Abschlussbericht

öffentliche Fassung

Version 1.0 vom 5. November 2009

verantwortlicher Partner: SUB Göttingen

TextGrid

Modulare Plattform für verteilte und kooperative wissenschaftliche Textdatenverarbeitung - ein Community-Grid für die Geisteswissenschaften

Bundesministerium für Bildung und Forschung

Projekt: **TextGrid**

Teil des D-Grid Verbundes und der deutschen e-Science Initiative

BMBF Förderkennzeichen: 07TG01A-H

Laufzeit: Februar 2006 - Mai 2009

Dokumentstatus: final

Verfügbarkeit: öffentlich

Autoren: TextGrid Konsortium

Niedersächsische Staats- und Universitätsbibliothek Göttingen	Neuroth, Heike Aschenbrenner, Andreas
Institut für Sprach- und Literaturwissenschaft, Technische Universität Darmstadt	Vitt, Thorsten
Institut für Deutsche Sprache	Witt, Andreas Zielinski, Andrea Kupietz, Marc
Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften an der Universität Trier	Rapp, Andrea Büdenbender, Stefan
Fachhochschule Worms, Fachbereich Informatik und Telekommunikation	Küster, Marc Wilhelm Ludwig, Christoph
Kompetenzzentrum EDV-Philologie an der Universität Würzburg	Wegstein, Werner
DAASI International GmbH	Gietz, Peter Funk, Stefan Haase, Martin
Saphor GmbH	Pempe, Wolfgang

Inhaltsverzeichnis

I. Kurze Darstellung

1. Aufgabenstellung

a) Projektziele

Über Jahrhunderte hinweg war die Arbeit eines Textwissenschaftlers zumeist durch die Auseinandersetzung des einzelnen Wissenschaftlers mit „seinem“ Text gekennzeichnet. Der Gesamtkomplex der wissenschaftlichen Textdatenverarbeitung wird nunmehr mit der Einführung und Bereitstellung kollaborativer Methoden und durch Nutzung von Netzwerken mit verteilten Ressourcen und standardisierten Werkzeugen auf eine neue Basis gestellt. Diese Zielvorstellung trägt den Tendenzen einer zunehmend interdisziplinären, mobilen und in virtuellen Arbeitsumgebungen operierenden Wissenschaft Rechnung, in der eine gridfähige, aber von den technischen Details des Grid abstrahierende Workbench für die philologische Bearbeitung, Analyse, Annotation, Edition und Publikation von Textdaten eine zentrale Rolle spielt. Dabei können Werkzeuge in einer für den Endnutzer völlig transparenten Weise von verschiedenen Institutionen und Fachdisziplinen eingebracht und (international) vernetzt werden.

Zu den verteilten Werkzeugen gehören verteilte Daten. Vor allem die Einbindung von Objekten in Bild- und Tonformaten erzeugt höchste Ansprüche an Speicher- und Netzkapazitäten. Die Verteilung der Ressourcen, die internationale Dislozierung von Forschungs- und Arbeitsgruppen und lokal bereitgestellte, potentiell integrationsfähige Module für die wissenschaftliche Bearbeitung bilden den zweiten Baustein für den Aufbau eines geisteswissenschaftlichen Community Grid. Nur auf diesem Weg entsteht eine Plattform, auf der zeit- und ortsunabhängig die weltweit gestreuten wissenschaftlichen Kompetenzen zur Bearbeitung von Texten gebündelt und mit einem Set von modularen Softwaretools ausgestattet werden, wobei bereits vorhandene Teillösungen adaptiert und integriert werden.

Wichtig war den Projektteilnehmern dabei auch von Anfang an die Zielvorstellung eines Semantic Grid. Die EDV-philologischen und textwissenschaftlichen Auswertungsmethoden, insbesondere hochgradig strukturierter Textsorten wie z.B. Wörterbücher, werden einen unverzichtbaren Grundstein für die zuverlässige Erfassung von Ontologien legen. Darauf wiederum werden über Mapping-Verfahren Instrumente entwickelt, die ihrerseits wesentliche Dienste für eine Strukturierung der digitalen Welt nach semantischen Kriterien sind.

b) Infrastruktur

Für die Wissenschaftler eröffnen sich durch die neue, in TextGrid aufgebaute verteilte Infrastruktur attraktive Möglichkeiten, hauptsächlich mit Blick auf die Weiterentwicklung von Methoden und die Verbesserung der Kommunikation. Die wissenschaftliche Diskussion wird in der Community bereits im Prozess der jeweiligen Vorhaben gefördert, in vielen Fällen, vor allem im interdisziplinären Kontext, überhaupt erst ermöglicht. Werkzeuge, deren bloße Existenz einem Fachwissenschaftler in vielen Fällen vorher gar nicht bekannt gewesen wäre, können nun einfach genutzt werden.

Für die textwissenschaftliche Klientel gehört dabei die einfache Nutzbarkeit (Usability) zu den Kernvoraussetzungen für die Akzeptanz der Infrastruktur. Deswegen bleibt ein Großteil der von uns aufgebauten Grid-Infrastruktur für den Endbenutzer zunächst einmal verborgen:

- die Mechanismen zur Einbindung verteilter Werkzeuge in Workflows, die der Forschungsfrage angepasst sind,
- die Verwaltung von Metadaten und multiplen Datenrepositorien etwa für kernkodierte Daten und ihre Originale,
- die Konvertierung von Forschungsdaten in die Kernkodierung,
- die Existenz zentraler Utilities etwa zur Authentifizierung oder zur Bearbeitung von Recherchen,
- die Verteilung von Daten auf TextGrid-Knoten,
- das Management von und die Ressourcenzuteilung für textwissenschaftliche Werkzeuge.

Selbst für Softwareentwickler, die TextGrid um domainspezifische Werkzeuge erweitern, ist ein Gutteil dieser Infrastruktur unsichtbar. Sie ermöglicht ihnen allerdings, ihre eigenen Daten schrittweise zu integrieren und neue Werkzeuge für weitere Bereiche beizusteuern, die dann auch für den Endnutzer sicht- und nutzbar sind.

2. Voraussetzungen, unter denen das Vorhaben durchgeführt wurde

Die Projektteilnehmer arbeiten seit 2003 in der Würzburger Arbeitsgruppe <philtag> an der Definition und Entwicklung philologischer und textwissenschaftlicher Software. Dort sind so verschiedene Kompetenzen wie die des Informatikers, des klassischen Philologen, des Mediävisten, des Linguisten, des Literaturwissenschaftlers und des Philosophen vereint. Ausgangspunkt für die Kooperation war die Erkenntnis, dass TEI sich als bestes Datenformat für die langfristige Archivierung von philologisch interessanten Texten erwiesen hat und weiterhin erweist. Alle Teilnehmer können auf Grund ihrer Beteiligung an Digitalisierungsprojekten auf weitreichende praktische Kenntnisse über die Verwendung von TEI zurückgreifen. Sie sind ebenfalls an der Weiterentwicklung der TEI-Richtlinien interessiert und verfolgen und gestalten diese aktiv. Die Orientierung an wesentlichen einschlägigen Standards bestimmt gleichzeitig die Mitarbeit an Gremien wie der Dublin Core Metadata Initiative (SUB im Advisory Board der DCMI) oder im Editorial Board des Metadata Encoding Transmission Standard (METS).

Bereits vor Projektbeginn waren die Antragsteller an der Digitalisierung und Erstellung von Editionen, von digitalen Wörterbüchern und von Korpora umfassend beteiligt (in Auswahl):

- Deutsches Wörterbuch von Jacob und Wilhelm Grimm: www.dwb.uni-trier.de (Trier)
- Mittelhochdeutsche Wörterbücher im Verbund: www.mwv.uni-trier.de (Trier)
- Das Heinrich-Heine-Portal: www.hhp.uni-trier.de (Trier)
- Der junge Goethe in seiner Zeit: www.jgoethe.uni-muenchen.de (Jannidis)
- Das Projekt „Historisches Korpus": Sammlung historischer deutscher Texte der Zeit zwischen dem Ende des 18. und der Mitte des 20. Jahrhunderts (IDS)

Im Rahmen dieser und anderer Projekte haben sich die Beteiligten dabei mit den Anforderungen an philologische Software im allgemeinen vertraut gemacht und die Einschränkungen der vorliegenden Programme kennen gelernt. Dabei sind sie zu der Einsicht gelangt, dass lokale Lösungen entscheidende Vorteile der Digitalisierung verschenken und somit die Verbreitung des kulturellen Erbes, wie es Netzwerkeffekte bewirken könnten, verzögern.

Ein erster Schritt der Würzburger Arbeitsgruppe bestand daher darin, die Anforderungen an textwissenschaftliche Software und die Architektur einer modularen integrativen Lösung zu definieren, was in einer Reihe von Vorträgen und Treffen sowie der Diskussion in einem Online-Forum geschah. Dies war der Ausgangspunkt für die Konzeption von Prototypen für einzelne Problemlösungen, die im Rahmen von Seminaren und Abschlussarbeiten entwickelt wurden.

DAASI International GmbH verfügt über einschlägige Technologieexpertise insbesondere in den Bereichen Middleware-Infrastruktur, Security und Informationsmanagement. Viele an Hochschulen durchgeführte Projekte können als Vorarbeiten angesehen werden, insbesondere weil hierdurch Open Source Module und Bibliotheken in den Bereichen Authentifizierung,

Autorisierung, Public Key Infrastructure und Verzeichnisdienste erstellt worden sind. Schließlich waren die Arbeiten im Rahmen des GGFs, insbesondere zu Grid Information Services, Grid PKI und CIM-basierende Schema-Modellierung sowie die generelle Expertise zu Grid-Technologien, die durch langjährige Mitarbeit in den Arbeitsgruppen und Diskussionen mit Gridexperten entstanden ist, innerhalb des Projekts sinnvoll einsetzbar.

Die Firma Saphor GmbH verfügt über langjährige Erfahrung in den Bereichen Datenstrukturierung und -konvertierung, Publishing und DTD-Entwicklung. Für die Konvertierung und Aufbereitung von Textdaten existieren zahlreiche Programmbausteine, die für die Entwicklung der geplanten Module herangezogen werden können. Zur Publikation strukturierter Daten (SGML/XML) liegen Satzroutinen für TEI nach xsl:fo sowie für TEI nach LaTeX vor, die streng an den jeweiligen Standards orientiert sind. Web-Publishing mit Cocoon und Webservices sind weitere Bereiche, in denen Saphor über entsprechende Kompetenzen verfügt.

Gemäß den Vorgaben der Förderlinie und des Projektträgers beruhte der Entwurf für die TextGrid-Middleware-Architektur auf einer engen Anknüpfung an die vom DGI bereitgestellte Integrationsplattform. Das Grid-Datenmanagement spielt eine zentrale Rolle in dem Projekt. Da die vom DGI angebotenen bzw. im Kern-D-Grid verfügbaren Lösungen nicht in allen Punkten den Anforderungen der Community genügten (z.B. vertrauenswürdige Speicherung durch ein nachhaltiges Betriebskonzept Speicher, Rechtemanagement), wurde die TextGrid-Middleware um Komponenten erweitert, die die gewünschten Funktionalitäten zur Verfügung stellen.

Ein entscheidender Faktor für die Entwicklung des Projekts war die projektübergreifende Zusammenarbeit im D-Grid-Kontext, sowohl im Rahmen von direkten Kontakten mit anderen Community-Grids (insbesondere MediGrid und C3-Grid), als auch im Rahmen von zentralen Veranstaltungen, wie den vom DGI organisierten Workshops. Von besonderem Interesse waren hier die Workshops zum Thema Grid-Middleware (Globus) und GAT, sowie zum Thema Nachhaltigkeit und Geschäftsmodelle.

3. Planung und Ablauf des Vorhabens

Als zentraler Faktor, der das Gesamtprojekt ausgezeichnet und der es so erfolgreich gemacht hat, ist die kooperative und offene Vorgehensweise aller Partner hervorzuheben. Sowohl die interne Kommunikation als auch die Kommunikation mit externen Partnern und Interessenten an TextGrid war ausgesprochen intensiv. Diese Interaktion war essenziell, damit sich das Projekt flexibel mit dem sich schnell ändernden Projektkontext (siehe Kapitel II.3,"Während der Durchführung ...") weiterentwickeln konnte.

Das Projekt propagierte die bewusste Öffnung zur wissenschaftlichen Community hin und suchte den Dialog über die klassischen wissenschaftlichen Kommunikationsmittel: ... hat auf relevanten Konferenzen präsentiert (siehe Kapitel II.1, AP6), Kooperationen mit externen Projekten initiiert (siehe Kapitel I.5), und auch externe Experten in interne Diskussionen mit aufgenommen. Diese Offenheit ist durch das Interesse der Community belohnt worden: neben regelmäßigen Einladungen und wichtigen Kooperationen haben mittlerweile rund 300 Personen den TextGrid Newsletter abonniert. Der abschließende TextGrid Summit unterstreicht mit

einer Vielzahl von Gästen aus Europa und den USA (etwa 130 Teilnehmer) die erreichten Ergebnisse.

