist, die handschriftlichen Befunde in der Edition möglichst originalgetreu wiederzugeben.¹⁵

Material, Medium, Inhalt

Die Notwendigkeit einer differenzierten und präzisen Materialautopsie auch für eine digitale Edition wird vor dem Hintergrund des Materialitätsdiskurses erneut bekräftigt, der spätestens mit der Tagung „Materialität in der Editionswissenschaft“¹⁶ einsetzte, die die Arbeitsgemeinschaft für germanistische Edition 2008 veranstaltet und die die Editionstheorie und -praxis inzwischen beeinflusst hat. Fontanes Notizbücher bieten sich hier ebenfalls an, über grundlegende Zusammenhänge zwischen dem Material, dem Medium und dem Inhalt nachzudenken. Es ist mittlerweile unbestritten, dass Nachlassdokumente wie Briefe und Tagebücher eine „sinnhafte Einheit von Text und Textmedium“ (Hurlebusch 1995: 25).¹⁷ bilden. So werden Umfang, Funktionen und Inhalte vom Medium Brief oder Tagebuch mit seiner jeweiligen

meinen Hinweise aufgrund von Einzelnachweisen präzisiert werden. Schließlich bietet die Wiedergabe von Faksimiles die beste Möglichkeit für das Lesepublikum, die subjektiven editorischen Entscheidungen zu überprüfen.

15 In einer früheren Publikation wurde von mir noch der Begriff *überlieferungsadäquat* verwendet, der ebenfalls den Editor/die Editorin zu einer möglichst originalgetreuen Textdarbietung verpflichtet; vgl. Radecke (2010). – Vgl. zum hier verwendeten Begriff der Authentizität Zeller (2002: 52). Auf eine kritische Diskussion des Authentizitätsbegriffs, der innerhalb der editionswissenschaftlichen Fachliteratur nach wie vor unterschiedlich definiert wird, muss hier verzichtet werden. Für die Methode der Materialautopsie ist Zellers letzte Begriffsbestimmung im Zusammenhang mit der Historisch-kritischen Edition des C. F. Meyer-Briefwechsels von 2002 sehr geeignet, weshalb sich darauf beschränkt wird.

16 Vgl. den gleichnamigen Sammelband von Schubert (2010). Inzwischen wurde eine weitere editionswissenschaftliche Konferenz zum Thema „Text, Material, Medium“ veranstaltet, die den Zusammenhang von Material und Inhalt fokussiert (vgl. Lukas et al. 2014).

17 Vgl. auch die Ergebnisse der Ausstellung und der Konferenz „Der Brief – Ereignis & Objekt“, die vom 11. September bis zum 16. November 2008 im Freien Deutschen Hochstift in Frankfurt/Main stattfand (vgl. Bohnenkamp/Wiethölter 2008).

„materialen Erscheinungsform“ (Rasch et al. 2013: 97) beeinflusst.¹⁸ Dasselbe gilt auch für Notizbücher, wie es 2013 in dem Beitrag „Notizbuch-Editionen“ erstmals auf breiter Basis diskutiert wurde, deren kompaktes Format (etwa 10 × 17 cm) beispielsweise sogenannte Unterwegszeichnungen begünstigt (vgl. Radecke 2013: 149–155.). Infolgedessen ist die Materialität einer Handschrift keine editorisch zu vernachlässigende äußerliche Rahmenbedingung, sondern gehört vielmehr zum „integralen Bestandteil“ (Rasch et al. 2013: 97) der Bedeutungserschließung von Briefen, Tage- und Notizbüchern. Der Berücksichtigung der Materialität kommt dabei nicht zuletzt deshalb eine große Bedeutung zu, weil die Kenntnis und Analyse der materialen Beschaffenheit, die man durch die Methode der Materialautopsie gewinnt, zur Erklärung der Funktionen und zur Ermittlung von Inhalten beiträgt.

Ergebnisse der Materialautopsie

Duktus

Zu einer der auffallendsten materialen Erscheinungsformen in Fontanes Notizbüchern gehört der Wechsel von der Standardschrift zur Reinschrift bzw. zu einem hastigen Duktus (vgl. Abb. 2, 3 und 4). Wie wichtig das Erkennen, Differenzieren und Benennen des sehr unterschiedlich ausgeprägten äußeren Schriftbildes für die Ermittlung und Interpretation von Textsorten ist, belegen mittlerweile mehrere Notizbuchfunde. So ist es erst aufgrund einer präzisen Analyse des Duktus gelungen, eine Vortragsmitschrift zu identifizieren, die Fontane in seinem Notizbuch C03 während seines Besuches auf der nahe bei Berlin gelegenen Pfaueninsel im Sommer 1870 festgehalten hat (vgl. Radecke 2008: 217–219; Radecke 2010: 101–103). In vielen Notizbüchern finden sich weitere Beispiele eines unruhigen Duktus, wie etwa Fontanes Aufzeichnungen in seiner Funktion als Kritiker der „Vossischen Zeitung“ während der Theateraufführungen im Königlichen Schauspielhaus am Gendarmenmarkt (vgl. Abb. 4, Blatt 8v; Notizen zu Hugo Bürgers Drama „Gold und Eisen“ vom 12. März 1882).

Ebenso verhält es sich mit den zahlreichen Notizen, die Fontane bei seinen Besuchen von Kunstausstellungen angefertigt hat, und die wie seine Bemerkungen zu den Theaterstücken im Rahmen der Notizbuch-Edition

18 Vgl. exemplarisch Radecke (2010: 95–106).

erstmal erschlossen und publiziert werden. Aufgrund einer differenzierten Materialautopsie wurden nun Fontanes umfangreiche Aufzeichnungen zu Carl Blechens Gemälden erstmals ermittelt, die im Zusammenhang mit seinen Vorbereitungen zu einer unvollendet gebliebenen Blechen-Biografie entstanden sind. Bekannt waren bisher nur wenige Notizen in Notizbuch B09, die Fontane am 23. Dezember 1881 bei einem ersten Besuch der 14. Sonderausstellung der Berliner Nationalgalerie niederschrieb, in der u. a. 370 Gemälde von Blechen gezeigt wurden.¹⁹ Hier wird einmal mehr Fontanes Verfahren deutlich, das er auch bei anderen Kunstausstellungsnotizen zwischen den 1860er- und 1880er-Jahren erprobte (vgl. etwa die Notizbücher D04, E02 und E03): Die mit unruhigem Duktus geprägten Notate entstanden vor Ort ohne feste Schreibunterlage und Fontane verzeichnete listenartig die Werknummern und Titel der Gemälde, die gelegentlich mit subjektiven Kommentaren ergänzt wurden. Am Schluss des Notizbuch-Eintrags vom Dezember 1881 hat er jedoch nur noch die Bildnummern, geordnet nach den Ausstellungsräumen und nach der Bildgröße, eingetragen (vgl. Abb. 3).

Warum hier lediglich die Nummern notiert wurden, erklärt sich, wenn man die Aufzeichnungen aus Notizbuch B08 hinzuzieht, die vermutlich während der anschließenden Besuche am 4., 13. und 17. Januar 1882 entstanden sind (vgl. Erler/Erler 1994: 146–150). Demnach hatte Fontane bereits im Dezember 1881 weitere Besichtigungen der Kunstausstellung beabsichtigt, bei denen er die zunächst offenbar nur oberflächlich betrachteten Gemälde ein zweites Mal etwas genauer anschauen und beschreiben wollte, deren Nummern er vorsorglich schon im Notizbuch B09 als Gedächtnisstütze festgehalten hat. Bei einem der nächsten Male hat er sich dann auch mit den 64 Blechen-Gemälden der Sammlung von Wilhelm Brose beschäftigt, wie es die „Brose“-Notiz auf Blatt 9r zeigt (vgl. Abb. 4). Ob sich Fontanes Aufzeichnungen zu den einzelnen Blechen-Werken auch auf Fontanes Besichtigung der Privatsammlung von Carl Ludwig Kutzt am 3. Januar 1882 und auf sein Treffen mit Broses Witwe am 10. Januar 1882 beziehen sowie auf sein Aktenstudium in der Akademie der Künste am 7. und 9. Januar 1882 (vgl. Streiter-Busch 2010: 147–149),²⁰ wird im Rahmen des Editionsprojekts ermittelt.

19 Vgl. hierzu die erste Auswertung der Aufzeichnungen in Streiter-Buscher (2010: 143 f.).

20 Vgl. die fundierte Rekonstruktion der Beschäftigung Fontanes mit Blechen in Streiter-Buscher (2010).

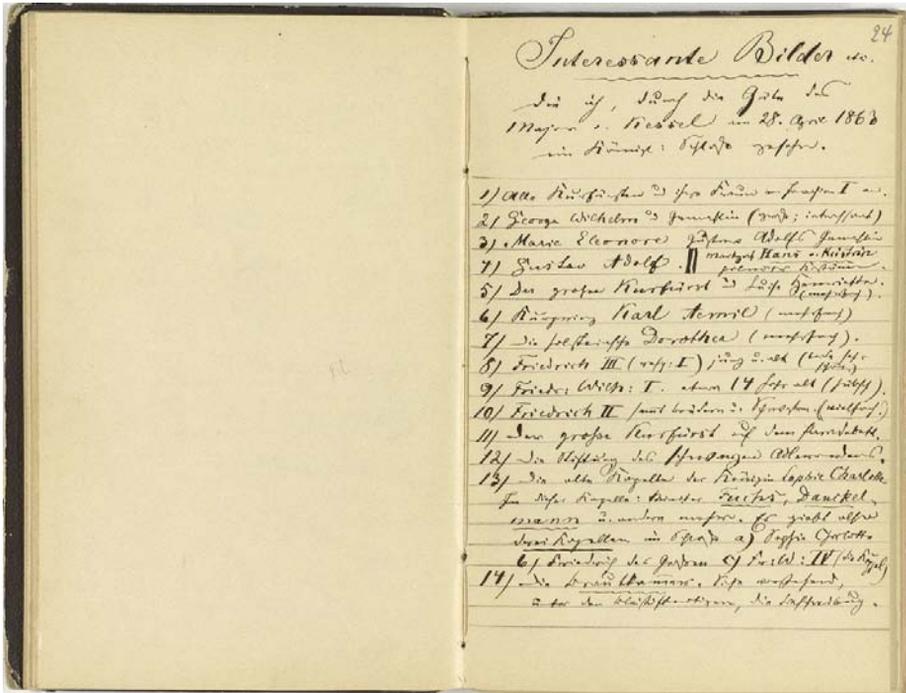


Abb. 2 Fontane, Notizbuch A07, Blatt 24r. Reinschrift in schwarzer Tinte

Jedenfalls steht jetzt schon fest, dass durch die Analyse des Duktus und der äußerlichen Anordnung der Einträge Fontanes Notizen zumindest eingegrenzt werden können. Für die exakte Verifizierung bedarf es freilich dann weiterer Quellen wie des Ausstellungskatalogs sowie Fontanes Briefe und Tagebuchaufzeichnungen. Das Notizbuch B08 ist im Zusammenhang mit Fontanes geplantem Blechen-Fragment vermutlich deshalb bisher nicht berücksichtigt worden,²¹ weil einerseits der Tagebucheintrag in der Edition nicht kommentiert (vgl. Erler/Erler 1994: 360), andererseits auf eine Materialautopsie der Notizbücher verzichtet wurde und man sich bei der Recherche lediglich auf das lücken- und fehlerhafte Inhaltsverzeichnis bezogen hat,

21 Vgl. zuletzt Streiter-Buscher (2010: 143–150). Fontanes Notizen zu den Blechen-Gemälden befinden sich in Notizbuch B08, auf den Vorderseiten von Blatt 1, 3–10, 12, 14–18 und 20–25 sowie auf der Rückseite von Blatt 5. Auf den anderen Verso-Seiten hat Fontane seine Theatermitschriften festgehalten.

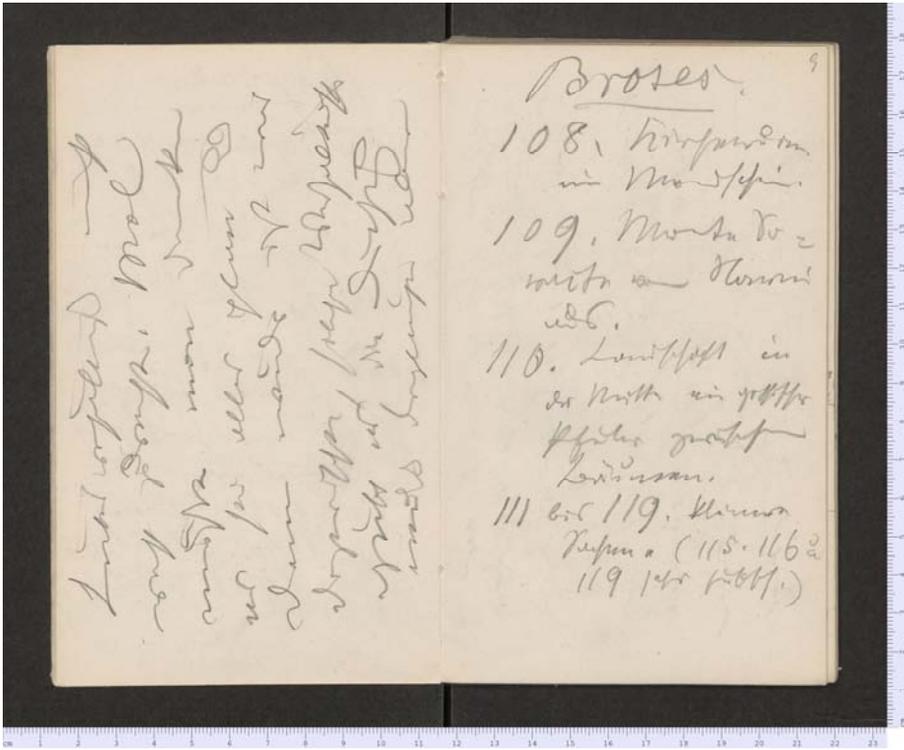


Abb. 4 Notizbuch B08, Blatt 8v/9r. Duktus *standard* („Broses“) und *hasty* in Bleistift; die Notizen entstanden im Januar 1882

Blattfragmente

Ein weiteres charakteristisches Merkmal der Notizbücher Fontanes sind die vielen Blattfragmente, die Reste von beschrifteten oder unbeschrifteten, herausgerissenen oder herausgeschnittenen Notizbuchblättern (vgl. Abb. 5). Bei der ersten Sichtung der Digitalisate waren diese Blattfragmente nicht zu erkennen; erst der Blick in die Original-Notizbücher verdeutlichte das Ausmaß der entfernten Blätter; so hat Fontane etwa allein aus Notizbuch E01 über siebzig Blatt herausgeschnitten. Außerdem wurde durch die Materialautopsie deutlich, dass die Erstdigitalisierung der Notizbücher durch das Theodor-Fontane-Archiv ohne Materialautopsie lediglich mit Fokussierung auf die Notizbuchinhalte erfolgte und infolgedessen für eine material- und me-

dienbasierte wissenschaftliche Notizbuch-Edition unbrauchbar war.²³ Hans Zeller hat bereits 1971 festgestellt, dass eine Edition dem Anspruch auf Objektivität am besten gerecht wird, wenn die Handschriften idealerweise neben den Edierten Texten mit abgedruckt werden können (vgl. Zeller 1971: 83 f.). Faksimilierte Handschriften innerhalb einer Edition stehen, wie es Klaus Hurlbusch dann 1995 formulierte, zur Ansicht von Transkription und Ediertem Text sowohl in einem „komplementären“ als auch in einem „kommentierenden Verhältnis“, denn sie bewahren den „graphisch manifestierte[n] private[n] Charakter der Aufzeichnungen“, der durch „Homogenisierung“ der Typografie verschwunden ist.²⁴ Damit den Benutzerinnen und Benutzern nun also durch die integrale Aufnahme von Handschriftenfaksimiles (in der Fontane-Notizbuch-Edition sind es Digitalisate) neben der Transkription und dem Edierten Text ein erweiterter Interpretationsspielraum gegeben werden kann, muss die Digitalisierung nach den Prinzipien der Materialität erfolgen und möglichst alle Ergebnisse der Materialautopsie bewahren.

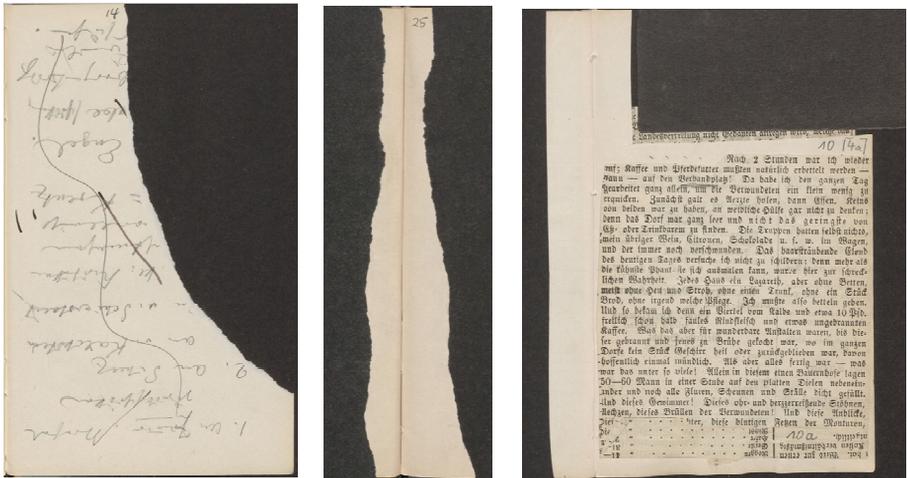


Abb. 5 Collage von Fragmenten herausgerissener und herausgeschnittener Blätter; angefertigt von Judith Michaelis

23 Vgl. hierzu weiterführend Radecke (2013: 168 f.). Zur Unterscheidung von Notizbuch-Editionen, die der Materialität und Medialität von Notizbüchern verpflichtet sind, und einer Edition von Notizbuchaufzeichnungen, die inhaltsorientiert lediglich ausgewählte Aufzeichnungen chronologisch und inhaltlich geordnet veröffentlicht, ohne den Notizbuch-Kontext zu berücksichtigen (vgl. Radecke 2013).

24 Alle Zitate sind entnommen aus Hurlbusch (1995: 34).

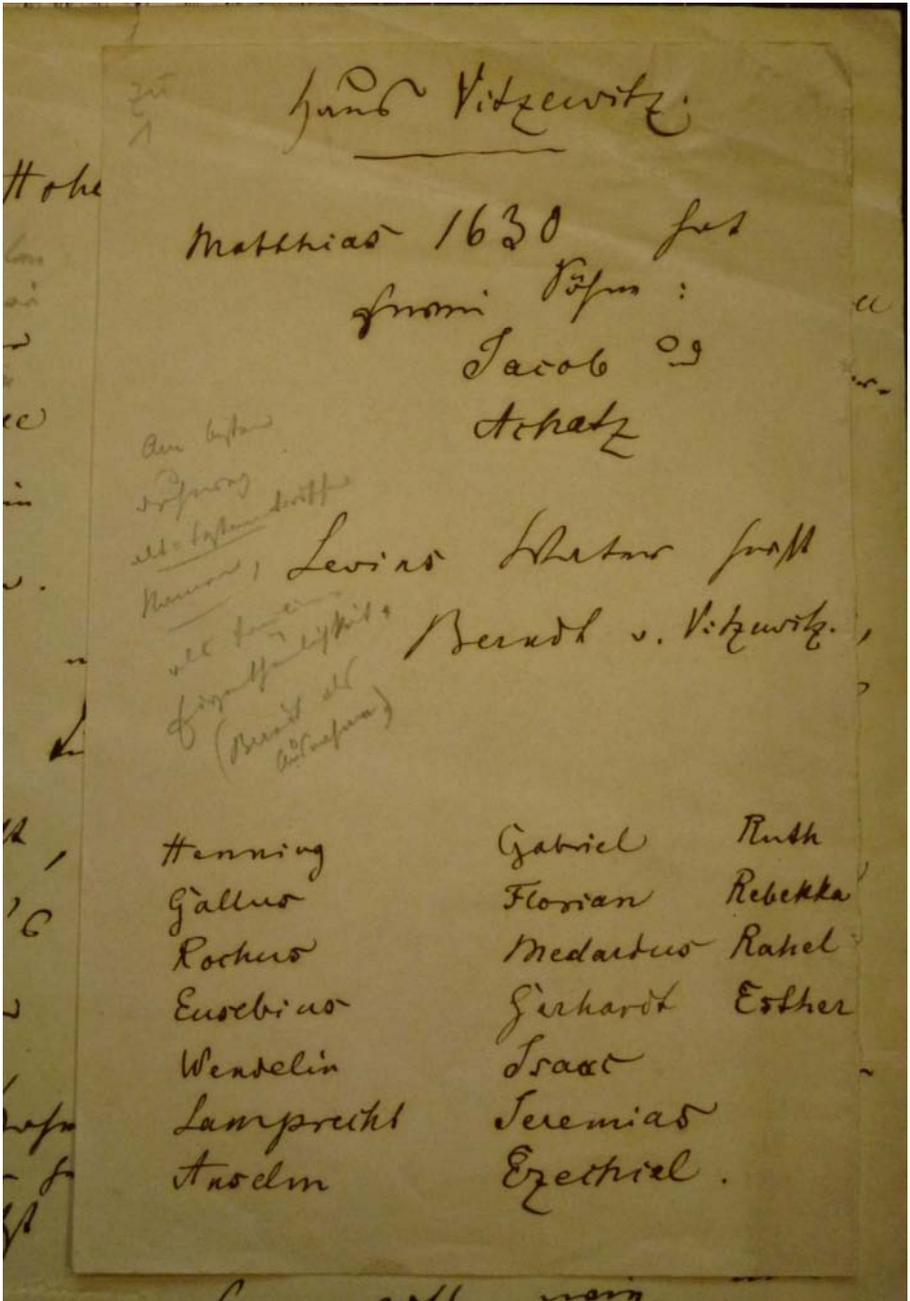


Abb. 6 Fontane: „Vor dem Sturm“, ehemaliges Notizbuchblatt.
 Stiftung Stadtmuseum Berlin, Inv.-Nr. V-67/870, Buch 1, Kap. 2, Blatt 1r

Das gilt auch für die Folierung, um etwa durch eine Zählung in arabischen und römischen Ziffern sowie durch Buchstaben alle Notizbuchblätter von den auf- und angeklebten Blättern, den Vorsatzblättern und Beilagen zu unterscheiden und somit eine Basis für die editorische Referenzierung zu erzielen. Für das Fontane-Projekt waren sowohl eine Neufolierung als auch eine Neudigitalisierung erforderlich, da die Erstfolierung sich als fehlerhaft und unsystematisch erwies. Zudem wurden bei der archivalischen Erstdigitalisierung nicht nur die Blattfragmente ignoriert, sondern blieben auch alle unbeschrifteten Seiten unerfasst, obwohl diese – ebenso wie die beschrifteten Blätter und Blattfragmente – als Bestandteil eines Notizbuchs über die Arbeitsweise, die Schreibchronologie, den Beginn und das Ende einer Niederschrift und insbesondere über die weitere Verwendung ehemaliger Notizbuchseiten Aufschluss geben. Denn Fontane hatte Gründe, die Notizbuchblätter aus den Notizbüchern herauszutrennen: Mittlerweile gibt es einige Funde in anderen Fontane-Konvoluten verschiedener Archive, Bibliotheken und Museen, die belegen, dass Fontane die ursprünglichen Notizbuchnotate insbesondere für sein erzählerisches Werk weiterverarbeitete, etwa für „Graf Petöfy“²⁵ oder auch für „Vor dem Sturm“ (vgl. Abb. 6).

Statt einer Zusammenfassung: Konsequenzen für die Edition

Im Unterschied zu den bisherigen, ausschließlich an inhaltlichen Kriterien orientierten Buchpublikationen ausgewählter Notizbuchaufzeichnungen Fontanes stellt das Konzept der Genetisch-kritischen und kommentierten Hybrid-Edition nunmehr die komplexe Überlieferung mit ihren materialen Kennzeichen und ihren medialen Bedingungen in den Mittelpunkt. Aus diesem Grunde werden die Notizbuch-Digitalisate in einer synoptischen Ansicht zusammen mit den diplomatischen Transkriptionen und zeichengetreuen

25 Bereits im Rahmen der Arbeit an der Großen Brandenburger [Fontane-] Ausgabe wurden ehemalige Notizbuchblätter gefunden, die Fontane für seine Arbeit am Roman „Graf Petöfy“ weiterverwendete; vgl. Radecke (2013: 162 f., Fußnote 55 und Abb. 5). Die ehemaligen Notizbuchblätter befinden sich heute nicht mehr bei den Notizbüchern, sondern in den Fontane'schen Roman- und Werk-Konvoluten im Stadtmuseum Berlin, in der Staatsbibliothek zu Berlin, Preußischer Kulturbesitz, und im Theodor-Fontane-Archiv der Universität Potsdam. Sie werden im Rahmen der Arbeit an der Notizbuch-Edition durch Autopsie, insbesondere durch Abmessungen und Abgleich der Schnittkanten, ermittelt und mit der Edition publiziert.

Edierten Texten veröffentlicht. Um dieses Ziel einer Notizbuch-Edition zu erreichen, war zunächst eine aufwendige Materialautopsie erforderlich und für das Gesamtvolumen von rund 10.000 Blatt wurde etwa ein Sechstel der Projektlaufzeit in Anspruch genommen.

Wenngleich im Rahmen dieses Beitrags lediglich zwei Beispiele, Duktus und Blattfragmente, ausgewählt werden konnten, so bestätigen diese bereits die Theorien zur editorischen Relevanz der Materialität nunmehr auch aus editionspraktischer Sicht. Dieses Ergebnis hat weitreichende Folgen für die Entscheidung, ob eine materialbasierte Edition analog als Buch oder digital in einem Portal publiziert werden soll. Wenngleich es nach wie vor viele Vorzüge gibt, Bücher zu edieren, weil sie Lesetexte nach inhaltlichen Kriterien linear und in einer geordneten Systematik aufbereiten, so erfordern gerade Editionen von heterogenem Material wie Notizbücher²⁶ eine digitale Veröffentlichung, in der – mehr als in gedruckten Ausgaben – die materialen Eigenschaften bewahrt, adäquat als farbige Digitalisate, diplomatische Transkriptionen und Edierte Texte ediert und somit auf unterschiedliche Weise rezipiert werden können. Zur Umsetzung der Ergebnisse der Materialautopsie und des daraus resultierenden Editionskonzeptes bedarf es neben den editionsphilologischen Methoden und Prinzipien freilich weiterer Methoden – insbesondere aus der Informatik und der Informationswissenschaft –, mit denen es schließlich erst gelingt, die detailreichen und sehr differenzierten materialen und genetischen philologischen Beschreibungen des Ausgangsmaterials in eine standardisierte und maschinenlesbare Auszeichnungssprache zu übertragen und die Ergebnisse zu visualisieren. Hinzu kommen Virtuelle Forschungsumgebungen wie TextGrid, die durch ihre vielfältigen Angebote wissenschaftliche Arbeitsprozesse entscheidend unterstützen und somit eine „wesentliche Voraussetzung für die effiziente und erfolgreiche Projektdurchführung“²⁷ bieten.

26 Zur Heterogenität von Notizbüchern vgl. Radecke (2013: 150 f.).

27 Burch (2014), Abschnitt „Virtuelle Forschungsumgebung“

Literaturverzeichnis

Theodor Fontanes Notizbücher sind Eigentum der Staatsbibliothek zu Berlin, Preussischer Kulturbesitz, Handschriftenabteilung, Signatur: Nachl._Theodor_Fontane,Notizbücher.²⁸

Websites des Fontane-Notizbuch-Projekts: <http://www.fontane-notizbuecher.de> und <http://www.uni-goettingen.de/de/303691.html>

Bohnenkamp, Anne; Wiethölder, Waltraud (Hrsg.) (2008): *Der Brief – Ereignis & Objekt. Veranstaltet vom Freien Deutschen Hochstift – Frankfurter Goethe-Museum und der Goethe-Universität Frankfurt am Main in Verbindung mit dem Deutschen Literaturarchiv Marbach*. Frankfurt: Stroemfeld.

Burch, Thomas (2014): Methoden, Verfahren und Werkzeuge zur Erstellung digitaler Editionen. Konzeption und Entwicklung ergonomischer Softwarelösungen zur Abbildung des Arbeitsablaufes in editionswissenschaftlichen Forschungsvorhaben. In: *Literaturkritik.de* (Sept. 2014). http://www.literaturkritik.de/public/rezension.php?rez_id=19691&ausgabe=201409.

Duden online: Stichwort „Autopsie“. <http://www.duden.de/suchen/dudenonline/Autopsie>.

Erler, Gotthard; Erler, Therese (Hrsg.) (1994): *Theodor Fontane: Tagebücher 1866–1882, 1884–1898* (Große Brandenburger Ausgabe, Tage- und Reisetagebücher; 2). Berlin: Aufbau Verlag.

Hurlebusch, Klaus (1995): Divergenzen des Schreibens vom Lesen. Besonderheiten der Tagebuch- und Briefedition. In: *editio* 9: 18–36.

Kerzel, Martina et al. (2013): Die Edition „Johann Friedrich Blumenbach – online“ der Akademie der Wissenschaften zu Göttingen. In: Neuroth, Heike et al. (Hrsg.): *Evolution der Informationsinfrastruktur: Kooperation zwischen Bibliothek und Wissenschaft*. Glückstadt: Verlag Werner Hülsbusch, S. 107–136.

Lukas, Wolfgang et al. (Hrsg.) (2014): *Text, Material, Medium. Zur Relevanz editorischer Dokumentationen für die literaturwissenschaftliche Interpretation* (Beihefte zu editio; 37). Berlin/Boston: Walter de Gruyter.

Neuroth, Heike et al. (2011): TextGrid – Virtual Environment for the Humanities. In: *The International Journal of Digital Curation* 2 (6): 222–231.

²⁸ Einige ehemalige Notizbuchblätter befinden sich im Stadtmuseum Berlin. Für die Erlaubnis zur Abbildung sei der Staatsbibliothek zu Berlin (Abb. 1–5) sowie dem Stadtmuseum Berlin (Abb. 6) sehr herzlich gedankt.

- Plachta, Bodo (2011): Edition. In: Lauer, Gerhard; Ruhrberg, Christine (Hrsg.): *Lexikon Literaturwissenschaft. Hundert Grundbegriffe*. Stuttgart: Philipp Reclam, S. 66–70.
- Radecke, Gabriele (2008): Theodor Fontanes Notizbücher. Überlegungen zu einer notwendigen Edition. In: Amrein, Ursula; Dieterle, Regina (Hrsg.): *Gottfried Keller und Theodor Fontane. Vom Realismus zur Moderne* (Schriften der Theodor Fontane Gesellschaft; 6). Berlin: Walter de Gruyter, S. 211–233.
- Radecke, Gabriele (2010): Theodor Fontanes Notizbücher. Überlegungen zu einer überlieferungsadäquaten Edition. In: Schubert, Martin (Hrsg.): *Materialität in der Editionswissenschaft* (Beihefte zu editio; 32). Berlin/New York: Walter de Gruyter, S. 95–106.
- Radecke, Gabriele (2013): Notizbuch-Editionen. Zum philologischen Konzept der Genetisch-kritischen und kommentierten Hybrid-Edition von Theodor Fontanes Notizbüchern. In: *editio* 27: 149–172.
- Radecke, Gabriele; Detering, Heinrich (2014): Zu dieser Ausgabe. In: *Theodor Fontane: Von Zwanzig bis Dreißig. Autobiographisches*. Hrsg. von der Theodor Fontane-Arbeitsstelle, Universität Göttingen (Große Brandenburger Ausgabe, Das autobiographische Werk; 3). Berlin: Aufbau Verlag, S. 893–897.
- Radecke, Gabriele et al. (2013): Theodor Fontanes Notizbücher. Genetisch-kritische und kommentierte Hybrid-Edition, erstellt mit der Virtuellen Forschungsumgebung TextGrid. In: Neuroth, Heike et al. (Hrsg.): *Evolution der Informationsinfrastruktur: Kooperation zwischen Bibliothek und Wissenschaft*. Glückstadt: Verlag Werner Hülsbusch, S. 85–105.
- Radecke, Gabriele et al. (2015; im Druck): Vom Nutzen digitaler Editionen – Die Genetisch-kritische und kommentierte Hybrid-Edition von Theodor Fontanes Notizbüchern erstellt mit der Virtuellen Forschungsumgebung TextGrid. In: Bein, Thomas (Hrsg.): *Vom Nutzen digitaler Editionen* (Beihefte zu editio; 39). Berlin/Boston: Walter de Gruyter.
- Rasch, Wolfgang et al. (2013): Gutzkows Korrespondenz – Probleme und Profile eines Editionsprojekts. In: Bohnenkamp, Anne; Richter, Elke (Hrsg.): *Brief-Edition im digitalen Zeitalter* (Beihefte zu editio; 34). Berlin/Boston: Walter de Gruyter, S. 87–107.
- Schubert, Martin (Hrsg.) (2010): *Materialität in der Editionswissenschaft* (Beihefte zu editio; 32). Berlin/New York: Walter de Gruyter.
- Streiter-Buscher, Heide (2010): Die nichtvollendete Biographie. Theodor Fontanes *Karl Blechen*-Fragment. In: Berbig, Roland (Hrsg.): *Fontane als Biograph* (Schriften der Theodor Fontane Gesellschaft; 7). Berlin/New York: Walter de Gruyter, S. 133–172.

Weber, Jutta (2013): Briefnachlässe auf dem Wege zur elektronischen Publikation. Stationen neuer Beziehungen. In: Bohnenkamp, Anne; Richter, Elke (Hrsg.): *Brief-Edition im digitalen Zeitalter* (Beihefte zu editio; 32). Berlin/Boston: Walter de Gruyter, S. 25–35.

Zeller, Hans (1971): Befund und Deutung. Interpretation und Dokumentation als Ziel und Methode der Edition. In: Martens, Gunter; Zeller, Hans (Hrsg.): *Texte und Varianten. Probleme ihrer Edition und Interpretation*. München: Beck, S. 45–89.

Zeller, Hans (2002): Authentizität in der Briefedition. Integrale Darstellung nichtsprachlicher Informationen des Originals. In: *editio* 16: 36–56.

Blumenbach digitalisieren

Einblicke in eine Digitalisierungswerkstatt

von *Martin Liebethuth*

„Hast du eben mal Zeit für eine Begutachtung?“, fragte mich Dr. Helmut Rohlfing, Leiter der Abteilung für Handschriften und Alte Drucke der Niedersächsischen Staats- und Universitätsbibliothek Göttingen (SUB), am Telefon.

Natürlich hatte ich Zeit. Gemeinsam mit Dr. Christian Fieseler, dem Stellvertreter in der Abteilungsleitung, gingen wir im September 2010 ins dortige Handschriftenmagazin. Vorsichtig blätterten wir bald durch die nachgelassenen Handschriften und Drucke, die in der Göttinger Universitätsbibliothek unter der Signaturgruppe „Cod Ms Blumenbach“ gesammelt und aufbewahrt werden. Schon ein erster Blick zeigte: Der konservatorische Erhaltungszustand der Bände, der Handschriften, der mit kleinen Zeichnungen versehenen Originale ließ zu wünschen übrig. Zu bedeutend waren die von Blumenbach niedergeschriebenen Erkenntnisse, um ungenutzt in den Magazinen zu lagern. Nein, diese Bände, diese Handschriften waren häufig benutzt, oft durchblättert. Damit war in den Lesesälen der Bibliothek in den vergangenen Jahrzehnten gearbeitet worden, denn hier bot sich WissenschaftlerInnen ein direkter Einblick in die ungefilterte Gedankenwelt eines bedeutenden Naturforschers des 18. Jahrhunderts.

Neben der intensiven Nutzung durch Besucher der Bibliothek trug auch der natürliche Alterungsprozess der Trägermaterialien, des Papiers und des Kartons, zum Zustand bei, wie wir ihn im Jahr 2010 in Augenschein nehmen mussten.

Und nun sollten diese Materialien digitalisiert werden, sollten einer physischen Belastung ausgesetzt werden, die derjenigen einer weiteren intensiven Nutzung im Lesesaal in nichts nachsteht. Und das in dem damals desolaten Allgemeinzustand.

Uns war bald klar: Ohne eine fachgerechte Restaurierung *vor* der Digitalisierung wäre das nicht zu verantworten. Wie leicht konnten die bereits jetzt am Rande eingerissenen Blätter, die teilweise auch verschmutzt und nachge-

dunkelt waren, während der Aufnahme weitere Schäden erfahren – trotz aller gebotenen Sorgfalt beim Umgang mit den Originalen.

Gemeinsam mit Martina Kerzel, der projektverantwortlichen Kollegin in der SUB für das Kooperationsprojekt *Johann Friedrich Blumenbach – online*, wurde schnell der Beschluss gefasst, dass die fachgerechte Restaurierung der Blumenbach'schen Dokumente Voraussetzung und Bestandteil des Projektes werden musste.

Nachdem auch die kooperierende Akademie der Wissenschaften zu Göttingen, die die Hauptverantwortung dieses Vorhabens trägt, von der Notwendigkeit des vorgeschlagenen Weges überzeugt war, begann der aufwendige Prozess der Restaurierung der Blumenbach'schen Bände.

Die Kolleginnen Magdalena Schumann und Manuela Kreis der Restaurierungsabteilung der SUB Göttingen bearbeiteten in den folgenden Monaten die 23 Bände, die im Projekt *Blumenbach – online* im Fokus der WissenschaftlerInnen stehen.

Dabei wurden zunächst die Einbände von den Buchblöcken gelöst, sodann reinigten die Kolleginnen die Seiten von Verschmutzungen und verstärkten die Blätter, wo es Einrisse oder Knicke gab.



Abb. 1

Mitglieder der Arbeitsgruppe „Blumenbach digital“ (v.l.: Martina Kerzel, Wolfgang Böker, Heiko Weber) besuchen die Restaurierungswerkstatt der SUB Göttingen

Nachdem die Buchblöcke so in einen wesentlich stabileren physischen Zustand gebracht worden waren, konnte der Prozess der Digitalisierung dieser ausgebondenen Titel erfolgen.

Dass ein Einband die Bände zu diesem Zeitpunkt noch nicht wieder schützend einfasste, erwies sich für die Bilddigitalisierung als Vorteil.

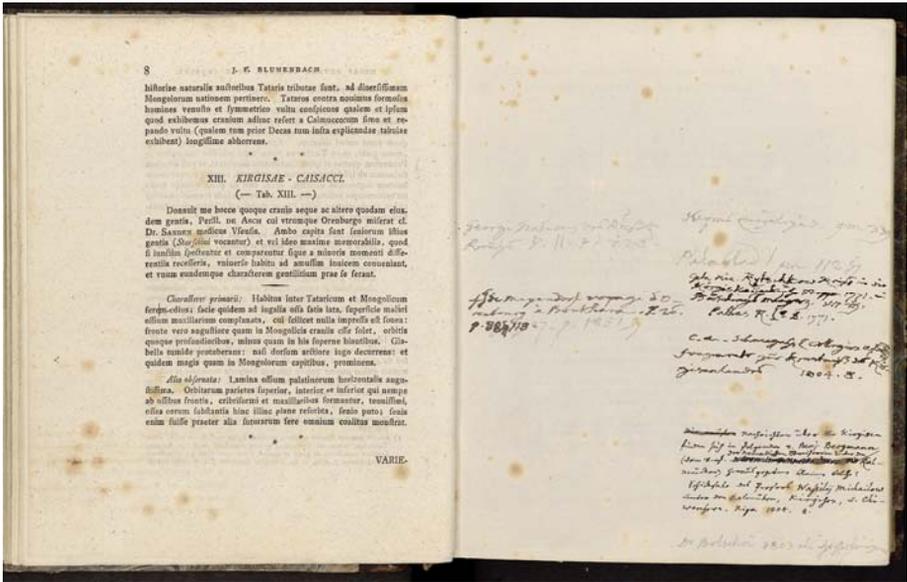


Abb. 2 Doppelseite aus Cod Ms Blumenbach 2. Auf der eingebundenen Leerseite gut erkennbar sind die handschriftlichen Notizen von Blumenbachs Hand.

Das Göttinger Digitalisierungszentrum (GDZ) der SUB verfügt über eine mehr als 15-jährige Erfahrung mit der fachgerechten, bestandsschonenden Digitalisierung unterschiedlichster Objekte aus dem Bestand von Gedächtnisinstitutionen, wobei sich der Schwerpunkt naturgemäß bei Text-Dokumenten verorten lässt.

Im Fall der Blumenbach'schen Handschriften und mit Handschriften durchsetzten Werke konnten wir einen hochwertigen Buchscanner einsetzen, der die unikaln Stücke mit einer überragenden Bildqualität aufnimmt. Dadurch wird gewährleistet, dass die Digitalisate der Originale nahezu alle Fragen der WissenschaftlerInnen beantworten, ohne dass die Handschriften noch einmal einer neuen, möglichst besseren Aufnahme-prozedur unterworfen werden müssten.

Als maßgebliche Richtlinie für Digitalisierungsprojekte in Deutschland haben sich die DFG-Praxisregeln „Digitalisierung“¹ bewährt. Diese werden dem jeweils aktuellen Stand von Wissenschaft und Technik angepasst und fortgeschrieben. Der derzeit gültige Stand stammt aus dem Frühjahr 2013.

Den Vorgaben aus den Praxisregeln sind wir auch bei den Aufnahmen der Blumenbach-Werke gefolgt. In der für Handschriften als Mindestwert ermittelten Auflösung von 400 ppi wurden alle Titel für das Projekt digitalisiert.

Damit werden sämtliche Zeichen, die Informationen darstellen, in die digitale Bilddatei übertragen und können später anhand dieser Datei für die Edition der Blumenbach-Werke ausgelesen und transkribiert werden.

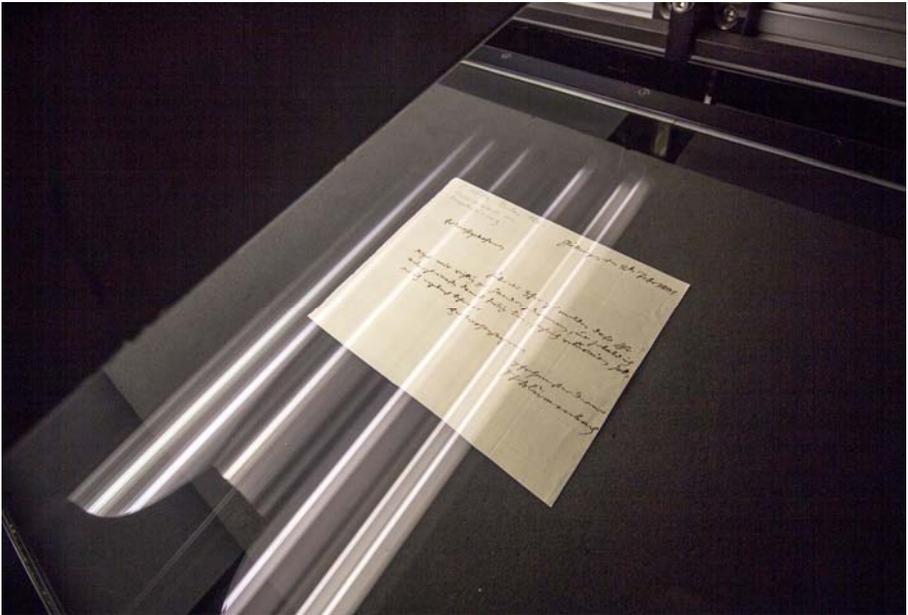


Abb. 3 Brief Blumenbachs im Buchscanner

Für den Prozess der Edition – und ebenso für die Präsentation von digitalisierten Büchern – hat sich als Standard herausgebildet, einzelne Seiten aufzunehmen und anzuzeigen. Bei der eigentlichen Bilddigitalisierung wird diese Anforderung durch die Software der Buchscanner unterstützt: Die Werke werden auf den Buchscanner gelegt; eine aufliegende Glasscheibe sorgt für eine gleichbleibende Fokusebene, für die optimale Präsentations-

1 http://www.dfg.de/formulare/12_151/12_151_de.pdf

form des Buches für die Kamera sowie für die konstante Position der Seiten. Die über der Glasscheibe angebrachte Digitalkamera nimmt das unten liegende Objekt auf und die Software teilt die beiden jeweils zu sehenden Buchseiten automatisch in zwei einzelne Bilddateien auf.

Auch wenn dieser Vorgang des Aufnehmens als eintönige, immer wieder die gleichen Handgriffe erfordernde Tätigkeit aufgefasst werden kann, so verlangt sie von den OperateurInnen doch höchste Präzision, Sorgfalt und Konzentration. Die Ergebnisse dieser Tätigkeiten sind nach der Aufarbeitung für alle Interessierten im Internet zu verfolgen, werden also maximal öffentlich zugänglich sein.

Ein Spezifikum des *Blumenbach – online*-Vorhabens besteht darin, einen Vielzahl der von Blumenbach gesammelten, erforschten und beschriebenen Objekte ebenfalls zu digitalisieren und – neben den digitalisierten Text-Dokumenten – gemeinsam in einer neuen, virtuellen Arbeitsplattform zu präsentieren. Mineralien, Schädel, zoologische Artefakte und Ethnographica werden im Projektkontext digitalisiert und später in die Virtuelle Forschungsumgebung TextGrid importiert. Dort kann auf innovative Weise eine Kontextualisierung von Objekten und Texten erfolgen, damit die Arbeits- und Sichtweisen von Blumenbach und seinen Zeitgenossen nachvollzogen und rekonstruiert werden. Darüber hinaus kann – exemplarisch beginnend bei Blumenbach und seinem Wirken in Göttingen sowie seinem Einfluss auf das damalige europäische Wissenschaftsdenken – die Rezeptionsgeschichte eines bedeutenden Naturforschers entdeckt, untersucht und dokumentiert werden.

Voraussetzung für die weiterführende Forschung ist aber die sorgfältige, komplette und verlässliche Erstellung und Bereitstellung von Digitalisaten der überlieferten Originale. Diese stehen den WissenschaftlerInnen auch immer noch vor Ort in den Bibliotheken und wissenschaftlichen Sammlungen für eine Untersuchung zur Verfügung, sind aber im Blumenbach-Projekt an einen neuen, jederzeit zugänglichen virtuellen Arbeitsplatz in das Internet-Zeitalter übertragen worden.

Usability in Forschungsinfrastrukturen für die Geisteswissenschaften

Erfahrungen und Einsichten aus TextGrid III

von *Ben Kaden und Simone Rieger*

In der geisteswissenschaftlichen Forschung werden Usability-Studien über Tools und Dienste bisher nur selten durchgeführt. Dies liegt zum einen daran, dass der Bereich der Digital Humanities – d.h. die geisteswissenschaftliche Forschung mit Unterstützung moderner Informationstechnologien und damit verbundenen eigenen Softwareentwicklungen – eine recht junge Disziplin mit neuen Anforderungen und Rahmenbedingungen darstellt. Zum anderen wird Usability gerade im Bereich der Entwicklung von Forschungsinfrastrukturen und elektronischen Arbeitsumgebungen immer noch als eher nachgeordnete Anforderung angesehen.

Jedoch ist die Usability gerade für die Akzeptanz von Forschungsinfrastrukturen und ihrer Komponenten von zentraler Bedeutung. Erst über sie wird die methodische und analytische Leistungsfähigkeit digitaler Werkzeuge und Infrastrukturen voll nutzbar.

Es stellte sich u.a. für TextGrid heraus, dass die Handhabung einer Forschungsumgebung durch die Nutzer überwiegend als sehr oder sogar zu komplex empfunden wird. Dies wird, wenigstens in der Einarbeitungsphase von TextGrid, zunächst häufig als Herausforderung empfunden.

In TextGrid III gab es daher zur Klärung der Frage, wie sich die Usability verbessern lässt, ein gesondertes Arbeitspaket (AP 5.3). Dieses reflektierte intensiv – u.a. in Kooperation mit dem AP 6.1 (Integriertes Monitoring) – die Rolle von Usability und erarbeitete besonders auf Basis eines Arbeitsworkshops im September 2014 mit Vertretern der Infrastrukturentwicklung in der Wissenschaft, der Digital Humanities sowie von Usability-Experten einen Überblick zum Stand der Wahrnehmung des Themas Usability im Kontext der Entwicklung von Forschungsinfrastrukturen. Ermittelt wurden vor allem Desiderate, wobei die Bedarfe (a) einer umfassenden Sensibilisierung für das Thema, (b) des Austausches von Expertise und schließlich (c) der Berücksichtigung in den Rahmenbedingungen projektbasierter Infrastrukturentwick-

lung besonders prominent herausstachen. Der Beitrag fasst die Ergebnisse dieser Diskussionen zusammen und dient zugleich als Grundlage für einen zweifellos notwendigen weiteren Austausch zum Thema.

Forschungsinfrastrukturen als Ökosystem

Forschungsinfrastrukturen lassen sich mit einem digitalen Ökosystem vergleichen. Diese Analogie beruht auf der Erkenntnis, dass sich in den Infrastrukturen dynamische und zum Teil auch entwicklungs offene Wechselwirkungen von Forschungsfragen, Werkzeugen und Rahmenbedingungen finden. Die Forschungsinfrastruktur gibt notwendigerweise als Rahmen das vor, was für mit ihr durchgeführte Forschungsprozesse essenziell ist. Im Idealfall betrifft dies aber nur bestimmte Forschungskonventionen und -routinen. Darüber hinaus sollte sie offen und erweiterbar sein, um Entwicklungen in den jeweiligen Disziplinen, ihren Gegenständen und Methodologien zureichend integrieren zu können. Forschungsinfrastrukturen bleiben für eine Wissenschaftsgemeinschaft nur dann relevant bzw. nutzbar, wenn sie die in dieser Gemeinschaft anzutreffenden Forschungsinteressen umfassend aufgreifen können. An diesem Punkt müssen sich Wissenschaftler, Infrastrukturanbieter und Entwickler (bzw. die Software-Maintenance) permanent verständigen. Konkrete Entwicklungsprozesse sollten aus diesem Grund iterativ und dialogisch erfolgen. Im laufenden Dauer- bzw. Routinebetrieb ist es zweckmäßig, Veränderungen im Anforderungsprofil an eine Forschungsinfrastruktur frühzeitig zu erkennen und zielführende Steuerungs- und Anpassungsmaßnahmen anzuregen.

Warum Usability?

Es gibt jedoch neben diesen forschungsspezifischen Interessen auch das forschungspraktische Interesse an einer einfachen, klaren und verlässlichen Benutzbarkeit der Funktionen, die eine Forschungsinfrastruktur bereitstellt. Wissenschaftler möchten sich vorrangig auf ihre Forschung konzentrieren. Werkzeuge, Infrastrukturen und Hilfsmittel sollen die Forschung unterstützen, selbst jedoch so wenig Aufmerksamkeit wie möglich auf sich ziehen. Der Einarbeitungs- und Anpassungsaufwand für solche Anwendungen sollte auf ein Mindestmaß begrenzt bleiben. Im Routineeinsatz sollten Forschungsinfrastrukturen eindeutig strukturierte und organisierte Interaktionsräume

sein, die den Nutzenden keinen besonderen Zusatzaufwand abverlangen. Eine entsprechende Optimierung von Anwendungen übernimmt das sogenannte Usability-Engineering, das neben den Betreibern einer Infrastruktur, den Wissenschaftlern und den Entwicklern als weiterer Akteur bei der Umsetzung und dem Betrieb von Forschungsinfrastrukturen aktiv wird.

Die Einsicht in die Bedeutung von Usability für den Erfolg von Forschungsinfrastrukturen ist weder überraschend noch neu. Warum jedoch ist es in der Realität so schwierig, diese Anforderungen zufriedenstellend umzusetzen? Und weshalb findet sich das Usability-Engineering bislang eher selten im Kontext der Entwicklung von Forschungsinfrastrukturen? Um hierauf eine Antwort zu finden, enthielt die dritte Projektphase einen entsprechenden Reflexionsschwerpunkt, den die Arbeitspakete AP 5.3 (Usability) und AP 6.1 (Integriertes Monitoring) aufgriffen.

Leitaspekte:

Modularisierung, Kompatibilität und Aktualisierbarkeit

Ein Trend lässt sich deutlich feststellen: Auf einen bestimmten Anwendungsfall begrenzte Einzellösungen sind u. a. auch ökonomisch selten sinnvoll und gewünscht. Gleiches gilt für wenig flexible Globallösungen, die notwendig nur einen bestimmten Grad der Spezifizierung der Forschung ermöglichen.

Generische Werkzeuge sind für häufig auftretende Routineprozesse sinnvoll. Im Forschungs-Workflow müssen aber zugleich Möglichkeiten eines Zuschnitts und einer methodologischen Öffnung gegeben sein.

Damit verbunden ist der Aspekt der Aktualisierbarkeit. Das Ökosystem einer Forschungsinfrastruktur ist heute mehr denn je mit anderen Ökosystemen im Internet vernetzt. Der Webbrowser wird häufig als zentraler Zugangspunkt auch zu Forschungssoftware benutzt. Web-Standards und webfähige Datenstandards werden damit auch für Forschungsinfrastrukturen verbindlich. In dem Maße, in dem sich solche Standards entwickeln, müssen auch Forschungsinfrastrukturen aktualisiert werden. Ein offensiverer Ansatz läge darüber hinaus darin, aus den Erfahrungen der direkten Anbindung an die Wissenschaftsgemeinschaft entsprechende Erkenntnisse unter anderem zur Usability abzuleiten, zu formulieren und in die Entwicklung von Web- und Usability-Standards einzubringen.

Bei der Software-Architektur zeigt sich eine modulare Umsetzung kombinierbarer Anwendungen als einzig gangbarer Weg. Eine Forschungsinfra-

struktur besteht bei diesem Modell aus einem stabilen Infrastruktorkern, an den flexibel Werkzeuge und Services angedockt werden können. Das Ziel ist ein freier Austausch von einzelnen Komponenten und spezifischen Anpassungen je nach den Bedürfnissen der Fachgemeinschaften.

Eine besondere Herausforderung ist die Entwicklung eines gebrauchsfähigen Systems unter Berücksichtigung der separaten Nutzbarkeit auch nur einzelner Werkzeuge und Services. Welche Werkzeuge und Services als Teile des jeweiligen digitalen Ökosystems berücksichtigt und aufgenommen werden, hängt also sehr von deren Zugänglichkeit und Strukturtransparenz (idealerweise als Open Source), Kompatibilität, Stabilität, Abgeschlossenheit und Wiederverwendbarkeit ab.

Ein konkreter Forschungszusammenhang lässt sich in der Regel am besten durch eine Kombination von bestimmten für die Forschungsfragen relevanten Softwarelösungen und -diensten adressieren. Um eine derartige Kombinierbarkeit im Sinne eines Baukastens zu gewährleisten, sind die Werkzeuge und Dienste zureichend generalisierbar, übertragbar und anpassbar zu gestalten. Zudem ist zu berücksichtigen, dass sie in solchen Kombinationsszenarien in Wechselbeziehung mit anderen Anwendungen funktionieren müssen. Softwarearchitektonisch ist Kompatibilität daher eine Grundbedingung.

Auf der Ebene der Usability betrifft die Kompatibilität besonders die direkte Interaktion zwischen den Nutzenden und den Anwendungen auf Bildschirm und/oder Displays. Den stabilen Kern der Infrastruktur bildet hier vor allem eine einheitliche integrierende Oberfläche. Die Usability umfasst hierbei vor allem die Vertrautheit mit Interaktions- und Orientierungselementen auf dieser Oberfläche und die Verlässlichkeit der Interaktionswirkungen. Das Prinzip Kompatibilität greift daher durchaus auch über die Software hinaus. Kurz gesagt: Die Anwendungen müssen mit den Nutzungsansprüchen und -erwartungen der Anwender kompatibel sein.

Erweiterungen und Anpassungen der Funktionalität sollten immer in einem stabilen Interaktionsrahmen geschehen. Um hier bedarfsnahe und möglichst passgenaue Lösungen erzeugen zu können, empfiehlt sich eine sehr enge und iterative Kooperation von Softwareentwicklung, Usability-Engineering und den Anwendern. Dies betrifft vor allem die stabilen Elemente einer Infrastruktur. Infrastrukturanbieter müssen sich an dieser Stelle ihre Verantwortung für die Wissenschaft bewusst sein: Eine Forschungsinfrastruktur wirkt immer als Dispositiv, also als Rahmen der möglichen Handlungen der Forschenden. Er muss daher unbedingt forschungsadäquat in Übereinstim-

mung mit den Bedürfnissen der Wissenschaftsgemeinschaft sein und sollte hier so wenige Vorgaben wie möglich beinhalten. Usability bedeutet in diesem Zusammenhang folglich auch, Überlegungen mit einzubeziehen, wie eine Infrastruktur an sich bestimmte Forschungsprozesse ermöglicht. Sie ist keinesfalls auf die Gestaltung von Interaktionsflächen begrenzt, sondern muss auch die passgenaue Abbildung der funktionalen Anforderungen einer Fachgemeinschaft in den Softwarelösungen sicherstellen.

Usability als Chance zur Optimierung

Ein optimales Usability-Engineering ist unbestreitbar aufwendig und zum Teil auch kostenintensiv. Auf lange Sicht erweisen sich diese Kosten jedoch als auch ökonomisch sinnvolle Investition, wenn die Usability-optimierten Infrastrukturen einerseits wirksam zur Verbesserung der Produktivität von Forschungsprozessen beitragen und andererseits den Möglichkeitsrahmen von Forschungsfragen und -methodologien erweitern, also Kreativität in der Wissenschaft fördern.

Insofern kann man schlussfolgern:

- *Usability schafft Akzeptanz*

Akzeptanz einer Software bei den Anwendern lässt sich an der freiwilligen und ausgiebigen Nutzung ablesen. Dies ist in der Regel dann erwartbar, wenn die Lösung einem Anwender einen Mehrwert im Rahmen seiner Forschungstätigkeit oder einen Erfolg für den weiteren Forschungsverlauf verspricht.

Der Mehrwert wird oft von der Anwendern dadurch ermittelt, dass sie den Einarbeitungs- und Anpassungsaufwand bei einer Software mit den gewünschten Resultaten (Output) und den Verbesserungen zur Erzeugung dieser Resultate bei der Anwendung dieser Software in Beziehung setzen. Ist der Einarbeitungs- und Anpassungsaufwand sehr hoch, steht einer Akzeptanz oft bereits die enge Taktung der Projektlaufzeiten im Weg. Ein Forschungsprojekt kann es sich zumeist nicht leisten, mehrere Monate in die Beschäftigung allein mit der Software zu investieren, besonders wenn unklar ist, ob die Software die angestrebten Ergebnisse in der gewünschten Form gewährleistet. In der Praxis greift man daher lieber auf bereits bekannte Lösungen, auch wenn diese veraltet sind. Neue Lösungen, wie sie beispielsweise auch TextGrid in der Produktivversion darstellte, stehen folglich zunächst in einer Bringschuld. Sie müssen anschaulich vermitteln, dass sie für den Forschungsprozess reibungs- und

aufwandsarm einsetzbar sind und zugleich die angestrebten Resultate des Forschungsprojektes hervorbringen können. Generell gilt: Je gebrauchsfreundlicher eine Software, desto eher wird sie genutzt. Das Erreichen einer Akzeptanz hängt von verschiedenen Faktoren ab. Was die Anwender als Mehrwert empfinden, variiert. Entsprechend ist auch hier von Bedeutung, die stabilen Elemente (Routinen) und die variablen Elemente sauber zu trennen und kombinierbar zu halten.

- *Usability hilft, Qualität zu sichern*

Gibt es keine Konkurrenz- oder Alternativprodukte, kann man in der Wissenschaft erfahrungsgemäß von einer vergleichsweise hohen „Leidenschaft“ und Einarbeitungsbereitschaft ausgehen. Projektpraktische Beschränkungen, wie zum Beispiel kurze Förderzeiten, führen allerdings bei schlechter Usability und hohem Einarbeitungs- bzw. Anpassungsaufwand oft dazu, dass die Ergebnisse eines Forschungsprozesses nicht die Qualität aufweisen, die eigentlich gewünscht wäre. Klare Strukturen, die Berücksichtigung von Standards und eine überschaubare Bedienbarkeit ermöglichen es den Wissenschaftlern, sich auf den Forschungsprozess, den Forschungsgegenstand bzw. die Forschungsdaten und die Ergebnisdarstellung (Daten- und Erkenntnispublikation) zu konzentrieren. Idealerweise beinhalten entsprechende Lösungen direkt Elemente zum Beispiel zur Ergebnisvalidierung oder anderer Formen der Qualitätssicherung. Das ist dann besonders relevant, wenn man Usability nicht allein als Software-, sondern auch als eine Art Forschungs-Usability definiert. Eine hohe Datenqualität, die eindeutige Nachvollziehbarkeit der Abläufe zur Datenerzeugung und auch die rechtliche Transparenz hinsichtlich der Nutzbarkeit der gewonnenen Daten ermöglichen ein hohes Maß an Nachnutzbarkeit. Usability sollte daher ganzheitlich verstanden werden und die Anwendungsergebnisse in Forschungsinfrastrukturen als potenzielle Nutzungsobjekte ebenfalls im Blick behalten.

- *Usability unterstützt Effizienz*

Eine einfache Grundregel der Usability für Forschungsinfrastrukturen lautet: Je weniger Zeit die Anwender für das Erlernen und die Bedienung aufwenden müssen, desto besser können sie sich auf die eigentliche Nutzung, nämlich die Forschung, konzentrieren. Wissenschaftler arbeiten in der Regel wissenschaftlich, zeitökonomisch und resultatsorientiert. Ihre Motivation, bestimmte digitale Produkte als Werkzeuge, Hilfsmittel oder Material einzusetzen, ist sehr hoch, wenn diese ihnen bei der Arbeit helfen. Die technische Struktur der Produkte interessiert sie dagegen im All-

tagsbetrieb weniger. Eine gute Usability unterstützt den motivierten und effizienten Gebrauch der Produkte. Wirkt Usability zeitsparend, steigen zudem die Chancen, die Anwender für den aktiven Austausch hinsichtlich einer weiteren Optimierung und Entwicklung der Infrastrukturen zu motivieren. Ein klares Verständnis für die Struktur, Funktionalität und Interaktionsoptionen einer Anwendung unterstützen idealerweise die prinzipielle Prosumer-Idee von Forschungsinfrastrukturen, also eine direkte Kooperation der oben benannten Stakeholder (Betreiber, Nutzer und Entwickler).

- *Usability hilft Kosten sparen*

Ein Usability-optimiertes Produkt, d.h. ein optimal nutzbares Produkt, kann zum einen schneller und ohne großen Mehraufwand, etwa durch zusätzliche Einarbeitungszeit oder Betreuungsaufwand, benutzt werden. Darüber hinaus eröffnet der geschilderte modulare Ansatz neue Möglichkeiten der Nachnutzung einer Entwicklung in anderen Forschungszusammenhängen. Doppelarbeit kann durch klare Standards, eine stabiles Grundgerüst der Infrastruktur und eindeutig beschriebene Interaktions- und Vernetzungsformen (zum Beispiel klare Schnittstellen) vermieden wären. In der Gesamtkostenbewertung dürfte ein Usability-optimiertes Angebot deshalb und auch angesichts der Arbeitskosten der damit tätigen Wissenschaftler für die Wissenschaft und auch die Wissenschaftsförderung kostengünstiger sein.

Zusammenfassend zeigt sich also deutlich, dass die Optimierung der Usability von Forschungsinfrastrukturen langfristig sowohl bei der Akzeptanz dieser Lösungen wie auch bei den Kosten positiv wirken kann.

Herausforderungen

Bei der Analyse der derzeitigen Berücksichtigung von Usability im Kontext virtueller Forschungsumgebungen kristallisierten sich folgende Aspekte heraus:

- Usability wird derzeit generell zu wenig beachtet.
- Die Bedeutung von Usability wird weitgehend unterschätzt bzw. es fehlt häufig eine Sensibilität dafür.
- Usability gilt gegenüber der unverzichtbaren prinzipiellen Funktionalität einer Lösung oft als „nice to have“ und weniger als grundlegende Anforderung.

- Häufig wird nicht verstanden oder anerkannt, dass Usability ein integraler Teil der Software- und Produktentwicklung ist und Funktionalität idealerweise über Usability vermittelt wird.

Erfahrungsgemäß geht man Projekte und Softwarelösungen in diesem Bereich sehr häufig mit Priorisierung der Funktionalität an sich an. Sowohl zeitlich wie auch finanziell begrenzte Ressourcenrahmen motivieren eine entsprechende Gewichtung, da man nachvollziehbar davon ausgeht, dass ein grundlegend seine Funktion erfüllendes Angebot einer optimierten Interaktionsoberfläche vorausgehen muss. Aus einem ganzheitlichen Verständnis von Usability bzw. Benutzbarkeit (bzw. auch von Funktionalität) erweist sich dieser Ansatz jedoch als nicht optimal. Die Funktionalität des Angebots ist nur dann wirklich gegeben, wenn sie von einer reibungs- und aufwandsarmen tatsächlichen Nutzbarkeit durch die Zielgruppe begleitet wird. Eine abgrenzende Priorisierung ist hier keinesfalls sinnvoll. Funktionalitätsentwicklung und Usability-Engineering sind im hier bevorzugten iterativen Entwicklungsmodell untrennbare und in Wechselbeziehung stehende Teile der Infrastrukturentwicklung und -optimierung.

- Werden Prozesse der Usability-Optimierung bzw. -Anpassung durchgeführt, geschieht dies oft zu spät und führt damit zu vergleichsweise hohen Kosten im Entwicklungsprozess.
- Wenn Usability-Tests und -Anpassungen erfolgen, dann geschieht dies meist bei Produkteinführung und in Einsteigersituationen. Nachbesserungen infolge von Erfahrungen aus dem Dauerbetrieb finden kaum statt. Trennt man die Entwicklung der Funktionalität und das Usability-Engineering, ergibt sich die Notwendigkeit, die Benutzbarkeit nachträglich auf bereits un- oder schwerveränderliche Realisierungen in der Software anpassen zu müssen. Strebt man dabei eine optimale Usability an, wird es unter Umständen nötig, bereits abgeschlossene Elemente wieder zu öffnen und umzuarbeiten. Dies ist in realen Entwicklungszusammenhängen oft nicht realisierbar, weshalb die Lösungen in einem suboptimalen Zwischenstadium bleiben. Dies führt zu Akzeptanzproblemen und steht häufig einer modularen Nutzung bzw. Nachnutzung im Weg. Die Nachordnung der Usability hinter die Funktionalität führt weiterhin dazu, dass die Bewertung der Benutzbarkeit (meist durch die Nutzer) umfassend erst beim Launch einer Produktivversion erfolgen kann. Daraus ergibt sich wiederum wenigstens bei projektbasierten Software- und

Infrastrukturentwicklungen das Problem, dass zu diesem Zeitpunkt Entwicklungsressourcen nur noch eingeschränkt zur Verfügung stehen. Selbst dort, wo Nachbesserungen technisch möglich wären, sind Gestaltungsmöglichkeiten nicht selten organisational begrenzt. Das generell nicht sonderlich ausgeprägte Bewusstsein für die Notwendigkeit des Usability-Engineerings zeigt sich zudem darin, dass der Fachdiskurs zu den entsprechenden Fragestellungen im Sinne von Best-Practice-Lösungen und Lessons-Learned-Erfahrungen häufig auf die Usability-Forschung selbst beschränkt bleibt. Diese forscht jedoch selten spezifisch zu Forschungsinfrastrukturen. Ein Wissenstransfer findet kaum statt. Es ist daher zu beobachten, dass sich viele Fehler von Projekt zu Projekt wiederholen. Für eine übergreifende Kommunikation zum Thema fehlt im Bereich der Entwicklung von Forschungsinfrastrukturen bisher das Forum. Möglicherweise lässt sich dies perspektivisch in Institutionen wie beispielsweise DARIAH-DE verankern.

- Die konkreten Anforderungen der Nutzer bzw. Zielgruppe an die Usability eines Angebotes sind in der Regel kaum bekannt und werden auch nicht systematisch ermittelt.

Wie gezeigt wurde, ist die Entwicklung von Forschungsinfrastrukturen idealerweise ein iteratives und agiles Entwicklungsgeschehen, bei dem die Anforderungen der Zielgruppe mit den Möglichkeiten der Betreiber und den Umsetzungen der Entwickler koordiniert werden müssen. Auf einer rein funktionalen Ebene findet dies meist in zufriedenstellender Weise statt. An diesem Punkt sind Wissenschaftler im Regelfall sehr gut in der Lage, ihre funktionalen Anforderungen präzise zu explizieren. Das erleichtert die Übersetzungsleistung zwischen den Anforderungsgebern und den Entwicklern erheblich. Hinsichtlich der Gestaltung der Usability trifft man dagegen häufig auf weniger scharfe Analogien („wie Word“, „wie im Browser üblich“), ohne dass Elemente konkret benannt werden können. Ein auf wissenschaftliche Anwendungen hin geschultes professionelles Usability-Engineering wirkt hier als Vermittler, konkretisiert die Anforderungen und koordiniert diese zugleich mit dem funktionalen Kern der entsprechenden Lösung.

- Die Wissenschaftler, die in entsprechenden Entwicklungs- und Forschungsprojekten mitwirken, verstehen oft nicht zureichend, dass sie nicht nur als forschende Akteure, sondern auch als Abnehmer des Produktes idealerweise auf Augenhöhe und durchgängig in den Entwick-

lungs- und Produktionsprozess eingebunden werden sollten. Oft fehlt auch den Entwicklern genau diese Einsicht.

Eine grundsätzliche Herausforderung bei der Vermittlung von Software- und Infrastrukturdienstleistung in die Wissenschaft besteht in dem Verständnis einer Consumer-Position. Da besonders Geisteswissenschaftler oft nur bedingt Kompetenzen im Bereich der Informatik besitzen, externalisieren sie nicht selten die Verantwortung für entsprechende Gestaltungsentscheidungen an die jeweiligen Experten der Softwareentwicklung. Nach einer initialen Anforderungsbeschreibung erfolgt bis zur Übergabe des Produktes oft keine weitere Interaktion. Die iterative Softwareentwicklung beruht dagegen darauf, die Wissenschaftler in jeden Gestaltungsschritt als Partner einzubinden. Diese permanenten Rückkopplungsprozesse sind auch für die Entwickler aufwendiger als eine lineare Umsetzung der Initialanforderungen. Die iterative und dialogische Umsetzung von Funktionalität und Usability führt jedoch zu passgenaueren und damit eher akzeptierten Lösungen. Das Verständnis dafür muss freilich oft noch sowohl an die Entwickler wie auch an die Wissenschaftler vermittelt werden. Akteure des Usability-Engineerings sind dagegen traditionell eher in diesem Zwischenraum verankert.

- In Forschungsanträgen wird der Aufwand für Usability-Anpassungen selten zutreffend eingeschätzt und vermerkt.
- Eine Bedarfs- und Aufwandsabschätzung (Requirement Engineering) sollte am besten vor Antragstellung vorgenommen werden, erfolgt aber in der Regel auch aufgrund mangelnder Ressourcen nicht oder nur erratisch.
- Förderanträge werden nicht selten sogar ohne Einbindung der Entwickler bzw. von Vertretern des Usability-Engineerings formuliert und eingereicht.

Auch wenn man vermutlich nicht von einer strukturellen Geringschätzung von Aspekten der Usability bei der Forschungsförderung sprechen kann, so fällt doch auf, dass derartige Aspekte vergleichsweise selten in Anträgen berücksichtigt und von Forschungsförderern explizit eingefordert werden. Dabei zeigt sich, wie oben ausgeführt, dass ein professionelles Usability-Engineering eher kostenintensiv ist und ohne entsprechende Ressourcen oft nicht in einem Projektkontext zureichend realisiert werden kann. Häufig beginnt das Problem bereits in der Antragsphase. Es ist im Gegensatz zum Beispiel zur Bewertung des Entwicklungsaufwandes

für die Funktionalität, bisher auch aufgrund fehlender (explizierter) Erfahrungen mit dem Usability-Engineering im Kontext von Forschungsinfrastrukturen kaum möglich, für die Usability-Optimierung notwendige Ressourcen a priori zu ermitteln bzw. abzuschätzen. Sofern es sich nicht um dezidierte Usability-Projekte handelt, wird Expertise in diesem Feld oft erst spät in der Laufzeit des Projektes hinzugezogen. Zu diesem Zeitpunkt sind die entsprechenden Gestaltungsmöglichkeiten allerdings bereits reduziert. Nachteilig wirkt zudem, dass vielfach gar nicht klar ist, wie man entsprechende Experten findet. Wie sehr Usability-Aspekte in einem Projektantrag und damit in einem Projekt berücksichtigt werden, steht in Abhängigkeit zu dem persönlichen Netzwerk bzw. zur individuellen fachlichen Sensibilität der Antragsteller.

- Da Usability nicht systematisch in den Projekten verankert ist, bleibt ihre Umsetzung häufig vom Zufall und der Sensibilität bzw. Kompetenz von Akteuren im Projekt abhängig.

Es fehlt, wie oben zum Fachdiskurs angedeutet, ein systematischer Überblick dazu, wie Usability in die Infrastrukturentwicklung eingebunden werden kann und wo sich entsprechende Akteure befinden. Im Gegensatz zur Softwareentwicklung ist das Usability-Engineering kein umfassend bekanntes Tätigkeitsfeld. Bislang hängt die Berücksichtigung von Usability-Aspekten in der Infrastrukturentwicklung weitgehend von Kenntnis und Bereitschaft der unmittelbar Beteiligten (Projektleiter, Entwickler) ab. Die Fachwissenschaft fordert ebenfalls nur dann eine Berücksichtigung ein, wenn die Beteiligten bereits Erfahrungen auf diesem Gebiet haben.

- Für die Pflege und Aktualisierung von Usability-relevanten Elementen eines Angebotes nach Auslaufen der Projektförderung und/oder in der Produktivnutzung stehen im Normalfall keine dezidierten Ressourcen zur Verfügung. Oft existiert nicht einmal ein diese Notwendigkeit überhaupt berücksichtigendes Planungsszenario.

Die Nachhaltigkeitssicherung ist bei Infrastrukturprojekten generell eine zentrale Herausforderung. Hinsichtlich der Usability und des Usability-Engineerings wiederholt sich hier häufig, was sich bereits bei der Entwicklung vorzeichnet: Es wird eher als Zusatzdienst betrachtet und daher hinter der allgemeinen Betriebssicherung zurückgestellt. Aus Sicht der Usability ist dies weder sinnvoll noch produktiv, da Usability-Optimierungen genauso wie die Software-Aktualisierung integraler Bestandteil des gesamten Lebenszyklus solcher Lösungen sind.

Empfehlungen

Aus Sicht eines Usability-Engineerings sind für die Verbesserung der Bedingungen für die Usability und damit für die Benutzbarkeit von Forschungsinfrastrukturen mehrere Schritte dringend geboten. Generell lässt sich festhalten, dass die Rahmenbedingungen im Wissenschaftsbetrieb und insbesondere auch für Drittmittelprojekte aktuell nicht sehr Usability-freundlich sind. An diesem Punkt ist eine grundsätzliche Veränderung von Einstellungsmustern gegenüber Fragen der Usability und der Rolle des Usability-Engineerings notwendig. Die Strukturen und Bedingungen für die Formulierung, Einreichung und Bewertung von erfolversprechenden Forschungsanträgen lassen, so die Erfahrung, die Beantragung vieler für das Usability-Engineering notwendiger Aspekte nicht zu.

Bei vielen Stakeholdern fehlen neben der entsprechenden Sensibilität zugleich Kompetenzen für die Planung und Umsetzung von Usability-Engineering. Daher werden Ressourcen und der Zeitaufwand für das Usability-Engineering in den Vorhaben oft unterschätzt bzw. nicht eingeplant. Ein Requirement Engineering für die Usability unterbleibt oft. Hier sind eine Kooperation mit Experten auf diesem Gebiet und Weiterbildungen wichtige Maßnahmen. Mit einem wachsenden Verständnis für Fragen der Usability geht idealerweise eine Verschiebung der Perspektive auf das Verhältnis von Funktionalität und Bedienbarkeit einher. Beides sollte bei der Entwicklung von Lösungen als interdependent verstanden und umgesetzt werden. Die Idee einer iterativen Softwareentwicklung, bei der die Forscher, die Entwickler, die Betreiber und nach Möglichkeit auch Vertreter des Usability-Engineerings in jedem Entwicklungsschritt kooperieren, bietet sich hierfür an. Höheren Initialkosten stehen dabei eine größere Akzeptanz, Nachhaltigkeit und Nutzbarkeit gegenüber. Generell ist eine Sensibilisierung für Usability-Forschung und -Engineering notwendig und sollte bereits in den Curricula der Fachdisziplinen und der Digital-Humanities-Ausbildungsgänge als Basiskompetenz erfasst sein.

Schließlich ist besonders hervorzuheben, dass es einer zentralen Anlaufstelle oder eines Kompetenzzentrums für Usability bei der Infrastruktur-, Werkzeug- und Dienstentwicklung für die Digitalen Geisteswissenschaften bedarf. Je stärker geisteswissenschaftliche Forschung auf digitalen Werkzeugen und Infrastrukturen aufsetzt, desto größer wird der Bedarf für eine entsprechende Usability-Optimierung. Anwendbarkeit und Erfolg von die Forschung unterstützenden Angeboten wie TextGrid hängen grundlegend davon

ab, wie das Ergebnis der Aufwand-Nutzen-Abschätzung für die jeweilige Lösung bei der Zielgruppe ausfällt. Ein Angebot, das die Forschung optimiert bzw. verlässlich, stabil und möglichst ohne viel Einarbeitung und Anpassung neue Forschungsmöglichkeiten eröffnet, setzt sich eher durch als eine komplizierte Anwendung, auch wenn diese funktional sehr elaboriert ist. Dabei muss Usability-Optimierung nicht unbedingt in einem Maximalszenario umgesetzt werden, um wirksam zu sein. Bisweilen reichen kleine Schritte, um eine erhebliche Verbesserung zu erzeugen. Aber auch um zu wissen, welches diese Schritte sind, bedarf es eines Sammelns, Erschließens und Vermittelns von Erfahrungen, Kompetenzen und Ansprechpartnern. Die Plattform von DARIAH-DE könnte dafür im Fall der Digital Humanities der richtige Ort sein.

Authentifizierung und Autorisierung im Kontext Virtueller Forschungsumgebungen in den Geisteswissenschaften

von Peter Gietz

In Zeiten, in denen auch im Bereich der Geisteswissenschaften Open-Source-Software und Open-Access-Veröffentlichung von Inhalten, die beide für eine kostenlose und anonyme Nutzung von Software bzw. Inhalten stehen, eine immer größere Bedeutung haben, stellt sich die Frage der Notwendigkeit von Zugriffskontrolle und der dazu notwendigen Technologien für Authentifizierung und Autorisierung. Der folgende Text reflektiert diese Fragen im Rahmen von Virtuellen Forschungsumgebungen in den Geisteswissenschaften und beschreibt die entsprechenden Arbeiten beim Aufbau der TextGrid-Authentifizierungs- und Autorisierungs-Infrastruktur (AAI).

Was sind Authentifizierung und Autorisierung?

Authentifizierung ist ein Prozess, in dem eine Entität – in unserem engeren Sinn eine Person – sich authentisiert, also eine Authentizität behauptet, und eine zweite Entität – in unserem Fall ein Authentifizierungsdienst – die Authentizitätsbehauptung verifiziert.

Eine einfache Implementierung dieses Prozesses ist die Abfrage eines Passworts, welches auf dem Authentifizierungsserver zu einer Identität gespeichert ist und entsprechend überprüft werden kann. Wusste ein Benutzer das Passwort, wurde „bewiesen“, dass es sich bei dem Benutzer um die behauptete Person handelt. Natürlich ist eine Passwort-Überprüfung kein hundertprozentiger Beweis, da es zum einen von einem Dritten erraten worden sein könnte, wenn es sich um ein sehr einfaches Passwort („secret“, „123“ oder etwa der Vorname des Partners) handelt, zum anderen kann das Passwort auf andere Weise von einem Dritten erschlichen worden sein, etwa indem ein Zettel gefunden wurde, auf dem es aufgeschrieben worden ist, oder indem es während einer Übertragung im Netz abgehört wurde. Schließlich ist relevant, welche „Identität“ bewiesen wird, so könnte der Besitzer des Passworts sich bei der Registrierung unter einem falschen Namen angemeldet haben.

Die Sicherheit der Authentizität bei der Authentifizierung ist also abhängig davon, dass die initiale Identifizierung korrekt war (etwa über Einsicht in den Personalausweis), die Übertragung von Geheimnissen (Passwort o.Ä.) nicht abgehört werden kann und dass die Geheimnisse nicht einfach zu erraten sind (erreichbar z.B. über die Erzwingung von entsprechenden Passwortbildungsregeln, *Password Policies*).

Es gibt sicherere Authentifizierungsverfahren als Passwort-Abfragen, wie etwa biometrische Verfahren (z.B. Scan der Iris oder des Fingerabdrucks), Hardware-Tokens (etwa das Handy) oder PKI-Zertifikate (s.u.). Man spricht hier auch von drei unterschiedlichen Faktoren: etwas wissen (Passwort), etwas im Besitz haben (Token), etwas sein (Biometrie). Die Sicherheit beim Authentifizierungsverfahren kann auch dadurch erhöht werden, dass mehrere Faktoren genutzt werden (*Multifactor Authentication*).

Grundsätzlich funktionieren aber alle Authentifizierungsverfahren ähnlich, sodass das Meiste, das oben bezüglich Passwörtern ausgesagt wurde, auch auf die anderen Mechanismen zutrifft.

Nachdem sich nunmehr ein Benutzer authentifiziert hat, wird über einen zweiten, Autorisierung genannten Prozess entschieden, auf welche Ressourcen der Benutzer Zugriff erhalten darf. Autorisierungsentscheidungen können entweder von dem Dienst, der die Ressource bereitstellt – dem sog. *Policy Enforcement Point* (PEP) – getroffen werden oder von einer zentralen Stelle – einem sog. externen *Policy Decision Point* (PDP) –, der der Dienst vertraut.

Autorisierungsentscheidungen werden in der Regel über Eigenschaften gefällt, die zu der digitalen Identität gespeichert wurden – also etwa, dass der Benutzer Mitglied einer Organisation ist (im Falle einer Hochschule etwa als Studierender, Angestellter, zur Fakultät zugehörig, etc.), dass er eine Rolle innehat (Administrator, Projektleiter, Lehrstuhlinhaber, etc.) oder dass er Mitglied in einer bestimmten Berechtigungsgruppe ist (z.B. Gruppe der TextGrid-Nutzer). Grundsätzlich ist es aber auch möglich, die Autorisierungsentscheidung vorwegzunehmen, indem nicht Eigenschaften, sondern auch Berechtigungen (sog. *Entitlements*) bei der digitalen Identität gespeichert werden, wie etwa „Darf im TextGrid-Wiki schreibend zugreifen“.

Für Authentifizierung und Autorisierung gab es zu Beginn von TextGrid eine Reihe von Standard-Technologien, die bei der Architektur für TextGrid berücksichtigt wurden:

- LDAP¹, ein IETF-Standard (vgl. Zeilenga 2006) zum Speichern von Benutzerinformationen und zum Abprüfen von Passwörtern (vgl. auch Gietz 2004).
- PKI² – *Public Key Infrastructure*, ein IETF-Standard, bei der man sich mit digitalen Zertifikaten (vgl. Housley et al. 1999) authentifizieren kann.
- SAML³, ein OASIS-Standard, um innerhalb einer Föderation Authentifizierungs- und Autorisierungsinformationen zu übertragen (vgl. Cantor et al. 2005).
- RBAC⁴, ein ANSI-Standard für rollenbasierte Zugriffskontrolle (vgl. Sandhu et al. 2000).

Im weiteren Verlauf dieses Beitrags werden diese Technologien noch ausführlicher thematisiert.

Wozu braucht man das in den Geisteswissenschaften?

Auch in den Geisteswissenschaften verbreiten sich die Ideen der Open-Access-Bewegung. Nicht nur Publikationen, sondern auch Forschungsdaten sollen hier der Allgemeinheit kostenlos zur Verfügung gestellt werden. Nur die korrekte Nennung der Urheberschaft bleibt selbstverständlich eine *conditio sine qua non*. Insbesondere da Forschung auch in den Geisteswissenschaften in der Regel staatlich finanziert wird, wird weder ein Grund gesehen, die Forschungsergebnisse der Öffentlichkeit vorzuenthalten, noch solle es Verlagen möglich sein, daraus kostenpflichtige Angebote machen zu können.

Dennoch gibt es während des Forschungsprozesses offensichtlich sehr gute Gründe, nicht gleich alle Daten zu veröffentlichen, sondern erst nach Abschluss und Veröffentlichung der Publikation. In der traditionellen Arbeitsweise, in der der Wissenschaftler für sich alleine und nur mit Daten auf seinem eigenen Computer arbeitet, sind die Daten über den Zugriff auf den entsprechenden Arbeitsrechner bzw. den gut zu hütenden Backup-Medien geschützt. Erst wenn man z. B. im Rahmen von VREs kollaborativ arbeitet bzw. die Forschungsdaten auf Servern speichern möchte, kommt Zugriffskontrolle und damit Authentifizierung und Autorisierung ins Spiel.

1 Lightweight Directory Access Protocol

2 Public Key Infrastructure, vgl. <http://datatracker.ietf.org/wg/pkix/charter/>

3 Security Assertion Markup Language; vgl. https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=security.

4 Role Based Access Control: ANSI INCITS: American National Standard 359-2004

Voraussetzungen im TextGrid-Projekt

Als es darum ging, eine Architektur für TextGrid zu entwickeln (vgl. auch Aschenbrenner et al. 2005), waren durch die D-Grid-Initiative, zu der TextGrid fachlich gehörte, eine Reihe von technischen Voraussetzungen gegeben, die zu berücksichtigen waren. Insbesondere gab es eine vordefinierte D-Grid-Plattform, die spezifizierte, dass für die Zurverfügungstellung von Computing- und Storage-Ressourcen eine von drei möglichen Grid-Middleware-Systemen⁵ zu nutzen sind (vgl. Fieseler/Gürich 2007).

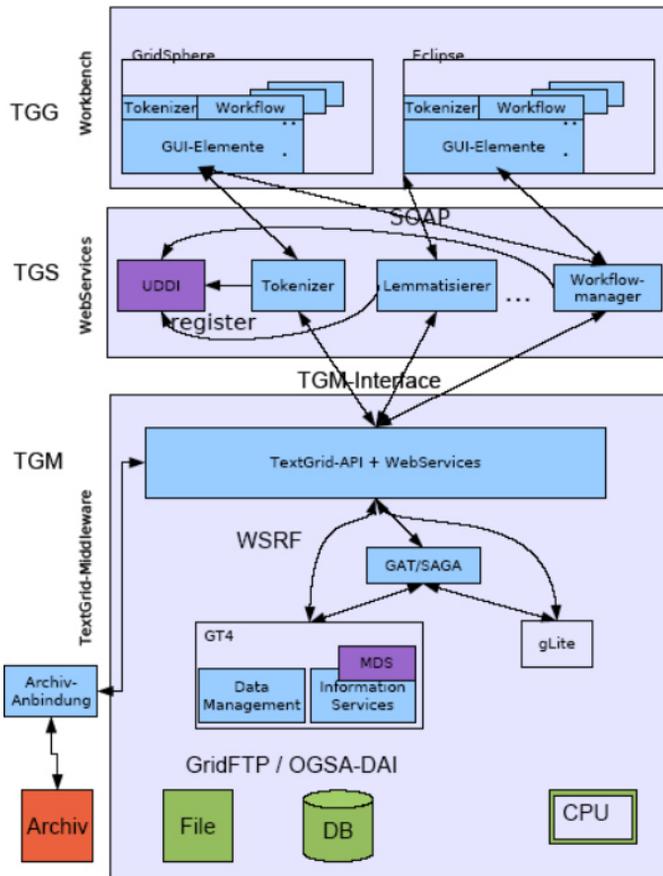


Abb. 1 Frühe TextGrid-Architektur

⁵ Globus Toolkit (<http://toolkit.globus.org/toolkit/>), gLite (<http://grid-deployment.web.cern.ch/grid-deployment/glite-web/>) und Unicore (<https://www.unicore.eu/>)

TextGrid entschied sich für Globus Toolkit. Um möglichst hohe Freiheitsgrade zu erlangen, wurde jedoch zusätzlich entschieden, die Grid-Middleware-Schicht über die generische Schicht GAT⁶ (vgl. Allen et al. 2003) zu abstrahieren, die in D-Grid grundsätzlich für die Anwendungsentwicklung empfohlen wurde. Bereits ein sehr frühes TextGrid-Architektur-Diagramm (vgl. Abb. 1) zeigt die verschiedenen Schichten: Benutzerinterface/GUI (TGG), Services (TGS), Middleware (TGM) und innerhalb der Middleware die Kapselung durch GAT.

Für das Sicherheitsmodell musste ebenfalls eine Brückentechnologie gefunden werden, da hier Grid-Computing (vgl. Foster et al. 1998; Welch et al. 2003) traditionell PKI-basiert war,⁷ es sich aber abzeichnete, dass die relativ komplexe Handhabung hierfür notwendiger Benutzerzertifikate für die Adressaten von E-Humanities nicht sinnvoll war. Deshalb setzte TextGrid von Anfang an auf die im deutschen Hochschulbereich bereits angewandte SAML-basierte Technologie und die darauf aufbauende Hochschul föderation DFN-AAI.

PKI und SAML: zwei Standards für Authentifizierungs-Infrastruktur

PKI und SAML sind zwei unterschiedliche Standards, um Authentifizierungs-Infrastrukturen aufzubauen. Bei diesem Thema muss man grundsätzlich zwischen Nutzer- und Server-Authentifizierung unterscheiden.

PKI basiert auf einem asymmetrischen Verschlüsselungsverfahren, bei dem ein Schlüsselpaar eines privaten und eines öffentlichen Schlüssels generiert wird: Mit dem öffentlichen Schlüssel können Texte verschlüsselt werden, die nur durch den privaten Schlüssel entschlüsselt werden können; mittels des privaten Schlüssels kann ein Text mit einer digitalen Signatur versehen werden, die mit dem öffentlichen Schlüssel verifiziert werden kann. Eine Infrastruktur (die PKI) gewährleistet, dass vertrauenswürdige Dritte (sog. *Certification Authorities*, CA) nach strengen Regeln (*Certification Policies*) die Identität einer Person feststellen und Zertifikate der digitalen Identität ausstellen. Im Zertifikat unterschreibt die CA den öffentlichen Schlüssel

6 Grid Application Toolkit (<http://gridlab.man.poznan.pl/>); die Nachfolgetechnologie ist SAGA (<http://radical-cybertools.github.io/saga-python/index.html>).