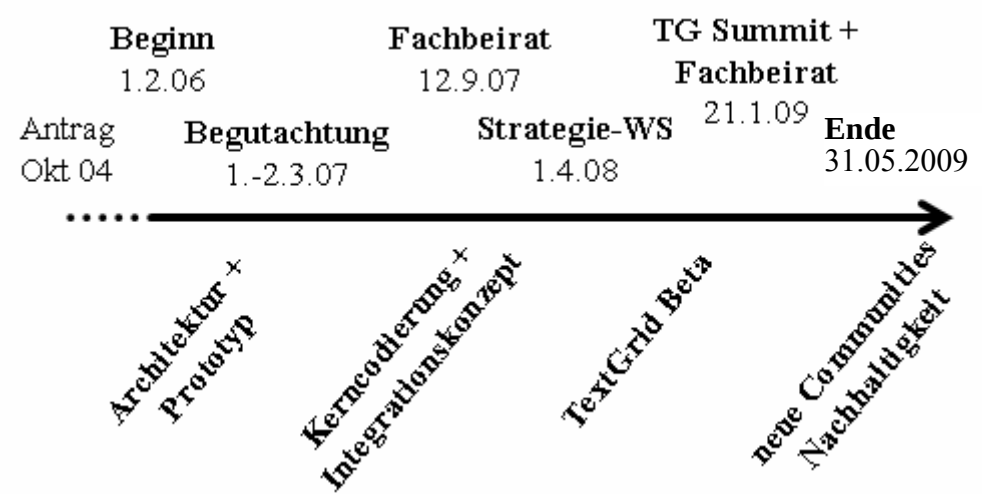


Abbildung 1: grobe Chronologie des Projekts

a) Projektüberblick

Das TextGrid-Konsortium besteht aus acht Partnern (siehe Annex A), die in allen sechs Arbeitspaketen sehr eng zusammengearbeitet haben. Wegen dieser Abhängigkeiten hat sich das Gesamtprojekt von Anfang an darauf geeinigt, den Projektbeginn gemeinschaftlich auf den 01.02.2006 festzulegen.

Zur Kommunikation wurde auf eine Reihe von parallelen, komplementären Kanälen gesetzt:

- Der interne Bereich der Website: die Projekt-**Wiki** enthält alle wichtigen Dokumente und archiviert detailliert Ablauf und Entscheidungen des Projektes.
- Verknüpft mit dem Wiki ist ein **WebDAV**, das vor allem im späteren Projektverlauf zum Austausch von Beispieldaten nützlich war.
- Eine Reihe von **Mailinglisten** für die jeweiligen Arbeitspakete, Arbeitsgruppen und das Gesamtprojekt ermöglicht zielgerichtete Kommunikation auch mit externen Partnern. Wie in den anderen D-Grid-Projekten auch wurde das Listensystem des DFN[1] verwendet, das zudem die gesamte Kommunikation archiviert hat.
- **Treffen** wurden regelmäßig abgehalten - für Arbeitsgruppen in den verschiedensten Konstellationen je nach Bedarf, für das Gesamtprojekt etwa einmal im Quartal.
- Neben virtuellen und persönlichen Arbeitsgruppentreffen haben sich speziell für die technische Entwicklungsarbeit "**Programmiersprints**" durchgesetzt. In Programmiersprints treffen sich Programmierer für einige Tage, um Module zu integrieren und ge-

[1] DFN Mailinglisten. http://www.listserv.dfn.de/

meinsam Probleme zu lösen. Schon eine kurze Zeit fokussierter Teamarbeit trägt wesentlich zur Kohärenz der Gesamtsoftware bei und verringert nachhaltig den Aufwand für den Einzelnen und das Gesamtprojekt.

- Zwischen den Treffen ermöglichten **Video- bzw. Telefonkonferenzen** den kontinuierlichen Austausch. Für Videokonferenzen stand das Videokonferenzsystem des DFN-Vereins[2] zur Verfügung. Für Telefonkonferenzen hat sich im Laufe des Projektes "skype" als Mittel der Wahl durchgesetzt, das auch kurzfristige Termine und Absprachen ermöglicht.
- Zur Außendarstellung dienen vor allem die **TextGrid-Website** und ein regelmäßig publizierter **Newsletter**. Die mitunter monatlicher mehr als 8.000 Besucher auf der TextGrid Website[3] und rund 300 Abonnenten des Newsletters unterstreichen das öffentliche Interesse, das TextGrid erzeugt hat.

b) Arbeitspakete und Arbeitsgruppen

In Abstimmung mit dem Projektträger wurden nach knapp der Hälfte der Projektlaufzeit die Deliverables (Meilensteine und Berichte) für die Arbeitspakete 2 und 3 angepasst, deren ursprüngliche Planung aus dem Projektantrag im Herbst 2004 stammte. Diese Anpassung schien durch die bis zu diesem Zeitpunkt gesammelten Erfahrungen der Partner dringend nötig, um die Einhaltung der Projektziele zu sichern und die Außendarstellung von TextGrid zu stärken. (Die Auflistung der Deliverables in Kapitel II.1. zeichnet die angepassten Deliverables explizit aus.) Der Zeitplan für die Deliverables wurde bis auf vernachlässigbare Verzögerungen eingehalten, zu Projektende liegen alle Deliverables nach Plan vor.

Neben der Arbeit in den Arbeitspaketen wurden - ebenfalls schon sehr früh im Projektverlauf - thematische Arbeitsgruppen gebildet, die quer über Arbeitspakete hinweg agierten und damit die Kohärenz des Gesamtprojekts sicherstellten.

- AG Textformate - Diskussion und Spezifikation von Abläufen und Anforderungen an Tools, Metadaten und - als wichtigstes Element für TextGrid-weite Interoperabilität - der TextGrid-Kernkodierung (baseline encoding).
- AG Wörterbücher - Untergruppe der AG Textformate speziell für Formate im Bereich Wörterbücher.
- AG Linguistische Korpora - Untergruppe der AG Textformate speziell für Formate im Bereich linguistischer Korpora.
- AG Architektur - technische Konzeptionsarbeit zur grundlegenden Infrastruktur von TextGrid (z.B. Schnittstellen, API's, Schichtenmodell) und Forum zur Verknüpfung der verschiedenen Module und Werkzeuge.

[2] Videokonferenzsystem des DFN-Verein.

[3] http://www.textgrid.de/webalizer (Zugriff nur über den passwortgeschützten internen Bereich)

Auch externe Experten (z.B. Tom Baker, W3C; Torsten Schaßan, Wolfenbüttel; Sebastian Rahtz, TEI; Kollegen aus D-Grid und andere mehr) wurden unbürokratisch in Diskussionen mit eingebunden, wenn dies für eine bestimmte Fragestellung als hilfreich erachtet wurde.

c) Fachbeirat

Neben der gezielten Einladung von Experten im Hinblick auf spezifische Aspekte des Projekts hat TextGrid von einem eigens eingerichteten Fachbeirat (siehe Annex D) profitiert. Dieser hat das Projekt kontinuierlich inhaltlich, technisch und strategisch begleitet. Der Fachbeirat hat einerseits dazu beigetragen, dass Maßnahmen und Entwicklungen im Projekt zielgerichtet erfolgt sind und dem aktuellen Stand der Forschung auf informationstechnologischem und fachwissenschaftlichem Gebiet entsprechen. Auf der anderen Seite waren unter den Fachbeiratsmitgliedern auch wichtige Multiplikatoren, die den Bekanntheitsgrad von TextGrid gesteigert haben.

Neben der ständigen Möglichkeit für Feedback und Diskussion über eine eigens eingerichtete Mailingliste gab es zwei Fachbeiratstreffen - eines in der Mitte (12.9.2007) und eines am Ende (22.01.2009) der Projektlaufzeit. Sowohl auf den Fachbeiratstreffen als auch auf einem e-Humanities-Strategietreffen (30.3.-2.4.2008) waren Vertreter des BMBF, des Projektträgers sowie der DFG vertreten.

Empfehlungen des ersten Fachbeirats-Treffens lauteten:

- Auf- und Ausbau nationaler und internationaler Kooperation,
- Verstärkung der Anstrengungen zum Community-Building,
- Aufbau eines gridbasierten Archivs zur Langzeitarchivierung (länger als 10 Jahre) als wesentlicher Infrastrukturdienst für die Geisteswissenschaften.

Diesen Empfehlungen wurden in der weiterführenden Projektlaufzeit gefolgt: Es wurde beschlossen, dass Endprodukt konzeptionell in zwei Teile zu gliedern, das TextGridLab als Virtuelle Arbeitsumgebung und das TextGridRep als erster Schritt zum gridbasierten Archiv zur Langzeitarchivierung von Forschungsdaten. Das Interesse, das dem Projekt seitens der Community entgegen gebracht wird, wurde nicht nur während des Summit im Januar 2009 deutlich: rund 130 internationale Fachwissenschaftler nahmen an dieser Tagung teil. Auch Anfragen um mögliche Kooperationen oder andere Möglichkeiten der Zusammenarbeit haben zugenommen, so dass auf der Projekt-Homepage eine Unterseite mit Informationen für Kooperationspartner angelegt wurde (auch englischsprachig).

Das zweite Fachbeiratstreffen fand im Rahmen des TextGrid-Summit statt und wurde dazu genutzt, dem Fachbeirat das TextGridLab und das TextGridRep vorzustellen. Die Anregungen konnten aufgegriffen und zum größten Teil bereits umgesetzt werden. Damit wurde die äußerst fruchtbare Zusammenarbeit mit dem Fachbeirat in dieser Zusammensetzung beendet.

d) Projektergänzungen

Das langfristige Ziel für TextGrid ist der Aufbau einer nachhaltigen e-Humanities-Infrastruktur für die Geisteswissenschaften, wenn möglich in Verbindung mit einer nachhaltigen Struktur für D-Grid. Dieses visionäre Ziel kann nicht unter den Rahmenbedingungen ei-

nes zeitlich stark begrenzten Projektes durchgeführt werden und muss sich in eine (inter)national entstehende e-Humanities-Infrastruktur einfügen. TextGrid hat sich schon sehr früh (international) vernetzt - ein kurzer Überblick über die wichtigsten Kontakte und Kooperationen findet sich in Kapitel 5 (Zusammenarbeit mit anderen Stellen). Hier sind einige Kooperationen herausgehoben, die einen wesentlichen Einfluss auf den Projektverlauf hatten, ferner Erweiterungen des Projektumfangs im Rahmen von D-Grid:

TextGrid hat die Entwicklung des geplanten **Kollationierers** in den **Interedition**-Verbund eingebracht (s. Kap. 5). Die Interedition-Partner aus Birmingham (u.a. Peter Robinson) haben langjährige Erfahrung bei der Entwicklung von Kollationierungstechnologien und diese Kooperation sichert die internationale Relevanz und Qualität des entstehenden Produktes. Obwohl die Abstimmung mit einem internationalen Partner zusätzlichen Aufwand bedeutete, ist diese Kooperation also ein wesentlicher Erfolg, der TextGrid aus fachwissenschaftlicher Sicht noch weit über die Projektlaufzeit hinaus beträchtlichen Einfluss auf internationale Entwicklungen gibt.

Auch der geplante "**Text Publisher Print**" konnte wesentlich erweitert werden. Diskussion und Feedback auf Konferenzen hat die Wichtigkeit dieses Werkzeuges in der Community deutlich gemacht, die bei Antragstellung in diesem Umfang nicht klar war. In Kooperation mit weiteren externen Partnern wie der MPDL[4] wurde ein DFG-Projekt initiiert[5], das den Erwartungen der Nutzer an ein XML-basiertes Print-Modul in vollem Umfang gerecht werden kann. Auch bei den Werkzeugen **Bibliografietool** und **OCR** konnten wichtige Kontakte geknüpft und zukünftige Projektkooperationen (u.a. IBM, DFKI bzw. Technische Universität Kaiserslautern) initiiert werden, die direkt auf die TextGrid-Infrastruktur aufgesetzt werden können.