7 Vgl. z. B. <http://toolkit.globus.org/toolkit/security/>.

des Benutzers mit seinem privaten Schlüssel. Der Benutzer kann somit über das Zertifikat authentifiziert werden, wenn er im Besitz seines privaten Schlüssels ist und die Signatur der CA überprüft werden konnte. Auch im Fall von Zertifikaten hat also der Benutzer in Form seines privaten Schlüssels – wie beim Passwort – ein Geheimnis in seinem Besitz, das nicht weitergegeben werden darf, das aber sicherer als ein Passwort etwa auf einem Hardware-Token, wie einer Chipkarte, gespeichert werden kann. Die Zusicherung der Identität wurde hierbei allerdings in der Vergangenheit erstellt. Deshalb hat ein Zertifikat nur eine begrenzte Gültigkeit, etwa von einem Jahr, und die Infrastruktur bietet zusätzlich die Möglichkeit, Zertifikate wieder zurückzuziehen, indem die CA sie in einer sog. *Certificate Revocation List* (CRL) veröffentlicht. Der Benutzer muss seinen privaten Schlüssel sicher verwahren und sich regelmäßig um eine Erneuerung des Zertifikats kümmern. Die Komplexität dieser beiden Prozesse ist der Hauptgrund, warum sich in vielen Bereichen Benutzerzertifikate nicht durchgesetzt haben. Schon lange und ubiquitär etabliert ist jedoch die Verwendung von Server-Zertifikaten, bei denen sich ein technisch versierter Server-Systemadministrator um die Zertifikate kümmert.

Das bisher zu PKI Gesagte bezieht sich nur auf den Vorgang der Authentifizierung. Für die Autorisierung muss eine parallele Infrastruktur – eine sog. *Privilege Management Infrastructure* – aufgebaut werden, die sog. Attribut-Zertifikate verwendet, die von sog. *Attribute Authorities* ausgestellt werden.

Bei SAML findet die Zusicherung der Identität gleichzeitig mit dem Authentifizierungsprozess statt. Die Identität wird bei der Erstellung eines Accounts in der Heimatorganisation, etwa bei der Anstellung eines Mitarbeiters oder der Immatrikulation eines Studenten, überprüft. Über das Passwort kann sich ein Benutzer immer wieder authentifizieren. Das kann auch bei einer Anwendung (oder einem *Service*) geschehen, die nicht von der Heimatorganisation betrieben wird. Das SAML-Protokoll ermöglicht es, dass die Benutzerverwaltung – in der Rolle eines *Identity Providers* (IdP) – der Anwendung, die die Rolle des *Service Providers* (SP) innehat, Zusicherungen macht, wie etwa „Dieser Benutzer konnte sich bei mir eben erfolgreich mit dem Passwort authentifizieren“ (*Authentication Assertion*) oder auch „Dieser Benutzer hat das Attribut ‚StudentIn an der Universität X‘“ (*Attribute Assertion*). Die Anwendung kann aufgrund solcher *Assertions* sehr aktuelle Autorisierungsentscheidungen treffen. Wenn derselbe Benutzer später wieder Zugriff auf die Anwendung verlangt, aber nicht mehr dieselben Attribute hat, kann die Anwendung eine neue Entscheidung fällen. Die Funktionen IdP und

SP werden als Server-Software abgebildet. Die Authentizität dieser Server – die wichtig ist, damit der SP der Zusicherung des IdP glauben kann, und damit der IdP darauf vertrauen kann, dass der SP mit den personenbezogenen Daten rechtmäßig umgeht – wird über Serverzertifikate gesichert. Die Verwaltung der Föderation (z. B. der deutschen Hochschul föderation DFN-AAI) besteht im Wesentlichen darin, eine Liste der Serverzertifikate der beteiligten IdPs und SPs zu verwalten, die sich in entsprechenden Verträgen zur Aktualität der Daten (bei IdPs) und zur Datenschutzkonformität (bei SPs) verpflichten. Das SAML-Protokoll ermöglicht zusätzlich *Single Sign On* (SSO), da der Benutzer für die Dauer der Gültigkeit einer Assertion bzw. der damit verbundenen Session (z. B. acht Stunden) nur einmal beim IdP das Passwort eingeben muss, wonach der IdP beliebig vielen SPs entsprechende *Assertions* ausstellen kann.

Eben weil es für den Benutzer wesentlich einfacher ist (keine Benutzerzertifikate, SSO), weil die Technologie die Autorisierung bereits integriert hat und weil sich im deutschen Hochschulbereich eine nachnutzbare entsprechende Föderation entwickelte, haben wir uns in TextGrid für SAML und gegen die im Grid üblichen Benutzerzertifikate entschieden.

RBAC und OpenRBAC

Um nicht für jeden einzelnen Benutzer Zugriffsrechte vergeben zu müssen, wurde der ANSI-Standard *Role Based Access Control* spezifiziert (vgl. Sandhu et al. 2000). Rechte werden für einzelne Rollen vergeben und Benutzer diesen Rollen zugeordnet. Dies vereinfacht das Rechtemanagement erheblich – insbesondere, weil Rollen auch aufgrund von Eigenschaften automatisch zugewiesen werden können (allen Studierenden einer Hochschule kann die Rolle StudentIn zugewiesen werden). RBAC ermöglicht es ebenfalls, die Autorisierungsentscheidung zu externalisieren. Die Entscheidungsregeln müssen also nicht jeweils in allen Anwendungen (PEP) gespeichert werden, sondern können zentral an einer Stelle (dem PDP) verwaltet werden. Die Anwendung kann dann über ein festgelegtes Protokoll Autorisierungsentscheidungen erfragen.

Im Rahmen einer von DAASI betreuten Informatik-Studienarbeit wurde eine stabile, alle wichtigen im Standard spezifizierten Funktionen unterstützende RBAC-Implementierung erstellt, die alle anfallenden Daten (zu Rollen, Benutzern, Sessions, Ressourcen und Rechten) in einem LDAP-Server

speichert. LDAP bot sich hierbei an, da es sich als Speicher für Benutzerverwaltungen bereits etabliert hatte. Diese OpenRBAC⁸ genannte Open-Source-Software wurde nun im TextGrid-Projekt verwendet und um entsprechende Schnittstellen erweitert. Als LDAP-Server kam die Open-Source-Software OpenLDAP⁹ zum Einsatz. In Widmer (2008) wird der Einsatz von OpenRBAC in TextGrid beschrieben.

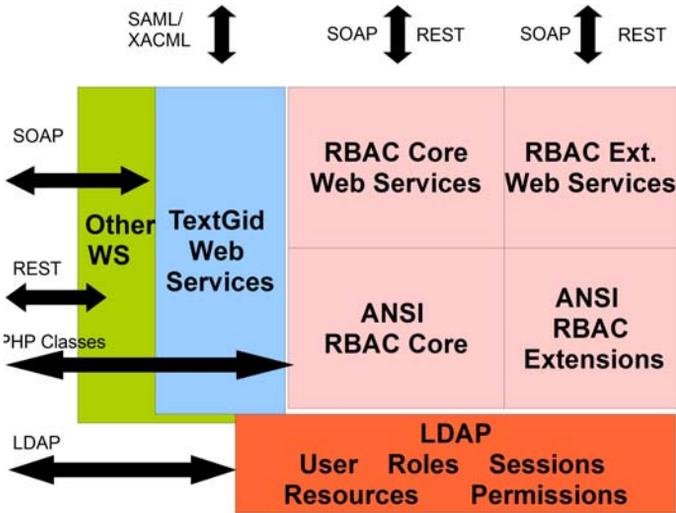


Abb. 2 Architektur von OpenRBAC

OpenRBAC zeichnet sich dadurch aus, dass sowohl die im Standard spezifizierten Managementfunktionen als auch die Abfrage von Zugriffsentscheidungen über verschiedene Protokolle abgerufen werden können. Abbildung 2 zeigt einerseits, dass sowohl der Basis-Standard (*RBAC Core*) als auch die Standarderweiterungen (*RBAC Extensions*) implementiert wurden, ferner, dass die Managementfunktionen sowohl über die Webservice-Protokolle SOAP¹⁰ und REST¹¹ als auch direkt über PHP-Klassen angesprochen werden können. Die Zugriffsentscheidungen können zusätzlich über ein spe-

8 Vgl. www.openrbac.org.

9 Vgl. www.openldap.org.

10 Simple Object Access Protocol – ein OASIS-Standard, der mittlerweile vom W3C weiter gepflegt wird, vgl. <http://www.w3.org/TR/soap/>

11 Representational State Transfer – ein Programmierparadigma für Webanwendungen, das direkt auf das Übertragungsprotokoll HTTP aufsetzt (vgl. Fielding 2000)

ziell hierfür entwickeltes XACML-Protokoll, das auf das SAML-Request- und Response-Protokoll aufsetzt, sowie direkt über das sehr performante LDAP-Protokoll (OpenLDAP erlaubt zigtausende Zugriffsabfragen pro Sekunde) abgefragt werden. Spezialfunktionalität für Projekte wie TextGrid wird über eigene Klassen/Webservices gekapselt.

Nutzungsszenarien und Implementierung von TG-auth*

Eine Umfrage unter den an TextGrid beteiligten geisteswissenschaftlichen Fachwissenschaftlern ergab, dass Rollen nur im Rahmen von Projekten sinnvoll einzusetzen sind. Ein TextGrid-Projekt ist ein Forschungsprojekt, das die TextGrid-Infrastruktur nutzt. Ein solches Projekt wird in OpenRBAC über eine Rolle abgebildet. Die Rollen innerhalb des Projekts werden über die in RBAC vorgesehene Rollenhierarchie integriert.

Folgende Projektrollen wurden definiert:

- **Projektleiter** hat die oberste administrative Funktion eines Projekts inne. Diese Rolle darf Teilnehmern in das Projekt bzw. in Projektrollen aufnehmen und wieder entfernen und administrative Funktionen delegieren. Außerdem darf nur der Projektleiter Ressourcen publizieren, also von jeglicher Zugriffskontrolle befreien.
- **Administrator** kann Ressourcen lesen und löschen.
- **Bearbeiter** darf neue Ressourcen anlegen, ändern und projektinterne Ressourcen lesen.
- **Beobachter** darf in der Regel Projektressourcen lesen.

Abbildung 3 stellt die gesamte Architektur der AAI in TextGrid dar. Über das Benutzerinterface (TextGridLab) authentifiziert sich der Benutzer. Hierbei gibt es die Möglichkeit, einen TextGrid-Account zu verwenden, der im TextGrid-LDAP-Server verwaltet wird oder aber über die DFN-Föderation und den Account an der Heimathochschule, also den Account, den der Benutzer für seine E-Mails verwendet (hier abgebildet als IdP2, IdP3).

Hierfür wurde für TextGrid ein mittels der Open-Source-Software Shibboleth implementierter SAML-IdP aufgesetzt (hier „homeless“ IdP), der an den TextGrid-LDAP angeschlossen ist, sowie ein zentraler ebenfalls auf Shibboleth aufbauender TextGrid-SP. Sobald der SP entsprechende Assertions von einem IdP erhalten hat, wird über den mittels OpenRBAC implementierten PDP entschieden, ob der Benutzer generell über seine Rollen Zugriff bekommen darf, sowie ein Sitzungs-Token (*Session ID*, SID) erstellt

und in dem in TextGridLab integrierten Webbrowser gespeichert. Über die SID spricht das Lab dann den zentralen Ressourcen-Verwaltungsdienst TG-crud (Create, Read, Update, Delete) an. TG-crud fragt bei jedem von dem Benutzer gewünschten Zugriff über die SID beim PDP an, ob sie entsprechende Berechtigungen hat und führt im positiven Falle über das dem TG-crud ausgestellte sog. Robot-Zertifikat die entsprechenden Dateioperationen im zunächst Globus-basierten Grid-Storage durch. In der Grid-Middleware selbst wird das Robot-Zertifikat mit einem generischen Benutzeraccount gemappt und die Operation im Dateisystem als dieser gemappte Benutzer durchgeführt.

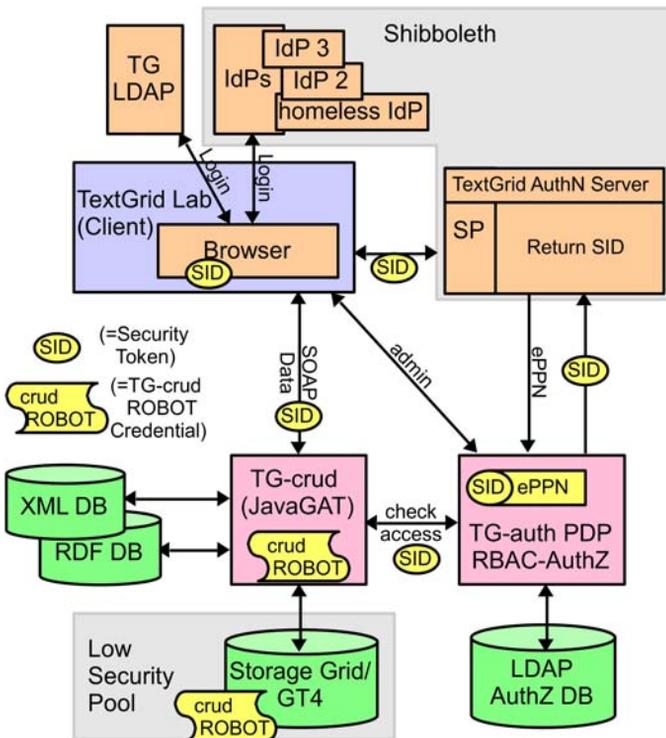


Abb. 3 Architektur von TG-auth*

Hier genau lag das Problem dieser Lösung, da alle TextGrid-Benutzer im Dateisystem des Ressourcen-Providers (theoretisch ein beliebiges im Rahmen von D-Grid engagiertes Rechenzentrum) als ein und derselbe Nutzer abgebildet wurden. Der Ressourcen-Provider hat also nicht die Möglichkeit, den eigentlichen Nutzer zu identifizieren. Deshalb wurde im Rahmen des

D-Grid-Projekts GAP-SLC¹² eine bessere Integration der TextGrid-Lösung in die Grid-Infrastruktur implementiert. Diese wird in Abbildung 4 dargestellt. Hierbei kamen sog. *Short Lived Credentials* (SLC) zum Einsatz. Grundsätzlich sind SLC normale PKI-basierte Benutzerzertifikate, mit dem einzigen Unterschied, dass sie automatisch nach einer erfolgreichen z.B. SAML-basierten Authentifizierung von einem sog. *SLC-Service* (SLCS) erstellt werden (und nicht von einer strengen CA) und dass sie nur eine Gültigkeit von wenigen Stunden haben (und nicht etwa von einem Jahr). Grundsätzliche Sicherheitsaspekte solcher Lösungen werden in Groeper et al. (2009) behandelt.

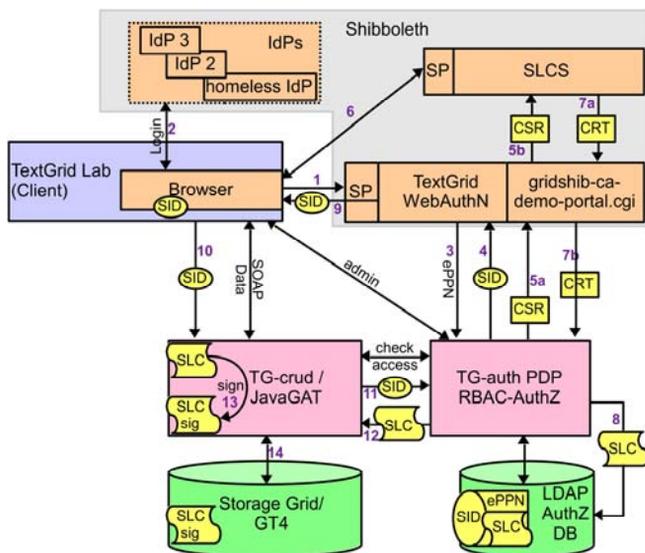


Abb. 4 Integration von Short Lived Credentials

In der TextGrid-Implementierung (vgl. auch Pinkernell et al. 2010) wird nach erfolgreicher Authentifizierung (die wie gehabt über SAML/Shibboleth erfolgt) von TG-auth* ein Schlüsselpaar (privater und öffentlicher Schlüssel) und ein Zertifizierungsantrag (*Certificate Signing Request, CSR*) erzeugt. Dieser CSR wird über eine im Rahmen des Gridshib-Projekts¹³ (vgl. Barton et al. 2005) erstellte Software an den SLCS weitergereicht, der das Zertifikat

12 Vgl. <http://gap-slc.awi.de/index.html>.

13 Vgl. <http://www.cilogon.org/gridshib> und <https://spaces.internet2.edu/display/GS/ProjectSummary>. Die ursprünglichen Seiten des Projekts unter Globus.org existieren leider nicht mehr.

(im Bild zunächst CRT und dann zusammen mit dem privaten Schlüssel SLC genannt) erstellt. Das Zertifikat wird auf dem gleichen Weg wieder zurückgereicht und von TG-auth* im LDAP-Server gespeichert. Wenn nun TG-crud wie gehabt beim PDP über die SID Zugriffentscheidungen erfragt, erhält es zusätzlich zur Entscheidung das SLC, das es dann an Stelle des Robot-Zertifikats verwendet, um die Dateioperationen auf dem Grid-Storage durchzuführen. Für jeden Benutzer kommt also ein eigenes Zertifikat beim Ressourcen-Provider an, das über das im Grid etablierte Mapping-Verfahren zu je eigenen Nutzer-Accounts übersetzt wird, sodass dort der Benutzer nachvollziehbar ist. Die über die Rollen gesteuerten Rechte können jetzt zusätzlich auf Grid-Storage-Ebene durchgesetzt werden. Hierzu wurde eine Software entwickelt, die die entsprechenden Ressourceninformationen aus dem LDAP-Server ausliest und in Unix-File-ACLs (*Access Control Lists*) übersetzt.

Zusammenfassung und Ausblick

Bei der Architektur der TextGrid-Middleware wurde großen Wert auf die Verwendung von Standards gelegt – sowie auf Kapselung und Interoperabilität unterschiedlicher Standards. Im Bereich der AAI kamen hierbei die Standards LDAP, SAML und RBAC zum Einsatz. Auf diesen Grundlagen konnte erfolgreich ein Produktivsystem entwickelt werden, das flexibel auf neue Anforderungen angepasst werden konnte. Auch die beschriebene Lösung zur Integration der SAML-basierten TextGrid-AAI in die PKI-basierte Grid-Welt wurde produktionsreif implementiert. Die entsprechende Funktionalität ist in TG-crud mittels Konfiguration an- und abschaltbar. Die Entwicklungen im Rahmen der D-Grid-Initiative und die Entscheidung innerhalb des TextGrid-Projekts, nicht mehr auf Grid-Ressourcen zu speichern, sondern in einem von einem einzigen Rechenzentrum zur Verfügung gestellten redundanten Speicher ohne Integration von Benutzerzertifikaten, bewirkte, dass diese Lösung letztendlich doch nicht mehr produktiv genutzt wurde. Auch die GAT-basierte Kapselung verschiedener Grid-Middlewares wurde hierdurch obsolet. Dennoch konnte die Kapselungsschicht in TG-crud sehr gut bei den aktuellen Bemühungen nachgenutzt werden, um die TextGrid-Middleware in die DARIAH-DE-Infrastruktur¹⁴ zu integrieren.

14 Vgl. <https://de.dariah.eu/>.

Die grundsätzliche Entscheidung, für die Benutzer-Authentifizierung anstelle PKI den damals recht neuen SAML-Standard zu verwenden, hat sich als langfristig richtig erwiesen. Die meisten der heutigen Forschungsinfrastrukturen setzen auf diesen Standard und mittlerweile hat sich nicht nur die DFN-AAI in Deutschland fest etabliert, sondern auch ähnliche Hochschul-föderationen in Europa und allen anderen Kontinenten. Über eduGain¹⁵ steht eine Interföderation zur Verfügung, wodurch all diese nationalen Föderationen miteinander verbunden werden können. In DARIAH wird konsequent eine solche internationale Integration angestrebt. Auch diesbezügliche Arbeiten fußen auf den im Rahmen von TextGrid gemachten Erfahrungen. Die TextGrid-Benutzerverwaltung konnte wegen der Verwendung derselben Standards relativ einfach in die DARIAH-DE-Benutzerverwaltung überführt werden, sodass ein nachhaltiger Betrieb auch nach Ende der TextGrid-Projekte sichergestellt werden kann. Ein ebenfalls wichtiger Nachhaltigkeitsaspekt war die konsequente Verwendung von Open-Source-Software (Shibboleth, OpenLDAP, OpenRBAC, etc.), die dazu führt, dass wesentliche Komponenten der Architektur auch außerhalb geförderter Forschungsprogramme weiter gepflegt werden.

Literaturverzeichnis

- Allen, Gabrielle et al. (2003): Enabling Applications on the Grid – A GridLab Overview. In: *International Journal on High Performance Computing Applications* 17 (4): 449–466.
- Aschenbrenner, Andreas et al. (2005): TextGrid – a modular platform for collaborative textual editing. In: Risse, T. et al. (Hrsg.): *Proceedings of the International Workshop on Digital Library Goes e-Science (DLSci06), September 21, 2006, Alicante, Spain*, <http://www.l3s.de/dlsci/dlsci2006.pdf>, S. 27–36.
- Barton, Tom et al. (2005): Identity federation and attribute-based authorization through the globus toolkit, shibboleth, gridshib, and myproxy. In: Polk, William T./Hastings, Nelson E./Seamons, Kent (Hrsg.): *5th Annual PKI R&D Workshop, October 2005*, NISTIR 7313, NIST, Gaithersburg MD, http://csrc.nist.gov/publications/nistir/ir-7313/NIST-IR-7313_Final.pdf, S. 54–67.
- Cantor, Scott et al. (2005): Assertions and Protocols for the OASIS Security Assertion Markup Language (SAML) V2.0. OASIS Standard, 15 March 2005, <http://docs.oasis-open.org/security/saml/v2.0/saml-core-2.0-os.pdf>.

15 Vgl. <http://services.geant.net/edugain/Pages/Home.aspx>.

- Fielding, Roy (2000): *Architectural Styles and the Design of Network-based Software Architectures*. Dissertation, University of California, Irvine.
- Fieseler, Thomas; Gürich, Wolfgang (2007): Core D-Grid Infrastructure. In: Lin, Simon C.; Yen, Eric (Hrsg.): *Grid Computing: International Symposium on Grid Computing (ISGC 2007)*. New York: Springer, S. 153–162.
- Foster, Ian et al. (1998): A security architecture for computational grids. In: *Proceeding CCS '98 Proceedings of the 5th ACM Conference on Computer and Communications Security*. New York: ACM, S. 83–92.
- Gietz, Peter (2004): Chancen und Risiken LDAP-basierter zentraler Authentifizierungssysteme. In: Thorbrügge, Marco (Hrsg.): *11. Workshop Sicherheit in vernetzten Systemen*. Hamburg: DFN-CERT, S. I-1–I-14
- Groeper, Ralf et al. (2009): A concept for attribute-based authorization on D-Grid resources. In: *Future Generation Computer Systems* 25 (3): 275–280.
- Housley, R[uss] et al. (1999): Internet X.509 Public Key Infrastructure Certificate and CRL Profile. IETF RFC 2459, January 1999, <https://www.ietf.org/rfc/rfc2459>.
- Pinkernell, Stefan et al. (2010): *Task 1: Deliverable D1 – BMBF Projekt „Nutzung von kurzlebigen Zertifikaten in portalbasierten Grids (GapSLC)“*, 6.5.2010, http://gap-slc.awi.de/documents/GapSLC_D1-v1.0.pdf.
- Sandhu, Ravi et al. (2000): The NIST Model for Role Based Access Control: Toward a Unified Standard. In: *Proceedings, 5th ACM Workshop on Role Based Access Control, July 26–27, 2000*. Berlin: ACM, S. 47–63.
- Welch, Von et al. (2003): Security for Grid Services. In: *Twelfth International Symposium on High Performance Distributed Computing (HPDC-12)*, Piscataway, NJ: IEEE Press, S. 48–57.
- Widmer, Markus (2008): Role Based Access Control mit OpenLDAP (TextGrid-interner Bericht), 20. Mai 2008, <http://www.openrbac.de/documents/rbacTextGrid.pdf>.
- Zeilenga, Kurt (2006): Lightweight Directory Access Protocol (LDAP): Technical Specification Road Map. IETF RFC 4510, June 2006, <https://www.ietf.org/rfc/rfc4510>.

Metadaten, LOD und der Mehrwert standardisierter und vernetzter Daten

von Martin de la Iglesia, Nicolas Moretto und Maximilian Brodhun

Neben der Tiefenerschließung der Texte, welche die Grundlage für die Beantwortung vielfältiger Forschungsfragen schafft, besteht ein wesentliches Ziel vieler Editionsprojekte darin, die erstellten digitalen Objekte so bereitzustellen, dass sie von WissenschaftlerInnen ebenso wie der Öffentlichkeit in vielfältiger Weise genutzt werden können. Der erste Aspekt, die Erschließung der Texte, wird bei TextGrid über den internationalen Standard der *Text Encoding Initiative* (TEI) (vgl. TEI Consortium 2014) umgesetzt. Die Bereitstellung der digitalen Objekte orientiert sich an den Grundsätzen von *Linked Open Data* (LOD), d.h. die Objekte sollen in einer offenen und digital vernetzbaren Form angeboten werden. Der Fokus liegt dabei nicht allein auf dem freien Zugang zu den Objekten, sondern gerade auch darauf, die Möglichkeit zu schaffen, auf andere digitale Objekte verweisen zu können und damit das Objekt im Kontext darzustellen und verständlich zu machen.

Linked Open Data liegt somit die Idee zugrunde, ein Netz aus Daten zu erstellen. Ähnlich wie das Internet ein Netz aus Websites darstellt, soll durch LOD eine vernetzte Struktur von digitalen Objekten erstellt werden. Um LOD effizient zu gestalten, prägte Tim Berners-Lee, der maßgebliche Begründer des World Wide Web, vier Regeln, die bei der Integration von Information in ein bestehendes Netz beachtet werden sollen:

“Use URIs as names for things

Use HTTP URIs so that people can look up those names

When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)

Include links to other URIs. so that they can discover more things.”

(Berners-Lee (2006)

In diesem Beitrag einige der von Berners-Lee angesprochenen Voraussetzungen der Bereitstellung von Texten als LOD im Kontext der Edition von Texten in TextGrid näher betrachtet und anhand einiger konkreter Praxisbeispiele illustriert werden. Im ersten Teil wird ausgeführt werden, wie Texte durch TEI tief erschlossen und mit *Uniform Resource Identifiern* (URI) aus-

gezeichnet werden. Im zweiten Teil wird dargestellt, wie mithilfe des *Resource Description Framework* (RDF)¹ diese erschlossenen Texte vernetzt und zueinander in Beziehung gesetzt werden können. Zudem wird die Rolle beleuchtet, die die Standardisierung von Metadaten und Formaten in diesen beiden Prozessen spielt.

Metadaten in TEI

Zu den wichtigsten Datenformaten, mit denen in TextGrid gearbeitet wird, zählt XML nach den Regeln der Text Encoding Initiative, welches hier schlicht als „TEI“ bezeichnet wird. Es ist daher an dieser Stelle angebracht zu betrachten, in welchen verschiedenen Ausprägungen Metadaten in TEI repräsentiert werden. Üblicherweise enthalten TEI-Dokumente Transkriptionen von gedruckten oder handschriftlichen Vorlagen. Als Metadaten im weiteren Sinne könnte man bereits dasjenige TEI-Markup dieser Texte bezeichnen, welches ihr Erscheinungsbild in der Vorlage (also typografische, strukturelle, buchbinderische, Layout-, Erhaltungszustands- oder andere physisch bzw. materiell evidente Merkmale) nicht nur beschreibt, sondern auch hinsichtlich der urheberischen Absicht (z. B. Tilgung, Hervorhebung, Ersetzung) interpretiert.

Eine tiefere Auszeichnungsstufe ist semantisches Markup, durch das Textteile als Aussagen über die reale Welt gekennzeichnet werden. Beim folgenden Beispiel handelt es sich um eine Datumsangabe, die man wie folgt auszeichnen könnte:²

1 Für eine allgemeine Einführung in RDF vgl. den RDF-Primer, <http://www.w3.org/TR/rdf-primer/>.

2 Dieses und die folgenden Beispiele sind Auszüge aus den noch in Bearbeitung befindlichen TEI-Dokumenten des Projekts *Genetisch-kritische und kommentierte Hybrid-Edition von Theodor Fontanes Notizbüchern basierend auf einer Virtuellen Forschungsumgebung* (www.fontane-notizbuecher.de). Vgl. Radecke (2010), Radecke (2013) sowie den Beitrag von Gabriele Radecke im vorliegenden Band.

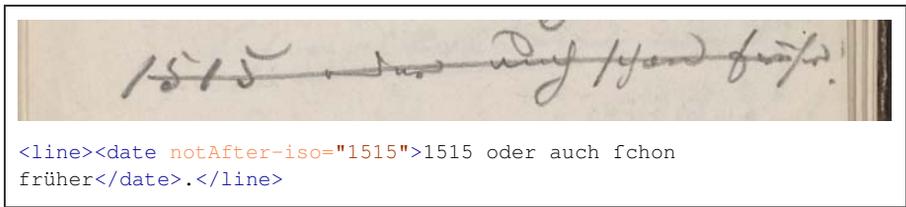


Abb. 1 Scan (Detail) einer Notizbuchseite Theodor Fontanes mit Auszug aus dazugehörigem XML/TEI-Code

Die Angabe, dass etwas sich im Jahr 1515 „oder auch schon früher“ ereignet hat, wird hier gemäß den TEI-Guidelines und dem dort vorgesehenen ISO-Standard 8601 normiert zu `<date notAfter-iso="1515">`.

Desweiteren lassen sich auch Erwähnungen von Personen, Orten und anderen Entitäten in TEI codieren. Grundsätzlich gibt es zur Standardisierung dieser Daten drei verschiedene Möglichkeiten: 1. die Verwendung interner (also frei gewählter und nicht allgemeingültiger) Identifikatoren, 2. die Verwendung von Identifikatoren aus einer externen (idealerweise explizit benannten) Quelle, und 3. die Verwendung von URIs, die auf geeignete Daten, z. B. in Form von *Linked Open Data*, verweisen. In der Praxis werden diese Methoden oft miteinander kombiniert, wie ein weiteres Beispiel aus den Notizbüchern Theodor Fontanes und deren TEI-Encoding zeigt:³

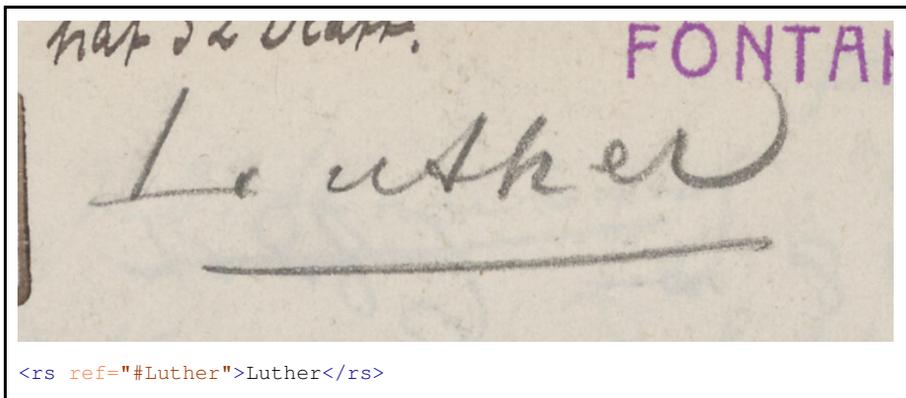


Abb. 2 Scan (Detail) einer weiteren Notizbuchseite Theodor Fontanes mit Auszug aus dazugehörigem XML/TEI-Code

³ Ein ähnlicher TEI-Mechanismus zur Codierung von referenzierten Entitäten findet sich beispielsweise in der Carl-Maria-von-Weber-Gesamtausgabe (2014).

Die Namensnennung „Luther“ wird als Verweis mit einer internen Kennung (hier ebenfalls „Luther“) codiert. Dieser wird an anderer Stelle im TEI-Dokument wie folgt aufgelöst:

```
<person xml:id="Luther">
  <idno type="GND">118575449</idno>
</person>
```

Abb. 3 weiterer Auszug aus einem XML/TEI-Dokument

Hier wird also zusätzlich zur internen Kennung eine externe referenziert, nämlich die Nummer eines Datensatzes in der Gemeinsamen Normdatei der Deutschen Nationalbibliothek.⁴ Dieser Datensatz identifiziert die im Notizbuch erwähnte Person eindeutig als den Reformator Martin Luther. Ein GND-Datensatz ist auch direkt per URI referenzierbar, was in TEI beispielsweise so aussehen könnte:⁵

```
<rs ref="http://d-nb.info/gnd/118575449">Luther</rs>
```

Abb. 4 alternatives Encoding zu dem in Abbildung 2 Gezeigten

Auf diese Weise lassen sich nicht nur im Volltext vorkommende Personen-, Orts- und Körperschaftsnamen sowie Zeitangaben standardisieren, sondern auch editorische Informationen wie zum Beispiel Bucheinbandbeschreibungen, Textgattungsklassifikationen, Provenienzmerkmalsangaben oder Drucktypenbeschreibungen. Voraussetzung dafür ist jeweils das Vorhandensein eines inhaltlich passenden und formal geeigneten Standards, dessen Klassen oder Deskriptoren möglichst leicht, d. h. konzis und eindeutig (idealerweise per URI), referenzierbar sein sollten. Hat man erst einmal ein TEI-Dokument derart mit standardisierten Metadaten ausgezeichnet, eröffnen sich zahlreiche Möglichkeiten, dieses Dokument weiterzuverarbeiten oder es mit anderen Dokumenten zu vergleichen, sofern diese ebenfalls mit ähnlichen Metadaten angereichert sind.⁶

4 <http://www.dnb.de/gnd>

5 Auf ähnliche Weise werden Entitätenreferenzen mit Normdatensätzen in der Online-Edition von Joachim von Sandrarts „Teutscher Academie“ im Projekt Sandrart.net verbunden; vgl. <http://ta.sandrart.net/data/tei-verbose.tgz>.

6 Siehe auch de la Iglesia/Göbel (im Erscheinen).

Ein weiterer Schritt in Richtung Semantic Web sind TEI-Codierungen wie die folgende:

```
<relation type="CRM" name="P74_has_current_or_former_residence"
active="http://d-nb.info/gnd/118575449"
passive="http://www.geonames.org/6548975"/>
```

Abb. 5 TEI-Encoding mit <relation>-Element

Die dadurch getroffene Aussage lautet, dass die durch die GND-Nummer 118575449 bezeichnete Person – Martin Luther – am Ort mit dem Identifikator 6548975 aus der geografischen Datenbank GeoNames⁷ – Wittenberg – gewohnt hat. Unschwer lässt sich an diesem TEI-Beispiel die typische RDF-Tripelstruktur aus Subjekt, Prädikat und Objekt erkennen. Das Prädikat „P74“ ist hierbei dem Standard CIDOC-CRM entnommen. Von einem solchen TEI-Code aus ist es nur noch ein kleiner Schritt zu einem RDF-Triplestore.

Linked Open Data

In TextGrid werden Objekte abgespeichert, die Editionen, Werke, Kollektionen und Items in XML-Format sowie die Volltexte in TEI-Format repräsentieren. Diese werden mit Metadaten erfasst und in einer Datenbank abgespeichert. Zwischen diesen Objekten bestehen Beziehungen, die in der Natur der Sache liegen. Jedes Werk ist eventuell mehreren Editionen zuzuordnen, welche wiederum verschiedene Volltexte haben und auch wieder Teil von anderen Editionen sein können oder auch Bestandteil einer Kollektion sind. Um diese Beziehungen adäquat abspeichern zu können und auch von der Datenbank abfragbar zu machen, spielt *Linked Open Data* (LOD) eine wesentliche Rolle für die technische Umsetzung.

In diesem Kontext setzt TextGrid auf die empfohlenen Standards von Tim Berners-Lee, um LOD effizient umzusetzen. Alle TextGrid-Objekte werden durch ihren Identifier, den TextGrid-URI, bezeichnet und lassen sich mittels http-Request gezielt abfragen. Die Metadaten zu den Objekten werden im Format RDF in XML-Repräsentation in einer Datenbank abgespeichert, um sie mittels der Abfragesprache SPARQL (SPARQL Protocol And RDF Query

⁷ <http://geonames.org/>

Language) abfragen zu können, und sie enthalten Links auf weitere Objekte, welche durch die Verwendung von URIs gesetzt werden. Als Datenbank wird ein Triplestore genutzt. Triplestores sind Datenbanken, die für die Speicherung von Beziehungen zwischen den Objekten optimiert sind.

Abbildung 6 zeigt den Auszug der Metadaten zur Edition von Jules Verne's „Der Einbruch des Meeres“ aus dem Projekt *Digitale Bibliothek*⁸. Durch das `<tg:isEditionOf rdf:resource>`-Feld wird die Edition dem Werk-Objekt zugeordnet, welches durch den TextGrid-URI „textgrid:wXPb.0“ bezeichnet wird.

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xmlns="http://www.tei-c.org/ns/1.0"
  xmlns:ttl="http://textgrid.info/namespaces/metadata/language/
2010"
  xmlns:xi="http://www.w3.org/2001/XInclude"
  ...
  xmlns:tei="http://www.tei-c.org/ns/1.0">
<rdf:Description rdf:about="textgrid:wXP9.0">
  <tg:inProject
rdf:resource="textgrid:project:TGPR-372fe6dc-57f2-6cd4-01b5-
2c4bbefcfd3c"/>
  <owl:sameAs rdf:resource="hdl:11858/00-1734-0000-0005-8745-
4"/>
  <tg:hasPid rdf:resource="hdl:11858/00-1734-0000-0005-8745-
4"/>
  <ore:aggregates rdf:resource="textgrid:wXjg.0"/>
  <ore:aggregates rdf:resource="textgrid:wSn7.0"/>
  <ore:aggregates rdf:resource="textgrid:wVdd.0"/>
  <ore:aggregates rdf:resource="textgrid:wSks.0"/>
  ...
  <dc:format
rdf:resource="textgrid:mime:text/tg.edition+tg.aggregation+xml"/>
  <tg:isEditionOf rdf:resource="textgrid:wXPb.0"/>
</rdf:Description>
</rdf:RDF>
```

Abb. 6 TextGrid-Metadatenauszug einer Edition

Ein technisches Beispiel, welches durch eine Anwendung der in RDF gespeicherten Beziehungen gelöst werden musste, ist eine Funktion der erweiterten Suche im TextGrid Repository.⁹ In dieser besteht die Möglichkeit, die

8 <http://www.textgrid.de/Digitale-Bibliothek>; vgl. auch den Beitrag von Katrin Betz im vorliegenden Band.

9 <http://textgridrep.de/advanced.html>

Autorensuche mit einem Volltext zu verknüpfen, um beispielsweise die Texte abzufragen, in denen Goethe das Wort „Liebe“ benutzt hat.¹⁰

Um die Informationen zusammenzuführen, welche/r AutorIn den einer Edition zugrundeliegenden Text verfasst hat und welcher Volltext zu einer Edition gehört, benötigt man mehrere Abfragen. Durch das Feld <isEditionOf> kann die Beziehung zu dem entsprechenden Werk-Objekt und dem/der zugehörigen AutorIn leicht abgefragt werden. Die Information, welcher Volltext zu der jeweiligen Edition gehört, benötigt komplexere Abfragen, um die URIs herauszufiltern und die Volltext-Dateien durch diese gezielt ansprechen zu können. In Abbildung 7 ist die SPARQL-Abfrage abgebildet, die in TextGrid genutzt wurde, um das beschriebene Ergebnis zu erreichen. Dabei werden zunächst alle URIs ausgewählt, die von einer explizit angegebenen Edition aggregiert werden, deren MIME-Type demjenigen einer Edition in TextGrid entspricht. Aus diesem Ergebnis wird mit der Klausel *FILTER NOT EXISTS* eine weitere Eingrenzung des Ergebnisses durchgeführt. Dabei werden die URIs herausgenommen, die Editionen repräsentieren, die wiederum in Editionen liegen und damit nicht einem zugrunde liegenden Volltext zugeordnet werden können. Ein solcher Fall kann eintreten, falls beispielsweise eine Edition eines Gedichts noch in einem Gedichtband veröffentlicht ist.