Ein weiteres, eng mit TextGrid verknüpftes Projekt wurde von einigen der TextGrid-Partner gemeinsam mit neuen externen Partnern initiiert, es handelt sich um das seit Oktober 2008 vom BMBF geförderte Projekt "Wechselwirkungen zwischen linguistischen und bioinformatischen Verfahren, Methoden und Algorithmen: Modellierung und Abbildung von Varianz in Sprache und Genomen" Verbundvorhaben im Rahmen des BMBF-Förderschwerpunktes "Wechselwirkungen zwischen Natur- und Geisteswissenschaften". Das Projekt geht davon aus, dass es prinzipielle strukturelle Gemeinsamkeiten zwischen sprachlichem Code und Genomcode gibt, insofern als Information und Wissen durch Gruppierung von Einzelzeichen nach bestimmten Regeln kodiert werden und für beide "Entwicklungsfähigkeit" und "Varianz" essenziell sind. Es sollen Methoden und Algorithmen entwickelt und erprobt werden, mittels derer Varianz erschlossen und geordnet werden kann. Kern des sprachwissenschaftlichen Zugangs ist eine Basis-Lemmaliste des Standarddeutschen, auf die klassifizierte Varietäten-Lemmalisten (z.B. historische Sprachstufen, regionale Varietäten, Autorvarietäten) abgebildet werden. Auf dieser Grundlage und basierend auf Genomdatenbanken wird eine "Grammatik der Varianz in Genomen und Sprache" in Ontologien modelliert. Die Ergebnisse

[4] Max Planck Digital Library - MPDL. http://www.mpdl.mpg.de/

[5] Das Projekt wurde im Oktober 2009 bewilligt und startet voraussichtlich am 1.2.2010.

werden visualisiert und in die TextGrid-Umgebung eingebracht. Ein differenzierteres Verständnis der Mechanismen von Entwicklung und Varianz ermöglicht präzisere Verfahren der Informationsgewinnung, der Daten-Speicherung, -Bearbeitung und -Auswertung. Das Projekt leistet damit also u.a. einen Beitrag zur präziseren semantischen Erschließung und Ordnung großer Textmengen.

Auch im Rahmen von D-Grid ist TextGrid über Projekte mit weiteren Aktivitäten verknüpft. So nahm ein TextGrid-Partner im D-Grid-Projekt IVOM (Interoperabilität und Integration der VO-Management-Technologien im D-Grid) Teil und brachte die TextGrid-spezifischen Anforderungen insbesondere in Bezug auf die Shibboleth-Integration mit ein. Des weiteren werden die TextGrid-Spezifika in dem kürzlich bewilligten D-Grid-Projekt SLC-GAP, in dem es u.a. um die Integration des vom DFN-Verein bereitgestellten Short Lived Credential Service (SLCS) geht, über einen TextGrid-Partner eingebracht.

Diese nationalen und internationalen Aktivitäten zur Publikation von TextGrid und für Kooperationen, wie auch die intensive Abstimmung innerhalb von D-Grid bedeuteten eine große Chance für TextGrid, gleichzeitig aber auch einen erheblichen Mehraufwand - sowohl zeitlich als auch finanziell. Das BMBF hat diese Bemühungen des TextGrid-Projektes gesehen und mit einer großzügigen **Aufstockung der Reisemittel** die Fortsetzung dieser Aktivitäten ermöglicht.

Durch eine substanzielle jährliche **Sonderinvestition** stellte das BMBF in den Jahren 2006 bis 2008 sicher, dass D-Grid auf eine gute Grundausstattung der Hardware-Infrastruktur aufbauen kann. Auch TextGrid hat sich in diesen Runden am Aufbau von D-Grid beteiligt. Die Hardware wurde am Rechenzentrum GWDG (Gesellschaft für wissenschaftliche Datenverarbeitung mbH[6]) im Rahmen der Göttingen-weiten Kooperation GoeGrid (HEP, Medizin, Informatik, GWDG) aufgebaut und dem D-Grid-Verbund zur Verfügung gestellt.

Letztlich hat das BMBF allen D-Grid-Projekten der ersten Runde eine Verlängerung ihrer Aktivitäten um einige Monate bis Ende Februar 2009 ermöglicht. Dadurch dass TextGrid mit Februar 2006 erst später als andere D-Grid-Projekte begonnen hat, bedeutete dies eine **kostenneutrale Verlängerung** um vier Monate bis Ende Mai 2009. TextGrid hat diese zusätzliche Zeit genutzt und seine Anstrengungen für eine nachhaltige, permanente Aufstellung fortgesetzt.

4. Wissenschaftlicher und technischer Stand, an den angeknüpft wurde

Obwohl in den Textwissenschaften der Computer für bestimmte Aufgaben bereits früh (in den 50er Jahren) Einzug gehalten hatte, war dies in der Regel der Arbeitsbereich einiger weniger Spezialisten. Die Folge war (und ist im Grunde bis heute) die Tendenz zu lokalen Installationen mit jeweils projektbezogenen Applikationen (Insellösungen), hinzu kommt das Fehlen verbindlicher gemeinsamer Standards und Schnittstellen für Digitalisate, Formate und Werk-

[6] http://www.gwdg.de/index.php

zeuge. Daraus resultieren erhebliche infrastrukturelle Nachteile für alle Überlegungen, die mit hohem Aufwand produzierten Ressourcen für zukünftige Vorhaben in geeigneter Weise und mit geeigneten Arbeitsinstrumenten nutzbar zu machen.

Insgesamt ergibt sich in Bezug auf den EDV-Einsatz in den Textwissenschaften noch immer ein uneinheitliches Bild, das wohl nicht zuletzt auch auf ein bisweilen ambivalentes Verhältnis zum Medium Computer zurückzuführen ist: Einerseits ist der Rechner als Hilfsmittel zur Bewältigung traditioneller Aufgaben des Faches nicht mehr wegzudenken, andererseits ist die Bereitschaft, grundlegende Methoden und Arbeitsweisen im Hinblick auf das volle Potenzial des neuen Mediums zu reflektieren und ggf. neu auszurichten, äußerst unterschiedlich ausgeprägt.

Während viele Einzelprojekte heute gut mit Standard-Arbeitsrechnern durchgeführt werden können, erweist sich aber bereits in diesem Zusammenhang die Tendenz zum Einzelplatzrechner oft als Hemmschuh. Neue, potenziell sehr ergiebige Gebiete und Verfahren – genannt sei etwa die automatische Verknüpfung von Wörterbuchartikeln – sind schwer zu etablieren, wo Umfang der Datenmengen und Komplexität der Algorithmen besondere Speicher- und Rechenleistungen erfordern, die mit Standardhardware nicht abzudecken sind. Bei der Erforschung neuer statistischer Verfahren zum automatischen semantischen Clustering und zur Verschlagwortung von Texten, um nur zwei Beispiele zu nennen, kommt ein weiterer Aspekt hinzu: Sie verbrauchen nicht nur sehr viel CPU-Zeit, sondern arbeiten z.T. zunehmend effizienter, je größer und reichhaltiger die Textgrundlagen sind, auf die sie sich stützen können. Die Möglichkeit, jederzeit dynamisch auf verteilte Korpora zugreifen zu können, die ihrerseits durch Verbesserungen und Ergänzungen ständig aktualisiert werden, ist für solche Ansätze von essenzieller Bedeutung.

Während hier im Bereich der Computerlinguistik, insbesondere mit Arbeiten, die unter den Begriff Semantic Web sowie unter automatisches Clustering von Textressourcen subsumiert werden können, durchaus Fortschritte erzielt werden konnten, die weitreichende Konsequenz auf computergestützte Geisteswissenschaften haben, existiert bisher keine tragfähige übergreifende Infrastruktur, die einen Zugriff auf verteilte Daten- und Computing-Ressourcen auf Grundlage verbindlicher Standards und Schnittstellen ermöglichen würde.

Dass eine solche Forschungsumgebung nicht nur ein Desiderat, sondern gleichzeitig auch eine äußerst schwierige Herausforderung darstellt, ist unbestritten, doch sind die Textwissenschaften hier nicht auf sich alleine gestellt. In vielen Bereichen der Naturwissenschaften, wie Kernphysik, Meteorologie, Genetik etc. haben sich Formen des verteilten Arbeitens und des Teilens von Ressourcen auf Grundlage von Grid-Technologien fest etabliert.

Einige dieser spezifischen z.T. bereits seit langem existierenden Einzel-Lösungen (wie TUSTEP, WordCruncher, Tact usw.) bieten zwar eine herausragende Sammlung von Wissen über Anforderungsprofile, Algorithmen und Lösungskonzepte, weisen jedoch aus heutiger Sicht auch gravierende Mängel auf (keine Unterstützung einschlägiger Standards wie z.B. TEI und Unicode, schwerfällige Handhabbarkeit, steile Lernkurve, mangelnde Netzwerk- und

Gridfähigkeit). Neben ersten Ansätzen für geisteswissenschaftliche eSciences ('Arts and Humanities Research Board[7]'; „A System for publishing Scientific Data“[8]) haben insbesondere die Arbeiten im Bereich der Computerlinguistik weitreichende Konsequenzen für die computergestützten Geisteswissenschaften. Auch im Global Grid Forum hat sich eine Research Group namens „Humanities, Arts, and Social Science RG[9] etabliert. Obwohl also erste Ansätze existieren, war das Grid-Computing noch nicht in den Geisteswissenschaften angekommen. Daher lag gerade hierin eine Chance für die D-Grid-Initiative, durch die Integration einer großen Anwendergruppe mit langer EDV-Tradition neue Akzente im Grid-Computing zu setzen.[10]

TextGrid ist daher angetreten, die Möglichkeiten des Einsatzes kollaborativer Methoden sowie der Nutzung verteilter Ressourcen und standardisierter Werkzeuge zu initiieren bzw. zu stärken. Als ein zentrales Ziel im Bereich der Tool-Entwicklung wurde eine gridfähige Workbench für die philologische Bearbeitung, Analyse, Annotation, Edition und Publikation von Textdaten identifiziert. Mit dem Arbeitspaket 1 stand dabei ein eigenes Modul bereit, das die wissenschaftliche Ausgangslage in Hinblick auf existierende Softwarelösungen dokumentierte und daraus Empfehlungen für das weitere Vorgehen ableitete. Dabei lag der Fokus auf dem Bereich der Textdatenverarbeitung und des Nutzer/Workflow/Wissensmanagements, Fragen zur Systemarchitektur wurden anderweitig behandelt (s.u.). Im Einzelnen wurden folgende Gebiete identifiziert und untersucht:

- Text Processing
- Linking
- Text Retrieval
- Publishing
- Management von Workflow, Access, Kommunikation und Nutzer
- Ontologien

Zu jedem dieser Punkte entstand ein ausführlicher Report, auf den an dieser Stelle nur verwiesen werden kann (vgl. eingefügte Hyperlinks). Zudem flossen die Ergebnisse in einen Aufsatz, der die Arbeitsbereiche und die Alblaufanalysen in TextGrid diskutiert, in das so genannte „Szenarienpapier“[11] mit ein, in dem typische Anwendungsszenarien und Vorschläge für Softwarelösungen zusammengestellt wurden, sowie in die daraus abgeleiteten Pflichtenhefte für die Entwicklung eigener Tools. Zusammenfassend lässt sich sagen, dass zu Beginn von TextGrid in den meisten Bereichen zwar bereits eine Auswahl von mehr oder

[7] http://www.nesc.ac.uk, http://www.ahrbict.rdg.ac.uk/informationresources/e_science.htm

[8] http://www.nesc.ac.uk/action/projects/project_action.cfm?title=195

[9] http://forge.gridforum.prg/projects/hass-rg

[10] Einen Überblick über die deutschen Aktivitäten im Bereich der Philologie gibt die Bibliographie über Deutschsprachige Literatur zur Computerphilologie: http://computerphilologie.uni-muenchen.de/jg99/bibliographie.html.

[11] http://www.textgrid.de/fileadmin/TextGrid/TextGrid-Szenarien_061212.pdf

minder verbreiteten Werkzeugen existierte, diese sich aber in keinem Fall direkt in die TextGrid-Infrastruktur integrieren ließen. Dies scheiterte teils an allgemeinen Kriterien (kein open source, undokumentierter Code, plattformabhängig), teils an spezifischeren (mangelnde GUI, mangelnde Unterstützung wesentlicher Standards oder Funktionalitäten) oder an einer Inkompatibilität zur Systemarchitektur.

Für die konkrete Toolentwicklung ergaben sich damit, abhängig von der jeweiligen Ausgangssituation, verschiedene Szenarien:

- Anpassung, Zusammenführung und ggf. Ausbau existierender Lösungen

Dieser Weg wurde insbesondere bei sehr komplexen Tools wie etwa dem XML-Editor gewählt und war insofern im Einzelfall mit erheblichem Arbeitsaufwand verbunden.

- Neuentwicklung, ggf. in separatem Projekt

Wo keine entsprechenden Lösungen vorhanden oder eine Anpassung nicht aussichtsreich war, wurden Tools vollkommen neu entwickelt. In einzelnen Fällen wurde nach grundlegenden Untersuchungen und Rücksprache mit der Community deutlich, dass die vorhandenen Anforderungen innerhalb des ursprünglichen Projektrahmens nicht sinnvoll zu bewältigen sein würden. So werden derzeit der Printpublisher und die Meta-Lemmaliste über spin-off Projekte zu einem wesentlich weiteren Grad entwickelt, als das in TextGrid möglich gewesen wäre.

- Kooperation mit nationalen und internationale Initiativen

In einigen Bereichen gab es bereits Entwicklungsansätze, die starke Synergien versprachen. So hat sich etwa im Rahmen des „Interedition“ Projekts ein fruchtbarer Austausch ergeben, von dem der Kollationierer und der grafische Linkeditor sehr profitieren, und der von der konzeptuellen Ebene bis hin zum gemeinsamen Erstellen von Programmcode reicht.

Die Sichtung der einschlägigen Literatur zeigte die breite Spannweite an philologischen Anwendungen, die Forscher entwickeln und für die TextGrid idealerweise ein geeignetes Framework bietet soll. Auch wenn es nicht im Rahmen dieses Projektes möglich sein wird, für all diese Anwendungen die notwendigen Module in TextGrid bereitzustellen, so müssen sie doch jetzt schon bei der Planung von TextGrid bedacht werden, um nicht künftige Lösungen zu verbauen – allein dafür lohnte sich bereits der für die Sichtung notwendige Aufwand, weil diese Punkte für die spätere Akzeptanz in den textwissenschaftlichen Communities mit entscheidend sein werden.

Darüber hinaus ergaben sich dabei wertvolle Hinweise auf frühere Softwareprojekte (ARCHway, GATE), die sich zwar noch nicht die Möglichkeiten des Grid Computings zu Nutze machten, aber ebenfalls einen modularen Ansatz verfolgten und z.T. ähnliche Frameworks zur GUI-Gestaltung einsetzten, wie sie für TextGrid vorgesehen sind. TextGrid plant, mit den Entwicklern dieser Projekte Kontakt aufzunehmen, um von den dort gemachten Erfahrungen zu profitieren. Ob es auch möglich sein wird, einzelne der dort entstandenen Module nach TextGrid zu portieren, lässt sich zum jetzigen Zeitpunkt noch nicht abschätzen.

Nach diesem Überblick über die vorgefundene Landschaft der *Digital Humanities* sollen nun noch einige Abschnitte zum Stand der Grundlagentechnik folgen, die eher unabhängig von

diesem speziellen Anwendungsgebiet sind – hier der Bereich Grid und Authentifizierung/Autorisierung. Oft werden verschiedene Ausrichtungen von Grid-Umgebungen hervorgehoben: *Computational Grid*, *Service Grid* und *Data Grid*. Ein Computational Grid wird errichtet, um rechenintensive Aufgaben wie z.B. Simulationen verteilt und möglichst schnell durchführen zu können. Bei dem Begriff Service Grid schwingen je nach Community unterschiedliche Ausrichtungen mit, allgemein versteht man unter Service Grid die Verwendung von Services und Konzepten der *Service Oriented Architecture* (SOA) zum Aufbau einer Grid-Umgebung, also die Endnutzer-orientierten Services und Tools. In einem Data Grid schließlich werden verteilte Ressourcen zu einem virtuellen Speichersystem zusammengeschlossen, in dem Daten zuverlässig, effizient und sicher abgelegt werden können. TextGrid bedient die Bereiche Service Grid und Data Grid. Hierunter fallen alle Services, die vom TextGrid-Lab (TG-lab) aus genutzt werden können, sowie das TextGrid-Repository (TG-rep), das für die Verwaltung der Daten verantwortlich ist.

Ein zentrales Paradigma für Grid Umgebungen war und ist die *Service Oriented Architecture* (SOA) [erl04]. Aktuelle Grid Standards basierten zu Projektbeginn auf einer solchen „loosely coupled“ Architektur, und die *Open Grid Services Architecture* (OGSA) [ogsa_v1] beinhaltet auch heute noch SOA-basierte grundlegende Konzepte für Grid-Architekturen. Eine erste detailliertere Spezifizierung der Konzepte von OGSA in der Form von *Open Grid Services Infrastructure* (OGSI) ist inzwischen allerdings vom *Web Service Resource Framework* (WSRF) abgelöst worden. Bei WSRF handelte es sich bei Projektbeginn noch um eine sehr neue Technologie, die noch nicht in einfach zu benutzenden und produktionsreifen Bibliotheken zur Verfügung stand. Daher hat sich TextGrid bemüht, die WSRF-Funktionalität vor dem Anwendungsprogrammierer zu kapseln. Alle Zugriffe auf das TextGrid-Repository werden mit einen CRUD-Service (create/read/update/delete) gekapselt (TG-crud), der unter anderem den Zugriff auf das Grid regelt. TG-crud ist als WebService implementiert.

Ein *WebService* ist ein internetbasierter serverseitiger Dienst, der Clients eine öffentliche Standard-basierte Schnittstelle zur Verfügung stellt, über die dieser den WebService ansprechen kann. Da in TextGrid Werkzeuge definiert und implementiert wurden, die im Grid verteilt sind, boten sich WebServices als Technologie für diese Werkzeuge an. Mit dem *entsprechenden Transport-Protokoll* SOAP können XML-basierte Informationen in einer dezentralisierten, verteilten Umgebung wie einem Grid ausgetauscht werden. Im Rahmen von WebServices sind diese XML-Informationen Aufrufe, Ergebnisse und Fehlermeldungen von WebService-Operationen. Um eine SOAP-Nachricht zwischen zwei Parteien zu verschicken, muss ein konkretes Netzwerk-Protokoll zum Transport der Nachrichten angegeben werden. Das immer noch am häufigsten verwendete Protokoll ist das *Hypertext Transfer Protocol* (HTTP). Im Rahmen von TextGrid bot es sich an, für den dazugehörenden Nachrichtenaustausch SOAP über HTTP zu verwenden. Nahezu alle in TextGrid genutzten Service-Schnittstellen wurden mit SOAP realisiert, so dass eine einheitliche Kommunikation innerhalb von TextGrid möglich ist. Darüber hinaus können jederzeit weitere SOAP WebServices an TextGrid angebunden werden. Das im Vergleich zu SOAP einfachere und deswegen performantere WebService-Protokoll REST (REpresentational State Transfer) wird in verschiedenen Teilen der TextGrid-Middleware ebenfalls unterstützt.

Zur Erstellung eines Datengrid werden unterschiedliche Aufgaben und Funktionalitäten gebraucht, die im Sinne einer modularen Architektur zumeist in einzelnen Komponenten gekapselt sind. Man kann hierbei unterscheiden zwischen den Bereichen *Datentransfer*, *Datenmanagement*, und *Information Services*. Zu Projektbeginn gab es bereits eine Vielzahl von Grid-Softwarepaketen, die bestrebt waren (und es noch sind), die in den vorherigen Abschnitten beschriebenen Standards zu implementieren, und so beim Aufbau eines Grids behilflich zu sein. Zum einen gibt es hier Low-Level-Softwarepakete, die lediglich Grid-Standards wie WSRF implementieren (z.B. Globus Toolkit 4), und zum anderen High-Level-Softwarepakete, die auf den ersteren aufbauen und dem Applikationsentwickler eine vereinfachte Schnittstelle zu den Low-Level-Softwarepaketen anbieten (z.B. GAT, SAGA und das JavaCoG Kit). Das *Globus Toolkit* ist ein Grid-Softwarepaket (eine sogenannte Grid-Middleware), welches eine Implementierung von WSRF/WSN beinhaltet und darauf aufbauend Low-Level-Softwarekomponenten – unter anderem in Form von WebServices – anbietet, die die oben genannten Aspekte eines Grids umzusetzen helfen.

Das *Grid Application Toolkit* (GAT) ist eine objektorientierte API, die bestrebt ist, Grid-Applikationsentwicklern einen einfachen Zugang zu Grid-Funktionalitäten anzubieten. GAT ist lediglich eine abstrakte Spezifikation einer API, zu deren Einsatz es einer konkreten Realisierung bedarf. Es müssen Vorkehrungen getroffen werden, die gewährleisten, dass beispielsweise die Operationen der Dateiverwaltung von GAT effektiv durchgeführt werden. Kandidaten für eine konkrete Realisierung solcher Operationen waren etablierte Grid Middleware Softwarepakete wie Globus Toolkit oder gLite. Zu Projektbeginn gab es bereits Adaptoren für viele Komponenten von Globus Toolkit 4, Adaptoren für gLite waren in Planung. Daher wurde in TextGrid GAT genutzt, um das Globus Toolkit für die Umsetzung des TextGrid Repositories anzusprechen: Hierbei kommen die GAT-Komponenten *FileManagement*, *FileStreamManagement* und *LogicalFileManagement* zum Einsatz.

Der Bereich Authentifizierung (Feststellung der Identität eine Benutzers) und Autorisierung (Regelung der Rechte des Benutzers) stellte sich zu Projektbeginn folgendermaßen dar: der SAML-Standard, der eine föderierte, also verteilte Authentifizierung und Autorisierung ermöglicht, war noch relativ neu. Die darauf basierende Software Shibboleth [shibboleth] begann sich langsam, insbesondere an US-amerikanischen Hochschulen und in der Schweiz, zu etablieren. In Deutschland wurde sie erstmals in Projekten im Umfeld von Universitätsbibliotheken (Freiburg, Regensburg) eingesetzt [vascoda]. Erst parallel zur Projektlaufzeit wurde die deutsche Föderation DFN-AAI aufgebaut, die dann entsprechend bei der Implementierung der TextGrid-Authentifizierung berücksichtigt wurde.

Während für die Authentifizierung föderierte Ansätze dem aktuellen Stand der Technik entsprachen, waren im Bereich Autorisierung hauptsächlich zentrale Ansätze vorherrschend, die auch schon eine Standardisierung hinter sich hatten. Als der wohl bekannteste und neueste ist hier der XACML-Standard [XACML] zu nennen, der verschiedene sehr feingranulare Möglichkeiten der Autorisierung vorsieht, aber auch äußerst komplex ist; bei vielen Implementierungen dieses Standards ist auch eine unzureichende Performanz zu verzeichnen. Nicht ganz so mächtig, aber trotzdem recht feingranular ist der RBAC-Standard (Role-Based Access

Control) [RBAC]. Hier werden einzelne Berechtigungen auf Ressourcen (z. B. lesen, schreiben) an Rollen geknüpft, und jeder Benutzer kann bestimmte Rollen innehaben. Dieser Standard wurde im Projekt umgesetzt.

Der Stand der Technik zur Repräsentation von Autorisierungsinformation kann nicht so einfach umrissen werden. Häufig kommen hier SQL- oder XML-Datenbanken zum Einsatz; es gibt hier aber auch Verzeichnisdatenbanken, die das LDAP-Protokoll (Light-weight Directory Access Protocol) [LDAP] implementieren. Ein solches Verzeichnis ist für häufige Lese-Zugriffe optimiert, was eine optimale Voraussetzung sowohl für die Repräsentation von Authentifizierungsinformationen als auch für Autorisierungsentscheidungen bietet. Folglich wurde sowohl für das TextGrid-Community-Verzeichnis (für Benutzer, die in der Föderation keinen anderweitigen Zugang haben), als auch für die zentrale Autorisierungsdatenbank als Kern des RBAC-Systems ein OpenLDAP-Server eingesetzt.

Weitere und detaillierte Informationen zur Technik sind in TextGrid-Report 3.1 enthalten [rep31].

Literatur

[erl04] Thomas Erl: ServiceOriented Architecture: A Field Guide to Integrating XML and Web Services. Prentice Hall, 2004. ISBN 0131428985.

[osga_v1] The Open Grid Services Architecture, Version 1.0., http://www.ggf.org/documents/GFD.30.pdf

[LDAP] Zeilenga, Kurt (Ed.): Lightweight Directory Access Protocol (LDAP): Technical Specification Road Map. RFC 4510ff, June 2006. Online: http://tools.ietf.org/html/rfc4510 (und von dort weiterführende Links).