```
PREFIX ore:<http://www.openarchives.org/ore/terms/>
PREFIX dc:<http://purl.org/dc/elements/1.1/>
SELECT * WHERE {
    ?edition ore:aggregates* <textgrid:wxp9.0> .
    ?edition dc:format
    <textgrid:mime:text/tg.edition+tg.aggregation+xml> .
    FILTER NOT EXISTS {
        ?edition ore:aggregates+/dc:format
        ore:aggregates+/dc:format
        <textgrid:mime:text/tg.edition+tg.aggregation+xml> }
    }
}
```

Abb. 7 Beispiel einer SPARQL-Abfrage

Bei der Speicherung der RDF-Daten, um Abfragen in SPARQL an diese formulieren zu können, setzt TextGrid auf den De-facto-Standard Sesame¹¹

10 <http://textgridrep.de/results.html?query=edition.agent.value%3A%22Goethe%22%20AND%20Liebe&target=both>

11 <http://www.openrdf.org/>

von OpenRDF. Diese SPARQL-Anfragen können alle NutzerInnen an die in TextGrid gespeicherten Dokumente stellen. Dafür werden Endpunkte sowohl für den öffentlichen¹² als auch für den nicht öffentlichen¹³ Bereich bereitgestellt. Für die Daten, die aus dem dynamischen und damit nicht öffentlichen Speicher bereitgestellt werden, wurden bestimmte Daten, wie zum Beispiel der Titel, weggelassen, sodass keine Rückschlüsse getätigt werden können.

Dieses Beispiel zeigt, dass in LOD großes Potenzial im Management von Daten und ihren Beziehungen untereinander liegt. Die Verwendung von Standards wie RDF und SPARQL ist eine Grundlage für eine fundierte Kollaboration über Grenzen von Institutionen und Projekten hinweg. Um dies zu erreichen, muss *Linked Open Data* auf der technischen Ebene suchbar, zugreifbar, interoperabel und nachnutzbar sein. Durch die Nutzung der hier aufgezeigten Standards kann dies unterstützt werden. Um eine Austauschbarkeit zu erreichen, sind jedoch noch weitere Faktoren relevant, wie die Qualität der jeweiligen Daten und Metadaten und die Zitierbarkeit der vorhandenen Daten.

Metadaten zwischen Standardisierung und Spezialanforderung

Damit diese Zusammenarbeit gelingen kann, müssen die kooperierenden Einrichtungen oder Projektpartner allerdings eine „gemeinsame Sprache“ sprechen. Erst durch die standardkonforme Verwendung der Metadaten-schemata wird ein Vergleich mit anderen Texten oder digitalen Objekten möglich, die nach demselben Standard angereichert wurden. Standardisierung bedeutet Interoperabilität und ermöglicht Austauschbarkeit der Daten. (vgl. van Hooland/Verborgh 2014: 49). Während Standardisierung bei ausschließlich intern verwendeten Bestandsdaten, die nur das eigene Projekt oder die eigene Einrichtung betreffen, noch eine untergeordnete Rolle spielt, stellt sich spätestens beim Austausch der Daten oder der Integration von Daten aus verschiedenen Quellen in ein gemeinsames Angebot die Frage nach gemeinsam befolgten Regeln bei der Beschreibung der Daten und ihres Kontexts. Sobald dann, ausgehend von den eigenen Anforderungen, die Entscheidung für einen Standard gefallen ist, bringt jede Abweichung von diesem Standard auch immer die Gefahr eines Informationsverlusts mit sich.

12 <http://www.textgridlab.org/1.0/triplestore/textgrid-public/>

13 <http://www.textgridlab.org/1.0/triplestore/textgrid-nonpublic/>

Um die Daten an die von anderen verwendeten Schemata anzugleichen, sind dann wieder aufwendige Übersetzungsvorgänge von einem Schema zum anderen erforderlich (vgl. Chan/Zeng 2006, Part II), manchmal ist dies auch gar nicht ohne weiteres möglich. Die Methoden, die zur Standardisierung von Metadaten angewendet werden können, sind vielfältig.¹⁴

Die Formate müssen andererseits jedoch auch zu den Anforderungen des eigenen Projekts oder Vorhabens passen, um für das Projekt nutzbar zu sein – und diese Anforderungen können dann wieder von der vom Standard vorgeschlagenen Lösung abweichen. Auch wenn ein Schema grundsätzlich gemäß dem Standard angewendet wird, bleibt regelmäßig ausreichend Spielraum für eine individuelle Interpretation oder Ausgestaltung eines Anwendungsprofils. Somit genügt selbst Standardkonformität oft nicht, um von übergreifend standardisierten Metadaten sprechen zu können. So wichtig der Vorsatz von Konformität also in der Theorie ist, so schwierig kann es sein, ihn in der Praxis konkret umzusetzen. Auch andere Gründe können Einrichtungen dazu bewegen, einen vom Standard abweichenden Weg zu wählen. Gewachsene Strukturen in den Einrichtungen zum Beispiel können bei der standardkonformen Aufarbeitung von Metadaten ebenso eine Rolle spielen wie das Kosten-Nutzen-Verhältnis der oft zeitaufwendigen Anpassungsprozesse. Standardisierung berührt damit auch Fragen von Personaleinsatz, Ausbildungscurricula und MitarbeiterInnenschulung.

Offene, transparent entwickelte und gut dokumentierte Standards mit aktiven Entwicklungs- und Nutzungs-Communitys erleichtern generell die Standardisierung. Auch eine Trennung von Bestandsformaten, in denen die Daten vorgehalten werden, und Formaten speziell für den Datenaustausch kann hier erste Abhilfe schaffen. Das Format ersterer muss nicht unbedingt einem Standard folgen, während bei letzteren auf Standardkonformität geachtet werden muss. Die beiden Formate müssen dabei nicht identisch sein, was mehr Gestaltungsfreiheit verschafft. Das Austauschformat sollte allerdings bei Auswahl oder Erstellung des Datenhaltungsformats von vornherein berücksichtigt werden, um den späteren standardkonformen Austausch zu erleichtern. Um die Standardisierung von Metadaten zu erleichtern, laufen zudem schon seit geraumer Zeit Bestrebungen in der deutschsprachigen Bib-

14 Siehe für einige Beispiele etwa Chan/Zeng (2006), Part I.

liotheklandschaft, Zertifikate zur Qualitätssicherung (analog bereits existierender ähnlicher Zertifikate für Repositorien) von Metadaten einzuführen.¹⁵

Standardisierung und die offene Vernetzung von Information

Standardisierung von Metadaten spielt auch für den Austausch und die Vernetzung der Daten eine wichtige Rolle, da auch diese nur über die Einhaltung von Standards funktionieren und ihr Potenzial entfalten können. Die Vernetzung löst die Daten vom „Datensilo“ und damit vom Ort und Format ihrer Erfassung. Als Standard für interoperable Metadaten erfüllt RDF einen Bedarf nach der Vernetzung heterogener Information aus verschiedenen Quellen. Dies ermöglicht es zunächst, eine Ressource aus verschiedenen Perspektiven zu betrachten, je nachdem, welcher „Anknüpfungspunkt“ an die Datensätze gewählt wird. Darüber hinaus erlaubt RDF die Nachnutzung einmal gesammelter und vernetzter Informationen in verschiedenen Kontexten, da es möglich ist, immer wieder neue Anknüpfungspunkte zu finden und neue Verbindungen herzustellen und damit das „Datennetz“ zu erweitern und „engmaschiger“ zu gestalten. Dieser Prozess trägt nicht zuletzt zur Qualitätskontrolle bei, da nützliche und qualitätsvolle Daten und komplex angelegte Datensammlungen sich durchsetzen, indem sie durch eine Vielzahl an Links auf andere verweisen und damit sichtbar werden. Zudem werden sie selbst auch wieder mit anderen Datensätzen verbunden, was die Sichtbarkeit erneut erhöht. Datensätze, die diese Voraussetzungen nicht erfüllen, bleiben hingegen unsichtbar (vgl. van Hooland/Verborgh 2014: 210). Damit dieser Prozess der wechselseitigen Datenaufwertung durch Vernetzung gelingen kann, ist die Befolgung des Standards – die gemeinsame Sprache – allerdings unabdingbare Voraussetzung.

Metadatenformate eröffnen bei der Beschreibung, Erschließung, Vernetzung und Verbreitung von Information vielfältige Möglichkeiten. Das volle Potenzial der Daten erschließt sich aber erst dann, wenn viele Projekte und Einrichtungen zusammenarbeiten. *Linked Open Data* kann schon der Definition nach nicht allein verwirklicht werden. Damit dies gelingen kann, sollte soweit wie möglich auf Standards gesetzt werden. Spezifisch auf Texte

15 Siehe hierzu z.B. Dokumente des Kompetenzzentrums Interoperable Metadaten, http://www.kim-forum.org/Subsites/kim/DE/Materialien/Dokumente/dokumente_node.html;jsessionid=A8E32E167AB3450F9CBB88D058DFD655.prod-wor-ker3#doc42066bodyText8.

bezogen bildet XML – selbst ein Standard und hier ausgedrückt nach den standardisierten Regeln von TEI – die Möglichkeit der Tiefenerschließung kulturbezogener Information. Diese Erschließung bildet gleichzeitig die Grundlage, auf die der RDF-Standard zur Vernetzung der Daten aufbaut. Als Ergebnis der Kombination dieser Schritte wird die wissenschaftlich fundierte Suche nach und die disziplinen- wie spartenübergreifende Verbindung von Informationen möglich. Die Vernetzung wiederum eröffnet neue Möglichkeiten, sich Informationen zu erschließen und stellt gleichzeitig ein Angebot an WissenschaftlerInnen (und im Falle von „Open“ Data auch an viele andere Interessierte) dar, einerseits ihre eigenen Daten in Beziehung zu anderen Daten zu setzen, andererseits über den Horizont der eigenen Disziplin hinauszublicken und damit auch neue Perspektiven auf das eigene Fach zu entdecken. Die Verbindung von Informationen über semantische Beziehungen zeigt Zusammenhänge auf, die bei isolierter Betrachtung der Informationseinheiten und digitalen Objekte gar nicht erkennbar waren oder nur mit großem Aufwand sichtbar gemacht werden konnten. Bisher unabhängig voneinander stehende Informationen auf diesem Wege zusammenzuführen, wird daher eine der Aufgaben sein, denen sich Kultureinrichtungen aller Sparten und Universitäten in den kommenden Jahren verstärkt werden stellen müssen, wenn sie die Auffindbarkeit und breite Nutzung ihrer Daten möglich machen wollen.

Literaturverzeichnis

- Berners-Lee, Tim (2006): Linked Data. <http://www.w3.org/DesignIssues/Linked-Data.html>.
- Carl-Maria-von-Weber-Gesamtausgabe* (2014). Digitale Edition, <http://www.weber-gesamtausgabe.de> (Version 1.2 vom 12.06.2014).
- Chan, Lois Mai; Zeng, Marcia Lei (2006): Metadata Interoperability and Standardization. A Study of Methodology. Part I: Achieving Interoperability at the Schema Level, <http://www.dlib.org/dlib/june06/chan/06chan.html>. Part II: Achieving Interoperability at the Record and Repository Levels, <http://www.dlib.org/dlib/june06/zeng/06zeng.html>. In: *D-Lib Magazine* 12 (6).
- de la Iglesia, Martin; Göbel, Mathias (im Druck): From Entity Description to Semantic Analysis. The Case of Theodor Fontane's Notebooks. In: *Journal of the Text Encoding Initiative*.

- Radecke, Gabriele (2010): Theodor Fontanes Notizbücher. Überlegungen zu einer überlieferungsadäquaten Edition. In: Schubert, Martin (Hrsg.): *Materialität in der Editionswissenschaft*. Berlin/New York: De Gruyter (Beihefte zu editio; 32), S. 95–106.
- Radecke, Gabriele (2013): Notizbuch-Editionen. Zum philologischen Konzept der Genetisch-kritischen und kommentierten Hybrid-Edition von Theodor Fontanes Notizbüchern. In: *editio* 27: 149–172.
- TEI Consortium (Hrsg.) (2014): TEI P5. Guidelines for Electronic Text Encoding and Interchange. Version 2.6.0. <http://www.tei-c.org/Guidelines/P5/>.
- van Hooland, Seth; Verborgh, Ruben (2014): *Linked Data for Libraries, Archives and Museums. How to clean, link and publish your metadata*. London: Facet.

III.
Kulturelles Erbe,
Vielfalt der Disziplinen:
Digitale Forschung mit TextGrid

Das Textdatenbank- und Wörterbuchprojekt des Klassischen Maya

Möglichkeiten und Herausforderungen digitaler Epigraphik

von *Christian M. Prager*

Einführung

2014 nahm die Bonner Arbeitsstelle „Textdatenbank und Wörterbuch des Klassischen Maya“¹ der Nordrhein-Westfälischen Akademie der Wissenschaften und der Künste unter der Leitung von Prof. Dr. Nikolai Grube, Abteilung für Altamerikanistik, Universität Bonn, in Kooperation mit der Forschungs- und Entwicklungsabteilung der Staats- und Universitätsbibliothek Göttingen (SUB) sowie der Universitäts- und Landesbibliothek Bonn (ULB) ihre Arbeit an der Erstellung eines maschinenlesbaren Textkorpus und eines darauf basierenden Wörterbuchs der Klassischen Mayasprache auf, die beide in der virtuellen Forschungsumgebung für die Geisteswissenschaften TextGrid angelegt werden (vgl. Neuroth/Lohmeier 2011). In dem auf fünfzehn Jahre angesetzten Vorhaben stehen somit die Sprache und Hieroglyphenschrift der Mayakultur und ihre erstmalige wissenschaftliche Erschließung mithilfe digitaler Methoden in einem online verfügbaren und gedruckten Wörterbuch im Mittelpunkt der Forschungstätigkeit.

Die Mayaschrift ist das bedeutendste lesbare Schriftsystem der präkolumbischen Amerikas und diente den Königen als Kommunikationsmedium für die Speicherung des kulturellen Gedächtnisses, ermöglichte mithilfe von Büchern und Monumenten die Kommunikation in Zeit und Raum und manifestierte in Einheit mit Bildwerken und Architektur eine öffentliche Symbolordnung, in der sich die Herrscher und die Angehörigen des Hofstaates darstellten und *vice versa* damit ihren Machtanspruch legitimierten. Auf dem Gebiet der Halbinsel Yukatan im Süden Mexikos, in Guatemala, Belize und im westlichen Honduras existierten während der Klassik zwischen dem dritten und neunten Jahrhundert Dutzende von unabhängigen Kleinstaaten, deren

1 <http://www.mayawoerterbuch.de>

Herrscher um regionale und überregionale Vorherrschaft und um die Kontrolle von Ressourcen konkurrierten (Abb. 1).

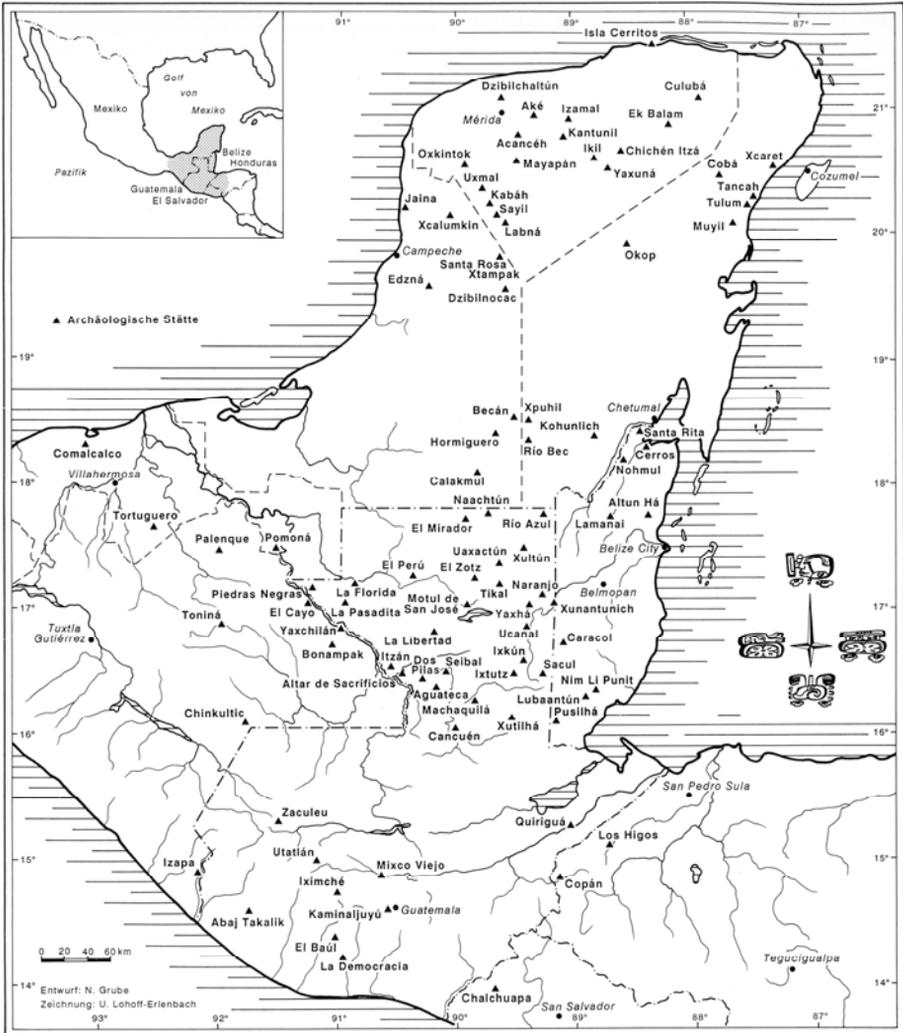


Abb. 1 Karte der Halbinsel Yucatan mit den bedeutendsten Fundstätten der Klassischen Mayakultur (Zeichnung von U. Lohoff-Erlenbach)²

In den aus Tempeln und Wohnbauten bestehenden Stadtzentren regierten Könige mit gottähnlichem Status (*k'uh ajaw* – „Gott-König“) mit ihren Familien, deren Autorität auf dem Fundament der Religion ruhte und ihre Macht

2 Bildquelle: Eggebrecht/Eggebrecht/Grube (1992: 42)

durch Heirats- und Bündnispolitik, Ressourcenkontrolle sowie Kriegsführung etablierten und konsolidierten (vgl. Grube 2010). Als Gottkönige reklamierten sie ihren Machtanspruch in Schrift und Bild und schufen im Zentrum ihrer von bäuerlichen Wohnsiedlungen umgebenen Stadtstaaten imposante Architektur, die den gottgleichen Königen als Kulisse für öffentliche Auftritte und als Ausdruck vollendeter Macht diente (vgl. Prager 2013: 2). Bereits im 4. Jahrhundert v. Chr. entwickelte die Mayabevölkerung im zentralen Mayatiefland ein Schriftsystem im eigentlichen Sinn, indem gesprochene Sprache mit zirka 900 Wort- und Silbenzeichen grafisch realisiert und auf rund 10.000 Textträgern manifestiert wurde.

Weil bis heute weder das Korpus der Maya-Inschriften vollständig inventarisiert noch die einzelnen Hieroglyphentexte maschinenlesbar sind, bleibt die exakte Zahl der Texte und Schriftzeichen (noch) unbestimmt. Für die Kompilation der Textdatenbank und des Wörterbuchs ist dies allerdings dringend erforderlich. Deshalb wird diese Forschungslücke im Rahmen des laufenden Vorhabens mit Methoden und Technologien der Digital Humanities geschlossen, sodass auf der Grundlage digitaler Epigraphik Sprache und Schrift der Maya vollständig erschlossen werden können.

Kontext und Inhalte der Hieroglyphentexte

Die frühesten Schriftträger stammen aus dem dritten vorchristlichen Jahrhundert, späteste Textzeugen sind drei Handschriften auf stückiertem Rindenbastpapier, die kurz vor oder während der spanischen Eroberung im frühen 16. Jahrhundert entstanden und anschließend als Beute nach Europa gebracht wurden; im Verborgenen existierte die Schrifttradition jedoch bis in das 17. Jahrhundert. Die ersten Inschriften mit Kalenderdaten datieren zwischen 36 v. Chr. und 126 n. Chr. und finden sich im Hochland und der Pazifikküstenregion Guatemalas. Während der Schriftgebrauch dort abbrach, setzte spätestens zu Beginn der Klassischen Periode im dritten Jahrhundert mit dem ersten sicher datierbaren Monument des Mayatieflandes eine lange und intensive Maya-Schrifttradition ein, die mindestens bis zur Ankunft der Spanier andauerte (vgl. Fahsen/Grube 2005). Das Gros der Hieroglyphentexte ist allerdings der Klassischen Periode von 300 bis 900 n. Chr. zuzuordnen. Die in dieser Zeit entstandenen Texte enthalten vorwiegend biografische Angaben über die Könige und die Angehörigen ihrer Hofstaaten. Dank Kalenderangaben sind die meisten Texte auf den Tag genau datiert, sodass bei der

Analyse raum-zeitliche Koordinaten vorhanden sind und sich dadurch Varietäten, Entwicklungsstadien oder die Verwendung von Sprache und Schrift genau nachzeichnen lassen. Im Zusammenspiel mit Erkenntnissen archäologischer Forschung sind es vor allem historisch-biografische Textbestandteile mit Berichten über Dynastiegründungen, Geburten, Thronbesteigungen, militärische Siege, Weihungen, Rituale und dynastische Beziehungen, die in den vergangenen Jahrzehnten zum Verständnis der Kultur und Politik der Königshöfe der klassischen Zeit beigetragen und unser heutiges Bild dieser Gesellschaft maßgeblich geprägt haben (vgl. Stuart 1995).

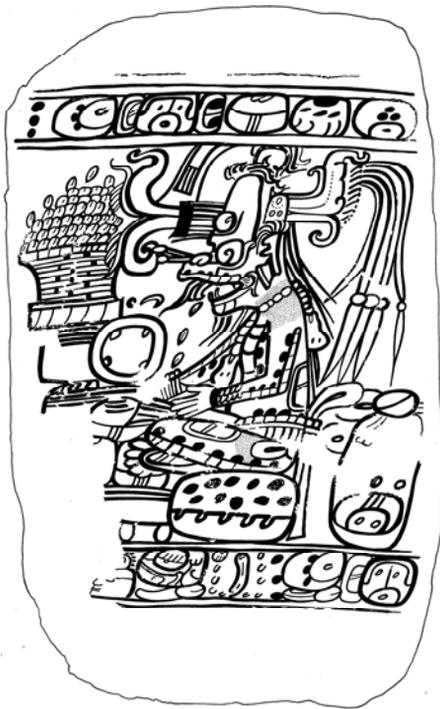


Abb. 2 Bemalter Gewölbedeckstein aus Dzibilnocac (600–900 n. Chr.) mit einer Darstellung des sitzenden K'awiil, der als Fruchtbarkeitsgott laut Inschrift über und unter der Abbildung einen großen Überfluss an Mais, Bohnen, Kakao und Aussaat bringt (Zeichnung von Christian Prager)

Schreib- und Lesefähigkeit war wohl nur wenigen Spezialisten am Königshof vorbehalten, die in einigen Fällen sogar Angehörige der herrschenden Familie waren (vgl. Houston/Stuart 1992). Als Schrift- und Bildträger nutzten die Schreiber (*ajtz'ib*) unterschiedliche Beschreibstoffe, doch in dem

tropisch-feuchten Klima haben sich vorwiegend Trägermaterialien aus Stein erhalten, die seit über 100 Jahren von der Forschung dokumentiert und weiterhin in gedruckter Form veröffentlicht werden (vgl. Graham 1975 ff.; Kerr 1989 ff.; Mayer 1978 ff.). Dabei bilden Online-Publikationen eher die Ausnahme. Erklärtes Ziel des Akademieprojektes ist demgegenüber, Inschriftenfunde systematisch, online und via *Open Access* in einer Inschriftendatenbank einer weltweiten Öffentlichkeit zugänglich zu machen.

Mit 210 beschriebenen und mit Bilddarstellung bemalten Seiten zählen die drei erhaltenen Mayahandschriften aus der Kontaktzeit insgesamt zu den längsten Texten der Mayakultur. Sie werden in Bibliotheken und Museen in Dresden, Madrid und Paris aufbewahrt und enthalten Ritual- und Weissagungskalender, Berechnungen über Venusphasen oder Mond- und Sonnenfinsternisse, Anleitungen zum Jahreswechsel, Bestimmung günstiger und ungünstiger Tage und weitere Themen (vgl. Grube 2012). Die Madrider Mayahandschrift enthält neben Angaben zu Jagd, Landwirtschaft, Nahrungsherstellung und den Einfluss der Götter auf diese Alltagshandlungen auch Anweisungen für die Priester zur Herstellung und Anbetung von Götterbildern oder machen umfangreiche Angaben über die Kultivierung von Bienenstöcken (vgl. Vail/Bricker 2003). Die leporelloartig gefalteten Bücher dienten religiösen Spezialisten bei Anfragen als Nachschlagewerk für ihre kultische Arbeit. Die schriftlichen Erzeugnisse der Maya stehen in ihrer kulturellen Funktion, Bedeutung und Ausfertigung in einer Reihe mit der altägyptischen Schrifttradition, den Keilschriftsystemen Mesopotamiens oder den bekannten Schrifttraditionen auf dem asiatischen Kontinent. Im Gegensatz zu diesen Schriftsystemen ist die Mayaschrift allerdings noch nicht vollständig entschlüsselt, die zugrunde liegende Sprache bis heute nur teilweise verstanden. Das Bonner und Göttinger Langzeitprojekt schließt diese Forschungslücke und setzt sich zum Ziel, den Sprachschatz dieser Kultur erstmals mithilfe digitaler Methoden und Technologien umfassend zu dokumentieren, zu analysieren und zu edieren, um damit die Grundlagen für die vollständige Entzifferung der Mayaschrift zu legen.

Im folgenden Abschnitt werden zentrale Fragen, Desiderate und Herausforderungen der Mayaschriftforschung diskutiert, mit denen sich das Projekt in den kommenden fünfzehn Jahren auseinandersetzen wird, indem es auf Methoden und Technologien der Digitalen Geisteswissenschaften zurückgreifen und diese problembezogen anwenden wird.

Forschungsfragen und -strategien

Die Erforschung der Mayaschrift hat in den vergangenen Jahrzehnten beträchtliche Fortschritte gemacht; die Entzifferung ist jedoch nicht abgeschlossen (vgl. Coe 1992; Houston/Chinchilla Mazariegos/Stuart 2001). Daher bestehen weiterhin Defizite in Bezug auf das Verständnis des Schriftsystems, seiner Struktur, seiner Entstehung und seines Verhältnisses zur Sprache. Diese Defizite gehen zurück auf die erheblichen Lücken in der Erforschung und Dokumentation der Schriftquellen, des Schriftsystems und seiner Entzifferung, der Sprache und der damit vermittelten Inhalte, des Graphem-Inventars und schließlich auf den Bereich der Lexikografie. Grundproblem ist dabei die Punktualität, d. h. die zeit-räumliche Eingeschränktheit bisheriger Forschung, die auf dem Problem beruht, dass die Texte nicht umfassend und maschinenlesbar erfasst sind. So konnten Schrift und Sprache bis heute nicht in ihrer zeitlichen Tiefe, geografischen Verbreitung und unter der Berücksichtigung soziopolitischer und kulturhistorischer Einwirkungen untersucht werden, da die vorhandene Datenmenge zu groß und eine datenbanktechnische Erschließung bisher nicht erfolgt ist.

Defizite in der Erforschung der Mayaschrift existieren im Bereich der systematischen Dokumentation und digitalen Edition des Textmaterials. Bis heute fehlt etwa ein vollständiges Verzeichnis der bekannten Texte oder ein digitales Archiv sämtlicher Textträger, das der Forschung als Referenzwerk online für die Recherche zur Verfügung steht. Die seit Anfang der 1970er-Jahre existierenden Dokumentationsprojekte – darunter das von Ian Graham initiierte *Corpus of Maya Hieroglyphic Programme*, Justin Kerrs *Maya Vase Database* und das Dokumentationsprojekt für Monumente unbekannter Herkunft von Karl Herbert Mayer mit dem Namen *Maya Monuments of Unknown Provenance* – haben sich auf die Dokumentation und Publikation bestimmter Textgruppen spezialisiert und veröffentlichen mit Ausnahme von Justin Kerr bis dato in gedruckter Form (vgl. Graham 1975 ff.; Kerr 1989 ff.; Mayer 1978 ff.). Die Untersuchung des analog publizierten Ausgangsmaterial wird erschwert und der Vergleich von Textpassagen ist ein langwieriges Unterfangen, das angesichts tausender von Textträgern in kurzer Zeit schwer und nur punktuell zu bewältigen ist.

Das Akademieprojekt hat in diesem Bereich eine Forschungslücke erkannt und es sich zum Ziel gesetzt, sämtliche bekannten Inschriftenträger zu sammeln, digital zu erfassen, maschinenlesbar zu machen und für die Forschung über die virtuelle Forschungs- und Editionsplattform TextGrid via

Open Access und als *Linked Data* zugänglich zu machen. Auf der Basis eines EDV-gestützten Textkorpus werden sich etwa paläografische oder inhaltliche Vergleiche von Textpassagen durchführen lassen, die es erlauben, parallele Textaussagen zu erkennen oder im Fall von Texten unbekannter Provenienz den Herkunftsort einzugrenzen. Zu diesem Zweck kooperiert die Bonner Arbeitsstelle unter anderem mit Karl Herbert Mayer in Graz, dessen umfangreiches Archiv mit über 40.000 Fotografien von Textträgern, Architektur und Keramiken derzeit in Bonn digitalisiert, mit Beschreibungs- bez. Metadaten angereichert und im Modell RDF (*Resource Description Framework*) in TextGrid eingepflegt wird.

Dabei handelt es sich um einen informationstechnologischen Standard zur Beschreibung von Metadaten in Form logischer Aussagen.³ Die Metadaten für die Beschreibung von Maya-Textträgern orientieren sich an den etablierten Beschreibungsstandards in der Mayaforschung, die für das Projekt in die Kategorien des CIDOC *Conceptual Reference Model* überführt werden.⁴ Basierend auf dem Prinzip der formalen Semantik, bietet es erweiterbare Ontologien für Begriffe und Informationen im Bereich des Kulturerbes.⁵ Auf diese Weise können epigraphische und inschriftliche Informationen den Begriffen des CIDOC CRM zugeordnet werden. Die u.a. auch Nachhaltigkeit gewährleistende Anwendung dieses Schemas erlaubt zum Beispiel den kontrollierten Austausch von Daten bzw. Informationen zwischen Projekten, Archiven, Bibliotheken oder Museen. Auf diese Weise wird garantiert, dass die Forschungsdaten des Projekts nach vereinbarten, etablierten Standards erfasst und mit anderen epigraphischen oder verwandten Projekten geteilt werden können. Heterogene Forschungsdaten werden auf diese Weise vereinheitlicht und unterstützen die Formulierung von Normen und Konventionen, die auf der anderen Seite wenig(er) Spielraum für individuelle Nomenklaturen und Beschreibungen zulässt.

Eine weitere Herausforderung des Forschungsprojekts, der defizitäre Entzifferungsstand der Mayaschrift, ist bereits erwähnt worden. Weite Teile der Maya-Inschriften sind heute zwar lesbar oder semantisch erschlossen, doch der Prozess der Schriftentzifferung ist mit rund 70 % der etwa 1000 bekannten Wort- und Silbenzeichen noch immer nicht abgeschlossen. Begründet

3 http://de.wikipedia.org/wiki/Resource_Description_Framework

4 <http://network.icom.museum/cidoc/working-groups/overview/>

5 http://de.wikipedia.org/wiki/CIDOC_Conceptual_Reference_Model

liegt dies u. a. darin, dass es sich bei den nicht oder unsicher entzifferten Zeichen häufig um singuläre Grapheme mit wenig Verbreitung oder einem vage verständlichem semantischem Ko- und Kontext handelt. Erschwerend kommt der Umstand hinzu, dass die Texte bislang noch nicht vollständig maschinenlesbar sind, sodass etwa mit korpuslinguistischen Methoden die sprachliche Entzifferung auf ihre sprachliche und semantische Plausibilität hin geprüft werden könnte. Eine Herausforderung bleibt außerdem, dass bestimmte Zeichenfolgen lautsprachlich vollständig gelesen werden, sich jedoch ihrer Übersetzung entziehen, weil die entsprechenden Korrelate in den kolonialzeitlichen und modernen Mayasprachen noch nicht identifiziert wurden oder im Lauf der Sprachgeschichte verloren gegangen sind. Ein Beispiel hierfür ist das Lexem *talim*, das in der Dresdner Mayahandschrift ein Säugetier bezeichnet, wofür in den bekannten Mayasprachen bisher noch kein Korrelat gefunden wurde (Abb. 3).



Codex Dresden, Seite 14, Abschnitt c, Bild 1

Lesefolge: A1, B1, A2, B2, C1, C2

| | | |
|---------------------------|-------------------|-------------------------|
| A1 | B1 | C1 |
| T565:24.502 | 126.552:24 | T121.533:24 |
| ta-li-ma | ya-ta-li | ?-?-li |
| <i>Talim</i> | <i>Yatil</i> | ?! |
| <i>Talim</i> | [ist] der Partner | ? |
| A2 | B2 | C2 |
| T152.1009:23 | T19:670:178.120 | TIII.158 |
| ?-?-na | AL-la-NE | OX W'IL |
| ?n | al ?n | ox w'il |
| [des Gottes] Itzamnaaj | Nachkommenschaft? | Überfluss an Nahrung |

„Talim ist [an diesen Tagen] der Partner des Gottes Itzamnaaj; [es gibt] Nachkommenschaft, ?, Überfluss an Nahrung“

Abb. 3 Ausschnitt aus der Dresdner Mayahandschrift (spätes 15. und frühes 16. Jahrhundert) mit hemerologischem Inhalt und einer Analyse (Koordinaten, Klassifikation, Transliteration, Transkription, wörtliche und sinngemäße Übersetzung)⁶

Das Akademieprojekt wird aus diesem Grund neben der Entwicklung einer virtuellen epigraphischen Forschungsumgebung sämtliche Wörterbü-

⁶ Bildquelle: Sächsische Landesbibliothek Dresden, <http://digital.slub-dresden.de/werkansicht/dlf/2967/14/>

cher moderner und kolonialzeitlicher Mayasprachen digitalisieren und die Einträge datenbanktechnisch erfassen. Die in TextGrid eingepflegten Wortlisten heutiger Mayasprachen lassen sich dadurch leichter durchsuchen und mit Wörtern des Klassischen Maya korrelieren. Durch ihre datenbanktechnische Erfassung lassen sich etwa Wortwurzeln, grammatische Suffixe oder Wortbedeutungen in modernen Mayasprachen effizient und schneller auffinden, um die Funktion und Bedeutungen von Lemmata des Klassischen Maya zu bestimmen. Neben der Kontextanalyse bildet damit der Vergleich mit modernen Mayasprachen eine zweite, zentrale Methode der Lexikografie des Klassischen Maya.

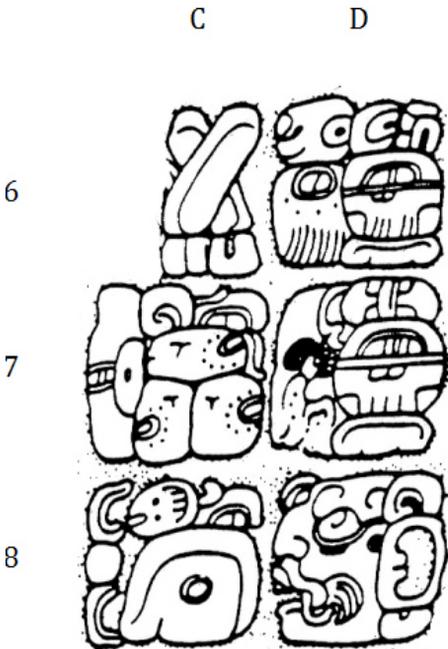


Abb. 4 Ausschnitt der Inschrift auf der zentralen Tafel im Kreuztempel von Palenque, die einen Schöpfungsmythos erzählt, wonach eine für Palenque zentrale Gottheit erst dann aus dem Himmel heruntargestiegen war (D7–D8), nachdem für sie am Horizont (ti'chan, D6b) neue Steine für den Herd (k'ob, D6a) gesetzt wurden (C6).⁷

Ein Problem bei der Erforschung des Klassischen Maya ist, dass bisher publizierte lexikalische Listen häufig auf falschen Entzifferungen beruhen –

⁷ Zeichnung von Linda Schele; aus Schele Drawing Collection, <http://research.famsi.org/uploads/schele/hires/01/IMG0093.jpg>

mit der Folge inkonsistenter Interpretationen und ihrer Verankerung in der Forschungsliteratur. Ein Beispiel hierfür ist die Entzifferung des Graphems mit der Bezeichnung T128 (vgl. Thompson 1962), das vor 20 Jahren als Wortzeichen CH'A für „darniederliegen, flach“ gelesen wurde (vgl. Abb. 4, D6b) (vgl. Schele 1991). Im Zuge der damaligen Diskussion über Schöpfungsmythen bei den vorspanischen Maya wurde auf der Basis der Entzifferung *ch'a chan* als „darniederliegender Himmel“ ein Konzept rekonstruiert, wonach die Maya die Vorstellung besaßen, dass der Himmel vor seiner Erhebung flach ‚darnieder lag‘ (siehe Abb. 4) (vgl. Freidel/Schele/Parker 1993). Obschon die Entzifferung selbst und die daraus abgeleitete Idee eines primordialen, flachen Himmels in der Zwischenzeit widerlegt wurde (siehe Legende zu Abb. 4), bewahren sich solche fehlerhaften Interpretationen bis in die gegenwärtige Forschungsliteratur (vgl. Hermann Lejarazu 2007: 91).

Das Akademieprojekt wird daher kritisch mit bisherigen Forschungsdaten und -ergebnissen umgehen, diese hinterfragen und diesem Umstand bei der Entwicklung der Virtuellen Forschungsumgebung Rechnung tragen. Die Bezugnahme auf die Lesungs- und Entzifferungsgeschichte von Texten und Hieroglyphen ist fundamental und soll in der Datenbank mittels Literaturverweisen und kritischen Kommentaren durchgeführt werden. Ziel der epigraphischen Forschung muss es aber sein, bestehende oder neue Lesungshypothesen zu verifizieren und auf ihre sprachliche und semantische Plausibilität zu prüfen. Es handelt sich dabei um ein dringendes Desiderat, das im Rahmen des Projekts angegangen wird. Grundbedingung hierfür ist allerdings, dass das Zeicheninventar der Mayaschrift vollständig erfasst und laufend aktualisiert wird, sobald Inschriftenfunde mit bislang nicht dokumentierten Zeichen das Gesamtkorpus ergänzen. Zeichenklassifikationen sind besonders bei Schriftsystemen, deren Entzifferung noch nicht abgeschlossen ist, ständig auf ihre Konsistenz und Korrektheit zu überprüfen, da bisher identifizierte Zeichen fehlerhaft klassifiziert, falsch erkannt oder nur unvollständig erfasst sein können oder mit anderen Zeichen ein komplexes Graphem mit einem eigenständigen Lautwert konstituieren können (vgl. Stuart 2009).

Da die bisher publizierten Zeicheninventare in diesen Bereichen unvollständig und fehlerhaft sind, ist die Erarbeitung eines dynamisch konzipierten, stets modifizierbaren digitalen Zeichenkatalogs notwendig. Zu diesem Zweck publiziert das Akademieprojekt auf seiner Website⁸ eine Konkordanz der

8 www.mayawoerterbuch.de

bisher publizierten Zeichenkataloge⁹ inkl. Anzeige der Originalabbildungen, aus der nicht nur die bisher publizierten Klassifikationen und Zeichenidentifikationen ersichtlich sind, sondern auch Kommentierungen der einzelnen Grapheme durch die Mitarbeiter der Forschungsstelle vorgenommen werden können. Auf der Basis der bisherigen Zeichenkataloge entwickelt das Forschungsprojekt ein dynamisch-kritisches Graphem-Inventar, das nicht nur die bisherige Forschung einbindet, sondern tagesaktuell neue Erkenntnis in den dynamisch angelegten Online-Katalog einbinden und publizieren kann. Eine derartige Flexibilität ist im Fall herkömmlicher Druckversionen nicht möglich.

Neben der Dokumentation von Textträgern und der Schrift-Entzifferung besteht auch im Bereich der Schriftstruktur ein intensiver Forschungsbedarf, der alleine auf der Basis digitaler Epigraphik in Angriff genommen werden kann. Schriftstruktur und Schreibkonventionen sind bisher nur oberflächlich untersucht und wurden in Grundzügen von Marc Zender zusammengefasst (vgl. Zender 1999). Bisherige Untersuchungen dazu haben allerdings punktuellen Charakter und konnten aufgrund fehlender digitaler Textkorpora nur an ausgewählten Beispielen vorgenommen werden, sodass zeiträumliche Parameter und die damit verbundene Frage von Kontinuität und Wandel wenig erforscht wurden. Die Rekonstruktion der Konventionen, nach denen sich ein Schreiber richtete, steht erst in den Anfängen. Die Kontroverse um die Bedeutung und Funktion vokalharmonischer und -disharmonischer Schreibungen sowie der Funktion von phonetischen Komplementierungen von logografischen Zeichen hat die bis heute andauernde Debatte um orthografische Konventionen in den letzten Jahren beherrscht (vgl. Houston/Robertson/Stuart 1998; Lacadena García-Gallo/Wichmann 2004). Ihr liegt die Frage zugrunde, ob und in welcher Form die Maya Vokallängen und Vokalqualitäten in der Schrift repräsentierten.

Unberücksichtigt blieben in der Diskussion allerdings die Dynamik des Schriftsystems: Sie erfasst bislang weder räumliche noch zeitliche Aspekte von Schrift- und Sprachphänomenen. Zudem kommt erschwerend hinzu, dass Hypothesen über die Konventionen des Schriftsystems auf problematischen sprachlichen Lesungen beruhen können, wie weiter oben dargestellt. Untersuchungen zur geografischen Verteilung, die Rückschlüsse auf intrakulturelle

9 Vgl. Evreinov/Kosarev/Ustinov (1961); Gates (1931); Grube (1990); Knorozov (1963); Knorozov et al. (1999); Macri/Looper (2003); Macri/Vail (2009); Rendón M./Spescha (1965); Ringle/Smith-Stark (1996); Thompson (1962); Zimmermann (1956).

oder diastratisch bedingte Varietäten oder Schrift- und Sprachstadien zulieben, wurden bislang nicht unternommen. Veränderungen der Zeichenformen, des Zeichenbestands, allomorphe Schreibungen oder die Kriterien, nach denen Zeichen im Block angeordnet wurden, sind bisher unzureichend beschrieben und analysiert. Forschungslücken sind auch auf dem Gebiet der Paläografie evident: Bislang liegt lediglich die Arbeit von Alfonso Lacadena vor (vgl. Lacadena García-Gallo 1995), die ausgewählte Beispiele aus dem Textkorpus behandelt. Eine umfassende paläografische Zeichenliste, wie sie etwa für das Sumerische von René Labat vorliegt (vgl. Labat 1994), stellt ein dringendes Forschungsdesiderat dar.

Das Forschungsprojekt wird diesem Umstand Rechnung tragen und neben Lexikografie und Korpuslinguistik auch graphetische und graphemische Fragestellungen bei der Konzipierung und Entwicklung der Virtuellen Forschungsumgebung berücksichtigen. Zu diesem Zweck wird die originale Schreibung der Hieroglyphen in die Forschungsumgebung von TextGrid eingepflegt und nach XML/TEI ausgezeichnet. Das Projekt wird dabei auf die Standards von EpiDoc zurückgreifen, die ein strukturiertes Markup von epigraphischen Dokumenten in XML/TEI ermöglichen und bereits in zahlreichen epigraphischen Projekten zur Anwendung kommen.¹⁰ So erlaubt das Auszeichnungsformat etwa, die Position von Graphemen im Block und deren Ausrichtung und Größe im Verhältnis zu den übrigen Zeichen im Hieroglyphenblock zu beschreiben, Textsegmente wie etwa Phrasen, Eigennamen oder Kalenderinformationen zu markieren oder die Relation von Text und Bild festzulegen, sodass bei einer Durchsuchung des Textmaterials die Möglichkeit besteht, Zeichen zu suchen, die um 90° gedreht und sich nur in Textbestandteilen befinden, die auf Körpern angebracht und graviert statt im Relief gearbeitet sind. Verteilung und zeitliche Dimensionen von Schrift und Sprache lassen sich durch das XML-basierte Markup ebenso gut bestimmen wie etwa auch die Anzahl von lesbaren oder auch entzifferten Zeichen im Text oder das Verhältnis von Kopf- und Standardvarianten in einem wählbaren Zeit- oder Verwendungszeitraum. Die Markierung von Zeichenfunktionen mithilfe von XML erlaubt es, orthografische Konventionen zu identifizieren und die Bedeutung vokalharmonischer oder -disharmonischer Schreibung zu erschließen.

¹⁰ <http://steinheim-institut.de/cgi-bin/epidat>

Indem der Kontext der Maya-Inschriften und ihre Metadaten ebenfalls datenbanktechnisch erfasst und mithilfe dieser Koordinaten ein Netzwerk an inschriftlichen und textlichen Informationen entsteht, wird das Projekt zudem die Frage beantworten, in welchem Kontext orthografische Konventionen entstanden oder verwendet wurden. Der Vorteil der digitalen Epigraphik liegt dabei darin, zeiträumliche und soziokulturelle Einflüsse und Dynamiken der Schriftkultur erfassen und auswerten zu können. Dies wäre mit herkömmlichen Methoden wie Zettelkasten und Papierarchiv nicht möglich.

Eine weitere Herausforderung der Maya-Inschriften- und -Sprachforschung besteht darin, dass bislang noch keine umfassende und detaillierte Sprachbeschreibung des Klassischen Maya existiert (vgl. Wichmann 2006). Begründet liegt dies im Fehlen grundlegender Kenntnisse über die Phonologie, über morphophonemische Prozesse, über den Bereich der Verbmorphologie und ihr Tempus-/Aspekt-/Modus-System, über die Bedeutung von Adverbien und Deixis. Auch diese Detailarbeit kann nur mithilfe korpuslinguistischer Methoden in Angriff genommen werden, die wiederum das Vorliegen maschinenlesbarer Texte zur Voraussetzung haben. Aus diesem Grund kann die Mayaforschung wichtige Entwicklungen innerhalb der Mayasprachen – wie etwa die Entstehung von aspektorientierter Teilergativität in verschiedenen Mayasprachen – noch nicht vollständig nachvollziehen. Diskursgrammatische Untersuchungen sind bislang erst an wenigen Texten unternommen worden (vgl. Josserand 1991). Auch die literarischen Genres und ihre sprachlichen Manifestationen sind nur exemplarisch erforscht worden (vgl. Hull/Carrasco 2012). Das Verhältnis von gesprochenen Sprachen zur Schriftsprache und die Prozesse der Innovation, Etablierung und Transformation innerhalb des Klassischen Maya sollten in Relation zu historischen, sozialen und kulturellen Prozessen analysiert werden. Der Eindruck, dass das Schriftsystem im Tiefland zu einem bestimmten Zeitpunkt weitestgehend homogen gewesen sei, basiert häufig auf punktuellen Untersuchungen und müsste im Rahmen umfangreicher Textanalysen verifiziert oder falsifiziert werden. Auch steht die Rekonstruktion der Sprachgeografie und Soziolektik des Klassischen Mayatieflands noch in ihren Anfängen, da umfassende Studien aufgrund eines fehlenden Lexikons des Klassischen Maya bislang noch nicht durchgeführt werden konnten.

Grundlagen dieser Untersuchung finden sich methodisch im Bereich der Korpuslinguistik wieder, die Äußerungen natürlicher Sprache, ihrer Elemente und Strukturen auf der Grundlage von Analysen authentischer Texte beschreibt (vgl. Lemnitzer/Zinsmeister 2006: 9). Korpusuntersuchungen arbei-

ten mit digitalen Textsammlungen, die mithilfe von Kookkurrenz- und Kollokationsanalysen ausgewertet und in KWIC- (*Keyword in Context*) und Volltext-Ansicht dargestellt werden. Dieses Verfahren kam in der Maya-Inschriftenforschung vor dem Computer-Zeitalter bereits zur Anwendung (vgl. Gates 1931; Zimmermann 1956), wobei die damals noch unlesbaren Zeichen mithilfe von Zahlencodes repräsentiert und die Kontexte auf Karteikarten gesammelt wurden. Ab 1961 kamen auch Computer zum Einsatz und unterstützten Forscher bei der Entzifferungsarbeit,¹¹ wobei diese Methode später kaum Widerhall gefunden hat und erst jetzt im Rahmen des Akademieprojekts intensiv zur Anwendung kommen wird. Während die bisherigen Projekte ausschließlich mit numerischen Codes für die Repräsentation der Schriftzeichen operiert haben, wird das Akademieprojekt die Originalschreibung der Texte in KWIC einbinden können und damit der Korpuslinguistik eine visuelle Repräsentationsform geben können: Bei der Korpusanalyse werden die Abfrageergebnisse in KWIC-Ansicht zeilenweise mit ihrem originalen Ko(n)text dargestellt, wobei „der Suchbegriff in der Mitte einer Textzeile dargestellt und üblicherweise grafisch hervorgehoben“ (Scherer 2006: 43) wird. Auf diese Weise lassen sich Kookkurrenzen und Kollokationen bestimmen, auf deren Basis Grammatik, Syntax und Semantik des Klassischen Maya in ihrem jeweiligen Kontext (Wort, Satz, Text, Inschrift) analysiert werden können. Kookkurrenzen bezeichnet dabei das „gemeinsame Vorkommen zweier oder mehrerer Wörter in einem Kontext von fest definierter Größe [...]“ (Lemmitzer/Zinsmeister 2006: 197).

Überträgt man diese Methode auf die Maya-Epigraphie, so lassen sich im Fall unentzifferter Textpassagen bedeutungsvolle Zeichenkombinationen abfragen und Querverweise zu anderen Texten mit vergleichbaren Kombinationen herstellen, deren Bedeutung im Gegensatz zum Ausgangstext schon erschlossen ist. Mithilfe bereits lesbarer Texte lassen sich systematisch Rückschlüsse auf die semantische und sprachliche Bedeutung unentzifferter Textpassagen ziehen und so die Schrift entziffern. Von Bedeutung für die Untersuchung des Lexikons des Klassischen Maya, aber auch bei der Analyse von Genre und Diskurs, ist das Aufdecken von Kollokationen, mit der „das wiederholte gemeinsame Vorkommen zweier [oder mehrerer] Wörter in einer strukturell interessanten Einheit bezeichnet [werden]. In einer Kolloka-

11 Vgl. Escalante Hernández (1971); Evreinov/Kosarev/Ustinov (1961); Riese (1971); Ringle/Smith-Stark (1996).

tion beeinflusst ein Wort die Auswahl eines anderen Wortes zuungunsten von Wörtern mit gleicher oder ähnlicher Bedeutung“ (ebd.: 196).

Korpuslinguistik und ihre Methoden sind daher von großer Bedeutung für die Untersuchung des Lexikons des Klassischen Maya, das bislang nur in Wortlisten und unvollständigen Wörterbüchern erfasst wurde (vgl. Boot 2002, 2009; Mathews/Bíró 2006 ff.; Montgomery 2002). Die bisher publizierten Verzeichnisse sind in Bezug auf das Ausgangsmaterial unvollständig und fehlerhaft; sie enthalten falsche Entzifferungen, Einträge sind aus ihrem Kontext herausgelöst und müssen eigenhändig in den Originalquellen referiert werden; sprachliche Lesungen, die Etymologie der Wörter oder grammatische Erklärungen bleiben häufig unbegründet oder werden nicht erläutert. Das entscheidende Defizit ist jedoch das Fehlen eines Vorkommensnachweises, der sämtliche Textquellen erfasst und dabei auch jene Textteile in ihrer originalen hieroglyphischen Schreibung und alphanumerischer Transkription einschließt, die bislang noch nicht gelesen werden können. In den bisher veröffentlichten katalogartigen Bestandsaufnahmen von hieroglyphischen Ausdrücken fehlen morphologische und grammatikalische Analysen der gelesenen hieroglyphischen Ausdrücke, z.T. zeiträumliche Referenzen, Erläuterungen zum kulturellen Kontext, Begründungen zur Übersetzung sowie bibliografische Angaben zu Vorarbeiten.

Zusammenfassend können wir festhalten, dass trotz digitaler Präsentationsform einiger Wörterbücher (vgl. Mathews/Bíró 2006 ff.) Varietäten und Entwicklungen von Lexikon, Grammatik und Schrift nicht fassbar werden. Ebenso wenig kann der Kontext aufgerufen werden – die Bedeutung von Wörtern liegt in deren Verwendung, die Offenlegung des Verwendungskontexts ist Grundbedingung für die Entzifferung der Mayaschrift und ihrer Zusammenfassung in einem Wörterbuch. Die zur umfassenden Erforschung von Schrift und Sprache erforderliche Interaktivität zwischen Textkorpus und analytischem Wörterverzeichnis findet sich in keinem der bislang publizierten Wörterbücher oder –listen und kann nur mithilfe digitaler Epigraphik in Angriff genommen werden. Hier liegt in den kommenden Jahren die Aufgabe des Textdatenbank- und Wörterbuchprojekts zum Klassischen Maya.

Projektziele im Blickfeld digitaler Epigraphik

Die im vorangegangenen Kapitel erörterten Forschungslücken und Herausforderungen lassen sich zusammenfassend damit begründen, dass das Quellen-

material unvollständig dokumentiert und digital (noch) nicht erschlossen ist. Ziel des Akademievorhabens ist es daher, das digitale Wörterbuch des Klassischen Maya auf der Basis eines EDV-gestützten Inschriftenkorpus, das nach dem Vorbild des digital zugänglichen *Thesaurus Linguae Aegyptiae* (TLA)¹² in Berlin, dem *Chicago Assyrian Dictionary* (CAD)¹³ und dem *Pennsylvania Sumerian Dictionary* (PSD)¹⁴ erstellt und der Forschung im Internet zugänglich gemacht wird. Das Klassische Maya nimmt in dieser Reihe eine besondere Stellung ein, da die Schriftentzifferung im Gegensatz zum Akkadischen, Sumerischen oder dem Altägyptischen noch nicht abgeschlossen ist. Daher wird das Projekt diesem Umstand besondere Rechnung tragen. Das geplante Wörterbuch arbeitet mit Visualisierungen und beruht daher auf der originalen hieroglyphischen Schreibung, der Umschrift und Übersetzung der Hieroglyphentexte. Grundlage des Wörterbuchs ist das EDV-gestützte Textkorpus, das alle bekannten Inschriftenträger für die computergestützte Recherche bereitstellt und komplexe Recherchen erlaubt.

Auf diese Weise kann eine lexikalische Datenbank eingerichtet werden, auf deren Grundlage das Wörterbuch erstellt wird. Das Textkorpus wird in TextGrid so angelegt, dass es regelmäßig erweitert und aktualisiert werden kann. Neue Texte können eingebunden, Entwicklungen in der Schriftforschung können verfolgt, Revisionen ergänzt und aktuelle Interpretationen und Beschreibungen integriert werden. Diese Vorgehensweise erlaubt es, alle Schrift- und Sprachdaten einschließlich unentzifferter Textstellen zu berücksichtigen, Zitate der jeweiligen Textstellen in originaler hieroglyphischer Schreibung abzubilden, zeitliche und räumliche Angaben einzubinden, epigraphische und sprachwissenschaftliche Analysen durchzuführen sowie kulturhistorische Kommentierungen und Verweise auf Sekundärliteratur einzubinden. Aufgrund der umfangreichen Datenmenge ist ein korpusbasiertes Wörterbuchprojekt und somit eine weitergehende Schrift-Entzifferung nur mittels der datenbanktechnischen Erfassung der Texte und mithilfe computergestützter Konkordanz- und Kollokationsanalyse sowie Textmustererkennung möglich.

12 <http://aew.bbaw.de/tla/>

13 <http://oi.uchicago.edu/research/publications/assyrian-dictionary-oriental-institute-university-chicago-cad>

14 <http://psd.museum.upenn.edu/epsd1/index.html>

Literaturverzeichnis

- Boot, Erik (2002): A Preliminary Classic Maya – English, English – Classic Maya Vocabulary of Hieroglyphic Readings. <http://www.mesoweb.com/resources/vocabulary/vocabulary.pdf>.
- Boot, Erik (2009): The Updated Preliminary Classic Maya – English, English – Classic Maya Vocabulary of Hieroglyphic Readings. <http://www.mesoweb.com/resources/vocabulary/Vocabulary-2009.01.pdf>.
- Coe, Michael D. (1992): *Breaking the Maya Code*. New York: Thames & Hudson.
- Eggebrecht, Eva; Arne Eggebrecht; Grube, Nikolai (Hrsg.) (1992): *Die Welt der Maya*. Mainz: Philipp von Zabern.
- Escalante Hernández, Roberto (1971): *Análisis de estructuras en el Códice de Dresde* (Centro de Estudios Mayas; Cuaderno 4). México, D.F.: UNAM.
- Evreinov, Eduard V.; Kosarev, Ju. G.; Ustinov, V. A. (1961): *Primenenie elektronnyh vyčislitel'nyh mašin v issledovanii pis'mennosti drevnich majja* (= *Die Anwendung elektronischer Rechenmaschinen bei der Erforschung des Mayaschrifttums*). Novosibirsk: Izd. Sibirskoe Otdelenija Akademia Nauk SSSR.
- Fahsen, Federico; Nikolai Grube (2005): The Origins of Maya Writing. In: Fields, Virginia M.; Reents-Budet, Dorie (Hrsg.): *Lords of Creation: The Origins of Sacred Maya Kingship*. London; Los Angeles, CA: Scala Publishers; Los Angeles County Museum, S. 74–79.
- Freidel, David; Schele, Linda; Parker, Joy (1993): *Maya Cosmos: Three Thousand Years on the Shaman's Path*. New York: William Morrow & Co.
- Gates, William E. (1931): *An Outline Dictionary of Maya Glyphs: With a Concordance and Analysis of Their Relationships* (Maya Society Publication; 1). Baltimore: Maya Society.
- Graham, Ian (1975 ff.): *Corpus of Maya Hieroglyphic Inscriptions*. Cambridge: Peabody Museum of Archaeology and Ethnology, Harvard University.
- Grube, Nikolai (1990): *Die Entwicklung der Mayaschrift: Grundlagen zur Erforschung des Wandels der Mayaschrift von der Protoklassik bis zur spanischen Eroberung* (Acta Mesoamericana; 3). Berlin: Karl-Friedrich von Flemming.
- Grube, Nikolai (2010): Das Gottkönigtum bei den Klassischen Maya. In: Assmann, Jan; Strohm, Harald (Hrsg.): *Herrscherkult und Heilserwartung*. München: Wilhelm Fink, S. 19–47.
- Grube, Nikolai (2012): *Der Dresdner Maya-Kalender: Der vollständige Codex*. Freiburg/Basel/Wien: Herder.

- Hermann Lejarazu, Manuel A. (2007): Símbolos de poder: un análisis comparativo entre la iconografía del Clásico maya y los códices mixtecos. In: *Estudios de Cultura Maya* 30: 79–106.
- Houston, Stephen D.; Chinchilla Mazariegos, Oswaldo; Stuart, David (Hrsg.) (2001): *The Decipherment of Ancient Maya Writing*. Norman: University of Oklahoma Press.
- Houston, Stephen D.; Stuart, David; Robertson, John (1998): Disharmony in Maya Hieroglyphic Writing: Linguistic Change and Continuity in Classic Society. In: Ciudad Ruiz, Andrés et al. (Hrsg.): *Anatomía de una civilización: Aproximaciones interdisciplinarias a la cultura maya*. Madrid: Sociedad Española de Estudios Mayas, S. 275–296.
- Houston, Stephen D.; Stuart, David (1992): On Maya Hieroglyphic Literacy. In: *Current Anthropology* 33 (5): 589–593.
- Hull, Kerry M.; Carrasco, Michael (2012): *Parallel Worlds: Genre, Discourse, and Poetics in Contemporary, Colonial, and Classic Period Maya Literature*. Boulder: University Press of Colorado.
- Josserand, J. Kathryn (1991): The Narrative Structure of Hieroglyphic Texts at Palenque. In: Fields, Virginia M. (Hrsg.): *Sixth Palenque Round Table, 1986*. Norman & London: University of Oklahoma Press, S. 12–31.
- Kerr, Justin (1989 ff.): *The Maya Vase Book: A Corpus of Rollout Photographs of Maya Vases*. New York: Kerr Associates.
- Knorozov, Yuri V. (1963): *Pis'mennost' indeitsev Maiia (= Schrift der Mayaindianer)*. Moskva-Leningrad: Akademia Nauk SSSR.
- Knorozov, Yuri V. et al. (1999): *Compendio Xcaret de la escritura jeroglífica descifrado por Yuri V. Knórosov*. Chetumal: Universidad de Quintana Roo/Promotora Xcaret.
- Labat, René (1994): *Manuel d'épigraphie akkadienne : signes, syllabaire, idéogrammes*. Paris: Paul Geuthner S.A.
- Lacadena García-Gallo, Alfonso (1995): *Evolución formal de las grafías escriturarias mayas: Implicaciones históricas y culturales*. PhD Thesis, Universidad Complutense de Madrid, Madrid.
- Lacadena García-Gallo, Alfonso; Wichmann, Søren (2004): On the Representation of the Glottal Stop in Maya Writing. In: Wichmann, Søren (Hrsg.): *The Linguistics of Maya Writing*. Salt Lake City: The University of Utah Press, S. 100–164.
- Lemnitzer, Lothar; Zinsmeister, Heike (2006): *Korpuslinguistik: Eine Einführung*. Tübingen: Gunter Narr Verlag.

- Macri, Martha J.; Lopper, Matthew G. (2003): *The New Catalog of Maya Hieroglyphs – Volume 1: The Classic Period Inscriptions* (The Civilization of the American Indian Series; 247). Norman: University of Oklahoma Press.
- Macri, Martha J.; Vail, Gabrielle (2009): *The New Catalog of Maya Hieroglyphs – Volume 2: The Codical Texts* (The Civilization of the American Indian Series; 264). Norman: University of Oklahoma Press.
- Mathews, Peter; Bíró, Péter (2006–). *Maya Hieroglyph Dictionary*, http://research.famsi.org/mdp/mdp_index.php.
- Mayer, Karl Herbert (1978 ff.): *Maya Monuments: Sculptures of Unknown Provenance*. Markt Schwaben: Saurwein.
- Montgomery, John (2002): *Dictionary of Maya Hieroglyphs*. New York: Hippocrene Books.
- Neuroth, Heike; Lohmeier, Felix (2011): TextGrid – Virtual Research Environment for the Humanities. In: *The International Journal of Digital Curation* 6 (2): 222 bis 231.
- Prager, Christian M. (2013): *Übernatürliche Akteure in der Klassischen Maya-Religion: Eine Untersuchung zu intrakultureller Variation und Stabilität am Beispiel des k'uh „Götter“-Konzepts in den religiösen Vorstellungen und Überzeugungen Klassischer Maya-Eliten (250–900 n. Chr.)*. Diss., Rheinische Friedrich-Wilhelms-Universität Bonn. <http://hss.ulb.uni-bonn.de/2014/3539/3539.htm>.
- Rendón M., Juan José; Spescha, Amalia (1965): Nueva clasificación «plástica» de los glifos mayas. In: *Estudios de Cultura Maya*, 5, S. 189–244.
- Riese, Berthold (1971): *Grundlagen zur Entzifferung der Mayahieroglyphen: Dargestellt an den Inschriften von Copan* (Beiträge zur mittelamerikanischen Völkerkunde; 11). Hamburg/München: Hamburgisches Museum für Völkerkunde und Vorgeschichte/Kommissionsverlag Klaus Renner.
- Ringle, William M.; Smith-Stark, Thomas C. (1996): *A Concordance to the Inscriptions of Palenque, Chiapas, Mexico* (Middle American Research Institute; 62). Tulane: Middle American Research Institute, Tulane University.
- Schele, Linda (1991): *Further Adventures with T128 Ch'a* (Copán Note; 9). Austin: Copán Acropolis Project & Instituto Hondureño de Antropología e Historia.
- Scherer, Carmen (2006): *Korpuslinguistik*. Heidelberg: Universitätsverlag Winter.
- Stuart, David (1995): *A Study of Maya Inscriptions*. PhD Thesis, Vanderbilt University, Nashville.
- Stuart, David (2009): Unusual Signs 2: The “Fringed Crossed-Bands” Logogram. In: *Maya Decipherment: A Weblog Devoted to the Ancient Maya Script*. <https://de->

cipherment.wordpress.com/2009/09/16/unusual-signs-2-the-fringed-crossed-bands-logogram/.

- Thompson, J. Eric S. (1962): *A Catalog of Maya Hieroglyphs* (The Civilization of the American Indian Series; 62). Norman: University of Oklahoma Press.
- Vail, Gabrielle; Bricker, Victoria R. (2003): New Perspectives on the Madrid Codex. In: *Current Anthropology* 44: 105–112.
- Wichmann, Søren (2006): Mayan Historical Linguistics and Epigraphy: A New Synthesis. In: *Annual Review of Anthropology* 35: 279–294.
- Zender, Marc (1999): *Diacritical Marks and Underspelling in the Classic Maya Script: Implications for Decipherment*. Master's thesis, University of Calgary, Alberta.
- Zimmermann, Günter (1956): *Die Hieroglyphen der Mayahandschriften* (Abhandlungen aus dem Gebiet der Auslandskunde; 62 – Reihe B, Völkerkunde, Kulturgeschichte und Sprachen; 34). Hamburg: Cram, de Gruyter & Co.

Die *Bibliothek der Neologie* (BdN)

Einblicke in ein aufklärungstheologisches Editionsprojekt

von *Bastian Lemitz*

in Zusammenarbeit mit *Larissa Figgen, Lena Goormann, Johannes Huck, Lars-Steffen Meier, Olga Söntgerath und Marco Stallmann*

Theologiegeschichtliche Verortung

Dem Editionsprojekt *Bibliothek der Neologie. Kommentierte kritische Auswahl-edition in zehn Bänden* (BdN) liegt ein kirchen- und theologiegeschichtliches Paradoxon zugrunde. Während die Aufklärungszeit in philosophie-, literatur- und allgemeinhistorischer Perspektive längst als bedeutendes Forschungsfeld erkannt ist, muss die Aufklärung in kirchen- und theologiegeschichtlicher Perspektive – obgleich sich die Situation innerhalb der letzten beiden Jahrzehnte durch Spezialstudien bzw. Tagungs- und Sammelbände merklich gebessert hat (vgl. Beutel 2009: 28–33) – noch immer als ein „Stiefkind der evangelischen Kirchengeschichtsschreibung“ (Schultze 1991: 61)¹ gelten.² Diese Situation widerspricht dem sich immer stärker durchsetzenden Bewusstsein, dass insbesondere der Protestantismus³ in wechselseitiger Beziehung mit der Aufklärung als „intellektuelle[r] Hauptbewegung des 18. Jahrhunderts“ (Burkhardt 1988: 106) einen fundamentalen Transforma-

1 Im Anschluss an Goethes legendarisches Sterbewort „Mehr Licht!“ formuliert Beutel (1996: 1), innerhalb der Kirchen- und Theologiegeschichte liege auf der Aufklärung ein „Schatten“.

2 Vgl. die ernüchternde Forschungsbilanz in Nowak (1999).

3 Zwar hat die Aufklärung konfessionsübergreifend auf Theologie und Kirche ausgestrahlt, christentumsgeschichtlich ist jedoch festzuhalten, dass sich die theologische Aufklärung früher und richtungsweisender auf protestantischer Seite vollzogen hat (vgl. Beutel 2009: 170–181). Zur katholischen Perspektive vgl. Klutening (1993) und Lehner/Printy (2010), verwiesen sei schon jetzt auf Wolf (2015). Einen Überblick über die jüdische Aufklärung (*Haskalah*) vermittelt Schuhmacher-Brunhes (2010).

tionsprozess durchlaufen hat,⁴ der den Übergang vom Alt- zum Neuprottestantismus markiert und Theologie und Kirche bis heute maßgeblich prägt.⁵

Zu den wesentlichen Problemen, vor die die theologische Forschung gestellt ist, zählt nicht zuletzt eine insgesamt als desolat einzuschätzende Quellsituation. Zwar stehen für einzelne Protagonisten der theologischen Aufklärung neben digitalen Zugriffsmöglichkeiten (etwa im Rahmen von VD 18⁶ u. Ä.) auch Nachdrucke bereit, doch fehlt es an umfassend angelegten und modernen editionsphilologischen Standards genügenden kritischen Quellenausgaben, um die positionelle und literarische Bandbreite der Aufklärungstheologie abzubilden. Hier soll die *Bibliothek der Neologie* (BdN) abhelfen.

Da sich die Aufklärung als gesamteuropäisches und in sich äußerst differenziertes Phänomen darstellt, ist eine sachgemäße Begrenzung von Texten und Autoren notwendig, die gleichzeitig jedoch als repräsentativ für die theologische Aufklärung gelten können. Betrachtet man die deutsche Aufklärungstheologie genauer, so werden Entwicklungslinien sichtbar, die eine weitere Zergliederung des betreffenden Zeitraumes erlauben. Danach spannt sich ein Bogen, der von den im 17. Jahrhundert einsetzenden Entwürfen der Physikotheologie bis zu den spätrationalistischen Positionierungen in der ersten Hälfte des 19. Jahrhunderts reicht (vgl. Beutel 2009: 89–169; Gericke 1989: 77–120).⁷ Als „konzeptionell innovativste Phase“ (Barth 2013: 107)

4 Hier ist v. a. Emanuel Hirsch zu nennen, der die in seiner Quellensammlung (Hirsch 1938) dokumentierten Transformationsprozesse in seinem in mehreren Auflagen erschienenen Hauptwerk (Hirsch 1949–1954) umfassend rekonstruiert hat (vgl. dazu Müller 1984; Beutel/Leppin 2004).

5 An dieser Stelle sei allein auf Wagner (1999) verwiesen.

6 Zum *Verzeichnis der im deutschen Sprachraum erschienenen Drucke des 18. Jahrhunderts* (VD 18) vgl. <http://www.vd18.de>.

7 Barth (2013) hält es für die Vergegenwärtigung der protestantischen Aufklärungstheologie für zweckmäßig, von dem Dreierschema Übergangstheologie (1720–1750) – Neologie (1720–1790) – theologischer Rationalismus (1790–1820) auszugehen, datiert den Beginn der Neologie, wie auch den Beginn der Übergangstheologie, jedoch in das Jahr 1720 (vgl. ebd.: 107). Beutel (2009: 115) hält es dagegen für „präziser und [...] aussagekräftiger, den materialen Auftakt in das Jahr 1748 zu datieren“, dem Erscheinungsjahr zweier „epochenprägende[r] Hauptwerke der Neologie“, Spaldings *Betrachtung über die Bestimmung des Menschen* sowie A. F. W. Sacks *Vertheidigter Glaube des Christentums*. Konsens besteht darüber, dass der Niedergang der Neologie zeitlich mit der Geltungsdauer des 1788 erlassenen Woellnerschen Religionsedikts zusammenfällt.

muss jedoch die Neologie (vgl. dazu Beutel 2009: 112–151)⁸ gelten, die vor diesem Hintergrund mit gutem Recht als „eigentliche Periode“ (Maurer 1999: 33) bzw. „reife Gestalt“ (Beutel 1998: 945; vgl. Beutel 2014: III) der deutschen Aufklärungstheologie bezeichnet werden kann.

Neben ihrer immensen theologischen Bedeutung ist für die Neologie editionsstrategisch zusätzlich ins Feld zu führen, dass sie zum einen mit der zweiten Hälfte des 18. Jahrhunderts nur wenige Jahrzehnte umfasst und sich zum anderen nur auf wenige maßgebliche Theologen (ohne dass im engeren Sinne von einem Schulhaupt gesprochen werden kann) und Zentren innerhalb des deutschen Sprachraumes (v. a. die Universitätsstädte Halle, Jena, Leipzig, Göttingen und Frankfurt/Oder sowie Metropolen wie etwa Hamburg, Braunschweig und v. a. Berlin) beschränkt (vgl. Beutel 2009: 112–150). Trotz dieser zeitlichen, personellen und geografischen Überschaubarkeit hat die mit der Neologie einhergehende Transformation des protestantischen Christentums nachhaltig auf nahezu alle Bereiche des öffentlichen und kulturellen Lebens ausgestrahlt. Ihre Wirkungen und Impulse lassen sich in der zeitgenössischen Literatur ebenso nachweisen wie in der Politik, Philosophie, Kunst oder Architektur. Und in gesamteuropäischer Perspektive ist zu bemerken, dass sich die Aufklärungstheologie im Allgemeinen und die Neologie im Besonderen ganz wesentlich aus der westeuropäischen Geistes- und Religionsgeschichte speist und auch nach ihrer eigentlichen Manifestationsphase richtungsweisend in diese zurückgewirkt hat.

Mit der *Bibliothek der Neologie* soll diese äußerst vitale und wegweisende Periode protestantischer Theologiegeschichte im Rahmen einer kritischen Quellenedition in ihrer thematischen, gattungsspezifischen und positionellen Bandbreite abgebildet und so weiterführende, nicht zuletzt auch interdisziplinär angelegte Forschung im Bereich der Aufklärungstheologie stimuliert werden.

Projektskizze

Die *Bibliothek der Neologie* entsteht in enger Zusammenarbeit des Lehrstuhls für Kirchengeschichte II (Reformation, Neuere und Neueste Kirchengeschichte) der Evangelisch-Theologischen Fakultät der Westfälischen Wil-

⁸ Als profilierteste Monografie speziell zur Neologie ist noch immer Aner (1929) zu nennen; eine konzise Darstellung findet sich in Pockrandt (2003: 5–11).

helms-Universität (WWU) Münster und der Abteilung Forschung und Entwicklung der Niedersächsischen Staats- und Universitätsbibliothek (SUB) Göttingen und ist ein von Prof. Dr. Albrecht Beutel (WWU Münster) initiiertes und gemeinsam mit Dr. Heike Neuroth (SUB Göttingen) beantragtes, DFG-gefördertes Langzeitprojekt, das auf zwölf Jahre angelegt ist und zum 1. April 2014 die Arbeit aufgenommen hat. Ziel der gemeinsamen Arbeit ist eine zehnbändige kritische Hybrid-Edition (Print- und Digitalversion) zentraler Werke der Neologie, wobei beide Projektpartner bereits in anderen Zusammenhängen ihre jeweiligen Kompetenzen unter Beweis gestellt haben.

Am Münsteraner Lehrstuhl für Kirchengeschichte II, zu dessen erklärten Schwerpunkten die Kirchen- und Theologiegeschichte der Aufklärungszeit gehört, ist mit der Kritischen Spalding-Ausgabe (Bde. 1–13, Tübingen 2001 bis 2013) in vorbildlicher Weise ein führender Neologe ediert und der Forschung zugänglich gemacht worden. In sachlicher Hinsicht, personell jedoch auch in Gestalt von Prof. Dr. Albrecht Beutel und Frau Olga Söntgerath, M.A., ist hier eine bewährte Kontinuität gewährleistet. Die Abteilung Forschung und Entwicklung gehört in Deutschland zu den größten ihrer Art und ist gemeinsam mit der ebenfalls an der SUB Göttingen angesiedelten Abteilung Metadaten und Datenkonversion eine der führenden Adressen für den Aufbau von virtuellen Forschungsumgebungen und Infrastrukturen für wissenschaftliche Daten und Dienste.

Innerhalb der *Bibliothek der Neologie* spiegeln sich diese Kompetenzbereiche in folgender Arbeitsteilung wider: Der umfassende informationstechnologische Support liegt, nicht zuletzt auch aufgrund einer hervorragenden Vernetzung im Bereich der Digital Humanities, auf Göttinger Seite. Zu den besonderen Schwerpunkten zählt die projektspezifische Anwendung und Modifikation des TEI-Schemas; anschließend werden die zur Edition vorgesehenen und mithilfe von Metadaten beschriebenen Texte im Rahmen eines Online-Portals aufbereitet und visualisiert. Dem Hybrid-Charakter entsprechend, wird auf Grundlage der so präparierten Daten gleichzeitig die Print-Version vorbereitet. Zur texttechnologischen Erschließung der Texte im XML-Format dient dabei die Virtuelle Forschungsumgebung TextGrid. Die Langzeitarchivierung der Forschungsdaten erfolgt im TextGrid Repository, sodass die Ergebnisse der *Bibliothek der Neologie* im Sinne des Open-

Access-Gedankens frühzeitig veröffentlicht und neben der Präsentation im Portal via SADE⁹ auch über das Repository nachgenutzt werden können.

In Ergänzung dazu erbringt die WWU Münster – neben der Erarbeitung editorischer Richtlinien, die im Online-Portal veröffentlicht, aber auch den einzelnen Bänden der *Bibliothek der Neologie* als technische Einleitung beigegeben werden – die fachwissenschaftliche Bearbeitung der Texte, zu der neben der Erstellung von forschungsrelevanten Registern (Personen- und Schlagwortregister, Bibelstellen, antike Autoren) auch die inhaltliche Kommentierung sowie eine theologiegeschichtliche Einleitung in das jeweilige Werk gehören.

Zu den wichtigsten Kommunikationsschnittstellen zwischen beiden Projektpartnern zählen regelmäßige Schulungen, in denen von Göttinger Seite theoretisch wie praktisch in den Umgang mit der Virtuellen Forschungsumgebung TextGrid eingeführt wird, sodass auf Münsteraner Seite neben der fachwissenschaftlich-theologiegeschichtlichen Verantwortung der Texte zusätzlich deren Auszeichnung nach XML/TEI-Standard, nicht zuletzt auch die Anreicherung mit Meta- und Normdaten, umgesetzt werden kann. Gleichzeitig lässt der rege Austausch über inhaltliche Fragen Probleme erkennbar werden, deren Lösung wertvollen Input bei der Anpassung und Weiterentwicklung der Virtuellen Forschungsumgebung, speziell im Hinblick auf geisteswissenschaftliche Editionsprojekte, liefert. Schließlich partizipieren beide Arbeitsstellen gemeinsam auch an den bundesweit stattfindenden TextGrid-Nutzertreffen.

Im Rahmen der *Bibliothek der Neologie* sind folgende Autoren respektive Texte zur Edition vorgesehen:

- I. C. F. Bahrdt / J. S. Semler: *Glaubensbekenntnisse*
1779 (24 S.) – 1779 (XIV + 119 S.) – 1779 (8 S.) – 1779 (56 S.) –
1780 (22 S.) – 1792 (XXIV + 382 S.)
- II. J. A. Ernesti: *Institutio interpretis Novi Testamenti*
¹1761 (VIII + 248 S.) – ⁵1809 (XXXVI + 452 S.)
- III. J. J. Griesbach: *Anleitung zum Studium der populären Dogmatik*
¹1779 (VIII + 128 S.) – ⁴1789 (XX + 255 S.)
- IV. G. Leß: *Sontags-Evangelia übersezt, erklärt, und zur Erbauung angewandt*
¹1776 (700 S.) – ³1781 (715 S.) + Zusätze (3. Aufl.) 1781 (75 S.)

⁹ SADE = Skalierbare Architektur für Digitale Editionen; vgl. <http://www.bbaw.de/teleta/software/sade/sade-1>

- V. A. H. Niemeyer: *Grundsätze der Erziehung und des Unterrichts*
¹1796 (512 S.) – ⁹1834 (XXII + 572 S.)
- VI. J. A. Nösselt: *Anweisung zur Bildung angehender Theologen*
¹1786/89 (VIII + 832 S.) – ³1818 / 19 (XLVIII + 806 S.)
- VII. A. F. W. Sack: *Vertheidigter Glaube der Christen*
¹1748–1753 (1131 S.); ²1773 (XXIV + 824 S.)
- VIII. G. S. Steinbart: *System der reinen Philosophie oder Glückseligkeitslehre des Christentums*
¹1778 (XVIII + 228 S.) – ⁴1794 (LII + 288 S.)
- IX. W. A. Teller: *Wörterbuch des Neuen Testaments zur Erklärung der christlichen Lehre*
¹1772 (XXVIII + 392 S.) – ⁶1805 (552 S.)
- X. J. G. Toellner: *Katechetischer Text, oder Unterricht vom christlichen Lehrbegriff für Unstudirte*
¹1765 (XXIV + 320 S.); ²1772 (XXIV + 340 S.)

Die Bandzahl innerhalb der *Bibliothek der Neologie* ergibt sich aus der alphabetischen Reihenfolge der Autoren, die Publikationsreihenfolge ist jedoch eine andere. Diese wird in der ersten Projektphase von Nösselts *Anweisung* (VI.) eröffnet, es folgen Griesbachs *Anleitung* (III.), Leß' *Sontags-Evangelia* (IV.) und Tellers *Wörterbuch* (IX.). Die übrigen Bände sind für die folgenden Projektphasen vorgesehen. Doch ist bereits jetzt abzusehen, dass jeder Band inhaltlich wie informationstechnologisch vor eigene Herausforderungen stellt.

Zur Editionspraxis

Allgemeines Vorgehen

Allen in die *Bibliothek der Neologie* aufgenommenen Texten ist gemeinsam, dass sie in gedruckter Form vorliegen und – mit Ausnahme von Ernestis *Institutio* (II.) – in deutscher Sprache verfasst sind. Dies mag im Vergleich zu Editionsprojekten, deren Quellengrundlage etwa aus Autographen besteht, von Vorteil sein, doch sind die Probleme, die sich aus den orthografischen, grammatischen und typografischen Besonderheiten des 18. Jahrhunderts ergeben, nicht zu unterschätzen.

Diese schlagen sich vor allem in der Ermittlung der Varianten nieder, die später unter Verwendung der bereits in der Kritischen Spalding-Ausgabe (s.o.) verwendeten Siglen unterhalb des Fließtextes vermerkt werden. Bei

diesem handelt es sich um eine einzelne Auflage, die aus sachlichen bzw. chronologischen Gründen als Leittext herangezogen wird und als Referenztext für die Ermittlung sämtlicher Varianten dient.¹⁰ Obgleich die Print-Edition jeweils einen Leittext bietet, lässt sich über den Nachweis sämtlicher Varianten die Textgestalt jeder einzelnen Auflage rekonstruieren, sodass eine vollumfängliche Zitierbarkeit gewährleistet ist. Online kann die Textgestalt jeder einzelnen Auflage als zusammenhängender Text ausgegeben werden.

Eine weitere Besonderheit besteht darin, dass die in der *Bibliothek der Neologie* zusammengefassten Texte – wiederum mit Ausnahme von Ernestis *Insitutio* (II.) – in Fraktur gedruckt sind. Antiqua ist allein lateinischen und modernfremdsprachlichen Begriffen und Passagen vorbehalten. Hinzu kommt – wie in theologischen Texten, zumal in solchen des 18. Jahrhunderts, zu erwarten – altgriechische und hebräische Schrift. Für die Editionsarbeit hat dies zur Folge, dass der zu edierende Text zunächst abgeschrieben werden muss, um einen elektronisch prozessierbaren Leittext zu erstellen.¹¹ Dieser wird dann in mehreren Korrekturrunden, sogenannten Leserunden, von Fehlern bereinigt.¹² Anschließend werden, ebenfalls über Leserunden, die

10 Für den Fall, dass gegenüber dem Leittext – entweder weil sie nicht im Leittext enthalten sind, oder aber, weil der Leittext derart viele Abweichungen enthält, dass diese aus Gründen der Übersichtlichkeit nicht einzeln im Apparat aufgeführt werden können – ganze Abschnitte geboten werden müssen, werden diese als Einschübe gekennzeichnet und dem Haupttext beigegeben.

11 Erste gute Ergebnisse im Hinblick auf Fraktur-OCR lassen jedoch darüber nachdenken, den Leittext elektronisch zu erstellen. Zwar müsste auch eine auf diese Weise erstellte Leitaufgabe mehrere Korrekturrunden durchlaufen, doch ließe sich im Hinblick auf das Abschreiben Arbeit und Zeit sparen. Anzudenken wäre dann auch, alle Auflagen eines Werkes auf diese Weise zu erfassen und die Kollationierung mittels Software (etwa das in TextGrid verfügbare Programm CollateX) durchzuführen. Auch wenn die Anwendung von Texterkennungs- und Kollationierungssoftware nicht bei allen Texten innerhalb der BdN zielführend erscheint, so ist doch zu prüfen, ob derartige Hilfsmittel nicht zumindest bei kurzen Werken mit geringer Auflagenzahl (absehbar wären dies Griesbachs *Anleitung* [III.] und Leß' *Sontags-Evangelia* [IV.]) mit Gewinn eingesetzt werden könnten.

12 Wie bereits in der Kritischen Spalding-Ausgabe (s.o.) erprobt, wird dies von zwei Projektmitarbeitern durchgeführt. Der erste Mitarbeiter liest laut aus dem Original vor, während der andere die Abschrift verfolgt und eventuelle Fehler korrigiert.

Varianten ermittelt¹³ und über die für die Arbeit mit einer Leitaufgabe besonders geeignete *double endpoint*-Methode inline ausgezeichnet.

Bereits der grobe Umriss der Editionspraxis lässt erkennen, dass die *Bibliothek der Neologie* nur mit großem Zeit- und Personalaufwand zu realisieren ist. An der WWU Münster sind die Aufgaben daher auf vier wissenschaftliche und drei studentische Mitarbeiter, in Göttingen auf zwei wissenschaftliche Mitarbeiter verteilt.

Ein abschließender Blick auf den ersten zu edierenden Band lässt die Art der Herausforderungen deutlich werden, vor die die *Bibliothek der Neologie* vor allem in informationstechnologischer Hinsicht gestellt ist.

Beispiel: J. A. Nösselts Anweisung (Halle ²1791)

Der Auftaktband der *Bibliothek der Neologie* – Johann August Nösselts *Anweisung zur Bildung angehender Theologen* (Halle ¹1786/1789 bis ³1818/1819) (VI.) – ist als ‚Leitfaden‘ für das Theologiestudium konzipiert¹⁴ und kann mit gutem Recht als „neologische Enzyklopädie“ (vgl. van Spankeren 2012: 155–238) bezeichnet und zu den einflussreichsten ihrer Art gezählt werden (vgl. ebd.: 237). Als Leittext dient die Zweitaufgabe (Halle ²1791), da sie als letzte zweifelsfrei und vollumfänglich von Nösselt besorgt worden ist. Die dritte Aufgabe (Halle ³1818/1819) firmiert unter dem Namen August Hermann Niemeyer, einem Schüler Nösselts, der den Text seines Lehrers etwas über zehn Jahre nach dessen Tod auf den neuesten Forschungsstand

13 Diese Art der Variantenermittlung ist ebenfalls in der Kritischen Spalding-Ausgabe (s.o.) erprobt. Im Falle der *Bestimmung des Menschen* wurden so immerhin elf Auflagen (¹1748–¹¹1794) parallel gelesen (vgl. J. J. Spalding [Ausg. 2006]), die BdN hat mit Niemeyers *Grundsätzen* (V.) maximal neun Auflagen gegenüberzustellen.

14 In der gegenüber der Zweitaufgabe umfangreicheren Vorrede zur Erstausgabe (Nösselt 1786/1789: III–XVI) formuliert Nösselt, er wolle mit der *Anweisung* „eine Anleitung [...] geben, was? worüber? warum? und wie man studieren sollte? um sich zu einem würdigen Lehrer der Religion zu bilden“, denn er habe vergeblich nach einem Buch gesucht (vgl. jedoch die in Nösselt 1791 I, S. 49–51 [§ 51] angeführten Titel), das ihm in diesen Fragen „zum Leitfaden diene, und den wirklichen Bedürfnissen unsrer Zeit, den grossen Fortschritten in den Wissenschaften, selbst in der Theologie, angemessen wäre“ (Nösselt 1786/1789: V). Weiter entfaltet wird dieses Anliegen in der Einleitung (vgl. Nösselt 1791 I, S. 1–56 [§§ 1–54]). Im Abschnitt zur Philosophie bemerkt Nösselt zudem, dass ihm an einer *historischen* (vgl. Nösselt 1791 I, S. 215 Anm. *), d. h. wohl nicht zuletzt auch im Sinne einer *objektiven*, Darstellung gelegen ist.

gebracht und mit teils umfangreichen Zusätzen versehen hat. Hier sind nicht nur orthografische, sondern auch zahlreiche inhaltliche Textveränderungen zu erwarten.

Die Hauptschwierigkeit der *Anweisung* ergibt sich jedoch aus ihrer Anlage. Eingebettet in eine systematische Zusammenordnung der theologischen Einzeldisziplinen, bietet Nösselt mehrere hundert bibliografische Angaben, die sich über das gesamte Gebiet der Theologie einschließlich der propädeutischen Disziplinen Philosophie und Philologie sowie der sog. Schönen Wissenschaften erstrecken und teils in den Fließtext integriert, teils in Form von Listen zusammengefasst sind. In einem ersten Schritt müssen diese bibliografischen Angaben einzeln nachrecherchiert und in einem zweiten Schritt möglichst gewinnbringend für das *Semantic Web* erschlossen werden. Hinsichtlich der Recherche stellt sich das Problem, dass nicht alle in der *Anweisung* genannten Werke – obwohl sich die Recherchemöglichkeiten im Hinblick auf das 18. Jahrhundert (hervorzuheben ist v. a. die Suche nach Zeitschriften bis hin zu einzelnen Beiträgen)¹⁵ ständig verbessern – eindeutig zu identifizieren sind. Dies liegt zum einen daran, dass Nösselts Angaben von der modernen bibliothekswissenschaftlichen Erfassung abweichen können, sodass hier wie dort von Fehlern ausgegangen werden muss;¹⁶ zum anderen sind Nösselts Angaben, ob beabsichtigt oder nicht, nicht selten unvollständig.

Wenn es etwa heißt, im Hinblick auf die Textkritik des Neuen Testaments seien „die *Millischen, Wetsteinischen, Griesbachischen*“¹⁷ Anmerkungen zu berücksichtigen, dann sind mit den Namen John Mill, Johann Jakob Wettstein und Johann Jakob Griesbach maßgebliche kritische Editionen des griechischen Neuen Testaments verbunden, die bis zum Erscheinen der *An-*

15 Verwiesen sei hier etwa auf die *Retrospektive Digitalisierung wissenschaftlicher Rezensionenorgane und Literaturzeitschriften des 18. und 19. Jahrhunderts aus dem deutschen Sprachraum* (vgl. <http://www.ub.uni-bielefeld.de/diglib/aufklaerung/>), den *Systematischen Index zu deutschsprachigen Rezensionenzeitschriften des 18. Jahrhunderts* (IdRZ 18) (vgl. <http://adw.sub.uni-goettingen.de/idrz/pages/Main.jsf>) und die Volltextverfilmung (Microfiche) *Deutsche Zeitschriften des 18. und 19. Jahrhunderts* (Hildesheim 1994–2000).

16 Im Falle Nösselts lassen sich die bibliografischen Angaben mit [Nösselt] (1810) – einer nach Nösselts Tod zu Verkaufszwecken gedruckten Aufstellung seiner umfangreichen Privatbibliothek – abgleichen. Allerdings lässt sich auf diese Weise einzig der Besitz, nicht aber die Benutzung der betreffenden Werke nachweisen.

17 Nösselt (1791) II, S. 47 (§ 35)

weisung in mehreren Auflagen vorlagen.¹⁸ Derartige Angaben lassen sich nach den *Functional Requirements for Bibliographic Records* (FRBR)¹⁹ weder auf eine konkrete Manifestation oder gar ein konkretes Exemplar bringen, noch kann ohne Einschränkung von einem Werk oder einer Expression gesprochen werden, da Mill, Wettstein und Griesbach in diesem Fall nicht als Autoren der neutestamentlichen Texte anzusprechen sind. Ein weiteres Problem besteht darin, dass Nösselt die bibliografischen Angaben häufig in einen Satz einbaut, sodass die betreffenden Titel zwar eindeutig zu verifizieren sind, in der von Nösselt gegebenen Gestalt jedoch nicht mit den benötigten Mark-Ups versehen werden können. So fährt Nösselt im Anschluss an die Aufzählung Mill, Wettstein und Griesbach (s. o.) fort, dass neben den Ausgaben des Neuen Testaments „auch des Letzern *Symbolas criticae* (Tom. prior. Halae 1785. 8.)“²⁰ mit Gewinn einzusehen seien. Zum einen ist an dieser Stelle Johann Jakob Griesbach als Autor zu ergänzen, zum anderen lautet der korrekte Titel *Symbolae criticae*. Der Akkusativ *Symbolas criticae* ergibt sich aus dem Satzzusammenhang. Dieses Beispiel zeigt also, dass gerade die häufig vorkommenden lateinischen Titel vor besondere Schwierigkeiten stellen.²¹ Und schließlich seien an dieser Stelle auch solche Titel erwähnt, die in der von Nösselt genannten Manifestation heute nicht mehr zu verifizieren sind, da sich keine der betreffenden Ausgaben im Besitz der zu Recherchezwecken konsultierten Bibliotheken²² befindet.

18 Vgl. auch die „*bossuet-cramerische* Einleitung, die *semmlerischen selecta Capita* [...] nebst der *schröckhischen* christl. Kirchengeschichte“ (Nösselt 1791 II, S. 136 f. [§ 103 Anm. 1]).

19 Vgl. <http://www.ifla.org/frbr-rg>.

20 Nösselt (1791) II, S. 48 (§ 35)

21 Ein weiteres Beispiel ist die bibliographische Angabe „*Ge. Henr. Nieupoort rituum, qui olim apud Romanos obtinuerunt, succincta explicatio*, Edit. 13. Berol. 1767 in gr. 8. auch Edit. 6. (Ultrajectina) curant. *Guil. Ottone et Jo. Freder. Reitzio* 1774 gr. 8.“ (Nösselt 1791 I, S. 163 [§ 142]). Hier werden die latinisierten Namen Wilhelm Otto und Johann Friedrich Reitz im Ablativ genannt. In diesen Fällen ist in die lateinische Nominalform bzw. in den deutschen Namen aufzulösen. Gleiches gilt von den latinisierten Ortsnamen (Berol. [Berolinum] = Berlin; Ultrajectina = Utrecht).

22 Für die Recherche werden v.a. die in VD 18 (vgl. Anm. 8) zusammengefassten Bibliotheken berücksichtigt. Nur in Einzelfällen werden die Titel über andere Bibliotheken nachgewiesen.

Um die bibliografischen Angaben dennoch so korrekt und präzise wie möglich zu erschließen, werden für die *Bibliothek der Neologie* unter Berücksichtigung des Personen- bzw. Autorennamens, des Titels, der Auflage und des Erscheinungsjahres projektspezifische Objekte angelegt und per URI mit dem Quelltext verknüpft. Auf diese Weise entsteht ein Verweisnetz nach dem *Resource Description Framework* (RDF), das den mit der *Anweisung* verbundenen bibliografischen Besonderheiten gerecht wird und zudem auch von anderen Projekten nachgenutzt werden kann.

Bereits der Einblick in die Arbeit an Johann August Nösselts *Anweisung* zur Bildung angehender Theologen zeigt, in welcher Weise sich die *Bibliothek der Neologie* mit den band-, aber auch zeitspezifischen Eigentümlichkeiten auseinanderzusetzen hat. Diesen sowohl fachwissenschaftlich als auch informationstechnologisch zu begegnen, wird das Hauptanliegen der *Bibliothek der Neologie* sein, die als theologisches Editionsprojekt in den kommenden zwölf Jahren ihren Beitrag zu den *Digital Humanities* leisten will.

Literaturverzeichnis

- Aner, Karl (1929): *Die Theologie der Lessingzeit*. Halle/Saale: Niemeyer.
- Barth, Ulrich (2013): Religion in der europäischen Aufklärung. In: ders.; Danz, Christian; Graf, Friedrich Wilhelm; Gräb, Wilhelm (Hrsg.): *Aufgeklärte Religion und ihre Probleme. Schleiermacher – Troeltsch – Tillich* (Theologische Bibliothek Töpelmann; 165). Berlin u.a.: De Gruyter, S. 91–112.
- Beutel, Albrecht (1996): *Lichtenberg und die Religion. Aspekte einer vielschichtigen Konstellation* (Beiträge zur Historischen Theologie; 93). Tübingen: Mohr Siebeck.
- Beutel, Albrecht (1998): Aufklärung II. Theologisch-kirchlich. In: Betz, Hans Dieter et al. (Hrsg.): *Religion in Geschichte und Gegenwart. Bd. 1*. 4. Aufl. Tübingen: Mohr Siebeck, S. 941–948.
- Beutel, Albrecht (2009): *Kirchengeschichte im Zeitalter der Aufklärung. Ein Kompendium*. Göttingen: Vandenhoeck & Ruprecht.
- Beutel, Albrecht (2014): *Johann Joachim Spalding. Meistertheologe im Zeitalter der Aufklärung*. Tübingen: Mohr Siebeck.
- Beutel, Albrecht; Leppin, Volker (Hrsg.) (2004): *Religion und Aufklärung. Studien zur neuzeitlichen „Umformung des Christlichen“* (Arbeiten zur Kirchen- und Theologiegeschichte; 14). Leipzig: Evangelische Verlagsanstalt.

- Burkhardt, Johannes (1988): Aufklärung A. Geschichtlich, In: Drehsen, Volker et al. (Hrsg.): *Wörterbuch des Christentums*. Gütersloh u.a.: Mohn u.a., S. 106–108.
- Gericke, Wolfgang (1989): *Theologie und Kirche im Zeitalter der Aufklärung* (Kirchengeschichte in Einzeldarstellungen; III/2). Berlin: Evangelische Verlagsanstalt.
- Hirsch, Emanuel (1938): *Die Umformung des christlichen Denkens in der Neuzeit. Ein Lesebuch*. Tübingen: Mohr Siebeck.
- Hirsch, Emanuel (1949–1954): *Geschichte der neuern evangelischen Theologie im Zusammenhang mit den allgemeinen Bewegungen des europäischen Denkens. Bde. 1–5*. Gütersloh: Bertelsmann.
- Klueting, Harm (Hrsg.) (1993): *Katholische Aufklärung – Aufklärung im katholischen Deutschland* (Studien zum achtzehnten Jahrhundert; 15). Hamburg: Meiner.
- Lehner, Ulrich L.; Printy, Michael (Hrsg.) (2010): *A Companion to the Catholic Enlightenment in Europe* (Brill's Companions to the Christian Tradition; 20). Leiden u.a.: Brill.