[RBAC] Role Based Access Control. ANSI Standard ANSI INCITS 359-2004

[shibboleth] Scavo, Tom., Cantor, Scott: Shibboleth Architecture – Technical Overview. Working draft, 02, 8 June 2005. Online: http://shibboleth.internet2.edu/docs/draft-mace-shibboleth-tech-overview-latest.pdf

[vascoda] Ruppert, Ato, Oberknapp, Bernd, Lienhard, Jochen, Borel, Franck: Das AAR-Projekt: Einsatz von Shibboleth im Bereich Wissenschaft und Lehre. 2. 4. 2007, Online: http://aar.vascoda.de/doc/management/management_report.pdf

[XACML] Moses, Tim (Ed.): eXtensible Access Control Markup Language (XACML) Version 2.0. OASIS Standard, 1 Feb 2005. Online: http://docs.oasis-open.org/xacml/2.0/access_control-xacml-2.0-core-spec-os.pdf

[rep31] TextGrid-Report 3.1: „Bericht über eine Evaluation von Grid Middleware Standards und von Grid Software Paketen“, August 2006, Online: http://www.textgrid.de/fileadmin/TextGrid/reports/TextGrid_Report_3_1.pdf

5. Zusammenarbeit mit anderen Stellen

Während des gesamten Projektzeitraumes fand ein intensiver Austausch mit nationalen und internationalen Projekten und Initiativen statt, dessen Ziel die Vernetzung von TextGrid, die Gewährleistung von struktureller und semantischer Interoperabilität der TextGrid-Komponenten sowie die Entwicklung einzelner Werkzeuge war.[12]

a) Interedition

So bildete sich im zweiten Jahr des Projektes mit Interedition[13] ein Verbund aus Editionswissenschaftlern, der über die nächsten vier Jahre die internationale Kooperation erleichtern will und darin durch COST gefördert wird. Interedition wird die Interoperabilität von Werkzeugen aus den Editionswissenschaften strukturell verbessern und den Nutzen von interoperablen Netzwerken aus Werkzeugen praktisch demonstrieren. Gemeinsam mit Interedition wurde ein Kollationierer entwickelt, dessen Basismodul zum Vergleich einzelner Textblöcke getestet werden kann[14].

b) CLARIN und DARIAH

Darüber hinaus wurde eine Kooperation zwischen TextGrid und CLARIN sowie DARIAH vereinbart. DARIAH strebt die Errichtung einer internationalen Infrastruktur für die e-Humanities an. DARIAH ist im Rahmen von ESFRI (European Strategy Forum for Research Infrastructures) langfristig finanziert und auf Nachhaltigkeit ausgerichtet. Durch die Mitarbeit in CLARIN sollen insbesondere Ressourcen (Tools, Daten, Services) für Sprachwissenschaftler verfügbar gemacht werden. CLARIN ist im Rahmen von ESFRI langfristig finanziert und auf Nachhaltigkeit ausgerichtet.

Zusammenarbeit bei der Entwicklung von Tools fand bzw. findet außerdem auch in Bereichen statt, die durch TextGrid identifiziert, jedoch innerhalb der Projektlaufzeit nicht adäquat bearbeitet werden konnten. So hat sich im Zuge der Vorarbeiten (Anforderungsanalyse, Literatursichtung, Evaluation existierender Werkzeuge) für das Modul Publisher Print herausgestellt, dass die Diversität der fachwissenschaftlichen Anforderungen sehr viel größer war, als ursprünglich angenommen. Zugleich ergaben Befragungen des anvisierten Benutzerkreises, dass die projektierte reine Batchlösung als nicht mehr zeitgemäß empfunden wurde. Ein Redesign des Printmoduls unter der Vorgabe, alle Benutzerwünsche abzudecken und interaktiv benutzbar zu sein, erwies sich im Rahmen des Projekts als nicht realisierbar. Die für das Print-Modul vorgesehenen Ressourcen wurden investiert, um ein interaktives sowie batch- und workflowfähiges Satzprogramm für komplexe XML-Daten mit Integration in TextGrid zu konzipieren und separat bei der DFG zu beantragen. Dieses im Oktober 2009 in vollem

[12] Vgl. auch: Understanding Global Activity in Higher Education and Research:
Primary, desk and web-based research, ed. Joint Information Systems Committee (JISC);
http://www.jisc.ac.uk/media/documents/publications/global_activity_in_he_final.pdf.

[13] http://interedition.huygensinstituut.nl/

[14] http://collatex.huygensinstituut.nl:2000

Umfang bewilligte Projekt wird auf den Ergebnissen von TextGrid aufbauen. Bewilligt wurde auch der Projektantrag „Wechselwirkungen zwischen linguistischen und bioinformatischen Verfahren, Methoden und Algorithmen: Modellierung und Abbildung von Varianz in Sprache und Genomen“, der anlässlich des BMBF-Calls „Wechselwirkungen zwischen Natur- und Geisteswissenschaften“ zur Unterstützung des Milestones M 5.3 Metalemmaliste eingereicht wurde. Im Rahmen dieses Projektes soll die Erstellung einer philologisch geprüften, umfassenden Metalemmaliste auf einer möglichst breiten Basis, insbesondere auch im Zusammenspiel von klassischen philologischen Verfahren und naturwissenschaftlichen Ansätzen (u.a. der Genomanalyse) vorangetrieben werden. Projektbeteiligte sind Partner von TextGrid, so dass die Kontinuität in diesem Zusammenhang gewährleistet ist.

Neben der kollaborativen Entwicklung konkreter Werkzeuge setzte TextGrid auf ständigen Austausch mit internationalen Infrastrukturprojekten. Da die vielfältigen internationalen Aktivitäten im Bereich der e-Humanities zu Projektbeginn nicht absehbar waren, wurden während der Projektlaufzeit die zu knapp kalkulierten Reisemittel aufgestockt. Dadurch konnten wichtige Kontakte geknüpft, mögliche Synergien identifiziert und vereinzelt bereits gemeinsame Aktivitäten mit internationalen Projekten und Initiativen angebahnt werden (dazu zählen u.a.: eSciDoc, Interedition, DARIAH, CLARIN, Perseus Digital Library, Bamboo). Für die Nachhaltigkeit von TextGrid ist die weitere internationale Vernetzung - auch die technische - durch Interoperabilitätsstandards und durch gemeinsame Entwicklungen essenziell.

Die Einhaltung und Weiterentwicklung von Interoperabilitätsstandards wurden von TextGrid außerdem durch die Mitarbeit in internationalen Initiativen und Gremien wie dem TEI-Consortium, dem Open Grid Forum (OGF) oder der IEEE vorangetrieben.

II. eingehende Darstellung

1. des erzielten Ergebnisses

a) AP-übergreifend: Die TextGrid-Architektur

Bei der Konzeption der TextGrid-Architektur standen drei Aspekte im Vordergrund:

1. Die Bedienung soll so einfach und intuitiv wie möglich sein.

2. Es muss mit einfachen Mitteln möglich sein, TextGrid den individuellen Bedürfnissen anzupassen.

3. TextGrid soll mit bestehenden Infrastrukturen und Werkzeugen zusammenarbeiten können.

Um der ersten Anforderung zu genügen, wurde und wird bei der Konzeption und Gestaltung der Benutzeroberfläche ein hoher Aufwand betrieben. Seit September 2008 laufen ausführliche Testreihen unter Beteiligung zahlreicher Fachwissenschaftler aus unterschiedlichen Disziplinen. Diese Testreihen werden intensiv evaluativ begleitet: Im Rahmen einer Dissertation wird ein entsprechendes Konzept entwickelt und erprobt, das Bildschirmaufzeichnungen, standardisierte Fragebögen sowie individuelle, ausführliche Leitfadeninterviews mit den Fachwissenschaftlern kombiniert. Darüber hinaus wurde versucht, die Komplexität des Grids in den tieferliegenden Architekturschichten zu kapseln. Für die Nutzer ist es in der Regel von wenig Interesse, auf welchen Servern die Daten liegen, wohin sie aus Gründen der Ausfallsicherheit und Performance repliziert werden, solange sie – unter Berücksichtigung der jeweils festgelegten Zugriffsrechte – zuverlässig verfügbar und durchsuchbar sind. Um dem zweiten und dritten Kriterium zu genügen, setzt TextGrid auf offene Standards und eine Service-orientierte Architektur (Service Oriented Architecture, SOA), ein modulares System verteilter, plattformunabhängiger Open-Source-Komponenten, die als Webservices angesprochen werden. Vorhandene Programme lassen sich relativ einfach in die TextGrid-Infrastruktur einbinden, indem sie als Webservices gekapselt werden. Der Zugriff erfolgt über Internet-Protokolle, so dass es unerheblich ist, auf welchem Server und auf welcher Plattform (Windows, Linux etc.) die Services vorgehalten werden. Lediglich Adresse und Parameter müssen bekannt sein.

Die TextGrid-Architektur gliedert sich, wie in Abbildung 2 dargestellt, vereinfacht in zwei Produkte: das TextGridLab, ein umfangreicher Werkzeugkasten aus Services und Tools zur wissenschaftlichen Textbearbeitung und -analyse, sowie das *TextGridRep*, ein Repository, in dem die über das Lab erzeugten Daten und Metadaten im Grid gespeichert, archiviert und indiziert werden sowie zu Verwaltung, Abruf und Analyse vorgehalten werden.

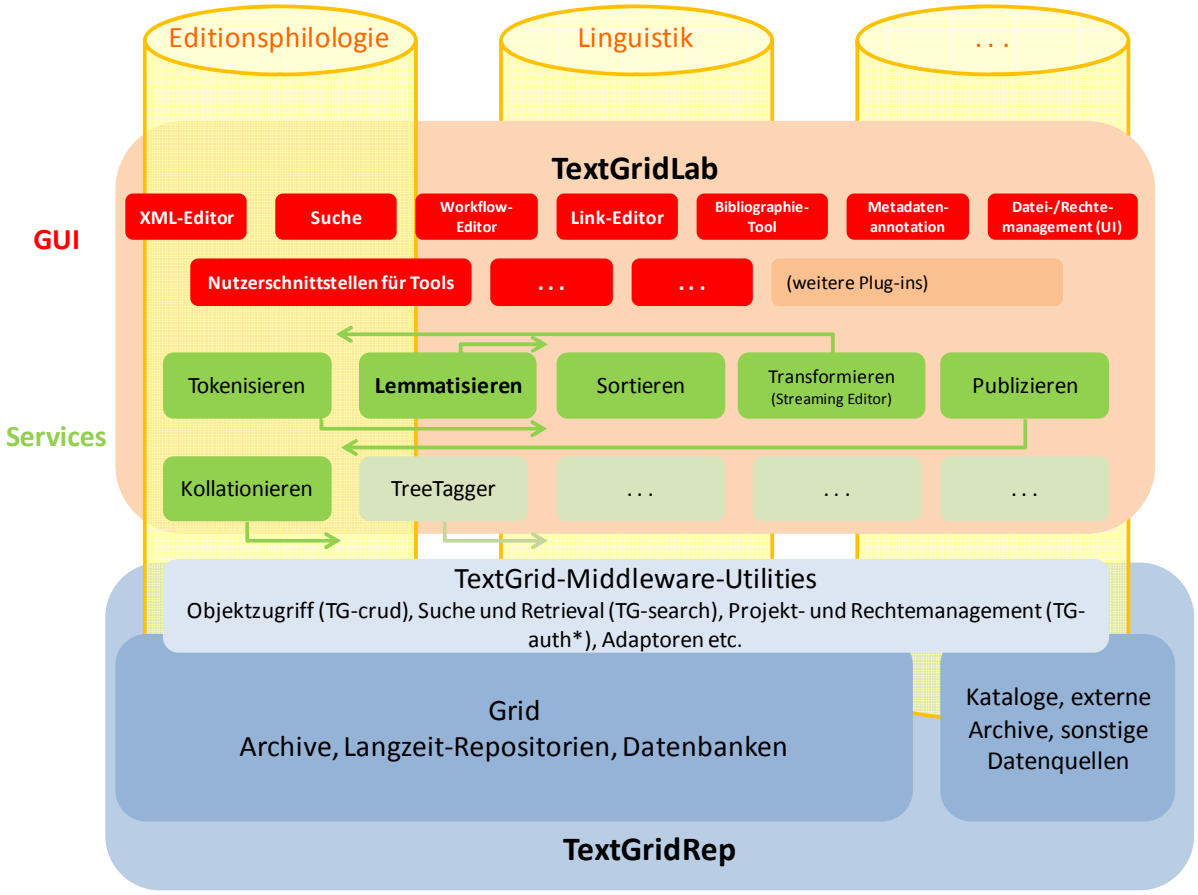


Abbildung 2: Überblick über die TextGrid-Architektur

Auf der untersten Ebene im TextGridRep befinden sich die Archive, in denen die Daten vorliegen, also Metadaten und die eigentlichen Inhalte. In TextGrid sind diese Daten auf verschiedene Gridknoten verteilt und die Architektur erlaubt sowohl die Nutzung weiterer Gridknoten aus dem D-Grid als auch die Anbindung externer, nicht notwendigerweise gridifizierter Archive und Datenquellen. Die *Utilities* der darüber befindlichen TextGrid-Middleware-Schicht bilden einheitliche Schnittstellen (für Authentifizierung, Suche, Schreiben/Lesen/Löschen etc.) zu den Gridressourcen und Archiven, die von den eingesetzten Implementierungen und Techniken abstrahieren und Webservice-Schnittstellen für die darüber befindliche Schicht anbieten. Dies wurde bewusst so gewählt, damit Serviceprogrammierer von den einfacheren und zahlreichen Webservice-Bibliotheken profitieren können.

Zusammen mit einigen Services aus der darüber liegenden Schicht bilden die Archiv- und die Middleware-Schicht das **TextGridRep**(ository), das eine strukturierte Speicherung und langfristige Aufbewahrung von wissenschaftlichen Daten im Grid ermöglicht. Texte werden dabei automatisch mit Metadaten verknüpft und semantisch indiziert. Neben Texten können weitere Dokumente, z.B. Bilddateien von Manuskript-Scans, ebenfalls mit Metadaten und anderen Objekten verknüpft werden.