- Maurer, Michael (1999): *Kirche, Staat und Gesellschaft im 17. und 18. Jahrhundert* (Enzyklopädie deutscher Geschichte; 51). München: Oldenbourg.
- Müller, Hans Martin (1984): Emanuel Hirsch und die Umformungskrise des neuzeitlichen christlichen Denkens. In: ders. (Hrsg.): *Christliche Wahrheit und neuzeitliches Denken. Zu Emanuel Hirschs Leben und Werk* (Theologische Beiträge und Forschungen; 4). Tübingen: Katzmann u. a. S. 7–14.
- Nösselt, Johann August (1786/1789): *Anweisung zur Bildung angehender Theologen*. Halle: Joh. Jac. Curts Wittwe.
- Nösselt, Johann August (1791): *Anweisung zur Bildung angehender Theologen*. 2. Aufl., Halle: Joh. Jac. Curts Wittwe.
- [Nösselt, Johann August] (1810): *Bibliotheca Nösseltiana sive Catalogus librorum e quocunque eruditionis genere, theologico in primis atque literario, quos Ioannes Augustus Nösseltus, celeberrimus nuper in Academia Halensi theologus, sibi comparaverat*. Halle: Orphanotropheum.
- Nösselt, Johann August (1818/1819): *Anweisung zur Bildung angehender Theologen*. 3. Aufl., hrsg. v. August Hermann Niemeyer. Halle: Curtsche Buchhandlung.
- Nowak, Kurt (1999): *Vernünftiges Christentum? Über die Erforschung der Aufklärung in der evangelischen Theologie Deutschlands seit 1945* (Forum Theologische Literaturzeitung; 2). Leipzig: Evangelische Verlagsanstalt.

- Pockrandt, Mark (2003): *Biblische Aufklärung. Biographie und Theologie der Berliner Hofprediger August Friedrich Wilhelm Sack (1703–1786) und Friedrich Samuel Gottfried Sack (1738–1817)* (Arbeiten zur Kirchengeschichte; 86). Berlin: De Gruyter.
- Schuhmacher-Brunhes, Marie (2010): Aufklärung im jüdischen Stil. Die Haskalah-Bewegung in Europa. In: *Europäische Geschichte Online* [EGO], hrsg. v. Institut für Europäische Geschichte [IEG]. Mainz, <http://www.ieg-ego.eu/schuhmacher-brunhesm-2010-de>.
- Schultze, Harald (1991): Evangelische Frauen in der deutschen Aufklärung. Desiderate kirchengeschichtlicher Forschung. In: *Berliner Theologische Zeitschrift* 8: 59–75.
- Spalding, Johann Joachim (Ausg. 2006): *Die Bestimmung des Menschen* (¹1748 bis ¹¹1794). Hrsg. v. Beutel, Albrecht; Kirschkowski, Daniela; Prause, Dennis unter Mitarb. v. Look, Verena; Söntgerath, Olga (SpKA I/1). Tübingen: Mohr Siebeck.
- van Spankeren, Malte (2012): *Johann August Nösselt (1734–1807). Ein Theologe der Aufklärung* (Hallesche Forschungen; 31). Halle: Verlag der Franckeschen Stiftungen/Harrassowitz.
- Wagner, Falk (1999): *Metamorphosen des modernen Protestantismus*. Tübingen: Mohr Siebeck.
- Wolf, Hubert (2015): Katholische Aufklärung? In: Beutel, Albrecht; Nooke, Martha (Hrsg.): *Religion und Aufklärung. Akten des Ersten Internationalen Kongresses zur Erforschung der Aufklärungstheologie. Münster, 30. März bis 2. April 2014*. Tübingen: Mohr Siebeck [in Vorbereitung].

TextGrid lernt Noten

Digitale Musikeditionen in der VRE

von Joachim Veit

„Der Schlüssel zu den Noten war der Text“ – dieser kryptisch wirkende Satz kennzeichnet den Beginn unserer Beteiligung am TextGrid-Projekt in dessen im Juni 2009 startender zweiter Phase. Als Fotis Jannidis im Frühjahr des vorausgehenden Jahres im Auftrag des TextGrid-Konsortiums bei uns angefragt hatte, ob wir bereit wären, bei der geplanten Erweiterung der beteiligten Fachdisziplinen in TextGrid II die Musikwissenschaft zu vertreten, ging dies vermutlich auf die Aktivitäten bei der Vorbereitung unserer im Dezember 2007 in Paderborn durchgeführten Tagung „Digitale Edition zwischen Experiment und Standardisierung“ zurück, an der eine ganze Reihe von Germanisten teilgenommen hatten (vgl. Stadler/Veit 2009). Dort stand u.a. die Codierung von Briefeditionen mit TEI zur Diskussion – ein Format, das ich erst fünf Jahre zuvor bei einem Workshop von Fotis Jannidis und Roland Kamzelak während der Aachener Tagung der Arbeitsgemeinschaft für germanistische Edition kennengelernt hatte. Die Auseinandersetzung mit der Anwendbarkeit dieses Formats für Briefeditionen führte zur Gründung einer entsprechenden Arbeitsgruppe, die von Werner Wegstein (Würzburg) und Kurt Gärtner (Trier) ins Leben gerufen worden war und die nach den Debatten der Paderborner Tagung in eine *Special Interest Group Correspondence* der TEI überführt wurde. Nahezu gleichzeitig war eine Arbeitsgruppe für Musik-Codierungsformate gegründet worden und bei der Tagung 2007 standen die beiden Vertreter der führenden XML-basierten Musikcodierungsformate MusicXML (Michael Good) und des an TEI angelehnten MEI (Perry Roland) auf dem Podium. Rasch war nach dieser Veranstaltung klar geworden, dass das Format der *Music Encoding Initiative* MEI ein für wissenschaftliche Zwecke sehr viel geeigneteres sei, und in einem gemeinsamen Projekt von NEH und DFG wurde das von Perry Roland an der University of Virginia (Charlottesville) entwickelte MEI zunächst durch die Beteiligung von Johannes Kepper und Daniel Röwenstrunk vom Musikwissenschaftlichen Seminar Detmold/Paderborn weiter ausgearbeitet und dann rasch in ein Projekt der

musikwissenschaftlichen Community überführt, das weiterhin von Charlottesville und Detmold aus maßgeblich betreut wurde.

War es also anfänglich eher der Textstandard TEI, durch den Detmold/Paderborn Tuchfühlung zur „digitalen Szene“ aufnahm, so stand bald die Ausarbeitung und Verbreitung des Musikcodierungsstandards MEI gleichberechtigt daneben. Hinzu kam, dass seit 2003 am Musikwissenschaftlichen Seminar mit DFG-Förderung an einem Werkzeugbündel zur digitalen Musikedition gearbeitet wurde, das sich unter dem Namen „Edirom“ rasch im Fach verbreitet hatte. Diese doppelte, für das Fach charakteristische Ausrichtung auf Text und Musik blieb dann auch kennzeichnend für die im Sommer 2009 beginnende Beteiligung im TextGrid-Projekt.

Bei den Überlegungen, was das Fach Musikwissenschaft¹ zu dem Verbundprojekt TextGrid beitragen könne, standen denn auch beide Bereiche zur Diskussion und die erste Beschreibung eines möglichen Arbeitspakets begann mit der Bemerkung:

Eine Beteiligung der musikwissenschaftlichen Editionsphilologie am TextGrid-Projekt war bisher allenfalls im Bereich von Texteditionen denkbar, da die Voraussetzungen für die Integration fachspezifischer Gegebenheiten fehlten. Aber auch fachintern stellt die Einbindung von Noten-Texten eine wesentliche Hürde für digitale musikwissenschaftliche Editionen dar.²

Die Probleme, Noten-Texte oder Notenbeispiele etwa in TEI-Texte einzubinden, ohne auf bloße Grafikformate zurückzugreifen, hatten sich aber bereits einigen anderen TextGrid-Partnern gestellt. Mit MEI war nun ein Format entstanden, das – TEI vergleichbar – eine detaillierte Codierung von Inhalten einschließlich auftretender Fassungen und Varianten erlaubte. Problematisch war jedoch einerseits die praktische Handhabung dieser außerordentlich umfangreichen Codierungen, andererseits die zunächst fehlende Möglichkeit, diese Codierungen wieder in „Notensatz“ zu visualisieren und damit für Otto-Normalmusiker lesbar zu machen. Daraus erwuchs die Idee, innerhalb von TextGrid einerseits eine Art grafischen MEI-Editor zu entwerfen, andererseits eine – wenigstens grobe – Anzeigemöglichkeit für die Spezifika dieses Formats zu schaffen (stets in dem Bewusstsein, dass Notensatz ein außerordentlich komplexes Unterfangen ist, sodass mit dem begrenzten Res-

1 Präziser müsste von dem Teilbereich Musikphilologie gesprochen werden, der von dem im Fach leider als eher isoliert wahrgenommenen Gebiet des *Music Information Retrieval* abgesehen am frühesten den Weg ins digitale Zeitalter beschritten hatte.

2 Zitat aus dem Antragstext von 2008, S. 101

sourceneinsatz des Projekts nur eine stufenweise Annäherung an halbwegs befriedigende Bildschirm-Darstellungsformen zu verwirklichen wäre). Beides – Erleichterung des Umgangs mit dem Format und rudimentäre Anzeige- und damit Kontrollmöglichkeiten – waren aber Voraussetzungen für die langfristige Akzeptanz dieses neuen Formats in der Community und damit für die Rechtfertigung des Etiketts eines neuen Standards und den Aufbau entsprechender Korpora. Mit dem Versuch, diese Probleme innerhalb von TextGrid II anzugehen, rückte die Geburtsstunde von MEISE näher.

MEISE – der Weg zum MEI-Score-Editor als Teil des TextGridLab

Als im Juni 2009 Julian Dabbert die Arbeit im TextGrid-Projekt aufnahm, bestand seine Aufgabe zunächst im Entwurf und der Integration von musikalischen Symbolen in die Fonts des dort benutzten XML-Editors. Durch eine Umfrage in den Forschungsinstituten wurde der Bedarf an musikalischen Einzelzeichen und kleinräumigen Zeichenkombinationen in Fließtexten ermittelt und als erster Schritt ein entsprechender Musikfont (MusicNotation TextGrid) in den *codepoints* der *private use area* der Unicode-Zeichentabelle zur Verfügung gestellt. Dieser auch außerhalb des Editors nutzbare Musikfont ist ein nach wie vor nützlicher „Notbehelf“, war aber nur als dringliche Abhilfe bei Problemen mit der Integration einfachster Zeichen (z. B. ♪ oder ♪) bzw. von Zeichenkombinationen (z. B. ♪♪ oder ♪♪♪ bzw. Einträgen wie *ff*, *ff*, *ff*, *ff* oder *ff*) innerhalb einzelner Textzeilen gedacht³ und sollte lediglich den Ausgangspunkt für die längerfristige Entwicklung eines auf MEI basierenden wirklichen „Noteneditors“ bilden.

Angesichts der Komplexität der Notenschrift wurden die Ziele dieses Editors von vornherein eng abgesteckt, da das Erstellen von Notensatzprogrammen in der Regel Arbeit für Dutzende von Programmierern bedeutet und dabei eher in Mannjahren denn -monaten zu denken ist. Der in die TextGrid-Werkzeuge einzubettende Editor sollte einerseits eine Möglichkeit zur

3 Vgl. <http://sourceforge.net/projects/musicnotationtg>; erst in jüngster Zeit wurde im Anschluss an die während der Mainzer MEI-Tagung 2013 geführten Diskussionen zu einem ausführlicheren, über die 220 Glyphen des von Perry Roland erstellten Unicode-Standards musikalischer Symbole weit hinausgehenden Repertoire an musikalischen Zeichen im sogenannten SMuFL-Projekt (*Standard Music Font Layout*) ein nahezu 2400 Glyphen einschließender Musikzeichensatz namens „Bravura“ als OpenType-Font entwickelt, der setzkastenartig auch eine individuelle Kombination von Zeichen erlaubt; zu der im Juni 2014 veröffentlichten Version 1.0 vgl. <http://www.smuf.org/>.

schlichten On-the-fly-Anzeige des in MEI codierten Notentextes bieten, um eine visuelle Kontrolle der umfangreichen Codierungsdaten vornehmen zu können. Zum anderen sollten Wissenschaftler im WYSIWIG-Stil Änderungen an einem als Notentext dargestellten MEI-Dokument so vornehmen können, dass diese unmittelbar wieder in den MEI-Code eingeschrieben wurden. Gleichzeitig sollte aber die Möglichkeit bestehen bleiben, für speziellere Eingriffe eine direkte Manipulation des Quelltextes in einem XML-Editor vorzunehmen. Schließlich sollte auch eines der Alleinstellungsmerkmale von MEI im editorischen Bereich, die Möglichkeit der Integration von Varianten, in der Anzeigeoberfläche des MEI-Score-Editors berücksichtigt werden – ein in professionellen Notensatzprogrammen bis dahin undenkbares Feature. Das Pflichtenheft, das Julian Dabbert im September 2009 erstellt hatte, trug also durchaus erschreckende Züge.

Es würde hier zu weit führen, über die einzelnen Schritte und die Erfolge und Rückschläge ausführlicher zu berichten. Ein Hinweis auf die Probleme, die sich durch eine in der Partitur sowohl innerhalb des Einzeltakts als auch zwischen verschiedenen Zeilen horizontal und vertikal zu synchronisierende Zeichenfolge ergeben, mag genügen bzw. vielleicht noch ergänzt werden durch die Bemerkung, dass die Zeichen dabei aus Einzelpartikeln zusammengesetzt sind, die – wie etwa die Balken – flexibel im Taktrahmen auch noch so zu platzieren sind, dass Balken und Notenhäse sich punktgenau decken, selbst wenn in einer anderen parallelen Stimme nachträglich eingefügte kleinere Werte die Taktbreite verändern. Mit bewundernswerter Umsicht und Energie umschiffte Dabbert viele Probleme, sah sich aber auch immer wieder gezwungen, einmal gefundene Lösungen wieder zu verwerfen, weil sie bei komplexeren Phänomenen zu ästhetisch wenig befriedigenden Ergebnissen führten. So entwickelte sich nach und nach ein System, das von gruppierten Einzelnoten ausging, dann durch Balken verbundene Notengruppen ermöglichte, die Synchronisierung paralleler Stimmen erlaubte, weitere musikalische Zeichen wie Akzente, dynamische Zeichen oder – als schwierigsten Part eines jeglichen Notensatzes – Artikulations- bzw. Phrasierungsbögen einschloss. Schon früh experimentierte Dabbert auch mit unterschiedlichen Möglichkeiten der Darstellung von Varianten in simultaner oder sukzessiver Form; letztere war dann am Ende die bevorzugte, da hierbei in einfacher Weise zwischen den Varianten umgeschaltet werden konnte.

Zugleich wurden die Voraussetzungen dafür geschaffen, den XML-Baum per Drag & Drop zu verändern bzw. auf diese Weise neu Elemente einzufügen. Dies hatte insbesondere Bedeutung für die Handhabung der strukturellen

MEI-Daten: Bis wirklich eine Note in MEI geschrieben werden kann, sind umfangreiche Voreinstellungen (zur Partituranordnung, zur Satzeinteilung, zu Takt- und Tonartangabe, Zahl der Stimmen in einem System usw.) zu definieren. Sowohl die Erstellung dieser Daten als auch die Einfügung neuer Elemente der Baumstruktur (Takte, Noten, Pausen) oder die Einfügung von Varianten war nun durch das MEISE-Werkzeug erleichtert. In der Abbildung 1 kann man rechts neben dem Notentext die Palette der einfügbaren Elemente sehen, im rechten oberen Fenster die Strukturansicht, in der eine Note in Takt 8 (Measure 8) im dritten System (Staff 3) der Hauptstimme (Layer 1) markiert ist, die in der Notendarstellung blau hervorgehoben wird (Notenzeile 3, 3. Takt). Die Eigenschaften dieser Note können nun im entsprechenden Fenster unten rechts verändert werden (etwa die angezeigte Tonhöhe e zu c), das Ergebnis dieser Änderung wird im Notentext angezeigt und mit der Bestätigung der Änderung zugleich im MEI-Code abgespeichert.

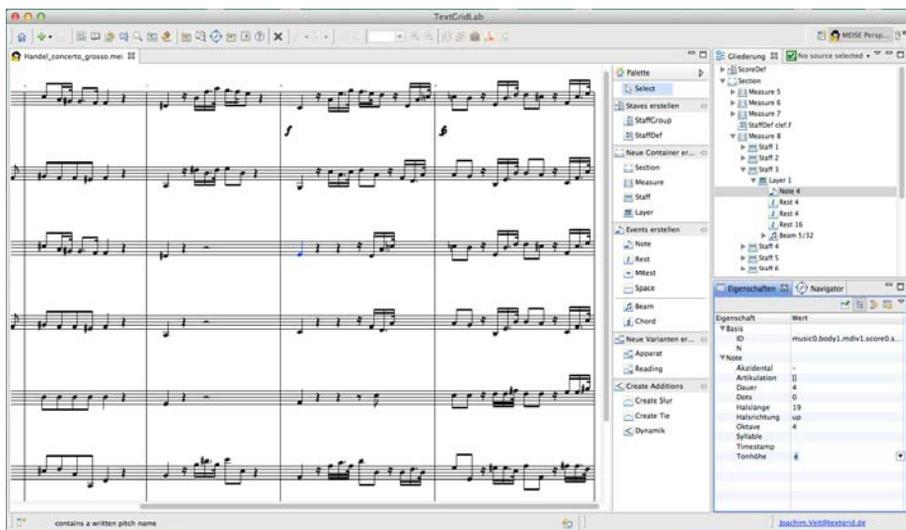


Abb. 1 Ausschnitt der Partituransicht eines *Concerto grosso* von Georg Friedrich Händel in MEISE

Die Hürden, die bei der Entwicklung von MEISE zu überwinden waren, hingen aber nicht nur mit den grundlegenden Schwierigkeiten der digitalen Darstellung von Notation zusammen, vielmehr kam erschwerend hinzu, dass MEI als ein sehr junges Format noch starken Veränderungen unterworfen war. Die Veröffentlichung eines komplett neuen Release von MEI, das nun eine längerfristig stabile Datengrundlage garantierte, aber natürlich eine Rei-

he von Anpassungen erforderte, verzögerte sich bis Mai 2010. Immerhin: „Das Umstellen auf 2010-05 ging eigentlich relativ schmerzarm – einige Architekturänderungen mussten vorgenommen werden, aber davon abgesehen läuft der Übergang reibungslos“, hielt Dabbert in seinem internen Arbeitstagebuch fest. Auch wenn durch MEI 2010-05 eine fixe Basis für die weitere Entwicklung geschaffen war – Klagen gab es weiterhin, doch diese bezogen sich nun immer häufiger auf Einschränkungen, die durch die Programmierumgebung der Eclipse Rich Client Platform entstanden. Technische Einschränkungen entstanden besonders durch das eingesetzte Graphical Editing Framework (GEF), da ein direktes Manipulieren des dargestellten Notentextes durch Maus-Interaktionen (Drag & Drop) im Rahmen dieser Technologie nicht bzw. nur mit großem Aufwand möglich war. Das war der Grund, warum als alternatives Eingabekonzept ein auf dem XML-Baum aufsetzender Gliederungseditor (Outline-View) in Kombination mit einem Attribut-Editor (Property-Editor) entwickelt werden musste, um Zugriffsmöglichkeiten auf sämtliche Bestandteile der Codierung zu erhalten, so wie dies oben bei Abbildung 1 beschrieben ist: Nicht das Ziehen der Note ändert dort die Tonhöhe in der Codierung, sondern umgekehrt die manuelle Änderung des Attributwerts führt zu einer abweichenden Darstellung im Notentext.

Innerhalb der technischen Rahmenbedingungen der Eclipse RCP war daher auch nicht denkbar, eine Verbindung zu der zunehmenden Zahl freier, auf Webtechnologien basierender, plattformunabhängiger Tools zu suchen, die eine dynamische Bearbeitung und Darstellung von Notentext bzw. direkte grafische Manipulationen erlauben. Hier wurde erstmals erkennbar, dass auch MEISE dauerhaft sinnvoller in eine Webanwendung zu überführen wäre, da dies auch Synergieeffekte durch die Einbeziehung existierender Tools ermöglichen würde. Um jedoch zunächst ein erstes grundlegendes Werkzeug für Editoren auf der Basis des schon Erreichten bereitzustellen, wurde das Tool auch bei der späteren Weiterführung im DARIAH-DE-Projekt (inzwischen auch als Standalone-Version nutzbar) zunächst in der gleichen Umgebung weiterentwickelt, bevor – auch vor dem Hintergrund jüngst erfolgter vielversprechender Tool-Entwicklungen in der MEI-Community – eine webbasierte Neukonzeption der MEISE in Angriff genommen wurde, die jedoch den ursprünglichen Zielsetzungen des Editors verpflichtet bleibt.⁴

4 Mit der Weiterführung in DARIAH-DE ist Nikolaos Beer betraut, der Erfahrungen mit MEI in der parallelen Arbeit an den Projekten *OPERA – Spektrum des europäischen*

Fiedellieder *plus* – ein Testfall für MEISE und das TextGridLab

Im Rückblick kann man es als eine günstige Konstellation bezeichnen, dass just zur Zeit der Entwicklung von MEISE ein sich über drei Semester erstreckendes Projektseminar am Musikwissenschaftlichen Seminar Detmold/Paderborn stattfand, in das sowohl dieses Werkzeug als auch verschiedene Tools des TextGridLab einbezogen werden konnten. Stefanie Rauch und der Autor dieses Beitrags erarbeiteten seit dem Wintersemester 2010/11 mit einer kleinen Studentengruppe eine Hybridedition der nach Texten von Theodor Storm komponierten „Fiedellieder“ des in Detmold und Hamburg wirkenden Violinvirtuosen und Komponisten Carl Louis Bargheer (1831–1902). Aufgearbeitet wurde dabei ein reichhaltiger Quellenfundus der Lippischen Landesbibliothek Detmold. Einerseits betraf dies textliche Vorlagen: Briefe und andere Ego-Dokumente des mit Johannes Brahms befreundeten Geigers waren ebenso zu erfassen wie der Text des Märchens „Das Sonnenkind“ von Ferdinand Röse, für das Storm ursprünglich einige Texte seines dann mehrfach überarbeiteten und erweiterten Liederzyklus konzipiert hatte. Für die Zusammenarbeit der Gruppe bei der Transkription, Kontrolle und Überarbeitung dieser Texte im TEI-Format boten sich die Werkzeuge des TextGridLab an, zumal dort für die Studierenden ein kostenlos zu nutzender leistungsfähiger XML-Editor zur Verfügung stand. Zugleich konnten im Repository für alle Beteiligten nicht nur die selbst erarbeiteten Texte, sondern auch sämtliche Digitalisate, die sowohl von den Textvorlagen als auch von den Musikalien angefertigt worden waren, zur Verfügung gestellt werden.

Die Editionsarbeiten an der Musik dieser für eine Singstimme, Soloviolone und Klavier komponierten Lieder erwiesen sich als besonders interessant, denn die in der Landesbibliothek erhaltenen Manuskripte dokumentierten mit ihren zahlreichen Überklebungen und Korrekturen fast lückenlos die Genese dieser Werke, die vom Komponisten immer wieder überarbeitet worden waren. In einer ersten Phase galt es zunächst die wichtigsten Unterschiede der Handschriften zu erfassen, um den Notentext für die im Druck erscheinende Neuauflage zu bestimmen. Dabei konnte zunächst das TextGridLab als eine Art Kollationsstool herangezogen werden, da sich Quellenfenster parallel über- oder auch nebeneinander öffnen und mit dem Textfenster des XML-Editors kombinieren ließen (vgl. Abb. 2). Der Text-Bild-

Link-Editor diente im übrigen parallel auch dazu, die Lithografien, die sich in dem gedruckten Märchentext Rösens befanden, mit den entsprechenden Textstellen zu verbinden.

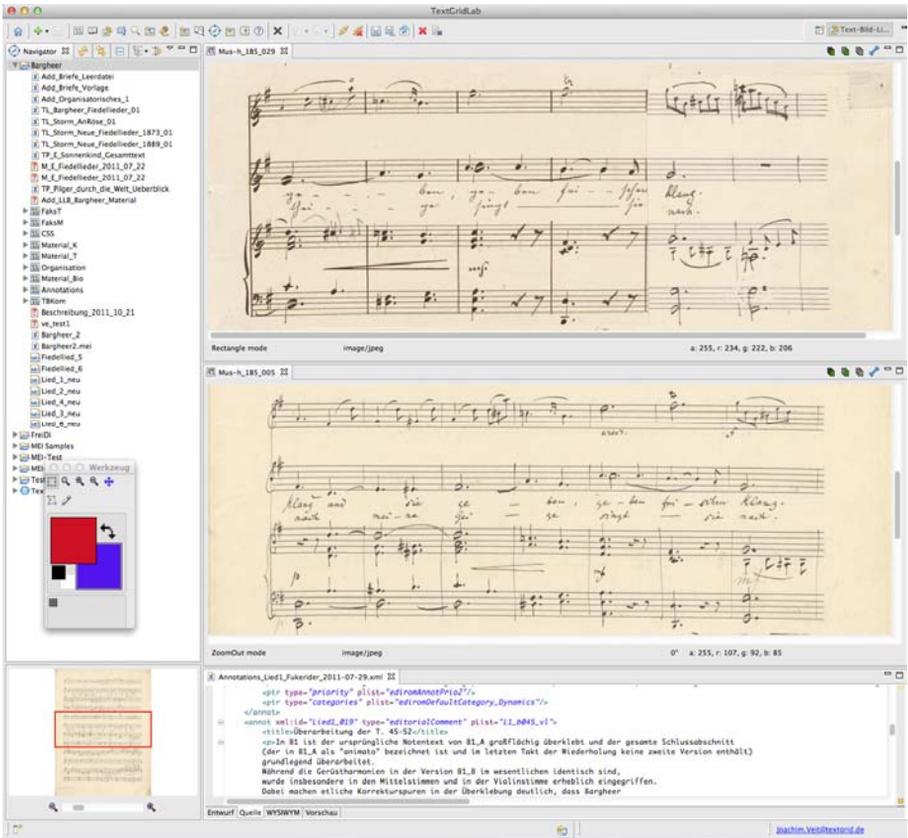


Abb. 2 Carl Louis Bargheer, „Fiedellieder“, Ausschnitt aus Nr. 1: oben überarbeitete Abschrift von Kopistenhand mit Korrekturen Bargheers, darunter Reinschrift des Komponisten, im unteren Bereich Anmerkungen zum Vergleich der Quellen

Bei der gemeinsamen Arbeit des Verzeichnens der Varianten, die später in die in Detmold entwickelte Edirom-Anzeige-Software⁵ eingebunden wer-

5 Vgl. dazu <http://www.edirom.de>; dieses weiter oben erwähnte DFG-Projekt zur Entwicklung von Werkzeugen für digitale Formen wissenschaftlich-kritischer Musikeditionen war in der Endphase der Förderung dabei, die frühere Anzeige-Version auf festen Datenträgern auf eine Online-Version umzustellen, die schließlich erstmals im Bargheer-Projekt genutzt werden konnte, sodass das studentische Unternehmen auch für diese Software einen Testfall darstellte.

den sollten, entstand schließlich der Wunsch, die im Detail nachvollziehbare Entwicklung der Kompositionen nicht nur durch die übliche Verbindung von grafischer Anzeige der Handschriftenfaksimiles und verbaler Beschreibung der Veränderungen innerhalb der Edirom zu verdeutlichen, sondern diese Vorgänge auch formalisiert innerhalb des Codierungsformats MEI zu erfassen. Nach einer grundlegenden Schulung in MEI und der Basis-Codierung der Lieder wurden dann in einer gemeinsamen Wochenendaktion die Entstehungsvarianten in die Codierung eingepflegt. Dies machte es nun möglich, innerhalb von MEISE die varianten Notentexte innerhalb von <reading>-Tags abzulegen und damit wieder in eine Anzeigeversion zu überführen. Abbildung 3 zeigt eine solche Ansicht, in der zwischen den verschiedenen zugrundeliegenden Manuskripten umgeschaltet werden kann (sichtbar ist die Version des zweiten Reinschrift-Autographs von Bargheer). Diese Codierungen sind bisher nicht in die vom Projekt erstellte Website aufgenommen worden, während die übrigen erarbeiteten Materialien innerhalb einer Anzeige-Version von „Edirom Online“ auf der projekteigenen Website dauerhaft zur Verfügung gestellt wurden.⁶

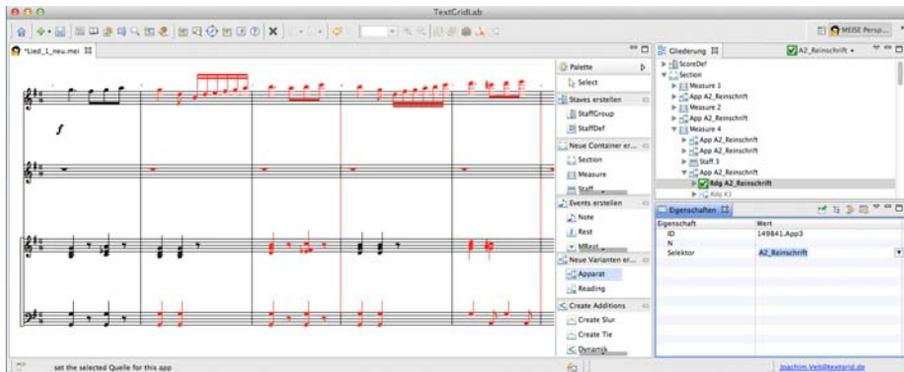


Abb. 3 Beginn von Nr. 1 der „Fiedellieder“ von C. L. Bargheer in der Varianten-Anzeige der MEISE; in Rot sind die Einträge aus der Reinschrift hervorgehoben, die von den übrigen Fassungen abweichen, ein Selektor erlaubt das Umschalten zwischen den Varianten

6 Vgl. <http://www.edirom.de/llb-bargheer>. Die gedruckte Edition der Noten erschien Ende 2012 im Carus-Verlag Stuttgart, hrsg. von Kostadin Delinikolov, Andreas Fukerider, Joachim Iffland, Ran Mo, Stefanie Rauch, Agnes Seipelt und Joachim Veit. Nachträglich wurde auch eine Tonaufnahme der Lieder erstellt, die auf der angegebenen Website abgerufen werden kann.

Ganz so reibungslos, wie dies nach der vorstehenden Beschreibung wirken mag, liefen die Arbeit mit den Werkzeugen des TextGridLab denn aber doch nicht ab. Die Hauptarbeitsphasen des studentischen Projekts fielen in eine Zeit verstärkter Testphasen der TextGrid-Komponenten zwischen der im Sommer 2011 bevorstehenden Veröffentlichung von Release 1.0 und der zunächst finalen Version 2.0 vom Mai 2012. Die Studierenden stellten sich freundlicherweise auch zweimal für einen Testtag zur Verfügung. Schwierigkeiten gab es zunächst wiederholt mit dem individuellen Zugang zum Lab – hier konnte durch Vermittlung Julian Dabberts dankenswerterweise meist rasch Abhilfe geschaffen werden. Andere Probleme betrafen die Stabilität der Tools (auch von MEISE), zudem bereiteten die etwas unübersichtliche Versionierung der Daten und die durch schwer kontrollierbare Down- und Upload-Vorgänge nur umständlich zu verwirklichende Offline-Arbeit mit den Daten Kopfzerbrechen. Auch führte der Versuch einer Integration eigener CSS-Dateien gelegentlich zu Abstürzen. Der scherzhafte Ausspruch eines Studenten, Vieles sei ja außerordentlich gut gedacht, wenn es denn nur auch funktionieren würde, war zu Anfang Ausdruck einer gewissen Skepsis – einer Skepsis, die dadurch verständlich wurde, dass die gewohnte Arbeit mit Word-Dateien in dieser Umgebung plötzlich nicht mehr möglich war ...

Dennoch: Die Tatsache, dass die auftretenden Schwierigkeiten oder die Anregungen aus dem Kreis der Bargheerianer von den Mitarbeiterinnen und Mitarbeitern des TextGrid-Verbands ernst genommen wurden, Probleme entweder umgehend behoben oder über Julian Dabbert auch Rückmeldungen aus dem Kreise der Techniker kamen, wenn die Lösung einmal nicht ganz so einfach war und ein paar Tage mehr in Anspruch nahm – dies alles beeindruckte und wurde dankbar als Ausdruck der Bemühungen verstanden, mit TextGrid wirklich Werkzeuge zum Nutzen der Forscher und damit auch für den wissenschaftlichen Nachwuchs zu konstruieren.

Fazit: Was bleibt nach dreijähriger Beteiligung an TextGrid?

Was erscheint im Rückblick auf unsere Beteiligung an der zweiten TextGrid-Phase wichtig? Vorab ist festzuhalten, dass zwar die unmittelbare Mitwirkung nach dieser Phase endete, nicht aber die Beziehung zu TextGrid. Einerseits wurde die Entwicklung von MEISE anschließend im DARIAH-DE-Verband weitergeführt, andererseits war in dem im Sommer 2012 gestarteten BMBF-Projekt *Freischütz Digital* von Anfang an vorgesehen, eine Anbin-

derung an TextGrid zu schaffen, um eine Verbindung zwischen den in Detmold entwickelten Edirom-Werkzeugen und den durch TextGrid entstandenen Möglichkeiten zu schaffen.⁷ Schließlich konnten wir in etlichen Beratungsgesprächen geplante oder neue Forschungsunternehmen auf die Eignung von TextGrid als Forschungsumgebung auch im musikwissenschaftlichen Umfeld hinweisen.

Selbst wenn innerhalb der dreijährigen TextGrid-II-Phase Julian Dabbert als Mitarbeiter in zahlreiche weitere TextGrid-Aktivitäten eingebunden war oder sich z. B. Daniel Röwenstrunk und Joachim Veit bei einem Leipziger Workshop der Akademien der Wissenschaften zu „Personendateien“ im September 2009 dafür einsetzten, das Grid für eine gemeinsame, vernetzte Arbeit an Personendaten zu nutzen, oder wenn TextGrid und MEISE stets eine Rolle auf den 2010 gestarteten Paderborner Edirom-Summer-Schools spielten – der Umgang „mit Noten“ blieb doch vordergründig der zentrale Gegenstand unseres Beitrags zu TextGrid. Sind die diesbezüglichen Ergebnisse zufriedenstellend? Salopp könnte man im Hinblick auf die Wirkung des entwickelten Tools sagen „Eine MEISE macht noch keinen digitalen Sommer“! Von Bedeutung scheint weniger der aus heutiger Sicht vermutlich eher unbefriedigende Leistungsumfang dieses Tools (das aufgrund der erkennbaren strukturellen Mängel daher derzeit auch neu konzipiert wird) als vielmehr die Signalwirkung, die von der Arbeit an der MEISE ausging: Gerade mit MEISE konnten wesentliche Vorteile des neuen Codierungsstandards – wie etwa der formalisierte Umgang mit Fassungen und Varianten – für ein über den engen Entwicklerkreis von MEI hinausgehendes Publikum verdeutlicht werden.

Darüber hinaus ebnete die Rückführung der Codierung in einen lesbaren Notentext dem Format den Weg zu einer breiteren Akzeptanz, trug dies doch wesentlich dazu bei, Verständnis für die auf Außenstehende zunächst eher kryptisch wirkenden XML-Codierungen zu wecken und Wissenschaftler behutsam an diese Strukturen und ihren Nutzen heranzuführen. Für die Durchsetzung von MEI als neuem wissenschaftlichem Standard zur Codierung von Musik hat MEISE insofern einen wichtigen Beitrag geleistet – die rasante Verbreitung des Formats in jüngster Zeit hängt auch damit zusam-

⁷ Zu dem Projekt vgl. www.freischuetz-digital.de; die im Projekt mit erarbeiteten und mit Edirom Online zu publizierenden Daten sind inzwischen nach einem ersten Release ins TextGrid Repository eingebunden und damit die Langzeitverfügbarkeit und Nachnutzungsmöglichkeiten garantiert.

men, dass mit MEISE sichtbar wurde, dass nützliche Werkzeuge für dieses Format zur Verfügung stehen, sodass nun umfangreichere, auf MEI aufbauende Editionen in Angriff genommen werden konnten.⁸ Dies hat auch dazu geführt, dass das „Rendern“ von MEI-Daten in vielen Projekten als Problem erkannt wurde und vielversprechende parallele Entwicklungen entstanden, die in jüngster Zeit dazu führten, dass sich die Beteiligten abstimmen und statt konkurrierender nun komplementäre Tools entstehen, sodass eine künftige Version von MEISE ebenfalls von solchen Synergieeffekten profitieren kann.⁹

Aus meiner Sicht wäre es jedoch verfehlt, ein Fazit der Mitarbeit in TextGrid alleine auf dieses kleine Tool – ja, nicht einmal auf das Werkzeugset insgesamt – zu beziehen. Für die langfristige Entwicklung der digitalen Methoden und der jungen Disziplin der Digital Humanities sehr viel bedeutender ist in meinen Augen die Tatsache, dass sich hier eine Schar gleichgesinnter Wissenschaftlerinnen und Wissenschaftler aus den verschiedensten geisteswissenschaftlichen Disziplinen und der Informatik zu einem gemeinsamen Unternehmen zusammengefunden hatte, das erstmals einen intensiven Austausch aller Beteiligten mit ihren so unterschiedlichen Voraussetzungen ermöglichte und das Verständnis für die Probleme der je anderen Seite beförderte. Dank der noch überschaubaren Zahl der beteiligten Projektpartner konnte man sich rasch untereinander, lernte Stärken und Schwächen einschätzen und fand gerade in den kritischen Phasen vor den jeweiligen Releases zu einer wunderbaren Gemeinschaft zusammen, die unter Einsatz aller Kräfte bestrebt war, den Nutzern eine problemlos funktionierende Virtuelle Forschungsumgebung zur Unterstützung ihrer (und letztlich auch der eigenen) Arbeiten bereitzustellen. Die kollegiale Atmosphäre, die wohlwollend freundliche Aufnahme, die wir „Notenwissenschaftler“ als eine Art „Exoten“

8 Neben Einzelprojekten wie *Freischütz Digital* haben z. B. im deutschsprachigen Raum größere Neuvorhaben inzwischen diesen Codierungsstandard aufgegriffen, so das an der Berliner Hochschule der Künste angesiedelte Projekt *A Cosmopolitan Composer in Pre-Revolutionary Europe – Giuseppe Sarti* und das Akademieprojekt *Beethovens Werkstatt*. Auch die *Digitale Mozartedition* in Salzburg und das Leipziger Projekt *Bach Digital* planen zzt. einen Wechsel zu MEI. Als Basis der Datenstrukturen der Edirom-Software spielt das Format ohnehin bei allen Projekten, die diese Werkzeuge nutzen, eine Rolle.

9 Absprachen in diese Richtung erfolgten sowohl bei der MEI-Konferenz in Charlottesville im Juli 2014 als auch bei der folgenden Digital-Humanities-Tagung in Lausanne Anfang Juli 2014.

in diesem Kreise stets fanden, und die vielen interessanten neuen Kontakte haben die eigenen Arbeiten beflügelt und waren mit dafür verantwortlich, dass wir immer mehr Kolleginnen und Kollegen des eigenen Faches von den Vorzügen digitaler Forschungsmethoden und Arbeitstechniken überzeugen konnten. Dies ist vielleicht das wichtigste Ergebnis unserer Beteiligung an TextGrid, denn gemäß dem Spruch „Der Ton macht die Musik“ hängt sehr viel von der Atmosphäre des Umgangs miteinander ab, in der erst lebendige Forschung gedeihen kann. Und eine Virtuelle Forschungsumgebung wächst vor allem mit dem Vertrauen, das die Nutzer im Umgang mit daran beteiligten Personen – und nicht mit anonymen Techniken – aufbauen.

Literaturverzeichnis

Stadler, Peter; Veit, Joachim (Hrsg.) (2009): *Digitale Edition zwischen Experiment und Standardisierung. Musik – Text – Codierung* (Beihefte zu editio; 31). Tübingen: Niemeyer.

Bilder als Quelle in TextGrid

von Robert Casties und Dirk Wintergrün

Text und Bild als Quelle

Nur ein beschränkter Teil der Quellengrundlagen in den historisch arbeitenden Wissenschaften – der allgemeinen Geschichte, Wissenschaftsgeschichte, Technikgeschichte, Kunstgeschichte – besteht aus textuellen Überlieferungen. Grundlage für die Erschließung historischer Kontexte und das Verständnis von Wissenssystemen ist in gleichem Maße das durch Artefakte, Zeichnungen und Pläne überlieferte historische Wissen.¹ Ein zentraler Aspekt der Quellenarbeit besteht aus der Kontextualisierung verschiedener Überlieferungstraditionen. Darauf muss eine Virtuelle Forschungsumgebung für die Geisteswissenschaften Rücksicht nehmen, auch wenn sie ihren eigentlichen Schwerpunkt auf die Arbeit mit Texten legt.

Zusätzlich liegen viele historisch relevante Texte nicht oder nur ausschnittsweise als Volltexte vor. Zugänglich sind in der Regel nur digitale Faksimiles, die von verschiedenen Bearbeitern mit verschiedenen disziplinären Hintergründen nur zur Beantwortung ihrer Fragestellungen nur in Ausschnitten transkribiert werden. Hier steht die fachwissenschaftliche Fragestellung im Zentrum und nicht der Anspruch, eine kritische Edition des Textes zu erstellen. Für die kritische Analyse und Überprüfung kann im Zweifelsfall immer auf das Faksimile zurückgegriffen werden. Eine Forschungsumgebung muss diese Arbeitsform unterstützen, will sie sich aus dem engen Rahmen der Textwissenschaften heraus bewegen. Sie kann dazu beitragen, fachwissenschaftliche analytische Arbeit an einer Textquelle und die textwissenschaftliche Erarbeitung einer kritischen Edition miteinander zu synchronisieren. Sie ermöglicht die zeitgleiche Arbeit an einer Quelle aus beiden Sichten – Mittler ist das Abbild der Quelle als digitales Bild. Die Herausforderung ist die Erstellung einer Arbeitsumgebung, die die Bild- und Textannotation ebenso wie die bidirektionale Verknüpfung von Text und Bild ermöglicht.

¹ Zur Rolle von Objekten in der Wissenschafts- und Wissensgeschichte siehe z.B. Hagner (1994), Renn (2014a) und als praktisches Beispiel die gerade in Open Access erschienenen drei Bände zu Wissensgeschichte der Architektur Renn (2014b).

Im Folgenden werden wir uns mit den Anforderungen an eine Annotationsumgebung für Bilder beschäftigen, die dafür notwendigen Voraussetzungen beleuchten und schließlich unsere technische Realisierung als Einzelanwendung und im Rahmen von TextGrid vorstellen. Bevor wir im Weiteren auf die Bedingungen und technischen Möglichkeiten von Bildannotationen im digitalen Zeitalter eingehen, erfolgt zunächst ein Rückgriff auf traditionelle Arbeitsweisen in der klassischen nicht-digitalen Welt.

Kommentare als Annotationen

Aus der Arbeitspraxis ist jedem analytisch mit einem Text oder einer Abbildung Arbeitenden intuitiv der Prozess des Annotierens ohne weitere Erläuterung verständlich – das gilt für das Schulkind wie den gestandenen Wissenschaftler oder die gestandene Wissenschaftlerin, unabhängig von der disziplinären Orientierung. Man nimmt einen Stift, wenn man das Dokument selbst verändern darf, markiert einen relevanten Teil und versieht diesen mit einer Bemerkung. Dieser Praxis verdanken wir historisch aufschlussreiche Erkenntnisse der Rezeption von Texten² – sei es in Form von Randbemerkungen in Druckwerken oder in der Annotationspraxis der Manuskripttradition.

Weniger invasiv wird – getrennt vom Dokument, auf einer Karteikarte, in einem Notizbuch und später in einer Fußnote – möglichst genau versucht zu beschreiben, wo ein annotierter Teil in der Quelle zu finden ist, eventuell verbunden mit einem Exzerpt. Bemerkungen werden also auf einem externen Medium verfasst. Alternativ wird die Annotation auf einer möglichst detailgetreuen Kopie des Originals oder auch auf einer über das Original gelegten transparenten Folie erstellt. Für alle diese Fälle finden sich in der aktuellen und in der historischen Praxis Beispiele und Bezüge. Verzichtet man auf den Eingriff in das Dokument, wie dieses in allem außer dem ersten Fall geschieht, eint diese Fälle das Problem, den Bezug von Annotation und Original möglichst genau wiederherzustellen, um die Annotationen nachvollziehbar zu machen und möglichst stabil an das originale Dokument anzubinden. Gemeinhin ist bei diesen Arbeitsweisen die Annotation stets nur ein Mittel zum Zweck für eine weitere Arbeit – als Materialsammlung für die eigentliche Publikation.

2 Ein herausragendes Beispiel findet sich in den Annotationen von Guidobaldo Del Monte an Benedettis und Jordanus' Mechanik (vgl. Damerow 2012).

Darüber liegt eine zweite Problemebene: Was ist das annotierte Objekt? Wie stark ist die Annotation an die konkrete Repräsentation des Originals gebunden? Nehmen wir das Beispiel eines Textes, der in verschiedenen Ausgaben vorliegt, z.B. mit unterschiedlicher Paginierung. Steht der Text selbst im Zentrum, so sollte die Annotation einen Bezugspunkt finden, der unabhängig von der tatsächlichen Repräsentation des Textes auf der jeweiligen (Buch-) Seite ist. Gelöst ist dies höchst stabil zumeist in religiösen Kontexten, insbesondere in der jüdischen Tradition, die feste Zählweisen der Verse aufweist; damit ist für die Annotation des Inhaltes jeweils ein fester Ankerpunkt vorgegeben. Fehlt ein solcher fester, durch Konvention mit dem Text verbundener Anker oder ist die Position des Textes auf einer Seite in einer spezifischen Ausgabe relevant, ist der übliche Weg die Angabe des Ankers in einem Koordinatensystem mit Text, Edition/Ausgabe, Zeile, Wort, Buchstabe als Basis. Der Inhalt ist damit unlösbar an seine Repräsentation gebunden. Fehlt eine der Angaben oder hat der Rezipient keinen Zugang zur genannten Edition, läuft die Annotation ins Leere, wird dekontextualisiert und damit letztendlich bedeutungslos.

Das Objekt selbst – das Exemplar – tritt in den Mittelpunkt, wenn es um die Sprache des Objektes geht, um seine Geschichte, die Form seines Gebrauchs. In diesem Fall reicht es nicht aus, den textlichen Inhalt zu annotieren, d.h. eine Referenz auf den abstrakten Text zu setzen. Die Annotation muss mit der konkreten Repräsentation verbunden sein, letztendlich auf dem Objekt selbst positioniert werden.

Gegenüberstellung als Annotation

Neben der Annotation als Kommentar ergibt sich eine andere Herausforderung aus der Gegenüberstellung von Objekten, der Verknüpfung von mehreren Objekten oder seiner Teile zu einem neuen Zusammenhang. Nur selten, zum Beispiel in Form von Ausstellungen, lässt sich dieses an den Objekten realisieren. Viel mehr noch als bei der Annotation des Einzelobjektes ist das Verfahren auf Reproduktion und genaue Referenzen angewiesen. In der klassischen Textarbeit spiegelt sich die Referenz im Normalfall in der Fußnote wider, doch ist diese naturgegeben stets nur eine halbe Referenz. Nie (oder fast nie) erfährt das referenzierte Objekt von seiner Referenzierung. Die Quelle „weiß“ nichts von ihrer Bearbeitung und Interpretation, der Rezipient der Quelle benötigt stets zusätzliches Wissen, um von Bearbeitungen und

Interpretation der Quelle durch Dritte zu erfahren. In der Kunstgeschichte hat die Fotografie es ermöglicht, räumlich getrennt Objekte in großem Maße gegenüberzustellen und damit neue Formen des visuellen Bildvergleichs eröffnet (vgl. Kieven 2014). Warburgs Displays³ zeigen eindrucksvoll, wie weit die Arbeit mit der fotografischen Reproduktion hin zu einem Netzwerk von Informationen geführt werden kann. Doch auch hier wird jede einzelne Reproduktion zunächst einmal zu einem Unikat, losgelöst vom ursprünglichen Objekt.

Vom Klassischen ins Digitale

Der Schritt von aus der klassischen Arbeitsweise gewohnten Verfahren der Annotation ins Digitale bedeutet zunächst einmal eine Übertragung der gewohnten Verfahren des Annotierens der geisteswissenschaftlichen Praxis in digitale Verfahren, auf der anderen Seite jedoch auch ein Überwinden von Grenzen, das nur die digitalen Verfahren ermöglichen. Neben dem unmittelbaren Zugriff auf die digitale Quelle wird die Möglichkeit eröffnet, Referenzen direkt im Dokument zu verankern und bidirektional Quelle und Dokument – allgemeiner: jedes digitale Objekt –, theoretisch in jeder beliebigen Granularität, mit jedem anderen Objekt zu verbinden.

Doch von der Möglichkeit zur Realität ist immer noch ein weiter Weg zu gehen. Die Realität der Hyperlinks im Web ist eine Welt der Einbahnstraßen und zerstörten Wege. Anker in Texten im Netz hängen noch immer zumeist von der Repräsentation eines Textes in HTML ab, mithin wie im klassischen Verfahren von der konkreten Darstellung, hier nicht im gedruckten Buch, sondern auf der Webseite. Auch wenn auf der Ebene der Beschreibung des eigentlichen Textes durch XML/TEI zusammen mit Techniken wie XPointer und XLinks das Problem gelöst ist, bleibt unbeantwortet, wie diese Anker beim Übergang vom TEI in HTML stabil gehalten werden können – insbesondere so stabil, dass auf jeweils spezifischen Browsern mit einem externen Annotationstool⁴ auf der Webdarstellung arbeitende Nutzer diese Verknüpfungen auch wieder auf den eigentlichen Primärtext in TEI verweisen lassen und dort festhalten können.

3 Vgl. Warnke (2008) bzw. <http://www.mediaartnet.org/werke/mnemosyne/>.

4 z. B. bei Tools wie Pundit (<http://thepundit.it>), das Annotationen in Overlaytechnik ermöglicht (vgl. Grassi 2013)

Auf der Ebene der Bild-Annotationen verschärft sich das Problem noch einmal. Hier fehlt es selbst auf der Quellebene noch an einem etablierten Standard, wie eine Marke, ein Overlay auf dem Bild unabhängig von der Darstellung des Bildes beschrieben werden können – etwa, wenn nur ein Ausschnitt eines Bildes dargestellt wird oder das Bild in verschiedenen Vergrößerungen vorliegt. Mit dem im Folgenden beschriebenen Bildbetrachtungs- und Annotationstool *digilib* sind wir einen Weg gegangen, der für dieses Problem einen Lösungsansatz bietet. Einen anderen Lösungsansatz zum Aufheben der Einbahnstraße zwischen Quelle und Text geht der Text-Bild-Link-Editor von TextGrid.

Faksimiles und Annotationen

Warum wurde und wird ein Bildserver wie *digilib* an einem wissenschafts-historischen Institut schon früh in Kooperation mit einem kunsthistorischen Institut entwickelt? Der Grund ist so offensichtlich wie einfach, doch soll er hier noch einmal genannt werden: Es fehlte an einem Tool, das es ermöglichte, mit Faksimiles von historischen Dokumenten – in unserem Falle von Drucken aus der frühen Neuzeit und mit Manuskripten – zu arbeiten und diese in Publikationen einzubauen. Eingang fand dieses Tool dann in eines der ersten größeren „Digital Library“-Projekte des DFG- und NFS-geförderten Projektes Archimedes und später in die große EU-geförderte ECHO-Initiative zur Publikation von kulturellem Erbe im Netz. Ein weiteres Ziel war, eine generische Lösung für das mit der Biblioteca Nazionale Centrale, Florence, dem Istituto e Museo di Storia della Scienza, Florence bereits 1998⁵ durchgeführte Projekt einer elektronischen Bild- und Text-Edition des Manuskripts von Galileo Galilei zu finden.

digilib

Der Bildserver *digilib*⁶ entstand ab 2000 an der Uni Bern und dem MPI für Wissenschaftsgeschichte aus anderen Vorarbeiten für netzwerkbasierte Betrachter für gescannte Dokumente. Während die vorherigen Betrachter, ähnlich wie der DFG-Viewer noch heute, Bilder in fester Auflösung benutzten,

5 MS72; s. http://www.mpiwg-berlin.mpg.de/Galileo_Prototype/MAIN.HTM

6 *Digilib* ist Open-Source-Software und unter <http://digilib.sourceforge.net/> zu finden.

bestand digilib von Anfang an aus einem die Bilder skalierenden Server, dem „Scaler“, sowie einem browserbasierten Frontend, das die Bilder oder Bildausschnitte nur in der jeweils sichtbaren Auflösung anfordert und darstellt. Dies macht es möglich, auch über langsame Netzwerke detailliert mit sehr großen Bildern zu arbeiten, da zu jedem Zeitpunkt nur jeweils ein kleines Bild an den Browser übertragen wird. Die Eignung zur Arbeit mit sehr großen Bildern wurde bereits ab 2002 in Zusammenarbeit mit Kunst- und Architekturhistorikern im Rahmen des ECHO-Projekts unter Beweis gestellt und weiterentwickelt (vgl. Raspe/Casties 2006). Aufgrund des server- und browserbasierten Ansatzes konnten zusätzliche, aus Anforderungen der Kunsthistoriker entstandene, Bildbearbeitungsfunktionen – wie die stufenlose Drehung von Bildern, Spiegelungen, Kontrast- und Helligkeitsveränderungen – schnell entwickelt und eingesetzt werden, da nur die Software auf dem Server ersetzt werden musste.

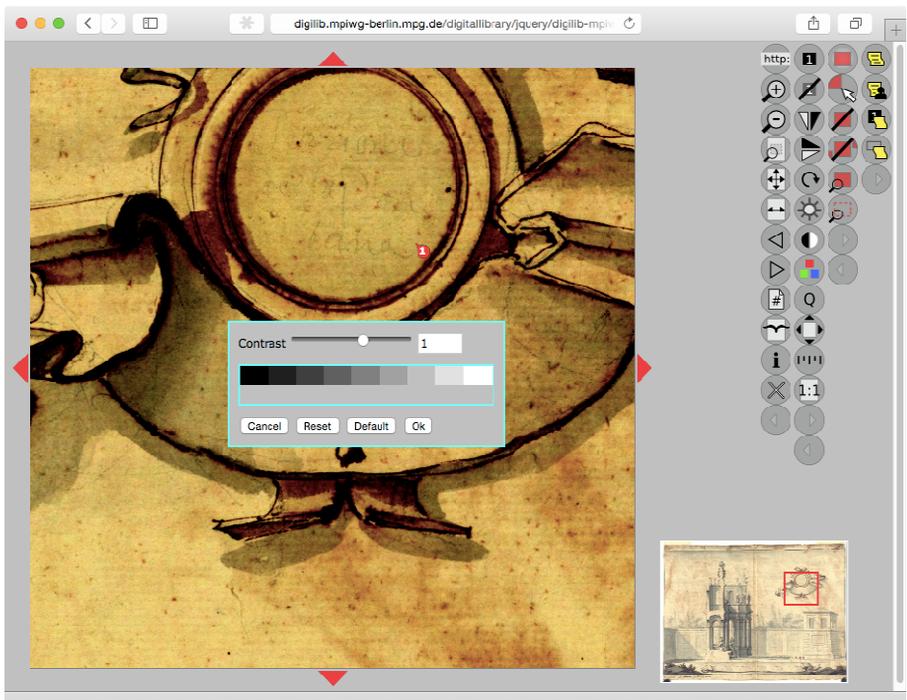


Abb. 1 Bildbearbeitungsfunktionen in digilib

Eine weitere Funktion, die digilib von Anfang an auszeichnete, ist die Fähigkeit, Markierungen auf dem Bild zu setzen und in Form einer URL zu

speichern. Diese Art der Annotation existiert nur lokal im Browser des Benutzers und enthält keine weiteren Informationen. Dennoch lässt sich eine solche URL zusammen mit einem Kommentar in einer E-Mail verschicken oder zur Illustration in einem PDF- oder HTML-Dokument verwenden. Diese Funktion wird nach wie vor viel benutzt.

Die URL kodiert den vollständigen Zustand von digilib – wie den betrachteten Ausschnitt, Rotationen, mögliche Bildbearbeitungen wie Kontrastverstärkungen und gesetzte Markierungen –, sodass beim Aufruf der URL im Browser diese spezielle Ansicht wiederhergestellt wird. Gleichzeitig bleibt das gesamte Bild durch Herauszoomen oder andere Bildbearbeitungen weiterhin erforschbar. Der Leser eines elektronischen Texts, der über einen solchen Link auf eine Quelle verwiesen wird, ist nicht auf den Bildausschnitt beschränkt, den der Autor verlinkt hat; er kann sich ein eigenes Bild des Kontexts machen und vielleicht weiteres entdecken.

Die Möglichkeit, mit der digilib-URL einen bestimmten Aspekt oder ein Bilddetail zu markieren und so zu annotieren, hat die Forscher von Anfang an begeistert. Sehr schnell kam in der Folge der Wunsch auf, mit der Annotation gleich einen Text zu verbinden, um sich beispielsweise in Gruppen über Bilder auszutauschen, ebenso wie der Wunsch, Annotationen anderer Benutzer zu sehen, die das gleiche Bild annotiert haben. Durch die Einweg-Natur der Web-Links waren diese Funktionen lange nicht praktikabel, da nur der Browser des jeweiligen Benutzers über die Information verfügt, dass ein Bild annotiert wurde – was dem Server nicht mitgeteilt werden kann. Diese Situation hat sich in der letzten Zeit mit dem Aufkommen verschiedener Web-Annotationssysteme verbessert, die Annotationen unabhängig vom Webserver auf eigenen Annotationsservern speichern und als Overlay über der existierenden Seite darstellen.⁷ Auch digilib bietet seit kurzem ein Annotations-Plug-in, mit dem Markierungen auf einem Bild mit kurzen Texten versehen und auf einem Annotationsserver gespeichert werden können.

⁷ Derzeit am weitesten verbreitet ist Annotator.js (<http://annotatorjs.org/>), auf dessen Code auch die permanente Annotationsfunktion für digilib beruht.

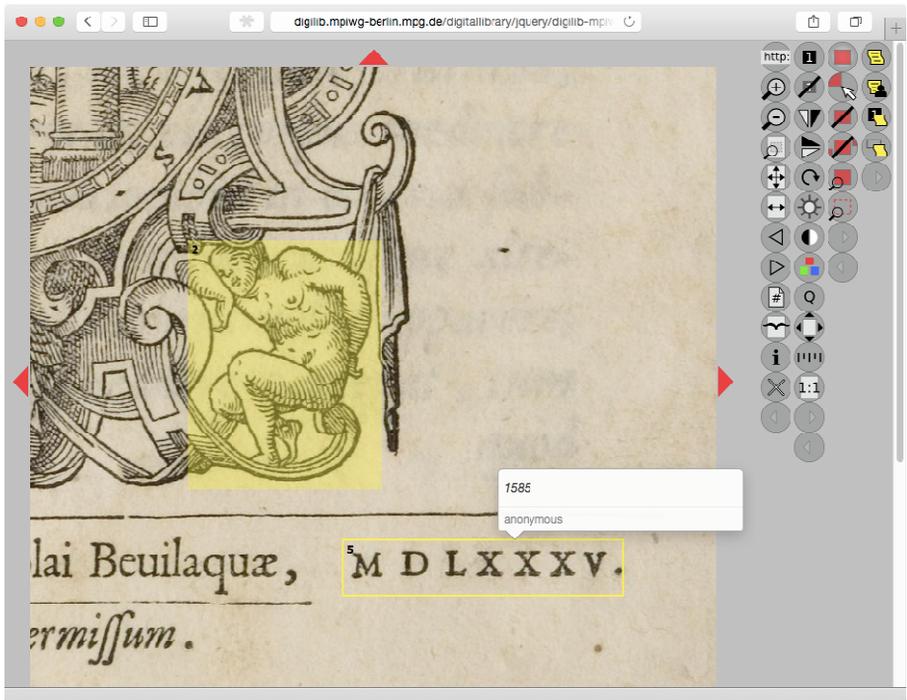


Abb. 2 Serverbasierte Annotationsfunktion in digilib

digilib in TextGrid

Seit 2011 bietet TextGrid digilib-Funktionen für Bilder in der Virtuellen Forschungsumgebung. Zur Implementation gehören zwei Teile: zum einen der digilib-TextGrid-Service auf dem TextGrid-Repository-Server, der die Funktionen des digilib-Scalers, wie die Skalierung der Bilder und weitere Bildbearbeitungsfunktionen, übernimmt; zum anderen ein digilib-Plug-in für das TextGridLab, das mit dem Service kommuniziert und die Darstellung der Bilder sowie die Steuerung der Bearbeitungsfunktionen übernimmt.

Mit dem digilib-Lab-Plug-in können beliebige Bilder aus dem Repository angezeigt und mit den Funktionen von digilib – wie Zoomen, Drehen, Spiegeln, Kontrast- und Helligkeitsveränderungen – bearbeitet werden. Die bearbeiteten Bildansichten können dann als digilib-Objekte gespeichert und wieder aufgerufen werden.

Der digilib-TextGrid-Service, der anfänglich nur über eine Schnittstelle zum Lab verfügte, erhielt im Zuge der Implementation des IIIF/SharedCan-

vas-Standards⁸ eine Web-Schnittstelle, die es ermöglicht, die Bilder aus dem TextGrid-Rep unter Benutzung der digilib-Funktionen in andere Webseiten wie die Browser-Ansicht des Repositorys einzubinden und die Bilder mit anderen Tools außerhalb des Lab zu verwenden.

Text-Bild-Link-Editor

Der Text-Bild-Link Editor (TBLE) im TextGridLab erlaubt es, Teile von Bildern mit Elementen eines strukturierten Texts zu verbinden. Im Bild können Regionen mit Rechtecken oder Polygonen markiert werden. Dann können im Text XML-Elemente oder beliebige Bereiche ausgewählt und mit der Bildregion verlinkt werden. Die entstandenen Links können dann als eigenständiges Objekt im TextGridRep gespeichert und wieder aufgerufen werden.

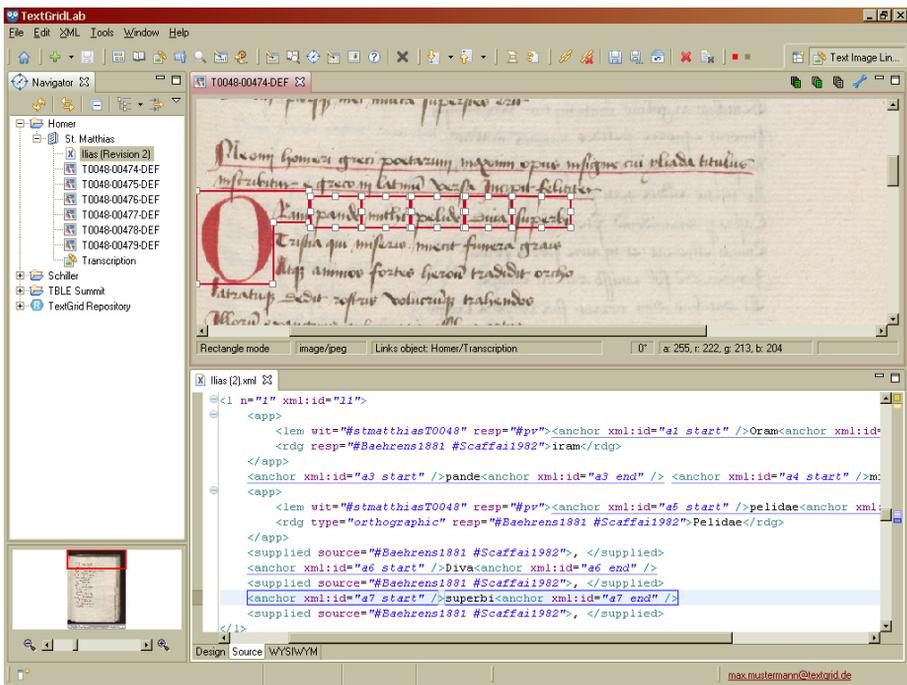


Abb. 3 Der Text-Bild-Link Editor (TBLE) in TextGrid

⁸ Das Projekt „TextGrid Implementation of SharedCanvas“ wurde von der Mellon Foundation (Grant 11300139) gefördert.

Mithilfe des TBLE können beispielsweise Transkriptionen von Manuskripten Wort für Wort oder Zeile für Zeile mit den gescannten Bildern der Seiten verbunden werden. Es ist auch möglich, eigene Notizen mit Bildern zu verlinken und auf diesem Weg Bild-Annotationen zu erstellen.

Die mit dem TBLE entstehenden Link-Objekte existieren im TextGridRep getrennt von den Bildern; für sie ebenso wie für die digilib-URLs gilt, dass es nicht ohne Weiteres möglich ist, von einem Bild ausgehend sämtliche Verweise auf dieses Bild zu finden und anzuzeigen.

Standards für Publikationen und kooperierende Werkzeuge

Beim Blick auf populäre Internetseiten scheinen digitale Bilder immer mehr zuzunehmen und in manchen Bereichen den Text als Kommunikationsform beinahe zurückzudrängen. Gleichzeitig stehen auch immer mehr historische Quellen gescannt als Bilder für die geisteswissenschaftliche Arbeit zur Verfügung. Zunehmend gibt es Werkzeuge wie digilib und Arbeitsumgebungen wie TextGrid, die es ermöglichen, wie oben beschrieben, wichtige Schritte geisteswissenschaftlicher Arbeit wie das Sammeln und Annotieren von Bildern ins Digitale zu übertragen – bislang sind diese Tools jedoch in der Regel noch wenig aufeinander abgestimmt. Eine stabile elektronische Publikation unter Verwendung von elektronischen Bildern aus verschiedenen Repositorien in aller Welt zu erzeugen, stellt weiterhin eine Herausforderung dar.

Der nächste Schritt der Entwicklung wird, so steht zu hoffen, ein Schritt hin zur Standardisierung und zur Kooperation der existierenden Werkzeuge sein. Im Bereich der Bild-Repositorien und -Annotationen liefern der IIIF-Standard und das OpenAnnotation-Modell⁹ internationale Standards, die zunehmend von großen Repositorien wie Museen und Sammlungen ebenso wie von freier Software unterstützt werden und die auch von kleinen Projekten eingesetzt werden können. Auch das TextGrid-Repositorium unterstützt diese Standards bereits und bietet sich damit als Quelle für die Entwicklung neuer Tools und nachhaltiger elektronischer Publikationen für die Geisteswissenschaften sowohl im Text- als auch im Bildbereich an.

⁹ Siehe <http://openannotation.org>.

Literaturverzeichnis

- Damerow, Peter; Renn, Jürgen (2012): The Equilibrium Controversy. <http://www.edition-open-sources.org/sources/2/index.html>.
- Grassi, Marco; Morbidoni, Christian; Nucci, Michele; Fonda, Simone; Piazza, Francesco (2013): Pundit: Augmenting Web Contents with Semantics. In: *Literary and Linguistic Computing* 28 (4): 640–59. DOI:10.1093/llc/fqt060.
- Hagner, Michael; Rheinberger, Hans-Jörg; Wahrig-Schmidt, Bettina [Editoren] (1994): *Objekte, Differenzen und Konjunkturen: Experimentalsysteme im historischen Kontext*. Berlin: Akademie Verlag.
- Kieven, Elisabeth (2013): Lineamenta – Eine Forschungsdatenbank für Architekturzeichnungen. In: *Zeitenblicke* 2 (1). <http://www.zeitenblicke.de/2003/01/kieven/>.
- Raspe, Martin; Casties, Robert. Digilib (2006): Wissenschaftliches Bildmaterial studieren und kommentieren im Internet. In: *Jahrbuch der Max-Planck-Gesellschaft* 2006. <http://www.mpg.de/411123/pdf.pdf>.
- Renn, Jürgen (2014): The Globalization of Knowledge in History and Its Normative Challenges. In: *Rechtsgeschichte Legal History – Zeitschrift des Max-Planck-Instituts für Europäische Rechtsgeschichte* 22: 52–60.
- Renn, Jürgen; Osthues, Wilhelm; Schlimme, Hermann (Hrsg.) (2014): *Wissensgeschichte der Architektur*, 2. Dezember 2014. <http://www.edition-open-access.de/studies/3/index.html>, <http://www.edition-open-access.de/studies/4/index.html>, <http://www.edition-open-access.de/studies/5/index.html>.
- Warnke, Martin; Brink, Claudia (2008): *Aby Warburg – Gesammelte Schriften – Studienausgabe: Der Bilderatlas MNEMOSYNE*. Gegenüber der 2. unveränderte Auflage. Berlin: Oldenbourg Akademieverlag.

Datenvisualisierung

Geo-Browser und DigiVoy

von Thomas Kollatz und Stefan Schmunk

Visual Turn und Visualisierung

„Nie sah Geisteswissenschaft cooler aus“, schrieb Sebastian Hermann in der Süddeutschen Zeitung vom 1. August 2014 in einem Beitrag mit dem Titel „Europas Besiedlung – Leben und Sterben in 2000 Jahren“ über das Forschungsprojekt des Kunsthistorikers Maximilian Schich an der Universität Texas (Dallas), bei dem 150.000 Geburts- und Sterbeorte von Künstlern, Wissenschaftlern, Politikern und weiteren wissenschaftlich, politisch und kulturell bedeutenden Persönlichkeiten der letzten 2.000 Jahren visualisiert wurden.¹ Im Fokus lagen die Migrations- und Wanderbewegungen und somit die geografische Mobilität der intellektuellen Elite der vergangenen 2.000 Jahre in ihren Raum- und Zeitbezügen. Erstmals ist so eine einzigartige Übersicht von Relationen und Kontextualitäten visuell erfassbar dargestellt worden, deren Ergebnisse nicht nur „cool aussahen“, sondern vor allem einen analytischen wissenschaftlichen Mehrwert generierten, der ohne den Einsatz digitaler Methoden und Verfahren nicht möglich gewesen wäre. Zwar wurde in der Vergangenheit in einer Vielzahl von Einzelstudien die geografische Mobilität von Personen oder spezifischer Gruppen – etwa des deutschsprachigen Exils der 1930er- und 1940er-Jahre – untersucht; eine Analyse von Daten und Datenreihen, die einen längeren Zeitraum oder auch größere heterogene sozio-ökonomische Relationen abdeckten, fehlte hingegen bis dato.

Der vorliegende Aufsatz will deshalb zwei Visualisierungs-Werkzeuge näher vorstellen, die im Rahmen von TextGrid und DARIAH-DE verwendet und teilweise weiterentwickelt werden, sowie deren Anwendungsmöglichkeiten und Chancen für geistes- und kulturwissenschaftliche Fragestellungen skizzieren. Dabei handelt es sich zum einen um DigiVoy, das das TextGrid

¹ <http://www.sueddeutsche.de/wissen/datenvisualisierung-europas-leben-und-sterben-im-zeitraffer-1.2071567>

Repository mit den unter dem Namen „Voyant“ von Stéfan Sinclair und Geoffrey Rockwell bereitgestellten Werkzeugen zur Textanalyse verbindet.² Das zweite Werkzeug ist der Geo-Browser, der zunächst im Rahmen des europeana4D konzipiert wurde und von DARIAH-DE in enger Zusammenarbeit mit anderen Projekten stetig weiterentwickelt wird. DigiVoy und der Geo-Browser sind generische Open-Source-Entwicklungen, die in die Virtuelle Forschungsumgebung von TextGrid eingebettet und stabil und langfristig in die digitale Forschungsinfrastruktur von DARIAH-DE integriert wurden, um als Teil einer digitalen Infrastruktur von Geistes- und KulturwissenschaftlerInnen genutzt werden zu können.

Datenvisualisierung in den Digital Humanities – eine Arbeitsdefinition

Zunächst ist zu klären, was unter Datenvisualisierung und visueller Repräsentation von Forschungsdaten in den Geistes- und Kulturwissenschaften verstanden wird. Der *Visual Turn* führte in den Geistes- und Kulturwissenschaften seit Ende der 1990er-Jahre dazu, dass vermehrt Bilder, Grafiken, Gemälde – kurzum: visuell erfahrbares und analysierbares Material – in den Fokus der Forschung gerieten. Neu war, dass nun die Visualisierung selbst das Forschungsergebnis von Untersuchungen war³ – mit der Folge, dass Erkenntnisse nicht mehr ausschließlich durch die Analyse von Text in neuem Text formuliert wurden und visuelles Material lediglich als illustrierendes Beiwerk diente – Text generiert Text –, sondern Forschungsergebnisse und Erkenntnisgewinn überhaupt erst durch die Visualisierung erreicht wurden (vgl. Haas 2006). Zugleich wurde es technisch möglich, größere und für die Forschung relevante Mengen an Daten analysieren zu können – bedingt einerseits durch die Entwicklung von digitalen Werkzeugen und Softwarelösungen, andererseits durch die seit rund 15 Jahren voranschreitende Digitalisierung von für Geistes- und Kulturwissenschaften forschungsrelevantem Quellen- und Datenmaterial.

Unter der Visualisierung von Daten versteht man in den Digital Humanities einen computergestützten Prozess, mit dessen Hilfe geistes- und kultur-

2 <http://docs.voyant-tools.org>

3 Dieser Ansatz erforderte aber auch ein neues methodisches Instrumentarium in den einzelnen Fachdisziplinen; vgl. Hessler/ Mersch (2009).

wissenschaftliche Daten so dargestellt und analysiert werden können, dass eine visuelle Repräsentation der inhärenten kontextualen bzw. inhaltlichen Zusammenhänge entsteht. Auf diese Weise können insbesondere größere Daten- und Quellenmengen analysiert werden, die von einzelnen ForscherInnen mit klassischen, nicht-digitalen Methoden nicht oder allenfalls nur mit erheblichem Zeit- und Ressourcenaufwand durchgeführt werden könnten. Die Datenvisualisierung kann neben (1) einer Visualisierung aber auch dazu dienen, (2) eine Exploration der Daten (*Visual Analytics*) und deren Abhängigkeiten und Relationen zu ermöglichen und (3) eine Emulation zu erstellen oder direkt Bildmaterial – beispielsweise durch Mustererkennung⁴ – zu analysieren.

Ein gutes Beispiel für Erkenntnisgewinn durch Datenvisualisierung stammt aus dem Bereich der Buch- und Publikationsgeschichte. Gerade bibliografische Daten eignen sich aufgrund ihrer hochgradigen Strukturierung in besonderer Weise für Visualisierungen. Die „Verzeichnisse der im deutschen Sprachraum erschienenen Drucke“, die für das 16., 17. und 18. Jahrhundert erstellt wurden, ermöglichen es erstmals, den Gesamtbestand der Publikationen im deutschen Sprachraum in dieser Periode zu analysieren.

Exemplarisch verdeutlicht der Ländervergleich der jährlichen Buchproduktionszahlen zwischen Deutschland (basierend auf VD 16 und VD 17) und England (basierend auf dem English Short Title Catalogue – ESTC)⁵ das für die geisteswissenschaftliche Forschung im Zusammentreffen von Digitalisierung und technischer Entwicklung liegende Potenzial.

Abbildung 1 zeigt am Beispiel der jährlich publizierten Drucke und Publikationen im deutschen und englischen Sprachraum, wie unter Zuhilfenahme von Verfahren der Informations- und Kommunikationstechniken (ICT) gesellschaftliche Entwicklungen und Veränderungen identifiziert werden können. Die ungeheure Steigerung der Publikationszahlen unmittelbar vor Beginn des Englischen Bürgerkrieges – geschuldet vor allem der politischen Auseinandersetzung zwischen Thron und Parlament und deren öffentlicher Rezeption, die 1649 in der Enthauptung Karls I. gipfelte –, zugleich aber auch der eklatante Rückgang der Publikationszahlen während des Dreißigjäh-

4 Auf diesen Ansatz wird im folgenden nicht näher eingegangen. Ein Beispiel für einen methodischen Ansatz zur Mustererkennung, dessen Potenzial und ein Anwendungsbeispiel liefert das BMBF-geförderte Projekt *eCodicology* (<http://www.ecodicology.org>).

5 Der English Short Title Catalogue beinhaltet sämtliche zwischen 1473 und 1800 in England und Nordamerika gedruckten Werke; siehe <http://estc.ucr.edu>.

rigen Kriegen in Deutschland können so erstmals empirisch valide belegt werden, ebenso wie die starke Abhängigkeit und gegenseitige Bedingtheit von Reformation und Flugblattdruck zu Beginn des 16. Jahrhunderts.

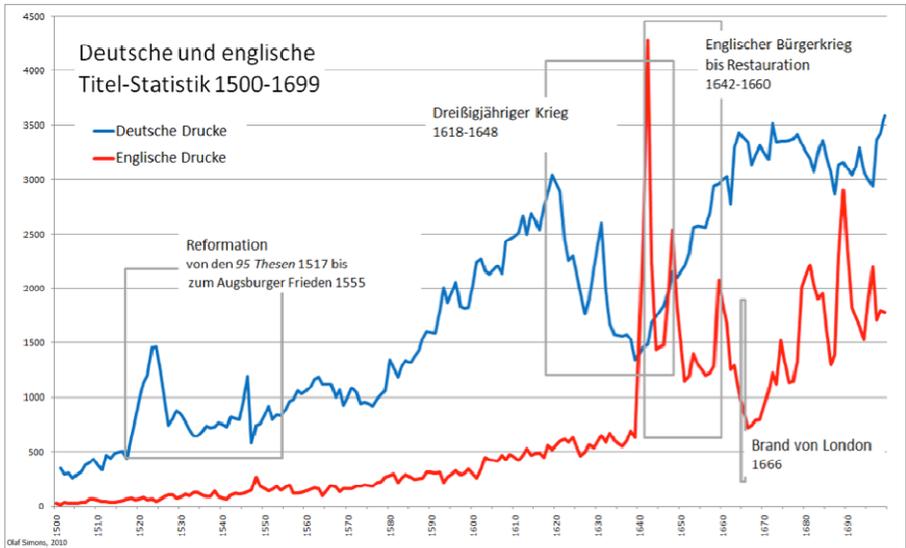


Abb. 1 Deutsche und englische Titel-Statistik, 1500–1699, basierend auf VD16, VD17 und ESTC⁶

Während das Beispiel in Abbildung 1 ausschließlich mit absoluten (Publikations-) Zahlen arbeitet, zeigt die folgende Abbildung 2, welchen weiteren analytischen Mehrwert eine dynamische visuelle Analyse bietet. Die epigraphische Datenbank epidat setzt den DARIAH-DE Geo-Browser zur Visualisierung der Raum-Zeit-Relationen von rund 20.000 datierten Grabmalen auf 145 historischen jüdischen Friedhöfen ein:

⁶ Bildquelle: „Deutsche und englische Titel-Statistik, 1500–1699“ von Olaf Simons – Eigenes Werk. Lizenziert unter Creative Commons Attribution-Share Alike 3.0 über Wikimedia Commons (http://commons.wikimedia.org/wiki/File:Deutsche_und_englische_Titel-Statistik,_1500-1699.png#mediaviewer/File:Deutsche_und_englische_Titel-Statistik,_1500-1699.png)

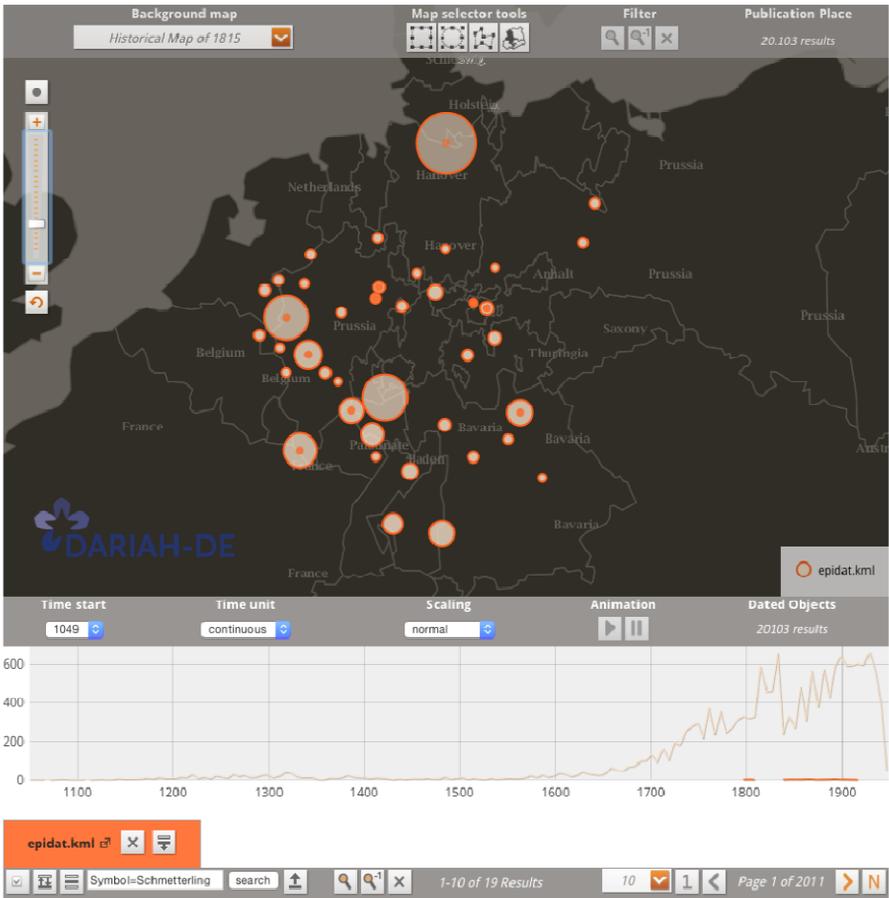


Abb. 2 Datierte jüdische Grabmale in ihren Raum- und Zeitrelationen⁷

Die Abbildung verdeutlicht auf Anhieb, dass das gesuchte Symbol „Schmetterling“ räumlich weitverteilt und zeitlich erstmals um das Jahr 1800 nachzuweisen ist. Die Visualisierung lädt ein zur Hypothesenbildung: Antike Symbolik taucht – vermutlich bedingt durch die Antikenrezeption seit Herder, Lessing und Weimarer Klassik – im jüdischen (wie christlichen) Grab schmuck gegen Ende des 18. Jahrhunderts auf; und da antike Symbolik nicht christlich konnotiert ist, konnte sie allmählich Eingang in die Symbolik jüdischer Sepulkralkultur finden.⁸ Beide Beispiele verdeutlichen, wie große

7 Bildquelle: <http://www.steinheim-institut.de/cgi-bin/epidat?info=geobrowser>

8 Ein Beispiel eines auf einen Ort (Worms) und den Zeitraum vor 1501 begrenztes Symbol gibt Brocke (2011).

Datenmengen durch digitale Tools erschlossen und zugleich durch Visualisierungen analysiert werden können, um auch nicht unmittelbar evidente strukturelle und inhaltliche Zusammenhänge zu erkennen und zu belegen.

DARIAH-DE Geo-Browser und Datasheet-Editor⁹

Wie das Beispiel in Abbildung 2 zeigt, ist eine raum-zeit-bezogene Visualisierung und Analyse von strukturierten Datenmengen geeignet, um Relationen sichtbar zu machen und Kontexte darzustellen. Der DARIAH-DE Geo-Browser¹⁰ besteht aus zwei miteinander korrelierenden Komponenten: einem Browser zur Daten-Visualisierung und dem Datasheet-Editor zur Dateneingabe. Der DARIAH-DE Geo-Browser wird seit 2011 basierend auf den Entwicklungen von europeana4D¹¹ und GeoTemCo¹² weiterentwickelt und steht ForscherInnen der Geistes- und Kulturwissenschaften für die Analyse von Raum-Zeit-Relationen von Daten zur Verfügung. Die SUB Göttingen und weitere Partnereinrichtungen von TextGrid und DARIAH-DE arbeiten bei der Entwicklung eng zusammen und kooperieren in einer im März 2013 gegründeten Arbeitsgruppe, um Entwicklungsprozesse abzustimmen, so z. B. mit VertreterInnen des MPI für Wissenschaftsgeschichte in Berlin, dem Fraunhofer IAIS in Sankt Augustin, der Staatsbibliothek Berlin und der Universität Leipzig, insbesondere mit Stefan Jänicke, der GeoTemCo und europeana4D gemeinsam mit Ralf Stockmann entwickelt hat.¹³ Eine enge Abstimmung und Entwicklungskooperation erfolgt insbesondere mit dem „Atlas der Innovation des Exzellenzcluster 264 – Topoi“¹⁴, dessen Geo-Browser-Modul¹⁵ ebenfalls öffentlich zugänglich ist und nachgenutzt werden kann. DARIAH-DE hat neben einer Umstellung der Codebasis des Geo-Browsers auf JavaScript im Jahr 2013 den Datasheet-Editor in den Jahren 2011 und 2012 zur einfachen Dateneingabe entwickelt und hostet darüber

9 <http://geobrowser.de.dariah.eu>

10 Eine Dokumentation mit Datenbeispielen des Geo-Browsers wird zurzeit erstellt unter: <https://dev2.dariah.eu/wiki/pages/viewpage.action?pageId=30376748>

11 <http://wp1187670.server-he.de/e4d/>

12 <http://www.informatik.uni-leipzig.de:8080/geotemco>

13 ausführlichere Informationen unter <https://de.dariah.eu/geo-browser-impressum>

14 <http://www.topoi.org/group/d-6/>

15 <https://github.com/skruse/PLATIN>

hinaus eine stabile Version des Geo-Browsers für interessierte WissenschaftlerInnen und Forschungsprojekte.

Dateneingabe und Anreicherung im DARIAH-DE Datasheet-Editor

Der Datasheet-Editor¹⁶ bietet NutzerInnen zwei Optionen, eigene Daten für die Visualisierung im Geo-Browser aufzubereiten: zum einen den Import und die anschließende Anreicherung bestehender CSV-Tabellen¹⁷, zum anderen die Direkteingabe raum- und zeitbezogener Daten. In der Regel werden die Datensätze direkt in den von TextGrid und DARIAH-DE gemeinsam genutzten DARIAH-DE-Storage überführt, gesichert und auch dort verwaltet – unabhängig davon, ob sie in den Datasheet-Editor importiert oder von Anfang an darin erstellt wurden. Für einen der kommenden Entwicklungsschritte ist darüber hinaus geplant, dass die Datensätze auch in einem Repository gespeichert und veröffentlicht werden können.¹⁸

Die Option zur Direkteingabe der Daten ist sehr einfach gehalten; erforderlich sind lediglich Orts- und Zeitangaben. Anschließend werden Geolokalisierungen (Längen- und Breitenangaben) unter Verwendung des Getty Thesaurus of Geographic Names (TGN)¹⁹, Open Geo Names (OGN)²⁰ und/oder Open Street Maps (OSM)²¹ (semi)automatisch ergänzt – ein Verfahren, das nicht nur Zeit spart, sondern bei dem die Ortsdaten zugleich mit den fehlenden Längen- und Breitenangaben und den entsprechenden eindeutigen Identifikatoren der Vokabulare angereichert werden. Anschließend können die im Datasheet-Editor angereicherten Daten nicht nur im Geo-Browser visualisiert und analysiert, sondern auch in weiteren Anwendungen nachgenutzt werden. Die Visualisierung direkt aus dem Datasheet-Editor

16 <http://geobrowser.de.dariah.eu/edit/>

17 Comma-separated values, siehe: [http://de.wikipedia.org/wiki/CSV_\(Dateiformat\)](http://de.wikipedia.org/wiki/CSV_(Dateiformat))

18 Vgl. den Beitrag von Stefan Funk und Wolfgang Pempe im vorliegenden Band.

19 <http://www.getty.edu/research/tools/vocabularies/tgn/>. Zur Optimierung des Zugriffs und der Skalierbarkeit bei der Verarbeitung von größeren Datenmengen wird bislang auf einen TGN-Dump zugegriffen, der von TextGrid und DARIAH-DE gehostet wird. Im ersten Quartal 2015 wird dies umgestellt und eine seit Sommer 2014 zugängliche Schnittstelle des TGN direkt abgefragt.

20 <http://www.geonames.org>

21 <http://www.openstreetmap.de>

heraus ermöglicht die Direktkontrolle auf den Karten des Geo-Browsers. Sollten Orte bei der automatischen Georeferenzierung falsch zugeordnet werden – wie etwa bei identischen Ortsnamen in unterschiedlichen Ländern oder Regionen (z.B. Paris/Texas und Paris/Frankreich; Frankfurt am Main und an der Oder) –, kann dies in der Nachbearbeitung leicht korrigiert werden, nämlich wiederum unter Zuhilfenahme der eingebundenen Thesauri. Alternativorte werden in einem Drop-down-Menü angezeigt, aus denen dann die korrekte Ortsangabe samt Koordinaten und Identifikator übernommen werden kann.

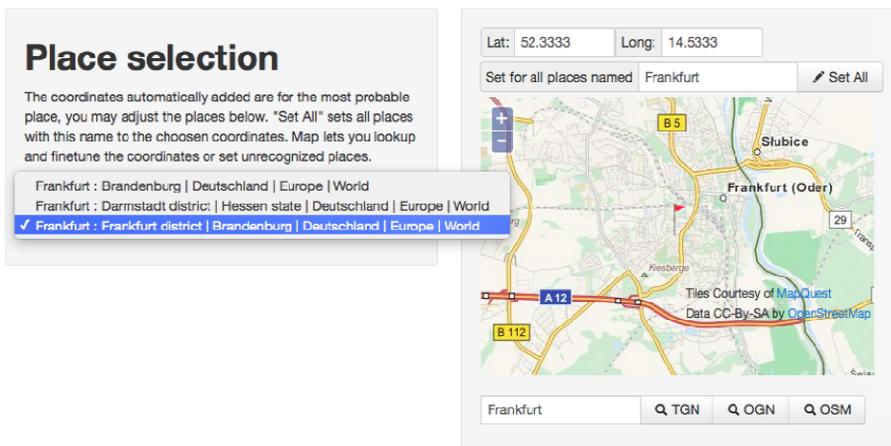


Abb. 3 Semiautomatische Ermittlung von Georeferenzen

Darüber hinaus können über das Kartenmodul zu Orten, die den angebotenen Thesauri unbekannt sind und zu denen diese also auch keine Georeferenzen ausgeben können – beispielsweise Dörfer und Städte, die seit dem Spätmittelalter als Wüstungen untergegangen sind –, durch eine Markierung auf der Karte die entsprechenden Längen- und Breitenangaben identifiziert und festgelegt werden. Damit dient der Datasheet-Editor zudem als Werkzeug, mit dem fachwissenschaftliche Fragestellungen beantwortet und analysiert werden können. Auch wenn es sich bei der Ortsreferenz nicht im eigentlichen Sinn um einen bewohnten Ort handelt, können beispielsweise auch archäologische Fundorte oder historische Bauwerke, etwa der Limes, georeferenziert und im Geo-Browser dargestellt werden (vgl. Romanello 2013).

Datenvisualisierung im DARIAH-DE Geo-Browser

Der Geo-Browser vereint drei korrelierende Elemente: eine Karte, eine Zeitleiste sowie die Dokumentation der visualisierten Datengrundlage. Bei den Karten kann interaktiv zwischen frei wählbaren zeitgenössischen und mehreren historisierenden Karten gewählt werden. Zudem besteht die Möglichkeit, eigenes georeferenziertes Kartenmaterial einzubinden. Das im Geo-Browser hinterlegt Standard-Kartenmaterial deckt einen Zeitraum von über 2.000 Jahren ab, sodass analog zur Periode der Datengrundlage meist eine entsprechende historisierende Karte zugeschaltet werden kann. Datensets aus dem beginnenden 20. Jahrhundert etwa können auf dem entsprechenden historisierenden Kartenmaterial mit der flächen- und grenzgetreuen Staatenwelt am Vorabend des Ersten Weltkrieges dargestellt werden.

Um die Vielzahl der Einzelorte bzw. der entsprechenden Datenmengen zu strukturieren, werden bei der Visualisierung im Geo-Browser Einzeldaten nach Dichte und Quantität zu regionalen Häufungen („heaps“) akkumuliert.

Die zeitliche und quantitative Dimension des Gesamtdatenbestandes wie der je aktuellen Auswahl wird in einem Graph auf der Zeitleiste dargestellt. Wird ein Punkt auf der Karte angesteuert, werden die entsprechenden Punkte der Zeitleiste sowie die Dokumentationsfelder hervorgehoben. Wird ein Zeitpunkt oder eine Zeitspanne auf der Zeitleiste ausgewählt, diese bewegt oder animiert, werden stets die korrelierenden Georeferenzen und Dokumentationsfelder hervorgehoben (siehe Abb. 4 oben).

Jederzeit kann aus dem ausgewählten und markierten Zwischenergebnis ein neues Datensample generiert werden – etwa, um Entwicklungen in unterschiedlichen Zeiträumen oder Regionen miteinander zu vergleichen. Zur lokalen Weiterverarbeitung oder als Grundlage weiterer Visualisierungsschritte kann das Datenset auch als KML-Datei²² exportiert werden (siehe Abb. 4 unten).

22 KML (Keyhole Markup Language) bezeichnet eine Auszeichnungssprache, die u.a. bei Google Earth verwendet wird (<https://support.google.com/mapsengine/answer/2505884?hl=de>). Seit 2008 ist KML vom Open Geospatial Consortium als Standard anerkannt worden.

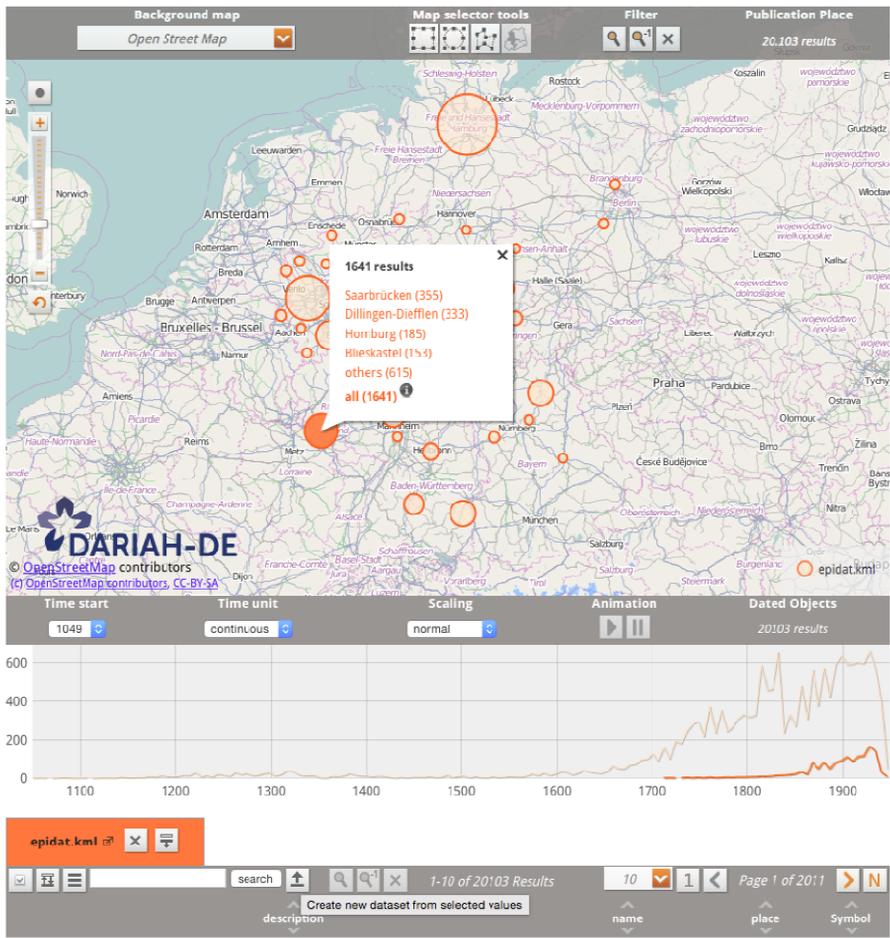


Abb. 4 Akkumulation von georeferenzierten Daten: Bei Auswahl wird die Häufung aufgeschlüsselt und die zeitliche Dimension auf der korrelierten Zeitleiste dargestellt.

Datasheet-Editor und Geo-Browser als Teil der digitalen Infrastruktur

Der Datasheet-Editor wurde von DARIAH-DE aus drei Gründen als generisches digitales Werkzeug auch für TextGrid-Anwendungen entwickelt: Erstens können WissenschaftlerInnen aus den Geistes- und Kulturwissenschaften relativ schnell manuell Datensets auf Basis ihrer eigenen Daten und Quellen zusammenstellen und diese mithilfe kontrollierter Vokabularien georeferenzieren; zweitens können geografische Orte unabhängig von ihrem Typus (Siedlungen, Städte oder beliebige sonstige Orte) mit Längen- und

Breitenabgaben versehen werden; und drittens können einzelne bereits digital vorliegende Datensets und Forschungsdaten wie z.B. das Shakespeare-Album²³ oder der Briefwechsel zwischen Friedrich Carl von Savigny und Johann Heinrich Christian Bang²⁴, aber auch XML-Daten von umfangreichen Editionsprojekten wie z.B. den Notizbüchern Theodor Fontanes²⁵, geolokalisiert und analysiert werden.

Aktuell wird deshalb die technische Verschränkung von TextGrid Repository²⁶ und Geo-Browser weiter ausgebaut. Im TextGrid Repository liegen hochstrukturierte Metadaten in Form des XML/TEI-Headers vor, der unter anderem auch geografische Informationen enthält. Gleiches gilt auch für den XML/TEI-Body, sofern die entsprechenden Informationen ausgezeichnet wurden. Exemplarisch wurde die Generierung von KML aus XML-Headerdaten im TextGrid Repository bereits im Rahmen des Editionsprojekts „Notizbücher Theodor Fontanes“ erfolgreich erprobt – mithin eine Funktionalität, die zukünftig allen Editionsprojekten im TextGrid Repository zur Verfügung stehen wird. Diese Beispiele belegen, dass strukturiert erfasste Forschungsdaten auch unabhängig von ihrem ursprünglichen Erfassungs- und Forschungskontext nachgenutzt werden können. Gerade die raum-zeitbezogene Darstellung und Analyse von Text- und Bildmaterial im Rahmen digitaler Editionen, wie sie der Geo-Browser leistet, wird eine neue qualitative Perspektive bei der Erstellung und wissenschaftlichen Nachnutzung von Editionen eröffnen; sie liefert zugleich ein gutes Beispiel für Synergieeffekte im Rahmen der angestrebten zunehmenden Zusammenführung der Angebote, Dienste und Funktionalitäten von TextGrid und DARIAH-DE.

Die Dienste DARIAH-DE Geo-Browser und Datasheet-Editor greifen im Hintergrund auf mehrere Komponenten des DARIAH-DE-Hosting- und -Service-Environments zu. Für die Weiterentwicklung wurde das Developer-Portal²⁷ verwendet; Geo-Browser und Datasheet-Editor setzen auf den DARIAH-DE-Hostingkomponenten und den operativen IT-Diensten²⁸ auf,

23 <http://www.shakespearealbum.de/es/geobrowser.html>

24 http://www.literatur.hu-berlin.de/aktuelles_idl/konferenzen_tagungen_vortraege/ontologien/poster-teilnehmer/poster_horn.pdf

25 <http://fontane-nb.dariah.eu/tei-conf/index/rom.pdf>

26 <http://www.textgridrep.de>

27 <https://de.dariah.eu/developer-portal>

28 <https://de.dariah.eu/operative-it-dienste>

die erhobenen Forschungsdaten werden in der DARIAH-DE-Storage-Umgebung gespeichert und repliziert.

DigiVoy – Visualisierung von Textdaten

Im Jahr 2012 wurde von DARIAH-DE der Service „DigiVoy²⁹ – TextGrids **DIGITale** Bibliothek mit **VOYant** entdecken“ als Demonstrator entwickelt.³⁰ DigiVoy ermöglicht eine direkte Verbindung zwischen und eine Analyse von im TextGrid Repository gespeicherten Inhalten und Forschungsdaten mit den Voyant-Tools³¹. Dabei handelt es sich um eine Web-Anwendung, mit der explorative Textanalysen des sogenannten *Distant Reading* (vgl. Moretti 2013) oder der Makroanalyse (vgl. Jockers 2013) durchgeführt werden können. DigiVoy beinhaltet unter anderem ein Word-Count-Tool, kann Tag-Clouds generieren, Konkordanzen und Häufigkeiten darstellen und stellt eine Vielzahl weiterer Tools bereit. Insbesondere die umfangreiche *Digitale Bibliothek*³² mit Werken von über 600 Autoren und weitere im Repository hinterlegte Texte bilden eine gute Grundlage, um diese mit verschiedenen Tools quantitativ zu analysieren und die Ergebnisse zu visualisieren. Die Nutzung von DigiVoy lässt sich in drei Schritten differenzieren: 1. Auswahl der zu analysierenden Texte, 2. Vorbereitung der Analyse inklusive Auswahl der zu verwendeten Voyant-Tools und 3. Visualisierung der ausgewählten Texte mit den gewünschten Präprozessierungen.

Beim ersten Schritt können im TextGrid Repository über die Suchfunktion gezielt einzelne Texte ausgewählt oder über die Basket-Funktion ein aus mehreren Texten bestehendes Textsample zusammengestellt werden, um diese in Voyant zu analysieren.

Beim zweiten Schritt – der Präprozessierung, mithin der Vorbereitung der Analyse – können gezielt bestimmte Textbestandteile ausgeschlossen werden; die Metadaten des XML-Headers werden grundsätzlich nicht in die Analyse aufgenommen. Ein- oder ausgeschlossen werden können beispielsweise Namen von Sprechern und von Akteuren in Bühnenwerken, aber auch

29 <https://dariah.zam.kfa-juelich.de/textgridrep-website/>

30 <https://de.dariah.eu/digivoy>

31 <http://voyant-tools.org>

32 Vgl. den Beitrag von Katrin Betz im vorliegenden Band.

Über- und Bildunterschriften.³³ Bei der Präprozessierung handelt es sich um das eigentliche Kernstück der DARIAH-DE- und TextGrid-Entwicklung von DigiVoy, da Texte auf diese Weise unabhängig von ihrem eigentlichen Erschließungs- und Auszeichnungsgrad so vorbereitet werden, dass diese in Voyant analysiert werden können. Hierbei werden diese Daten allerdings nicht angereichert bzw. weitererschlossen, sondern das Ziel ist vielmehr, vorhandene Daten unabhängig von ihrem Erschließungs- und Auszeichnungsgrad für eine Analyse so aufzubereiten, dass die originären Daten nicht verändert werden. Nach TEI-P5-Standard erschlossene Texte, die im TextGrid Repository langzeitarchiviert werden, sind somit zugleich aus dem Repository heraus für weiterführende Analysen verwendbar.

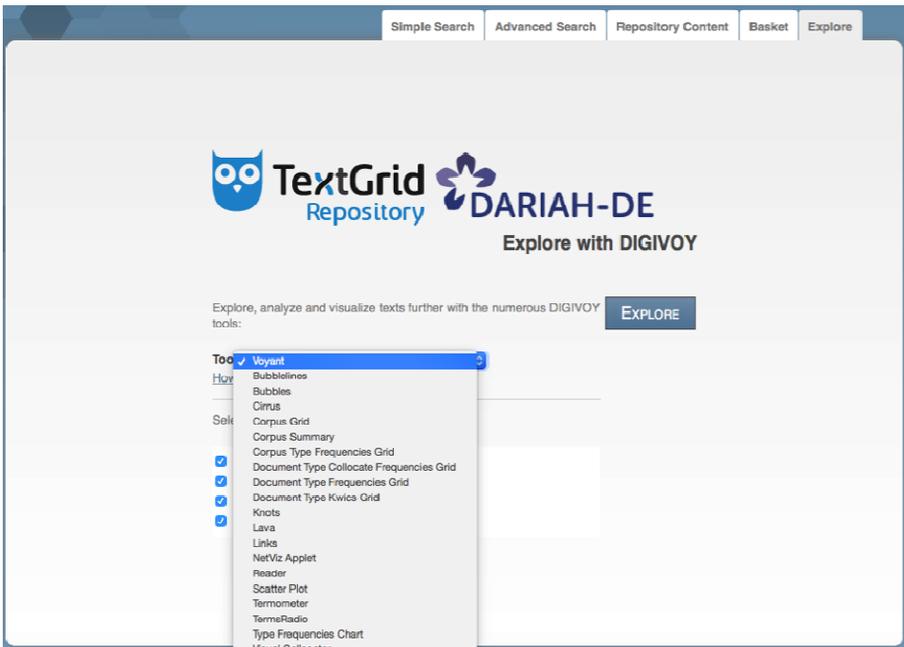


Abb. 5 Suchmaske im TextGrid Repository mit der Möglichkeit, das Suchergebnis mit DigiVoy durch Voyant-Tools zu analysieren

Darüber hinaus können bestimmte Voyant-Tools, wie in Abbildung 5 dargestellt, ausgewählt werden, mit denen dann die eigentliche Analyse in

33 Eine genaue Aufstellung der Textbestandteile, die ausgeschlossen werden können, findet sich unter <https://de.dariah.eu/digivoy>.

Voyant durchgeführt wird.³⁴ Ansonsten wird ein Basis-Toolset verwendet. Die eigentliche Visualisierung und Analyse erfolgt schließlich als dritter und letzter Schritt in Voyant.

Der fachwissenschaftliche Mehrwert von DigiVoy liegt darin, dass unmittelbar auf die Inhalte des TextGrid Repositorys zugegriffen werden kann, die wiederum mit nur geringem Aufwand zu präprozessieren sind; Datenformate müssen nicht verändert werden, um diese mit den angebotenen Tools zu analysieren.

Fazit

Beide hier vorgestellten Werkzeug-Umgebungen eignen sich hervorragend zur Hypothesenbildung und Hypothesenüberprüfung auf breiter Datengrundlage sowie zur forschungsbegleitenden Analyse. Während der Geo-Browser ein Werkzeug im eigentlichen Sinne ist, mit dem Raum-Zeit-Relationen visualisiert und analysiert werden können, handelt es sich bei DigiVoy um einen Service, mit dem Textdaten des TextGrid Repositorys für die quantitativ-visuelle Datenanalyse vorbereitet werden. Beide Tool-Angebote belegen dabei die Stärke standardisiert erfasster Daten (TEI P5). Der Weg zur unabhängigen Nachnutzung durch Dritte ist geebnet, Daten können mit digitalen Tools auch in ganz anderen Erkenntniskontexten analysiert werden. Die stärkere Vernetzung von Forschungsdaten-Repositoryn – erstellt in Virtuellen Forschungsumgebungen und mittels digitaler Forschungsinfrastrukturen, die nachhaltig Möglichkeiten der Nachnutzung schaffen – zeigt das Potenzial dieser Entwicklungen.

Ausblickend sei darauf hingewiesen, dass sich die von DARIAH-DE angebotenen Dienste auch kreativ in ganz eigener Weise nutzen lassen. So wird der im Datasheet-Editor des Geo-Browsers implementierte TGN Rest Service von DARIAH-DE auch als eigenständige Infrastrukturkomponente unterhalten. Diesen Dienst wiederum nutzt die WebApp „Orte jüdischer Geschichte“, um bei der Verifizierung ambivalenter Ortseingaben zu helfen und die entsprechenden Geokoordinaten zu liefern.

34 Eine ausführliche Dokumentation mit Übersicht aller Funktionalitäten bietet Sinclair/Rockwell (o.J.).

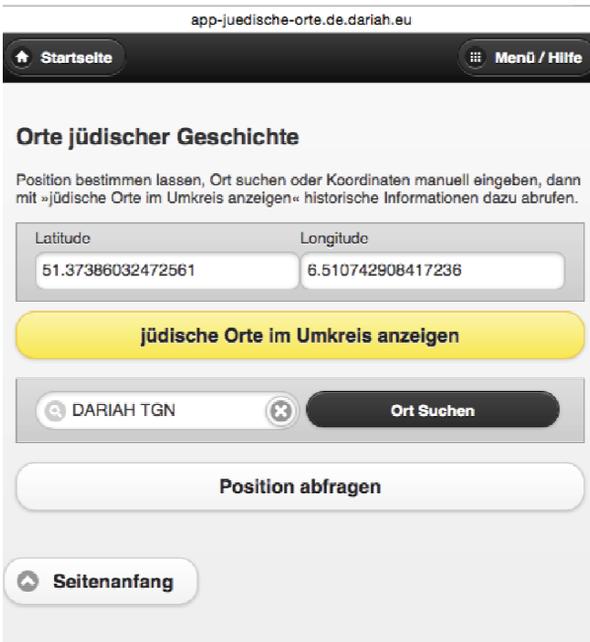


Abb. 6 WebApp „Orte jüdischer Geschichte“

Die WebApp „Orte jüdischer Geschichte“³⁵ bietet nicht nur Wissenschaftlern die Möglichkeit, bereits erschlossene digitale Forschungsdaten – unabhängig ihrer Provenienz und ihres Speicherortes – zu finden und zu analysieren und unter Umständen für eigene Forschungsfragen zu nutzen. In diesem Sinne dienen auch die vorgestellten Werkzeuge nicht nur der Visualisierung, sondern ermöglichen zugleich neue Möglichkeiten der offenen und nachhaltigen Publikation von Forschungsergebnissen und Erkenntnissen.

Literaturverzeichnis

Brocke, Michael (2011): The lilies of Worms. In: *Zutot* 8 (1): 3–13.

Craig, Hugh (2004): Stylistic Analysis and Authorship Studies, in: Schreibman, Susan; Siemens, Ray; Unsworth, John (Hrsg.): *A Companion to Digital Humanities*. Oxford: Blackwell, <http://www.digitalhumanities.org/companion/>.

35 Eine ausführliche Beschreibung der App – <http://app-juedische-orte.de.dariah.eu> – findet sich bei <https://de.dariah.eu/web-app-orte-juedischer-geschichte> sowie bei Lordick (2014).

- Elliot, Tom; Gillies, Sean (2009): Digital Geography and Classics. In: *Digital Humanities Quarterly* 3 (1), <http://www.digitalhumanities.org/dhq/vol/3/1/000031/000031.html>.
- Franco Moretti: *Graphs, Maps, Trees: Abstract Models for Literary History*, London/New York 2005.
- Haas, Stefan (2006): Vom Schreiben in Bildern. Visualität, Narrativität und digitale Medien in den historischen Wissenschaften. In: *Zeitenblicke* 5 (3), http://www.zeitenblicke.de/2006/3/Haas/index_html, URN: urn:nbn:de:0009-9-6482.
- Hermann, Sebastian (2014): Europas Besiedlung – Leben und Sterben in 2000 Jahren. In: *Süddeutsche Zeitung* vom 1. August 2014, <http://www.sueddeutsche.de/wissen/datenvisualisierung-europas-leben-und-sterben-im-zeitraffer-1.2071567>.
- Hessler, Martina; Mersch, Dieter (2009): *Logik des Bildlichen. Zur Kritik der ikonischen Vernunft*. Bielefeld.
- Jockers, Matthew L. (2013): *Macroanalysis: Digital Methods and Literary History*. University of Illinois.
- Kohle, Hubertus (2013): *Digitale Bildwissenschaft*. Glückstadt.
- Krause, Celia; Reiche, Ruth (2013): Ein Bild sagt mehr als tausend Pixel? Über den Einsatz digitaler Methoden in den Bild- und Objektwissenschaften. In: *Kunstgeschichte Open Peer Reviewed Journal*, http://www.kunstgeschichte-ejournal.net/354/1/EinBildsagtmehralstausendPixel_CK_RR.pdf.
- Lordick, Harald (2014): Smartphone-Web-App „Orte jüdischer Geschichte“, <http://djgd.hypotheses.org/36>.
- Moretti, Franco (2013): *Distant Reading*, New York.
- Romanello, Matteo (2013): DARIAH Geo-Browser: Exploring Data through Time and Space. http://upload.wikimedia.org/wikipedia/commons/1/1f/Wiki-data_trifft_Archäologie_-_DARIAH_Geo-browser_-_Romanello.pdf.
- Schich, Maximilian; Song, Chaoming; Ahn, Yong-Yeol; Mirsky, Alexander; Martino, Mauro; Barabási, Albert-László; Helbing, Dirk (2014): A network of cultural history. In: *Science* 345 (6196): 558–562.
- Sinclair, Stefan; Rockwell, Geoffrey (o.J.): *A Practical Guide to Text Analysis with Voyant*, <http://docs.voyant-tools.org>.
- Telea, Alexandra T. (2014): *Data Visualization: Principles and Practice*. Boca Raton, http://books.google.de/books?id=2WoLBAAAQBAJ&printsec=frontcover&dq=telea+data+visualisation&hl=de&sa=X&ei=73ZDVIP_LMOgPbKugJgF&ved=0CDUQ6AEwAA#v=onepage&q=telea%20data%20visualisation&f=false.
- Thomas, James T.; Cook, Kristin A. (Hrsg.) (2005): *Illuminating the Path. The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Center, http://vis.pnnl.gov/pdf/RD_Agenda_VisualAnalytics.pdf.

IV.

Store it, share it, use it:
**Forschungsdaten archivieren
und zugänglich machen**

Hochverfügbarkeit und Performanz: Beiträge zum langfristigen Betrieb von TextGrid

von Philipp Wieder¹

TextGrid ist für viele seit Jahren ein Begriff, wenn es um digitales Forschen, Edieren und Archivieren geht, und das TextGrid Laboratory² ist der Einstiegspunkt für die Nutzung der vielfältigen Angebote, die TextGrid ausmachen. Dabei sind die eigentlichen Software- und Infrastrukturkomponenten – mithin das, was sich „hinter den Kulissen“ befindet und gerne die TextGrid-Middleware genannt wird³ – weitestgehend unsichtbar. Wie bei den meisten IT-Diensten, die wir tagtäglich nutzen, ist dies gewollt. Das Gros der Nutzer interessiert sich weit mehr für ansprechende Benutzerführung, Dienste, die immer erreichbar sind, und kurze Antwortzeiten als für die verwendeten Technologien, Sicherheitsfeatures oder Server-Standorte. Bei Diensten wie E-Mail oder Suchmaschinen ist das offensichtlich: Niemand will nur zu Bürozeiten E-Mails verschicken können oder zwei Minuten auf die Antwort zu einer Suchanfrage warten. Aber auch bei einem wissenschaftlichen Dienst wie TextGrid erwartet mittlerweile die Nutzer-Community, dass der Dienst möglichst immer nutzbar ist, sodass Fristen bei wissenschaftlichen Projekten, Arbeiten oder Veröffentlichungen eingehalten werden können, die Durchführung von Schulungen oder der Zugriff auf relevante Daten jederzeit möglich ist.

Im vorherigen Abschnitt wurden drei der wichtigsten Anforderungen an IT-Dienste bereits genannt: (i) einfache Nutzbarkeit, (ii) hohe Verfügbarkeit

1 An der Entwicklung und Realisierung der TextGrid-Infrastruktur haben in zehn Jahren zahlreiche Personen mitgewirkt. Stellvertretend sollen hier jene genannt werden, die während der dritten Förderphase von TextGrid für die aktuelle Architektur und Infrastruktur verantwortlich zeichneten: Fatih Berber, Maximilian Brodhun, Stefan Funk, Peter Gietz, Martin Haase, Oliver Schmitt, Ubbo Veenster und Thorsten Vitt. Der Autor möchte allen für ihre Beiträge zu TextGrid danken und nicht unerwähnt lassen, dass hier nur in wenigen Worten zusammengefasst wird, woran andere viele Jahre gearbeitet haben.

2 Vgl. hierzu den Beitrag von Mathias Göbel im vorliegenden Band.

3 Vgl. hierzu die Beiträge von Stefan Funk, Wolfgang Pempe und Peter Gietz im vorliegenden Band.

und (iii) gute Performanz. Im Vergleich zur Industrie hat sich im akademischen Umfeld die Erkenntnis, dass Angebote an Wissenschaftlerinnen und Wissenschaftler diese Anforderungen erfüllen sollten, noch nicht überall durchgesetzt. Allerdings sind mittlerweile fast alle wissenschaftlichen Disziplinen immer stärker auf IT-Infrastrukturen angewiesen, sodass ein Umdenken, sofern es noch nicht eingesetzt hat, vermutlich nicht allzu lange auf sich warten lassen wird. Von der Erkenntnis zur Umsetzung struktureller, organisatorischer und technischer Maßnahmen, die die Bereitstellung von IT-Diensten gemäß der genannten Anforderungen zulassen, ist es allerdings ein langer Weg. TextGrid ist hier Vorreiter und ein gutes Beispiel dafür, was zu leisten ist, um für eine wissenschaftliche Community einen Dienst anzubieten, der höchsten Anforderungen genügt und gleichzeitig auf nachhaltigen Betrieb ausgelegt ist.⁴

Dieser Beitrag befasst sich mit den technischen Maßnahmen, die im Rahmen von TextGrid umgesetzt wurden, um die Verfügbarkeit und die Performanz des Dienstes sicherzustellen. Dabei wird auf eine ausführliche Vorstellung der einzelnen TextGrid-Komponenten verzichtet; diese sind umfassend dokumentiert.⁵ Vielmehr beschreibt dieser Beitrag die dem Hochverfügbarkeitskonzept von TextGrid zugrunde liegende Blaupause sowie eine Beispielkomponente, die maßgeblich zu einer guten Performanz des Gesamtsystems beiträgt.

Das Hochverfügbarkeitskonzept von TextGrid

Glücklicherweise ist die Vorstellung, dass wissenschaftliche Infrastrukturen wie TextGrid auf einem Personal Computer unter dem Schreibtisch eines Entwicklers betrieben werden, nicht mehr allzu weit verbreitet. Allerdings ist der tatsächliche Aufwand, der notwendig ist, um ein System wie TextGrid hochverfügbar zu betreiben, häufig außerhalb der Gruppe der Software-Entwickler und IT-Spezialisten nicht bekannt. Daher widmet sich dieser Abschnitt dem Hochverfügbarkeitskonzept, das dem Betrieb von TextGrid zugrunde liegt.

4 Vgl. hierzu die Beiträge von Heike Neuroth sowie von Andreas Witt et al. im vorliegenden Band.

5 Siehe Anhang „TextGrid im Netz“.

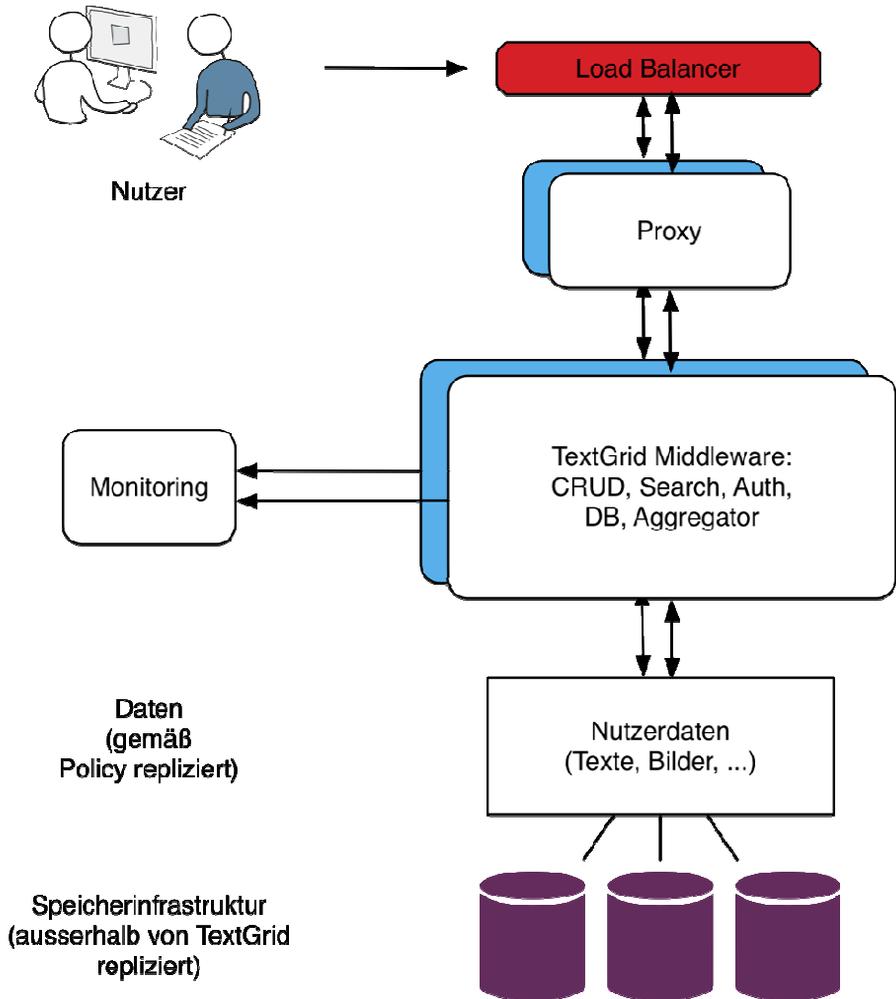


Abb. 1 Schematische Darstellung des Hochverfügbarkeitskonzeptes von TextGrid

Aus Abbildung 1 wird ersichtlich, dass für die Betrachtung der Hochverfügbarkeit grundsätzlich zwei Hauptaspekte ausschlaggebend sind: Daten und Dienste. Was die Daten betrifft, so sind im unteren Bereich der Abbildung die Nutzerdaten sowie die Speicherinfrastruktur repräsentiert. Die unpublizierten und publizierten Daten werden von den TextGrid-Nutzern erzeugt, bearbeitet und schließlich veröffentlicht. Die TextGrid-Middleware legt die Daten dann in einer Speicherinfrastruktur ab. Diese Daten stellen den wichtigsten Wert sowohl für die Nutzer als auch für die Betreiber von TextGrid dar. Zum einen müssen die Daten bestmöglich vor Verlust ge-

schützt werden; zum anderen soll der Zugriff auf sie zu jeder Zeit möglich sein.⁶ Um eine hohe Verfügbarkeit zu gewährleisten, ist es notwendig, dass beim Wegfall von Teilen der unterliegenden Speicherinfrastruktur weiterhin der Datenzugriff möglich ist. Dies wird durch die Replikation der Daten mittels Policies erreicht. Dabei handelt es sich um Regeln, die besagen, welche Daten wie repliziert werden sollen. So werden für gewöhnlich alle Daten an zwei geografisch unterschiedlichen Standorten gespeichert, um ihre Verfügbarkeit zum Beispiel auch im Falle eines Stromausfalles an einem Standort zu gewährleisten. Soll eine noch höhere Verfügbarkeit sichergestellt werden, so können weitere Standorte hinzugenommen werden. Auch die Dauer der Datenspeicherung, gemäß der DFG-Empfehlung (vgl. DFG 2013: 21 ff.) zum Beispiel mindestens zehn Jahre, kann mittels einer Policy definiert werden. Weitere Strategien zur Erhöhung der Verfügbarkeit werden in der Regel auch auf der Ebene der Rechenzentren, die die Speicherinfrastruktur für TextGrid zur Verfügung stellen, umgesetzt. So führt heutzutage der Ausfall einer einzelnen Festplatte normalerweise nicht mehr dazu, dass Daten nicht verfügbar sind.

Auf Dienstebene wird eine höhere Verfügbarkeit durch den Betrieb von mehreren – aktuell zwei – Instanzen der verschiedenen Dienste, die TextGrid ausmachen, erreicht. Dazu werden über einen Load-Balancer Anfragen der Nutzer an verschiedene Proxys verteilt, die diese dann zur Verarbeitung an die jeweilige Instanz der TextGrid-Middleware weiterreichen. Auf diese Weise kann das System den Ausfall einer Instanz kompensieren, indem der Load-Balancer in einem solchen Fall Anfragen nur an die funktionstüchtige Middleware weitergibt. Durch das Hinzufügen weiterer Instanzen kann die Verfügbarkeit zusätzlich erhöht werden. Ergänzt wird die TextGrid-Infrastruktur durch ein Monitoring-System, welches in die Überwachung bei den jeweiligen Rechenzentren integriert ist und es erlaubt, auf den Ausfall einzelner Instanzen zeitnah zu reagieren.

Die Umsetzung des beschriebenen Konzeptes erfordert den Betrieb mehrerer Server sowie die Bereitstellung der notwendigen Speicherinfrastrukturen. Es ist offensichtlich, dass dies im Vergleich zu einem nicht replizierten Betrieb einer TextGrid-Instanz höhere Investitions- und Betriebskosten bedingt. Mit jeder weiteren Instanz und jedem weiteren Replikat der Daten steigen die Kosten entsprechend. Mit Sicht auf die Verfügbarkeit ist eine

6 Auf das Backup von Daten oder andere Maßnahmen zum Schutz vor Datenverlust wird an diese Stelle nicht eingegangen.

derartige Investition zwar wünschenswert; in der Regel wird aber, wie auch bei TextGrid, eine Kosten-Nutzen-Rechnung aufgestellt und im Rahmen einer Analyse das tragbare Risiko definiert. Eine hundertprozentige Verfügbarkeit ist dabei in den seltensten Fällen erforderlich oder finanzierbar.

Zusätzlich zu den Infrastrukturmaßnahmen erfordert der parallele Betrieb mehrerer Instanzen der TextGrid-Middleware auch Software-Entwicklungen und damit weitere Kosten. So muss auf verschiedenen Ebenen der TextGrid-Middleware – wie beispielsweise TG-crud oder TG-auth*⁷ – sichergestellt werden, dass die notwendigen Daten synchron gehalten werden und eine Instanz den Ausfall einer anderen kompensieren kann. Offensichtlich wird das beispielsweise im Falle von TG-auth*, wo Nutzer und deren Rechte in beiden Instanzen identisch hinterlegt sein müssen, damit die Verfügbarkeit von TextGrid für den einzelnen Nutzer nicht von der Middleware-Instanz abhängig ist, die der Load-Balancer zuweist. Um dies für alle Funktionen von TextGrid und deren Zusammenspiel zu gewährleisten, waren und sind eine Vielzahl von Entwicklungen und entsprechende Investitionen notwendig.

Verbesserung der Performanz – Fallbeispiel Suche

Die TextGrid-Middleware wird, wie heute bei der Softwareentwicklung üblich, regelmäßig Tests unterzogen. Dabei wird neben der Prüfung, ob die Software fehlerfrei arbeitet, auch die Performanz getestet und evaluiert, ob es Möglichkeiten der Verbesserung gibt. Im Laufe der dritten Förderphase von TextGrid wurde eine solche Evaluierung systematisch für die gesamte Middleware durchgeführt und es wurden eine Reihe von Verbesserungsmöglichkeiten herausgearbeitet. In diesem Abschnitt werden die Suchfunktion von TextGrid, genannt TG-search, genauer betrachtet und die umgesetzten Verbesserungen, bezogen auf die Performanz des Gesamtsystems, erläutert.

TG-search ist ein Kerndienst von TextGrid, welcher die Suche über alle Daten erlaubt. So ist die Volltextsuche über alle TEI-Daten ebenso möglich wie die Suche nach Metadaten. Vor 2013 griff TG-search dazu auf eine eXist-db-Datenbank⁸ zu. Diese Lösung verursachte häufiger Probleme, beispielsweise bei der Re-Indizierung oder Restaurierung des Datenbestandes;

⁷ Vgl. hierzu den Beitrag von Stefan Funk und Wolfgang Pempe im vorliegenden Band.

⁸ <http://exist-db.org/exist/apps/homepage/index.html>

zudem wurde die Performanz der Suche als nicht ausreichend angesehen. Eine Auswertung verfügbarer Alternativen führte zu der Entscheidung, das Open-Source-Produkt *elasticsearch*⁹ zu verwenden.

Bei *elasticsearch* handelt es sich um eine verteilte Such- und Analyseplattform für große Datenbestände, die die Anforderungen von TextGrid an eine Suchtechnologie erfüllt. Es verfügt über verschiedene Erweiterungen und Schnittstellen zur Anbindung von Datenbanken, webbasierten Diensten und Analysefunktionen. Dabei ist die Realisierung komplexer Suchanfragen mit Facettierung genauso möglich wie die Volltextsuche in bekannten Dateiformaten wie XML, JSON, CSV, PDF oder Microsoft-Office-Dokumenten. *Elasticsearch* folgt bei der Indexierung von Daten und Metadaten dem aus dem NoSQL-Bereich bekannten Ansatz JSON-formatierter Dokumente. Um Lasten zu verteilen und die Aufgabenbearbeitung zu beschleunigen, kann *elasticsearch* flexibel skaliert werden, sodass sich auch TG-search in das Hochverfügbarkeitskonzept von TextGrid integrieren lässt.

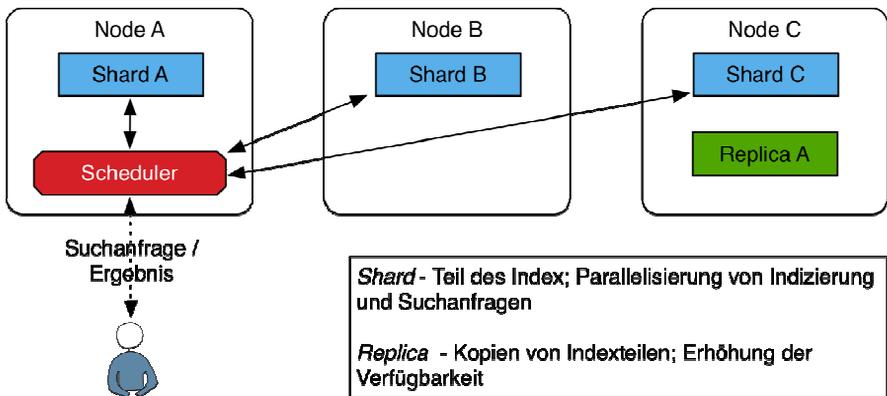


Abb. 2 Clusterbetrieb von *elasticsearch*

Abbildung 2 zeigt schematisch, wie *elasticsearch* die gewünschte Verbesserung der Performanz ermöglicht. Dazu werden in einem *elasticsearch*-Cluster eine Reihe von Nodes betrieben, die jeweils Teile des Index beinhalten. Dadurch können Suchanfragen parallel abgearbeitet werden, indem der Scheduler diese entsprechend verteilt. Zudem sieht *elasticsearch* die Verwendung von Replicas vor, die im Falle von Ausfällen einzelner Knoten die

9 <http://www.elasticsearch.org/>

jeweiligen Suchanfragen bedienen können.¹⁰ Durch den Einsatz von elasticsearch, welches in die aktuelle Version von TextGrid integriert ist, konnten die Antwortzeiten für Suchanfragen und damit die Performanz signifikant verbessert werden. Zudem fügt sich das Cluster-Konzept von elasticsearch gut in das Hochverfügbarkeitskonzept von TextGrid ein.

Resümee und Zukunftsperspektiven

Die Entwicklung von IT-Diensten hat in den letzten zwanzig Jahren eine rasante Entwicklung erlebt, die neben wachsenden Anforderungen seitens der Nutzer auch zu einem nahezu unübersichtlichen Angebot geführt hat. Cloud-Dienste, Big Data, Internet of Things und Quantified Self stellen dabei nur den aktuellen Zwischenstand einer Evolution dar, die in Richtung allgegenwärtiger und jederzeit verfügbarer IT-Dienste weiter voranschreitet. Auch wenn die Entwicklungsgeschwindigkeit im akademischen Umfeld, sicherlich abhängig von der jeweiligen Disziplin, in der Regel langsamer erfolgt als in der kommerziellen Landschaft, so setzen sich auch dort etablierte Entwicklungsmethoden, Testverfahren und Software-Produkte durch. TextGrid ist ein gutes Beispiel einer solchen akademisch getriebenen IT-Infrastruktur, die sich kontinuierlich weiterentwickelt und Forschenden hochverfügbare und leistungsstarke Dienste bereitstellt.

Die Weiterentwicklung von TextGrid erfordert den kontinuierlichen Abgleich mit den Anforderungen der wissenschaftlichen Nutzer, die bei ihrer Arbeit in wachsendem Maße die gleiche Dienstgüte erwarten, wie bei der Nutzung von E-Mail, Cloud-Speichern oder sozialen Netzwerken. TextGrid erfüllt diese Anforderungen bezüglich einer Vielzahl von Kriterien; exemplarisch wurden in diesem Beitrag Verfügbarkeit und Performanz betrachtet. Dazu kommen weitere Maßnahmen wie die Integration in das Monitoring der beteiligten Rechenzentren oder die systematische Erfassung und Abarbeitung von Nutzeranfragen, die das Angebot abrunden.

An dieser Stelle soll nicht unerwähnt bleiben, dass der langfristige Betrieb einer Infrastruktur wie TextGrid im akademischen Umfeld nicht selbstverständlich ist. Das hat eine Reihe von Gründen, auf die hier nicht näher einge-

¹⁰ In Abb. 2 ist hier exemplarisch ein Replica des Shard A in Node C gezeigt, welches einen Ausfall des Node A kompensieren kann. Auf die Abbildung der komplexeren Standardkonfiguration eines elasticsearch-Clusters wurde bewusst verzichtet; hier sei auf elasticsearch (2015) verwiesen.

gangen werden kann. Für diejenigen, die mit der wissenschaftlichen Landschaft vertraut sind, seien exemplarisch Drittmittelfinanzierung und das Hochschulrahmengesetz als Hindernisse für einen nachhaltigen Betrieb und die langfristige Beschäftigung qualifizierter Entwickler genannt. Den an TextGrid Beteiligten ist es trotz dieser Widrigkeiten gelungen, für ihre Community eine performante und hochverfügbare Lösung zu entwickeln und damit eine hervorragende Basis für zukünftige Entwicklungen und die Erfüllung kommender Anforderungen zu schaffen.

Literaturverzeichnis

DFG (2013): *Sicherung guter wissenschaftlicher Praxis. Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“*. Weinheim: Wiley-VCH.

Vom Konzept zur Umsetzung – Einblicke in die Entstehung des TextGrid Repository

von *Stefan E. Funk und Wolfgang Pempe*

Im Rahmen eines Jubiläumsbandes bietet es sich an, einen Moment innezuhalten und die Ereignisse, Diskussionen und Entwicklungen der letzten zehn Jahre Revue passieren zu lassen. In diesem Zusammenhang möchten wir anhand einiger Bilder aus dem TextGrid-Familienalbum die Entstehungsgeschichte der Architektur des TextGrid Repository¹ illustrieren.

Lehnen Sie sich also mit uns zurück und folgen Sie uns auf der Reise durch die Vergangenheit des TextGrid Repository ...

„Das Ding“ und TG-crud

Eine der ersten Entscheidungen, die in TextGrid getroffen wurden, war die Etablierung zweier Arbeitspaket-übergreifender Arbeitsgruppen, der AG Textformate (Schnittstelle Fachwissenschaft – Technik) und der AG Architektur. Letztere dient als Kommunikationsforum für alle am Projekt beteiligten Technikerinnen und Techniker und war stets der Ort, an dem die Konzeptionsarbeit zur technischen Infrastruktur geleistet wurde. Dies geschieht damals wie heute vornehmlich bei Face-to-Face-Meetings und Programmiersprints, in deren Rahmen ein bestimmtes Schwerpunktthema behandelt wird.

Die Entwicklung der technischen Infrastruktur zu der seit 2011 annähernd unverändert in Betrieb befindlichen Lösung war ein langwieriger und iterativer Prozess, in dessen Folge nicht nur zahlreiche Modelle erarbeitet und teilweise wieder verworfen, Software erprobt und Prototypen entwickelt, sondern auch im Rahmen von regelmäßigen Treffen viele hundert Liter Kaffee sowie nicht unerhebliche Mengen an konzentrationsfördernden Snacks (Kekse, Schokolade, Chips)² sowie bisweilen auch vitaminreicher Kost (Äpfel,

1 <http://www.textgridrep.org>; siehe auch die entsprechenden Reports auf <http://textgrid.de/ueber-textgrid/materialien/antraege-und-berichte/>.

2 ... sofern Wolfgang Pempe mit der Beschaffung betraut war

Weintrauben, Bananen)³ konsumiert wurden – nicht zu vergessen die anschließenden abendlichen Restaurantbesuche, bei denen die losen Enden des Arbeitstages üblicherweise wieder aufgenommen und weitergesponnen wurden.

So fand sich auch Anfang 2007 eine Gruppe zu einem Middleware-Treffen in Tübingen zusammen, in dessen Verlauf das in Abbildung 1 dargestellte Tafelbild entstand.

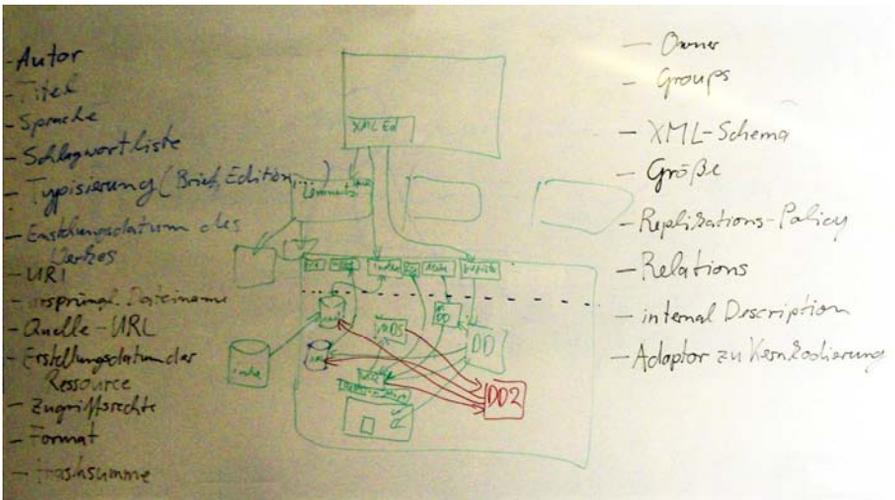


Abb. 1 Tafelbild „Das Ding“ (DD)

Was hier noch ein wenig wirr und für nicht Eingeweihte wenig erhellend wirkt, präsentiert sich wenig später in einer (internen) Präsentation schon sehr viel übersichtlicher (siehe Abb. 2).

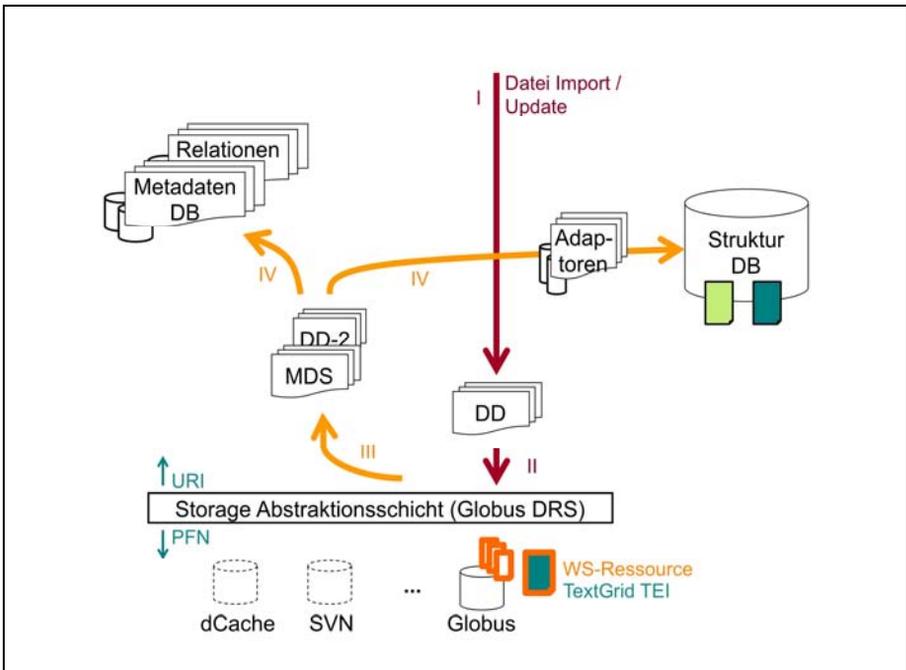


Abb. 2 Am Anfang war „Das Ding“

Der Name „Das Ding“ (DD) bezeichnet dabei eine Komponente, deren umfassende Funktion soweit klar war, deren konkrete Implementierung aber noch ausstand. Tatsächlich gab es zwei solcher Komponenten, „Das Ding“ und „Das Ding 2“ (DD-2). Hier ein kurzer Überblick über den angedachten Ablauf:

- I. Einspielen von Daten zunächst über „Das Ding“ direkt in den Grid-Storage, der
- II. über eine Storage-Abstraktionsschicht angesprochen wird.
- III. Über diese Abstraktionsschicht wird dann der Monitoring & Discovery Service (MDS) des Globus-Toolkit benachrichtigt, der über verfügbare Grid-Ressourcen und deren Status Aufschluss gibt.
- IV. „Das Ding 2“ bekommt über den MDS mitgeteilt, dass es neue (oder veränderte) Ressourcen gibt und verteilt die Daten sodann in die XML- und RDF-Datenbanken für Metadaten und Relationen und nutzt diverse „Adaptoren“, um die Daten (Volltexte) zusätzlich in eine Strukturdatenbank, ebenfalls XML, einzuspielen.

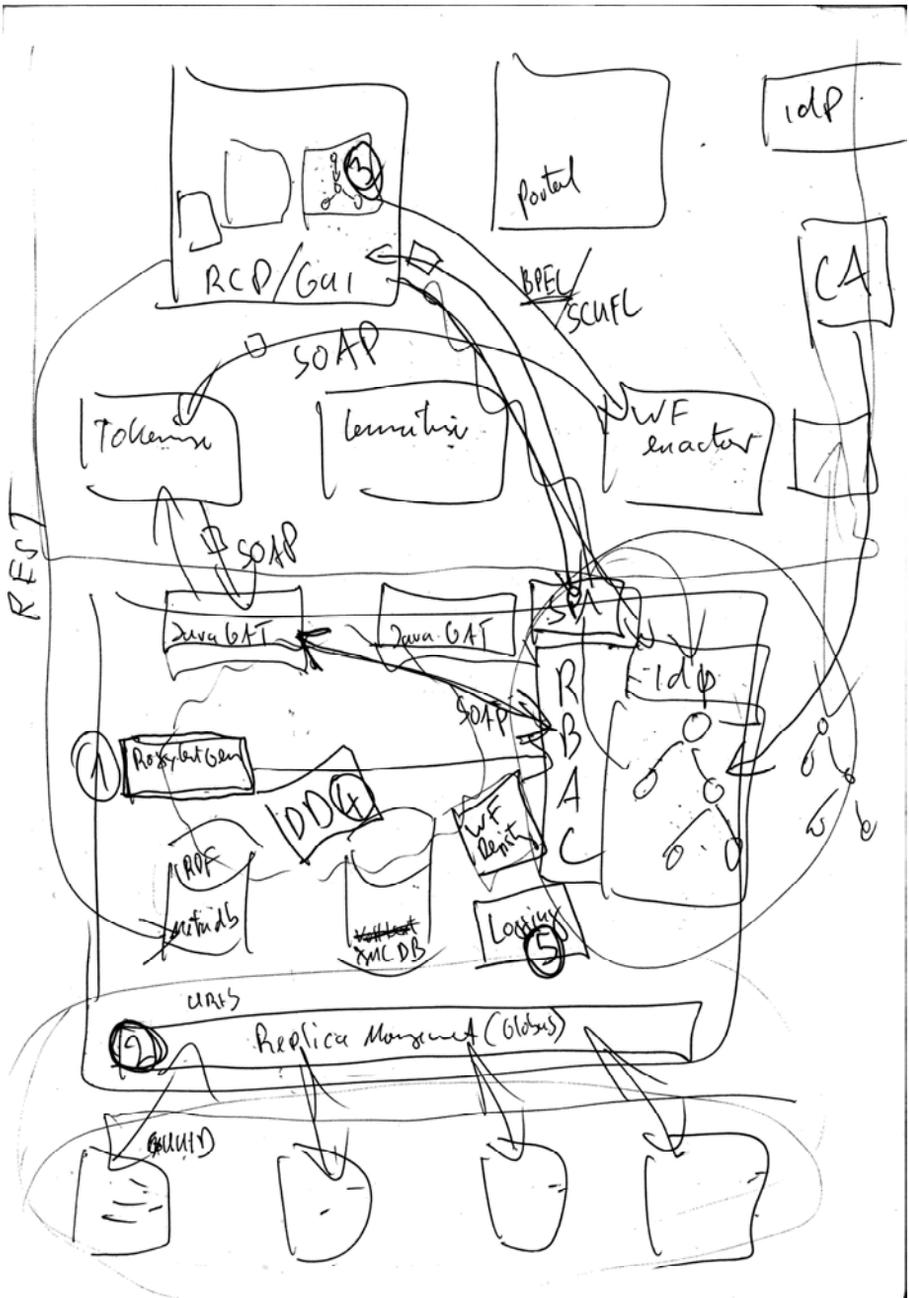


Abb. 3 „TextGrid-Geskribbel“

Mit der Skizze (Abb. 3) sollte einem der Autoren der Einstieg in die TextGrid-Architektur erleichtert werden, der erst in der zweiten Jahreshälfte des Jahres 2007 in das TextGrid-Team eingestiegen war und der die Arbeiten an der Middleware-Implementation fortsetzen sollte. Hierbei ging es in erster Linie um die Grid-Anbindung und die (Weiter-)Entwicklung des „Dings“. Das TextGridLab⁴ steckte noch in den Kinderschuhen: Obwohl schon mit der Eclipse-RCP implementiert, war die Anbindung der Middleware bei weitem noch nicht abgeschlossen. Es gab bereits einige Grid-Speicher-Services – die Fileservices, die mithilfe von JavaGAT auf das Grid zugegriffen –, sowie XML- und RDF-Datenbanken, die jedoch noch nicht umfassend implementiert und erst später von einem zentralen Service angebounden wurden. Die Skizze war tatsächlich sehr hilfreich, denn sie wurde sozusagen während des Zeichnens erklärt und so konnte der Aufbau der TextGrid-Architektur sinnvoll dargestellt und erläutert werden.

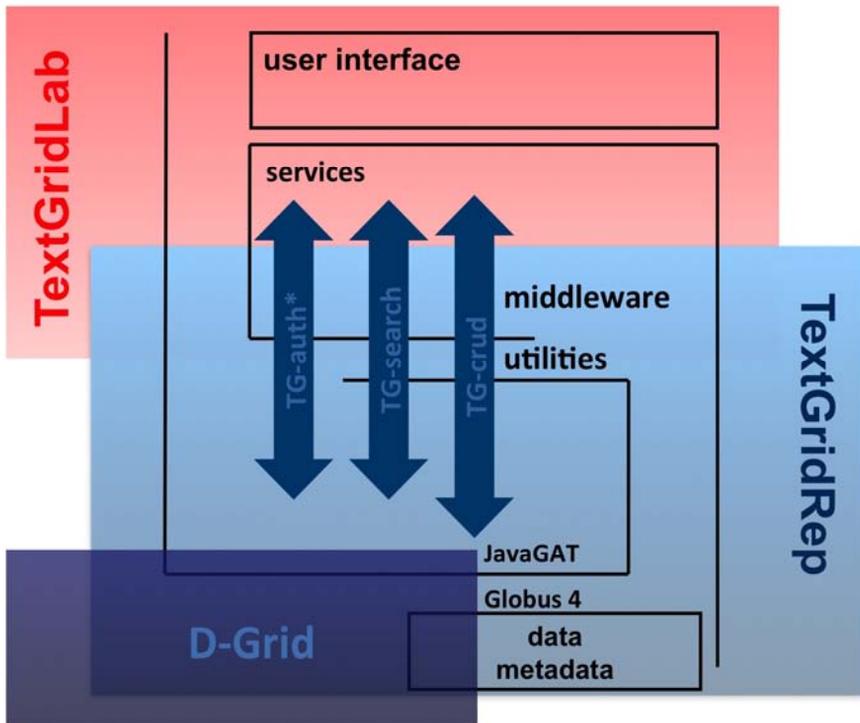


Abb. 4 Die drei Säulen der TextGrid-Middleware

⁴ Vgl. TextGrid – Download und Installation, <http://textgrid.de/registrierungdownload/download-und-installation/>.

Das war der Anfang des in der Folgezeit entwickelten TextGrid-crud-Dienstes, der zunächst grundlegende Funktionen wie das Anlegen, Abrufen, Aktualisieren und Löschen – Create, Retrieve, Update und Delete – zur Verfügung stellte und somit einen Großteil der Funktionen „des Dings“ abdeckte. Aus ursprünglich „T-crud“ wurde sehr schnell TG-crud⁵, dementsprechend wurden auch (als weitere Utilities) die AAI- und Indexdienste in TG-auth*⁶ und TG-search⁷ (um)benannt: Die drei Säulen, wie in Abbildung 4 zu sehen, wurden zu den Grundpfeilern der TextGrid-Infrastruktur.

„Die Folie“ und TG-publish

Während im Laufe der zweiten Förderphase der Produktivbetrieb in greifbare Nähe rückte, wurde allen Beteiligten bald klar, dass das TextGrid Repository in jedem Fall zwei globalen Anwendungsfällen genügen muss:

- Sowohl der/die einzelne Fachwissenschaftler/in als auch Datenanbieter müssen mithilfe des Rechtemanagements jederzeit in der Lage sein, den Zugriff auf ihre Daten steuern zu können, aber
- die (in/mit TextGrid erstellten) Forschungsdaten (auch Zwischenergebnisse) sollen auch langfristig öffentlich zugänglich und zitierfähig sein.

Daher wurden im TextGrid Repository zwei Speicherbereiche mit zugehörigen Suchindizes für die Daten vorgesehen. Das Tafelbild in Abbildung 5 dokumentiert einen Zwischenstand der konzeptionellen Vorarbeiten für diese Lösung.

Der dynamische Speicherbereich (links, etwas unleserlich), ist für die Arbeit an eigenen, unpublizierten Daten mit dem TextGrid Laboratory vorgesehen. Hier ist auch bei der Suche das Rechtemanagement (TG-auth*) aktiv.

Der statische Speicherbereich (rechts) ist für öffentliche, publizierte Daten vorgesehen. Wenn Nutzer in TextGrid publizieren, d.h. ihre Forschungsdaten öffentlich zugänglich machen wollen, dann werden bei diesem Publikationsvorgang die Daten eingefroren, d.h. sie sind inhaltlich nicht mehr änderbar, werden in einem separaten Speicherbereich abgelegt und mit Persistent Identifiern (PID) versehen, sodass sie unter einer festen Adresse langfristig zitierbar sind. Dies ist auch der Bereich des TextGrid Repository, der für die Langzeitarchivierung (LZA) vorgesehen ist.

5 Vgl. TextGrid-Wiki, TG-crud. <https://dev2.dariah.eu/wiki/display/TextGrid/TG-crud>.

6 Vgl. TextGrid-Wiki, TG-auth*, https://dev2.dariah.eu/wiki/display/TextGrid/TG-auth*.

7 Vgl. TextGrid-Wiki, TG-search, <https://dev2.dariah.eu/wiki/display/TextGrid/TG-search>.

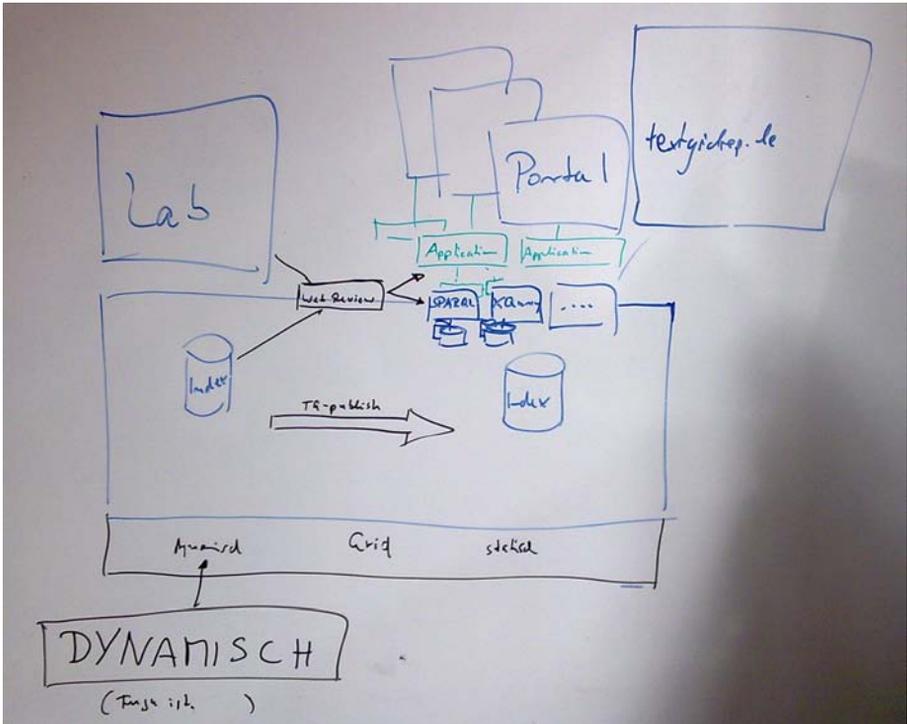


Abb. 5 Tafelbild – Vorarbeiten zu „Die Folie“

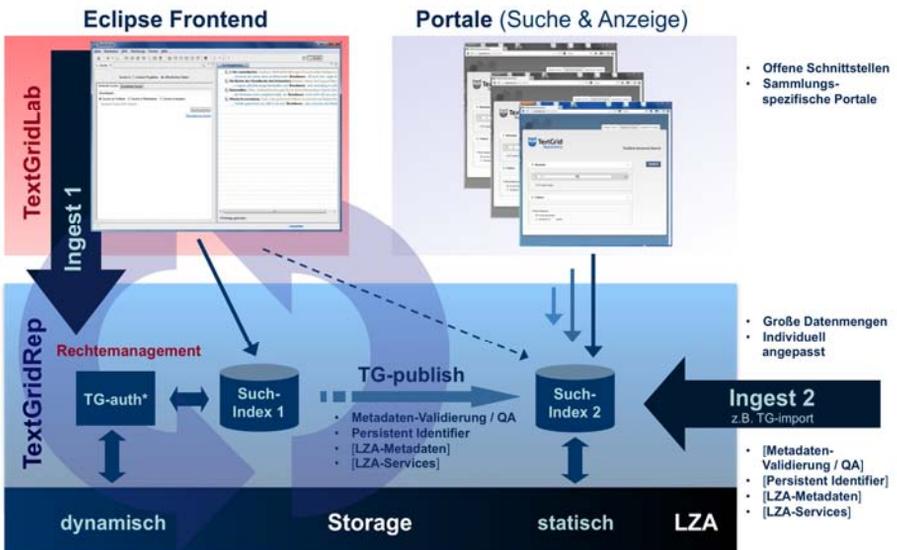


Abb. 6 „Die Folie“ – TextGrid-Architektur Stand 2010 bis 2014

Der Publikationsvorgang, der zahlreiche Arbeitsschritte in einem komplexen und mitunter mehrere Stunden dauernden Workflow zusammenfasst, wird von einer weiteren TextGrid-Utility, TG-publish, durchgeführt.

Das Schema in Abbildung 6 dokumentiert den abschließenden Stand dieses Konzepts, das Anfang 2011 umgesetzt wurde und in dieser Form nahezu unverändert bis heute besteht. Das gleiche gilt auch für die zugehörige Folie, die seitdem in zahlreichen Präsentationen und Publikationen zu finden ist und im Laufe der Zeit nur marginale Veränderungen erfahren hat. Zumindest SUB-intern wurde sie deshalb immer nur als „Die Folie“ bezeichnet.

TextGrid und die Nachhaltigkeit – das DARIAH-DE-Repository

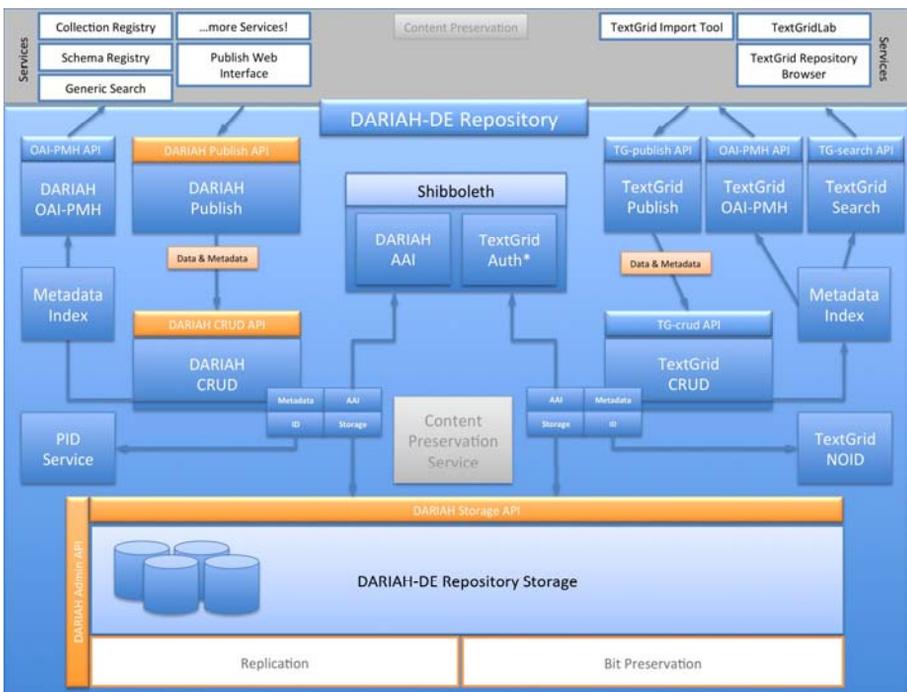


Abb. 7 Das DARIAH-DE-Repository und seine Beziehung zu TextGrid

Im Zuge der Nachhaltigkeitsdiskussion sollen die grundlegenden Dienste und Bestandteile der TextGrid-Middleware von DARIAH-DE als E-Humanities-Infrastrukturprojekt weiter gepflegt und betrieben werden. So wurde bereits in August 2014 die Authentifizierung von TextGrid und DARIAH-DE⁸

8 Vgl. Digital Humanities – DARIAH-DE. <https://de.dariah.eu/>.

– beides über Shibboleth bzw. SAML2 möglich – zusammengeführt, die Autorisierung wird zunächst weiterhin über das rollenbasierte Zugriffskontrollsystem OpenRBAC realisiert. Das DARIAH-DE-Repositorium bietet Storage- und Publikationsmöglichkeiten für alle DARIAH-Nutzer, insbesondere für kleinere E-Humanities-Projekte, um deren Forschungsdaten langfristig der Öffentlichkeit verfügbar und auffindbar zu machen.

Für die Implementierung des Repositoriums werden Kernkomponenten der TextGrid-Infrastruktur nachgenutzt, allerdings wird für DARIAH-DE zunächst nur ein statisches Repositorium implementiert. Um die nachhaltige Pflege der Repositoriums-Komponenten zu ermöglichen, werden außer der Authentifizierungs- und Autorisierungs-Infrastruktur (AAI) noch die Storage-Komponenten gemeinsam genutzt – über die DARIAH-Storage-API – sowie die Metadaten-Indizes, die OAI-PMH-Data-Provider-Implementation und ein PID-Service, der die Persistent Identifier für das TextGrid- und das DARIAH-Repositorium erzeugt und verwaltet.

Zusammenfassung

Um auch im wörtlichen Sinne im Schema zu bleiben, sei hier ein solches an den Schluss gestellt:

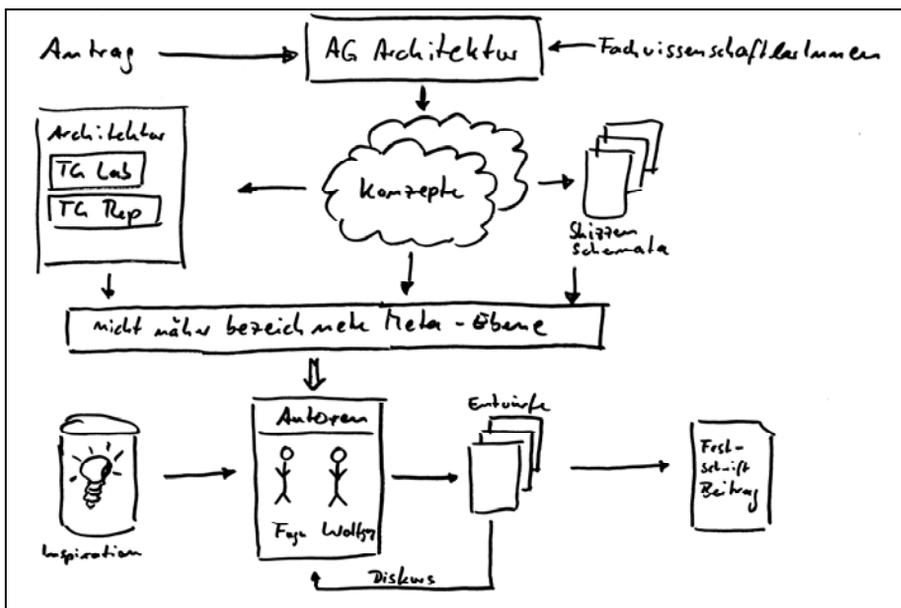


Abb. 8 Zusammenfassung

Literaturverzeichnis

TextGrid – Download und Installation, <http://textgrid.de/registrierungdownload/download-und-installation/>

TextGrid Repository, <http://textgridrep.de/>

TextGrid – Virtuelle Forschungsumgebung für die Geisteswissenschaften. Anträge und Berichte, <http://www.textgrid.de/ueber-textgrid/materialien/antraege-und-berichte/>

TextGrid-Wiki. TG-auth*, https://dev2.dariah.eu/wiki/display/TextGrid/TG-auth*

TextGrid-Wiki. TG-crud, <https://dev2.dariah.eu/wiki/display/TextGrid/TG-crud>

TextGrid-Wiki. TG-search, <https://dev2.dariah.eu/wiki/display/TextGrid/TG-search>

***Share It* – Kollaboratives Arbeiten in TextGrid**

von Andreas Aschenbrenner

Eine der Leitideen im Aufbau von TextGrid war die neue Möglichkeit zur vertieften Kooperation unter Wissenschaftlern. Das betrifft Projektverbünde mit gemeinsamem Ziel, die vielleicht durch ihre Größe, Verteilung der Arbeitsplätze oder Komplexität ihre Forschungsabläufe in einer Virtuellen Forschungsumgebung konzentrieren und strukturieren wollen. Gleichzeitig betrifft es aber auch Interessensgemeinschaften, die z.B. durch ähnliche Forschungsfragen, Quellen oder Forschungsmethoden ein inhaltliches Interesse an den Arbeiten eines anderen Forschungsprojektes haben.

Einen zentralen Baustein der Kooperation bildete im Rahmen der D-Grid-Initiative¹, der auch TextGrid seit 2006 angehörte, die Vernetzung von Speicher- und Rechenleistung. Obwohl TextGrid in der Infrastruktur Funktionen zur Föderation von Archiven und zur Virtualisierung von Rechenleistung vorgesehen hat, basieren Kooperationen zwischen geisteswissenschaftlichen Forschungsvorhaben aber nicht auf Steckplätzen in den Servern eines Rechenzentrums. Mehr noch, idealerweise sollte der Nutzer von dieser technischen Infrastruktur gar nichts mitbekommen. TextGrid geht mit der Perspektive der Geisteswissenschaften konzeptuell eine Ebene höher und unterstützt Kooperationen im Austausch von Daten, in der gemeinsamen Entwicklung von Werkzeugen sowie in der wissenschaftlichen Reflexion von Forschungsmethoden und -ergebnissen.

Daten – vom Archivieren zum Teilen und Nachnutzen

Geisteswissenschaftliche Forschung geschieht in und auf Daten: Editionsprojekte verknüpfen Digitalisate mit dem transkribierten Text und erstellen Annotationen, in den Musikwissenschaften werden Notentexte ediert oder beispielsweise Briefnetzwerke von Komponisten erstellt, in der Kunstgeschichte digitalisierte Bilder mit Annotationen auf Bildausschnitten untereinander vernetzt. Die Quelldaten, z. B. Digitalisate von historischen Texten oder Objekten, sind dabei oft nicht scharf von wissenschaftlichen Anreicherung

¹ <http://www.d-grid.de/>

rungen auf den Daten trennbar. Letztlich spiegeln sich in den Daten Forschungsfrage, Forschungsmethode und Forschungsergebnis wider.

Die Motivation zur Veröffentlichung und zum Austausch von Forschungsdaten ist vielfältig: Offene Daten erzeugen Transparenz und sind die Basis guter wissenschaftlicher Praxis². Die Nachnutzbarkeit von Forschungsdaten aus früheren Vorhaben spart Zeit und Redundanzen in der Quellenerschließung. Eine wachsende Datenbasis ermöglicht es, Analysen teils ad hoc durchzuführen, die früher überhaupt nicht oder nur in jahrelanger Kleinarbeit möglich waren.

Praktisch sind der Austausch und die Nachnutzbarkeit von Forschungsdaten allerdings oft schwierig zu bewerkstelligen, schon allein wegen der datentechnischen Ausgangsposition und dem wissenschaftlichen Selbstverständnis in den Geisteswissenschaften: Geisteswissenschaftliche Forschungsdaten sind nicht technisch testbar und teilweise widersprüchlich in unterschiedlichen Quellen; Forschungsfragen und Forschungsmethoden, die sich in den Daten widerspiegeln, sind heterogen und im Fluss. Selbst dort, wo Projekte gemeinsame Standards für Daten, Metadaten und Schnittstellen nutzen, ist Interoperabilität nicht gesichert.

TextGrid erzwingt daher keine Interoperabilität zwischen unterschiedlichen Projekten, sondern ermöglicht sie und bietet entsprechende Anreize durch semantische Tiefe in Zusatzfunktionalitäten. Die folgenden Funktionalitäten von TextGrid sind daher wechselseitig aufeinander abgestimmt:

Die **Archivierung** der Daten im TextGrid Repository³ ist die Basis für Kooperation. Die Archivierung aufwendig erstellter und daher auch wertvoller Daten ist nicht nur eine intrinsische Motivation für Forscher, sondern wird von Förderinstitutionen inzwischen auch eingefordert.⁴ Maßnahmen zur Langzeitverfügbarkeit wie die Dokumentation und die Kontextualisierung von Daten sind inzwischen Teil „guter wissenschaftlicher Praxis“ und werden in

2 Vgl. die gleichnamige Leitlinie der DFG (1998, 2013).

3 Zum Repository vgl. u. a. auch den Beitrag von Stefan Funk und Wolfgang Pempe im vorliegenden Band.

4 Für die DFG siehe DFG (2009); ein Überblick über unterschiedliche Förderinstitutionen auf <http://www.forschungsdaten.org/index.php/Förderorganisationen>. Siehe auch Blümm (2012), wie die Förderinstitutionen die Entwicklung entsprechender Infrastrukturen fördern.

TextGrid durch strukturierte Metadaten auf Objekt- und Projektebene sowie durch die Möglichkeit zur Verknüpfung freier Dokumentation unterstützt.

Idealerweise archiviert der Forscher dabei nicht nur finale Ergebnisse nach jahrelanger Projektarbeit, sondern legt auch Zwischenresultate (nach relevanten Meilensteinen) ab. Die durch die **Versionierung** entstehenden Verknüpfungen zwischen Objekten sind in 1 skizziert. Verknüpfungen existieren in TextGrid bereits in Einzelprojekten mit unterschiedlichen Objekten, sind aber besonders ausgeprägt in großen Projekten, deren Daten auch in Partnerprojekten nachgenutzt werden. Die semantisch modellierten Beziehungen geben Einblick in die Entstehungsgeschichte von Forschungsarbeiten, und sie erlauben technisch die Reduzierung von Redundanzen in projektübergreifenden Analysen in TextGrid (z. B. ist bei einer Suche in sämtlichen TextGrid-Sammlungen initial die Darstellung der letzten Version eines Objektes ausreichend).

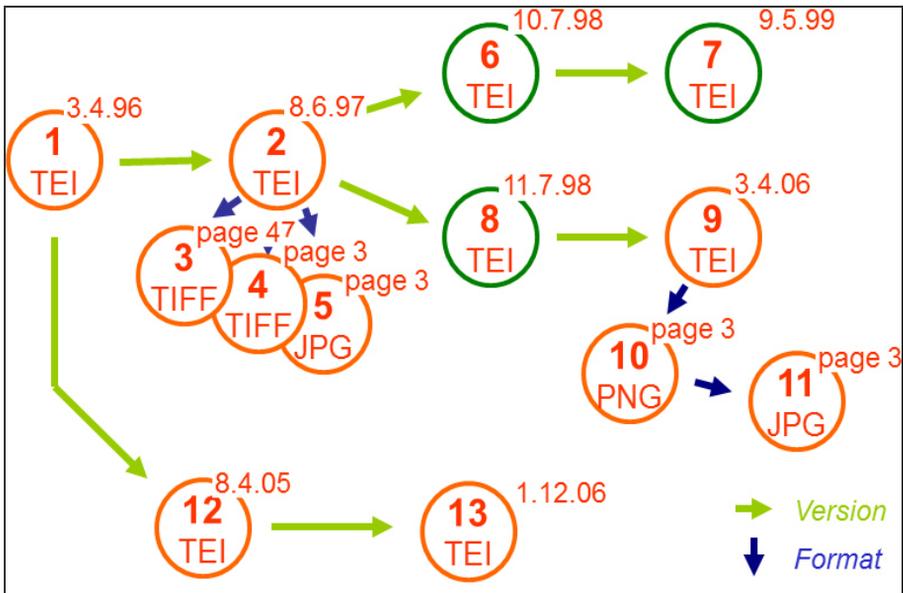


Abb. 1 Beziehungen zwischen Objekten in TextGrid (stilisiertes Modell) – Grüne Pfeile skizzieren Versionen derselben XML/TEI-Datei; Weichen in mehrere Richtungen könnten auf unterschiedliche Forschungsmethoden, -schwerpunkte oder Nachnutzungen in neuen Projekten hinweisen. Bilder werden mit einem Objekt direkt verknüpft, sind aber potenziell für alle Objekte relevant; z. B. könnten die Bilder 3–5 auch digitalisierte Quellen der TEI-Daten in Pkt. 13 sein. Eine Suche würde für sämtliche Objekte nur einen Eintrag bzw. nach Drill-down jedes einzelne Objekt darstellen.

Archivierte Daten müssen stabil referenzierbar sein, um in wissenschaftlichen Publikationen zitiert und in neuen Forschungsvorhaben nachgenutzt werden zu können. Persistent Identifier (PIDs) erlauben die **Referenzierung** publizierter Daten in TextGrid. Die bei der Referenzierung zu erreichende Granularität ist dabei schon bei der Datenmodellierung zu beachten: Sollen z.B. in einem Wörterbuchprojekt einzelne Lemmata referenziert werden, müssen diese entsprechend ausgezeichnet werden – bei XML/TEI offensichtlich, bei Retrodigitalisaten oder einer SQL-Datenbank eventuell etwas trickreich.

Wesentlich in allen Punkten ist die **Flexibilität** und gleichzeitig die Kontrolle der Forscher über ihre Daten. Der Forscher entscheidet selbst, welche Daten (aus dem TextGrid Laboratory) ins Archiv überführt und publiziert werden; die Datenmodellierung und in großen Teilen auch die Metadaten-Modellierung liegt bei ihm. Darüber hinaus gibt es **Anreize**, Daten nachnutzbar zu gestalten und freizugeben: Durch Publikation und Dokumentation erhöht sich dabei die Visibilität der Forschungsdaten und durch Daten- und Metadaten-Standards können neben den Daten auch die datenverarbeitenden Werkzeuge nachgenutzt.

Werkzeuge – Offenheit, Interoperabilität oder beides

Die Werkzeuge des TextGrid Laboratory benötigen bestimmte Datenstrukturen für ihre Funktionalität. In der einfachsten Form ist das ein Datenformat wie ein Text, ein Bild oder ein Video. Mitunter verknüpfen Werkzeuge dabei auch unterschiedliche Datenformate oder legen eine zusätzliche Verweisschicht über die Daten. So erlaubt etwa der Text-Bild-Link-Editor die Verknüpfung von Textsegmenten (XML) mit Bildausschnitten, indem er Verweise auf die jeweiligen Text-/Bild-Ausschnitte in einem separaten Objekt speichert; Videos oder Ton werden bislang nicht unterstützt.

Technische Voraussetzungen, die Werkzeuge in Forschungsdaten erfordern, existieren auf der Ebene der Metadaten und Relationen sowie direkt in Struktur und Inhalt der Forschungsdaten. Obwohl TextGrid grobe Cluster von Werkzeugtypen mit gemeinsamen Anforderungen bietet, kann jedes Werkzeug für bestimmte Funktionalitäten individuelle Anforderungen stellen.

TextGrid-Objekte müssen einen Kern an **Metadaten und Relationen** haben, um in TextGrid verwaltet werden zu können. Dazu zählen auf der einen Seite Rechte oder andere administrative und technische Metadaten, aber auch

deskriptive Metadaten, die die Suche und Darstellung der Objekte z.B. im TextGrid-Navigator ermöglichen. Das TextGrid-Metadaten-Schema basiert auf internationalen Metadaten-Standards wie Dublin Core, FRBR und MODS und sämtliche Konzepte der Metadaten-Community zu Interoperabilität und Nachnutzbarkeit treffen analog für TextGrid zu. Relationen in TextGrid wurden in Abbildung 1 skizziert; neben Versionierung (*isVersionOf*) und Format-Derivaten (*isAlternativeFormatOf*) werden Relationen auch für andere Zwecke genutzt (*hasSchema*, *refersTo*, u.a.). Zusätzlich zum (möglichst kompakten) TextGrid-Metadaten-Schema können spezifische Metadaten vergeben werden, die auch projektübergreifend durch Werkzeuge genutzt werden können.

Wie z.B. der Text-Bild-Link-Editor stellt ein Werkzeug mitunter auch Anforderungen an **Struktur und Inhalt der Forschungsdaten** selbst. Obwohl Werkzeuge maximal flexibel sein und hinsichtlich der Daten möglichst keine Strukturen antizipieren sollen, müssen für die Nutzung von semantisch tieferegreifenden Funktionalitäten einige Voraussetzungen in den Daten und Metadaten gewährleistet sein. So kann etwa der XML-Editor jegliches XML verarbeiten; für die WYSIWYG-Darstellung allerdings muss der Editor die Struktur des XML zu einem gewissen Grad verstehen und erwartet spezielle TEI-Elemente.

Für Darstellung (unter Nutzung von SADE⁵), Analyse (z. B. mittels Voyant⁶) und Druck⁷ von XML/TEI-Daten hat TextGrid die **Kernkodierung** (vgl. TextGrid 2007–2009) entwickelt. Ansätze für eine projektübergreifende Codierung in TEI liegen von verschiedener Seite vor (vgl. Geyken 2013); die TextGrid-Kernkodierung unterscheidet sich hier, indem sie keine allgemeinen Empfehlungen für eine umfassende Codierung gibt, sondern ganz spezifisch die Mindestanforderungen von Tools codiert. Dadurch ist die TextGrid-Kernkodierung schlanker und flexibler als andere Ansätze⁸ und schafft einen gemeinsamen Nenner zwischen unterschiedlichen Projekten, ohne diese durch weitreichende Vorgaben einzuschränken. Die Möglichkeiten, die eine

5 SADE – Scalable Architecture for Digital Editions, <http://www.bbaw.de/telota/software/sade/sade-1>; vgl. auch den Beitrag von Ubbo Veenster im vorliegenden Band.

6 Voyant, DIGIVOX: <https://de.dariah.eu/digivoy>

7 XML Print: www.xmlprint.de

8 Vgl. z.B. IDS-XCES, DTA-BF, beschrieben in Geyken (2013); TEI Tite (Trolard 2011), TEI Lite (Burnard/Sperberg McQueen 2006).

semantisch tiefgreifende Integration durch die Kerncodierung eröffnet, werden auch in TextGrid noch nicht umfassend genutzt; weitere Entwicklungsschritte für neue wissenschaftliche Werkzeuge und Methoden stehen den Digital Humanities diesbezüglich noch offen.

Sowohl die kompakten Metadaten als auch die minimalen Vorgaben an die Forschungsdaten liefern größtmögliche Freiheiten und stellen gleichzeitig auf inhaltlicher Ebene eine Basis-Interoperabilität zwischen Projekten sicher. Diese Interoperabilität bietet schließlich den größten Anreiz im Teilen (*share it!*) und in der Nutzung einer verteilten Virtuellen Forschungsumgebung wie TextGrid: Existierende Projekte können unmittelbar nachgenutzt, Analysen (soweit von der TextGrid-Kerncodierung unterstützt) korpusübergreifend automatisiert durchgeführt werden.

Nicht nur hinsichtlich der Daten, auch bezüglich der Werkzeuge hat TextGrid bereits bei der Entwicklung der Basisarchitektur eine feine Balance zwischen Flexibilität und Interoperabilität gesucht. Werkzeuge können in TextGrid auf zwei Ebenen eingebunden werden: (1) in die Service-Schicht⁹ und auf die REST-basierten Services für Speicherung/Archivierung (TG-crud), Authentifizierung (TG-auth*) und Suche (TG-search)¹⁰ sowie (2) in die Eclipse-basierte Nutzerumgebung des TextGrid Laboratory¹¹. Für jede Ebene liegen u. a. mit API-Spezifikationen, OAI-PMH, Vorgaben für Service-Beschreibungen etc. sämtliche notwendigen technischen Ansatzpunkte vor, um Eigenentwicklungen direkt in die TextGrid-Umgebung einzubinden. Letztlich können also Werkzeuge ähnlich wie Daten in der TextGrid-Umgebung gemeinschaftlich entwickelt und geteilt werden.

Mit der Öffnung der Schnittstellen gibt TextGrid auch einen gewissen Grad an Kontrolle an die Nutzer-Community ab: Die Stabilität externer Dienste kann in einer offenen Service-Umgebung nicht garantiert werden und ohne einen zentralen Gesamtüberblick können mögliche fachliche oder technische Lücken eventuell übersehen bleiben. Um diese Risiken zu mindern und um Stabilität und Sicherheit aus der Infrastruktur heraus zu sichern, hat TextGrid gemeinsam mit den größten Rechenzentren Deutschlands umfangreiche Infrastrukturarbeit in die Kernkomponenten (TG-crud, TG-auth* und TG-search) investiert.

9 <https://dev2.dariah.eu/wiki/display/TextGrid/Writing+a+Service+Description>

10 http://www.textgrid.de/fileadmin/TextGrid/reports/R3_5-manual-tools.pdf

11 <https://dev2.dariah.eu/wiki/display/TextGrid/User+Manual+2.0>

In der Praxis werden die offenen Schnittstellen derzeit primär von TextGrid-Partnerprojekten bzw. mit individueller Anleitung und Support von TextGrid genutzt. Ähnliches kann auch bei anderen geisteswissenschaftlichen Infrastrukturinitiativen beobachtet werden. Vielleicht ist die Community der geisteswissenschaftlichen Forscher mit Programmiersprachenkenntnissen derzeit noch nicht groß genug, um eine Bewegung ähnlich derjenigen um das Web 2.0 aus Rest-APIs und Mashups zu entfachen. Vielleicht aber liegen die großen Entwicklungsschritte zum Teilen und Nachnutzen unmittelbar vor uns.

Sicherheit und Teamstrukturen

Das Teilen und Nachnutzen von Daten und Werkzeugen in TextGrid ist prinzipieller Bestandteil der Forschungsumgebung; wesentlich hierbei ist jedoch die Nutzer-Kontrolle über Zeitpunkt und Umfang der Veröffentlichung:

(1) Nicht jeder Bearbeitungsschritt soll öffentlich einsehbar sein. Objekte in TextGrid werden daher nur durch einen expliziten Publikationsschritt öffentlich zugänglich gemacht. Der Ersteller eines Objektes in TextGrid (genauer: Projektmanager, siehe nachfolgend) kann diesen Publikationsschritt auslösen und er führt damit eine automatische Qualitätskontrolle aus, versieht das Objekt mit einem Persistent Identifier und archiviert es permanent. Dieser Schritt ist irreversibel und garantiert die dauerhafte Zitierbarkeit der publizierten Daten.

Anders als bei Print-Publikationen wird der TextGrid-Nutzer zu häufigem Publizieren ermuntert. Das jeweilige Projekt gewinnt Visibilität und ermöglicht frühe Interaktion mit der wissenschaftlichen Gemeinschaft; darüber hinaus kann ein wissenschaftlicher Diskurs schon in frühen Projektphasen die Qualität des Projektergebnisses entscheidend verbessern.

(2) Wenngleich die Ergebnisse der Bearbeitungsphase bzw. deren Zwischenstände häufig nicht vollständig publiziert werden sollen, arbeiten mehrere Projektmitarbeiter an denselben Objekten. Die Projekt- und Nutzerverwaltung erlaubt es, Projektteilnehmern Rollen zuzuteilen und damit Rechte für Nutzerverwaltung, Löschen, Schreiben und Lesen. Die daraus entstehenden Teamstrukturen skalieren auch auf große Projekte an unterschiedlichen Standorten.

(3) Wenn Daten aufgrund von Copyright-Bestimmungen, Datenschutz oder anderen rechtlichen Einschränkungen nur begrenzt zugänglich sind, können sie auch nach Veröffentlichung in separaten TextGrid-Infrastruktur-

Knoten (vgl. TG-licence) nur einer begrenzten Nutzerschaft zugänglich gemacht werden. Auch zeitliche Begrenzungen von Zugriffssperren (Moving Walls) sind technisch machbar.

Entsprechend dem ursprünglichen Gedanken einer offenen Infrastruktur fördert TextGrid die Veröffentlichung, Archivierung und Nachnutzung von Forschungsdaten in einer wachsenden und vernetzten Wissenschaftsgemeinde. Die Möglichkeit zur Publikation von Zwischenständen fördert die Transparenz von Forschungsprojekten und den wissenschaftlichen Diskurs.

Reflexion – *Share It!*

Teilen führt nicht, wie vielfach befürchtet, zu einer Einebnung von Vielfalt und freiem Denken, sondern führt umgekehrt zu mehr Wissenschaftlichkeit durch Vergleich und Diskussion zwischen Projekten, aber auch zu neuen Einblicken durch eine größere Datenbasis und eine breitere Werkzeugpalette.

Mit Teilen alleine ist es aber nicht getan. Erst Interoperabilität und offene Dokumentation ermöglichen die Nachnutzung von Daten, die offene Entwicklung von Werkzeugen, komplexe Abläufe und übergreifende Analysen. In der Entwicklung von TextGrid mussten die Partner eine Balance zwischen unterschiedlichen und zum Teil widersprüchlichen Zielen finden:

- Flexibilität in Datenmodellierung vs. projektübergreifende Interoperabilität,
- Kontrolle beim Nutzer vs. Transparenz und Offenheit zur Nachnutzbarkeit,
- Stabilität durch Zentralisierung vs. wachsendes Netzwerk aus gleichberechtigten Partnern.

Die Balance wurde in TextGrid durch eine intensive Diskussion unter gleichberechtigten technischen und fachwissenschaftlichen Projektpartnern gefunden. Dabei haben sie eine gemeinsame Vision einer vernetzten Forschungslandschaft entwickelt, in der auch Forschungsdaten, Werkzeuge und Zwischenstände von Projektergebnissen ein Teil des wissenschaftlichen Publikationszyklus sind. Mechanismen dafür sind in TextGrid vorhanden und diese teilt TextGrid (inzwischen) mit anderen aktuellen Initiativen. Wie diese Mechanismen letztlich genutzt werden und ob die Wissenschaftsgemeinde vielleicht sogar darüber hinausgeht, liegt bei jedem einzelnen Forscher. In den ersten zehn Jahren TextGrid sind wichtige Schritte dahingehend gemacht worden.

Literaturverzeichnis

- Blümm, Mirjam (2012): Zur Rolle der Deutschen Forschungsgemeinschaft (DFG) und des Joint Information Systems Committee (JISC) innerhalb der deutschen und englischen Bibliothekslandschaft exemplifiziert an ausgewählten Tätigkeitsbereichen und Förderprojekten. In: *Perspektive Bibliothek* 1 (1): 147–168.
- DFG (1998, 2013): Sicherung guter wissenschaftlicher Praxis. Denkschrift. http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf.
- DFG (2009): Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten. http://www.dfg.de/download/pdf/foerderung/programme/lis/ua_inf_empfehlungen_200901.pdf.
- Geyken, Alexander (2013): Wege zu einem historischen Referenzkorpus des Deutschen: Das Projekt Deutsches Textarchiv. In: Hafemann, Ingelore (Hrsg.): *Perspektiven einer corpusbasierten historischen Linguistik und Philologie. Internationale Tagung des Akademienvorhabens „Altägyptisches Wörterbuch“ an 'der Berlin-Brandenburgischen Akademie der Wissenschaften, 12.–13. Dezember 2011*. Berlin 2013, S. 221–234. <http://edoc.bbaw.de/volltexte/2013/2442/pdf/Geyken.pdf>.
- Küster, Marc Wilhelm; Ludwig, Christoph; Aschenbrenner, Andreas (2009): TextGrid: eScholarship und vernetzte Angebote. In: *it – Information Technology* (Themenheft „Informatik in den Philologien“) 51 (4): 183–190.
- TextGrid (2007–2009): TextGrid Baseline Encoding for Text Data in TEI P5. <https://dev2.dariah.eu/wiki/download/attachments/7439482/baseline-all-en.pdf>.

Publizieren mit TextGrid

SADE und XML-Print

von Ubbo Veentjer und Marc Wilhelm Küster

TextGrid bietet neben der langfristigen Archivierung geisteswissenschaftlicher Forschungsdaten im TextGrid Repository weitere Wege, die im TextGridLab generierten Forschungsdaten zu publizieren – digital wie analog. Dazu zählt zum einen die „Skalierbare Architektur für Digitale Editionen“ (SADE), die es ermöglicht, XML/TEI-Daten aus dem Lab in einem eigenen Online-Portal zu veröffentlichen. Zum anderen bietet sich mit dem Plug-in „XML-Print“ ein ergonomisches Satzsystem speziell für kritische Editionen auf XML/TEI-Basis.

SADE – Skalierbare Architektur für Digitale Editionen: Ein Publikations-Framework für TEI-Daten

Bei Präsentationen von TextGrid wird oft die Frage gestellt: „Wo können wir uns die digitale Edition ansehen?“ Dahinter steht der verständliche Wunsch, Daten aus dem TextGrid Repository komfortabel in einem eigenen Webportal zu präsentieren. Noch in der ersten Projektphase wurde der Prototyp einer Komponente mit dem Namen „WebPublisher“ entwickelt. Diese bestand aus einem TextGridLab-Plug-in, in dem die zu publizierenden Inhalte zusammengestellt werden konnten, und einer serverseitigen Komponente, bei der die XML-Datenbank eXist¹ für die Darstellung der publizierten Ordnerstruktur, für die HTML Ansicht der TEI Daten sowie für die Volltextsuche genutzt wurde. Für die Darstellung von Bildern kam der IIP Image Service² zusammen mit dem IIPMooViewer³ zum Einsatz.

Die Entwicklungen dieses ersten Prototypen wurden dann von der TextGrid-Entwicklung überholt. So wurde im TextGridLab und -Repository die

1 <http://exist-db.org>

2 <http://iipimage.sourceforge.net/>

3 <http://iipimage.sourceforge.net/documentation/iipmooviewer-2-0/>

Möglichkeit geschaffen, Objekte in sogenannten Aggregationen (einer Art Ordnerstruktur) zusammenzustellen. Hier konnte für den TextGridLab-Aggregationseditor Quelltext der TextGridLab-Webpublisher zur Zusammenstellung von Objekten nachgenutzt werden. Der damalige TextGrid-Entwickler Wolfgang Pempe und Alexander Czmiel von der BBAW stellten dann bei einem Gespräch fest, dass das an der BBAW im Rahmen des TELOTA-Projektes entwickelte SADE⁴ dieselbe Zielsetzung wie der Webpublisher aufweist und ähnliche Technologien nutzt. Bei SADE kam eXist für die HTML- und digilib⁵ für die Bilddarstellung zum Einsatz. Um SADE als Nachfolger des Webpublishers einsetzen zu können, sodass mit einem TextGridLab-Plug-in komplette Aggregationen publiziert werden können, vereinbarte man eine Zusammenarbeit. Wurde beim Webpublisher noch IIIP eingesetzt, erwies sich die digilib-Nutzung von SADE als Vorteil: Bereits seit der zweiten Projektphase wird digilib auch für TextGrid eingesetzt und angepasst.

Seit 2012 beteiligen sich auch die Österreichische Akademie der Wissenschaften (ÖAW), das Max-Planck-Institut für Wissenschaftsgeschichte (MPIWG) und das Cologne Center for eHumanities (CCeH) an der Weiterentwicklung von SADE.

SADE steht als Akronym für „Skalierbare Architektur für digitale Editionen“ und bietet ein Programmpaket, welches verschiedene Komponenten umfasst, die zum Entwickeln und Darstellen einer digitalen Edition besonders geeignet sind. Der Fokus liegt dabei auf der Präsentation von XML-Dokumenten, die den Richtlinien der Text Encoding Initiative (TEI) folgen. Es hat zum Ziel, eine modulare und vom Nutzer einfach an eigene Bedürfnisse anpassbare Publikationsplattform für digitale Editionen anzubieten. Zu diesem Zweck wurde in SADE ein Modulsystem verankert, mit welchem sich für ein Portal genutzte Komponenten zusammenstellen lassen. Zur leichteren Anpassung des Portals an eigene Designs kommt die eXist Template Engine zum Einsatz. Dadurch werden Funktionalität und Layout von Modulen getrennt, so dass sich Module komfortabel in eigenen Webseiten-Designs nutzen lassen. SADE liefert darüber hinaus einige Kernmodule, beispielsweise für die TEI-Darstellung oder für die facettierte Suche (Abb. 1):

4 Scalable Architecture for Digital Editions – <http://www.bbaw.de/telota/software/sade/sade-1>

5 <http://digilib.sourceforge.net/>

The screenshot shows the SADE web interface. At the top, there's a search bar containing the word 'Frosch'. Below the search bar, there are navigation tabs: 'SADE', 'Dokumentation', 'Weiterführende Links', and 'Info'. On the left side, there are four main categories for filtering results: 'Person', 'Schlagwort', 'Zeit', and 'Publikationsort'. Each category has a list of items with blue circular icons and counts. The main content area shows search results for 'Frosch'. It includes a pagination bar (1, 2, >, >>), a result count 'Zeige Treffer 1 - 10 von 15 Treffern', and sorting options 'Sortierung nach Relevanz' and 'Aufsteigend'. The first result is '1. Der Froschkönig oder der eiserne Heintich' with a snippet: '... die Stimme käme, da erblickte sie einen **Frosch**, der seinen dicken häßlichen Kopf au ...'. Other results include 'Der kleine Tuk', 'Die Störche', 'Unpolitische Erinnerungen', and 'Abenteurer und Fahrten des Huckleberry Finn', each with a snippet containing the word 'Frosch' in red.

Abb. 1 Publikation von Aggregationen aus dem TextGrid Repository (*Digitale Bibliothek*) in SADE: Facettierte Suche und Darstellung der Suchergebnisse mit KeyWord In Context (KWIC)

Die Nutzung von eXist ermöglicht die einfache Entwicklung von XML-basierten Applikationen mit XML-Technologien wie XQuery oder XSLT. Digilib bietet auch aufgrund der guten Integration in das TextGrid Repository einen Mehrwert für die Bilddarstellung. So können Scans im TIFF-Format im TextGrid Repository abgelegt werden, für die Darstellung in SADE on the fly für das Web optimierte JPEG-Bilder generiert und ausgeliefert werden.

Um SADE zu nutzen, wird zunächst ein ZIP-Ordner von der Webseite⁶ heruntergeladen. Dieser wird auf dem eigenen Rechner oder dem Server entpackt und gestartet. Anschließend kann mit der Web-Publikation von Daten in ein eigenes Portal begonnen werden. Um die Möglichkeiten von SADE vor einer Veröffentlichung zu testen, können Daten zunächst auf die TextGrid-Referenzinstanz⁷ publiziert werden. Im TextGridLab kann hierfür das SADE-Publish-Plug-in verwendet werden (vgl. Abb. 2):

6 <http://www.bbaw.de/telota/software/sade/documentation/getting-started>

7 <https://dev2.dariah.eu/wiki/display/TextGrid/SADE+TextGrid+reference+installation>

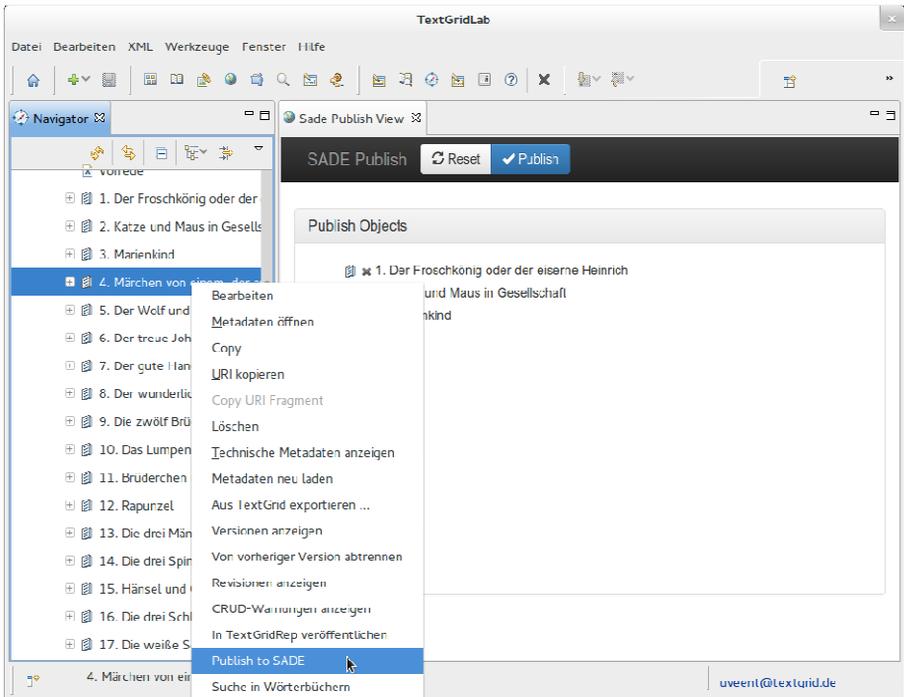


Abb. 2
Das SADE-Publish-Tool im TextGridLab und das Kontextmenü des Navigators

SADE dient nicht dazu, die Publikation in das TextGrid Repository zu ersetzen, es stellt vielmehr eine Ergänzung zu dieser Veröffentlichungsform dar. Durch die Publikation im TextGrid Repository sollen Forschungsdaten langfristig und nachhaltig zugänglich, durchsuch- und zitierbar sein; es garantiert die nachhaltige Verfügbarkeit und Zugänglichkeit der geisteswissenschaftlichen Forschungsdaten sowie eine optimale Vernetzung. In der Darstellung der Datensätze ist die TextGrid-Repository-Webseite hingegen eingeschränkt; Möglichkeiten zur individuellen, projektspezifischen Anpassung sind nicht gegeben. SADE hingegen bietet Forschungsvorhaben die Möglichkeit, die Präsentation und Darstellung ihrer Forschungsdaten beliebig anzupassen und etwa durch eigene Visualisierungen aufzubereiten.⁸ Heute nutzen bereits verschiedene Projekte erfolgreich diese Möglichkeit, darunter

⁸ Eine ausführliche Dokumentation von SADE in TextGrid liefert der Report 4.2.2: http://textgrid.de/fileadmin/TextGrid/reports/TG-R-4_4_2-SADE_Dokumentation_und_Referenzinstanz.pdf.

die *Bibliothek der Neologie* (vgl. Abb. 3), die *Hybrid-Edition von Theodor Fontanes Notizbüchern* und *Johann Friedrich Blumenbach – online*.⁹

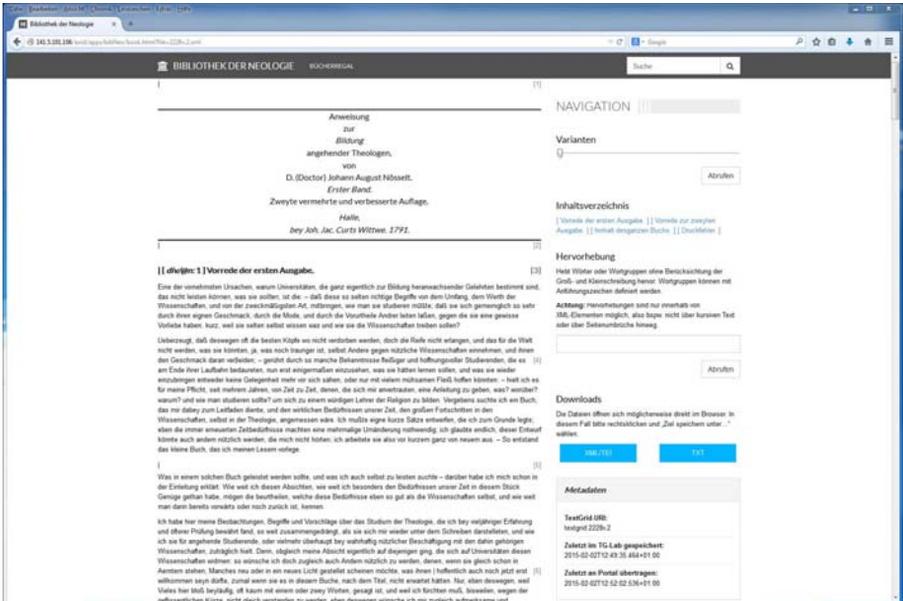


Abb. 3 SADE-Portalansicht eines Kapitels aus der *Bibliothek der Neologie* (Universität Münster), Testversion

Ubbo Veenjtjer

Publizieren mit TextGrid: XML-Print¹⁰

XML-Print ist ein von der Deutschen Forschungsgesellschaft (DFG) gefördertes Open-Source-System, um XML-strukturierte Daten nach PDF zu konvertieren und zu drucken. Es bietet neben der eigentlichen Implementierung der Textsatz-Algorithmen eine intuitive Oberfläche, um die semantische XML-Struktur auf Formatvorlagen abzubilden und den Satzprozess zu steuern.

⁹ Vgl. dazu die Beiträge von Gabriele Radecke, Bastian Lemitz und Martin Liebethuth im vorliegenden Band.

¹⁰ Die Ausführungen dieses Abschnitts basieren u. a. auf Burch et al. (2012), Küster et al. (2013) und Georgieff et al. (2014). Diese auf drei aufeinanderfolgenden Digital-Humanities-Konferenzen dem Publikum präsentierten Vorträge geben gleichzeitig einen Überblick über die in die Projektrealisierung und -leitung primär involvierten KollegInnen.

Zu drucken? Ist Papier nicht ein Anachronismus, Relikt einer Gutenberg-Galaxie¹¹, die wir längst hinter uns gelassen haben? Gehört die Gegenwart nicht ausschließlich Online-Editionen, vielleicht noch in XML, vielleicht gleich ausschließlich in HTML5?

In manchem ist Papier wirklich ein Überbleibsel einer vergangenen Zeit. Gedruckte Texte können immer nur einen Teil des Reichtums zumindest einer guten digitalen, etwa in TEI (vgl. Burnard et al. 2007) strukturierten kritischen Edition nachbilden. Beispielsweise sind Annotationen in einer Druckausgabe nur mühsam über Apparate zu imitieren. Im Gegensatz zum elektronischen Medium hat Papier physische Beschränkungen, die es schwer machen, Anmerkungen in beliebiger Länge zu realisieren. Links zwischen Dokumenten, die in der elektronischen Welt selbstverständlich sind, lassen sich im Druck nur über ausgefeilte Referenzsysteme ausdrücken. Ebenso sind Verweise innerhalb des Dokuments selbst kompliziert – Papier favorisiert lineare, hierarchisch strukturierte Darstellungen. Versuche, Hypertexte auf Papier zu visualisieren, sind naturgemäß Notlösungen – und kritische Editionen sind in ihrer Essenz genau das: Hypertexte, die den gegebenen Lesetext mit den Varianten anderer Textzeugen und Hintergrundinformationen verknüpfen, u. a. unter Einsatz verschiedener Apparate (vgl. Abb. 4).

Dennoch ist Papier auch heute noch aus dem Alltag eines Geistes- und speziell eines Editionswissenschaftlers nicht wegzudenken. Für viele Editionen bleibt Druck der Maßstab. Oftmals gibt es Editionsreihen seit Jahrzehnten, teils wie die *Inscriptiones Graecae* oder das *Corpus Inscriptionum Latinarum* seit weit über einem Jahrhundert. In dieser Zeit haben sich strikte Konventionen etabliert, wie der Druck darzustellen ist, dem sich auch neue Editionen beugen müssen.

Gleichzeitig hat Papier unbestreitbare Vorteile für die Langzeitarchivierung. Bücher überleben weitgehend problemlos Jahrhunderte, elektronische Editionen oftmals nicht einmal wenige Jahre. Letzteres wird sich mittelfristig sicher ändern, denn entsprechende Projekte zur Langzeitarchivierung geisteswissenschaftlicher Texte sind in der Entwicklung.¹² Es gibt aber bislang keine Dienste, die diese Archivierung für diese Art von Texten mit derselben

11 so der Titel von McLuhan (1962)

12 Vgl. z. B. Neuroth et al. (2010) für eine Übersicht über einschlägige Projekte. Eine nähere Diskussion dieser Thematik ist nicht Gegenstand dieses Kapitels.

peut lui échapper. De là ce besoin d'isoler son peuple et des souvenirs du
 229 passé, et des séductions du présent^a. | De là ces lois sévères contre des
 230 vaincus plus nombreux que les vainqueurs^b. De là | ces châtimens effroya-

l'intention manifeste de Moïse, a souvent embarrassé les théologiens. Dieu voulait, dit saint
 Philippe¹ (Monarchie des Hébreux), recevoir des hommages de son peuple à quelque prix
 que ce fût. Dieu paraît, dit Spencer, avoir, dans l'institution des rites mosaïques, été forcé et
 subjugué par une sorte de nécessité qui l'entraînait presque malgré lui, *quasi coactus*.
 (SPENCER, de leg. rit. Hæb. I, 196.)

^a Israël habitera seul et en sûreté. (Deuter., XXXIII, 28, Gen. XLIII, 32.) Je suis le Seigneur
 votre Dieu qui vous ai séparés des autres peuples, pour que vous ne fussiez qu'à moi.
 (Lévit, 9, 20, 24, 25, 26.) La plupart des lois rituelles des Hébreux finissent par ces mots :
 «Observez cette loi, car elle est un signe entre vous et moi. (Exod. 31, 13.) Vous n'agirez ni
 selon les coutumes du pays d'Égypte où vous avez demeuré, ni selon les mœurs du pays de
 Chanaan dans lequel je vous ferai entrer. Vous ne suivrez ni leurs lois ni leurs règles. (Lévit.
 XVIII, 3 et suiv.)» L'intention de Moïse s'aperçoit dans les désignations des lieux particu-
 15 liers pour les sacrifices, et dans les peines prononcées contre ceux qui en offriraient ailleurs.
 Dans la partie même des lois qui se rapportent aux causes d'impureté, partie manifestement
 empruntée ou imitée de l'Égypte, le législateur cherche encore des lignes de séparation.
 (SPENCER, I, 115, 195.) C'est ainsi que s'expliquent mille interdictions qui semblent arbi-
 20 traires, celles de semer dans les vignes, de faire cuire le chevreau dans le lait de sa mère, de
 manger de la chair crue, etc., etc. (Deuter. XXII, 9 et suiv.) Toutes ces interdictions étaient
 motivées sur quelque usage des nations voisines ; il en est de même de la défense
 de labourer avec un bœuf et un âne. Si l'on veut voir en détail combien ce but est marqué dans
 les lois mosaïques, il faut lire le traité de Spencer que nous avons cité. (I, 277, 587.)

^b Il est à remarquer que dans ces mesures rigoureuses, Moïse a presque toujours la nécessité
 25 pour excuse. Condamnés à conquérir un sol qui les nourrit, les Hébreux étaient forcés de
 détruire les tribus qui, revenues de leur première épouvante et se réunissant contre eux, les
 auraient tôt ou tard détruits eux-mêmes. La dévastation marchait donc inévitablement avec
 la conquête. Tout autre peuple en eût fait autant. Ce n'est point la religion de Moïse, c'est
 sa position qu'il faut accuser. Mais Moïse prévoit une époque où plus d'indulgence sera
 30 possible. «Quand vous approcherez d'une ville pour l'assiéger,» dit-il, «vous lui offrirez la

23 et un âne] ou un âne *Rel. II, I, corrigé dans l'erratum*

sances s'étendaient, était à un comparatisme qui paraît aujourd'hui hâtif et superficiel. Jones
 avait donné l'exemple dans son essai *On the Gods of Greece, Italy and India*, publié en
 1799, mais communiqué dès 1785 à la Société asiatique, dans lequel il mettait l'accent sur
 les ressemblances des religions classiques avec celles de l'Inde, idée féconde, mais qui,
 chez d'autres, n'était pas fondée sur des connaissances aussi solides. Constant avait trouvé
 des exemples chez Spencer. Les comparaisons entre la Grèce et l'Égypte ou l'Inde abondent
 chez les contemporains de Constant, et datent même, pour la Grèce et l'Égypte, de l'anti-
 quité.

¹ Le nom «saint Philippe» est une erreur de l'imprimeur pour Saint Philippe, francisation de :
 Bacallar y Sanna, Vicente, marques de San Felipe (1669–1726), diplomate et historien
 espagnol, auteur de *Monarchia hebraea*, La Haye : Van der Kloot, 1727. Traduction fran-
 çaise par Antonin de Labarre de Beaumarchais : *La monarchie des Hébreux, ouvrage post-
 175 hume tiré de la Bible et de Flavius Josèphe*, La Haye, 1727.

Abb. 4 Beispiel einer kritischen Edition¹³

13 Beispiel aus der von K. Kloocke besorgten Edition Constant (1999) mit zwei Apparat-
 ten

Selbstverständlichkeit und Zuverlässigkeit wahrnehmen, wie es Bibliotheken für gedruckte Bücher seit Jahrhunderten machen.¹⁴

Vor diesem Hintergrund ist das gedruckte Buch vielleicht nicht mehr das absolute Leitmedium der Editionswissenschaft. Es bleibt aber auf absehbare Zeit ein in vielen Fällen unverzichtbares Ausgabeformat. Gleichzeitig aber steht das Buch nicht mehr allein. Im Zentrum steht die XML-Version der Edition. Das Buch ist genau wie die Online-Edition nur *ein* Ausgabeformat dieser XML-Edition unter mehreren.

Es bleibt das Problem, wie man vom XML zu einer ansprechenden Druckausgabe kommen kann. Es gibt zwar einige kommerzielle¹⁵ und freie¹⁶ Programme, die einen Teil der Aufgabe erledigen. Sie sind allerdings größtenteils entweder zu teuer für viele Editionsprojekte in den Geisteswissenschaften oder sehr komplex in der Benutzung.

Was?

Überblick

Vor diesem Hintergrund entstand XML-Print als von der DFG gefördertes gemeinsames Projekt der Hochschule Worms und der Universität Trier. XML-Print hat zwei ineinander verzahnte Open-Source-Werkzeuge entwickelt, die beide auf <http://sourceforge.net/projects/xml-print/> zum Download und zur Weiterentwicklung verfügbar sind:

- eine in Java als Eclipse-Plug-in entwickelte interaktive grafische Oberfläche (GUI), um Regeln zu definieren, die festlegen, wie die XML-Elemente eines Texts im Druck dazustellen sind,
- eine als Kommandozeilen-Werkzeug realisierte Satz-Engine, geschrieben in F#, um den XML-Text als PDF für den Druck auszugeben.

Für die Verwendung der beiden Komponenten sind zwei Möglichkeiten vorgesehen. Normalerweise interagieren Nutzer mit der GUI, um XML-Strukturen auf Layout-Regeln abzubilden. In diesem Fall bleibt die eigentliche Satz-Engine für sie weitestgehend unsichtbar. Die Oberfläche ruft hinter den

¹⁴ In spezialisierten Bereichen wie etwa der Gesetzgebung gibt es sehr wohl entsprechende Dienste.

¹⁵ etwa Arbortexts 3B2 und Adobes InDesign

¹⁶ Vgl. etwa Ott (2008), Knuth (1986), Lamport (1986) und Apache Foundation (2013).

Kulissen die Engine auf und zeigt dem Nutzer das Resultat. Er kann die Regeln dann ggf. solange anpassen, bis das Endergebnis seinen Vorstellungen entspricht. Zur Nutzung etwa in einem Web-Service lässt sich die Engine aber auch losgelöst von der Oberfläche einsetzen. Wir werden weiter unten Beispiele für beide Nutzungsszenarien sehen.

Als Schnittstelle zwischen Oberfläche und Satz-Engine fungiert eine um für Editionen unersetzliche Elemente erweiterte Untermenge der XSL-FO-Spezifikation (vgl. Berglund 2006), die wir XSL-FO+ genannt haben.

Funktionalitäten

Sowohl die Oberfläche als auch die Satz-Engine unterstützen natürlich die wesentlichen Eigenschaften von XSL-FO wie die Auswahl von Seitenformaten, Schriftarten und -schnitten, Absatzgestaltung usw. So zeigt der folgende Dialog die Konfiguration von Absatzparametern, in diesem Fall den gewünschten Zeilenabstand vor einem Absatz. Alle anderen Werte bleiben in diesem Beispiel unverändert:

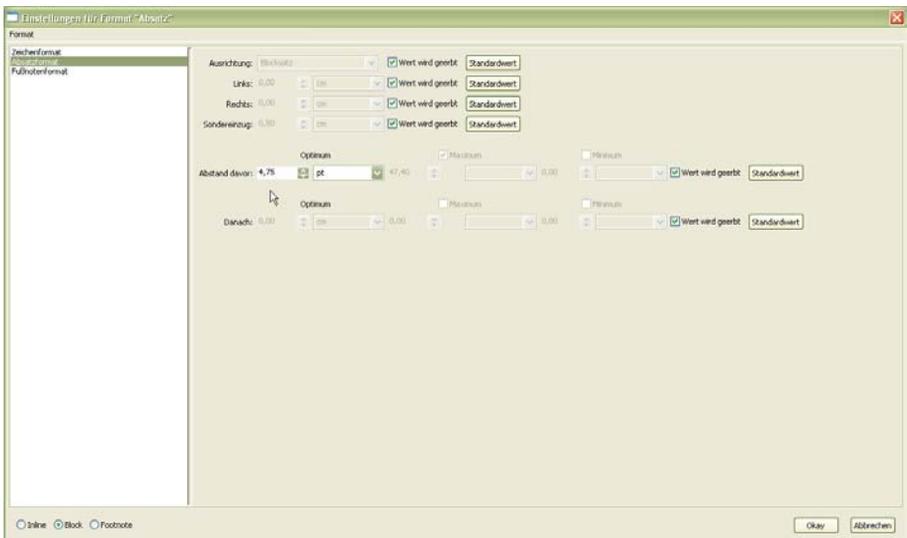


Abb. 5 Absatzformatierung mit XML-Print

Darüber hinaus haben wir den Fokus aber gerade auf die speziellen Anforderungen kritischer Editionen gelegt:

- mehrspaltiger Satz,

- kritische Apparate,
- Verweise innerhalb des Textes.

Wie?

Architektur

Wie bereits erwähnt, entkoppelt XML-Print die für die Nutzer sichtbare Oberfläche (GUI) von der eigentlichen Satz-Engine, um einerseits die Komplexität beider Einzelkomponenten zu reduzieren, andererseits beide Komponenten getrennt voneinander einsetzbar zu machen. Nutzer legen in der GUI die Regeln fest, nach denen bestimmte Elemente in den XML-Daten formatiert werden sollen. Basierend darauf generiert die GUI eine Reihe von Transformationsregeln als XSL-Datei sowie das Resultat der Transformation als XSL-FO+. Beide zusammen werden samt der Ausgangsdaten in ein Zip-Archiv verpackt; sie dienen der Satz-Engine als Eingabe. Diese wiederum generiert schlussendlich die PDF-Datei mit dem Satzergebnis, die der Nutzer in der Oberfläche inspizieren kann.

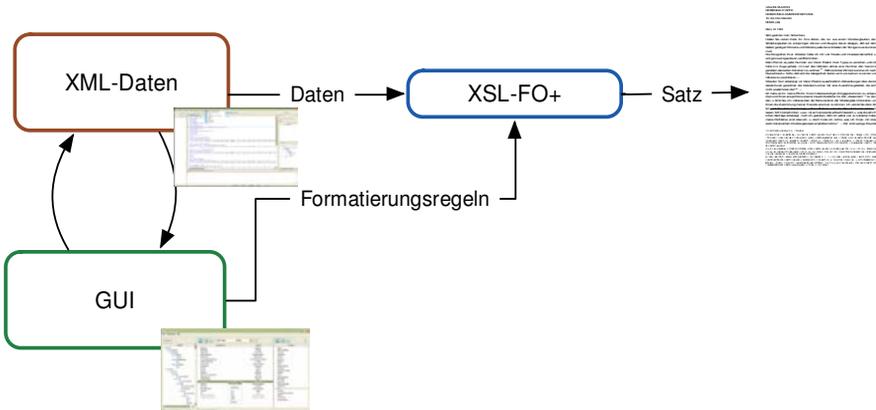


Abb. 6 Architektur von XML-Print

Jeder dieser Zwischenschritte ist für die Nutzer im Normalfall unsichtbar. Sie können aber problemlos die im Prozess erzeugten Dateien analysieren und bei Bedarf von Hand weiter anpassen.

Die GUI ist als Eclipse-Plug-in implementiert, um einerseits ein Maximum an Reaktivität zu ermöglichen, andererseits aber auch die Integration mit andere auf Eclipse aufbauenden Plattformen, vor allem TextGrid, zu vereinfachen.

Interaktion mit Nutzern

Nutzer starten üblicherweise mit der Baumdarstellung der XML-Struktur ihres Dokuments. Sie können jedes Element auswählen und mit einer Formatvorlage verknüpfen oder aber auch Elemente für den Satz ignorieren. Sollte eine Eins-zu-eins-Zuordnung nicht ausreichen, können Nutzer auch beliebige XPath-Ausdrücke mit einer Formatvorlage verbinden, um etwa Absätze, die direkt nach einer Überschrift kommen, anders zu behandeln als Absätze im Fließtext. Sie können im Nutzer-Interface auch generell Aspekte des Satzes wie die Seitenparameter und die gewünschte Silbentrennung festlegen.

Konzeptionell entsprechen die Formatvorlagen weitgehend ihren Äquivalenten in üblichen Office-Paketen. Nutzer können bestimmen, dass etwa die Formatvorlage „Gedichtzeile“ in einem bestimmten Schriftschnitt und mit anderen Layout-Eigenschaften zu formatieren ist. Sie können auch festlegen, ob diese Formatvorlage als Inline-Element, Absatz oder aber als Fußnote oder Apparat zu behandeln ist.

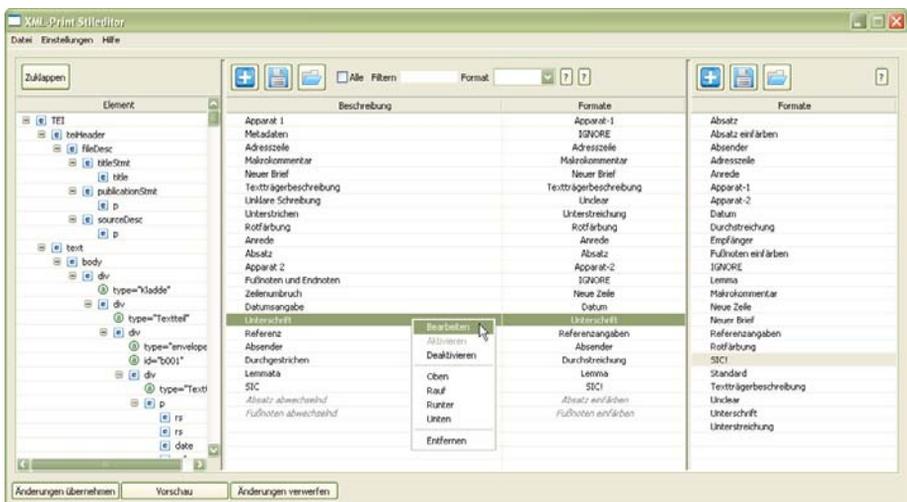


Abb. 7 Abbildung der XML-Struktur auf Formatvorlagen in XML-Print

Die für diese Ausgangsdatei gewählten Formatvorlagen und ihre Abbildungsregeln auf XML-Elemente werden auch im Zip-Archiv abgelegt und können damit anderen Nutzern verschickt oder in zukünftigen Arbeitssitzungen verfeinert werden.

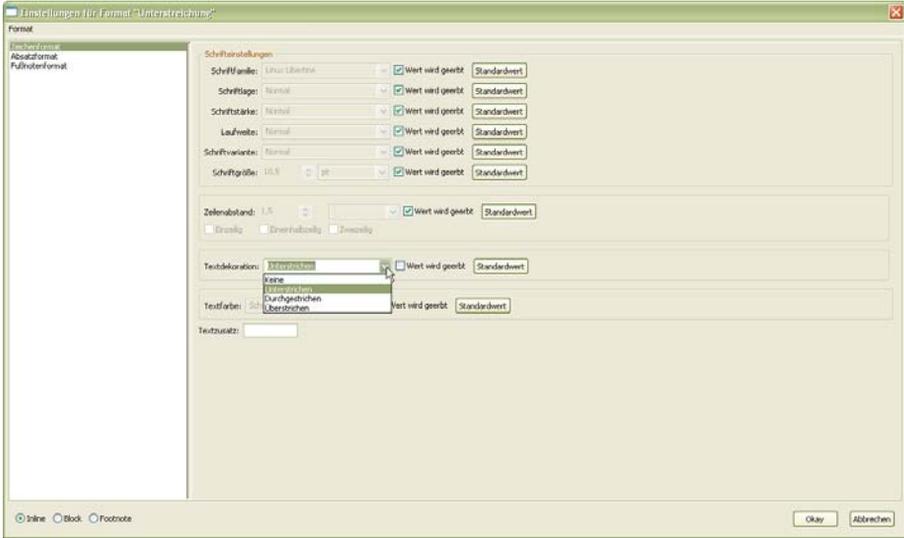


Abb. 8 Einstellungen in XML-Print

Algorithmen

Zusätzlich war es ein Ziel, als Informatiker zu testen, wie sich Satz-Algorithmen unter Einsatz einer modernen funktionalen Programmiersprache verbessern lassen. Moderne Entwicklungsumgebungen und entsprechende Bibliotheken bieten Grundfunktionalitäten etwa in der Fontbehandlung und in der Erzeugung von PDF, die vor vierzig Jahren – als Systeme wie TUSTEP und TeX ursprünglich entwickelt wurden – noch nicht vorhanden waren.

Nach einer ersten Analysephase fiel die Auswahl auf die funktionale Sprache F# (vgl. Smith 2012), die von Microsoft für das .NET-Framework entwickelt wurde. F# baut auf der etablierten funktionalen Sprache OCaml auf und ergänzt sie insbesondere um Funktionen zur Interoperabilität mit .NET. Gleichzeitig ist F# auch kompatibel zu mono, der Open-Source-Implementierung der .NET-Umgebung. Dadurch laufen in F# geschriebene Anwendungen auf allen Plattformen, für die auch mono existiert, insbesondere auf MS Windows, Mac OS X und üblichen Linux-Distributionen. Alle für .NET geschriebenen Programm-bibliotheken sind ebenfalls problemlos nutzbar, sofern sie rein in einer .NET-Sprache entstanden sind.

Zur Behandlung von Schriften setzt das Projekt auf OpenType (vgl. Microsoft/Adobe 2013), insbesondere auf die für .NET verfügbare Open-Source-iText-Bibliothek. Wir haben diese Bibliothek um die Unterstützung

fortgeschrittener OpenType-Eigenschaften wie echte Kapitälchen und bidirektionale Schriftsysteme erweitert. Dieselbe Bibliothek nutzen wir auch zur PDF-Generierung.

Funktionale Sprachen verzichten weitgehend oder ganz auf veränderbare Variablen und Zustände. Algorithmen, die nur auf unveränderbare Variablen zurückgreifen, sind weniger fehlerträchtig, kompakter und vor allem leicht zu parallelisieren. XML-Print kann auf diese Weise den Satz einzelner Kapitel parallelisieren, was auf den immer üblicheren Multicore-Prozessoren erhebliche Geschwindigkeitsvorteile bringt.

XML-Print hat außerdem neue Zeilen- und Absatzumbruchalgorithmen implementiert, die beide über die Möglichkeiten von Plass und Knuth (1980), wie sie aktuell in TeX implementiert sind, hinausgehen.¹⁷

Ebenso hat XML-Print Neuland betreten, was die Behandlung von verschiedenen Fußnotenapparaten angeht. Die Engine kann eine beliebige Zahl von Fußnotenapparaten verwalten und kümmert sich um deren Umbruch auch über Seitenwechsel hinaus.

Normen

XML-Print baut auf dem ganzen Stack von XML-Technologien auf, angefangen von XML selbst über XSLT-T, um XML in Satzinstruktionen zu transformieren, bis zu XSL-FO, um die typografische Sicht auf XML zu erlauben. Allerdings implementiert das Projekt nicht die volle Funktionalität von XSL-FO, da bestimmte Funktionalitäten für kritische Editionen nicht einschlägig sind. Andererseits fügen wir im oben erwähnten XSL-FO+ Elemente und Attribute hinzu, um die Bedürfnisse von Apparaten und Verweissystemen abzubilden.

Für die Ausgabe generiert XML-Print die für die Langzeitarchivierung vorgesehene PDF-Variante PDF/A-1 nach ISO 19005-1:2005.

Wo?

Schwitters-Briefe

Von einer frühen Projektphase an haben Mitarbeiter des Editionsprojekts „Kurt Schwitters: Wie Kritik zu Kunst wird“ (vgl. Kocher et al. 2013) XML-

¹⁷ Vgl. Georgieff et al. (2014) für Details zu den implementierten Algorithmen.

VALORI PLASTICI
RASSEGNA D'ARTE
DIREZIONE E AMMINISTRATIONE
10, Via Ciro Menotti
ROMA (49)

Rom, XI 1919

Sehr geehrter Herr Schwitters.

Haben Sie vielen Dank für Ihre Zeilen, die nur aus einem Künstlerglauben, der Schöpferglauben ist, entspringen können und Zeugnis davon ablegen, daß auf dem Gebiet geistigen Wirkens und Waltens jedes ferne Erbeben der Stringens verstummen muß.

Die Fotografien Ihrer Arbeiten habe ich mit viel Freude und Interesse betrachtet u. will gerne einiges davon veröffentlichen.

Mein Plan ist es, jeder Nummer von Valori Plastici ihren Typus zu verleihen, und ich hatte in's Auge gefasst, im Lauf des nächsten Jahres eine Nummer den hervorragendsten deutschen Künstlern zu widmen ^{a)}. Während des Winters komme ich nach Deutschland u. hoffe, daß sich die Gelegenheit bieten wird uns kennen zu lernen und näheres zu vereinbaren. –

Was den Text anbelangt, ist Valori Plastici ausschließlich Abhandlungen über die bildende Kunst gewidmet: die Kubistennummer hat eine Ausnahme gebildet, die sich nicht wiederholen darf ^{b)}. –

Ich halte es für meine Pflicht, Ihrem lebenswürdigen Entgegenkommen zu entsprechen und Ihnen einige Fotos unserer Hauptmitarbeiter für den „Zweemann“ ^{c)} zu senden, u. bitte Sie, mir näheres über die Remuneration der Wiedergabe mitzuteilen, um Ihnen die Zustimmung meiner Freunde erwirken zu können. Ich würde Sie dann bitten „mit Erlaubnis des Verlages „Valori Plastici“ unter die Reproduktionen setzen zu lassen. Mit Holzschnitten, u.s.w. ist es hierzulande schlecht bestellt u. was die schriftlichen Beiträge anbelangt, muß ich gestehen, daß ich selbst viel zu kämpfen habe; meine Pultfächer sind übertoll, u. doch finde ich nichts, was ich Ihnen mit klassisch-italienischem Künstlergewissen empfehlen könnte ¹. – Wir sind wenige Freunde

1 empfehlen könnte] So im Original

a) deutschen Künstlern zu widmen] In Valori Plastici II, 7-8 (1920) erschien ein Artikel über George Grosz und Paul Klee (S. 88f.), jedoch keine Reproduktionen von Werken dt. Künstler, sondern ausschließlich solche des italien. Künstlers Giorgio di Chirico. KS wird mit keinem Wort erwähnt und es erschienen auch in späteren Ausgaben weder Artikel von oder über ihn noch Abbildungen seiner Werke in Valori Plastici.

b) die Kubistennummer...wiederholen darf] Valori Plastici I, 2-3, (Feb./März 1919) war dem franz. Kubismus in bildender Kunst und Poesie gewidmet, obwohl sich die Zeitschrift eigentlich ausschließlich mit den bildenden Künsten auseinandersetzte.

c) Fotos unserer Hauptmitarbeiter... „Zweemann“] Im → Zweemann wurden keine Bilder der „Hauptmitarbeiter“ von Valori Plastici – italienische Künstler des Magischen Realismus – veröffentlicht. Allerdings schrieb Theodor Däubler (Triest/Österreich 1876-1934 St. Blasien/Schwarzwald), einer der Haupttexter von Valori Plastici, auch für den Zweemann.

Print eingesetzt, um ihre XML-Transkripte zu testen. In einer späteren Projektphase haben sie Formatvorlagen und Mappings für den textkritischen und den Erläuterungsapparat hinzugefügt.

Wörterbuchnetz

Das „Deutsche Wörterbuch von Jacob und Wilhelm Grimm“ ist eine digitalisierte Version des führenden deutschen Wörterbuchs mit mehr als 300.000 in XML strukturierten Einträgen. Nutzer der Onlineausgabe wünschen sich eine PDF-Ausgabe insbesondere zum Druck der oftmals langen Einträge. Das PDF wird dabei mit der über einen Web-Service eingebundenen Satz-Engine on-the-fly generiert.

The screenshot shows a web browser displaying the online German dictionary 'Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm'. The search results for the letter 'A' are shown, with a list of words on the left and a detailed entry for 'A' on the right. The entry for 'A' includes a definition: 'A, der edelste, ursprünglichste aller laute, aus brust und kehle voll erschallend, den das kind zuerst und am leichtesten hervor bringen lernt, den mit recht die alphabete der meisten sprachen an ihre spitze stellen. a hält die mitte zwischen i und u, in welche beide: A'.

Deutsches Wörterbuch von Jacob und Wilhelm Grimm. 16 Bde. in 32 Teilbänden. Leipzig: 1854-1961. Quellenverzeichnis Leipzig 1971. Online-Version vom 08.07.2014.

Abb. 10 Automatischer Satz eines Wörterbuchartikels des Grimm'schen Wörterbuchs

Wohin?

Die Hauptförderphase des Projekts endete im Mai 2014. Das Projekt selbst ist jetzt, wie bereits im Ursprungsantrag vorgesehen, in einer Phase der Programmpflege und punktueller Weiterentwicklung. Zum jetzigen Zeitpunkt stehen insbesondere ein verfeinertes System von Verweisen innerhalb des Texts, weitere Optionen zur Formatierung von Apparaten und die Unterstützung von lebenden Kolummentiteln auf dem Programm.

Neben der Weiterentwicklung des Systems selbst arbeiten wir daran, die Community auf SourceForge zu stärken. Aktuell haben wir im Mittel über fünfzig monatliche Downloads. Gleichzeitig versuchen wir, weitere Projekte zu gewinnen, die XML-Print praktisch einsetzen und damit an seiner zukünftigen Evolution mitarbeiten können. Ein wesentlicher Schritt dahin ist auch die Integration in TextGrid, um die Nutzung von XML-Print innerhalb der TextGrid-Gemeinschaft zu erleichtern.

Literaturverzeichnis

- Neuroth et al. (Hrsg.) (2010): *nestor-Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.3*. <http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:0008-2010071949>.
- Apache Foundation (2013): Apache FOP – a print formatter driven by XSL formatting objects and an output independent formatter. <http://xmlgraphics.apache.org/fop/>.
- Berglund, A. (Hrsg.) (2006): Extensible Stylesheet Language (XSL) Version 1.1. <http://www.w3.org/TR/2006/REC-xsl11-20061205/>.
- Burch, T. et al. (2012): XML-Print: an Ergonomic Typesetting System for Complex Text Structures. In: *Digital Humanities*, Hamburg, S. 375–379. http://www.dh2012.uni-hamburg.de/wp-content/uploads/2012/07/HamburgUP_dh2012_BoA.pdf.
- Burnard, L.; Bauman, S. (2007): TEI P5: Guidelines for Electronic Text Encoding and Interchange. Text Encoding Initiative.
- Constant, B. (1999): *De la religion, considérée dans sa source, ses formes et ses développements. Tome 2*, hrsg. v. Pierre Deguise und Kurt Kloocke. Berlin: de Gruyter.
- Georgieff, L. et al. (2014): XML-Print. Typesetting arbitrary XML documents in high quality. In: *Digital Humanities*, Lausanne, S. 169–171. https://dh2014.files.wordpress.com/2014/07/dh2014_abstracts_proceedings_07-11.pdf.
- Knuth, D. E. (1986): *The TEXbook* (Computers & Typesetting; Vol. A). Reading, Mass. u. a.: Addison-Wesley.
- Kocher, U.; Schultz, I. (2013): Wie Kritik zu Kunst wird: Kurt Schwitters' Strategien der produktiven Rezeption. <http://www.avl.uni-wuppertal.de/forschung/projekte/wie-kritik-zu-kunst-wird.html>.

- Küster, M. W. et al. (2013): XML-Print: Addressing Challenges for Scholarly Typesetting. In: *Digital Humanities*, Lincoln, NE, S. 269–272. http://dh2013.unl.edu/abstracts/files/downloads/DH2013_conference_abstracts_print.pdf.
- Lamport, L. (1986): *LaTeX: User's Guide & Reference Manual*. Reading, Mass. u. a.: Addison-Wesley.
- McLuhan, M. (1962): *The Gutenberg galaxy. The making of typographic man*. University of Toronto Press.
- Microsoft/Adobe (Hrsg.) (2013): Microsoft Typography – OpenType Specification. <http://www.microsoft.com/typography/otspec/default.htm>.
- Ott, W. (2008): *TUSTEP-Handbuch*. Ausgabe 2008. Tübingen: Universität Tübingen.
- Plass, M. F.; Knuth, D. E. (1980): Breaking paragraphs into lines. In: *Software: Practice and Experience* 11: 1119–1184. <http://bowman.infotech.monash.edu.au/~pmoulder/line-breaking/knuth-plass-breaking.pdf>.
- Smith, C. (2012): *Programming F# 3.0*. 2. Aufl., Sebastopol, CA: O'Reilly Media.

Ein virtuelles Bücherregal: Die *Digitale Bibliothek* im TextGrid Repository

von *Katrin Betz*

Die *Digitale Bibliothek* ist eine umfangreiche Sammlung von Texten vom Anfang des Buchdrucks bis zu den ersten Jahrzehnten des 20. Jahrhunderts. Insbesondere für die Literaturwissenschaft ist die digitale Textsammlung von großem Interesse: Enthalten sind zahlreiche literaturhistorisch relevante Texte und nahezu alle kanonisierten Texte, deren urheberrechtliche Schutzfrist abgelaufen ist. Als Digitalisierungsquelle für die Sammlung wurden zum größten Teil Studienausgaben oder Erstdrucke verwendet, sodass die Texte zitierfähig sind. Ähnliches gilt für die Philosophie und Kulturwissenschaften.¹

Die Texte der *Digitalen Bibliothek* wurden von Editura digitalisiert und mit einem layout-basierten XML-Markup strukturiert.² TextGrid hat die XML-Daten erworben und stellt sie der Allgemeinheit nach verschiedenen Aufbereitungsschritten in einer semantisch-orientierten TEI-Auszeichnung unter der Creative Commons Lizenz „by“ Version 3.0 zur Verfügung.³

Ein Großteil der Sammlung (der Literaturordner) ist bereits in das TextGrid Repository importiert. Der Rest des Datenbestandes ist derzeit in Bearbeitung und soll ebenfalls in das Repository integriert werden.

Im folgenden Beitrag wird die Aufbereitung des Literaturordners beschrieben. Dabei wird auf die Struktur der Originaldaten, die Zielstruktur, ausgewählte Bearbeitungsschritte und die Abbildung der Datenstruktur auf das TextGrid-Metadatenmodell eingegangen.⁴

1 Vgl. <http://textgrid.de/ueber-textgrid/digitale-bibliothek/>.

2 Siehe <http://www.zeno.org>.

3 Vgl. <http://textgrid.de/ueber-textgrid/digitale-bibliothek/> zu detaillierten Informationen zur Lizenzierung.

4 Die Stylesheets zur Transformation des Literaturordners nach TEI wurden größtenteils von George Dogeru entworfen. Der Entwurf der Zielstruktur und die Auszeichnung der Werkebene wurden von Mirjam Blümm bearbeitet.

Die Originaldaten

Die Texte der erworbenen Sammlung gliedern sich in folgende Themenbereiche:

- Geschichte (14 Autorendateien)
- Kulturgeschichte (113 Autorendateien)
- Kunst (zwölf Autorendateien)
- Literatur (693 Autorendateien)
- Märchen (58 Autorendateien)
- Musik (81 Autorendateien)
- Naturwissenschaften (20 Autorendateien)
- Philosophie (248 Autorendateien)
- Soziologie (eine Autorendatei)
- Nachschlagewerke (27 Autorendateien)

Ordnerstruktur

Für jeden Themenbereich ist ein eigener Ordner vorhanden. Innerhalb der thematisch sortierten Ordner sind die Texte wiederum in einzelne Dateien aufgeteilt; die Werke jeweils eines Autors sind in einer einzigen Datei gespeichert.⁵

Zu jedem Themenbereich ist außerdem eine Katalogdatei vorhanden, in der die bibliografischen Angaben der Digitalisierungsquelle gespeichert sind. Die Katalogdateien sind ebenfalls XML-kodiert. Dieses Markup ermöglicht den gezielten Zugriff auf einzelne bibliografische Angaben und die Verknüpfung mit den Texten in den Autorendateien. Die Information zu einer bestimmten Digitalisierungsquelle liegt dabei als String vor, d. h. einzelne Metadatenfelder sind nicht ausgezeichnet.

Struktur des XML-Markups der Texte

Die Daten sind durch die rekursive Verwendung eines Elements (*article*) strukturiert. Dieses Element wird verwendet, um mehrere Werke einer bestimmten Textsorte wie Briefe, Romane, Gedichte oder Dramen zu umfassen,

⁵ Die Zusammenstellung der Autoren- und Textauswahl der Sammlungen basiert nicht auf textwissenschaftlichen Kriterien.

aber auch um Einzelwerke oder Texteinheiten wie Kapitel, Akte und Szenen auszuzeichnen.

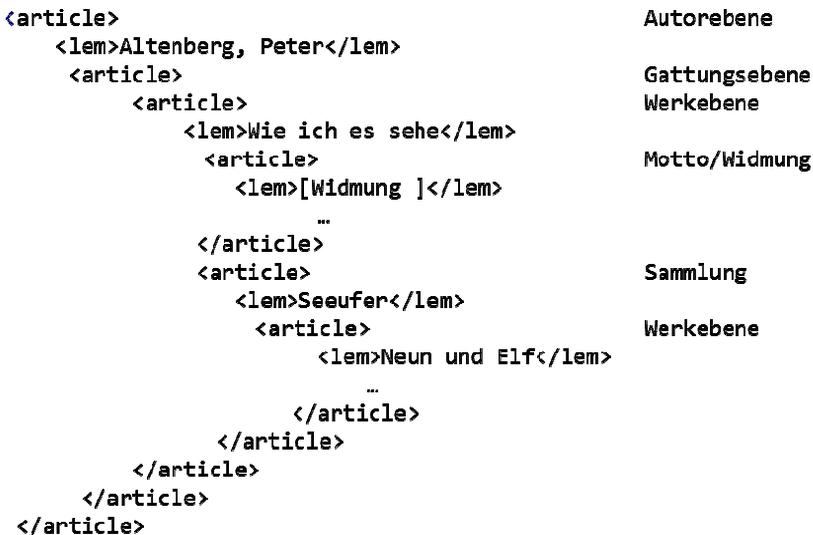


Abb. 1 *article*-Struktur der *Digitalen Bibliothek*

Die Texte selbst sind mit einem stark layout-basierten Markup versehen, das teilweise an (X)HTML erinnert. So werden z. B. die Absätze eines Romans ebenso mit einem *p*-Tag kodiert wie die einzelnen Verszeilen eines Gedichtes. Die Strophen eines Gedichtes sind durch ein *br*-Element voneinander abgesetzt. Weitere Layout-Anweisungen sind in Form von *class*-Attributen vorhanden.⁶

Aufbereitung der Daten

Bei der Aufbereitung der Daten wurden zwei Ziele verfolgt: Einerseits muss die rekursive Struktur des *article*-Elements auf eine *TEI/teiCorpus*-Struktur abgebildet werden, die wiederum kompatibel zum TextGrid-Metadatenmodell ist. Andererseits soll das layout-basierte Markup in ein semantisch-orientiertes TEI-Markup überführt werden. Im Folgenden wird zunächst anhand

⁶ Vgl. für Beispiele zur Auszeichnung der Feinstruktur und deren Transformation den Abschnitt „Aufwärtstransformation in ein semantisch-orientiertes TEI-Markup“ weiter unten.

von Beispielen die Feinauszeichnung der Daten beschrieben und anschließend die Entwicklung der *TEI/teiCorpus*-Struktur und deren Abbildung auf das TextGrid-Metadatenmodell erläutert. Die tatsächliche Aufbereitung der Daten erfolgt in insgesamt elf Transformationsschritten. In den folgenden Kapiteln wird nur ein Ausschnitt der Aufbereitung beschrieben, der auf die oben genannten Ziele fokussiert. Die einzelnen Unterschritte der Implementierung werden hier nicht aufgezeigt.

Aufwärtstransformation in ein semantisch-orientiertes TEI-Markup

Spezifische Aufwärtstransformation wurden für die Textgattungen Drama, Roman und Gedicht durchgeführt. Die Transformationen basieren auf Heuristiken, deren Funktionsweise exemplarisch am Beispiel eines Dramas gezeigt werden soll. Die folgenden Zeilen zeigen XML-Ausschnitte eines Dramas in der Originalform und der Zielstruktur.