Eine Ebene darüber befinden sich die Services. Hierbei handelt es sich um eine Reihe von allgemeinen und genrespezifischen Tools, die mit der Middleware interagieren und von der

obersten Ebene, der Benutzeroberfläche, interaktiv angesprochen werden können. Andockpunkte für eine Erweiterung der Funktionalitäten von TextGrid sind hauptsächlich hier vorhanden, denn es können prinzipiell beliebige Webservices eingebunden werden, mit und ohne Zugriff auf die Utilities. Die oberste Ebene wird durch die Benutzeroberfläche gebildet. Im Projekt wird dafür eine Rich-Client-Technologie auf Basis von Eclipse eingesetzt. Neben rein interaktiven Tools sind hier die Benutzeroberflächen zu den einzelnen Services, aber auch zum Zugriff auf die Utilities vereinigt und integriert. Somit wird dem nicht technisch versierten Benutzer ohne Grid-Erfahrung ein intuitiver und interaktiver Zugang zu Gridtechnologien ermöglicht. Die beiden oberen Ebenen werden zusammengenommen als **TextGrid-Lab**(oratory) bezeichnet, weil sie über die Basisdienste des TextGridRep hinaus und auf diese aufbauend die tatsächlichen Werkzeuge für die Arbeit des Benutzers bieten.

Fachdisziplinen und Projekte, aber auch einzelne Nutzer können im TextGridLab vorhandene Module zu spezifischen Workflows kombinieren und die existierenden Tools durch eigene Programme ergänzen. Das Gleiche gilt auch für die Komponenten der Benutzeroberfläche. Eigene Datenbestände lassen sich bei Bedarf vollständig in die Grid-basierte Storage-Infrastruktur des TextGridRep integrieren. Auf diese Weise lassen sich fachspezifische „Säulen“ (siehe Grafik) errichten, die den jeweiligen fachwissenschaftlichen Ansprüchen in besonderem Maß genügen – im Sinne des oben erwähnten wissenschaftlichen Nutzens.

b) AP1: Inhaltliche Studie mit Empfehlungen über die Nachnutzbarkeit internationaler Editionstools

In diesem Arbeitspaket wurden international relevante Werkzeuge aus den Anwendungsbereichen Publishing, Text Processing, Text Retrieval und Linking sowie Workflow Tools sichten, auf ihre Nutzbarkeit für TextGrid überprüft und relevante Tools auf ihre Verwendbarkeit in TextGrid getestet.

Generelle Zeitplanung:

#	Inhalt	Beginn (Monat)	Ende (Monat)
M1.1	Text Processing	1	4
M1.2	Linking	5	8
M1.3	Text Retrieval	9	12
M1.4	Publishing	13	16
M1.5	Management von Workflow, Access, Kommunikation und Nutzer	17	20
M1.6	Ontologien	21	24

Milestones und Deliverables

R1.1 / M1.1 Text Processing

Bericht öffentlich verfügbar seit 30.7.2006.[15]

Im Rahmen von AP 1 wurden im 1. Halbjahr 2006 zahlreiche Literaturbeiträge und existierende Werkzeuge für die wissenschaftliche Textverarbeitung gesichtet und auf ihre Relevanz für TextGrid hin evaluiert. Dieser Bericht schildert die wichtigsten Ergebnisse für die im Projektantrag genannten Module Tokenizer, Kollationiere, Lemmatisierer, Sortierer, XML-Editor und OCR sowie einige für TextGrid generell wichtige Ergebnisse der Literaturrecherche.

In der Folge dieser in R1.1 dokumentierten Literaturrecherche wurden zentrale Werkzeuge identifiziert und näher analyisert. Darüber hinaus ergaben sich dabei auch für TextGrid wertvolle Hinweise auf frühere Softwareprojekte im eHumanities-Umfeld, besonders ARCHway und GATE, die sich zwar noch nicht die Möglichkeiten des Grid Computings zu Nutze machten, aber ebenfalls einen modularen Ansatz verfolgten und z.T. ähnliche Frameworks zur GUI-Gestaltung einsetzten. TextGrid konnte diese Erkenntnisse in die eigene Entwicklungsarbeit aufnehmen.

R1.1/M1.2 Linking

Bericht öffentlich verfügbar seit 22.12.2006.[16]

Für wissenschaftliche Texte enthalten regelmäßig zahlreiche Verweise, sowohl innerhalb des Textes selbst als auch auf externe Quellen. Elektronische Werkzeuge erlauben es, die referentielle Integrität solcher Verweise sichderzustellen, auch wenn der Autor noch Änderungen am Text vornimmt, die einen anderen Seitenumbruch bewirken, oder ganze Textteile umstellt.

Tatsächlich werden unter dem Oberbegriff *Linking* eine Sammlung von heterogenen Aufgaben zusammengefasst, die lediglich gemein haben, dass sie eine Information oder eine Stelle in einem Dokument mit einer anderen verknüpfen. Der für M1.2 erstellte Bericht betrachtet den State of the Art für Link-Editoren aus den Bereichen Text-Annotation, in dem Verweise zwischen Texten verwaltet werden, Manuskriptannotation, bei der wenigstens eine Seite der Verknüpfung in ein Bilddokument – typischer-, aber nicht notwendigerweise ein Scan einer Manuskriptseite – verweist, sowie Bibliographieverwaltung.

Der Bericht hat diese Fragestellungen analysiert und daraus wiederum Anforderungen für die TextGrid-Entwicklung, insbesondere für AP2, abgeleitet.

R1.3 / M1.3 Text Retrieval

Bericht öffentlich verfügbar seit 30.5.2007[17]

Wenn die Anwender aus den in TextGrid verfügbaren Daten Nutzen ziehen sollen, benötigen Sie spezialisierte Werkzeuge, um in den TextGrid-Ressourcen zu suchen, die über eine Volltextsuche mit

[15] http://www.textgrid.de/fileadmin/TextGrid/reports/TextGrid_Report_1_1.pdf

[16] http://www.textgrid.de/fileadmin/TextGrid/reports/TextGrid_Report_1_2.pdf

[17] http://www.textgrid.de/fileadmin/TextGrid/reports/TextGrid_Report_1_3.pdf

regulären Ausdrücken und booleschen Verknüpfungen hinausgehen. Orthographische Varianten, Transkriptionen und Transliterationen, Anforderungen an den Kontext der Fundstellen sowie Einschränkungen, die nur bei Auswertung der Metadaten eines Textes berücksichtigt werden können, machen deutlich, dass die Recherchekomponente in TextGrid hochkomplexere Anfragen bedienen muss.

R1.3 hat diese Herausforderungen herausgearbeitet und einen Überblick über die verfügbaren einschlägigen Programme und Softwarebibliotheken gegeben. Diese Einforderungen wiederum sind insbesondere in das Recherchetool (s. AP2) eingeflossen.

R1.4 / M1.4 Publishing

Bericht öffentlich verfügbar seit 30.8.2007[18]

Publikationen von wissenschaftlichen Editionen erfolgen heute meist in einer oder mehreren von drei Formen: In traditioneller Weise als Print-Edition, als CD-ROM- bzw. DVD-Edition sowie zunehmend auch oder sogar vor allem als Online- oder Web-Edition. Aus Sicht von TextGrid sind hierbei besonders die Print- und die Web-Publikation von Interesse, weil beide unmittelbar aus den in TextGrid gespeicherten Daten erzeugt werden sollen. Bei CD-ROM-Editionen werden die Daten zwangsläufig aus TextGrid exportiert und in ein für „TextGrid-fremde“ Programme geeignetes Format überführt.

R1.4 konzentriert sich vor diesem Hintergrund auf die Print- und die Web-Publikation. Der Bericht arbeit wiederum die für sehr spezifischen textwissenschaftlichen Anforderungen heraus und blickt besonders auf die Print- und Web-Bedürfnisse kritischer Editionen. Daraus werden wiederum konkrete Anfoderungen für das Publishing-Tool in AP2 extrahiert.

R1.5/M1.5 Management von Workflow, Access, Kommunikation und Nutzer

Bericht öffentlich verfügbar seit 10.1.2008.[19]

In jedem nichttrivialen textwissenschaftlichen Projkt, z. B. einer kritischen Edition, durchlaufen die Ausgangsdaten neben der manuellen Anreicherung und Kommentierung durch den Editor sehr viele automatisierte Transfomationen, ehe daraus die fertige gedruckte oder elektronische Edition wird – erfasste Texte werden in Tokens zerlegt, Varianten eines Textes werden kollationiert, extrahierte Register werden sortiert usw. Diese Arbeitsschritte müssen typischerweise sehr häufig und immer wieder in der gleichen Reihenfolge durchlaufen werden, etwa nach Korrekturen in den Ausgangsdaten oder wenn die Parameter eines der zuvor genutzen Werkzeuge variiert werden sollen. Es bilden sich also festgelegte Arbeitsprozesse heraus, die sich nur in den Daten, auf die sie angewandt werden, und der Konfiguration der beteiligten Werkzeuge unterscheiden.

[18] http://www.textgrid.de/fileadmin/TextGrid/reports/TextGrid-R1_4_Publishing.pdf

[19] http://www.textgrid.de/fileadmin/TextGrid/reports/TextGrid_Report_1_5.pdf

R1.5 untersucht dazu in seinem ersten Teil, welche Unterstützung potentielle TextGrid-Anwender erwarten, um diese repetitiven Arbeiten ergonomisch, nachvollziehbar und weniger fehleranfällig zu machen.

Workflows betreffen üblicherweise Daten, die räumlich verteilt sind und oftmals zu unterschiedlichen Projekten mit unterschiedlichen Rechten gehören. In diesen Fällen ist die Kommunikation aller Mitarbeitenden essentiell, damit sich die Beteiligten nicht gegenseitig behindern oder gar Daten zerstören. Der Bericht betrachtet deshalb, wie der notwendige Informationsfluss in einem in TextGrid umgesetzten Projekt realisiert werden kann.

Schlussendlich studiert R1.5 die für TextGrid benötigte *Authentication and Authorization Infrastructure (AAI)*. Der Bericht diskutiert die Anforderungen an die Nutzerverwaltung und AAI , die sich auch im Hinblick auf Usability in einer Grid-Community und insbesondere in TextGrid stellen, und gibt einen Überblick über die verfügbaren Lösungen.

Basierend auf dieser Analyse definert R1.5 die beste Strategie für TextGrid, insbesondere auch für seine zweite Projektphase.

M1.6 Ontologien

Bericht öffentlich verfügbar seit 27.4.2008.[20]

Semantische Technologien haben spätestens seit dem Schlagwort *Web 2.0* , das u. a. semantisch angereicherte Informationen bieten soll, sehr viel Aufmerksamkeit erregt. Um Daten semantisch auswerten zu können, bedarf es grundsätzlich geeigneter Ontologien, weil nur in deren Kontext aus den Daten echte Informationen werden.

Semantische Anwendungen und speziell auch Ontologien sind für die TextGrid -Zielgruppe als Werkzeug und als eigenständiger Forschungsgegenstand gleichermaßen relevant. In diesem Report betrachten wir allerdings ausschließlich technische Lösungen zum Umgang mit Ontologien und für die Verwaltung von semantischen Informationen.

R1.6 definiert die Begriffe *Ontologie*, *Referenzmodell* und *Referenzarchitektur* im TextGrid-Kontext und stellt dann die einschlägigen Normen wie RDF und Topic Maps sowie die zugehörigen Abfragesprachen vor. R1.6 dokumentiert dann die in TextGrid entwickelte Lösung der Abbildung projektspezifischer Ontologien zur Textauszeichnung für den Zweck korpusübergreifender Recherchen.

Bemerkung: Die Fragen, wie TextGrid die Arbeit an bzw. die Nutzung von Ontologien als textwissenschaftliche Objekte unterstützen kann, wurden bereits ausführlich in Report 5.1 [BÜ08] diskutiert.

c) AP2: Das TextGridLab

Ziel dieses Arbeitspaketes war die Entwicklung einer gridfähigen Workbench für die Erstellung, Bearbeitung, Annotation und Analyse von XML-kodierten Textdaten. Diese Workbench dient dazu, die sehr großen Datenmengen, die durch die Bildung einer virtuellen Nationalbib-

[20] http://www.textgrid.de/fileadmin/TextGrid/reports/TextGrid_Report_1_6_Ontologien.pdf

liothek anfallen, den Textwissenschaftlern unter einer einheitlichen Arbeitsoberfläche zugänglich zu machen und die sehr rechenzeitintensiven Anwendungen zur Annotation und Analyse von Texten zu verteilen. Die internationale Community von Textwissenschaftlern hat in den letzten Jahrzehnten eine Reihe von Programmen entwickelt, die für kleinere bis mittlere lokal verfügbare Korpora verwendet wurden und die einige ausgewählte der im folgenden beschriebenen Funktionen zur Verfügung stellen.