```
<p class="zenoPC">
    <i>Feldlager. Nacht. Walther von Merheim
    und Ulrich Zenger ruhen, in ihren Mäntel
    gehüllt, auf dem Boden.
    </i>
</p>
<p>ZENGER.</p>
<p class="zenoPLm4n4"> Eine wahre Hexennacht! Die Nebel jagen...
</p>
```

Abb. 2 Ausgangsstruktur der *Digitalen Bibliothek*

```
<stage hi="italics"> Feldlager. Nacht. Walther von Merheim und Ulrich Zenger
ruhen, in ihre Mäntel gehüllt, auf dem Boden.
</stage>
<sp>
    <speaker>ZENGER.</speaker>
    <p> Eine wahre Hexennacht! Die Nebel jagen </p>
</sp>
```

Abb. 3 Zielstruktur der *Digitalen Bibliothek*

Am ersten Abschnitt kann die Auszeichnung der Bühnenanweisung als *stage* nachvollzogen werden: In den Originaldaten ist eine Bühnenanweisung durch die Verschachtelung von *p/i* kodiert, sodass diese Elementkombination leicht in ein *stage*-Element überführt werden kann.

Zur Auszeichnung der Sprechereinsätze werden zwei Struktureigenschaften überprüft: Einerseits werden die *p*-Elemente der Dramen auf ihre Länge getestet. Ist der Text des *p*-Elementes sehr kurz, handelt es sich möglicherweise um die Angabe des Sprechernamens. Zusätzlich wird überprüft, ob der Text nur aus Majuskeln besteht. Treffen beide Eigenschaften zu, wird angenommen, dass es sich um eine Sprecherangabe handelt. Auf dieser Basis werden Gruppierungen gebildet, wobei eine Gruppe jeweils mit einem Sprecher-*p* beginnt und dann mit *sp* ausgezeichnet wird.

Auf der Basis dieser und ähnlicher Mechanismen wurden alle Dramen, Gedichte etc. ausgezeichnet. Problematisch hierbei ist, dass zwischen den Layout-Anweisungen und einer semantischen Interpretation kein 1:1-Verhältnis vorliegt, sodass in den Daten nicht alle Sprecher als solche erkannt und andererseits Texteinheiten fälschlicherweise als Sprecher markiert wurden. Ähnliche Fehlerquellen sind bei der Auszeichnung anderer semantischer Einheiten wie *lg*, *closer* oder *head* hinderlich.

Entwicklung einer TEI/teiCorpus-Struktur

Aufgrund der rekursiven Verwendung des *article*-Elementes ist es nicht möglich, maschinell herauszufiltern, bei welchen Einheiten es sich um Werke handelt. Diese Information ist aber wiederum notwendig, um Werke als solche in das Repository zu importieren. Deshalb wurde eine Browser-Oberfläche entworfen, die eine schnelle manuelle Auszeichnung der Werkebene ermöglicht.

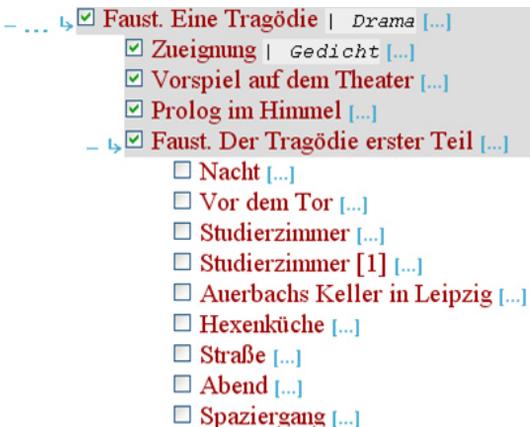


Abb. 4 Oberfläche zur manuellen Auszeichnung der Werkebene

Nach der Werkauszeichnung kann die *article*-Struktur in eine *teiCorpus/TEI*-Struktur überführt werden. Um dabei Konflikte mit dem TextGrid-Metadatenmodell zu vermeiden, gleichzeitig aber eine gültige TEI-Struktur zu erhalten, mussten hierbei auch Texteinheiten wie Widmungen oder Vorwörter als eigenes TEI-Element (anstelle von beispielsweise *front*) kodiert werden, da zumeist nicht zu unterscheiden war, um welche Struktureinheiten es sich tatsächlich handelt.

Extraktion der Metadaten

Parallel dazu wurde in einem weiteren Transformationsschritt für jedes Werk aus der Katalogdatei die zugehörige bibliografische Angabe der Digitalisierungsquelle extrahiert und in den *teiHeader* geschrieben. Da die Angabe der Metadaten in der Katalogdatei als String vorliegt und in diesem die Reihenfolge oder auch die Trennzeichen nicht immer einheitlich sind, war es nur begrenzt möglich, Metadaten in eine strukturierte Form zu bringen. Die Extraktion beschränkt sich deshalb auf Informationen wie den Publikationsort und die Daten der Entstehung oder der Erstveröffentlichung. Die Publikationsorte wurden halbmanuell aus den Daten extrahiert.

Die Datumsangaben wurden entweder als konkrete Zahl oder in Form von Zeiträumen kodiert. Konnte aus einem eindeutig identifizierbaren String ein Datum in Verbindung mit einem eindeutigen Schlüsselwort (z. B. *entstanden* oder *Erstveröffentlichung*) ausfindig gemacht werden, wurde das entsprechende Datum kodiert. Sind in dem Metadaten-String mehrere Daten ohne zuordenbare Schlüsselwörter vorhanden, wurden aus dem String jeweils das größte und das kleinste Datum herausgefiltert. Diese Daten wurden mit den Geburtsdaten des Autors verglichen. Kodiert wurde in diesem Fall in den beiden Attributen *notBefore* und *notAfter* der kleinste sinnvolle Zeitraum.

Abbildung der Struktur auf das TextGrid-Metadatenmodell

Das TextGrid-Metadatenmodell stellt für den Datenimport die Objekttypen *item*, *edition*, *work* und *collection* zur Verfügung. Jedes der Objekte besteht jeweils aus einer Metadatendatei und einer Inhaltsdatei. Ein *item* kann als Inhaltsdatei eine XML-Datei, eine *jpeg*-Datei oder aber eine Aggregation enthalten. Die Metadatendatei des Items enthält als obligatorische Angabe nur den Rechteinhaber. Ein *work*-Objekt besteht aus einer leeren Inhaltsdatei,

die zugehörige *work*-Metadaten-datei enthält Informationen über den Autor, das Entstehungsdatum etc. Zu einem *work*-Objekt muss notwendigerweise zusätzlich ein *edition*-Objekt erstellt werden. Die Inhaltsdatei dieses Objektes ist im Falle der *Digitalen Bibliothek* eine Aggregation, die das jeweilige Werk aggregiert. Besteht das Werk aus mehreren Einzelwerken und/oder mehreren *items*, werden diese aggregiert. In der Metadaten-datei der *edition* sind die Angaben zu der Digitalisierungsquelle kodiert. Die Inhaltsdatei der *edition* verweist auf das *item* und kann zusätzlich weitere Unterwerke oder *items* aggregieren. Abbildung 5 gibt einen Überblick über die verwendeten Objekte.⁷

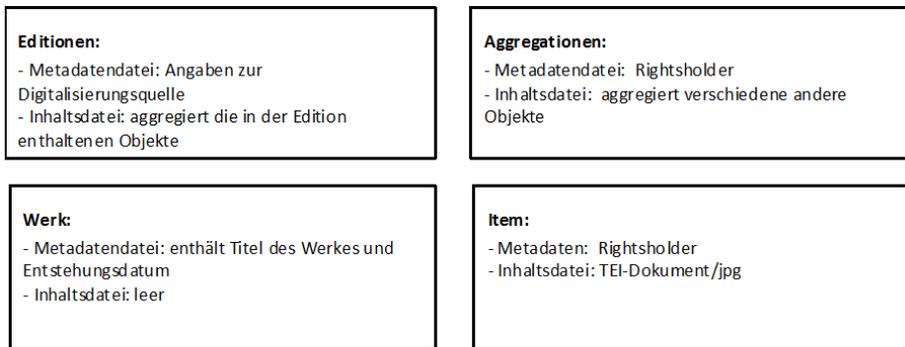


Abb. 5 Übersicht über die verwendeten Objekte des TextGrid-Metadatenmodells

Zur Überführung der *TEI/teiCorpus*-Struktur in das TextGrid-Metadatenmodell wurden die einzelnen Autorendateien gesplittet. Dabei wurden folgende Abbildungsregeln entwickelt:

- Jedes *TEI*-Element, das als Werk im Repository vorliegen soll, wird als eigenständige *TEI*-Datei importiert. Dazu müssen die Dateien für ein *item*-Objekt, ein *work*-Objekt und ein *edition*-Objekt erstellt werden.⁸ Die *item*-Inhaltsdatei ist die *TEI*-Datei selbst. Deren Metadaten-datei sowie die Metadaten-dateien der zugehörigen *work*- und *edition*-Objekte werden auf der Basis der Metadaten des *teiHeaders* generiert.

⁷ Vgl. zum Metadatenmodell auch Pempe (2012) und die Dokumentation unter <https://dev2.dariah.eu/wiki/display/TextGrid/Metadata>.

⁸ Zur logischen Struktur und zur Verbindung zwischen den einzelnen Objekttypen vgl. <https://dev2.dariah.eu/wiki/display/TextGrid/Metadata>.

- *TEI*-Elemente, die kein Werk sind, werden auf *items* abgebildet. Die *TEI*-Datei ist wieder die Inhaltsdatei und die Metadatenfile wird zusätzlich erzeugt.
- *teiCorpus*-Elemente, die selbst Werke sind, werden bei der Erstellung der TextGrid-Metadaten auf ein *edition*-Objekt abgebildet. Die Inhaltsdatei der *edition* aggregiert die Unterwerke oder *items*. Die Metadaten der *edition* werden auf der Basis der im *teiHeader* kodierten Daten erzeugt.
- *teiCorpus*-Elemente, die selbst nicht als Werk gelten, werden auf ein *item* abgebildet, dessen Inhaltsdatei wiederum eine Aggregation ist.
- *teiCorpus*-Elemente werden vollständig durch die Aggregationsmechanismen des TextGrid-Metadatenmodells ersetzt.

Die so aufbereiteten Daten werden anschließend innerhalb eines *collections*-Objektes aggregiert und über CoLibRI in das TextGrid Repository importiert. Im Laufe des Imports wurden PIDs vergeben und weitere Metadaten automatisch generiert (vgl. hierzu auch Funk et al. 2013).

Veröffentlichung des Literaturordners

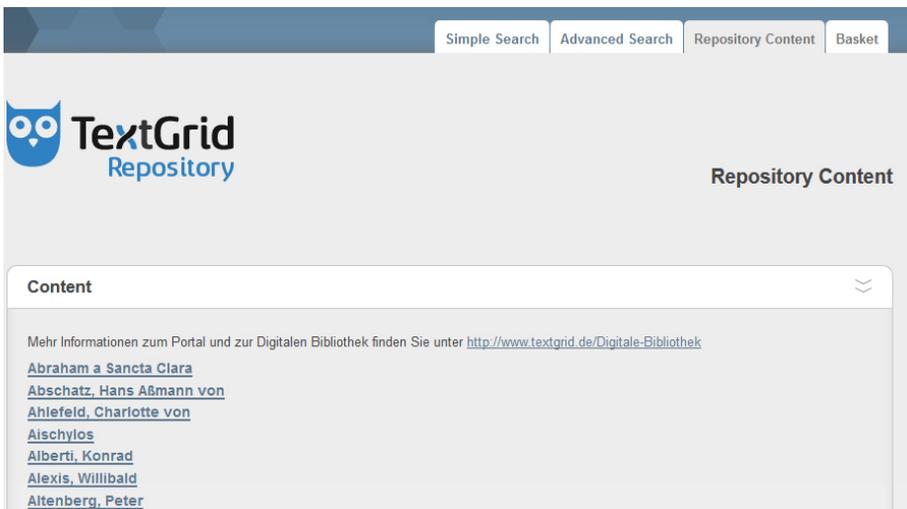


Abb. 6 Die *Digitale Bibliothek* im TextGrid Repository

Der Literaturordner wurde im Juli 2011 in das Repository importiert.⁹ Im Repository können die Texte gelesen und durchsucht werden. Die Verknüp-

⁹ <http://textgridrep.de/>

fung des Repository mit Tools von Digivoy ermöglicht zudem Auswertungen der Daten, die über einen Aggregator zusammengestellt werden können.¹⁰

Zusätzlich wird der gesamte Inhalt des Literaturordners als Datendump zum Download angeboten.¹¹ In diesem Dump sind die Daten nicht gesplittet, d. h. für jeden Autor ist ebenso wie in den Ursprungsdaten eine Datei vorhanden. Außerdem sind die Daten nicht mit PIDs versehen.¹²

Umfang des Literaturordners

Grundlage der Zählung ist der Dump unter <http://textgrid.de/ueber-textgrid/digitale-bibliothek/>:

- Autorenanzahl: 697 (693 Textdateien, z. T. enthält eine Datei Werke von mehreren Autoren, z. B. Gebrüder Grimm; Anonym zählt als ein Autor);
- *work*-Objekte im Literaturordner: 93.462;
- Wörter (Folge von Nicht-Whitespace-Zeichen, die zwischen Whitespace-Zeichen steht): 134.316.246 Wörter;
- 860.932.040 Bytes in den Textdateien.

Ausblick

Der Literaturordner ist weiterhin in Bearbeitung und soll in einer zweiten Version veröffentlicht werden. Für diese Version wurden gemeldete Bugs behoben und die Heuristiken zur Extraktion der Metadaten ergänzt.¹³

Parallel dazu werden die restlichen Ordner der *Digitalen Bibliothek* bearbeitet. Hierbei werden die vorhandenen Transformationsroutinen angewendet und an die jeweiligen Daten angepasst. Bei der Übertragung wird zusätzlich darauf geachtet, die Stylesheets sukzessive modularer und generischer zu gestalten, um die Anwendung zu erleichtern und den Code lesbarer und flexibler zu gestalten.

10 Vgl. hierzu <https://de.dariah.eu/web/guest/digivoy>.

11 <http://textgrid.de/ueber-textgrid/digitale-bibliothek/>

12 Für eine ausführliche Übersicht über die Unterschiede zwischen den Daten im Repository und dem Datendump vgl. <http://textgrid.de/ueber-textgrid/digitale-bibliothek/>.

13 Eine Liste bekannter Bugs findet sich unter <http://textgrid.de/ueber-textgrid/digitale-bibliothek/>.

Literaturverzeichnis

- Funk, Stefan E.; Veentjer, Ubbo; Vitt, Thorsten (2013). Die Virtuelle Forschungs-
umgebung TextGrid – Status quo und neue Entwicklungen. In: Neuroth, Heike
et al. (Hrsg.): *Evolution der Informationsinfrastruktur: Kooperation zwischen
Bibliothek und Wissenschaft*. Glückstadt: Verlag Werner Hülsbusch, S. 277–299.
http://webdoc.sub.gwdg.de/univerlag/2013/Neuroth_Festschrift.pdf.
- Pempe, Wolfgang (2012): Geisteswissenschaften. In: Neuroth, H. et al. (Hrsg.):
Langzeitarchivierung von Forschungsdaten – Eine Bestandsaufnahme. Boizen-
burg: Verlag Werner Hülsbusch, S. 137–159. [http://nestor.sub.uni-goettingen.de/
bestandsaufnahme/nestor_lza_forschungsdaten_bestandsaufnahme.pdf](http://nestor.sub.uni-goettingen.de/bestandsaufnahme/nestor_lza_forschungsdaten_bestandsaufnahme.pdf).
- TEI Consortium (Hrsg.) (2011): TEI P5: Guidelines for Electronic Text Encoding
and Interchange, <http://www.tei-c.org/Guide-lines/P5/>.

V.
**Von der Community
für die Community:
Potenziale und Perspektiven**

Das vielschichtige und komplexe Thema Nachhaltigkeit

Ein Interview mit Heike Neuroth



Frau Neuroth, welche Aspekte umfasst das Thema Nachhaltigkeit für Sie?

Heike Neuroth: Ich konnte dazu in den letzten Jahren viel lernen, nicht immer ganz freiwillig und für mich und die beteiligten fachwissenschaftlichen Kolleginnen und Kollegen zum Teil auch unerwartet. Dazu später aber mehr ...

Das Thema Nachhaltigkeit ist vielschichtig und aufgrund mehrerer Faktoren komplex. Ich unterscheide mittlerweile folgende Aspekte der Nachhaltigkeit und bin zu der Ansicht gekommen, dass wir für die einzelnen Bereiche verschiedene Lösungen mit unterschiedlichen Experten, Stakeholdern etc. brauchen und auch unterschiedlich weit vorangekommen sind:

- die fachwissenschaftliche Nachhaltigkeit,
- die datentechnische Nachhaltigkeit,
- die technische Nachhaltigkeit,
- die betriebliche Nachhaltigkeit und
- begleitende fachwissenschaftliche Aktivitäten

Die **fachwissenschaftliche** Nachhaltigkeit umfasst die beiden Themenbereiche Forschung und Lehre. Die Aktivitäten im Bereich „Lehre und Qualifizierung“ haben zum Beispiel das Ziel, neue Studiengänge im Bachelor- und Masterbereich zu etablieren, spezifische Doktoranden- oder Stipendienprogramme zu entwickeln und insgesamt in den Curricula dafür Sorge zu tragen, dass die Studierenden mit den verschiedenen Aspekten der digitalen Transformation und den Digital Humanities vertraut werden, gerade auch im Hinblick auf neue Berufsfelder. Der Bereich Qualifizierung umfasst zum Beispiel themenbezogene Workshops und (Nutzer-) Schulungen. Dies ist ein besonders wichtiger Aspekt, wie sich in TextGrid herausgestellt hat. Hier ist es im Laufe der letzten Jahre gelungen, die unterschiedlichen Fachdisziplinen

an „einen Tisch zu bringen“ und einen inspirierenden und notwendigen Austausch unter den Forschungsprojekten anzuregen. Das Feedback – z. B. Kritik hinsichtlich unverständlicher Hilfetexte oder der Wunsch nach ansprechenden Video-Tutorials –, Gemeinsamkeiten und Unterschiede in den Forschungsfragen und -methoden sind als wertvoller Input in die Weiterentwicklung und den Ausbau der Forschungsinfrastruktur sowie der Dienste und Werkzeuge geflossen. Gerade weil die geisteswissenschaftlichen Disziplinen untereinander zum Teil nicht besonders vernetzt sind, sind qualifizierende Angebote und Plattformen für den Austausch sehr wichtig, um die Community zu erreichen und zu verstehen. Uns ist es in unseren Tätigkeiten immer sehr wichtig gewesen, gemeinsam mit der jeweiligen Fach-Community die Forschung und Entwicklung in TextGrid und später auch in DARIAH-DE voranzutreiben.

Die *datentechnische* Nachhaltigkeit beschäftigt sich zuvorderst mit dem Nachweis, der Nachnutzung und der persistenten Speicherung von wissenschaftlichen Sammlungen und Forschungsdaten der Geistes- und Kulturwissenschaften. Dies soll idealerweise so realisiert sein, dass auf der Ebene der Forschungsdaten(sammlungen) Suchanfragen durchgeführt werden können und ein direkter Zugriff auf die relevanten wissenschaftlichen Sammlungen bzw. Daten erfolgt. Zurzeit ist es zum Beispiel in den Geistes- und Kulturwissenschaften (noch) nicht möglich, sich mittels eines One-stop-Zugriffs einen Überblick über die „Landschaft“ der bereits digital vorliegenden fachspezifischen Sammlungen zu verschaffen. Darüber hinaus ist es wichtig, Art, Qualität und Medialität von wissenschaftlichen Sammlungen einheitlich zu dokumentieren, um ihre Eignung für die Forschungsfrage z. B. im Text- und Data-Mining-Bereich abklären zu können (Image-Digitalisierung, „schmutzige“ Volltexterfassung, hochwertige Volltexterfassung zur Eignung von Text- und Data-Mining-Fragestellungen etc.).

Des Weiteren spielen rechtliche Aspekte immer noch eine ganz entscheidende Rolle und sind insofern überhaupt noch nicht geklärt, als dass sich die meisten Forschenden bei der Nachnutzung, Anreicherung und Weiterprozessierung von Daten Anderer in einer extrem unsicheren bzw. unklaren Rechtslage sehen. Wir sind immer noch weit entfernt von einheitlichen Lizenzbedingungen für wissenschaftliche Sammlungen und für Forschungsdaten unterschiedlichster Provenienz. Ein Netzwerk von relevanten Forschungsdaten-Archiven in den Geistes- und Kulturwissenschaften in Deutschland, Europa oder gar international ist noch nicht in Sicht, auch wenn die EU durch die Anschubförderung von zwei ESFRI-Projekten (DARIAH,

CLARIN) sicherlich schon einen entscheidenden Schritt in diese Richtung unternommen hat. Zu der datentechnischen Nachhaltigkeit gehört auch, sich quer über die verschiedenen Fachdisziplinen über eine gemeinsame Sprache und Terminologie im Bereich Forschungsdatenmanagement zu verständigen. In DARIAH-DE gibt es z.B. eine Expertengruppe, die fachübergreifend einen „Research Data Lifecycle“ entwickelt, der unterschiedliche fachwissenschaftliche Anforderungen an das Forschungsdatenmanagement sammelt und bei der Entwicklung des Referenzmodells berücksichtigt.

Die *technische* Nachhaltigkeit befasst sich mit der Stabilisierung der fachwissenschaftlichen Dienste, der verlässlichen Bereitstellung von Hardware-Komponenten wie z.B. Speicherplatz, Rechenleistung und Virtuelle Maschinen. Insgesamt muss der komplette „Service-Lifecycle“ behandelt werden, von der Konzeption über die Entwicklung bis hin zur dauerhaften Bereitstellung inklusive pflegerische und innovative (Weiter-)Entwicklung der Dienste und Werkzeuge. Zurückschrecken sollte man auch nicht vor der Abwicklung bzw. dem Abschalten von Diensten. Dies sollte natürlich auch mit „Sinn und Verstand“, einer sorgfältigen Analyse und der richtigen Kommunikationsstrategie geschehen. Neue (fachwissenschaftliche) Dienste, die innerhalb oder außerhalb von Projekten und Forschungsinfrastrukturen entwickelt werden, gilt es nach definierten Qualitätskriterien und Policies zu integrieren. Nachhaltigen Betrieb erfordert ebenso die Speicher-Infrastruktur für fachwissenschaftliche Forschungsdaten, die technische Langzeitarchivierung und -verfügbarkeit leisten muss und die optimalerweise redundant, ortsverteilt, stabil und dauerhaft angeboten werden soll. Idealerweise bietet die technische Infrastruktur eine Art „Menükarte“, mit der sich Forschungsverbände im Vorfeld darüber informieren können, welche Dienste und welcher technische Support angeboten werden und was dies kostet. Diese (teil)automatisch generierte Übersicht könnte dann Drittmittel-Anträgen beigelegt werden; die Förderer könnten dies mitbegutachten lassen und als eine Art „Daten-Pauschale“ oder „Infrastruktur-Pauschale“ mitfinanzieren. Unserer Erfahrung nach sind gerade kleinere Verbundprojekte oder Vorhaben an Standorten mit Infrastruktureinrichtungen (Bibliothek, Rechenzentrum), die noch keine umfangreiche Erfahrung mit Digital Humanities sammeln konnten, sehr froh, einen Ansprechpartner zu haben, der ihnen in allen Belangen der technischen Nachhaltigkeit hilft. Großer Bedarf besteht zurzeit in der datentechnischen Nachhaltigkeit inklusive Speicherung, Nachnutzung und Interpretierbarkeit von Forschungsdaten. Mit der zunehmenden Digitalisie-

rung der Wissenschaft werden aber auch die anderen Bereiche wie Portale, Werkzeuge etc. immer relevanter.

Die *betriebliche* Nachhaltigkeit stellt sicher, dass Digital-Humanities-Projekte idealerweise schon vor Antragseinreichung via „Menükarte“ die benötigten digitalen Dienste aussuchen und dann eine Liste der möglichen Optionen (inklusive Kosten) erhalten. Hier soll dafür gesorgt werden, dass nach Ablauf der Förderung von TextGrid durch das BMBF das Forschungs-Infrastrukturprojekt DARIAH-DE betriebswirtschaftliche und rechtliche Kompetenz aufbaut, damit im Rahmen von TextGrid entwickelte Funktionalitäten in der zukünftigen Service-Unit von DARIAH-DE dauerhaft bereitgestellt werden können. So wird sichergestellt, dass die AntragstellerInnen bereits bei der Projektplanung eine solide Finanzierung für die Nutzung der Forschungsinfrastruktur berücksichtigen können. Ein Vorteil dieser Herangehensweise besteht auch darin, dass sich die Forschungsprojekte zu Beginn ihrer Projektlaufzeit nicht erst langwierig um die Einrichtung von Servern, Virtuellen Maschinen, Datenbanken etc. kümmern müssen, sondern sich ganz auf ihre Forschungsfragen konzentrieren können. Rückmeldungen assoziierter Forschungsprojekte zu TextGrid und DARIAH-DE ergaben, dass die Forschenden offenbar noch nie so schnell alle IT-relevanten Fragen beantwortet und umgesetzt hatten. Austausch, Angebote und konkrete Bedarfsanalysen beschleunigen also den Forschungsprozess und unterstützen nachhaltige Aspekte im erheblichen Umfang.

Begleitende wissenschaftliche Aktivitäten knüpfen an aktuelle Diskussionen zu Forschungsinfrastrukturen und Virtuellen Forschungsumgebungen an. Sie fragen nach Erfolgskriterien für die neuartigen Aufgaben, Methoden und Verfahren der Digital Humanities. So soll beispielsweise analysiert werden, wie der Impact digitaler Forschungsinfrastrukturen überhaupt gemessen werden kann und ob Nutzerzahlen oder gesellschaftlich relevante Forschungsthemen tatsächlich das „Maß aller Dinge“ darstellen. Eine nicht zu unterschätzende Herausforderung stellt der Bereich „Usability“ dar – insbesondere, wenn sie sich an FachwissenschaftlerInnen unterschiedlich ausgeprägter IT-Affinität und Fachdisziplinen wendet. Die Forderung der geistes- und kulturwissenschaftlichen Community nach intuitiver Bedienungsfreundlichkeit ist ernst zu nehmen, da die Akzeptanz von Digital-Humanities-Angeboten auch von diesem Faktor abhängt. Diese Fragestellungen sollten ebenfalls berücksichtigt werden, um auch von dieser Perspektive her Forschungsinfrastrukturen und Virtuelle Forschungsumgebungen zu stabilisieren und so zu ihrer Nachhaltigkeit beizutragen.

Einige andere Aspekte, die ich jetzt nicht explizit genannt habe, spielen natürlich auch eine Rolle. Neue Berufsfelder müssten etabliert werden. Im Rahmen von Digital-Humanities-Lehrveranstaltungen höre ich oft die Frage „Was kann ich denn mit meinem Abschluss anfangen, wo kann ich später mal arbeiten?“. Dies bedeutet für mich, dass wir einerseits die jungen Menschen viel früher im Studium mit allen Aspekten der digitalen Transformation vertraut machen müssen und uns andererseits wesentlich stärker darum kümmern müssen, die gesellschaftspolitische Verzahnung zwischen Hochschule und Gesellschaft hinzubekommen. Es ist ja nun auch so, dass nicht alle für den Rest ihres Berufslebens in der akademischen Welt verbleiben, sondern im Gegenteil nur ein kleiner Prozentsatz eine Dauerstelle an einer Hochschule oder Forschungseinrichtung erlangt.

Wir sehen, dass das Thema Nachhaltigkeit vielschichtig ist und ganz unterschiedliche Komponenten beinhaltet. Meines Erachtens kann es aber nur im Zusammenspiel gelingen, eine Nachhaltigkeit für digitale Forschungs- und Informationsinfrastrukturen aufzubauen und zu gewährleisten, so wie es mit analogen Forschungs- und Informationsinfrastrukturen ebenfalls zum Teil bereits seit Jahrtausenden gelungen ist – denken wir nur an die Bibliotheken mit ihrer vertrauenswürdigen und zuverlässigen (Grund-) Versorgung von Schriften bzw. Schriftträgern und Informationen.

Wenn Sie an die Anfänge digitaler geisteswissenschaftlicher Forschungsprojekte denken, spielte Nachhaltigkeit zu diesem Zeitpunkt schon eine Rolle?

Heike Neuroth: Ich kann diese Frage natürlich nur vor meinem eigenen Erfahrungshintergrund beantworten. In der Computerlinguistik, die bereits seit Jahrzehnten mit digitalen Methoden und Verfahren arbeitet, stellt sich der Aspekt Nachhaltigkeit sicherlich ganz anders dar.

Also, als wir damals 2004 überlegten, uns im Rahmen der D-Grid-Initiative¹ auch mit einem geisteswissenschaftlichen Projekt zu beteiligen – obwohl die Geisteswissenschaften weder angesprochen noch eingeladen waren –, hatten wir ganz andere Dinge im Kopf. Ich glaube, wir fühlten uns als Pioniere, die sich mutig auf den Weg machen, die digitale, geisteswissenschaftliche Welt zu erobern – und zwar unter den Bedingungen, die wir für wichtig erachtet haben, wie z. B. den Einsatz internationaler Standards, innovativer Technologien und Open-Source-Software(entwicklungen), der Verwendung offener Schnittstellen etc. Wir waren damals sicherlich etwas naiv

¹ <http://www.d-grid.de>

und wähten uns auch in dem Glauben, dass beispielsweise die gesamte technische Basis-Infrastruktur von D-Grid geliefert würde – und zwar natürlich angepasst an unsere Wünsche und Bedürfnisse. Dies mussten wir sehr schnell korrigieren und uns selber einarbeiten in die Grid-Technologien. Im Nachhinein muss ich sagen, dass es eine unglaubliche Stärke des Konsortiums war, dass wir nicht auseinandergebrochen sind: auf der einen Seite die IT-Experten, die sich selber in Technologiebereiche vorwagen mussten, die noch relativ unerforscht waren – und auf der anderen Seite die einzelnen Fachwissenschaftler, die, natürlich sehr stark von ihren eigenen Forschungsfragen geprägt, ihre speziellen Vorstellungen eingebracht hatten. Und zusammen saßen wir dann in den verschiedenen Gremien der Grid-Initiative unter lauter – ich will auch mit meinem eigenen Hintergrund als Naturwissenschaftlerin mal sagen – Hardcore-Wissenschaftlern mit ihren riesigen Datenmengen und „potenten“ Rechenzentrumsleitern, die mit verantwortlich waren für ein Gesamtbudget von insgesamt 120 Mio. Euro über die gesamte Laufzeit der D-Grid-Initiative. Also, nein, Nachhaltigkeit hat damals keine Rolle für uns gespielt, zumindest nicht in den ersten drei Jahren. Da lag der Schwerpunkt eindeutig in der Konzeption und Entwicklung der Virtuellen Forschungsumgebung, dem gemeinsamen Basteln und Ausprobieren, sich insgesamt besser kennenzulernen und verstehen – wir hatten ja auch ganz unterschiedliche Sprachen und Terminologien.

In der zweiten Förderphase (2009–2012), die weiterhin parallel zur D-Grid-Initiative lief, war uns die zunehmende Stabilisierung und dauerhafte Bereitstellung der technischen Infrastruktur und Dienste sowie der vertrauensvolle Aufbau der Community bereits ein großes Anliegen. Wir haben die Angebotspalette von TextGrid erheblich erweitert und uns auf das Forschungsdaten-Repository gestürzt, auch vor dem Hintergrund der nachhaltigen Bereitstellung und intellektuellen Nachnutzung der geisteswissenschaftlichen Forschungsdaten. Damit meine ich, dass der Nachhaltigkeitsaspekt sich hier überwiegend auf die Dienste und Daten bezogen hat und wir uns eher vorsichtig und zögernd den betriebswirtschaftlichen und strukturellen Aspekten der Nachhaltigkeit genähert haben. Es ist aber natürlich schwierig, sich diesen Bereichen – ich will mal sagen, mit Hingabe – zu widmen. Innovative Entwicklung und Forschung zu betreiben und sich gleichzeitig mit der gleichen Energie den zum Teil trocken anmutenden Aspekten der Nachhaltigkeit zu beschäftigen, ist schon eine besondere Herausforderung – zumal, wenn man auf diesem Gebiet ziemlich unerfahren ist. Nichtsdestotrotz haben wir in der zweiten Förderphase zumindest eine Über-

sicht über die Kosten auf der Ausgabenseite gewonnen, ohne allerdings eine Idee zu haben, wie wir diese dann auf der Einnahmeseite ausgleichen könnten. Zum Glück waren wir vor ca. drei Jahren dann aber offenbar in der Förderlandschaft in einer Phase, in der das Thema „Wie finanziert man erfolgreiche digitale Forschungsinfrastrukturen“ sehr virulent geworden ist. Mehrere Drittmittel-Projekte standen damals an dem Scheideweg, ob und wie sie ihre Werkzeuge, Dienste, Daten etc. nach Ablauf der Förderphase weiter aufrechterhalten können. Es gab nur wenige Fachdisziplinen, die dazu ermuntert wurden, einen Nachfolgeantrag einzureichen.

Die dritte und letzte Förderphase (2012–2015) hatte dann ganz konkret das Ziel, gemeinsam mit dem Förderer BMBF zu eruieren, wer sich unter welchen Bedingungen an der anteiligen Finanzierung von TextGrid – und später auch DARIAH-DE – beteiligen kann. Es war relativ schnell klar, dass wir hier über ein hybrides Finanzierungsmodell nachdenken müssen, also eine Mischfinanzierung, die aus mehreren Komponenten besteht. Da wäre zum einen der Bund, der weiterhin signalisiert, dass er sich vorstellen kann, sich an einer Co-Finanzierung zu beteiligen. Allerdings müssten dann auch einige Bundesländer Interesse an dem „Produkt“ und an der anteiligen Finanzierung zeigen. Wir haben daher in einem kleinen Kreis mit einigen vielversprechenden Länder-Ministerien Kontakt aufgenommen und konzentrierte Gespräche geführt. Die Ergebnisse sind insgesamt sehr ermunternd und deuten eine Verstetigungsrichtung an, die in das Modell der hybriden Finanzierung passt. Auch einige Partner-Einrichtungen von TextGrid und DARIAH-DE signalisieren eine grundsätzliche Bereitschaft, sich an der Co-Finanzierung durch die Bereitstellung von dauerhaften Eigenmitteln zu beteiligen. Der Bereich der innovativen (Weiter-) Entwicklung könnte wie bisher auch durch das Einwerben von Drittmitteln abgedeckt werden. Nicht vergessen darf man hier natürlich auch die fast 50 assoziierten Forschungsprojekte, die TextGrid und DARIAH-DE ganz oder teilweise für ihre Vorhaben nutzen und ganz wesentlich dazu beitragen, dass sich die Digitalen Geisteswissenschaften in diesem Bereich weiterentwickeln.

Flankiert wird dies zum Glück durch eine Reihe weiterer Maßnahmen wie z.B. die Gründung des deutschsprachigen Digital-Humanities-Verbundes, der vermehrten Ausschreibung dezidierter Digital-Humanities-Professuren wie ganz aktuell in Köln, Göttingen, Stuttgart, Darmstadt etc. und der vermehrten Etablierung von BA- und MA-Studiengängen. Dies sind natürlich Entwicklungen, die die Nachhaltigkeit wesentlich unterstützen.

Welche Herausforderungen gibt es aus Ihrer Sicht beim Thema Nachhaltigkeit digitaler geisteswissenschaftlichen Forschungsinfrastrukturen?

Heike Neuroth: Dies ist eine schwierige Frage, die ich natürlich nur aus meinem persönlichen Hintergrund heraus beantworten kann. Aber lassen Sie mich ehrlich antworten: Ich bin sehr verwundert, dass das Thema Nachhaltigkeit so ganz ohne politische Unterstützung angegangen werden muss. Ich finde es schon ziemlich viel verlangt, dass sich die FachwissenschaftlerInnen so breit und tief mit dem Thema Nachhaltigkeit auseinandersetzen müssen und sich für das Thema engagieren sollen. Wir dürfen ja nicht vergessen, dass eine akademische Karriere ganz anderen Erfolgskriterien unterliegt als zum Beispiel die eines Wissenschaftsmangers oder einer Direktorin einer großen Forschungs- oder Infrastruktureinrichtung. In der Akademia zählen die Anzahl der Publikationen, die Einwerbung von Drittmitteln, die optimaerweise linear verläuft zum Erkenntnisgewinn, und die Reputation insgesamt, die sich zum Beispiel auch in Form von Einladungen als Keynote-Speaker oder in (inter)nationale Programm-Komitees wichtiger Konferenzen ausdrückt. Ein Engagement im Bereich Wissenschaftspolitik und dann auch noch zum Thema Nachhaltigkeit steht sicherlich nicht ganz oben auf der Prioritätenliste von WissenschaftlerInnen.

Ich persönlich bin daher sehr dankbar, dass wir in den Projekt-Konsortien immer genug Raum und Ideen für derartige Diskussionen hatten. Was mir bisher fehlt, ist ein klares Bekenntnis von Seiten der politisch Verantwortlichen. Es kann einfach nicht sein, dass wir so viel Forschungs- und Entwicklungsgeld in die digitale Transformation investieren, in den Aufbau und die Entwicklung fachspezifischer Informations- und Forschungsinfrastrukturen, um sie dann nach Ablauf der Förderphase still und heimlich in der Versenkung verschwinden zu lassen und bei einem neuen Hype-Thema – wie z. B. gerade *Big Data* – wieder von vorne anfangen zu müssen. Was für die analogen Infrastrukturen selbstverständlich ist – denken wir an die Strom-, Schienen- oder Straßen-Infrastrukturen – muss auch für die digitalen Infrastrukturen selbstverständlich sein. Dabei darf uns in Deutschland der Föderalismus nicht im Wege stehen. Die Forschung wird vernetzter, die Nutzung von Diensten und Daten kooperativer und die Fachdisziplinen arbeiten internationaler. Was nützt es uns da, wenn z. B. das Bundesland Baden-Württemberg eine vorbildliche E-Science-Initiative startet oder die Leibniz-Gemeinschaft eine Vorreiterrolle im Bereich digitaler Infrastrukturen einnimmt?

Die Forschungsverbände mit ihren WissenschaftlerInnen aus mehreren Bundesländern und aus unterschiedlichen Wissenschaftsorganisationen orientieren sich längst an nationalen, wenn nicht gar internationalen Gepflogenheiten. Einen solchen Forschungsverbund aber von der Nutzung einer weltweit zugänglichen digitalen Forschungsinfrastruktur auszuschließen mit der Begründung, das zugehörige Bundesland oder die jeweilige Wissenschaftsorganisation habe sich nicht an den ursächlichen Aufbau- und Entwicklungskosten beteiligt, ist nicht vermittelbar. Ich übertreibe hier bewusst, um den föderalen Irrsinn herauszustreichen und auch, um die Herausforderungen, denen wir in unseren Gesprächen immer wieder begegnen, auf den Punkt zu bringen. Wir verlieren europäisch und international gesehen nicht nur Zeit, sondern vor allem auch die Kompetenz und Erfahrung exzellenter Experten im Bereich digitale Forschungsinfrastruktur, die in einem langfristigen Engagement auf diesem Gebiet kaum Sinn sehen – wird es doch weder in Form klarer Karrierewege mit etablierten Berufsfeldern belohnt, noch ist es zielführend für die akademische Karriere. Mein Wunsch ist, dass die politischen Entscheidungsträger sich der Problematik bewusst werden und schnell und entschieden handeln. Der Ende letzten Jahres gegründete Rat für Informationsinfrastruktur mit Vertretern aus Bund, Land, Wissenschaftsorganisationen, hochrangigen Fachwissenschaftlern etc. setzt hier sicherlich ein gutes Zeichen – aber ob das ausreichen wird? Exzellente Forschung braucht exzellente analoge und digitale Forschungsinfrastrukturen, die dauerhaft und verlässlich bereitstehen und kontinuierlich an den Bedarfen der Fachdisziplinen weiterentwickelt werden – und zwar hier und heute!

Das Interview führten Michelle Rodzis und Sibylle Söring.

Das TextGrid Laboratory: Zehn Jahre Software-Entwicklung

von Mathias Göbel

Die Virtuelle Forschungsumgebung TextGrid funktioniert nach dem Client-Server-Prinzip. Prozesse und Speicherplatz werden zentral organisiert und sind möglichst losgelöst von dem PC oder der Oberfläche, über die man Zugang erhält. In diesem Zusammenhang stellt das “TextGridLab [...] TextGrid’s client application” (TextGrid Konsortium 2009–2010: 6) dar. Es bietet für viele Dienste die grafische Oberfläche: vom Import digitaler Daten über die Er- und Bearbeitung technischer und inhaltlicher Dokumente bis zur Recherche in den Beständen und Publikationen des TextGrid Repository. Es ist die Software-Komponente, die auf den Computern der Nutzer installiert wird, und damit die unmittelbare Schnittstelle der Bearbeiter zur Virtuellen Forschungsumgebung. Es verbindet Content Provider mit der digitalen Infrastruktur und ist den damit arbeitenden wissenschaftlichen Projekten alltäglicher Begleiter.

Daraus ergeben sich Anforderungen, die während der neun Jahre, die die offizielle Förderung durch das BMBF andauerte, immer wieder geprüft wurden. Das bedeutet auch, dass es nun Zeit ist, die Anwendung zu testen und die einst genannten Ziele und Kriterien mit der Ist-Situation – dem TextGrid-Lab in Version 2.2 – abzugleichen.

Lab-Evolution

Die Basis für das TextGridLab bildet die *Eclipse Rich Client Platform* (RCP). Sie liefert ein Grundgerüst für diverse Softwareprodukte, ist anpassbar und für verschiedene Systeme verfügbar. Man kann auf vorhandene Bibliotheken zurückgreifen und verfügbare Module nach Bedarf einbauen. Gleich zu Beginn der Förderung musste die Entscheidung für ein Software-Ökosystem getroffen werden – der im Juli 2006 (Beginn der Förderung war Februar 2006) veröffentlichte Report unter dem Titel „Text Processing“ deutet die Wahl von Eclipse als Entwicklungsumgebung erstmals an:

[ARCHway¹] hat demonstriert, dass das Eclipse SDK nicht nur für die Softwareentwicklung, sondern auch als ergonomische Umgebung für textwissenschaftliche Werkzeuge geeignet ist und aufgrund seines Plug-in Konzeptes für modulare Anwendungen prädestiniert ist. Es steht deshalb zu erwarten, dass sich Eclipse auch für das GUI zu den TextGrid-Modulen bewähren wird. (TextGrid Konsortium 2006: 3)

Neue Module wurden gefertigt, die auf die TextGrid-Architektur abgestimmt sind und viele grundlegende Services bereitstellen. Dazu zählen (unter anderen) grafische Oberflächen für die Navigation durch gespeicherte Daten (der „Navigator“), das Management der Projekte, wozu auch die Vergabe von Nutzerrechten mittels der „Benutzer-Verwaltung“ gehört, oder zum Beispiel die grafischen Oberflächen zur Konfiguration von Publikationsprozessen (tg-publish oder SADE Publish Tool)². Zu diesen Programmen, innerhalb des Labs „Perspektive“ oder „Ansicht“ genannt, kommen Middleware-Komponenten, welche die Kommunikation zwischen Lab und Server-Komponenten (zum Beispiel TextGrid Repository) regeln. Das TextGridLab ist ein Produkt aus dem Zusammenspiel all dieser Komponenten, was dazu führt, dass Änderungen an der Server-Infrastruktur auch Änderungen am Lab nach sich ziehen können. Entsprechend mussten zunächst alle Komponenten in eine stabile Version gebracht werden. Seit dem Sommer 2008 lag intern, im folgenden Herbst dann auch öffentlich zugänglich, ein „TextGridLab Beta“ vor. Mit Abschluss der ersten Projektphase wurden Rückmeldungen der Beta-Tester (der ersten TextGrid-Nutzer) gesammelt (vgl. TextGrid Konsortium 2008: 5) und in eine Version „TextGridLab 1.0“ eingearbeitet, welche ihr Release im Juli 2011 erfuhr und neben Verbesserungen und beseitigten Fehlern auch mit neuen Funktionen wie der integrierten Suche in den Beständen des Trierer Wörterbuch-Netztes aufwartet.

Lab-Evaluation

TextGrid wendet sich explizit an die Gesamtheit der geisteswissenschaftlichen Forscher, die im Einzelnen aber hochspezialisierten Fachgruppen angehören und konkrete Anforderungen stellen. Zudem wird die Berücksichtigung der im Fach etablierten Standards erwartet. Eine Software so generisch

1 ARCHway: Architecture for Research in Computing for Humanities through Research, Teaching, and Learning

2 Vgl. zu SADE den Beitrag von Ubbo Veenster im vorliegenden Band.

zu gestalten, dass sie an ein gemeinsames Repository angeschlossen werden kann und zugleich gemeinsam nutzbare Tools bereitstellt, erfordert an die Fachwissenschaften angeschlossene Entwicklungsprozesse. Diese lassen sich an den Berichtspflichten ablesen, in denen Muss-, Wunsch- und Abgrenzungskriterien für einzelne Komponenten aufgeschlüsselt wurden. Da die Letztgenannten jene Punkte umfassen, die nicht in den Entwicklungsprozess einfließen, bleiben diese im Folgenden unberücksichtigt.

Für das TextGridLab und generell für die RCP finden sich unter den Musskriterien (vgl. TextGrid Konsortium 2007: 9) die schon angesprochene Plattformunabhängigkeit (1), die mit Einsatz von Eclipse erreicht wurde und nur wenige punktuelle Einschränkungen erfährt. So müssen Windows-Nutzer bei den aus dem Marketplace nachzuladenden Tool „SADE Publish Tool“ auf die Drag-and-drop-Funktion im Zusammenspiel mit dem Navigator verzichten und der Text-Text-Link-Editor kann nur nach Installation proprietärer Software von Drittanbietern³ genutzt werden. Weiter wird die Möglichkeit des Online- und Offline-Arbeitens samt einer Synchronisierungsfunktion (2) genannt, was ebenfalls als erreicht gilt. Einst exportierte Daten können mittels der Importfunktion wieder aufgenommen werden. Dabei wird wahlweise die vormals exportierte Revision überschrieben oder eine neue Revision angelegt.

Als erstes Kriterium wird die „[i]ntuitive Bedienung und Konfiguration“ (TextGrid Konsortium 2008: 9) (3) genannt. Hier gilt es, den schmalen Grat zwischen Komplexität und Funktionsvielfalt einerseits und der notwendigen Einfachheit andererseits zu finden. Eclipse selbst bietet viele – manch einem vielleicht zu viele – Funktionen, die Darstellung, Verschachtelung oder Sortierung der Perspektiven zu regeln. Diese Optionen sind unabhängig von den einzelnen Komponenten. Das bringt eine große Flexibilität mit sich, sorgt aber auch dafür, dass Ansichten – je nach Nutzung anderer Werkzeuge im Lab – ihre Position wechseln oder soweit verschwinden, dass diese nur noch als Symbol in kleinen, abgegrenzten Leisten auftauchen. Andererseits kann man auf viele Hilfen bei der Steuerung zurückgreifen. So stehen Tastenkombinationen zur Verfügung, die fast jede Perspektive aufrufen, ohne Menüs durchzugehen. Mit dem „Schnellzugriff“ – via Strg + 3 zu erreichen – steht einem ein sehr hilfreiches Auswahlm Menü zur Verfügung; es bietet zudem ein Eingabefeld, mit dem auch Untermenüs und Funktionen, denen kein Key-shortcut zugewiesen wurde, schnell zu erreichen sind. Das bedeutet auch,

3 Apple QuickTime for Windows, nur mit 32-bit-Versionen nutzbar

dass mit einem bestimmten Nutzungsinteresse oder einer konkreten Aufgabe die Wege im Lab kurz sind – gleich, welcher Nutzergruppe man sich zuordnet oder welche Erfahrung man mitbringt. Die definitionsgemäßen Kriterien zur Usability (zum Beispiel nach ISO 9241-11: “effectiveness, efficiency, and satisfaction”); Burghardt/Manuel 2014: 70) werden in einem eigenen Report⁴ verhandelt.⁵ Ein weiteres Musskriterium ist mit „Einhaltung von Konventionen und Schnittstellen der zugrundeliegenden Plattform (Eclipse), um eine Integration mit von dritter Seite für diese Plattform entwickelten Komponenten / Plug-ins zu erleichtern“ (TextGrid Konsortium 2007: 9) (4) benannt. Beispielhaft sei die Nutzung gemeinsamer Perspektiven genannt und explizit die Nutzung einer Perspektive innerhalb des Labs, in der die Status- oder Fehlermeldungen aller laufenden Komponenten zusammengetragen werden.

Unter den Wunschkriterien findet sich der Punkt „Fortschrittsanzeige bei der Verarbeitung längerer Jobs“ (ebd.), was gleichsam übererfüllt wurde. So lässt sich nicht nur der Status abrufen, anstehende Aufgaben können auch im Hintergrund weiterlaufen, was dazu führt, dass in Wartezeiten andere Aufgaben im Lab bearbeitet werden können. Weiterhin wird die Integration aller Tools in das Lab als ein Ziel (vgl. ebd.: 10) formuliert.

Die einzige Anforderung, die diese RCP an einen Computer stellt, bildet ein installiertes Java Runtime Environment (JRE). Will man große Datenmengen verarbeiten und zum Beispiel mithilfe des Text-Bild-Link-Editors mehrere Bilder mit einem XML-Dokument verknüpfen, so benötigt man ausreichend Arbeitsspeicher. Dessen Akquise wird von einer Konfigurationsdatei geregelt, die im lokalen Verzeichnis liegt. Hier ist ein geringer Wert festgelegt, der das Ausführen des Labs auf weniger gut ausgestatteten Maschinen ermöglicht. Von den Nutzern sollte dieser nicht angepasst werden. Dennoch ist oft schon nach Öffnen weniger Bilder der Maximalwert der Nutzung erreicht. Diese externen Variablen sind kaum kontrollierbar und die Entscheidung für eine auf allen Systemen lauffähige Einstellung, wie sie derzeit besteht, ist zu bevorzugen.

Auch für einzelne Tools wurden derartige Kriterien formuliert. Hier sei der XML-Editor exemplarisch herausgegriffen, da diesem Werkzeug eine sehr große Bedeutung zukommt. Er ist das zentrale Be- und Erarbeitungstool. Die Richtlinien der Text Encoding Initiative stellen einen der Standards dar,

4 TextGrid Konsortium (2013): Report 5.3.1

5 Vgl. dazu auch den Beitrag von Simone Rieger und Ben Kaden im vorliegenden Band.

die von den Textwissenschaftlern eingefordert wurden. „Die wichtigste Funktion des XML-Editors ist es, Benutzer, die keine PC-Fachleute sind, in ihrer Arbeit mit XML-Texten möglichst weitgehend zu unterstützen“ (ebd.: 26). Die Umsetzung eines „Visual Editors for XML“⁶, der sich bereits durch die Eclipse-Community in Entwicklung befand, hat sich in diesem Punkt erfüllt. Durch den What-You-See-Is-What-You-Mean-Editor wird das Arbeiten mit XML von der Quellcode-Ansicht in eine Ansicht überführt, die ohne spitze Klammern auskommt und weitere Darstellungsoptionen zulässt, die im Nutzerhandbuch beschrieben stehen.⁷

Der XML-Editor muss XML-Dokumente entsprechend einem in einer der gängigen XML-Schemasprachen (DTD, W3C XML Schema, Relax NG) formulierten Schema bearbeiten und validieren können. Der Benutzer soll bei seiner Arbeit möglichst weitgehend unterstützt werden, indem ihm angezeigt wird, welche XML-Elemente an einer bestimmten Position erlaubt sind und außerdem, welche Elemente er bereits häufiger verwendet hat. Hinzu kommt die Möglichkeit zum Auszeichnen vorhandenen Texts per Maus. (TextGrid Konsortium 2007: 26)

All diese Punkte sind mit dem derzeitigen Entwicklungsstand abgedeckt und in vollem Umfang realisiert. Die unterstützte Schemasprache ist schließlich W3C XML Schema, was per Konversion aus anderen Typen generiert werden kann (vgl. TextGrid Konsortium 2010: 18). Da Relax-NG das von der Text Encoding Initiative bevorzugte Format darstellt, ist man auf die Konversionsroutinen angewiesen, was unter bestimmten Umständen⁸ zu Fehlern in der Validierung von xml:id-Attributen führt. Mit oXygen steht ein kommerzieller Editor zur Verfügung, der nach Bedarf über den Marketplace installiert wird und in der Lage ist, Relax NG vollständig zu unterstützen.

Der TextGrid-XML-Editor ist darüber hinaus zur Erstellung weiterer technischer Dokumente geeignet. Das Schema selbst kann mit diesem Programm entwickelt werden; auch Transformationsroutinen mit XSLT lassen sich generieren und im Lab nutzen (vgl. Radecke et al. 2013: 101). Damit findet eine Anforderung der informationswissenschaftlichen Community ihre Umsetzung im Lab. Ein konkretes Schema gehört zu jeder XML-Datei, insbesondere, wenn diese archiviert werden soll, denn es liefert zur jeweiligen

6 <https://eclipse.org/vex/>

7 Vgl. <https://dev2.dariah.eu/wiki/display/TextGrid/Features+of+the+WYSIWYM+View.>

8 Vgl. [http://tei-1.970651.n3.nabble.com/ODD-problem-xml-id-missing-or-wrong-usage-in-XSD-tp4024120.html.](http://tei-1.970651.n3.nabble.com/ODD-problem-xml-id-missing-or-wrong-usage-in-XSD-tp4024120.html)

XML-Auszeichnung die zugehörige Definition, optional auch weitere Beschreibungen und Beispiele, und sorgt damit für die spätere Interpretierbarkeit der Daten.

Lab-Feedback

Rückmeldungen aus der Fachcommunity werden aktiv eingeholt: bei Vorträgen, Workshops, Schulungen und im persönlichen Gespräch. Bereits in den konzeptionellen Phasen spielte die Community eine große Rolle, was die 23-malige Nennung (Z-Score = 1.2)⁹ des Begriffs „Community“ im Antrags-text für die erste Förderphase unterstreicht. Das Feedback wurde dann in Folgeanträge eingebettet und teils in den laufenden Entwicklungsprozess eingebunden. Schließlich sind es auch die einzelnen Projekte, die mit TextGrid arbeiten und jederzeit über verschiedene Kanäle Rückmeldungen liefern können. Sie können direkt in ein Bugtrackersystem eintragen, wodurch die Entwickler unmittelbar über aktuelle Probleme, Schwierigkeiten und Wünsche informiert werden. Bugtrackersysteme sind oftmals mit der Softwareentwicklung verbunden, sodass bestimmte Bearbeitungen und Status einsehbar sind, aber auch Antworten gegeben und Rückfragen gestellt werden können. Diese Ticketsysteme knüpfen an den technischen Details der gesamten Infrastruktur an. Nutzern werden hier Auswahlmenüs für die von einem bestimmten Fehler betroffene Komponente oder Zielversionen präsentiert, wobei es sich um Fragestellungen handelt, die von Anwendern nicht beantwortet werden müssen. Eine Moderation hat sich an dieser Stelle als geeignetes Mittel erwiesen. Diese wird von einem Support-Team übernommen, das auf entsprechende Anfragen, die per E-Mail, Foren oder auf anderen Wegen eingehen, antwortet und diese in das Bugtrackersystem übersetzt. Diese Person ist idealerweise in den einzelnen Projekten selbst angesiedelt, sodass Anpassungen und projektspezifische Workflows und Workarounds bei derartigen Meldungen bereits berücksichtigt werden und eine direkte Kommunikation mit den Entwicklern stattfinden kann, wovon Infrastruktur als auch die Projekte selbst profitieren.¹⁰

9 Vgl. <http://goo.gl/cP7qHZ>.

10 Vgl. Radecke et al. (2013) sowie den Beitrag von Sibylle Söring und Mirjam Blümm im vorliegenden Band.

Bezieht man die für eine Antragstellung notwendigen Vorüberlegungen mit ein, so blickt TextGrid nun auf zehn Jahre Entwicklungszeit zurück. Wo vor zehn Jahren die Eclipse Rich Client Platform stand, finden sich heute mehr und mehr die Webbrowser. Die Technologien, die Webapplikationen ermöglichen, standen damals erst am Anfang. Das heute weit verbreitete JavaScript-Framework JQuery erfuhr im August 2006 die erste Veröffentlichung (vgl. Wikipedia o.J.) – die Quelle, die dieses Datum nennt, kannte am 1. Januar 2006 gerade 895.000 Artikel. Heute entnimmt man diese Information einem von fast 5 Millionen Einträgen.

Die erreichten Anforderungen zeigen zudem, dass das TextGridLab als wichtigste Schnittstelle zur Infrastruktur stabil funktioniert – ein Verdienst der konzeptionellen und technischen Entwickler, die gemeinsam mit Fachwissenschaftlern Anforderungen definiert und deren Umsetzung konsequent auf diese ausgerichtet haben. Das Ergebnis ist eine stabile Infrastruktur im produktiven Betrieb, die mit dem TextGridLab komplexe digitale Verfahren und Methoden ermöglicht – unabhängig von den hier gezeigten Kritikpunkten. Transparente Entwicklung und die Open-Source-Lizenzierung lassen Folgeprojekten alle Optionen offen, einzelne Komponenten herauszugreifen, anzupassen, weiterzuentwickeln oder zu verbessern.

Literaturverzeichnis

- Al-Hajj, Yahya Ahmed Ali; Küster, Marc Wilhelm (2013): The text-image-link-editor: A tool for linking facsimiles and transcriptions, and image annotations. In: *Literary and Linguist Computing* 28 (2): 190–198. <http://llc.oxfordjournals.org/content/28/2/190.full.pdf+html>.
- Burghardt, Manuel (2014): *Engineering annotation usability – Toward usability patterns for linguistic annotation tools*. Dissertation, Universität Regensburg. <http://epub.uni-regensburg.de/30768/>.
- Burnard, Lou; Sperberg-McQueen, C. M. (2012): TEI Lite: Encoding for Interchange: an introduction to the TEI. http://www.tei-c.org/release/doc/tei-p5-exemplars/html/tei_lite.doc.html.
- Funk, Stefan E.; Veentjer, Ubbo; Vitt, Thorsten (2013): Digitale Werkzeuge in den digitalen Geisteswissenschaften: Die Virtuelle Forschungsumgebung TextGrid – Status quo und neue Entwicklungen. In: Neuroth, Heike et al. (Hrsg.): *Evolution der Informationsinfrastruktur: Kooperation zwischen Bibliothek und Wissenschaft*. Glückstadt: Verlag Werner Hülsbusch, S. 277–300.

- Kiernan, Kevin; Jaromczyk, Jerzy W.; Dekhtyar, Alex; Porter, Dorothy Carr; Hawley, Kenneth; Bodapati, Sandeep; Iacob, Ionut Emilnd (2005): The ARCHway Project: Architecture for Research in Computing for Humanities through Research, Teaching, and Learning. In: *Literary & Linguistic Computing* 20: 69–88. <http://llc.oxfordjournals.org/content/20/supplement/69.full.pdf+html>.
- Krasnay, John (lead developer); Andersson, Marcus; Austin, John; Berkovits, Adrian; Gautier, Martin; Holroyd, David; Thienel, Florian (lead); Hiesserich, Carsten; Voormann, Holger; Carver, David; Campista, Igor Jacy Lino (2008 ff.): Vex – A Visual Editor for XML. <https://eclipse.org/vex/>.
- Radecke, Gabriele; de la Iglesia, Martin; Göbel, Mathias (2015): Vom Nutzen digitaler Editionen. In: Bein, Thomas (Hrsg.): *Vom Nutzen der Editionen*. Berlin u.a.: de Gruyter. Im Druck.
- Radecke, Gabriele; Göbel, Mathias; Söring, Sibylle (2013): Theodor Fontanes Notizbücher: Genetisch-kritische und kommentierte Hybrid-Edition, erstellt mit der Virtuellen Forschungsumgebung TextGrid. In: Neuroth, Heike et al. (Hrsg.): *Evolution der Informationsinfrastruktur: Kooperation zwischen Bibliothek und Wissenschaft*. Glückstadt: Verlag Werner Hülsbusch, S. 85–106.
- TEI Consortium (Hrsg.) (2014): TEI P5: Guidelines for Electronic Text Encoding and Interchange. Version 2.7.0, 16. Sept. 2014. <http://www.tei-c.org/Guidelines/P5/>.
- TextGrid Konsortium (Hrsg.) (2006): Text Processing. <http://textgrid.de/fileadmin/berichte-1/report-1-1.pdf>.
- TextGrid Konsortium (Hrsg.) (2007): TextGrid Tools I. <http://textgrid.de/fileadmin/berichte-1/report-2-1.pdf>.
- TextGrid Konsortium (Hrsg.) (2008): TextGrid Tools II. <http://textgrid.de/fileadmin/berichte-1/report-2-2.pdf>.
- TextGrid Konsortium (Hrsg.) (2009–2010): R 2.3: User’s Manual TextGrid-Tools. <http://textgrid.de/fileadmin/berichte-1/report-2-3.pdf>.
- TextGrid Konsortium (Hrsg.) (2009–2011): Abschlussbericht. <http://textgrid.de/fileadmin/berichte-1/abschlussbericht-1.pdf>.
- TextGrid Konsortium (Hrsg.) (2010): AP6: TextGrid-Tools I (Pflichtenhefte) (R 6.0.1). <http://www.textgrid.de/fileadmin/berichte-2/report-6-0-1.pdf>.
- TextGrid Konsortium (Hrsg.) (2012): User Manual 2.0. <https://textgrid.de/fileadmin/dokumentation/user-manual-2.pdf>.
- TextGrid Konsortium (Hrsg.) (2014): SADE – Dokumentation und Referenzinstanz. http://textgrid.de/fileadmin/TextGrid/reports/TG-R-4_4_2-SADE_Dokumentation_und_Referenzinstanz.pdf.
- Wikipedia (o. J.): Stichwort „jQuery“. <http://en.wikipedia.org/w/index.php?title=jQuery&oldid=642872897> [Abruf am 28.1.2015].

Grenzgänge

An der Schnittstelle von Geisteswissenschaft, Informatik und Informationswissenschaft

von Sibylle Söring und Mirjam Blümm

Der Punkt, an dem zwei Themengebiete, zwei Disziplinen, zwei Kulturen – zwei Galaxien, könnte man auch sagen – zusammenstoßen, sollte kreative Gelegenheiten erzeugen. In der Geschichte der geistigen Tätigkeit war dies immer der Ort, an dem es zu einem der Durchbrüche kam. Nun gibt es solche Gelegenheiten.
C. P. Snow, Die zwei Kulturen¹

Interdisziplinarität, Vielfalt der Fächer, ihrer Arbeitskulturen, ihrer Terminologien, ihrer Forschungsfragen, -verfahren und -methoden – kaum ein Tag in zehn Jahren TextGrid, der vor dem Hintergrund solcher „Grenzgänge“ nicht auf die eine oder andere Weise vor Herausforderungen gestellt hätte. Viel hatte man sich vorgenommen mit dem Vorhaben, GeisteswissenschaftlerInnen, InformationswissenschaftlerInnen und InformatikerInnen über mehrere Jahre, über verschiedene Disziplinen und verteilte Orte hinweg an einen Tisch zu setzen, um eine digitale Forschungsumgebung aufzubauen, die Geistes- (und hier v. a. Editions-) wissenschaftlerInnen die kollaborative Erschließung textbasierter Quellen und die sichere, langfristige Speicherung ihrer Forschungsdaten und -ergebnisse ermöglichen sollte. Anforderungen mussten erhoben, Aufwände abgeschätzt, Mittel eingeworben, Personal gefunden werden, das nicht nur über die bis dato kaum an bundesdeutschen Hochschulen gelehrt Digital-Humanities-Kenntnisse und -Kompetenzen verfügte, sondern das auch bereit war, sich – ohne die Sicherheit und das Renommee, geschweige denn die Methodenlehre eines etablierten, traditionsreichen Faches – in immer wieder neue Disziplinen und Methoden einzuarbeiten. Es mussten Zwischenergebnisse getestet und evaluiert, Anpassungen und Erweiterungen vorgenommen und schließlich das Angebot an eine fachlich wie hinsichtlich ihrer technischen Vorkenntnisse und last, but not least,

¹ Snow (1959)

bezüglich ihrer Aufgeschlossenheit neuen Technologien gegenüber überaus heterogene Zielgruppe kommuniziert werden.

Dass bei einem solchen Vorhaben Herausforderungen, Stolpersteine und Fallstricke gleichsam systemimmanent sind – mehr noch vielleicht als bei „analogen“ Forschungsvorhaben –, liegt auf der Hand. Sie sind dem Umstand geschuldet, dass hier jeweils unterschiedliche Arbeitsverfahren und -methoden, oft auch Arbeitstechniken ebenso wie unterschiedliche Forschungskulturen, Terminologien und sogar (Fach-) Sprachen immer wieder neu ins Einvernehmen zu setzen sind über das, was relevant und richtig ist. Was C. P. Snow in den 60er-Jahren harsch den „zwei Kulturen“, der Geisteswissenschaft einerseits und der Naturwissenschaft andererseits, attestierte, nämlich eine Unfähigkeit, miteinander zu sprechen,² das gilt mitunter auch für die beiden zunächst einander so fremden Disziplinen Geisteswissenschaften und Informatik. Dabei hatte alles so gut angefangen: Bereits in den 60er-Jahren waren es LinguistInnen, die die Vorteile computerunterstützter Verfahren frühzeitig erkannten, beförderten und unter dem Label der Computerlinguistik der Sprachwissenschaft pragmatisch eine neue Disziplin und Methodik beordneten, die sich in weitaus geringerem Maße Fragen der Selbstlegitimation als eigenständiges Fach ausgesetzt sah, als es heute, ein halbes Jahrhundert später, die Digital Humanities tun. Der Brückenschlag schien gelungen.

Zu den Kernanwendungen der Digital Humanities zählen heute digitale Editionen. Längst sind diese nicht mehr zuvorderst text-, sondern in hohem Maße auch bildbasiert. Die Digitalisierung des kulturellen Erbes hat dieser Entwicklung, zusammen mit der zunehmenden Verfügbarkeit kostengünstiger Speicherlösungen, großen Vorschub geleistet. Ohne die synoptische Ansicht von Faksimile und Transkription sind digitale Editionen kaum mehr vorstellbar; zunehmend werden zudem digitale Werkzeuge nachgefragt – und angeboten –, die das kollaborative Annotieren von Bild und Text ermöglichen, die bei der Verlinkung von Text und Bild unterstützen, die komfortables Edieren ermöglichen, die bei der Erschließung nach TEI unterstützen und die schließlich Technologien wie *Linked Open Data* erlauben. Dass hier gleichermaßen hoher fachwissenschaftlicher, informationswissenschaftlicher

² Zwar seien, so Snow, die Gelegenheiten zu solchen „kreativen Durchbrüchen“ durchaus gegeben; sie existierten jedoch „[...] sozusagen in einem Vakuum, denn die Angehörigen der zwei Kulturen können nicht miteinander sprechen“ (Snow 1959).

und informatischer Input gefragt ist, Know-how und Expertise, ist offensichtlich. TextGrid – von Beginn an mit der Prämisse gestartet, aus den Anforderungen von FachwissenschaftlerInnen heraus zu entwickeln und gemeinsam neue Werkzeuge und Dienste zu gestalten – lässt sich dabei im je unterschiedlichen Wortsinne als Schnittstelle verstehen. In der täglichen Zusammenarbeit im Projektverbund, in der internen Kommunikation ebenso wie in der Außendarstellung und Öffentlichkeitsarbeit, aber auch in der Kooperation mit Forschungsvorhaben muss dies immer wieder neu gelingen.

Partnerprojekte: *Communication matters*

Seit der Veröffentlichung des TextGridLab in der Produktivversion 1.0 im Sommer 2011 haben zahlreiche geisteswissenschaftliche Forschungsvorhaben bei Förderinstitutionen wie BMBF, DFG und Akademienunion Mittel für die je projektspezifische Anpassung und Erweiterung von TextGrid im Rahmen ihrer Arbeiten beantragt. Spätestens mit der ersten eHumanities-Förderlinie des BMBF 2011³ und der Veröffentlichung eines Kriterienkatalogs für die Begutachtung wissenschaftlicher Editionen durch die DFG⁴ lagen für die Förderung geisteswissenschaftlicher Forschungsprojekte Richtlinien vor, die dezidiert die Anwendung von und Umsetzung mit digitalen Informationstechnologien und -infrastrukturen wie TextGrid und DARIAH-DE⁵ (BMBF) sowie die Veröffentlichung z. B. einer wissenschaftlichen Edition nicht nur in Buchform, sondern auch als digitale Edition nach internationalen Auszeichnungsstandards wie XML/TEI, umfassten (DFG). Ausdrücklich ging es dabei um die Förderung solcher interdisziplinärer Ansätze, „in denen geistes[...]wissenschaftliche Fächer in Kooperation mit informatiknahen Fächern neue Forschungsansätze in ihren Fachdisziplinen entwickeln“⁶.

Im Zuge solcher und anderer Förderlinien ist bis heute eine Reihe von Forschungsvorhaben bewilligt, beantragt und/oder für die Begutachtung

3 <http://www.bmbf.de/foerderungen/21126.php>

4 http://www.dfg.de/foerderung/grundlagen_rahmenbedingungen/informationen_fachwissenschaften/geistes_sozialwissenschaften/kriterien_begutachtung_wissenschaftlicher_editionen/index.html

5 <https://de.dariah.eu/>

6 http://www.dfg.de/foerderung/grundlagen_rahmenbedingungen/informationen_fachwissenschaften/geistes_sozialwissenschaften/kriterien_begutachtung_wissenschaftlicher_editionen/index.html

vorbereitet worden, die in Kooperation mit Partnerinstitutionen aus dem TextGrid-Forschungsverbund – wie der TU Darmstadt, der SUB Göttingen oder dem Steinheim-Institut Essen – durchgeführt werden – zumeist in Form von Stellenanteilen, die die projektspezifische Einrichtung, Anpassung und Erweiterung der Virtuellen Forschungsumgebung, die Datenmodellierung bzw. Entwicklung eines (XML/TEI-) Metadatenschemas, die (durchsuchbare) Darstellung, Visualisierung und Publikation der Forschungsdaten in einem Web-Portal und schließlich ihre Langzeitarchivierung im TextGrid Repository zum Gegenstand haben. Spezifische, auf die Anforderungen des jeweiligen Projekts zugeschnittene Schulungen in der Arbeit mit dem TextGridLab sowie in der Datenauszeichnung (XML/TEI) sind ebenfalls Bestandteil solcher Aufgabenprofile, die in der Regel bereits möglichst frühzeitig im Rahmen einer Antragstellung gemeinsam mit FachwissenschaftlerInnen, Informationswissenschaftlerinnen und InformatikerInnen erarbeitet werden.

Bereits in diesem ersten Stadium der Arbeits- und Aufgabenplanung gilt es, quer durch die Disziplinen, eine gemeinsame Sprache zu finden, Anforderungen möglichst genau zu identifizieren und die mit der Umsetzung verbundenen Aufwände zeitlich, fachlich und finanziell zu quantifizieren. Und der Kommunikationsbedarf bleibt hoch – gerade dann, wenn, wie in aller Regel der Fall, ein Projekt ortsverteilt aufgestellt ist, zwei oder mehr Institutionen bundesweit oder gar international beteiligt sind. „Keine Telefonkonferenz“, so Franziska Diehr, Informationswissenschaftlerin im Projekt *Textdatenbank und Wörterbuch des Klassischen Maya*⁷, „ersetzt ein Face-to-face-Meeting“. Die MitarbeiterInnen der SUB Göttingen treffen sich mit den KollegInnen der Universität Bonn in monatlichen Abständen – zusätzlich zu wöchentlichen Telefonkonferenzen, einem intensiv genutzten Wiki und anderen arbeitsunterstützenden Tools wie einem gemeinsamen Online-Kalender, einem Issue Tracker und Etherpad. Im von der Akademie der Wissenschaften und der Künste über eine Laufzeit von 15 Jahren geförderten Projekt arbeiten AltamerikanistInnen, ComputerlinguistInnen, InformationswissenschaftlerInnen und Informatiker. Der Abstimmungsbedarf ist hoch, aber die Neugierde für die Einarbeitung in jeweils fachfremde Stoffe, Fragestellungen und Methoden ist es ebenso. „Als Informatiker sitzt man immer an einer Schnittstelle“, erläutert Maximilian Brodhun, Software-Architekt im gleichen Projekt.

7 Vgl. hierzu den Beitrag von Christian Prager im vorliegenden Band.

Zum Tagesgeschäft gehört es dabei, die jeweils erforderlichen Anforderungen und Aufwände auch für die jeweils andere Fachdisziplin verständlich zu kommunizieren. Dass eine fachwissenschaftlich vergleichsweise marginal erscheinende Anforderung wie etwa eine bestimmte Datenbank-Funktion einen hohen technischen Aufwand bedeuten kann, und vice versa – auch das ist Gegenstand einer solchen interdisziplinären Zusammenarbeit.

Im Rahmen digitaler Editionsprojekte – einer der Kernanwendungen der Digital Humanities – sind informationswissenschaftliche Kompetenzen wie die texttechnologische Erschließung via XML/TEI oder die nachhaltige Strukturierung digitaler Daten und Objekte oft erst noch aufzubauen bzw. zu erwerben. Auch Vorbehalte gegenüber der „Technifizierung“ der geisteswissenschaftlichen Forschungsarbeit und/oder ihrer Veröffentlichung im digitalen Medium – vielleicht gar schon in einem frühen, noch unfertigen Stadium – spielen gerade in den ersten Monaten eines noch jungen Kooperationsprojekts mitunter eine nicht zu unterschätzende Rolle. Hier gilt es, „die gewohnte geisteswissenschaftliche Praxis der Projektpartner mit informationswissenschaftlichen Fragestellungen anzureichern“, so Uwe Sikora, Metadatenpezialist an der SUB Göttingen im mit TextGrid umgesetzten Projekt *Bibliothek der Neologie*⁸ der Universität Münster. Regelmäßiger, ausführlicher Austausch, Aufgeschlossenheit und Neugierde – das sind letztlich Grundbedingungen wissenschaftlicher Arbeit, ob tradiert oder experimentell, analog oder digital, geistes- oder informationswissenschaftlich.

Den interdisziplinären Projektverbund managen: Interne Projektkommunikation und -organisation

Das TextGrid-Konsortium setzte sich in den letzten zehn Jahren aus 15 unterschiedlichen institutionellen Partnern zusammen – fünf davon waren über die gesamte Laufzeit dabei, vier über zwei Projektphasen und sechs nur eine. Beteiligt waren sechs Universitäten, drei Forschungseinrichtungen, zwei kommerzielle Unternehmen, eine Fachhochschule, ein Rechenzentrum, eine Akademie der Wissenschaften und eine wissenschaftliche Bibliothek.⁹

8 Vgl. hierzu den Beitrag von Bastian Lemitz im vorliegenden Band.

9 Für eine Übersicht der Partner in TextGrid vgl. den Anhang im vorliegenden Band sowie <http://www.textgrid.de/ueber-textgrid/forschungsverbund/>.

Hinter diesen verschiedenen Institutionen standen und stehen natürlich die MitarbeiterInnen mit ihren unterschiedlichen fachlichen Hintergründen und Kulturen: FachwissenschaftlerInnen mit mehr oder minder ausgeprägter Digital-Humanities-Kompetenz, InformatikerInnen mit und ohne geisteswissenschaftliche Vorerfahrung sowie InformationswissenschaftlerInnen mit einer dritten Sicht auf die Dinge.

Gerade in der Anfangsphase von TextGrid stellten die teilweise gegensätzlichen Erwartungshaltungen und vor allem Kommunikationsgewohnheiten eine große Herausforderung für die Zusammenarbeit im Konsortium dar. Vermeintlich gleiche Themenfelder können in den verschiedenen Disziplinen unterschiedliche Bedeutungen haben und Entscheidungen sich entsprechend schwierig gestalten. Hier galt es in besonderem Maße, viel zu kommunizieren, um eine gemeinsame Sprache zu finden.

Eine weitere Herausforderung angesichts des vielfältig zusammengesetzten Konsortiums war es, jeder bzw. jedem eine Stimme darin zu verleihen. Die Erfahrung in TextGrid hat gezeigt, dass das Zusammenspiel umso besser funktionierte, je genauer Entscheidungsstrukturen definiert wurden, da auf diese Weise sichergestellt werden konnte, Entscheidungen transparent und nachvollziehbar für alle Beteiligten zu gestalten.

Dafür wurde ein zentrales Entscheidungsgremium, die Steering Group, etabliert, in dem alle Projektpartner sowie die Konsortialleitung gleichberechtigt vertreten sind und das sich regelmäßig austauscht (Telefon- und Videokonferenzen; Face-to-Face-Treffen), um gemeinsam das weitere Vorgehen zu planen, etwaige Probleme zu besprechen und Workflows festzulegen. Entscheidungen und Abstimmungsprozesse werden seither in einem internen Wiki ausführlich dokumentiert, sodass alle MitarbeiterInnen darauf Zugriff haben.

Die Projektarbeit organisieren die Partner entsprechend ihrer jeweiligen Aufgaben- und Arbeitspakete weitestgehend eigenverantwortlich. Hierbei haben sich flache Hierarchien, eine enge Abstimmungs- und Entscheidungskultur sowie eine transparente, komfortabel up to date zu haltende Dokumentation (Wiki) bewährt. Auch softwareentwicklungsspezifische Managementinstrumente wie Programmiersprints und Bug-Tracking-Systeme kommen zum Einsatz. Auf diese Weise kann ein durch Mitarbeiterfluktuation bedingter Wissensverlust zumindest teilweise abgemildert werden.

Die Konsortialleitung behält den Gesamtprojektfortschritt im Auge und informiert alle MitarbeiterInnen in monatlichen, über interne Mailinglisten verschickte Berichte über den aktuellen Stand des Projekts und anstehende

Arbeiten. Bewährt haben sich zudem halbjährlich stattfindende Konsortialtreffen, bei denen Status und Planung in großer Runde präsentiert und Entscheidungen gemeinsam getroffen werden können.

Neue Wege bei Außendarstellung und Öffentlichkeitsarbeit

Interdisziplinarität, Vielfalt der Fächer, ihrer Arbeitskulturen, ihrer Forschungsfragen und -methoden – das galt bei TextGrid nicht nur für den Forschungsverbund selbst, seine interne Kommunikation und Zusammenarbeit, sondern in noch höherem Maße auch für die Zielgruppe von TextGrid – seine Nutzerinnen und Nutzer, die geisteswissenschaftlichen Forschungsprojekte verschiedenster Provenienz und Zielsetzung ebenso wie die InformatikerInnen und InformationswissenschaftlerInnen, die TextGrid nutzen und weiterentwickeln. Um diese Community anzusprechen – mehr noch: um einen vitalen Austausch zwischen „Machern“ und „Nutzern“ zu etablieren, der die Gestaltung des TextGrid-Angebots auf der Grundlage von Anforderungen und Wünschen der Community selbst ermöglicht – ist TextGrid in der Außendarstellung des Projekts und seines Angebots ebenso wie in der Öffentlichkeitsarbeit Wege gegangen, die z.T. bewusst über das in diesem Rahmen Übliche – schriftliche Dokumentationen, Flyer, Webpräsentationen – hinausgehen. So lag es nahe, ein digitales Angebot auch digital zu vermitteln: Verschiedene interaktive digitale Plattformen – Foren, Blogs –, aber auch audiovisuelle Formate wie Image-Filme und sogenannte „Erklärclips“, dienen dazu, komplexe Themenfelder rund um TextGrid und die Digital Humanities im weitesten Sinne einfach, verständlich und anregend in eine vielfältige Community zu vermitteln.

Von Anfang an, spätestens aber seit der Veröffentlichung der ersten TextGrid-Produktivversion im Sommer 2011, betrieb TextGrid durch zahlreiche Maßnahmen Öffentlichkeitsarbeit, angefangen bei Logo und Eulen-Maskottchen, Flyern, Pressemitteilungen und einer Website¹⁰ über Vorträge und Poster bei nationalen und internationalen Fachveranstaltungen bis hin zu Merchandising-Materialien wie Tassen, Lanyards und USB-Sticks.¹¹

Mit dem Ende 2011 in Kooperation mit DARIAH-DE und dem Max Planck-Institut für Wissenschaftsgeschichte Berlin eingerichteten Blog

10 <http://www.textgrid.de/>

11 Vgl. R 8.1: PR-Konzept, <http://www.textgrid.de/fileadmin/berichte-2/report-8-1.pdf>.