Auf die TextGrid-Architektur (vgl. Abbildung 2) bezogen, ist das Arbeitspaket für das *TextGridLab* zuständig, also für die Schichten des User Interface und der fachwissenschaftlichen Services.

Dabei ist eine Reihe von Tools entstanden, die auf die in AP3 entwickelten Utilities zurückgreifen und fachspezifische Funktionalitäten anbieten. Dies sind zum einen Services zum nicht-interaktiven Verarbeiten von Texten *(Streaming-Tools)* – die etwa die Wörter in einem Eingabe-Dokument um Lemmainformationen anreichern – und zum anderen interaktive Tools, mit denen die Nutzer die Dokumente im TextGrid analysieren und bearbeiten und die Services und Utilities ansprechen können.

Ausführliche Beschreibungen finden sich in den TextGrid-Reports 2.1[21], 2.2[22] und 2.3[23].

Streaming-Tools

Diese nicht-interaktiven Tools werden als Batchprozesse gestartet und verfügen ggf. über eine interaktive GUI-Komponente zur Konfiguration. Darüber hinaus lassen sie sich über eine SOAP- oder REST-Schnittstelle direkt ansprechen und so zu Workflows arrangieren und in andere Applikationen integrieren.

- Der **Tokenizer** zerlegt einen Text in eine Folge logischer Einheiten (Tokens), üblicherweise Wörter und Satzzeichen. Diese werden durch Anfang- und Endemarkierungen gekennzeichnet. Die hierbei zu verwendenden Elemente lassen sich ebenso wie vordefinierte Tokens wie Abkürzungen, Eigennamen oder reguläre Ausdrücke (z.B. für Datumsangaben) in der Tool-Konfiguration definieren.
- Der **Lemmatizer** analysiert - abhängig vom jeweils hinterlegten Vokabular - einzelne Wortformen morphologisch und liefert als Ausgabe das zugehörige Lemma (die Wortform wird auf ihre grammatische Grundform zurückgeführt), die zugehörige Wortart *(Part of Speech)* und weitere morphologische Merkmale (Numerus, Genus etc.).
- Mit Hilfe des **Sortiertools** kann ein Anwender gegebene Folgen von Zeichenketten gemäß kulturellen und fachlichen Erwartungen anordnen. Die dabei zu berücksichtigenden Sortierschlüssel und Sortierkonventionen kann der Nutzer frei vorgeben. Zur Vereinfachung der Handhabung erlaubt das Sortiertool aber auch, sich direkt auf einschlägige nationale und europäische Standards zu beziehen.

[21] http://www.textgrid.de/fileadmin/TextGrid/reports/TextGrid_Report_2_1.pdf

[22] http://www.textgrid.de/fileadmin/TextGrid/reports/TextGrid-R2.2_ToolsII.pdf

[23] http://www.textgrid.de/fileadmin/TextGrid/reports/Report_2.3_final.pdf

- Der **Streaming Editor** ermöglicht Transformationen von Dateien aufgrund von Regeln, z.B. automatisierte Anreicherung potentiell unstrukturierter Texte mit XML-Strukturen. Bei den Eingabedaten muss es sich nicht notwendig um XML handeln, sondern es können beliebige Textformate (etwa OCR-Rohdaten, reiner Text) verarbeitet werden. Die Ausgabe kann, muss aber nicht XML sein.
- Der **Text Publisher Web** dient Projekten dazu, Ihre Daten und Ergebnisse im Rahmen eines Internetauftritts zu präsentieren. Hierfür stellt TextGrid vorgefertigte Komponenten und eine Schnittstelle zur Datenanbindung an das TextGridRep (s. Architektur) zur Verfügung.

Interaktive Tools

Die interaktiven Tools werden in einer gemeinsamen Benutzungsoberfläche (auf der Basis von Eclipse) zusammengefasst – dem TextGrid-Client *TextGridLab* – und sind dort eng miteinander verknüpft. In der Regel gibt es dabei eine Teilung zwischen dem für die Benutzerinteraktion verantwortlichen Teil, der auch im Client implementiert ist, und einem Backend auf Web-Service-Basis, das die Daten verwaltet bzw. die eigentliche Arbeit erledigt: Etwa ein Workflow-Editor zum Bearbeiten und Definieren der Abläufe und ein Workflow-Enactor (in der Middleware), der die Workflows dann ausführt.

- Der **XML-Editor** bildet eine der zentralen Komponenten des TextGrid-Clients: Er ist ein interaktives Tool, das zur Neuerfassung wie zur nachträglichen (manuellen) Annotation von Texten in XML-basierten Formaten dient. Darüber hinaus wird er mit anderen TextGridLab-Tools integriert und kann so etwa für Streaming-Tools zur Ergebnisdarstellung verwendet werden.
- Das **Recherchetool** dient der Recherche und dem Retrieval von Struktur- und Metadaten. Eine RDF-basierte, semantische Suche ist ebenfalls möglich. Es integriert Werkzeuge zum Zugang, zur Suche und zum Blättern in den im Grid gespeicherten Texten und Metadaten oder zu Teilmengen davon. Zusätzlich können diverse statistische Analysen durchgeführt werden: In wie vielen Texten wird ein bestimmtes Wort verwendet? Wie viele Autoren verwenden die Phrase xy? Existieren zu einem Text noch weitere Versionen, welche Texte verweisen darauf?
- Der **Workflow-Editor** ist ein interaktiver Bestandteil der Benutzerumgebung und erlaubt es, die Automatisierung von Arbeitsabläufen (Workflows) zu definieren (orchestrieren). Wie oben erwähnt, werden fertige Workflows von einem Enactor in der Middleware ausgeführt.
- Die **Metadaten-Annotation** ist ein generisches Werkzeug, das dazu dient, strukturierte Daten über eine definierbare Maske zu erfassen und an einer frei bestimmbaren Position in einer Datei einzufügen / abzuspeichern. Dazu gibt es zwei Tools: einen *Maskengenerator*, der es dem Projektleiter / Techniker erlaubt, eine Maske vorzudefinieren, und einen *Masken-Editor*, der die vordefinierte Maske darstellt, die Eingabe überprüft, und die strukturierten Daten an einer bestimmten Position in einer Datei und im Grid speichert.

- Der **grafische Linkeditor** (man könnte ihn auch Topographie-Editor nennen) hat die Aufgabe, den Editor bei der Alignierung von Text und Bildelementen etwa aus dem Digitalisat einer Handschrift zu unterstützen. Ziel ist die Erstellung einer Ausgabedatei, die die Textelemente und die topographische Beschreibung enthält. Dabei sollen wie üblich Standards zur Anwendung kommen.
- Der **Link-Editor Text** ist eine Eingabehilfe für Links in XML-Dateien und verbindet Elemente von Recherchetool und XML-Editor. Benutzer haben die Möglichkeit, Links zu beliebigen Elementen in TextGrid-Dokumenten in die aktuelle TEI-Datei einzugeben, in dem über Recherchetool, Projektbrowser, XML-Outline-Darstellung oder direkt im XML-Editor das Zieldokument bzw. -element gewählt und über Kontextmenü oder ggf. Drag & Drop eine entsprechende Funktion ausgewählt wird; es wird dann die passende URI samt Fragment zum Einfügen in das Quelldokument des Links generiert.
- Der **Kollationierer**, ein interaktives Werkzeug, unterstützt den Philologen beim Vergleich von mehreren Dokumenten auf ihre Unterschiede hin. Er dient der Analyse und Annotation von Entstehungs- und Überlieferungsvarianten in Texten. Die Textüberlieferung kann ausgesprochen komplex sein, da sie Überarbeitungen, Auslassungen, Einfügungen und Umstellungen über weite Textstrecken hinweg enthalten kann, die wiederum in allen Textzeugen unterschiedlich ist. Gerade das automatische Erkennen von Textveränderungen, die zusammengehören, stellt eine wirkliche Herausforderung an die Algorithmen zur Varianzdetektion dar. Hinzu kommen disziplinspezifische Anforderungen, etwa diejenigen Varianten, die aus philologischer Sicht unbedeutend sind, von solchen zu unterscheiden, die notierungswürdig sind, und die Markierung dieser Unterschiede möglichst benutzerfreundlich zu gestalten.
- Ein **Navigationstool** ermöglicht das einfache Browsen durch die eigenen Objekte bzw. durch Materialien aus den Projekten, an denen man teilnimmt.
- Die **Projektverwaltung** ermöglicht die Erstellung neuer Projekte durch den Projektleiter und die Zuordnung von weiteren Personen (in bestimmten Rollen) zu einem Projekt. Außerdem sind andere zentrale Punkte zur Konfiguration von Projekten hier vereint.

Ein beträchtlicher Aufwand floss auch in die Entwicklung von Basisfunktionalität des TextGridLab selbst. So wurde insbesondere ein Modell für die Objekte des Repositories entwickelt und ein Frontend für den Zugriff auf TextGrid-Objekte geschrieben, das die TextGrid-spezifischen Utility-Schnittstellen auf das Eclipse-eigene Modell zum Dateizugriff abbildet. So ist es möglich, bereits existierende Eclipse-Komponenten wie etwa den im Rahmen der *Eclipse Web Standard Tools*[24] entwickelten XML-Schema-Editor ohne weitere Veränderung in das TextGridLab einzubinden.

[24] Eclipse Web Standard Tools: http://www.eclipse.org/webtools/wst/main.php

Milestones und Deliverables

M 2.1 Tokenizer, Workflow-Editor

M2.2 Lemmatisierung, XML Editor, Rich Client Platform (GUI)

M 2.3 Recherchetool, Streaming-Editor I, Datei-/Rechtemanagement, Metadaten-Annotation, grafischer Link-Editor, Bild-Segmentierung, Link-Editor Text, Bibliographietool, Sortieren

M 2.4 Streaming Editor II, Kollationierung

M 2.5 Text Publisher (Print), Text Publisher (Web), OCR

R 2.1 TextGrid-Tools I

R 2.2 TextGrid-Tools II

R 2.3 Dokumentation

d) AP3: Das TextGridRep

Das dritte unter unseren Arbeitspaketen hat die Middleware-Utilities des *TextGridRep* (vgl. Abbildung 2) entwickelt. Ziel war die Anbindung der in den anderen Arbeitspaketen erstellten Softwarebausteine über eine TextGrid-spezifische Middleware-Infrastruktur an die D-Grid Grid-Infrastruktur. Hierzu musste eine Schnittstelle zu den Tools spezifiziert und implementiert werden sowie eine Schnittstelle zu den Grid-Diensten.

Hierbei wurden folgende Komponenten der TextGrid-Middleware (TextGridRep) aufgebaut:

- *TG-auth** bietet eine föderierte Authentifizierung und feingranulare Autorisierung von einzelnen Ressourcen, einschließlich der Verwaltung von Projekten und Rollen.
- *TG-crud* für die Koordination von Create-, Read-, Update- und Delete-Operationen auf Ressourcen im Grid. Diese Komponente ist zentral, sie aktualisiert bei den genannten Operationen auch die verschiedenen Hilfsdaten für TG-search und kann dazu konfigurierbare *Adaptoren* anstoßen.
- *TG-search* erlaubt die performante Suche sowohl im Volltext als auch in der Struktur der XML-Dokumente. Hierzu wird je eine Relations-, Metadaten- und Strukturdatenbank verwendet und über TG-crud verwaltet.
- *TG-log* stellt einen zentralen Logging-Service bereit.
- *Workflow Manager:* Eine Workflow Engine, der Grid Workflow Execution Service (*GWES*), wurde in die Middleware integriert und ist vom Workflow Editor des TextGridLabs ansprechbar.
- *Ontology Manager:* Konzepte hierfür wurden in AP5 erarbeitet.

Die Konzepte und Beweggründe, die hinter der Realisierung dieser Dienste standen, lassen sich wie folgt zusammenfassen:

- *Kapselung der Dienste der Grid-Middleware*: es erfolgte früh eine Festlegung auf die Verwendung der GRID-Middleware Globus Toolkit 4 (*GT4*). GT4-Dienste sind WSRF-basiert, jedoch standen zu Projektbeginn nur unvollständige

Implementierungen von WSRF-Bibliotheken zur Verfügung. Außerdem sollte die Implementierung und Integration von neuen Diensten für Entwickler vereinfacht werden. Diese beiden Gründe waren die Motivation zur Entwicklung des TG-crud Services, der eine einfache SOAP-basierte Web-Service-Schnittstelle anbietet, Speicherung der Nutzdaten aber per JavaGAT in einem GT4-basierten Grid ermöglicht. Somit können oberhalb des CRUD statt WSRF die deutlich einfacher zu implementierenden SOAP Webservices verwendet werden.