*DHd – Digital Humanities im deutschsprachigen Raum*¹² wurden neue Wege beschritten. Der Blog sollte weniger zur Außendarstellung des Projekts dienen, sondern vielmehr als Plattform für alle Digital-Humanities-Themen für und vor allem aus der Community.



Abb. 1 Slide DHd-Blog¹³

Über 90 AutorInnen aus rund 60 Institutionen und Städten in Deutschland, Österreich und der Schweiz haben in den vergangenen drei Jahren mit insgesamt 532 Beiträgen und zahlreichen Kommentaren für einen regen Austausch gesorgt – das ist im Schnitt alle zwei Tage ein neuer Artikel.¹⁴

Besonders häufig genutzt wird der Blog für Veranstaltungshinweise und die Verbreitung von Stellenanzeigen; zunehmend gehen die Beiträge über den reinen Informationscharakter hinaus und eröffnen inhaltliche Diskussionen zu aktuellen Forschungsthemen und -entwicklungen, Veranstaltungsberichten und Rezensionen.

Ein weiterer multimedialer Schritt in Sachen Öffentlichkeitsarbeit war im März 2014 die Einrichtung eines eigenen DHd-Kanals auf YouTube,¹⁵ wieder in Zusammenarbeit mit DARIAH-DE, der zunächst eigenes Filmmaterial

12 <http://dhd-blog.org/>

13 Vgl. Slideshow <https://de.dariah.eu/>.

14 Stand: Januar 2015

15 <https://www.youtube.com/user/dhdkanal>

wie z. B. Mittschnitte vom von TextGrid initiierten Digital-Humanities-Festakt 2011 in Göttingen¹⁶ oder die TextGrid-Videotutorials¹⁷ zeigte.



Abb. 2 DHd-Kanal auf YouTube

Inzwischen stehen gut vier Stunden Filmmaterial bereit, der Kanal zählt 84 Abonnenten¹⁸ und ist bereits über 6.000 Mal aufgerufen worden. Mit der Playlist Digital Humanities Videoclips¹⁹ ist eine Sammlung von Videos zu DH-Themen anderer Institutionen hinzugekommen, die kontinuierlich erweitert wird. Der Schwerpunkt liegt allerdings nach wie vor auf eigenem Material.

Um das komplexe Thema „Digitales Edieren in einer Virtuellen Forschungsumgebung“ leicht verständlich zu erklären, hat TextGrid in Zusammenarbeit mit der explainity GmbH²⁰, einem mehrfach ausgezeichneten Unternehmen, im Herbst 2014 einen komplexitäts-reduzierten geisteswissenschaftlichen Kurzfilm in Schwarzweiß-Legetechnik produziert.

Der Clip zeigt am Beispiel einer Briefedition, wie Virtuelle Forschungsumgebungen WissenschaftlerInnen bei der computergestützten Erstellung digitaler Editionen unterstützen. Damit knüpfte TextGrid an die im Frühjahr 2014 von DARIAH-DE veröffentlichten explainity-Filme zu den Themen „Kollaborativ arbeiten“²¹ und „Datensicherheit“²² an.

16 http://www.youtube.com/playlist?list=PL_aqAE00s6w0a8EZES_-WQFNSG_Y6IEU_

17 http://www.youtube.com/playlist?list=PL_aqAE00s6w10Y9zklSv7H9VvYZByych5-

18 Stand: Januar 2015

19 http://www.youtube.com/playlist?list=PL_aqAE00s6w3fjcKNjuaCRdeXf07Wc_K-

20 <http://www.explainity.com/>

21 „Kollaborativ arbeiten“: <http://youtu.be/AWa661KfQG0> (englische Version: <http://youtu.be/Tqr8eoCvnhY>)

22 „Datensicherheit“: <http://youtu.be/ivzIQHkJfq4> (englische Version: <http://youtu.be/LMVLlcGjQCs>)

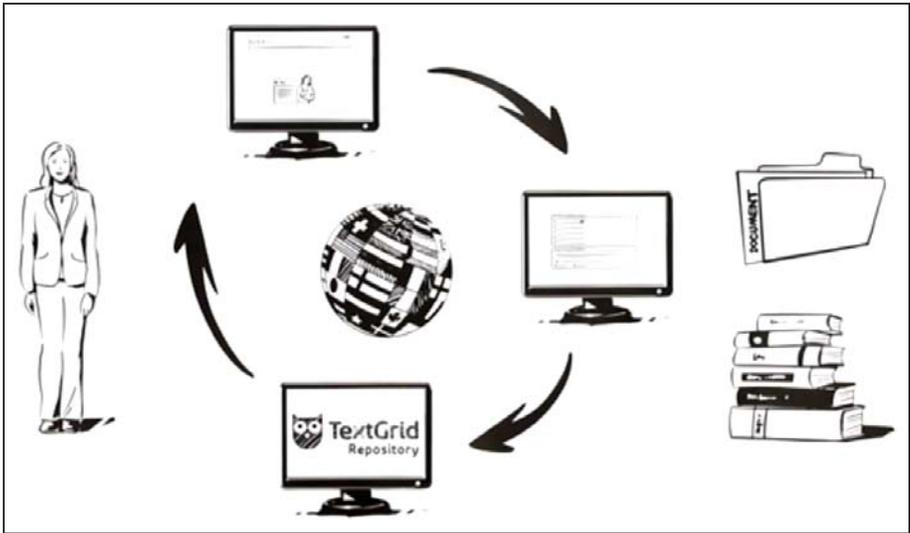


Abb. 3 Ausschnitt aus dem explainity-Clip „Digitales Edieren in einer Virtuellen Forschungsumgebung“

Grundidee aller drei Clips war es, relevante, aber hochkomplexe Themen unkonventionell aufzubereiten und damit einen leichteren Zugang zur vielschichtigen Thematik zu schaffen. Digital Humanities, Virtuelle Forschungsumgebungen, Forschungsdatenmanagement – all das sind Aspekte, mit denen die wenigsten GeisteswissenschaftlerInnen heute vertraut sind, obgleich der weit überwiegende Teil bereits seit langem mit digitalem Material arbeitet, digitale Tools nutzt und von den Vorteilen einer digitalen Präsentation seiner Forschungsergebnisse – wie etwa einer digitalen Edition – profitieren möchte. Ehrgeiziges Ziel der Clips: in drei bis vier Minuten FachwissenschaftlerInnen zu sensibilisieren, sie auf ihrem jeweiligen Wissensstand „abzuholen“, mögliche Vorbehalte abzubauen, das Ganze verständlich und unterhaltsam. Trotz des überwiegend positiven Feedbacks aus der Community bleibt diese Art der Außendarstellung eine Gratwanderung – besteht doch die Gefahr, Komplexität und Vielfalt der Themen einzuebnen und hinter dem z. T. bereits recht hohen Kenntnisstand der Zielgruppe zurückzubleiben.

Mehr Zeit, nämlich knapp acht Minuten, nimmt sich der neue Digital-Humanities-Film des TextGrid- und DARIAH-DE-Teams „Digitale Wissens-

räume: Virtuelle Forschungslabore der Geisteswissenschaften“²³, der am 24. Dezember 2014 Premiere feierte.

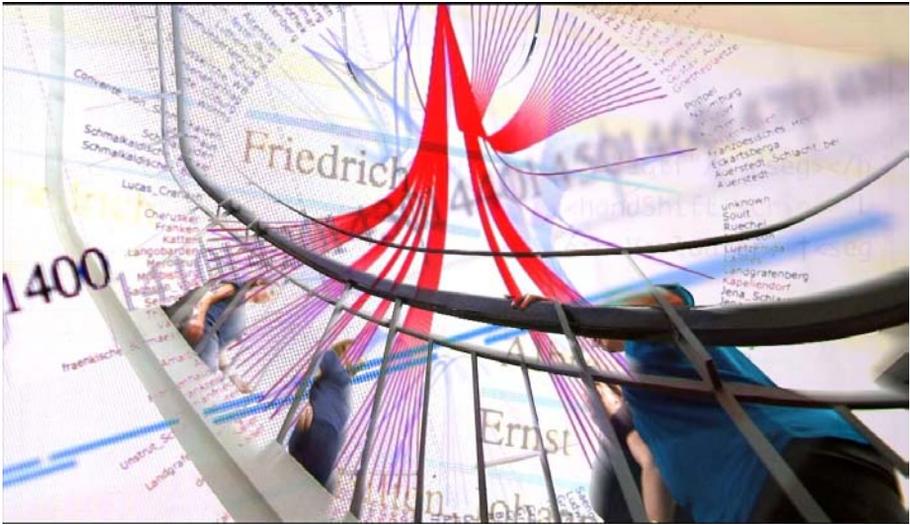


Abb. 4 Ausschnitt aus „Digitale Wissensräume: Virtuelle Forschungslabore der Geisteswissenschaften“

Hatte sein Vorgänger „Virtuelle Forschungswelten: Neue Technologien in den Geisteswissenschaften“²⁴ zuvorderst eine Einführung in die Digital Humanities zum Gegenstand, so stellt der aktuelle Film Forschungsvorhaben vor, die mit TextGrid umgesetzt werden: aus der Editionsphilologie, der Wissenschaftsgeschichte, der Judaistik, Musikwissenschaft und Altamerikanistik. Die Rolle von Infrastruktureinrichtungen wie wissenschaftliche Bibliotheken ebenso wie die Verankerung der Digital Humanities in der Ausbildung des geisteswissenschaftlichen Nachwuchses über Lehre und Curricula und schließlich die aktive, vielfältige und stetig wachsende Community sind weitere Themen des Films.

Grenzgänge und Brückenschläge: *It's all too human*

Die vorangegangenen Beispiele haben gezeigt: Wo der Brückenschlag zwischen den Disziplinen gelingen soll, wo Vielfalt gepflegt und neue Wege

23 <http://youtu.be/tMBq7dlnuLg> (englische Version: http://youtu.be/qedGW_c-nBQ)

24 <http://youtu.be/mQyn3D566Ew> (englische Version: <http://youtu.be/BY9XqXJMbc8>)

gemeinsam beschritten werden wollen, ist Kommunikation essenziell – im Rahmen von Projektverbänden, von Forschungsvorhaben, nach innen und nach außen, auf unterschiedliche Arten, Formaten, Kanälen. Die Herausforderungen, der sich die hier beschriebenen „Grenzgänge“ zwischen Geisteswissenschaft, Informatik und Informationswissenschaft gegenüber sehen, sind zum überwiegenden Teil solche, die in vergleichbaren Konstellationen – unabhängig von spezifischen wissenschaftlichen Disziplinen – ebenfalls auftreten: überall dort nämlich, wo Menschen unterschiedlichen Hintergrunds, unterschiedlicher Spezialisierungen, unterschiedlicher Vorhaben, Arbeitsverfahren und -methoden zusammenarbeiten, um neue, noch kaum etablierte Wege zu gehen. TextGrid mag zukünftigen Forschungsvorhaben der Digital Humanities als Beispiel dafür dienen, wie ein solcher Weg aussehen kann – und weiterhin als Plattform dafür, die sprichwörtlichen „zwei Kulturen“ voneinander profitieren zu lassen.

Literaturverzeichnis

Snow, C. P. (1959): Die zwei Kulturen. In: Kreuzer, Helmut (Hrsg.): *Die zwei Kulturen. Literarische und naturwissenschaftliche Intelligenz. C. P. Snows These in der Diskussion*. München 1987.

TextGrid in Zahlen

NutzerInnen, MitarbeiterInnen, Finanzen

von Nadja Grupe und Martina Kerzel¹

Seit den ersten Ideen für TextGrid sind mehr als zehn Jahre vergangen. „TextGrid hatte 2004 das Ziel, die digitalen Editionswissenschaften zu unterstützen und dafür eine Virtuelle Forschungsumgebung zu entwickeln, ohne dass wir das damals ‚Digital Humanities‘ genannt haben.“² Dass TextGrid schon längst aus den „Kinderschuhen“ herausgewachsen ist und sich zu einer umfassenden kollaborativen Software für verschiedenste geisteswissenschaftliche Communitys entwickelt hat, ist vor allem der Verdienst dieser Communitys, genauer gesagt der NutzerInnen. Denn die Vielfalt geisteswissenschaftlicher Forschungsfragen spiegelt sich auch in der Funktionsvielfalt von TextGrid wider. Doch nicht nur die TextGrid-Nutzenden trugen und tragen zur (Weiter-)Entwicklung und zum Ausbau von TextGrid und seiner Komponenten bei. Ein kurzer Überblick über ausgewählte Faktoren und Rahmenbedingungen anhand des uns zugänglichen Zahlenmaterials zu TextGrid bietet die Möglichkeit, die geleistete Arbeit und auch das investierte Vertrauen in die gemeinsame Vision zu würdigen.

NutzerInnen

Was wissen wir eigentlich über die TextGrid-Nutzenden? Und vor allem: woher? Um für die in TextGrid vorgehaltenen Daten die etablierten wissenschaftlichen Qualitätsstandards – wie z. B. die Nutzung zu ausschließlich wissenschaftlichen Zwecken – zu gewährleisten, ist vor dem Herunterladen

1 Die Autorinnen danken den Kolleginnen und Kollegen Maren Köhlmann, Claudio Leone, Hannes Riebl und Michelle Rodzis für die Unterstützung bei der Zusammenstellung des Zahlenmaterials.

2 Heike Neuroth im Digital-Humanities-Film von TextGrid und DARIAH-DE „Digitale Wissensräume. Virtuelle Forschungslabore für die Geisteswissenschaften“, erschienen im Dezember 2014, <http://youtu.be/tMBq7dlNuLg>

der Software eine Registrierung erforderlich; vor der manuellen Freischaltung der Accounts wird die Affiliation der sich registrierenden Person geprüft.³ Die bei der Registrierung angegebenen E-Mail-Adressen geben bereits Aufschluss über die Herkunft eines Accounts. Die Registrierung ist sowohl über die TextGrid-Homepage als auch über die Authentifikations- und Autorisierungs-Infrastruktur des Deutschen Forschungsnetzes (DFN-AAI) möglich. Nähere Informationen z. B. über den Zeitpunkt der Account-Vergabe oder die Herkunft der verwendeten E-Mail-Adressen können nur bezüglich der Registrierungen über die TextGrid-Homepage getroffen werden, die knapp 95 % der Gesamtregistrierungen umfassen.

Die nun folgenden Angaben zu Account-Herkunft und Registrierungszeitpunkt beziehen die Registrierungen über DFN-AAI nicht mit ein.⁴

*Registrierungen*⁵

Von April 2010 bis einschließlich Juni 2014 wurden – mit einem Monatsdurchschnitt von 32 – insgesamt 1.642 Accounts vergeben.

Bis zur Veröffentlichung von TextGrid 1.0 im Juli 2011 wurden insgesamt etwa 350 Accounts eingerichtet; dabei handelt es sich hauptsächlich um Test-Accounts, die im Rahmen des Software-Entwicklungsprozesses zu Test- und Optimierungszwecken genutzt werden. Nach der Bereitstellung von TextGrid 1.0 im Juli 2011 werden für diesen Monat über 80 Registrierungen verzeichnet. Weitere Registrierungsspitzen folgen im Januar und Februar 2012; die Vergabe der insgesamt rund 80 Accounts steht in Verbindung mit dem ersten TextGrid-Nutzertreffen im Februar 2012. Neben den schon mit TextGrid arbeitenden NutzerInnen nehmen daran auch interessierte Fachcommunitys, EntwicklerInnen und ForscherInnen teil, die TextGrid bisher noch nicht verwenden, eine Nutzung für ihr Forschungsvorhaben planen und sich anlässlich ihrer Teilnahme am Treffen registrieren, um mit der Software arbeiten zu können.⁶ Auf die erste Produktivversion im Sommer 2011 folgen

3 Mit der Bereitstellung der ersten Beta-Version von TextGrid wird seit April 2010 eine Statistik zur Erfassung der Registrierungen geführt. Seit Januar 2013 findet auch eine Dokumentation der Downloadzahlen des TextGrid Laboratory statt.

4 Siehe hierzu den Beitrag von Peter Gietz im vorliegenden Band.

5 Stand: Juli 2014

6 Siehe hierzu den Beitrag von Oliver Schmid et al. im vorliegenden Band.

im Frühjahr 2012 der Betrieb des TextGrid Repository und die Bereitstellung des TextGrid Laboratory zum Download in Version 2.0, was zu einem Anstieg der Registrierungen auf 40 für Mai 2012 führt. Im Juni 2012, ein Jahr nach Veröffentlichung von TextGrid 1.0, liegt die Anzahl der über die Homepage vergebenen Accounts bei rund 680. Im kommenden Jahr werden besonders viele Registrierungen zur Veröffentlichung der TextGrid-Version 2.0.2 im November 2012 und zum zweiten TextGrid-Nutzertreffen im Juni 2013 verzeichnet. Insgesamt zählt TextGrid im Juni 2013, zwei Jahre nach Veröffentlichung der ersten Produktivversion, knapp 930 Accounts. Im nächsten Jahr erfolgen die meisten Account-Vergaben im Juli und August 2013 zur Veröffentlichung der TextGrid-Version 2.0.4, im Januar und Februar 2014 aufgrund des dritten TextGrid-Nutzertreffens sowie im Juni 2014.

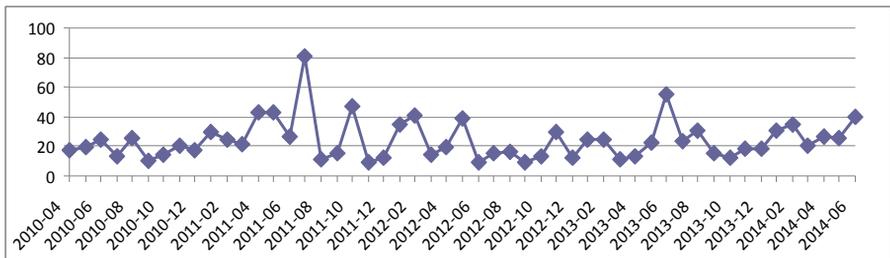


Abb. 1 Monatliche Neu-Registrierungen

Die Registrierungsspitzen der Statistik zeigen deutlich die Auswirkung der regelmäßig stattfindenden Nutzertreffen und der Bereitstellung weiterentwickelter Produktivversionen auf die Anzahl der Accountvergaben. TextGrid kann bis einschließlich Juni 2014 1.564 Registrierungen über die TextGrid-Homepage verzeichnen. Dazu kommen 78 Registrierungen über DFN-AAI, was zu der bereits genannten Gesamtsumme von 1.642 eingerichteten Accounts führt. Die zwischenzeitlich gelöschten, in den ersten Monaten eingerichteten Test-Accounts wurden in dieser Darstellung nicht berücksichtigt, sodass es sich bei der genannten Gesamtzahl um aktive Accounts handelt.

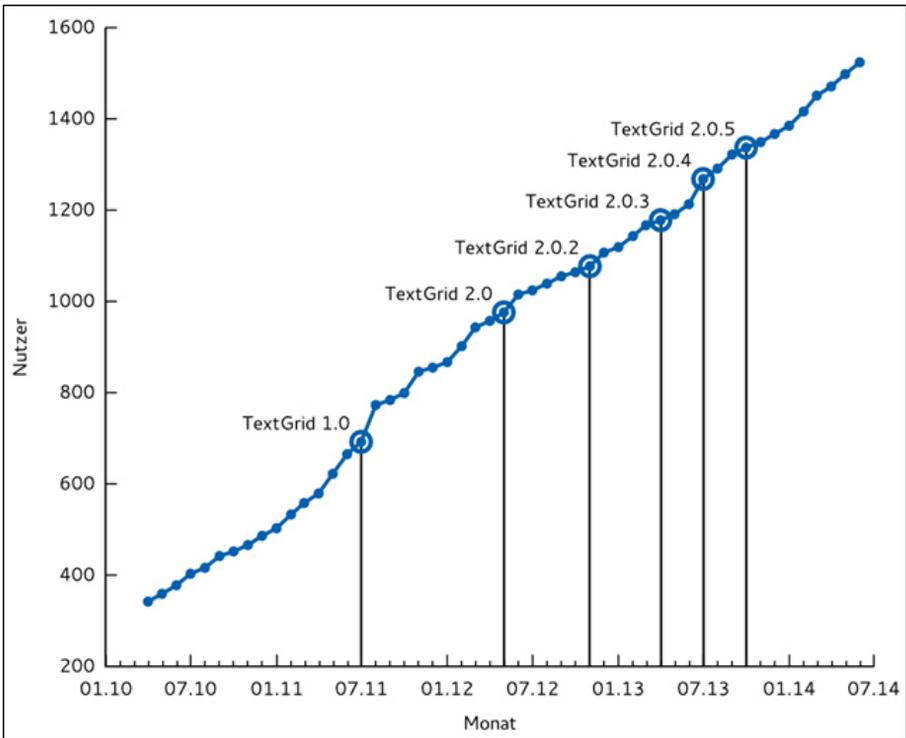


Abb. 2 Registrierungs-Verlauf 2010–2014

*Herkunft der Accounts*⁷

Vor dem Hintergrund, dass TextGrid sich zunächst zuvorderst an eine sprach- und literaturwissenschaftliche Fach-Community im deutschsprachigen Raum richtete, überrascht es nicht, dass 70% der TextGrid-Nutzenden deutschsprachigen Institutionen angehören (Domain „de“). Fünf Prozent der E-Mail-Adressen von TextGrid-Nutzenden stammen aus den USA (Domain „edu“) und etwa 7% der NutzerInnen haben universelle E-Mail-Adressen (Domain „com“, „net“, „org“), die keinem Land zugeordnet werden können. Die übrigen 18% stammen vor allem aus Nachbarländern wie Österreich, der Schweiz, Frankreich und den Niederlanden sowie aus weiteren europäischen Ländern wie Großbritannien, Italien und Belgien. Insgesamt verzeichnet TextGrid 43 verschiedene Domains, verteilt über die ganze Welt – beispielsweise auch aus Namibia, Thailand, Mexiko, China, Brasilien und Israel.

⁷ Stand: Juli 2014

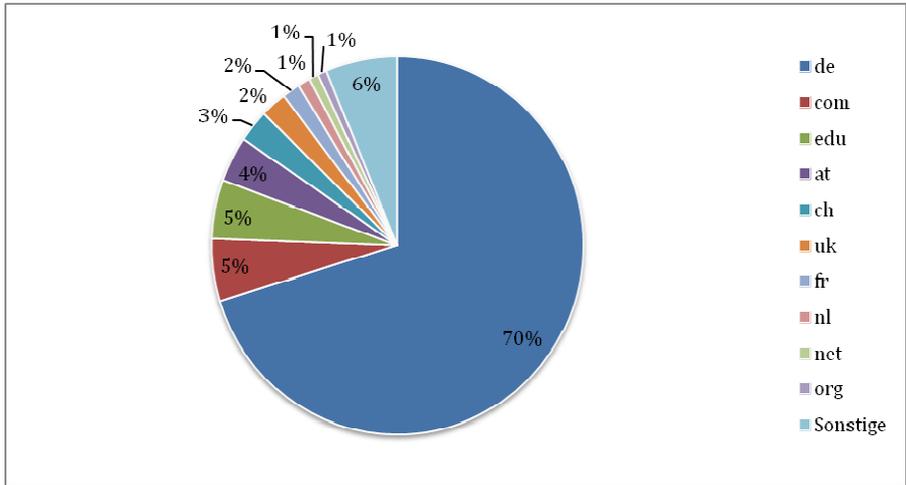


Abb. 3 Herkunft der Accounts nach Länder-Domains

*Downloads*⁸

Seit dem Beginn der statistischen Erfassung der Downloads des TextGrid Laboratory im Januar 2013, zwei Monate nach dem Release der Version TextGrid 2.0.2, bis Juli 2014 wurde die Software insgesamt 1.668 Mal heruntergeladen. Der überwiegende Anteil der Downloads (70%) erfolgte für auf Windows basierende Systeme. Beinahe gleichrangig fallen die Anteile der Mac-Nutzenden mit 16 % und Linux-Nutzenden mit 14 % aus.

Spitzen stehen bei den Downloads ebenso wie bei den Registrierungen mit konkreten Ereignissen in Verbindung. Neben der Veröffentlichung der TextGrid-Version 2.0.5 im Oktober 2013 haben auch hier insbesondere die TextGrid-Nutzertreffen (Juni 2013, Februar 2014) eine positive Auswirkung auf die Download-Zahlen.

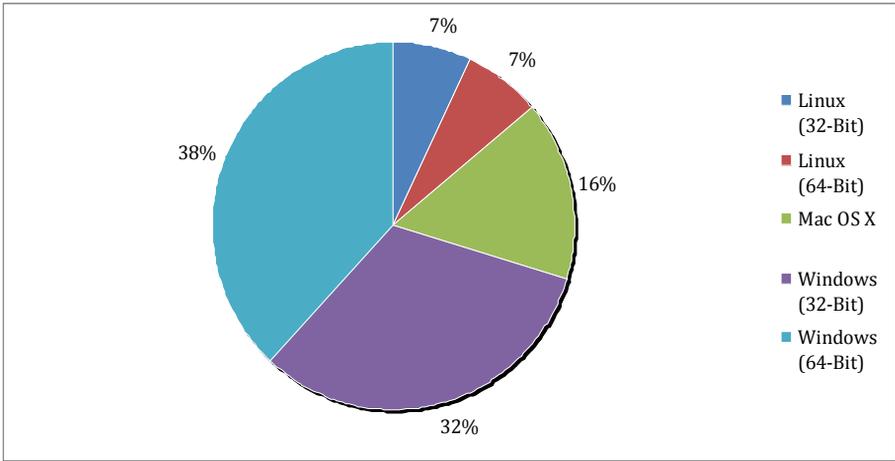


Abb. 4 Downloads nach Betriebssystem

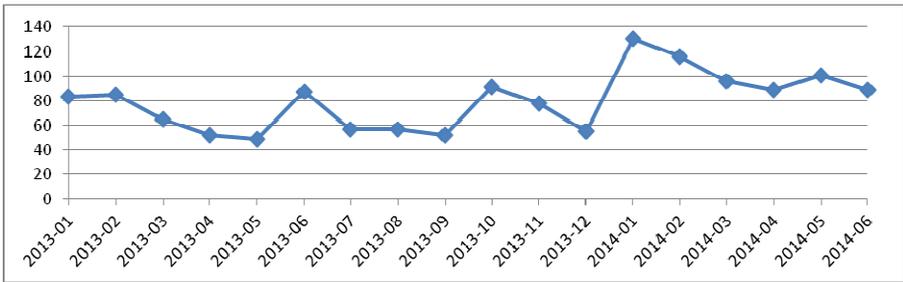


Abb. 5 Monatliche Downloads

Kooperationsprojekte⁹

Derzeit haben rund 63% der 24 aktiven¹⁰ Forschungs- und Entwicklungsprojekte, die mit oder auf Basis von TextGrid zusammenarbeiten, digitale Editionen zum Gegenstand. Dazu zählen historisch-kritische Ausgaben (meist als Hybridpublikation) eines Einzelautors, Handschriftensammlungen, musikwissenschaftliche Editionen, digitale geowissenschaftliche Sammlungsob-

⁹ Stand: November 2014

¹⁰ „Aktiv“ bezieht sich in diesem Abschnitt ausschließlich auf bewilligte oder in Durchführung befindliche Kooperationsprojekte. Weitere Kooperationsprojekte, die sich in Vorbereitung oder im Begutachtungsverfahren befinden, werden weiter unten in die Betrachtung einbezogen.

jekte, Briefeditionen und andere thematische Sammlungen. 17 % der aktiven Kooperationsprojekte konzentrieren sich auf die Weiterentwicklung der Forschungsinfrastrukturkomponenten – seien es beispielsweise Werkzeugneuentwicklungen, -spezifizierungen oder die Zusammenarbeit beim Aufbau eines gemeinsamen Datenrepositoriums. Der Anteil an Projekten, der sich ausschließlich Forschungsfragen widmet, die mithilfe von Textmining-Verfahren oder Visualisierungsmethoden bearbeitet werden, beträgt etwa 13 % der aktiven Kooperationsprojekte. Etwa 8 % der Projekte befassen sich mit der Erstellung digitaler Wörterbücher.

Bezieht man Forschungsvorhaben mit ein, die sich noch in der Vorbereitungs- oder Begutachtungsphase befinden, so ergibt sich ein ähnliches Bild mit leichten Verschiebungen gegenüber den oben erwähnten Anteilen: Der Bereich der digitalen Editionen inklusive Briefeditionen, musikwissenschaftlichen Editionen und Sammlungen dominiert mit rund 69 %. Auf die Weiterentwicklung der Forschungsinfrastruktur und den Bereich der Projekte, die primär Textmining-Verfahren und Visualisierungsmethoden einsetzen, fallen jeweils etwa 11 %. Sie belegen gemeinsam jeweils den zweitgrößten Bereich der prozentualen Verteilung. Die verbleibenden 9 % der laufenden und geplanten Kooperationsprojekte haben Wörterbücher bzw. Lexikografien zum Gegenstand.

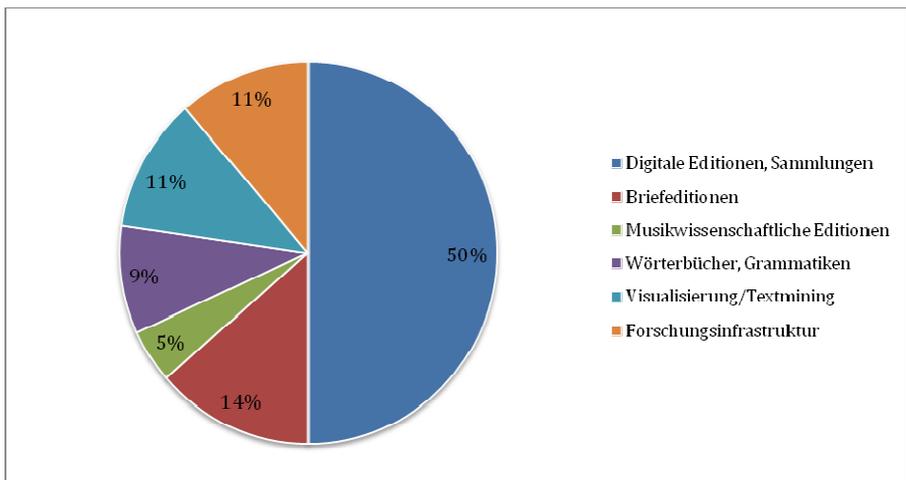


Abb. 6 Laufende und geplante Kooperationsprojekte. Prozentuale Verteilung der Forschungs- und Entwicklungsschwerpunkte

Insgesamt lässt sich ein stabil überwiegender Anteil an Editionsprojekten unterschiedlichster Fachdisziplinen (inkl. interdisziplinärer Fragestellungen) feststellen. Der Anteil an Projekten, die Textmining-Verfahren und Visualisierungsmethoden zur Beantwortung ihrer Forschungsfragen einsetzen, wächst, während der Anteil an Projekten, die sich der Weiterentwicklung von Forschungsinfrastrukturelementen widmen, sinkt. Ob dies ein Trend ist, der sich weiter fortsetzt, darf bei der schnellen Abfolge von Technologieerneuerungszyklen bezweifelt werden. Ohne diejenigen Phasen, in denen der Anteil an Forschungs- und Entwicklungsprojekten mit dem Fokus auf der Weiterentwicklung von Forschungsinfrastrukturelementen zunimmt, wird TextGrid einen stabilen Service bieten, aber nur schwer an aktuelle Entwicklungen anknüpfen können.

Der Forschungsverbund

Während der bisherigen Projektlaufzeit haben insgesamt mehr als 100 FachwissenschaftlerInnen, InformationswissenschaftlerInnen und InformatikerInnen aus 18 verschiedenen Institutionen verteilt über ganz Deutschland gemeinsam daran gearbeitet, TextGrid aufzubauen, beständig weiterzuentwickeln und zu verbessern. Bereits in der ersten Förderphase von Oktober 2005 bis Mai 2009, die den Aufbau einer „modularen Plattform für verteilte und kooperative wissenschaftliche Textverarbeitung“ zum Ziel hatte, war der TextGrid-Forschungsverbund mit acht nationalen Partner-Institutionen und insgesamt 35 MitarbeiterInnen unterschiedlichster fachlicher Hintergründe breit aufgestellt. In der zweiten Projektlaufzeit von Juni 2009 bis Mai 2012, in der der Ausbau inklusive Einbettung des generischen Forschungsdaten-Repositoriums in die TextGrid-Umgebung sowie die Community-Bildung im Vordergrund standen, verdoppelte sich mit 71 KollegInnen aus insgesamt zehn Institutionen die Größe des TextGrid-Konsortiums. Dies spiegelt die fundamentale Bedeutung der zweiten Förderphase für den Gesamterfolg des TextGrid-Projektes wider: Mit dem Ausbau des Laboratory, der Implementierung des Repository und dem Aufbau einer aktiven Community wurden hier die Grundpfeiler von TextGrid gelegt und gefestigt. In der dritten und letzten Förderphase von Juni 2012 bis Mai 2015, deren Fokus auf der Etablierung eines nachhaltigen Dauerbetriebs und dem Ausbau einer langfristigen Nutzerbasis liegt, sind insgesamt 47 Personen aus elf Einrichtungen in

TextGrid beschäftigt. Die Anzahl der Partnerinstitutionen ist also nochmals gestiegen.

In TextGrid I und II setzte sich das Konsortium jeweils aus drei bis fünf Universitäten, einer Fachhochschule, einer Bibliothek, ein bis zwei Forschungseinrichtungen und ein bis zwei kommerziellen Partnern zusammen. In TextGrid III gehören außerdem noch eine Akademie der Wissenschaften und ein Rechenzentrum zum Forschungsverbund.

In den vergangenen zehn Jahren hat sich die Anzahl der Partnerinstitutionen in jeder Projektphase erhöht; zugleich hat der Forschungsverbund im Hinblick auf weitere, in den Blick zu nehmende Fragestellungen (z. B. Usability-Aspekte, Nachhaltigkeit) an Vielfalt gewonnen. Die Aufnahme des Göttinger Rechenzentrums GWDG in das Konsortium spiegelt die wachsende Bedeutung einer verlässlichen und nachhaltigen technischen Infrastruktur wider.

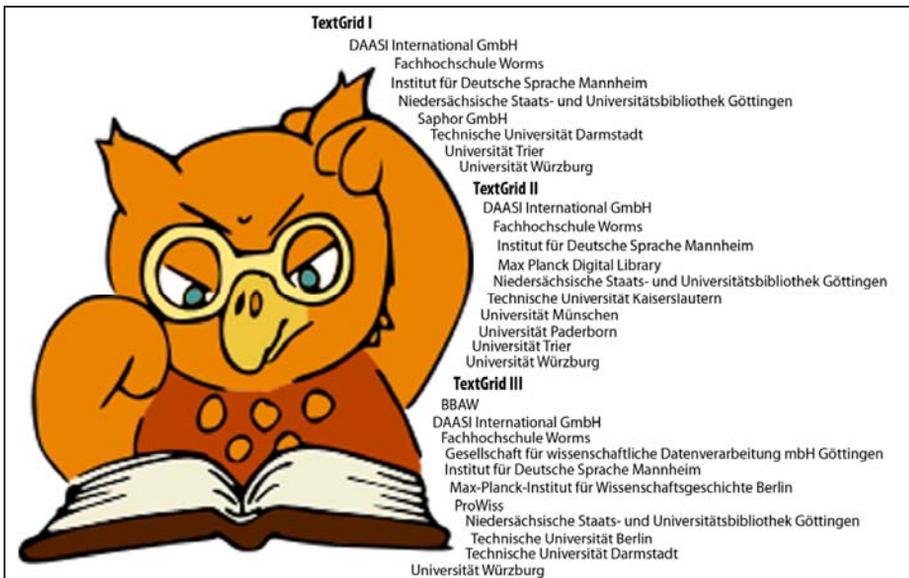


Abb. 7 Das TextGrid-Konsortium in drei Projektphasen

Beständig gewachsen ist zudem der Anteil der Mitarbeiterinnen im Forschungsverbund: Während in TextGrid I nur 11 % der Mitarbeitenden weiblich waren, stieg die Zahl in TextGrid II bereits auf 21 % und in TextGrid III noch einmal auf insgesamt 23 %. Abbildung 7 zeigt, dass einige Partner und Mitarbeiter TextGrid bereits seit der ersten Stunde begleiten; andere sind im

Laufe der Jahre ausgeschieden, neue sind hinzugekommen. Dies ist zu einem großen Teil den sich verändernden Schwerpunkten innerhalb der drei unterschiedlichen Förderphasen geschuldet, der Suche nach neuen Impulsen sowie nach neuen Fachdisziplinen, die durch ihre Forschungsfragen das Anwendungsspektrum von TextGrid immer wieder überprüfen und so die Weiterentwicklung der Forschungsinfrastruktur vorantreiben. Die gleichzeitig überwiegende personelle und institutionelle Kontinuität in der Zusammensetzung des Konsortiums bildet demgegenüber die Grundlage für einen lückenlosen Wissenstransfer und die langfristige Verfolgung und Umsetzung von Zielen – nicht zuletzt bewegt man sich auf bewährten Pfaden in einem gut eingespielten Team.

Fördermittel

TextGrid wird inzwischen in der dritten, im Mai 2015 endenden Förderphase durch das BMBF¹¹ finanziert. Insgesamt waren und sind für den Aufbau dieses Forschungsinfrastrukturbausteins rund acht Millionen Euro an Fördermitteln inklusive Projektpauschale bewilligt. Die Projektpauschale stellt das BMBF seit dem Jahr 2011 mit 10% und seit 2012 mit 20% im Rahmen von Erstbewilligungen für deutsche Hochschulen bereit.¹² Abzüglich der Projektpauschale, die für die Deckung indirekter Kosten an den Hochschulen vorgesehen ist, werden bei Projektende insgesamt 7,7 Millionen Euro für Personal und Sachkosten aufgewendet worden sein. Mit jeweils rund drei Millionen Euro stellen die zweite und dritte Förderphase die kostenintensivsten in der TextGrid-Geschichte dar: In diesen Phasen war eines der Hauptanliegen des Projekts die Weiterentwicklung des Prototypen hin zu einem skalierbaren Produktivbetrieb, bevor in der dritten Phase außerdem der nachhaltige Betrieb zu größerer Bedeutung gelangte. Die Neu- und Weiterentwicklung fachwissenschaftlicher Tools in TextGrid bildete dabei in jeder Phase einen wesentlichen Arbeitsbereich.

In allen Phasen nimmt der Bereich der Personalmittel den Hauptanteil der Aufwendungen ein – ein Phänomen, das in Projekten der geistes- und sozialwissenschaftlichen Disziplinen nicht überraschen dürfte. Im Vergleich zu manchen natur- und lebenswissenschaftlichen Disziplinen entfällt hier z. B. die Investition in Großgeräte für die Grundlagenforschung.

11 Bundesministerium für Bildung und Forschung, <http://www.bmbf.de/>

12 Vgl. <http://www.bmbf.de/de/1398.php>.

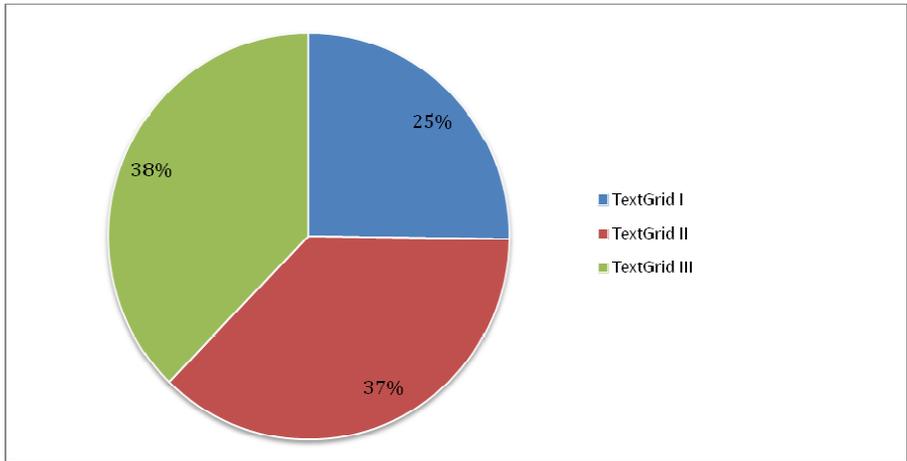


Abb. 8 Prozentuale Verteilung der Gesamtfördersumme von rund 8 Millionen Euro über die drei Förderphasen

Der Personalkostenanteil lag mit über 90 % des Gesamtbudgets in der ersten TextGrid-Förderphase am höchsten. In der darauf folgenden Projektphase betrug er rund 80 %, um in der laufenden dritten Phase auf rund 65 % abzusinken. Damit verbunden ist ein Anstieg der Sachkostenpositionen: von 10 % in Phase I über 20 % in Phase II und schließlich auf rund 35 % für Phase III. Diese Schwankungen lassen sich mit der inhaltlichen Perspektive der in den einzelnen Phasen schwerpunktmäßigen Arbeiten (vereinfacht dargestellt) erklären: Phase I: Konzeption und Prototyping – Phase II: Stabilisierung, Skalierbarkeit, Aufbau der Community – Phase III: Nachhaltigkeit, Kooperationsausbau, Zusammenführung mit Forschungsinfrastrukturen wie z. B. DARIAH-DE¹³. Der Aufbau der Community und damit auch die zunehmenden Kooperationsanfragen führten zu einer Erhöhung des Sachmittelbudgets z. B. für gemeinsame Workshops, Konferenzen und Reisen sowie für Aufträge, die ein professionelles Design von Print- und Webmaterial gewährleisten, und die es erlauben, auch einmal neue, kreative Wege der Außendarstellung zu beschreiten, etwa über Image-Filme oder sogenannte „Erklär-Videos“.¹⁴

13 <https://de.dariah.eu/>

14 Filme: „Virtuelle Forschungswelten: Neue Technologien in den Geisteswissenschaften“ (<https://www.youtube.com/watch?v=mQyn3D566Ew>), 2012, sowie „Digitale Wissensräume: Virtuelle Forschungslabore der Geisteswissenschaften“ (<https://www.youtube.com/watch?v=tMBq7dlnuLg>), 2014; Erklär-Video: „Digitales Edieren

Die Vielfalt der Aufgaben im Rahmen des TextGrid-Projekts erforderte und erfordert Expertise in den unterschiedlichsten Bereichen – nämlich fachwissenschaftliche, informationswissenschaftliche, informatische, aber auch betriebswirtschaftliche und förderpolitische Kompetenz. Diese Bandbreite spiegelt sich in sämtlichen Projektphasen in der Beteiligung von Partnern unterschiedlicher Fachrichtungen und Institutionen wider.

Bei den Hochschulen¹⁵ liegen mit 65 % rund zwei Drittel des Gesamtbudgets über die Gesamtförderdauer, gefolgt von kommerziellen Partnern¹⁶ und Facheinrichtungen¹⁷ mit jeweils 13%. Je vier Prozent des Gesamtfördervolumens erhielten die Max-Planck-Gesellschaft¹⁸ und die Fachhochschule Worms. In der letzten Förderphase konnte die Berlin-Brandenburgische Akademie der Wissenschaften für das TextGrid-Projekt gewonnen werden (rund 1 % des Gesamtfördervolumens).

in einer Virtuellen Forschungsumgebung“ (<https://www.youtube.com/watch?v=jJJgrZ4AaVE>), 2014

15 Zu den Hochschulen im TextGrid-Projektverbund zählen innerhalb der Gesamtförderdauer die TU Berlin, die TUs Darmstadt und Kaiserslautern, die Universitäten Würzburg, München, Paderborn (Musikwissenschaftliches Seminar Detmold/Paderborn) und Trier sowie die Universität Göttingen, vertreten durch die Niedersächsische Staats- und Universitätsbibliothek, die die Konsortialleitung innehat. Bei der Konsortialleitung liegen jeweils Mittel für das Gesamtkonsortium, wie z.B. Budgets für Fachveranstaltungen, Öffentlichkeitsarbeit etc.

16 kommerzielle Partner: DAASI International GmbH, Saphor

17 Facheinrichtungen: Institut für Deutsche Sprache, Gesellschaft für wissenschaftliche Datenverarbeitung mbH

18 Beteiligt waren in verschiedenen Phasen der Projektlaufzeit zunächst die Max Planck Digital Library und das Max-Planck-Institut für Wissenschaftsgeschichte.

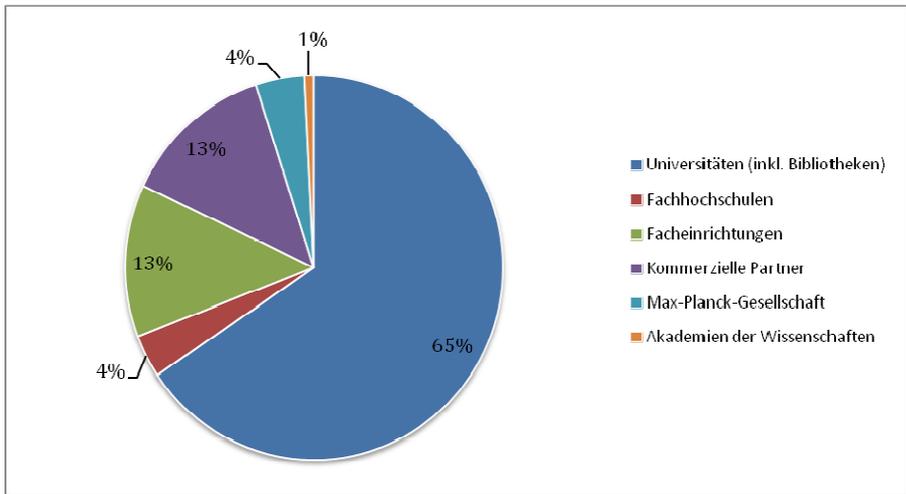


Abb. 9 Prozentuale Verteilung der Gesamtfördersumme auf die Einrichtungen

Die Verknüpfung von Forschung und Lehre, die Zusammenarbeit mit der Community und die Förderung von NachwuchswissenschaftlerInnen sind einige der zentralen Motivationselemente des Projekts, wie u. a. diese Zahlen belegen. Andererseits ist die Anbindung an etablierte wissenschaftliche Einrichtungen, die eine versierte und nachhaltige Bereitstellung von Forschungsinfrastrukturbausteinen – wie TextGrid Repository und Laboratory – sicherstellen, die notwendige Basis, um mit der Community und für die Community arbeiten zu können.

TextGrid-Code

Zählbar im Projektzusammenhang ist neben finanzieller Ausstattung und Personen, die TextGrid fördern, nutzen und weiterentwickeln, letztendlich das Ergebnis, das als ein Teil der gemeinsamen Errungenschaft gelten kann – nämlich die programmierten Codezeilen. Bis Juni 2014 sind knapp 400.000 Zeilen Code für TextGrid Repository und TextGrid Laboratory entstanden. In etwa denselben Umfang hat auch die Kernsoftware eines NASA-Shuttles.¹⁹ Dabei nimmt den Hauptanteil mit 89 % das TextGrid Laboratory ein; die restlichen 11 % entfallen auf das TextGrid Repository.

¹⁹ Vgl. http://www.nasa.gov/mission_pages/shuttle/flyout/flyfeature_shuttlecomputers.html. Weitere anschauliche Vergleiche von Codezeilenmengen sind hier zu finden: <http://www.informationisbeautiful.net/visualizations/million-lines-of-code/>.

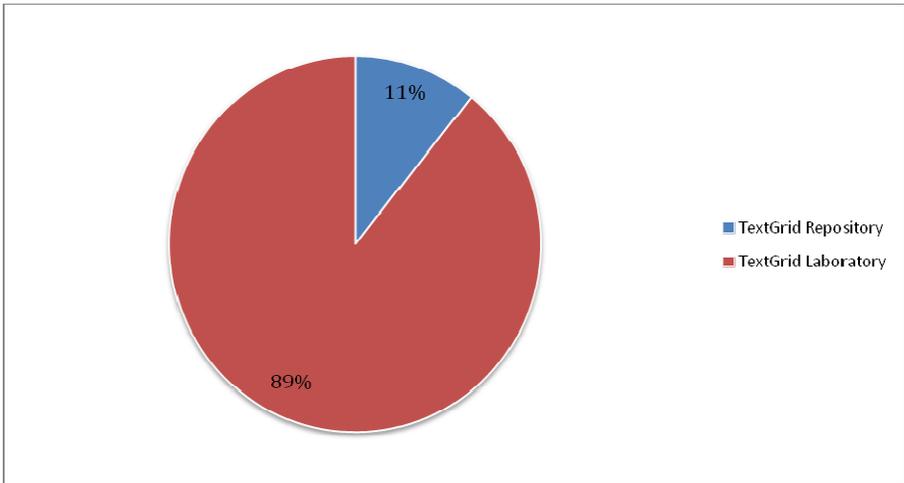


Abb. 10
Prozentuale Verteilung der Codezeilenanzahl auf TextGridRep und TextGridLab

Fazit

Zehn Jahre TextGrid: Von der Community – für die Community in Zahlen:

- 1.642 aktive Accounts
- 70 % der TextGrid-Nutzenden gehören einer deutschsprachigen, 30 % einer internationalen wissenschaftlichen Institution an
- 1.668 Downloads des TextGridLab, davon 70 % für Windows-basierte Betriebssysteme
- 69 % der aktiven, geplanten und im Begutachtungsprozess befindlichen Kooperationsprojekte entfallen auf den Bereich „Digitale Edition“, inklusive Briefeditionen, musikwissenschaftlichen Editionen und Sammlungen
- Rund 400.000 Zeilen Code liegen dem TextGrid Laboratory und dem Repository zugrunde.
- Rund acht Millionen Euro stellte das BMBF inklusive Projektpauschale für das Verbundvorhabens über drei Förderphasen bereit; zwei Drittel des Fördervolumens gehen an Hochschulen.



- Mehr als 100 FachwissenschaftlerInnen, InformationswissenschaftlerInnen und InformatikerInnen aus 18 verschiedenen Institutionen, verteilt über ganz Deutschland, haben in den letzten zehn Jahren im TextGrid-Projektverbund zusammengearbeitet; der Frauenanteil hat sich während dieser Zeit verdoppelt.



Der Aufbau einer Virtuellen Forschungsumgebung für die Geistes- und Kulturwissenschaften wie TextGrid reagierte auf eine zunehmende Nachfrage seitens der FachwissenschaftlerInnen nach digitalen Werkzeugen v. a. des philologischen Edierens und kollaborativen Arbeitens. Nutzer- und Downloadzahlen sowie die Anzahl an fachwissenschaftlichen Kooperationsprojekten bestätigen den Bedarf und zeigen, dass mittlerweile eine sehr lebendige, aktive und gut vernetzte TextGrid-Community entstanden ist, die stetig wächst. TextGrid ist ein Ausnahmeprojekt, sowohl hinsichtlich seiner Ausrichtung – es ist eines der wenigen, die Geistes- und Kulturwissenschaften ansprechenden Projekte, die eine komplette Forschungsinfrastruktur für textbasierte Daten anbieten – als auch hinsichtlich Fördervolumen und -dauer sowie in Bezug auf die Größe des Forschungsverbundes. Über zehn Jahre und drei Projektphasen hinweg wurden gemeinsam mit dem Förderer, dem BMBF, und dem Projektträger, dem DLR²⁰, den NutzerInnen und MitarbeiterInnen kleinere und größere Hürden genommen, Erfolgserlebnisse gefeiert und täglich dazulernt.

²⁰ Projektträger im Deutschen Luft- und Raumfahrtzentrum, Fachbereich Geistes- und Sozialwissenschaften: <http://pt-dlr-gsk.de/de/994.php>

Der Weg in die TextGrid-Community

Schulungen, Nutzertreffen, Support

*von Oliver Schmid, Philipp Vanscheidt, Celia Krause,
Andrea Rapp und Simon Rettelbach*

Erfahrungsgemäß sind die ersten Schritte die schwersten – die Arbeit mit der Virtuellen Forschungsumgebung TextGrid macht hier keine Ausnahme. TextGrid bietet aus diesem Grund verschiedene Formen von Schulungen an, um die zukünftigen NutzerInnen bei ihren ersten Schritten zu begleiten. Die Erfahrungen aus den Schulungen und Nutzertreffen wurden von Rapp et al. (2014) im Report „Erfahrungen aus den Nutzerschulungen, Empfehlungen für die weitere Ausgestaltung des Konzepts“ bereits ausführlich dokumentiert. In diesem Beitrag werden die Erkenntnisse aus dem TextGrid-Report R 3.3.1 zusammengefasst und durch weitere, von der TextGrid-Community eingebrachte Aspekte ergänzt.

Rückblick

Im Jahr 2008 wurde bei der Formulierung des Antrags für die zweite Förderphase von TextGrid einer der wichtigsten Faktoren für die Etablierung der Virtuellen Forschungsumgebung mit einem eigenen Arbeitspaket bedacht: die Nutzer-Community. Um TextGrid nicht nur *für* die Community zu entwickeln, sondern auch *mit* ihr, ging man damals einen Schritt weiter, als reine Öffentlichkeitsarbeit zur Bewerbung des innovativen Konzepts zu betreiben. So sollten Schulungen und Workshops angeboten werden, um sowohl NutzerInnen als auch Interessierten die Möglichkeit zu bieten, von MitarbeiterInnen des Projekts in die Arbeit mit der Software eingeführt zu werden.

In den vergangenen fünf Jahren hat sich die Annahme bestätigt, dass das Angebot von Schulungen für WissenschaftlerInnen, die sich für die Nutzung von TextGrid interessieren, ein unverzichtbarer Bestandteil der Öffentlichkeitsarbeit und ein grundlegender Erfolgsfaktor für die Etablierung der Virtuellen Forschungsumgebung ist. Auf der einen Seite wurde EinsteigerInnen

die Möglichkeit geboten, bei den ersten Schritten am ungewohnten digitalen Arbeitsplatz angeleitet zu werden und damit die größte Hürde auf dem Weg zum Arbeiten in der kollaborativen Forschungsumgebung zu überwinden. Auf der anderen Seite konnten erfahrene TextGrid-NutzerInnen in Workshops für Fortgeschrittene ihre Kenntnisse erweitern und ihren Erfahrungsschatz ausbauen.

In der Anfangsphase erwies es sich noch als schwierig, Schulungen ohne größere technische Probleme durchzuführen, da vor der Veröffentlichung von Version 1.0 nur die Beta-Version zur Verfügung stand. Neben der für Beta-Versionen üblichen technischen Instabilität war auch der Funktionsumfang der Software noch sehr eingeschränkt. Dennoch wurden bereits mit der Beta-Version Schulungen angeboten, um TeilnehmerInnen, die an weiteren Schulungsmaßnahmen partizipierten, die Möglichkeit zu geben, die Fortschritte in der Entwicklung des TextGridLab zu verfolgen. Auf diese Weise konnten frühzeitig Rückmeldungen aus der Nutzer-Community in den Entwicklungsprozess einfließen und Verbesserungen an der Software den NutzerInnen der ersten Stunde zur weiteren Begutachtung vorgelegt werden.

Auch nach der Veröffentlichung von Version 1.0 stand zunächst nur eine eingeschränkte Auswahl an Werkzeugen und Funktionen zur Verfügung, allerdings wurden die Fortschritte hinsichtlich der Stabilität und Zuverlässigkeit der Virtuellen Forschungsumgebung für die NutzerInnen deutlich erkennbar. Diese Weiterentwicklung in der Qualität der Software wirkte sich in verschiedener Hinsicht auf die Schulungen aus, deren Ablauf immer seltener vor technische Probleme gestellt war. Während zu Beginn noch Möglichkeiten gesucht werden mussten, TextGrid-Einführungen im Rahmen anderer Veranstaltungen anzubieten, stieg im Laufe der Zeit mit Stabilität und Funktionsumfang der Software auch die Nachfrage nach Schulungen und Workshops.

Obwohl Schulungen primär didaktische Ziele verfolgen, dienen sie im Rahmen von TextGrid auch als ein wichtiger Grundpfeiler der Public Relations. Durch das gute Zusammenspiel mit der Öffentlichkeitsarbeit des Projekts, die hauptsächlich auf die Außendarstellung wie beispielsweise Werbung und Pressearbeit fokussiert, hat sich in den vergangenen fünf Jahren gezeigt, wie eng diese beiden Konzepte im Projekt miteinander verzahnt sind. Die gute Zusammenarbeit zwischen Öffentlichkeitsarbeit und Community-Betreuung bzw. Nutzer-Kommunikation hat erheblich zur Verbreitung und zum Bekanntheitsgrad von TextGrid beigetragen.

Zahlreiche Gespräche mit TextGrid-NutzerInnen haben in den vergangenen Jahren gezeigt, dass die ausführliche Online-Dokumentation, ein gedrucktes User Manual sowie Video- und Text-Tutorien oft nur der erste Schritt sind, um Geistes- und KulturwissenschaftlerInnen die komplexe Software näherzubringen. Eine persönliche, auf das eigene Forschungsvorhaben zugeschnittene Einführung in die Arbeit mit der Virtuellen Forschungsumgebung wird von potenziellen NutzerInnen explizit gewünscht, um im Rahmen von Schulungen die Software und ihre Anwendungsmöglichkeiten praktisch und unter kompetenter Anleitung ausprobieren zu können. Schulungen und Workshops sind in idealer Weise geeignet, TextGrid nachhaltig in der Forschungsgemeinde zu etablieren.

Schulungen und Workshops

Für die Vermittlung von TextGrid-Kenntnissen haben sich im Laufe der Zeit verschiedene Schulungsformen etabliert, die je nach den unterschiedlichen Vorkenntnissen der TeilnehmerInnen im Bereich der Informationstechnologie variieren. Weitere relevante Faktoren für die Schulungsgestaltung sind die fachliche Zusammensetzung des Teilnehmerfeldes sowie die Teilnehmerzahl. In einigen Fällen bietet sich auch eine Kombination der Schulungsformen an, wenn beispielsweise ein Workshop für Fortgeschrittene, der bestimmte Themen vertieft, auf einer Schulung für EinsteigerInnen aufbauen soll.

Zweck der Schulungen

Für EinsteigerInnen aus den Geisteswissenschaften ist die Eclipse-basierte grafische Nutzeroberfläche in der Regel zunächst ungewohnt. Daher ist eine Einführung in die Besonderheiten von Eclipse im Rahmen der TextGrid-Schulungen für EinsteigerInnen unverzichtbar.

Ein wichtiger Gesichtspunkt bei Softwareschulungen ist der Abbau von Hemmschwellen. Bei Geistes- und KulturwissenschaftlerInnen sind diese erfahrungsgemäß größer als bei Vertretern IT-naher Wissenschaftsbereiche wie beispielsweise den Ingenieurs- oder Naturwissenschaften. WissenschaftlerInnen aus den sogenannten MINT-Fächern sind naturgemäß eher bereit, eine Software nur durch autodidaktisches Erlernen der Bedienung einer grafischen Oberfläche oder durch den Einsatz der Hilfefunktion auszuprobieren.

Hier bieten Schulungen neben ausführlicheren Beschreibungen insbesondere die Möglichkeit, Fragen zu stellen. Text- und Video-Tutorials bieten die detaillierte Schilderung von Arbeits- und Funktionsabläufen, die inzwischen auch für einige Grundlagen und bestimmte Arbeitsschritte im TextGridLab angeboten werden.¹ Eine Schulung können sie indessen nicht in vollem Umfang ersetzen. Abgesehen von der Möglichkeit für die an TextGrid interessierten TeilnehmerInnen, Rückfragen stellen zu können, bieten Schulungen dem Projekt die Chance, auf Basis des Feedbacks der SchulungsteilnehmerInnen Veränderungen an der Software und Infrastruktur vorzunehmen. Bei mündlichen Rückmeldungen lassen sich teilweise Verbesserungswünsche zwischen den Zeilen herauslesen, die nie als schriftliche Support-Anfrage formuliert worden wären.

Der unmittelbare Kontakt mit Forschenden und ihren Vorhaben ermöglicht es den Schulenden, die jeweilige Wissenschaftskultur der TeilnehmerInnen kennenzulernen. Da deren Kenntnisse in Bereichen wie XML und XSLT zudem recht unterschiedlich, in der Regel aber oft nicht sehr ausgeprägt sind, ist es wiederum unumgänglich, Bezüge zwischen diesen Techniken und dem jeweiligen Forschungsinteresse und der Sprache der NutzerInnen herzustellen. Umgekehrt müssen diese Interessen als Anforderungen wieder in die Sprache der EntwicklerInnen übersetzt und weitergetragen werden, damit sich der Kreis von der Entwicklung über Schulungen und Rückmeldungen zurück zur Softwareentwicklung schließen kann. Dieser Kreislauf ist wesentlich, um den Erfordernissen der NutzerInnen auch mit der Entwicklung der Software nachhaltig entsprechen zu können. Zudem wird durch die fortlaufende Rückkopplung mit den NutzerInnen bei den TextGrid-MitarbeiterInnen eine Kompetenz für die Wissenschaftskulturen und Arbeitspraktiken der Nutzer-Community aufgebaut, die den nachhaltigen inhaltlichen Hintergrund für die technische Weiterentwicklung der Software und ihre Anpassung an neue Technologien der Software bildet.

Schulungen versus Workshops

Einführungsveranstaltungen für EinsteigerInnen haben häufig Schulungscharakter, da der Aufbau und die Inhalte unabhängig von der Zielgruppe immer relativ ähnlich sind und die Grundlagenvermittlung den Schulenden,

¹ Die Themen „Dokumentation“ und „Tutorials“ werden im Abschnitt „Support“ weiter unten noch einmal aufgegriffen.

abgesehen vom genutzten Datenmaterial, wenig Spielraum bietet. Workshops für Fortgeschrittene bieten dagegen erheblich mehr Freiheiten bezüglich der Bandbreite behandelter Werkzeuge, der inhaltlichen Ausgestaltung des Stoffs sowie des Anspruchsniveaus und der Detailtiefe. Daher können Veranstaltungen für Fortgeschrittene eher wie ein Workshop gestaltet werden, auf dessen Ablauf die TeilnehmerInnen sowohl in der Vorbereitung als auch während der Durchführung mehr Einfluss nehmen können als in einführenden Schulungen, bei denen eine interaktive Mitgestaltung nur in geringerem Maße möglich ist. Die Möglichkeiten, die sich den Schulenden bei Workshops für Fortgeschrittene eröffnen, werden genutzt, um die Veranstaltungen abwechslungsreich zu gestalten.²

Rahmenbedingungen

Einen entscheidenden Einfluss auf die Planung und Durchführung einer Schulungsmaßnahme hat die Zusammensetzung des Teilnehmerfelds. Zentraler Faktor ist dabei, abhängig vom jeweiligen Veranstaltungsrahmen, die Homogenität bzw. Heterogenität der Gruppe. Daneben nehmen auch die Gruppengröße, die für die Schulung zur Verfügung stehenden technischen und personellen Ressourcen, der zeitliche Rahmen und die räumlichen Gegebenheiten Einfluss auf die zu präferierende Veranstaltungsform.

Schulungen heterogener Teilnehmergruppen zeichnen sich gegenüber Veranstaltungen für Arbeitsgruppen oder Projekte dadurch aus, dass sie ein breiteres Publikum anziehen und dadurch üblicherweise auch die Auswahl der gestellten Fragen und die Bandbreite der jeweiligen Bedürfnisse facettenreicher ist. Allerdings bringen Schulungen für ein heterogen zusammengesetztes Auditorium auch Herausforderungen mit sich, denen Schulende im Rahmen von Schulungen bei Projekten mit einer homogenen Teilnehmergruppe selten bis gar nicht begegnen. So bringt die heterogene fachliche Zusammensetzung des Teilnehmerfelds oft unterschiedliche Voraussetzungen für und Erwartungshaltungen an die Schulung mit sich. Anforderungen für die wissenschaftliche Arbeit wie beispielsweise die Verfügbarkeit bestimm-

² Dieses Vorgehen ist insbesondere im Zusammenhang mit den Nutzertreffen sinnvoll und wichtig, zu denen aktive NutzerInnen wiederholt anreisen und durch immer neue Facetten der Arbeit mit TextGrid bei jedem Nutzertreffen einen neuen Anreiz erhalten, weitere Veranstaltungen zu besuchen und Impulse für ihre Arbeit mit TextGrid zu bekommen.

ter Unicode-Zeichen oder Alphabete weisen ebenfalls unterschiedliche Komplexitätsgrade auf. Begnügen sich Mediävisten für die korrekte Transkription eines deutschen Textes aus dem Mittelalter teilweise schon mit der Verfügbarkeit einiger Sonderzeichen – z. B. des langen s („ſ“) – so benötigen Orientalisten neben anderen Zeichensätzen auch die Möglichkeit, abweichende Textrichtungen korrekt darstellen zu können. Oft können Fragen und Wünsche, die in Schulungen und Workshops geäußert werden können, bereits im Vorfeld antizipiert und mit Verantwortlichen und EntwicklerInnen diskutiert werden. Auf diese Weise ist es auch möglich, noch nicht umgesetzte Funktionalitäten, die für die TeilnehmerInnen von Interesse sein dürften, anzukündigen oder zur Diskussion zu stellen.

Gilt es bei fachlicher Heterogenität, die Schulung für das gesamte Auditorium inhaltlich interessant zu gestalten, so stellt die Heterogenität bezüglich der Vorkenntnisse die größere Herausforderung für die Schulenden dar. Solche Probleme können in Schulungen vom Betreuungspersonal aufgefangen werden, zudem wächst mit der Homogenität des Teilnehmerfelds naturgemäß auch das Potenzial, sich auszutauschen und voneinander zu lernen. Dabei kann von einer homogenen Gruppe gesprochen werden, wenn eine Schulung beispielsweise für ein Projekt oder eine Forschungsgruppe durchgeführt oder als Teil einer universitären Lehrveranstaltung angeboten wird, bei der die TeilnehmerInnen einander bereits kennen. Teilweise sind die Gruppen homogen in Bezug auf ihre Kenntnisse und Fertigkeiten, vor allem aber verfolgen ihre Mitglieder im Allgemeinen ein gemeinsames Interesse.

In diesem Zusammenhang hat sich gezeigt, dass mit zunehmender Größe und Heterogenität der Teilnehmergruppe mehr betreuendes Personal sinnvoll und meistens auch nötig ist, um den Lernerfolg auch für weniger erfahrene und IT-affine TeilnehmerInnen zu gewährleisten. Bei einer relativ homogenen Gruppe können einzelne Schulende bis zu ungefähr zehn TeilnehmerInnen ohne Unterstützung betreuen, bei größerer Teilnehmerzahl und Heterogenität ist jedoch die Unterstützung durch weitere betreuende Personen ratsam.

Ablauf und Inhalte

Unabhängig von den meisten Rahmenbedingungen bestehen TextGrid-Schulungen im Normalfall aus zwei grundlegenden Teilen. Der allgemeine Teil bzw. Theorieteil bildet üblicherweise die Einleitung, der spezielle Teil

bzw. interaktive Praxisteil meist den darauf aufbauenden Schwerpunkt der Veranstaltung. Der Umfang der jeweiligen Teile wird dabei stets an das Auditorium angepasst und kann – beispielsweise im Zusammenhang mit universitären Lehrveranstaltungen – um weitere, nicht TextGrid-spezifische Inhalte wie eine XML/TEI-Schulung ergänzt werden.

In der theoretischen Einführung hat es sich als sinnvoll erwiesen, zunächst das Projekt vorzustellen sowie seine Entwicklung über die verschiedenen Projektphasen hinweg zu skizzieren. Im Anschluss werden grundlegende Konzepte wie die Modularität des TextGridLab oder die Zugriffsmöglichkeiten auf die im TextGridRep gespeicherten Daten ausführlich erläutert. An dieser Stelle werden auch die für neue NutzerInnen oft ungewohnten Besonderheiten der Eclipse-basierten grafischen Nutzeroberfläche wie die Gegenüberstellung von Perspektiven und Sichten sowie das Konzept von Projekten und Objekten erklärt. Sowohl in der einführenden Präsentation als auch in der interaktiven Schulung hat es sich als zielführend erwiesen, Fragen nicht nur jederzeit zuzulassen, sondern bereits zu Beginn der Veranstaltung die TeilnehmerInnen dazu zu animieren, solche zu stellen. So ist sichergestellt, dass niemand aufgrund von nicht verstandenen Grundlagen den Anschluss und damit im schlimmsten Fall das Interesse verliert. Im interaktiven Praxisteil der Schulung werden typische Arbeitsschritte im TextGridLab vorgeführt, die dann von den TeilnehmerInnen am eigenen Computer nachvollzogen werden sollen. Dabei erhalten sie von den betreuenden Personen und bei leichter zu lösenden Problemen oft auch von anderen TeilnehmerInnen Unterstützung.

Bei Schulungen für heterogene Gruppen wird der behandelte Stoff eher allgemein gehalten. In der Regel haben die Schulungen einen thematischen Schwerpunkt, insbesondere wenn sie in einen größeren Rahmen wie eine Tagung eingebettet sind, die einen solchen Schwerpunkt erwarten lässt, wie beispielsweise der XML-Editor im Rahmen einer TEI-Konferenz. Wenn sich hingegen die Inhalte einer Schulung bei inhaltlich homogenen Teilnehmergruppen wie Projekten an deren Zielen orientieren, dann wird die Gewichtung der vorgestellten Werkzeuge entsprechend der Interessenlage angepasst.

Bei einem fachlich heterogenen Teilnehmerfeld hat es sich zudem als sinnvoll erwiesen, die Veranstaltung in mehrere einzelne Schulungen verschiedener Formen inklusive Spezialschulungen aufzuteilen. Diese Vorgehensweise ist beispielsweise bei Schulungsmaßnahmen im Rahmen von Konferenzen sinnvoll, um ein größeres Themenspektrum abzudecken und dadurch ein breiteres Publikum erreichen zu können. Beispielsweise fand im

Rahmen des Digital-Humanities-Festakts in Göttingen im Juli 2011 am Vormittag eine Grundlagenschulung statt, die als Einstieg für alle Interessierten vorgesehen war, am Nachmittag wurden drei parallele Schulungen veranstaltet, um den unterschiedlichen Interessen der TeilnehmerInnen gerecht zu werden. In diesen Veranstaltungen wurden die Themengebiete XML-Editor und Metadaten, Text-Bild-Link-Editor sowie Noten-Editor MEISE vertieft. Die Aufteilung in Schulungen und Workshops im Rahmen der Nutzertreffen basiert ebenfalls auf dieser Überlegung.

Nutzertreffen

Nutzertreffen stellen eine besondere Schulungsform dar, da sich ihr Programm aus einer Kombination von Vorträgen, Schulungen und Workshops zusammensetzt. Dadurch bilden die Nutzertreffen gewissermaßen selbst den Tagungsrahmen für ihre eigenen Schulungen.

Entstehungsgeschichte

Die Idee für das erste Nutzertreffen entstand aus der Überlegung heraus, einen interessanten Rahmen für eine Veranstaltung der in der zweiten Projektphase durchzuführenden Workshops anzubieten, der die unterschiedlichen Zielgruppen gleichermaßen anspricht. So wurde mit dem ersten Nutzertreffen eine Kommunikationsplattform für Content-Provider, FachwissenschaftlerInnen und EntwicklerInnen ins Leben gerufen, deren zentrales Ziel der Austausch zwischen diesen Teilnehmergruppen war. Das Angebot, im Rahmen der zweitägigen Veranstaltung Kontakte zu knüpfen, fand großen Zuspruch. Zudem wurde Projekten aus dem TextGrid-Umfeld durch Präsentationen in Form von Vorträgen, Demonstrationen und Postern die Möglichkeit geboten, ihre Arbeit vorzustellen und sich mit interessierten KollegInnen besser zu vernetzen. Bestätigt durch die positiven Rückmeldungen aus dem Teilnehmerfeld hat sich das „TextGrid-Café“ daher als fester Bestandteil der Nutzertreffen etabliert. Der Erfolg des ersten Nutzertreffens führte zu dem Entschluss, solche Nutzertreffen bei der Beantragung der dritten Projektphase als festen Bestandteil in das Community-Konzept von TextGrid aufzunehmen und der ersten Tagung dieser Art in den folgenden drei Jahren weitere sechs Nutzertreffen folgen zu lassen.

Weiterentwicklung

Seit dem zweiten Nutzertreffen haben diese jeweils spezifische Themen zum Gegenstand. Dabei konnten bei jeder Veranstaltung besondere Stärken von TextGrid in den Mittelpunkt gestellt werden. Auf die Themen „Digitale Edition“ und „Bilder in TextGrid“ folgten „Anpassungen und Erweiterungen“ sowie „Metadaten und Normdaten“; beim sechsten Treffen wird im März 2015 als Ausnahme von der Regel eine „Grand Tour“ durch die Forschungs-umgebung veranstaltet. Die Besucher können sich je nach Interesse an verschiedenen Stationen informieren und in den verschiedenen Werkzeugen üben. Den vorläufigen Abschluss der Schulungen im bisherigen Rahmen wird das Themenfeld „Publizieren und Archivieren“ bilden. Darüber hinaus werden die Schulungen und Workshops seit dem zweiten Nutzertreffen als mehrteilige Veranstaltungen hierarchisch strukturiert. Inzwischen werden die Nutzertreffen mit einer Schulung für EinsteigerInnen eröffnet, um an TextGrid Interessierten den Einstieg zu erleichtern; im Laufe der Veranstaltung bauen weitere Workshops auf die erworbenen Grundkenntnisse auf und finden abhängig von der inhaltlichen Ausrichtung parallel oder sequenziell statt.

Beim zweiten Nutzertreffen war zunächst nur aktiven oder abgeschlossenen Projekten die Möglichkeit geboten worden, ihre Arbeit und Ergebnisse vorzustellen, wobei die Bandbreite von studentischen Abschlussarbeiten bis hin zu Projekten im Akademien-Programm reichte. Im Rahmen des dritten Nutzertreffens wurden mit der Vorstellung von Projektideen hingegen neue Wege beschritten, da nicht nur Projektskizzen mit den dazugehörigen Fragen und Problemen präsentiert wurden, sondern auch die mit TextGrid entwickelten Lösungen für die im Vortrag geschilderten Fragestellungen in einem darauf aufbauenden Workshop aufgezeigt wurden. Auf diese Weise war durch die inhaltliche Verknüpfung von Vorträgen und Workshops in Form eines Dialogs mit Fragen und Antworten auch ein „roter Faden“ während des Nutzertreffens erkennbar. Die Reaktionen der TeilnehmerInnen bestätigten in diesem Zusammenhang das von uns bereits bei Schulungen beobachtete Phänomen, dass unter bestimmten Voraussetzungen die Hemmschwelle sinkt, Fragen zu stellen. Da die Vortragenden in ihren Beiträgen selbst mehr Fragen als Antworten vorgestellt hatten, sank die Hemmschwelle bei den ZuhörerInnen, selbst Fragen zu stellen und damit die eigenen Defizite an Kenntnissen und Erfahrung vor Publikum einzugestehen. Des Weiteren wird seit dem dritten Nutzertreffen mit dem Untertitel „Das Treffen für

TextGrid-NutzerInnen und alle, die es werden wollen“ bereits bei der Ankündigung unterstrichen, dass mit den Nutzertreffen nicht nur erfahrene TextGrid-NutzerInnen angesprochen werden sollen, sondern auch Interessierte ganz ohne oder mit wenig Erfahrung im Umgang mit der Virtuellen Forschungsumgebung willkommen sind. Diese Darstellung hat mit Sicherheit dazu beigetragen, auch EinsteigerInnen eine entspannte Atmosphäre zu bieten und die Angst vor „dummen Fragen“ abzubauen.

Erkenntnisse

Kommunikation ist ein wichtiger Erfolgsfaktor, in der Forschung gleichermaßen wie in anderen Aufgabenbereichen, die auf Zusammenarbeit zwischen Personen angewiesen sind. Daher sind sowohl der Austausch von Erfahrungen als auch das Knüpfen von Kontakten wichtige Gesichtspunkte der TextGrid-Nutzertreffen. Insofern gehört das inzwischen etablierte „TextGrid-Café“ als Plattform zum Austausch der NutzerInnen untereinander, aber auch mit den MitarbeiterInnen von TextGrid, in Verbindung mit den Vorträgen und Workshops zu den drei tragenden Säulen des Konzepts. In diesem informellen Rahmen kommen TextGrid-NutzerInnen verschiedener Fachrichtungen mit unterschiedlichen Kenntnissen miteinander ins Gespräch und so können sowohl fortgeschrittene NutzerInnen Erfahrungen austauschen als auch EinsteigerInnen in einer ungezwungenen Atmosphäre von erfahrenen WissenschaftlerInnen lernen.

Dass die EinsteigerInnen bei den Nutzertreffen eine wichtige Rolle spielen, zeigen die Zahlen der Anmeldungen zu den Schulungen und Workshops im Rahmen der Nutzertreffen. So hatten sich beim dritten Nutzertreffen drei Viertel der TeilnehmerInnen für die einführende Schulung zu Beginn des Nutzertreffens angemeldet, wengleich sich unter diesen Personen auch zahlreiche NutzerInnen befanden, die bereits an einer anderen Schulungsmaßnahme teilgenommen hatten und demnach schon Vorkenntnisse besaßen. Dennoch unterstreicht die Teilnahme zahlreicher neuer Interessierter das stetig wachsende Interesse an TextGrid und dessen (schulenden) Veranstaltungen und bestätigt, welch wichtiger Bestandteil der Nutzertreffen die Einsteiger-Schulungen sind.

Support

Als weitere Formen der Wissensvermittlung bietet TextGrid neben den interaktiven Angeboten eine umfassende Dokumentation sowie Tutorials an, die auf der TextGrid-Website zur Verfügung gestellt werden. Das Benutzerhandbuch dient als umfangreiches Nachschlagewerk und ist neben der downloadbaren PDF-Variante als kontextbezogener Hilfetext in das TextGridLab integriert. Die Dokumentation wird in einem Wiki gepflegt, aus dem die beiden Formen des Handbuchs resultieren. Im Gegensatz dazu fokussieren die Text-Tutorials auf die schrittweise detaillierte Beschreibung bestimmter Arbeitsabläufe in der Virtuellen Forschungsumgebung. Sie stehen, ergänzt durch zahlreiche Screenshots, als Webseiten und druckbare Fassung in Form von PDF-Dateien zur Verfügung,³ die Video-Tutorials werden zusätzlich über den DHd-Kanal auf YouTube angeboten.⁴ Handbücher und Tutorials bieten jedoch – abgesehen von der Kommentarfunktion im Wiki – keine Kommunikationsmöglichkeiten zur Klärung von Fragen oder für Anregungen.

Kommunikationswege

An elektronischen Kommunikationsplattformen zur Klärung von Fragen und Problemen können die NutzerInnen zwischen der Fehlerverwaltung Chili, dem Nutzerforum und der Support-Mailingliste von TextGrid wählen. Dabei hat keine der drei Optionen einen schulenden Charakter, vielmehr sollen sie als Instrument zur Klärung von Fragen, zur Meldung von aufgetretenen Fehlern und zum Austausch von Informationen und Erfahrungen dienen.

Mit der Software Chili können Fehler und ihre Lösungen sowie Verbesserungsvorschläge dokumentiert werden, allerdings wird die webbasierte Kommunikationsplattform primär von TextGrid-MitarbeiterInnen genutzt. Dagegen ist das TextGrid-Nutzerforum hauptsächlich nutzerbetrieben und nicht vollständig moderiert. Die Antworten auf Anfragen stammen von anderen TextGrid-NutzerInnen, MitarbeiterInnen des Projekts verfassen in diesem Kontext ihre Beiträge in ihrer Rolle als NutzerInnen.

3 Dokumentation und Tutorials: <http://www.textgrid.de/ueber-textgrid/dokumentation-tutorials/>.

4 YouTube-Kanal für Digital Humanities im deutschsprachigen Raum: <http://www.youtube.com/user/dhdkanal>.

Die seit 2009 bestehende Support-Mailingliste ist der am häufigsten gewählte Kommunikationsweg und wird hauptsächlich von den MitarbeiterInnen der Bereiche Nutzer-Kommunikation und Öffentlichkeitsarbeit gepflegt. Dieser Kommunikationsweg richtet sich sowohl an aktive TextGrid-NutzerInnen als auch an -Interessierte, primär genutzt wird das Angebot aber von NutzerInnen, die bereits mit TextGrid arbeiten. Dabei umfasst die Bandbreite der Anfragen unter anderem technische Probleme inklusive der im TextGridLab erstellten Fehlermeldungen, Fragen zur Oberfläche und deren Funktionen sowie zu XML/TEI, aber auch Anfragen für TextGrid-Schulungen und -Workshops. Bei Bedarf werden bei Fehlermeldungen die EntwicklerInnen des Projekts zur Lösung des Problems zu Rate gezogen, meistens werden diese Mitteilungen auch im Chili dokumentiert, wenn beispielsweise für die Lösung eines Problems mehr Aufwand betrieben werden muss als der Neustart eines nicht mehr fehlerfrei arbeitenden Dienstes.

Häufig werden die NutzerInnen aber auch außerhalb dieser Kommunikationswege unterstützt, wenn bereits ein persönlicher Kontakt zu ihnen besteht. So nehmen NutzerInnen teilweise lieber Kontakt zu TextGrid-MitarbeiterInnen auf, die sie persönlich kennen, als ihre Anfrage über die „anonyme“ Mailingliste zu versenden. Dieses Verhalten lässt sich möglicherweise mit einer niedrigeren Hemmschwelle erklären, wie sie bereits im Zusammenhang mit dem Verhalten in Schulungen und Nutzertreffen beschrieben wurde: Bedenken, mit einer Frage Defizite oder Schwächen kundzutun, spielen bei einem vertrauten Umfeld eine untergeordnete Rolle. Da die MitarbeiterInnen des Supports durch den regelmäßigen Kontakt zu den NutzerInnen Erfahrung mit der Bandbreite ihrer IT-Kenntnisse haben, sind sie auch im Umgang mit weniger IT-affinen GeisteswissenschaftlerInnen erfahren und können bei beliebigen Anfragen Problemlösungen fast immer einfach und nachvollziehbar erklären.

Ausblick

Mit der Verbreitung der Virtuellen Forschungsumgebungen werden sich die damit verbundenen Arbeitstechniken, Technologien und Software-Kenntnisse bei den NutzerInnen etablieren. Dazu gehören auch in Zukunft die Fortführung der Grundlagenschulung, die kontinuierliche Verbesserung der Usability der Software auf Basis der Rückmeldungen aus der Community und die Spezialisierung der Software anhand der Kooperation mit konkreten

Forschungsprojekten. Den Schulungen und Nutzertreffen kommt dabei neben den didaktischen Aufträgen auch die wichtige Rolle zu, der „virtuellen Umgebung“ ein reales Pendant gegenüberzustellen, in dem eine persönliche Didaktik stattfindet und der unmittelbare zwischenmenschliche Austausch ermöglicht wird.

Schulungsmaßnahmen sind nicht nur im Rahmen von TextGrid sinnvoll und wichtig, sondern allgemein, wenn es um den effektivsten Weg geht, AnwenderInnen den Umgang mit komplexen Werkzeugen nahezubringen. Dies gilt insbesondere für Werkzeuge mit besonders großem Funktionsumfang, denn so reich an Funktionen der Text-Bild-Link-Editor als Bestandteil des TextGridLab ist, so zeitaufwendig ist auch der Prozess, den Umgang mit diesem Werkzeug zu erlernen. Ähnlich verhält es sich mit dem von DARIAH-DE angebotenen Geo-Browser, dessen Anwendungsmöglichkeiten ebenso reichhaltig sind wie der Lernprozess lang. In welcher Form die Werkzeuge von DARIAH-DE und TextGrid kombiniert werden können, wurde im Rahmen des dritten Nutzertreffens „Bilder in TextGrid“ eindrucksvoll präsentiert. Auf lange Sicht gesehen ist die Verknüpfung von Werkzeugen beider Projekte wünschenswert, zumal sich auch die Schulungsmaßnahmen von TextGrid und DARIAH-DE gut vereinigen ließen. Beispiele für kombinierte Anwendungsmöglichkeiten gibt es schließlich genug, wie schon die Präsentation von DigiVoy im Rahmen des zweiten TextGrid-Nutzertreffens gezeigt hat. Diesem Vorbild können weitere Projekte folgen, sodass die Vernetzung TextGrids mit anderen Projekten ähnlicher Ausrichtung auch zukünftig zu Synergien und neuen Angeboten auch im Bereich der Schulungen führen kann.

Im Mai 2015 endet die dritte Förderphase von TextGrid, doch auch danach wird es weiterhin Angebote für die stetig wachsende TextGrid-Community geben. Weitere Veranstaltungen wie z. B. ein Nutzertreffen im zweiten Halbjahr 2015 sind bereits in der Planung. Gemeinsame Aktivitäten von TextGrid und DARIAH-DE werden im Rahmen von Schulungen und Workshops eine wichtige Rolle bei der Unterstützung der DH-Community spielen.

Literaturverzeichnis

Rapp, Andrea; Schmid, Oliver; Vanscheidt, Philipp; Krause, Celia (2014): Report R 3.3.1: Erfahrungen aus den Nutzerschulungen, Empfehlungen für die weitere Ausgestaltung des Konzepts, http://www.textgrid.de/fileadmin/TextGrid/reports/TextGrid_Report_3_3_1_Final.pdf.

Virtuelle Forschungsumgebungen nachhaltig organisieren – ein Erfahrungsbericht

von Andreas Witt, Ludwig Eichinger und Norman Fiedler

Als der Ausgräber Walter Andrae auf Geheiß der Deutschen Orientgesellschaft im Jahr 1908 seine Arbeiten in der alten, der Welt bestenfalls aus Randnotizen in der Heiligen Schrift bekannten mesopotamischen Königsstadt Assur wieder aufnahm, war noch nicht abzusehen, dass sich in den Trümmern eines eingestürzten Hauses eine jener vielzitierten archäologischen Sensationen verbergen sollte, die gleichsam Regalreihen wie Fernsehsendungen zu füllen vermögen. Zutage traten die Überreste einer Bibliothek aus divinatorischen Keilschrifttexten, die – einfache in der Sonne gehärtete und im Zerstörungsbrand ihrer Mutterstadt für 2600 Jahre konservierte Lehmklumpen als Textträger benutzend – einen Gutteil der im frühen Maßstab als wissenschaftlich zu bezeichnenden Fertigkeiten ihrer Zeit beinhaltete: eine Sensation, deren Bedeutung sich nicht nur über inhaltliche Erkenntnisse zur Befriedigung forschersicher Neugier des Archäologen erschließt, sondern auch in ihren bloßen Fundumständen zu suchen ist.

Der Motivation und Fantasie der unter der Mühsal der Entzifferung solcher großenteils noch ungelesener, in tausend Bruchstücke zerbrochenen Schrifttafeln leidenden Studienanfänger wird in den Hörsälen heutiger Universitäten bisweilen mit einem zugegebenermaßen kühnen, aber nicht minder eingängigen Gedankenspiel auf die Sprünge geholfen: Die Texte, die man aus dem Staub ihrer Zeit befreite, erlauben heute einen signifikanten Einblick in die Profession früher Gelehrter, die mithilfe von Himmels- und Erdbeobachtung, Eingeweideschau und anderer sich in ihrer eigenen Logik bewegender Empirie die Geschicke ihres Landes zu lenken versuchten. Dass sie sich noch heute in den großen Museen zu Berlin, Paris und London befinden, ist einzig und allein der Robustheit ihrer vergleichsweise recht einfachen Textträger zu verdanken. Welche Schlüsse aber zögen künftige Archäologen, wenn sie aus den Ruinen unserer urbanen Zentren nicht etwa die längst vergangenen papiernen Zeitungen und Literaturwerke, noch unsere wieder in ihre Urbestandteile zerfallenen, mit Datenmaterial angefüllten Disketten und Chips bärgen, sondern eben jene Tonklumpen grauer Vorzeit in den Händen

hielten, die in ihrer Genügsamkeit auch die Unstetigkeiten der nächsten zwei Jahrtausende überstehen dürften? Wären wir in ihren Augen dann eine Gesellschaft von Traumdeutern und Königsuntertanen, welche Entscheidungen über Feldzüge und Personalien voll und ganz auf das Verhalten von Schlangen und Vögeln oder auf die Bewegungen von Öl auf der Oberfläche von Wasser stützte, oder eben jene von Fortschritt und Vernunft getragene, zu intellektuellen und technischen Höchstleistungen fähig, für die wir uns dann und wann selbst halten? Kurzum sollte man sich in Zeiten ebenso diffiziler wie filigraner Wissensspeicherung die Frage nach der Beständigkeit unseres Erkenntnisschatzes stellen.

Grundsätzliches

Mit dem alles durchdringenden Aufkommen digitaler Methoden steht eine Gesellschaft beim Versuch, die Integrität ihrer digitalen Erzeugnisse nachhaltig zu wahren, vor besonderen Herausforderungen. Die zunehmende Entwicklung gerade geisteswissenschaftlicher Forschung zu einer durchweg empirisch gestützten Wissenschaft brachte ein Anwachsen von digitalen Forschungsdaten mit sich (Digital Humanities), das zwar die Verifizierbarkeit von Forschungsergebnissen optimiert, jedoch den Bedarf nach stabilen Lösungen für deren langfristige Verfügbarkeit jenseits rein technischer Aspekte deutlich gesteigert hat. Forschungsdaten leben hierbei meistens in einem spezifischen Biotop, einer Infrastruktur, die ihrer langfristigen Aufbewahrung und ihrer Bearbeitung dient. Welche technische Lösung des Forschungsdatenmanagements auch gewählt wird – das Projekt TextGrid bedient sich dem Konzept einer Virtuellen Forschungsumgebung (VFU)¹ – sie sieht sich gezwungen, die Kernaspekte einer nachhaltig funktionsfähigen, nutzbaren und sicheren Forschungsinfrastruktur zu adressieren und zu finanzieren. Insbesondere der Aspekt der Nachnutzbarkeit von Daten und infrastrukturellen Anwendungen zum Zwecke der Replizierbarkeit und Falsifizierbarkeit von Forschungsergebnissen erfordert nicht nur technische Flexibilität und Entwicklungspotenzial mit Blick auf sich ändernde Nutzeranforderungen und Formate, sondern auch einen nachhaltigen Betrieb rahmender Infrastrukturen, um mittelfristig das Wegbrechen dieser empirischen Basis der Forschung zu verhindern. Diese Problematik stellt die Frage nach umfassenden Lö-

¹ Zum Leistungsspektrum von Virtuellen Forschungsumgebungen vgl. Dickmann et al. (2010: 4 ff.).

sungsansätzen für Forschungsinfrastrukturen, die, auf die Komponenten einer ganzheitlichen Nachhaltigkeitsstrategie abgebildet, sämtliche Aspekte ihres dauerhaften Betriebes ansprechen.

Das Nachhaltigkeitskonzept des Projektes TextGrid ruht auf den drei Säulen der *technischen*, *fachwissenschaftlichen* und *organisatorischen* Nachhaltigkeit einer VFU (Abb. 1), die es textbasiert arbeitenden Nutzern nicht nur erlaubt, digitale Forschungsdaten zu erzeugen, sondern diese auch dauerhaft abzulegen, zu bearbeiten, kollaborativ zu analysieren, zu edieren und für kontinuierliche Forschungstätigkeiten nachzunutzen.

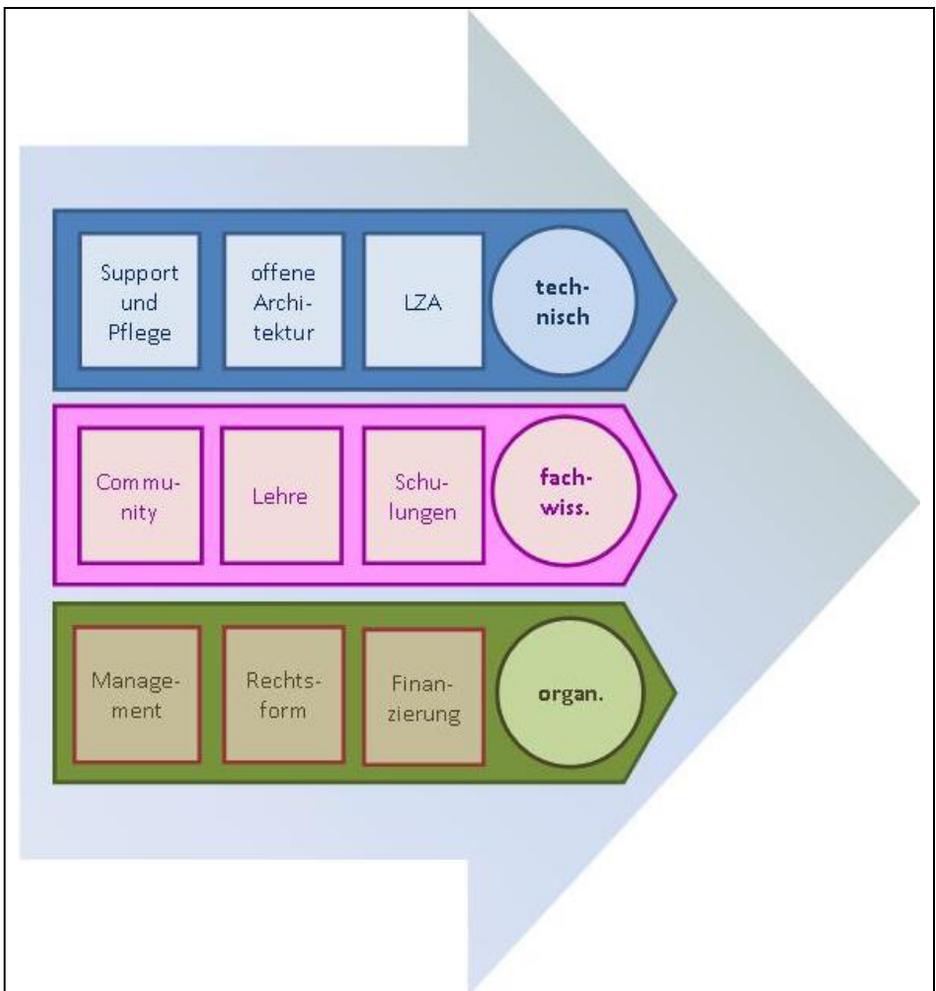


Abb. 1 Das Nachhaltigkeitskonzept in TextGrid

Um dieses Ziel zu erreichen, stellt TextGrid eine offene Architektur bereit, die den zentralen Aspekten und Herausforderungen eines verstetigten Betriebs gerecht wird. Hierbei gilt es, bestimmte Faktoren zu adressieren, die sich in dem skizzierten Spannungsfeld des dreistufigen Nachhaltigkeitsansatzes bewegen. Während sich die technischen Aspekte der Nachhaltigkeit, die in diesem Papier nicht zur Sprache kommen sollen, vornehmlich auf Fragen einer stabilen Datenkuration, der Instandhaltung und Sicherheit der technischen Infrastruktur sowie der Integration neuester technischer Entwicklungen und heterogener Nutzeranforderungen ausdehnen (vgl. Dickmann et al. 2010: 8 ff.), zielen Strategien der fachwissenschaftlichen Verankerung von Virtuellen Forschungsumgebungen zunächst auf den Wissenschaftler selbst: Wissenschaftliche Nutzer müssen für den Mehrwert eines solchen, auf digitale Methoden gestützten, Ansatzes sensibilisiert und im Umgang mit ihnen unterstützt werden. Die Ausbildung des wissenschaftlichen Nachwuchses sollte demnach von Beginn an die Nutzung solcher Werkzeuge mitdenken und entsprechende Lehreinheiten in die Curricula aufnehmen (vgl. Dickmann et al. 2011: 27 f.). Das Konzept der organisatorischen Nachhaltigkeit hat hingegen die Aufrichtung geeigneter Rahmenbedingungen für einen reibungslosen wissenschaftlichen und technischen Betrieb der VFU TextGrid im Sinne. Hierzu zählen neben der üblichen administrativen und kommunikativen Arbeit, die im Rahmen dieser Ausführungen nicht näher betrachtet werden sollen, vor allem eine geeignete Rechtsform sowie eine gesicherte Finanzierung von Aufbau, laufendem Betrieb und weiterer Entwicklung eines solchen Unterfangens.

Virtuelle Forschungsumgebungen aufbauen

Virtuelle Forschungsumgebungen zu konzipieren und aufzubauen, ist eine Herausforderung, die verschiedenste Faktoren und Eigenarten dieses Lösungsansatzes zu berücksichtigen hat. So sind sie zunächst standortübergreifend und disziplin- bzw. methodenspezifisch zu entwerfen.² Es steht somit zu erwarten, dass nicht nur eine technische Lösung für eine Vielzahl von Forschungsrichtungen entstehen wird. Um die Flexibilität, Anpassungsfähigkeit und das Weiterentwicklungspotenzial einer VFU jedoch nicht zu überdehnen, empfiehlt sich ein gewisses Maß an Spezialisierung auf disziplinergemeinsame

² Zur Heterogenität disziplinspezifischer Anforderungen vgl. Dickmann et al. (2011: 8 ff.).

Anforderungen bzw. auf ein übergreifendes Methodenspektrum (vgl. Dickmann et al. 2010: 6). Vor diesem Hintergrund spricht TextGrid nicht nur primär Geistes- und Kulturwissenschaftler an, es schränkt sein Repertoire vielmehr auf die Erstellung von textbasierten Forschungsdaten und deren Edition ein. Bei der Findung geeigneter Lösungen hat sich die Kooperation von Fachwissenschaftlern und Informatikern bewährt, um technisches Know-how mit Methodenwissen zu verschränken.

Die Finanzierung der Aufbauphase dürfte sich i. d. R. im Rahmen eines projektbasierten Anschubes gestalten. Erfahrungsgemäß bewegen sich solche Mittelbewilligungen durch die Öffentliche Hand – eine Finanzierung sollte anhand von Zielen und Partnern innerhalb der VFU genau abwägen, ob eine Bundes- oder Länderförderung erfolgen soll – in einem Zeitrahmen von bis zu sechs Jahren. Die erste Konzeption des Leistungsspektrums einer jeden VFU sollte dieses Zeitfenster vor Augen haben, sodass sie nach dem Verstreichen der initialen Förderung einen Stand erreicht hat, der eine Überführung in den Dauerbetrieb ermöglicht (vgl. ebd.: 14 ff.). Nichtsdestoweniger stellt sich die Frage, inwieweit Förderer künftig noch weitere individuelle Lösungen – auch in Gestalt von Virtuellen Forschungsumgebungen – unterstützen, oder aber die synergetische Nutzung und Adaption bestehender Angebote empfehlen werden. Die Einbeziehung der Nutzung von Virtuellen Forschungsumgebungen bei künftigen Projektanträgen erfordert jedoch eine entsprechende finanzielle Berücksichtigung innerhalb der Projektanträge (vgl. ebd.: 29) und geeignete Steuerungsmaßnahmen in der Förderpolitik (s. u.) (vgl. ebd.: 18 ff.).

Virtuelle Forschungsumgebungen finanzieren

Die großen Geldgeber des Bundes und der Länder nehmen das Kriterium der Verstetigung beantragter Forschungsvorhaben zusehends in die Formalitäten ihrer Ausschreibungen auf – eine Mindestvorhaldedauer von Daten und Werkzeugen von zehn Jahren gilt bei der Deutschen Forschungsgemeinschaft (vgl. DFG 2013: 21). Die Frage nach der Finanzierbarkeit der Aufbewahrung und Pflege immer größer werdender digitaler Datenarchive stellt sich auch jenseits gewöhnlicher Projektlaufzeiten insbesondere kleineren Disziplinen immer dringlicher.³ Es wird hierbei zusehends deutlicher, dass eine kurzfris-

3 Zum Verlauf der Diskussion vgl. etwa Alliance for Permanent Access (2008: 2 ff.).

tige, initial ausgelegte Finanzierung nicht mehr ausreicht, um die langfristige Aufgabe der Bestandhaltung zur Verhinderung von Erkenntnisverlust und Vermeidung kostspieliger, redundanter Neuerhebungen zu stemmen (vgl. Beucke 2010: 26). Der Übergang von einer Einzelförderung in eine institutionelle Förderung ist damit unumgänglich (vgl. Dickmann et al. 2010: 23; 26 f.).

Sowohl Förderer als auch künftige Antragssteller, die eine VFU zu nutzen beabsichtigen, sind folglich an einer genauen Aufschlüsselung potenziell entstehender Kosten interessiert, um die Förderpolitik entsprechend anpassen und Anträge auf Forschungsförderung glaubhaft ausformulieren zu können (vgl. Fiedler et al. 2014: 4 f.). Hierbei wägen Förderer den Mehrwert einer Nutzung gegen die zu erwartenden Kosten ab. Vage oder instabile Kostenpläne seitens der wissenschaftlichen Nutzer sind einer Förderung abträglich. Perspektivisch kommen Antragsteller also nicht um die genaue Einschätzung der Kosten der Nachhaltigkeit herum, wollen sie Fördergelder in seriös beziffertem Umfang einwerben (vgl. DFG 2013: 21 f.; Jensen 2012: 18; Rathmann 2013: 6).

Auf der Grundlage der im Projekt *Radieschen* (vgl. Rathmann 2013: 9) durchgeführten Kalkulationen, die auf einer Interviewreihe zentraler Institutionen zur Kostenentwicklung im Forschungsdatenmanagement fußen – Langzeitarchivierung (LZA) steht zumeist im Fokus derartiger Untersuchungen (vgl. Beargie et al. 2008: 24 ff.; California Digital Library 2014: 5 ff.), während Virtuelle Forschungsumgebungen aufgrund der ausgeprägten Heterogenität ihrer Ausrichtung kaum betrachtet werden (vgl. KII 2011: 34) –, wurde eine induktive Methode angewandt, um von einzelnen Arbeitsschritten zu einer Gesamtkostenaufstellung zu gelangen.

Für die in TextGrid vorgenommene Kostenkalkulation steht in der Folge die Betriebsphase der VFU im Mittelpunkt der Betrachtungen, von der, wie erwähnt, nicht zu erwarten steht, dass sie durchgängig projektbasiert gefördert wird. Vielmehr hat sich eine strikte Trennung von Aufbau- und Betriebsphase bei der Kostenabschätzung nicht bewährt: Die bloße Aufrechterhaltung des Status quo im Repertoire von Forschungsinfrastrukturen im Sinne eines fertigen Produkts ohne Anpassungen an dynamisch geänderte Technologien oder Bedarfe vermag ihr nachhaltiges Bestehen nicht zu sichern (vgl. Fiedler et al. 2014: 12).

Zur vollständigen Kalkulation aller Kosten wurden den Ansätzen in *Radieschen* folgend einzelne *Kostenarten* (vgl. California Digital Library 2014: 6 ff.) identifiziert, wie sie in Form einzelner Dienste und Werkzeuge

vorliegen, die in einen Dauerbetrieb überführt werden sollen. Auf diese Weise konnte eine Momentaufnahme der in TextGrid entstehenden laufenden Kosten entstehen, die (Weiter-) Entwicklungskosten zwangsläufig ausblenden musste. Da sich solche zusätzlichen Aufwände modular zu den kontinuierlich anfallenden Betriebskosten verhalten und perspektivisch bei einer erfolgreichen Integration neuer Komponenten ggf. zu einer Vergrößerung des ständigen Etats führen, stellt sich umso mehr die Notwendigkeit der Konzeption einer offenen Architektur auch für die Kostenplanung, im Rahmen derer Basisdienste und Werkzeuge flexibel aufzunehmen sind (vgl. Fiedler et al. 2014: 12 ff.).

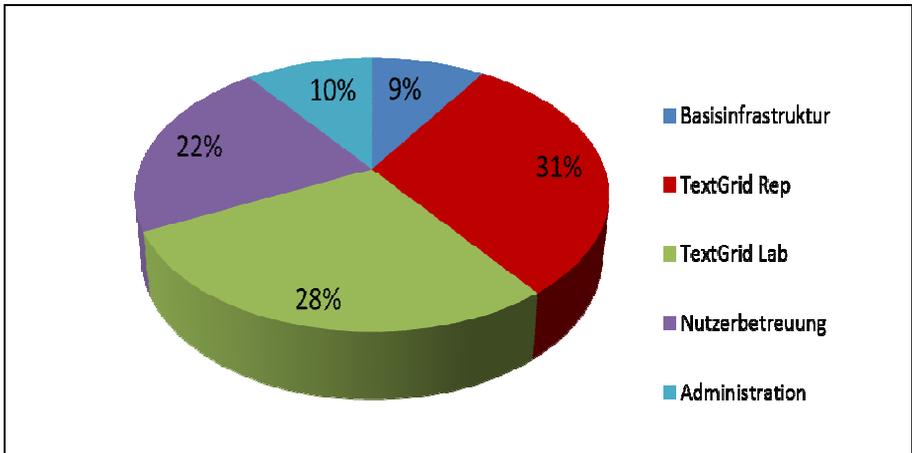


Abb. 2 Kostenverteilung in TextGrid (Stand: Dezember 2013)

Bei den durchgeführten Berechnungen überraschte es kaum, dass die Aufwände für Mitarbeiter der VFU den Löwenanteil der Gesamtkosten ausmachen: 85 % aller Mittel des laufenden Betriebs fließen in Personalkosten (vgl. Beucke 2010: 30). Bei einem Gesamtumfang von zwölf Vollzeit-äquivalenten pro Jahr entfallen im Rahmen der Kostenverteilung nach Kostenarten (vgl. Abb. 2) je ein Drittel der aufzubringenden Mittel auf die *Langzeitarchivierung* der Forschungsdaten (Pflege und Support für das TextGrid Rep) sowie Pflege und Support der *Werkzeuge und Dienste* (TextGrid Lab), während sich das letzte Drittel einerseits auf *Nutzerschulungen und -betreuung* (fachwissenschaftliche Nachhaltigkeit) zu fast einem Viertel der Gesamtkosten und andererseits den Erhalt der *Basisinfrastruktur* (Server, Sicherheit, Storage, Persistent Identifiers etc.) bzw. den *administrativen Betrieb*

(Projektmanagement, Öffentlichkeitsarbeit, Rechtsform, Dienstreisen etc.) verteilt.

Jedes Finanzierungsmodell muss der modularen Struktur der technischen, organisatorischen und fachwissenschaftlichen Komponenten der VFU TextGrid Rechnung tragen. Zur Aufrechterhaltung der zur Fundierung des laufenden Dauerbetriebs notwendigen Basisinfrastruktur wird es unumgänglich sein, eine institutionelle Finanzierung sicherzustellen. Da dies nicht allein durch eine dedizierte Dauerfinanzierung seitens Bund und Ländern zu erreichen sein dürfte, sind potente Schlüsselakteure mit einschlägigen Fertigkeiten zu identifizieren und auf Grundlage eines gemeinsamen organisatorischen Rahmenwerks für die Betreuung einzelner Leistungselemente zu gewinnen. Dies kann nur gelingen, wenn die Ansätze der technischen, fachwissenschaftlichen und organisatorischen Nachhaltigkeitsstrategie ineinandergreifen und potenziellen Partnern ein überzeugendes Gesamtkonzept vorgestellt werden kann, das nicht nur angepasste und zeitgemäße technische Lösungen bietet, sondern auch deren wissenschaftlichen Mehrwert vermitteln kann und rechtlich stabile organisatorische Rahmenbedingungen schafft. Erfahrungsgemäß werden die Mitglieder der wissenschaftlichen Communitys, die das Portfolio der VFU um eigene Entwicklungen zu ergänzen oder lediglich vorhandene Dienste und Werkzeuge zu nutzen gedenken, kaum durch ein umlagefinanziertes Gebührenmodell eingebunden werden können. Vielmehr muss sich die Förderlandschaft dahingehend anpassen, als wohlbegründete, durch einen geänderten fachwissenschaftlichen Bedarf oder eine fortschreitende technische Dynamik motivierte Neu- und Weiterentwicklungen in die Strukturen existierender technischer Lösungen wie TextGrid integriert und zu diesem Zweck entsprechende Mittel bereitgestellt werden. Es liegt demnach nahe, dass Drittmittelausschreibungen in ihrem Finanzrahmen, etwa in Gestalt eventueller Overhead-Pauschalen, die Nutzung Virtueller Forschungs-umgebungen von Beginn an berücksichtigen. Mit der Aufstellung eines Kostenrahmens trägt TextGrid zur Einschätzung der notwendigen Investitionen auf Fördererseite bei.

Virtuelle Forschungsumgebungen verwalten

TextGrid hat seit Beginn seiner Förderung das Projektende im Blick gehabt. Die VFU sollte nicht das Schicksal vieler Projekte teilen, die nach Auslaufen ihrer Förderung zwar für die wissenschaftliche Nutzung freigegeben, aber

nicht weiter gepflegt werden und somit schnell nicht mehr nutzbar sind. Zur Sicherung des Fortlebens von TextGrid und der Arbeit der maßgeblichen Träger und Entwickler des Projekts wurde nicht nur eine aktive und lebendige Community geschaffen, die ihren Fortbestand durch beständige Heranziehung wissenschaftlichen Nachwuchses garantiert, sondern auch ein rechtlicher Rahmen aufgerichtet, der die künftige Plattform des gemeinsamen Handelns bilden und sämtliche Funktionalitäten des bisherigen Projektkonsortiums übernehmen kann.

Derzeit (2014) befinden sich verschiedene Institutionalisierungslösungen in der Diskussion. Bis eine endgültige Entscheidung getroffen wird, hat sich das Projekt entschieden, vorerst in der Rechtsform eines eingetragenen, nichtwirtschaftlichen Vereins (e.V.) aufzutreten, der dem Konsortium als Verstetigungsoption zur Verfügung steht.⁴ Die Entscheidung, sich auf einen e.V. festzulegen, obgleich an anderer Stelle für eine gemeinnützige GmbH⁵ plädiert wurde oder sich mit den Infrastrukturprojekten CLARIN-D und DARIAH-DE Alternativen in Gestalt einer länderübergreifenden Rechtsform des European Research Infrastructure Consortium (ERIC) (vgl. European Commission 2010) anbieten, ist anhand einiger zentraler Kriterien (vgl. Fiedler/ Witt 2014: 5 ff.) gefällt worden, die für den nachhaltigen Betrieb der VFU unter organisatorischen Aspekten ausschlaggebend waren:

- Die Möglichkeit, die Rechtsform als *gemeinnützig* anerkennen zu lassen, war grundlegende Voraussetzung für ihre Akzeptanz, da sie es dem Verein erlaubt, für die VFU auch künftig als Empfänger von Drittmittelförderung agieren zu können.
- TextGrid wird seinen Nutzern weiterhin ein breites Spektrum an Diensten und Werkzeugen zur Verfügung stellen, welche perspektivisch deren aus Forschungsdaten bestehende empirische Grundlage verwalten und verwahren. Dies macht es notwendig, sämtliche Arbeitsabläufe zwischen Nutzern und Diensteanbietern rechtlich abzusichern. Für die Nutzung sämtlicher TextGrid-Ressourcen wurde eine verbindliche Nutzungsordnung erlassen, die das Projekt unter der Ausnahme fahrlässigen oder absichtlichen Fehlverhaltens von jeglichen Regressansprüchen befreit. Um schwer einzuschätzende Finanzrisiken auszuschließen, wurden ferner be-

4 <http://www.textgrid-verein.de/>.

5 Vgl. Dickmann et al. (2010: 24 f.); PricewaterhouseCoopers Legal (2010: 1); Dickmann et al. (2011: 28).

schränkte *Haftungsbedingungen* eruiert, die Rechtsansprüche ausschließlich auf das Körperschaftvermögen reduzieren.

- Die *Mitgliedschaft* in der Körperschaft sollte sich möglichst einfach gestalten und mit geringem finanziellem Aufwand verbunden sein – die Rechtsform ERIC sieht beispielsweise den Beitritt einzelner Mitgliedsstaaten der Europäischen Union vor. Trotz der angestrebten Transparenz der Entscheidungswege auch innerhalb der Gremien sollte ein Höchstmaß an Eigenständigkeit in der Bestimmung der Arbeitsstrategie und der Aufnahmekriterien für potenzielle Partner gewährleistet werden.
- Mindesteinsätze an *Kapitaleinlagen* sollten möglichst ganz vermieden werden, da sich TextGrid definitionsgemäß nicht auf kommerzielle Aktivitäten einlassen kann und nicht alle Finanzierungsmodelle eine hinreichende Kapitalausstattung zu garantieren vermögen.
- Mit Blick auf die zur Disposition stehenden Ertragsmodelle zur Finanzierung des laufenden Betriebes sollten sich Kosten und administrativer Aufwand während des Betriebs und ggf. auch bei der Auflösung der Körperschaft – angesichts geringer Budgets und zeitlich sehr eingespannten wissenschaftlichen Personals ohne juristischen Hintergrund – in Grenzen halten. Das zuvor erwähnte Kriterium der Gemeinnützigkeit hat darüber hinaus entscheidenden Einfluss auf den Wegfall von Körperschafts- und Gewerbesteuer und ermöglicht die Einnahme von Spendengeldern, die zweckgebunden dem Betrieb der Körperschaft zugefügt werden können. Unternehmerische Aspekte sind daher ein eher zu meidendes Kriterium bei der Wahl der Rechtsform.

Der TextGrid-Verein konnte alle in ihn gesetzten Erwartungen, wie sie sich anhand der Auswahlkriterien ergaben, erfüllen. Bis zur Übergabe der Verantwortung vom Projektkonsortium an den Verein, wodurch sämtliche für das alltägliche Funktionieren der VFU im Realbetrieb als unabdingbar identifizierten Fähigkeiten auf die neue Körperschaft übergehen werden, werden sich beide in ihrer Arbeit und Funktion ergänzen. Mit dieser Zielsetzung konnten bereits 17 akademische Institutionen angesprochen und als Unterstützer gewonnen werden, die sämtlich als einflussreiche Akteure die Digital Humanities im deutschsprachigen Raum gestalten. Gegen Ende des Projekts wird eine Übergangsphase anzusetzen sein, im Rahmen derer Zuständigkeiten und Aufgaben zugewiesen werden (vgl. ebd.: 10 ff.).

Indes ist der Verein trotz recht breiter Schultern auf ein flankierendes Finanzierungskonzept (s. o.) angewiesen, da ein stabiler Betrieb nur aus Mitgliedschaftsbeiträgen, wie sie der Verein inzwischen ermöglicht hat, nicht

auf Dauer zu gewährleisten sein wird. Weitere Bedarfe (vgl. ebd.: 14 ff.) zur Rüstung der Körperschaft als Trägerinstanz einer VFU haben sich aus dem TextGrid-eigenen Nachhaltigkeitskonzept ergeben. Zwar sind die wesentlichen Gremien im Rahmen der Vereinssatzung ermöglicht worden, um bisherige Entscheidungsinstanzen abzubilden und ein durchweg demokratisches Gefüge zu gewährleisten. Dennoch empfiehlt es sich, außerhalb der hoheitlichen Gremienstruktur Funktionseinheiten anzusiedeln, welche sowohl den technischen und administrativen Grundbetrieb (operative Einheiten) als auch die fachwissenschaftliche Verankerung ermöglichen. So ist die Einbindung von Disziplinen und Fachwissenschaftlern innerhalb institutionalisierter Fachgruppen ebenso erforderlich, um Bedarfe und Methoden der VFU zu formulieren, wie die Etablierung dedizierter Arbeitsgruppen, welche sich der Lösung bestimmter, aus den Nutzerbedarfen hergeleiteter Aufgaben widmen (s. Abb. 3).

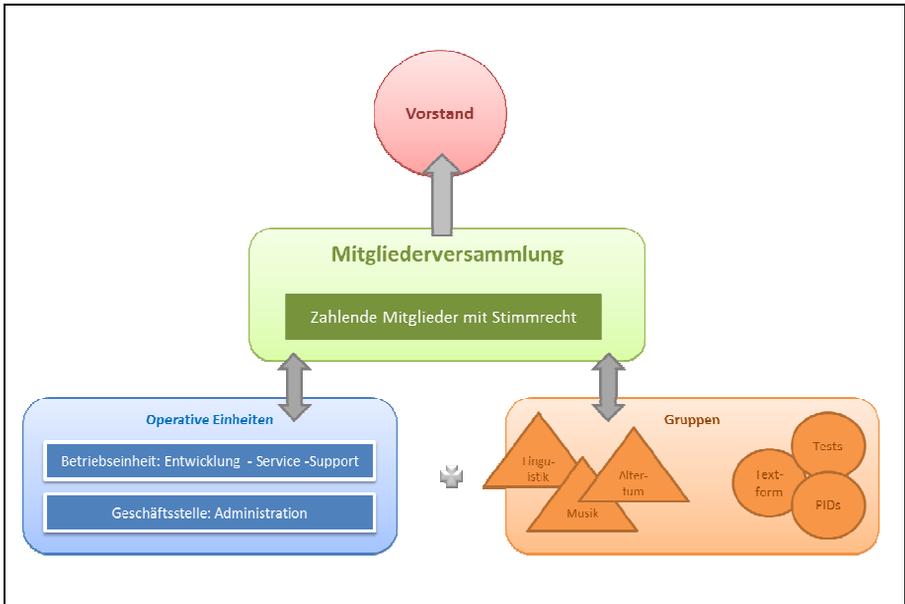


Abb. 3 Umsetzung der Nachhaltigkeitsstrategie am Beispiel des Vereins

Gerade mit Blick auf die Bündelung fachwissenschaftlichen Wissens und informatischer Kompetenz in einer institutionalisierten Kommunikationsplattform – ein Ansatz, der die Dreiheit des hier beschriebenen Nachhaltigkeitsgedankens aus technischer, fachwissenschaftlicher und organisatorisch/finanzieller Stoßrichtung materialisiert – wird in Zukunft weiterhin Ge-

sprächsbedarf darüber bestehen, welche rechtlichen Rahmenbedingungen hierbei ein Optimum aus Transparenz und Stabilität zu erzeugen vermögen. Insbesondere wird darüber zu reden sein, inwieweit die jetzt schon sehr fruchtbringende Zusammenarbeit der großen nationalen Verbände der Digital Humanities – DARIAH-DE, CLARIN-D und TextGrid – weiter vertieft und deren organisatorische wie rechtliche Spezifika mitgenutzt werden können. Denkbar wäre gerade mit Blick auf die Etablierung zusätzlicher funktionaler Einheiten für die weitere Entwicklung eine verstärkte Integration dieser Hauptakteure. Die synergetische Nutzung spezieller Fertigkeiten eines Projektes durch alle Fach-Communitys wird zur Vermeidung von Redundanzen ein erster entscheidender Schritt sein.

Ausblick

Um den Bedarf der wissenschaftlichen Community sinnvoll und vollständig abdecken zu können, setzt TextGrid auf ein modulares Betriebskonzept mit autonomen Servicekomponenten, das der Philosophie des Projekts hinsichtlich technischer, fachwissenschaftlicher und organisatorischer Nachhaltigkeit einer VFU entspricht. TextGrid bietet damit eine flexible und offene Struktur, die – den nachhaltigen Bestand aller Komponenten vorausgesetzt – nicht nur den Anforderungen von Nutzern aus dem Bereich der Digital Humanities an ein modernes Forschungsdatenmanagement entgegenkommt, sondern den Wissenschaftlern auch eine organisatorisch stabile Umgebung vorhält und sie mit den teils komplexen Herausforderungen im Umgang mit Forschungsdaten nicht allein lässt.

Dieses von Beginn an konsequent beibehaltene und umgesetzte Vorgehen hat sich in der Praxis als robust und handhabbar erwiesen. Es gewährleistet vor allem die notwendige Durchlässigkeit, um im Spannungsfeld der Agilität von technischem Potenzial und wissenschaftlicher Methode zu bestehen, und vermeidet bewusst jede Form der Starre, welche ein Schritthalten jeglicher Aspekte der VFU TextGrid mit der Entwicklung moderner Forschung behindern könnte. Die Grundkonstante dieses Konzepts ist steter Wandel, Flexibilität und Anpassung. Wenn demnach auch nicht zu erwarten steht, dass TextGrid nach Verstreichen der eingangs erwähnten Mindestvorhaltefrist von zehn Jahren, wie sie die Förderer voraussetzen, noch dasselbe Gesicht zeigen wird, wie es heute der Fall ist, so stehen die Chancen gut, dass es auch

darüber hinaus dank dieser erlernten Geschmeidigkeit seinen Beitrag zur Wahrung des Gedächtnisses der Gegenwart leisten wird.

Literaturverzeichnis

- Alliance of Permanent Access (2008): Keeping the records of science accessible: can we afford it? Report on the 2008 Annual Conference of the Alliance for Permanent Access, Budapest, 4 November 2008, http://www.alliancepermanent-access.org/wp-content/uploads/2010/12/documenten_Alliance2008conference_report.pdf.
- Beargie, Neil et al. (2008): Keeping Research Data Safe. A cost model and guidance for UK universities, <http://www.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf>.
- Beucke, Daniel (2010): *Geschäftsmodelle für die digitale Langzeitarchivierung. Das Beispiel Forschungsdaten* (Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft; 278), <http://edoc.hu-berlin.de/series/berliner-handreichungen/2010-278/PDF/278.pdf>.
- California Digital Library (2014): Total Cost of Preservation (TCP). Cost and Price Modeling for Sustainable Services, https://confluence.ucop.edu/download/attachments/163610649/TCP-cost-price-modeling-for-sustainable-services-v2_2_1.pdf?version=1&modificationDate=1397258795000.
- DFG (2013): Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“. Vorschläge zur Sicherung guter wissenschaftlicher Praxis. Ergänzte und aktualisierte Empfehlungen, http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf.
- Dickmann, Frank et al. (2010): Konzept eines Betriebsmodells für Virtuelle Forschungsumgebungen. Zweck, Form und Inhalt eines nachhaltigen Betriebs Virtueller Forschungsumgebungen, http://www.wissgrid.de/publikationen/deliverables/wp1/WissGrid_API_D1_D1-4_final_v2.pdf.
- Dickmann, Frank et al. (2011): Aspekte der Implementierung eines Betriebsmodells für Virtuelle Forschungsumgebungen, http://www.wissgrid.de/publikationen/deliverables/wp1/WissGrid_API_D15_v101_final.pdf.
- European Commission (2010): Legal framework for a European Research Infrastructure Consortium – ERIC, http://ec.europa.eu/research/infrastructures/pdf/eric_en.pdf.
- Fiedler, Norman; Witt, Andreas (2014): Erfahrungsbericht Rechtsform: Praxisbewährung und Nutzeranforderungen (R 1.1.2), <http://www.textgrid.de/file-admin/berichte-2/report-1-1-2.pdf>.

- Fiedler, Norman et al. (2014): TextGrid Report 1.1.1. Identifikation von Kostenfaktoren und Erarbeitung von Kostenmodellen, https://www.textgrid.de/fileadmin/user_upload/TG_R111_final.pdf.
- Jensen, Uwe (2012): Leitlinien zum Management von Forschungsdaten. Sozialwissenschaftliche Umfragedaten, http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2012/TechnicalReport_2012-07.pdf.
- KII (2011): Gesamtkonzept für die Informationsinfrastruktur in Deutschland. Empfehlungen der Kommission Zukunft der Informationsinfrastruktur im Auftrag der Gemeinsamen Wissenschaftskonferenz des Bundes und der Länder, http://www.allianzinitiative.de/fileadmin/user_upload/KII_Gesamtkonzept.pdf.
- PricewaterhouseCoopers Legal (2010), <http://www.d-grid-gmbh.de/fileadmin/downloads/Grid-Empfehlung-Rechtsform.pdf>.
- Rathmann, Torsten (2013): Projekt RADIESCHEN. Rahmenbedingungen einer disziplinübergreifenden Forschungsdateninfrastruktur. Preise, Kosten und Domänen. Entspricht dem Report D4.3 „LZA-Kostenstruktur“ nach Projektantrag, http://gfzpublic.gfz-potdam.de/pubman/item/escidoc:117052:2/component/escidoc:117199/ProjektRadieschen_PreiseKosten_und_Domaenen.pdf.

Anhang

Verzeichnis der Autorinnen und Autoren

| Name | Institution |
|--|--|
| Aschenbrenner, Andreas | Berlin-Brandenburgische Akademie der Wissenschaften, Berlin |
| Betz, Katrin | Universität Würzburg – Institut für Deutsche Philologie |
| Blümm, Mirjam, Dr. | Niedersächsische Staats- und Universitätsbibliothek Göttingen |
| Brodhun, Maximilian | Niedersächsische Staats- und Universitätsbibliothek Göttingen |
| Casties, Robert, Dr. | Max-Planck-Institut für Wissenschaftsgeschichte, Berlin |
| de la Iglesia, Martin | Niedersächsische Staats- und Universitätsbibliothek Göttingen |
| Eichinger, Ludwig M., Prof. Dr. Dr. h.c. mult. | Institut für Deutsche Sprache, Mannheim |
| Mittler, Elmar, Prof. em. Dr. Dr. h.c. mult. | Niedersächsische Staats- und Universitätsbibliothek Göttingen |
| Fiedler, Norman, Dr. | Institut für Deutsche Sprache, Mannheim |
| Figgen, Larissa | Westfälische Wilhelms-Universität, Münster |
| Funk, Stefan E. | DAASI International GmbH, Tübingen; Niedersächsische Staats- und Universitätsbibliothek Göttingen |
| Gietz, Peter | DAASI International GmbH, Tübingen |
| Göbel, Mathias | Niedersächsische Staats- und Universitätsbibliothek Göttingen |
| Goormann, Lena | Westfälische Wilhelms-Universität, Münster |
| Grupe, Nadja | Niedersächsische Staats- und Universitätsbibliothek Göttingen |
| Huck, Johannes | Westfälische Wilhelms-Universität, Münster |
| Jannidis, Fotis, Prof. Dr. | Universität Würzburg – Institut für Deutsche Philologie |

| | |
|---------------------------------|---|
| Kaden, Ben | Technische Universität Berlin – Zentrum Technik und Gesellschaft |
| Kerzel, Martina | Niedersächsische Staats- und Universitätsbibliothek Göttingen |
| Kollatz, Thomas | Salomon Ludwig Steinheim-Institut für deutsch-jüdische Geschichte, Universität Duisburg-Essen |
| Krause, Celia | Technische Universität Darmstadt – Institut für Sprach- und Literaturwissenschaft |
| Küster, Marc Wilhelm, Prof. Dr. | Fachhochschule Worms |
| Lemitz, Bastian | Westfälische Wilhelms-Universität, Münster |
| Liebethuth, Martin | Niedersächsische Staats- und Universitätsbibliothek Göttingen |
| Meier, Lars-Steffen | Westfälische Wilhelms-Universität, Münster |
| Moretto, Nicolas | Niedersächsische Staats- und Universitätsbibliothek Göttingen |
| Neuroth, Heike, Dr. | Niedersächsische Staats- und Universitätsbibliothek Göttingen |
| Pempe, Wolfgang | Deutsches Forschungsnetzwerk, Stuttgart |
| Prager, Christian M., Dr. | Rheinische Friedrich-Wilhelms-Universität Bonn – Institut für Archäologie und Kulturanthropologie |
| Quennet-Thielen, Cornelia | Bundesministerium für Bildung und Forschung, Bonn |
| Radecke, Gabriele, Dr. | Georg-August-Universität Göttingen – Seminar für Deutsche Philologie, Theodor Fontane-Arbeitsstelle |
| Rapp, Andrea, Prof. Dr. | Technische Universität Darmstadt – Institut für Sprach- und Literaturwissenschaft |
| Rettelbach, Simon, Dr. | Deutsches Institut für Internationale Pädagogische Forschung – Frankfurter Forschungsbibliothek |
| Rieger, Simone | Max-Planck-Institut für Wissenschaftsgeschichte, Berlin |
| Schmid, Oliver, Dr. | Technische Universität Darmstadt – Institut für Sprach- und Literaturwissenschaft |

| | |
|--------------------------|--|
| Schmunk, Stefan, Dr. | Niedersächsische Staats- und Universitätsbibliothek Göttingen |
| Söntgerath, Olga | Westfälische Wilhelms-Universität, Münster |
| Söring, Sibylle | Niedersächsische Staats- und Universitätsbibliothek Göttingen |
| Stallmann, Marco | Westfälische Wilhelms-Universität, Münster |
| Vanscheidt, Philipp | Technische Universität Darmstadt – Institut für Sprach- und Literaturwissenschaft |
| Veentjer, Ubbo | Niedersächsische Staats- und Universitätsbibliothek Göttingen |
| Veit, Joachim, Prof. Dr. | Universität Paderborn – Musikwissenschaftliches Seminar Detmold/Paderborn |
| Wieder, Philipp, Dr. | Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen |
| Wintergrün, Dirk | Max-Planck-Institut für Wissenschaftsgeschichte, Berlin |
| Witt, Andreas, Prof. Dr. | Institut für Deutsche Sprache, Mannheim |

Projektpartner

Partner in TextGrid I (2006–2009):

- DAASI International GmbH
- Fachhochschule Worms
- Institut für Deutsche Sprache Mannheim
- Niedersächsische Staats- und Universitätsbibliothek Göttingen
- Saphor GmbH Tübingen
- Technische Universität Darmstadt – Institut für Sprach- und Literaturwissenschaft (Germanistische Computerphilologie)
- Universität Trier
- Universität Würzburg, Institut für deutsche Philologie

Partner in TextGrid II (2009–2012):

- DAASI International GmbH
- Fachhochschule Worms
- Institut für Deutsche Sprache Mannheim
- Ludwig-Maximilians-Universität München
- Max Planck Digital Library
- Niedersächsische Staats- und Universitätsbibliothek Göttingen
- Technische Universität Kaiserslautern
- Universität Paderborn
- Universität Trier
- Universität Würzburg, Institut für deutsche Philologie, Lehrstuhl für Computerphilologie und Neuere deutsche Literatur

Partner in TextGrid III (2012–2015):

- Berlin-Brandenburgische Akademie der Wissenschaften
- DAASI International GmbH
- Fachhochschule Worms
- Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen
- Institut für Deutsche Sprache Mannheim
- Max-Planck-Institut für Wissenschaftsgeschichte
- Niedersächsische Staats- und Universitätsbibliothek Göttingen

- Technische Universität Berlin – Zentrum Technik und Gesellschaft
- Technische Universität Darmstadt – Institut für Sprach- und Literaturwissenschaft (Germanistische Computerphilologie)
- Universität Würzburg, Institut für deutsche Philologie

Kooperationsprojekte

Die nachfolgende Liste bietet eine Auswahl der uns bei Redaktionsschluss bekannten, mit TextGrid arbeitenden Projekte. Auf <http://textgrid.de/community/> finden Sie eine ständig aktualisierte Übersicht sowie die Möglichkeit, Ihr eigenes Projekt vorzustellen.

- Das Archiv jüdischer Autorinnen und Autoren in Berlin 1933–1945 als Virtuelle Forschungsumgebung, Europa-Universität Viadrina
- Bibliothek der Neologie (BdN), Westfälische Wilhelms-Universität Münster
- Johann Friedrich Blumenbach – online, Akademie der Wissenschaften zu Göttingen
- Briefwechsel Grimm-Savigny-Bang, TU Darmstadt
- Carl Louis Bargheer, „Fiedellieder“ von Theodor Storm, Musikwissenschaftliches Seminar Detmold/Paderborn
- COMPutergestützte Untersuchung von VARIabilität im KirchenSLAVischen (SlaVaComp), Universität Freiburg
- Deutsch-jüdische Publizistik des 19. Jahrhunderts, Steinheim-Institut
- Digital Medieval Projects, Stanford University
- eCodicology – Algorithmen zum automatischen Tagging mittelalterlicher Handschriften, TU Darmstadt
- Edition der Briefe von Ernst H. Kantorowicz, Deutsches Literaturarchiv Marbach
- ePoetics – Korpuserschließung und Visualisierung deutschsprachiger Poetiken (1770–1960) für den „Algorithmic Criticism“, Universität Stuttgart
- Georg Greflinger – Digitale historisch-kritische Edition, FU Berlin/Universität Oslo
- Genetisch-kritische und kommentierte Hybrid-Edition von Theodor Fontanes Notizbüchern, Universität Göttingen, Theodor Fontane-Arbeitsstelle
- Letters to Lord Nelson from the Iberian Peninsula (1803–1805), Europa-Universität Viadrina
- Online-Lexikon jüdischer Autorinnen und Autoren im nationalsozialistischen Deutschland, Europa-Universität Viadrina
- C. M. von Webers „Freischütz“ – Paradigmatische, genuin digitale Musikedition, Musikwissenschaftliches Seminar Detmold/Paderborn
- Relationen im Raum, Steinheim-Institut Essen

- Textdatenbank und Wörterbuch des Klassischen Maya, Universität Bonn
- TextGrid Shared Canvas, SUB Göttingen, MPIWG Berlin, Stanford University, CA/U.S.
- Virtuelles Skriptorium St. Matthias, KoZE Trier/TU Darmstadt
- Wissenschaftliches Netzwerk: Von Kanaan nach Gibraltar und zurück – die Phönizier im Mittelmeerraum, TU Darmstadt

Meilensteine und Reports

Meilensteine

TextGrid I (2006–2009)

- M 1.1 Text Processing
- M 1.2 Linking
- M 1.3 Text Retrieval
- M 1.4 Publishing
- M 1.5 Management von Workflow, Access, Kommunikation und Nutzer
- M 1.6 Ontologien
- M 2.1 Basiswerkzeuge: Tokenizer, Sortieren, XML Editor
- M 2.2 Textverarbeitung: Streaming-Editor, Lemmatisierung, Metadaten-Annotation
- M 2.3 Editor-Erweiterung, Workbench, Text Retrieval
- M 2.4 Query Interface, Kollationierung
- M 2.5 Druck, OCR, Integration aller Tools in Workbench
- M 3.1 Ermittlung der Middleware Anforderung
- M 3.2 Implementierung eines Prototyps der TextGrid-Middlewareplattform mit Anbindung an die IP
- M 3.3 Unterstützung und Evaluation bei der Anwendung des Prototyps für Musterapplikationen
- M 3.4 Implementierung einer Produktivversion der TextGrid-Middleware
- M 4.1 Erstellung der relevanten Corpora (Digitalisierung) als Testbed
- M 4.2 Prototypischer Durchlauf der Corpora durch die verschiedenen Tools mit Evaluation (Empfehlungen für weitere Anpassungen/Entwicklungen) Mathe + Geisteswissenschaften
- M 5.1 WB-Link zur Verknüpfung von Wörterbüchern
- M 5.2 WB-Link zur Verknüpfung von Wörterbüchern und Primärquellen unter Verwendung von GermaNet
- M 5.3 Meta-Lemmalisten
- M 5.4 Lemmatisierung von Belegstellen-Zitaten im Verbund mhd. Wörterbücher und im Deutschen Wörterbuch

- M 5.5 Erstellung von Umkehrwörterbüchern
- M 5.6 Entwicklung geeigneter Metadata Application Profiles
- M 6.1 Freischaltung der Projekthomepage
- M 6.2 Newsletter Version 1.0

TextGrid II (2009–2012)

- M 1.1.1 Übersetzung eSciDoc PubMan sowie weiterer Services in TextGrid-Release 1.5
- M 1.1.2 Übersetzung von TextGrid-Services in eSciDoc, insbesondere Kollationierer
- M 1.1.3 Adaptoren TextGrid ↔ eSciDoc
- M 1.2.2 Prototypischer Zugriff auf TextGrid-Services über ein Portal
- M 1.3.1 Auswahl und Dokumentation einer Testmethodik für Usability-Tests
- M 1.3.2 Anpassung der Usability-Evaluierungen anhand Release 1.5 und neuer Werkzeuge
- M 1.4.1 Aufbau eines Entwicklerportals
- M 1.4.2 Aufbau einer Umgebung zum öffentlichen Test von Release Candidates
- M 1.5.1 TextGridLab 1.5
- M 1.5.2 TextGridLab 1.9
- M 1.5.3 TextGridLab 2.0
- M 2.2.1 Integration TextGrid/Fedora
- M 2.2.2 AAI und Rechtemanagement
- M 2.2.3 Nachhaltige Repositorien-Föderation
- M 2.3.1 Langzeitarchivierungsdienste
- M 3.4.1 Konsultationen mit geisteswissenschaftlichen Entscheidungsträgern, Empfehlungen zur finanziellen Nachhaltigkeit für den Betrieb einer eHumanities-Infrastruktur in Deutschland
- M 3.6.1 Implementierung des überarbeiteten TextGrid-Geschäftsmodells (Ergänzung zum gemeinsamen GM), Gründung einer Organisation zur nachhaltigen Umsetzung des Geschäftsmodells
- M 4.1.1 Erstellung des TextGrid-Portals
- M 4.1.2 Etablierung organisatorischer Strukturen, Bestimmung von Ansprechpartnern, Klärung von Verantwortlichkeiten; Open-Source-Projekt:

Definition eines Releaseplanes sowie eines Schemas zur Beschreibung der Komponenten und ihrer Versionen, Diskussion eines Community Models

- M 4.2.1 TextGridVZ-Grundfunktionalitäten im TextGrid-Portal (Profil, FOAF, Foren)
- M 4.2.2 Ausbau, Anschluss an die TextGrid-Authentifizierungs- und Autorisierungsinfrastruktur und Produktivbetrieb TextGridVZ
- M 4.3.1 Erstellung und laufende Pflege einer Liste mit fachwissenschaftlich relevantem Content und Priorisierung in Rücksprache mit der geisteswissenschaftlichen Community
- M 4.4.1 Erstellung und laufende Aktualisierung einer Dokumentation bzw. Handreichung zu TextGrid-Funktionalitäten für geisteswissenschaftliche Forschungen aus fachwissenschaftlicher Perspektive
- M 4.5.1 Durchführung eines internationalen Tutorials und der eHumanities-Session im Rahmen der IEEE DEST 2009
- M 4.5.2 Ausrichtung eines internationalen Entwickler-Workshops zu TextGrid im Rahmen der Digital Humanities
- M 4.6.1 Laufende Dokumentation TextGrid, Erstellung von Tutorials
- M 5.1.1 Erarbeitung eines ersten Vorschlags für eine gemeinsamen Ontologie für Dienste und andere einschlägige Ressourcen
- M 5.1.2 Aktualisierung eines nun mit anderen Projekten und Communities abgestimmten Entwurfs für eine gemeinsamen Kern-Ontologie für Dienste und andere einschlägige Ressourcen sowie für die TextGridspezifischen Erweiterungen
- M 5.1.3 Publikation von technischen Regeln für die Verlinkung/Föderation von Dienstknoten basierend auf entstehenden europäischen und internationalen Normen
- M 5.1.4 Release der TextGrid-Registry auf der Basis der in Worms entwickelten Open-Source Engine Isidorus, die im Rahmen dessen angepasst wird
- M 5.1.5 Publikation von Regeln für die Organisation der Registry
- M 5.2.1 TextGrid-interne Richtlinien und Best practices
- M 5.2.2 Publikation der mit anderen D-Grid-Communities und eHumanities-Vorhaben abgestimmten Version der Richtlinien
- M 6.0 Entwickler-Koordinationstreffen

- M 6.1 Erste übergreifende Integration
- M 6.1.0 Planung und Durchführung eines internen Entwickler-Koordinationsworkshops
- M 6.1.1 Release einer ersten Version von TextGrid-Tools
- M 6.1.2 Koordination des gemeinsamen Zwischenreleases TextGridLab 1.5
- M 6.1.3 Release 1.9
- M 6.1.4 Produktionsreifes Abschlussrelease 2.0
- M 6.2 Release 1.5
- M 6.2.1 Erster Integrationsschritt
- M 6.2.2 Fertigstellung des Linkeditors Text-Bild und des Kollationierers
- M 6.2.3 Produktionsreifes Abschlussrelease der beiden Tools
- M 6.3 Release 1.9
- M 6.3.1 Webserviceintegration
- M 6.3.2 Fertigstellung der Webserviceintegration auf Metadatenebene
- M 6.3.3 Fertigstellung der gesamten Webserviceintegration
- M 6.4 Release 2.0
- M 6.4.1 Definition des Architekturmodells zur Anbindung von Digilib
- M 6.4.2 Fertigstellung der ersten integrierten Version von Digilib
- M 6.4.3 Zweite Version der Anbindung von Digilib/Integration der Repräsentationsformate
- M 6.4.4 Vollständige Integration der entwickelten Tools
- M 6.5.1 Erarbeitung eines adäquaten Auszeichnungssystems für glossierte Texte und Integration in TEI (Text Encoding Initiative)
- M 6.5.2 Weiterentwicklung des Text-Bild-Linkeditors (Image-Mapping) für die Anforderungen glossierter Texte; Erstellung des Präsentationstools
- M 6.5.3 Fertigstellung des Editionstools für glossierte Texte
- M 6.5.4 Einarbeitung von Feedback in Präsentations- und Editionstool; Produktionsreife; Integration in die TextGrid-Plattform (XML-Editor)
- M 6.6.1 Erste TextGrid-Integration und Dokumentation
- M 6.6.2 Integration des Notenanzeigetools in die TextGrid-Infrastruktur
- M 6.6.3 Fertigstellung des Anzeige- und Eingabetools
- M 6.6.4 Abschluss von Integration und Schulungen

- M 6.8.1 Erste Integration
- M 6.8.2 Release 1.5
- M 6.8.3 Release 1.9
- M 6.8.5 Release 2.0
- M 7.1.1 Erstellung und fortlaufende Modifikation der Fächermatrix in
Absprache mit den zahlreichen geisteswissenschaftlichen Disziplinen
- M 7.2.1 Dokumentation und Pflege aller relevanten Projekte, Initiativen und
Netzwerke im Bereich der eHumanities in Form standardisierter
Beschreibungen
- M 7.3.1 Aufbau und fortlaufende Erweiterung einer Clearingstelle für ein
eHumanities-Netzwerk in Deutschland
- M 8.1 (Re-)Launch der Homepage
- M 8.2 Organisation und Durchführung verschiedener Konferenzen und
(Experten-) Workshops
- M 8.3 Milestone-Ereignisse

TextGrid III (2012–2015)

- M 1.2.1 Geschäftsstelle Verein
- M 1.3.2 Blaupause zur Integration zukunftsfähiger Entwicklungen
- M 2.1.1 (Laufende) Kartierung der heterogenen Nutzercommunity
- M 2.1.2 Einrichtung eines Nutzerforums und dauerhafte Pflege
- M 2.2.2 Bereitstellung des Market Space
- M 2.2.1 Entwicklerworkshop mit externen Projektpartnern
- M 3.3.1 Nutzertreffen mit je spezifischen Schwerpunkten
- M 4.4.1 Update SADE (Dokumentation und Referenzinstanz)
- M 4.5.1 Architekturskizze eines produktiven Regelbetriebs
- M 5.1.1 Technische Dokumentation
- M 5.2.1 Laufende funktionale Tests
- M 5.3.1 Usability, Zusammenfassung der bisherigen Nutzerszenarien und
Identifikation der Testgruppen
- M 5.4.1 Konzepte und Workflows zur Integration von Tools in das
TextGridLab
- M 5.5.1 Vollständige Veröffentlichung der aufbereiteten Daten

- M 5.5.3 Usability, Abschluss des Fallbeispiels und sich daraus ergebende Anforderungsanalyse
- M 6.1.3 Workshop mit Projektpartnern zum erarbeiteten Konzept des integrierten Monitoring
- M 6.1.4 Implementierung des integrierten Monitoring, Report „Integriertes Monitoring einer digitalen Forschungsinfrastruktur“
- M 7.3.1 Webauftritt TextGrid

Reports

Die nachstehenden Reports finden Sie unter <https://textgrid.de/dokumentation/projektdokumentation> zur Lektüre und zum Download.

TextGrid I (2006–2009)

- R 1.1 Text Processing
- R 1.2 Linking
- R 1.3 Text Retrieval
- R 1.4 Publishing
- R 1.5 Management von Workflow, Access, Kommunikation und Nutzer
- R 1.6 Ontologien
- R 2.1 Dokumentation für Editor-Erweiterung
- R 2.2 Linking
- R 2.3 Dokumentation für Query Interface und Druck, OCR
- R 3.1 Bericht über Evaluation der vorhandenen Grid-Middleware-Standards und Software-Pakete, unter Berücksichtigung der geplanten Dienste der Integrationsplattform und der in M 3.1. ermittelten Anforderungen
- R 3.10 Veröffentlichung der revidierten Fassung der TextGrid Processing Markup Language
- R 3.11 Businessmodell zum Betrieb einer Community Infrastruktur
- R 3.2 Spezifikation der Architektur für die TextGrid-Middleware
- R 3.3 Spezifikation einer ersten Version einer GridServices Application Interface
- R 3.4 Spezifikation einer ersten Version einer TextGrid Processing Markup Language

- R 3.5 Spezifikation der von den Projektpartnern zu verwendenden Middleware-Schnittstelle
- R 3.6 Spezifizierung aller weiteren von der TextGrid-Middleware zu bedienenden Grid-Schnittstellen/Standards zur Anbindung an die IP
- R 3.7 Implementierungsplan
- R 3.8 Spezifizierung eines Testszenarios unter Berücksichtigung der Musterapplikationen
- R 3.9 Testbericht
- R 4.1 Handbuch: Dokumentation der Tools und potentiellen Workflows
- R 5.1 Reports über Ontologie-Software
- R 5.2 Report über Tests mit philologisch-linguistischer Verfahren zur Erstellung der Meta-Lemmaliste
- R 5.3 Abschlussreport
- R 6.1 Konzept für die Öffentlichkeitsarbeit

TextGrid II (2009–2012)

- R 1.3.1 Testbericht zur ersten Usability-Testrunde, Feedback an Entwickler
- R 1.3.2 Testbericht zur zweiten Usability-Testrunde, Feedback an Entwickler
- R 1.4.1 Dokumentation des Releaseplanes, der Schemata zur Beschreibung der Softwarekomponenten und ihrer Versionen sowie der identifizierten Softwarepaketen mit ihren jeweiligen Maintainern
- R 1.4.2 Abschlussreport: Interoperabilität eSciDoc-TextGrid, Richtlinien bzw. Best Practices für neue Dienste, Perspektiven und Ergebnisse Browser-basierter TextGrid-Lösungen, Erfahrungen und Empfehlungen Usability als Faktor der Akzeptanz und Communitybildung in den Geisteswissenschaften
- R 2.1.1 Langzeitarchivierungsanforderungen der Nutzer
- R 2.1.2 Modellierung und Systemanalyse
- R 3.3.1 Draft-Version Organisationsmodell
- R 3.2.1 Musterverträge
- R 3.6.1 Abschließende Fassung des TextGrid-Geschäftsmodells inkl. Wirtschaftsplan für das Jahr 2012, Organisationsmodell und Satzung/Statuten der zu gründenden TextGrid-Organisation

- R 4.1.1 Erhebung Zielgruppen-spezifischer Anforderungen an ein TextGrid-Portal
- R 4.1.2 Dokumentation der bei der Umsetzung der Milestones gesammelten Erfahrungen, Empfehlungen für den Entwurf eines Organisationsmodells
- R 4.2.1 Analyse der Anforderungen an das TextGridVZ (einschl. der Integration in das TextGrid-Portal)
- R 4.5.1 Bericht zu eHumanities-Themen mit Relevanz für die allgemeine Informatik
- R 4.5.2 Bericht zu notwendigen Veränderungen und Ergänzungen bei den Foren, Issue-Trackern etc. für Entwickler
- R 4.7.1 GAP Analyse und Formulierung von Empfehlungen für Forschungsprojekte, die den nachhaltigen Betrieb eines eHumanities-Netzwerkes in Deutschland sichern
- R 6.1 Status und Dokumentation der Tools des ersten Jahres, Ergebnisse des Entwicklerkoordinationstreffens und der ersten übergreifenden Koordination
- R 6.2 Status und Dokumentation der Tools des zweiten Jahres, Ergebnisse des Zwischenreleases 1.5
- R 6.3 Abschlussbericht: Dokumentation der Tools
- R 8.1 PR-Konzept
- R 8.2 PR Materialien
- R 8.3 Newsletter

TextGrid III (2012–2015)

- R 1.1.1 Identifikation von Kostenfaktoren und Erarbeitung von Kostenmodellen
- R 1.1.2 Erfahrungsbericht Rechtsform: Praxisbewährung und Nutzeranforderungen
- R 1.1.3 Finalisierung Kosten- und Finanzierungsmodelle
- R 1.3.1 Machbarkeitsstudien zu neuen technischen Entwicklungen für TextGridLab
- R 2.2.2 (Laufende) Dokumentation von externen Plugins
- R 3.1.1 Einrichtung von Sandboxes, Entwicklung von use-cases, Zusammenstellung von Demo-Materialien

- R 3.1.2 Dokumentation für Anwender
- R 3.3.1 Report über die Erfahrungen aus den Nutzerschulungen, Empfehlungen für die weitere Ausgestaltung des Konzepts
- R 4.4.1 Dokumentation/Leitfaden Import (technisch/inhaltlich)
- R 4.4.2 SADE Dokumentation, Referenzinstanz
- R 4.5.1 Architekturskizze inklusive Identifikation der Lücken und der Optimierungsmöglichkeiten
- R 4.5.2 Update der in Monat 6 erstellten Architekturskizze, Potenzial der Optimierungen, erste Umsetzungen
- R 5.2.1 Laufende Funktionale Tests
- R 5.3.1 Usability, Zusammenfassung der bisherigen Nutzerszenarien und Identifikation der Testgruppen
- R 5.5.1 Tiefenerschließung der Wörterbücher und Enzyklopädien
- R 5.5.2 Usability: Anforderungen an TextGrid aus der Arbeit mit den Testgruppen
- R 5.5.3 Usability: Abschluss des Fallbeispiels und sich daraus ergebene Anforderungsanalyse
- R 6.1.1 Report zur Analyse „Dimensionen des integrierten Monitoring einer digitalen Forschungsinfrastruktur“
- R 6.1.2 Konzept Beurteilungsmethodik, Beurteilungskategorien und Governance Werkzeuge
- R 6.1.5 Abschließender Bericht zur Implementierung inklusive Evaluierungen
- R 6.2.1 Report „Ergebnisse der Nutzerbefragung“
- R 7.2.1 Konzept/Organisation Experten-, Strategiegelgespräche und TextGrid Summit etc.
- R 7.2.2 Kooperation mit Fachgesellschaften
- R 7.3.1 Öffentlichkeitsarbeit
- R 7.3.2 Social-Media Konzept, Umsetzung & Betreuung
- R 7.3.3 Marketingkonzept & Umsetzung

Nutzertreffen und Summits 2011–2015

Nutzertreffen

- **TextGrid Nutzertreffen I**
22./23. Februar 2012, Technische Universität Darmstadt
- **TextGrid Nutzertreffen II**
21./22. Juni 2013, Akademie der Wissenschaften und der Literatur, Mainz
- **TextGrid Nutzertreffen III**
07./08. November 2013, Max-Planck-Institut für Wissenschaftsgeschichte, Berlin
- **TextGrid Nutzertreffen IV: „Bilder in TextGrid“**
20./21. Februar 2014, Technische Universität Darmstadt
- **TextGrid Nutzertreffen V: „TextGrid anpassen und erweitern“**
04./05. August 2014, Albert-Ludwigs-Universität Freiburg, Slavisches Seminar
- **TextGrid Nutzertreffen VI: „Norm- und Metadaten“**
25./26. November 2014, Steinheim-Institut Essen
- **TextGrid Nutzertreffen VII: „TextGrid Grand Tour“**
05. März 2015, Harnack-Haus der Max-Planck-Gesellschaft, Berlin
- **TextGrid Nutzertreffen VIII: „Archivieren und Publizieren“**
11.–13. Mai 2015, Sternwarte, Göttingen

Summits

- **Digital Humanities Festakt**
12./13. Juli 2011, SUB Göttingen
- **TextGrid Summit**
14./15. Mai 2012, Darmstadttium, Darmstadt
- **DH Summit 2015**
03./04. März 2015, Harnack-Haus der Max-Planck-Gesellschaft, Berlin

Publikationen von und über TextGrid

- Al-Hajj, Yahya Ahmed Ali; Küster, Marc Wilhelm (2013): The text-image-link-editor: A tool for linking facsimiles and transcriptions, and image annotations. In: *Literary and Linguist Computing* 28 (2): 190–198.
- Aschenbrenner, Andreas (2008): Feature – Editing, analyzing, annotating, publishing: TextGrid takes the a, b, c to D-Grid. In: *iSGTW* 54. <http://www.isgtw.org/feature/isgtw-feature-editing-analyzing-annotating-publishing-textgrid-takes-b-c-d-grid>.
- Aschenbrenner, Andreas; Blanke, Tobias; Dunn, Stuart; Kerzel, Martina; Rapp, Andrea; Zielinski, Andrea (2007): Von e-Science zu e-Humanities – Digital vernetzte Wissenschaft als neuer Arbeits- und Kreativbereich für Kunst und Kultur. In: *Bibliothek. Forschung und Praxis* 31 (1): 11–21. https://www.b2i.de/fileadmin/dokumente/BFP_Bestand_2007/Jg_31-Nr_1/Jg_31-Nr_1_Aufsaeetze/Jg_31-2007-Nr_1-S_11-21.pdf.
- Aschenbrenner, Andreas; Blanke, Tobias; Chue Hong, Neil P.; Ferguson, Nicholas; Hedges, Mark (2009): A Workshop Series for Grid/Repository Integration. In: *D-Lib Magazine* 15.
- Aschenbrenner, Andreas; Gietz, Peter; Küster, Marc Wilhelm; Ludwig, Christoph; Neuroth, Heike (2006): TextGrid – a modular platform for collaborative textual editing. In: *Proceedings of the International Workshop on Digital Library Goes e-Science (DLSci06), September 21, 2006, Alicante, Spain*, S. 27–36.
- Aschenbrenner, Andreas; Küster, Marc Wilhelm; Ludwig, Christoph; Vitt, Thorsten (2009): Open eHumanities Digital Ecosystems and the Role of Resource Registries. In: *Proceedings of the 3rd IEEE International Conference on Digital Ecosystems and Technologies, Istanbul*, S. 745–750.
- Aschenbrenner, Andreas; Meffert, Katja (2008): Wissenschaftliche Infrastruktur in den Geisteswissenschaften? – Eine Wegbeschreibung. In: *Jahrbuch für Computerphilologie – online*. <http://computerphilologie.tu-darmstadt.de/jg07/aschmeff.html>.
- Blanke, Tobias; Aschenbrenner, Andreas; Küster, Marc; Ludwig, Christoph (2008): No Claims for Universal Solutions – Possible Lessons from Current e-Humanities Practices in Germany and the UK. In: *e-Humanities – An Emerging Discipline. Workshop at the 4th IEEE International Conference on e-Science*. December 2008.
- Blümm, Mirjam; Dogaru, George; Fischer, Peter M.; Jannidis, Fotis; Vitt, Thorsten; Witt, Andreas (2010): Transforming a large text collection into a TEI-conformant

- text resource: a case study. In: *Akten des TEI Member's Meeting 2010*, 11.–14. November, Zadar, Kroatien.
- Büdenbender, Stefan; Leuk, Michael (2009): Daten als Dienste: Wörterbücher als Erschließungsinstrumente in der virtuellen Arbeitsumgebung „TextGrid“. In: *it – Information Technology* (Themenheft „Informatik in den Geisteswissenschaften“) 51 (4): 191–196.
- Carusi, Annamaria; Reimer, Torsten (2010): TextGrid – Virtual Research Environment in the e-Humanities. In: *Virtual research environment collaborative landscape study – A JISC funded project. Januar 2010*, S. 92–94. <http://webarchive.nationalarchives.gov.uk/20140702233839/http://www.jisc.ac.uk/media/documents/publications/vrelandscape-report.pdf>.
- DARIAH (Hrsg.) (2010): German Grid Activities Related to DARIAH. In: *DARIAH Newsletter* No. 5 (Febr. 2010), S. 4–6.
- EHRI (Hrsg.) (2011): ‘If You Build It, They Will Come’ – Inauguration of Göttingen Centre for Digital Humanities. In: *EHRI Newsletter*, Nov. 2011. <http://www.ehri-project.eu/content/if-you-build-it-they-will-come>.
- Enders, Florian; Krause, Celia; Stotzka, Rainer; Tonne, Danah; Vanscheidt, Philipp (2014): Nach der Digitalisierung. Zur computergestützten Erschließung mittelalterlicher Handschriften. In: Philippi, Sabine; Vanscheidt, Philipp (Hrsg.): *Digitale Rekonstruktionen mittelalterlicher Bibliotheken*. Wiesbaden: Reichert Verlag, S. 81–98.
- Gießmann, Lukas; Küster, Marc Wilhelm; Ludwig, Christoph (2009): Isidorus-UI: Generating a User Interface with Topic Maps Constraint Language and JavaScript Object Notation. In: *Paper for the TMRA 2009*, S. 207–218.
- Gietz, Peter; Aschenbrenner, Andreas; Büdenbender, Stefan; Jannidis, Fotis; Küster, Marc Wilhelm; Ludwig, Christoph; Pempe, Wolfgang; Vitt, Thorsten; Wegstein, Werner; Zielinski, Andrea (2006): TextGrid and eHumanities. In: *Second IEEE International Conference on e-Science, December 4–6, 2006, Amsterdam*. DOI: 10.1109/E-SCIENCE.2006.261066.
- Graiger, Christian (2006): <philtag 5>, TEI P5: Encoding dictionaries & more. <http://textgrid.de/fileadmin/publikationen/graiger-2006.pdf>.
- GridTalk (Hrsg.) (2009): Digitising culture: Grids and eHumanities. In: *Grid Briefings – Grid Computing in five minutes* 10 (Dec. 2009), S. 2.
- Harms, Patrick; Grabowski, Jens (2011): Usability of Generic Software in e-Research Infrastructures. In: *Journal of the Chicago Colloquium on Digital Humanities and Computer Science* 1 (3). <https://letterpress.uchicago.edu/index.php/jdhcs/article/view/89/98> (15.01.2015).

- Hedges, Mark; Neuroth, Heike; Smith, Kathleen M.; Blanke, Tobias; Romary, Laurent; Küster, Marc; Illingworth, Malcolm (2013): TextGrid, TEXTvre, and DARIAH: Sustainability of Infrastructures for Textual Scholarship. In: *Journal of the Text Encoding Initiative* Issue 5 (June 2013). <http://jtei.revues.org/774>; DOI: 10.4000/jtei.774.
- Kamzelak, Roland S. (2009): Materialwirtschaft. In: *editio – Internationales Jahrbuch für Editionswissenschaft* 23: 159–168.
- Kerzel, Martina; Mittelbach, Jens; Vitt, Thorsten (2009): TextGrid – Virtuelle Arbeitsumgebung für die Geisteswissenschaften. In: *Künstliche Intelligenz* (Themenheft „Kulturerbe und Künstliche Intelligenz“) 4: 36–39.
- Klaver, Marie-José (2009): De Ub als digitale Bibliotheek. In: *SPUI 30, UvA Alumni Magazine* 2: 10–13.
- Küster, Marc Wilhelm; Ludwig, Christoph; Aschenbrenner, Andreas (2007): TextGrid as a Digital Ecosystem. In: *IEEE DEST 2007, 21.–23. Februar, Cairns, Australien*, Special Session 3: e-Humanities for Digital Eco-systems: A Social, Cultural, Economic and Political Agenda.
- Küster, Marc Wilhelm; Ludwig, Christoph; Aschenbrenner, Andreas (2009): TextGrid: eScholarship und vernetzte Angebote. In: *it – Information Technology* (Themenheft „Informatik in den Philologien“) 51 (4): 183–190.
- Küster, Marc Wilhelm; Ludwig, Christoph; Al-Hajj, Yahya; Aschenbrenner, Andreas (2010): TextGrid: eScholarship und der Fortschritt der Wissenschaft durch vernetzte Angebote. In: Sieglerschmidt, Jörg; Ohly, H. Peter (Hrsg.): *Wissenspeicher in digitalen Räumen*. Würzburg: Ergon Verlag, S. 193–205.
- Moreira, André (2011): LEXUS on the TextGrid infrastructure – exploring new potentialities. In: *Language Archiving Technology – News*, 19.12.2011. <https://tla.mpi.nl/tla-news/lexus-on-the-textgrid-infrastructure-exploring-new-potentialities/>.
- Müller, Marina; Minn, Gisela (2010): Der wissenschaftliche Arbeitsplatz der Zukunft – Internationale Tagung: „Virtuelle Forschungsplattformen in Geisteswissenschaften“. In: *Trierer Unijournal* 36 (1): 40–41.
- Neuroth, Heike; Aschenbrenner, Andreas; Lohmeier, Felix (2007): e-Humanities – eine virtuelle Forschungsumgebung für die Geistes-, Kultur- und Sozialwissenschaften. In: *Bibliothek. Forschung und Praxis* 31 (3): 272–279.
- Neuroth, Heike; Jannidis, Fotis; Rapp, Andrea; Lohmeier, Felix (2009): Virtuelle Forschungsumgebungen für e-Humanities. Maßnahmen zur optimalen Unterstützung von Forschungsprozessen in den Geisteswissenschaften. In: *Bibliothek. Forschung und Praxis* 33 (2): 161–169. https://www.b2i.de/fileadmin/dokumente/BFP_Bestand_2009/Jg_33-Nr_2/Jg_33-Nr_2_Aufsaeetze/Jg_33-2009-Nr_2-S_161-169.pdf.

- Neuroth, Heike; Lohmeier, Felix; Smith, Kathleen Marie (2011): TextGrid – Virtual Research Environment for the Humanities. In: *The International Journal of Digital Curation* 2 (6): 222–231.
- o. V. (2006): Tekstontsluiting op het Grid. In: *E-Data & Research* 1 (3): 2 (Bericht über den <philtag>-Workshop; in niederländischer Sprache).
- o. V. (2006): TextGrid – Modulare Plattform für verteilte und kooperative wissenschaftliche Textverarbeitung für die Geisteswissenschaften. In: *digital library forum*, 14.2.2006. http://www.dl-forum.pt-dlr.de/deutsch/projekte/projekte_2513_DEU_HTML.htm.
- o. V. (2007): TextGrid: Ein Community Grid für die Geisteswissenschaften. In: Neuroth, Heike; Kerzel, Martina; Gentzsch, Wolfgang (Hrsg.): *Die D-Grid Initiative*. Göttingen: Universitätsverlag, S. 64–66. http://webdoc.sub.gwdg.de/univerlag/2007/D-Grid_de.pdf.
- o. V. (2009): TextGrid – Ein Community-Grid für die Geisteswissenschaften. In: Herpay, Lajos; Neweling, Sonja; Schwiegelshohn, Uwe (Hrsg.): *D-Grid. Die Deutsche Grid-Initiative. Vorstellung der Projekte. Veröffentlichung im Rahmen des D-Grid All Hands Meeting, März 2009*. Dortmund, S. 38–39. <http://www.d-grid-ggmbh.de/downloads/BroschuereAHM2009.pdf>.
- Rapp, Andrea (2007): Das Projekt „TextGrid. Modulare Plattform für verteilte und kooperative wissenschaftliche Textdatenverarbeitung – ein Community-Grid für die Geisteswissenschaften“. Chancen und Perspektiven für eine neue Wissenskultur in den Geisteswissenschaften. In: Arbeitsgemeinschaft historischer Forschungseinrichtungen in der Bundesrepublik Deutschland (Hrsg.): *Jahrbuch der historischen Forschung in der Bundesrepublik Deutschland: Berichtsjahr 2006*. München: Oldenbourg, S. 61–68.
- [Rapp, Andrea] (2009): *Perspektiven des Digitalen Publizierens*. Interview mit Andrea Rapp im Deutschlandfunk am 6.7.2009.
- Rings, Thomas; Aschenbrenner, Andreas; Grabowski, Jens; Kálmán, Tibor; Lauer, Gerhard; Meyer, Jörg; Quadt, Arnulf; Sax, Ulrich; Viezens, Fred (2010): An Interdisciplinary Practical Course on the Application of Grid Computing. In: *Proceedings of the 1st Annual IEEE Engineering Education Conference – The Future of Global Learning in Engineering Education (EDUCON 2010)*. http://gerhardlauer.de/files/9713/2665/8806/rings_educon-2010.pdf.
- Rockenberger, Annika (2011): Digital-Humanities-Festakt in Göttingen. Eröffnung des Göttingen Centre for Digital Humanities (GCDH) und TextGrid Release 1.0, Göttingen, 12./13. Juli 2011. In: *edition* 25: 219–224. <http://www.textgrid.de/fileadmin/publikationen/rockenberger-2011.pdf>.
- Schneiker, Christian; Seipel, Dietmar (2009): Declaratively Creating and Processing XML/TEI Data. In: *Paper for the TEI Conference 2009: Text encoding in the era*

of mass digitization, Nov. 9–15, 2009. http://www1.pub.informatik.uni-wuerzburg.de/databases/papers/nlpcs_2009.pdf.

Schneiker, Christian; Seipel, Dietmar; Wegstein, Werner (2009): Schema and Variation: Retro-Digitizing Printed Dictionaries. In: *Proceedings of the Third Linguistic Annotation Workshop, ACL-IJCNLP 2009*, S. 82–89.

Schneiker, Christian; Seipel, Dietmar; Wegstein, Werner; Prätör, Klaus (2009): Declarative Parsing and Annotation of Electronic Dictionaries. In: Sharp, Bernadette; Zock, Michael (Hrsg.): *Natural Language Processing and Cognitive Science. Proceedings of the 6th International Workshop on Natural Language Processing and Cognitive Science – NLPCS 2009 (in conjunction with ICEIS 2009)*. Mailand: SciTePress, S. 122–132.

Schwiegelshohn, Uwe (2008): D-GRID: The communities of Germany's D-Grid. In: *eStrategies Projects 7*: 14–18. <https://www.textgrid.de/fileadmin/publikationen/schwiegelshohn-2008.pdf>.

Smith, Kathleen; Söring, Sibylle; Veentjer, Ubbo; Lohmeier, Felix (2012): The TextGrid Repository: Supporting the Data Curation Needs of Humanities Researchers. Abstract/Poster DH2012, Hamburg, 16.–22.07.2012.

Wegstein, Werner (2008): Vorüberlegungen zu einem parallelen polnisch-deutschen Textkorpus, in: Czachur, Waldemar; Czyzewska, Marta (Hrsg.): *Vom Wort zum Text. Studien zur deutschen Sprache und Kultur. Festschrift für Professor Józef Wiktorowicz zum 65. Geburtstag*. Warszawa: Inst. Germanistyki, S. 145–152.

Zielinski, Andrea; Pempe, Wolfgang; Gietz, Peter; Haase, Martin; Funk, Stefan; Simon, Christian (2009): TEI Documents In The Grid. In: *Literary and Linguistic Computing Advanced Access 14*. <http://llc.oxfordjournals.org/content/early/2009/05/14/llc.fqp016.full.pdf>.

Abschlussarbeiten

Aschenbrenner, Andreas (2009): *Reference Framework for Distributed Repositories – Towards an Open Repository Environment*. Dissertation, Göttinger Zentrum für Informatik, 25.11.2009.

Hausner, Roman (2009): *Semantic Text Mining – linguistische Tools im Preprocessing von Text Mining Methoden*. Bachelorarbeit, Göttinger Zentrum für Informatik, 29.4.2009.

Veentjer, Ubbo (2009): *Determination of Relevant Words within a Text and Representation in the User Interface TextGrid-Workbench*. Bachelorarbeit, Göttinger Zentrum für Informatik, 29.4.2009.

TextGrid im Netz

Website

<http://www.textgrid.de>

Download TextGridLab

<http://www.textgrid.de/download>

TextGrid Repository

<http://www.textgridrep.de>

Dokumentation und Hilfe

<https://textgrid.de/dokumentation>

Online-Manual

<http://www.textgrid.de/doku>

Nutzerforum

<http://www.textgrid.de/forum>

Informationen für Entwickler / technische Dokumentation

<http://www.textgrid.de/dev>

Help Desk und Bug Reports

<http://www.textgrid.de/support>

DHd-Blog – Digital Humanities im deutschsprachigen Raum

<http://www.dhd-blog.org>

Allgemeine Fragen und Schulungen

<http://textgrid-extern@gwdg.de>

Twitter @textgrid

DHd-Kanal TextGrid / DARIAH-DE auf YouTube

<https://www.youtube.com/user/dhdkanal>

Filme

Virtuelle Forschungswelten: Neue Technologien in den Geisteswissenschaften (2012)

<http://youtu.be/mQyn3D566Ew>

Virtual Research Worlds: New Technologies in the Humanities (2012) [engl.]

<http://youtu.be/BY9XqXJMbc8>

Digitale Wissensräume: Virtuelle Forschungslabore der Geisteswissenschaften (2014)

<http://youtu.be/tMBq7dlnuLg>

Digital Knowledge Spaces: Virtual Research Laboratories in the Humanities (2014) [engl.]

http://youtu.be/qedGW_c-nBQ

explainity erklärt: Digitales Edieren in einer Virtuellen Forschungsumgebung (2014)

<http://youtu.be/jJJgrZ4AaVE>

explainity explains: Digital Editing in a Virtual Research Environment (2014) [engl.]

<http://youtu.be/rchdyegdkZc>

Weitere Titel aus dem vwh-Verlag (Auszug)

nestor Handbuch Eine kleine Enzyklopädie der digitalen Langzeitarchivierung
Version 2.0, hrsg. v. Heike Neuroth, Achim Oßwald, Regine Scheffel, Stefan
Strathmann, Mathias Jehn

Juni 2009, ISBN 978-3-940317-48-3, 620 S., 24,90 € (D), 25,60 € (A), 37,90 CHF

Eine ausgewählte Pressestimme: *Rundbrief Fotografie* 16 (2009), Heft 4, S. 30–32:
*Die Sammlung an Kompetenz ist beachtenswert, das Thema dringlich. Man kann
sich nur wünschen, dass mit der Herausgabe dieses Handbuches auch in der Öffent-
lichkeit das Bewusstsein für die Bedeutung der Bewahrung des digitalen kulturellen
Erbes gefördert wird. Für direkt Betroffene und Fachleute ist es eine unerlässliche
Arbeitshilfe im Kampf gegen das „Dark Age“, den Verlust digitalen Kulturerbes
(oder auch „nur“ der persönlichen digitalen Daten).*

Langzeitarchivierung von Forschungsdaten Eine Bestandsaufnahme

hrsg. v. Heike Neuroth, Stefan Strathmann, Achim Oßwald, Regine Scheffel,
Jens Klump, Jens Ludwig im Rahmen des Kooperationsverbundes nestor

Febr. 2012, ISBN 978-3-86488-008-7, 382 S., 29,90 € (D), 30,74 € (A), 36,90 CHF

Forschungsdaten und der langfristige Zugriff auf sie sind für Wissenschaftler aller
Disziplinen von großer Bedeutung: als Nachweis des Forschungsprozesses und sei-
ner Ergebnisse, zur Nachnutzung im Rahmen inner- und interdisziplinärer Koopera-
tionen oder im Hinblick auf vergleichende Analysen. Dies alles ist nur möglich, wenn
Forschungsdaten gesichert, für eine langfristige Nachnutzung archiviert und zur Ver-
fügung gestellt werden. Angesichts rasant anwachsender digitaler Datenmengen ist
die Langzeitarchivierung von Forschungsdaten daher eine wichtige begleitende In-
frastrukturaufgabe. In dieser Untersuchung haben WissenschaftlerInnen aus elf Fach-
disziplinen systematisch den Stand und die aktuelle Praxis im Umgang mit der Lang-
zeitarchivierung von Forschungsdaten in ihrer jeweiligen Disziplin aufgearbeitet.

Prof. Dr. Stefan Büttner in der *iwp – Information. Wissenschaft & Praxis* 65 (2014),
Heft 4–5, S. 299–300: *Besonders hervorzuheben ist die bereits erwähnte Orientie-
rung auf die Fachdisziplinen sowie die stringente, sehr gut strukturierte Darstellung.
[...]Mit „Langzeitarchivierung von Forschungsdaten“ liegt eine weitere Publikation
vor, die sich mit Sicherheit als Standardwerk für die Langzeitarchivierung etabliert.*

Leitfaden zum Forschungsdaten-Management

Handreichungen aus dem WissGrid-Projekt hrsg. von Jens Ludwig und Harry Enke
Jan. 2013, Hardcover, 120 S., Abb. S/W, ISBN 978-3-86488-032-2, 15,80 €

Digitale Forschungsdaten sind eine unverzichtbare Grundlage moderner Wissen-
schaft. Mit ihnen sind eine Reihe von Datenmanagement-Fragen verbunden: Wie
lange sollen die Daten aufbewahrt werden? Welche Kontextinformationen müssen
erfasst werden, um die Daten zu einem späteren Zeitpunkt noch sinnvoll benutzen zu
können? Wie viel kostet die Aufbewahrung? Das Buch stellt mit einem Leitfaden
und einer Checkliste einfach handhabbare Instrumente für fachwissenschaftliche oder
Infrastruktur-Einrichtungen bereit, um die wichtigsten Aufgaben und Fragen im For-
schungsdaten-Management strukturiert beantworten und effizient planen zu können.

Digital Curation of Research Data Experiences of a Baseline Study in Ger-
many hrsg. von Heike Neuroth, Stefan Strathmann, Achim Oßwald, Jens Ludwig
Nov. 2013, Broschur, 94 S., ISBN 978-3-86488-054-4, 12,80 €

Bei dieser englischsprachigen Broschüre handelt es sich um eine Kurzfassung der
umfangreichen deutschsprachigen Publikation *Langzeitarchivierung von For-
schungsdaten* (s. o.) vom Februar 2012.

Reihe „Schriften zur Informationswissenschaft“

M. Heckner: Tagging, Rating, Posting
2009, 27,90 €, ISBN 978-3-940317-39-1

S. Mühlbacher:

Information Literacy in Enterprises
2009, 32,90 €, ISBN 978-3-940317-45-2

M. Maßun: Collaborative Information
Management in Enterprises
2009, 28,90 €, ISBN 978-3-940317-49-0

T. Memmel: User Interface Specifica-
tion for Interactive Software Systems
2009, 33,90 €, ISBN 978-3-940317-53-7

A. Ratzka: Patternbasiertes User Inter-
face Design für multimodale Interak-
tion 2010, 33,90 €, 978-3-940317-62-9

M. Prestipino: Die virtuelle Gemein-
schaft als Informationssystem
2010, 30,90 €, ISBN 978-3-940317-69-8

A. Warta: Kollaboratives Wissens-
management in Unternehmen
2011, 30,90 €, ISBN 978-3-940317-90-2

J. Griesbaum/T. Mandl/C. Womser-
Hacker (Hg.): Information und Wis-
sen: global, sozial und frei?
Proceedings ISI 2011

2011, 36,50 €, ISBN 978-3-940317-91-9

M. Görtz: Social Software as a Source
of Information in the Workplace
2011, 31,90 €, ISBN 978-3-86488-006-3

C. Carstens: Ontology Based Query
Expansion Retrieval Support for the
Domain of Educational Research
2012, 34,90 €, ISBN 978-3-86488-011-7

A. Köstlbacher: Eine Informations-
system-Infrastruktur für das Thera-
peutische Drug Monitoring
2012, 27,90 €, ISBN 978-3-86488-019-3

S.-J. Untiet-Kepp: Adaptive Feedback
zur Unterstützung in kollaborativen
Lernumgebungen

2012, 30,90 €, ISBN 978-3-86488-023-0

H.-C. Hobohm (Hg.): Informations-
wissenschaft zwischen virtueller Infra-
struktur und materiellen Lebenswelten
Proceedings des 13. Internationalen Sym-
posiums für Informationswissenschaft
(ISI 2013), Potsdam, 19.–22. März 2013

2013, 34,50 €, ISBN 978-3-86488-035-3

N. Henrichs: Menschsein im Informa-
tionszeitalter Informationswissenschaft
mit Leidenschaft und missionarischem Eifer
2014, 33,90 €, ISBN 978-3-86488-061-2

Reihe „E-Humanities“

C. Russ: Online Crowds Massenphäno-
mene und kollektives Verhalten im Inter-
net 2010, 31,50 €, 978-3-940317-67-4

M. Janneck/C. Adelberger: Komplexe
Software-Einführungsprozesse gestal-
ten: Grundlagen und Methoden Am
Beispiel eines Campus-Management-Sys-
tems 2012, 26,90 €, 978-3-940317-63-6

H. Kohle: Digitale Bildwissenschaft
2013, 16,80 €, ISBN 978-3-86488-036-0

C. Krause/R. Reiche: Ein Bild sagt
mehr als tausend Pixel? Digitale For-
schungsansätze in den Bild- und Objekt-
wissenschaften Erscheint Febr./März
2015, ISBN 978-3-86488-076-6

Reihe „Web 2.0“

J. Moskaliuk (Hg.): Konstruktion und
Kommunikation von Wissen mit Wikis
2008, 27,50 €, ISBN 978-3-940317-29-2

H. Frohner: Social Tagging
2010, 26,90 €, ISBN 978-3-940317-03-2

R. Bauer: Die digitale Bibliothek
von Babel Über den Umgang
mit Wissensressourcen im Web 2.0
2010, 26,90 €, ISBN 978-3-940317-71-1

J. Jochem: Performance 2.0
Zur Mediengeschichte der Flashmobs
2011, 24,90 €, ISBN 978-3-940317-98-8

G. Franz: Die vielen Wikipedias
Vielsprachigkeit als Zugang zu einer globali-
sierten Online-Welt
2011, 27,50 €, ISBN 978-3-86488-002-5

R. Sonnberger: Facebook im Kontext
medialer Umbrüche
2012, 29,50 €, ISBN 978-3-86488-009-4

J. Brailovskaia:
Narzisstisch und sensationssuchend?
Eine Studie zum Zusammenhang zwischen
Persönlichkeitsmerkmalen und Online-
Selbstdarstellung am Beispiel von *studivZ*
2013, 24,50 €, ISBN 978-3-86488-039-1

Reihe „Medientheorie“

W. Drucker: Von Sputnik zu Google
Earth Über den Perspektivenwechsel
hin zu einer ökologischen Weltsicht
2011, 25,90 €, ISBN 978-3-940317-82-7

S. Brugner: Über die Realität
im Zeitalter digitaler Fotografie
2012, 23,90 €, ISBN 978-3-86488-018-6

J. Ahl: Web-TV
Entstehungsgeschichte, Begriffe, Ästhetik
2014, 33,90 €, ISBN 978-3-86488-074-2

Reihe „E-Learning“

- T. Strasser: Moodle im Fremdsprachenunterricht
2011, 28,50 €, ISBN 978-3-940317-92-6
- M. Nagl: Game-Assisted E-Learning in der Umweltbildung
2011, 28,50 €, ISBN 978-3-940317-94-0
- C. Biel: Personal Learning Environments als Methode zur Förderung des selbstorganisierten Lernens
2011, 24,90 €, ISBN 978-3-86488-001-8
- K. Himpsl-Gutermann: E-Portfolios in der universitären Weiterbildung
2012, 30,90 €, ISBN 978-3-86488-014-8
- M. Beißwenger/N. Anskeit/A. Storrer (Hg.): Wikis in Schule und Hochschule
2012, 36,90 €, ISBN 978-3-86488-017-9
- P. Bettinger: Medienbildungsprozesse Erwachsener im Umgang mit sozialen Online-Netzwerken
2012, 25,90 €, ISBN 978-3-86488-020-9
- C. Lehr: Web 2.0 in der universitären Lehre Ein Handlungsrahmen für die Gestaltung technologiegestützter Lernszenarien
2012, 27,90 €, ISBN 978-3-86488-024-7
- J. Wagner/V. Heckmann (Hg.): Web 2.0 im Fremdsprachenunterricht Ein Praxisbuch für Lehrende in Schule und Hochschule
2012, 27,50 €, ISBN 978-3-86488-022-3
- E. Blaschitz et al. (Hg.): Zukunft des Lernens Wie digitale Medien Schule, Aus- und Weiterbildung verändern
2012, 23,50 €, ISBN 978-3-86488-028-5
- U. Höbarth: Konstruktivistisches Lernen mit Moodle Praktische Einsatzmöglichkeiten in Bildungsinstitutionen - 3. Aufl. -
2013, 31,50 €, ISBN 978-3-86488-033-9
- A. Klampfer: E-Portfolios als Instrument zur Professionalisierung in der Lehrer- und Lehrerinnenausbildung
2013, 27,90 €, ISBN 978-3-86488-034-6
- M. Hielscher: Autorentools für multimediale und interaktive Lernbausteine Architektur und Einsatzszenarien von LearningApps.org
2013, 26,50 €, ISBN 978-3-86488-041-4

- C. Koenig: Bildung im Netz Analyse und bildungstheoretische Interpretation der neuen kollaborativen Praktiken in offenen Online-Communities
2013, 31,90 €, ISBN 978-3-86488-042-1
- B. Getto: Anreize für E-Learning Eine Untersuchung zur nachhaltigen Verankerung von Lerninnovationen an Hochschulen
2013, 26,90 €, ISBN 978-3-86488-052-0
- M. Gecius: Game-based Learning in der Schule Hamlet als Computerspiel – Die Einsatzmöglichkeiten der computerbasierten Spielform des Adventures im literaturbezogenen Englischunterricht
2014, 37,50 €, ISBN 978-3-86488-066-7
- M. Kirchner: Social-Software-Portfolios im Einsatz Zwischen Online-Lernen und Medienkompetenz im selbstgesteuert-konnektiven Lernalltag
2015, 32,90 €, ISBN 978-3-86488-075-9

Reihe „Medienwirtschaft“

- K. Huemer: Die Zukunft des Buchmarktes Verlage und Buchhandlungen im digitalen Zeitalter
2010, 24,90 €, ISBN 978-3-940317-73-5
- J.-F. Schrape: Gutenberg-Galaxis Reloaded? Der Wandel des deutschen Buchhandels durch Internet, E-Books und Mobile Devices
2011, 17,90 €, ISBN 978-3-940317-85-8
- B. Blaha: Von Riesen und Zwergen Zum Strukturwandel im verbreiteten Buchhandel in Deutschland und Österreich
2011, 24,90 €, ISBN 978-3-940317-93-3
- J. Stiglhuber: Macht und Ohnmacht der Unabhängigen Independent-Verlage und ihre Verortung
2011, 26,90 €, ISBN 978-3-86488-003-2

Kleine Schriften

- J.-F. Schrape: Wiederkehrende Erwartungen Visionen, Prognosen und Mythen um neue Medien seit 1970 [Softcover]
2012, 11,90 €, ISBN 978-3-86488-021-6
- weitere Reihen im vwh-Programm (s. www.vwh-verlag.de):*

- TypoDruck
- AV-Medien
- Game Studies
- Multimedia
- E-Business
- E-Collaboration
- Schriften des Innovators Club



Aktuelle Ankündigungen, Inhaltsverzeichnisse und Rezensionen finden sie im vwh-Blog unter www.vwh-verlag.de.

Das komplette Verlagsprogramm mit Buchbeschreibungen sowie eine direkte Bestellmöglichkeit im vwh-Shop finden Sie unter www.vwh-verlag-shop.de.