- *Individuelles Metadatenschema:* Um die Belange der Geisteswissenschaften zu repräsentieren, wurde ein TextGrid-spezifisches Metadatenschema entwickelt, das ein die Nutzdaten beschreibendes Metadatenformat spezifiziert. Metadaten werden redundant (im GRID und einer XML-Datenbank) gespeichert und von *TG-crud* verwaltet.
- *Suche auf der TextGrid-Kernkodierung*: Textdaten werden in den Geisteswissenschaften häufig im XML-Dialekt *TEI* kodiert. TEI sieht generische Strukturelemente vor, so dass ein Benutzer für eine Suche auf TEI-Daten deren spezielle Struktur kennen muss. Um dieses Problem zu lösen, wurde eine Kernkodierung (*baseline encoding*) für exemplarische Textsorten entwickelt, die eine generische strukturelle Suche mit der Utility *TG-search* erlauben. Die Repräsentation dieser kernkodierten Daten erfolgt ebenfalls getrennt vom GRID in einer XML-Datenbank. Auch diese Daten werden von *TG-crud* verwaltet.
- *Relationen*: Objekte in TextGrid können zueinander in Beziehungen stehen. Diese Beziehungen werden ebenfalls von *TG-crud* verwaltet und in einer Relationsdatenbank niedergelegt.
- *Föderierte Authentifizierung*: Bei der Authentifizierung setzt *TG-auth** die SAML-basierte Software *Shibboleth* ein. Somit können die Angehörigen derjenigen Hochschulen in Deutschland, die Teil der *DFN-AAI* sind, ohne eine vorherige TextGrid-spezifische Registrierung sich mit ihrem örtlichen Campus-Account im *TextGridLab* anmelden. Für alle anderen Benutzer (Mitglieder anderer Institutionen, Privatnutzer, oder Nutzer aus dem Ausland) steht der TextGrid-eigene Community Authentication Server zur Verfügung.
- *Feingranulare Zugriffskontrolle*: Die Autorisierungskomponente von *TG-auth** verwendet Role-based Access Control (*RBAC*) zur Zugriffskontrolle. Auf den Ressourcen sind die Zugriffsrechte somit nicht direkt festgelegt, sondern über Rollen, die wiederum Benutzer innehaben und einzeln je nach Arbeitskontext aktivieren können. Die Dienste von *TG-auth** stehen ebenfalls als SOAP Web Service zur Verfügung und können sowohl von *TG-crud* als auch direkt vom *TextGridLab* abgefragt werden.

Weitere zeit- und entwicklungsintensive Arbeiten dienten der konzeptuellen Vorbereitung und der Stabilisierung der dargelegten Architektur. So wurden im ersten Abschnitt des Projekts die TextGrid-Middleware-Fileservices als Vorgänger des CRUD implementiert. Im Zuge dessen erfolgten Spezifizierung der WebService-Schnittstellen, Interoperabilitätstests, In-

tegration verschiedener Komponenten, Evaluation von Globus Toolkit 4, WSRF sowie GAT/SAGA. Nach der Spezifizierung der gegenwärtigen Architektur wurden umfangreiche Arbeitsressourcen für die Implementierung der Konzepte, der Interaktion zwischen den einzelnen Komponenten, die Schaffung von Test- und Produktivplattformen und die Wartung dieser Installationen verwendet.

Milestones und Deliverables

Nach der Konzeption der aktuellen Architektur wurde ein Umbau der Report- und Milestone-Struktur notwendig. Die Anpassung wurde aufgrund der Erkenntnisse durchgeführt, die sich aus dem ersten Jahr ergaben, mit den folgenden Zielen:

a. Einbettung der Erfahrungen und im Rahmen der Projektarbeit getroffener Entscheidungen.
b. Nutzer-Orientierung - die zwei neuen Berichte R 3.5 und R 3.6 dokumentieren die Nutzung der TextGrid-Middleware für unterschiedliche Nutzergruppen (jeweils Anbindung von Archiven und Werkzeugen).
c. Engere Abstimmung zwischen den Arbeitspaketen - um eine enge Kooperation und einen reibungslosen Gesamtablauf zu garantieren, wurden Module der Middleware gemeinsam mit der Entwicklung der Werkzeuge (AP 2) umgesetzt und iterativ verbessert. Die diesbezüglichen Aktivitäten in der ursprünglichen Struktur (siehe R 3.7 und M 3.5) wurden daher vorgezogen.

Diese Anpassung wurde dem Projektträger unterbreitet und von diesem akzeptiert. Die angepassten Milestones und Reports werden im Folgenden aufgelistet.

M 3.1 Ermittlung der Middleware-Anforderung aus den anderen APs

M 3.2 Spezifikation der von den Projektpartnern zu verwendenden Middleware-Schnittstellen

M 3.3 Implementierung eines Prototyps der TextGrid-Middleware-Plattform mit Anbindung an die IP

M 3.4 Unterstützung und Evaluation bei der Anwendung des Prototyps für Musterapplikationen (AP4)

M 3.5 Implementierung einer Produktivversion der TextGrid-Middleware

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 3.2 Architektur für die TextGrid-Middleware

R 3.3 Spezifikation aller von der TextGrid-Middleware zu bedienenden Grid-Schnittstellen / Standards zur Anbindung an die IP (Version 1)

R 3.4 Middleware-Tests, unter Berücksichtigung der Werkzeuge (AP2) und Musterapplikationen (AP4)

R 3.5 User-Manual zur Anbindung von Werkzeugen an die TextGrid-Middleware

R 3.6 User-Manual zur Installation eines Datengrid-Knotens (für assoziierte Textarchive)

e) AP4: Entwicklung der Community Muster-Applikation

Das Projekt soll die Testmaterialien bereitstellen, an denen die Nutzbarkeit der grid-fähigen Werkzeuge für die Arbeitsbereiche Publishing, Textverarbeitung, Text Retrieval und Linking in der erforderlichen Komplexität erprobt und in Tests mit Anwendern unterschiedlicher Profile evaluiert werden kann, so dass die Ergebnisse iterativ in den Entwicklungsprozess der Software zurückfließen.

Als Testmaterialien wurden im Antrag drei unterschiedliche Materialtypen gewählt:

1. Der Typus ‚Wörterbuch' als Modell für stark strukturierte Textdaten, demonstriert am Beispiel von Joachim Heinrich Campes Wörterbüchern.

2. Der Typus ‚Historisch-kritische Edition' als Modell für Textdaten mit besonderen Anforderungen an die Präsentationsform. Als Editionsbeispiel sollte hier die neue Historisch-kritische Edition des Romans ‚Hesperus' von Jean Paul dienen.

3. Der Typus ‚Bilddaten' als Modell für den Datentyp mit hohen Anforderungen an Qualität und Metadatenverwaltung. Als Beispiele bieten sich hierfür einerseits die Bilddateien zu Campes Wörterbüchern (ca. 6.000 Seiten) und die Druck- und Nachlass-Materialien zu Jean Pauls Roman ‚Hesperus' an.

Ergebnisse der Arbeitsmodule:

1. Erstellen der Textgrundlage zur Erprobung der Arbeitsumgebung an stark strukturierten Daten

Im Rahmen von AP4 wurde mittels des Double-keying-Verfahrens der Text von Joachim Heinrich Campes Wörterbuch der deutschen Sprache vollständig elektronisch erfasst und für die weitere EDV-Verarbeitung aufbereitet:

Wörterbuch der Deutschen Sprache. Veranstaltet und herausgegeben von Joachim Heinrich Campe. Erster Theil. A - bis – E. Braunschweig 1807. In der Schulbuchhandlung. XXIV, 1023 S.

Wörterbuch der Deutschen Sprache. Veranstaltet und herausgegeben von Joachim Heinrich Campe. Zweiter Theil. F - bis – K. Braunschweig 1808. In der Schulbuchhandlung. IV, 1118 S.

Wörterbuch der Deutschen Sprache. Veranstaltet und herausgegeben von Joachim Heinrich Campe. Dritter Theil. L - bis – R. Braunschweig 1809. In der Schulbuchhandlung. IV, 908 S.

Wörterbuch der Deutschen Sprache. Veranstaltet und herausgegeben von Joachim Heinrich Campe. Vierter Theil. S - und – T. (Nebst einer Beilage) Braunschweig 1810. In der Schulbuchhandlung. IV, 944 S.

Wörterbuch der Deutschen Sprache. Veranstaltet und herausgegeben von Joachim Heinrich Campe. Fünfter und letzter Theil. U - bis – Z. Braunschweig 1811. In der Schulbuchhandlung. IV, 979 S.

Ferner als Ergänzungsband:

Wörterbuch zur Erklärung und Verdeutschung der unserer Sprache aufgedrungenen fremden Ausdrücke. Ein Ergänzungsband zu Adelung's und Campe's Wörterbüchern. Neue Starkvermehrte und durchgängig verbesserte Ausgabe von Joachim Heinrich Campe. Braunschweig 1813. In der Schulbuchhandlung. XIV, 674 S.

2. Erstellen von geeigneten Evaluationsszenarien für die historisch-kritische Jean-Paul-Edition.

Als vorbereitende Arbeit zu den Evaluationsszenarien wurde ein Mengengerüst der Jean-Paul-Drucke und wichtigsten Jean-Paul-Ausgaben erstellt und eine Magisterarbeit im Fach ‚EDV-Philologie vergeben:

Aylin Chaban: Digitalisierungskonzepte für Drucke in Frakturschrift aus dem 18. Jahrhundert, Magisterarbeit Würzburg, Februar 2007.

Darüber hinaus wurde die vom DFG-Projekt ‚Jean-Paul-Edition' beauftragte Digitalisierung von rd. 8000 Seiten des Jean-Paul-Nachlass an der Staatsbibliothek zu Berlin mit einem Datenvolumen von 681 GB technisch betreut. Über die Nutzung dieser Daten auch als Teil der TextGrid-Testbasis für Bilddaten mit hohen Anforderungen an Qualität und Metadatenverwaltung wurde ein Nutzungsvertrag zwischen der Staatsbibliothek und dem TextGrid-Konsortium, vertreten durch die SUB Göttingen, abgeschlossen.

Im Verlauf der Arbeiten zeichnete sich ab, dass – gegenüber den Vorstellungen, die bei der Beantragung von TextGrid zugrundegelegt wurden – die Edition nicht so zeitgerecht fertiggestellt sein wird, dass sie im Rahmen des TextGrid-Projekts als Testmaterial für Textdaten mit besonderen Anforderungen an die Präsentationsform genutzt werden könnte. Daher wurde auf dem Projekttreffen in Trier im September 2007 beschlossen, das Textbeispiel für den Modellfall ‚Historisch-kritische Edition' auszutauschen gegen eine unter Betreuung von Fotis Jannidis entstandene Edition, deren Text bereits zur Verfügung steht und ohne urheberrechtliche Einschränkungen genutzt werden kann: Bettina Ulrike Bauer, Historisch-kritische Edition des von Maximiliane und Friedmund von Arnim verfassten erläuternden Gedichtes zum Huldigungsbild für König Friedrich Wilhelm IV., Darmstadt 2007.

3. Dokumentation der Tools und Beschreibung des Workflows

Die Dokumentation der Tools wurde im Hinblick auf Überschneidungen mit Reports zu AP2 dorthin ausgegliedert. Auf dem Projekttreffen im November 2008 in Mannheim wurde dazu ferner festgelegt, den Titel des Reports 4.1 für die Beschreibung des Workflows dementspre-

chend anzupassen und umzubenennen in „Zur Digitalisierung von Primärquellen für die TextGrid-Umgebung: Modellfall Campe-Wörterbuch“[25].

4. Entwicklung und Auswertung der Tests der TextGrid-Benutzeroberfläche mit Studierenden

Die TextGrid-Konzepte und das TextGridLab wurden mit Studierenden aus unterschiedlichen Studiengängen und in unterschiedlichen Stadien ihres Studiums in vier Testformen erprobt und für die Entwicklung nutzbar gemacht:

1. Über die Beteiligung am Workshop <philtag n=“6“> im Oktober 2007 in Würzburg (TextGridLab-Demo und Hands-on session zur Wörterbuch.Kodierung mit Video-Mitschnitten). Mitschnitte und Resultate wurden als Modell für ‚TEI Education‘ auf der TEI Jahrestagung „TEI@20“ 2007 in College Park im Rahmen eines Workshops zum Thema ‚TEI-Education‘ vorgestellt. [12 studentische Teilnehmer]

Links:

http://www.phil1.uni-wuerzburg.de/institutelehrstuehle/ institut_fuer_deutsche_philologie/ lehrstuehle/ lehrstuhl_fuer_deutsche_sprachwissenschaft/forschung/

kompetenzzentrum_fuer_edv-philologie/philtag/philtag_n6/

Panel session 3.5 TEI Education (Convenor: Werner Wegstein) http://www.lib.umd.edu/dcr/events/teiconference/program.html#body.1_div.2 sowie die Meeting session der Special Interest Group ‚Education‘ (Susan Schreibman and Werner Wegstein).

2. Durch das Angebot (zwischen August 2007 und Oktober 2008), das Pflichtpraktikum der Aufbaustudiengänge ‚Linguistische Informations- und Textverarbeitung‘ und ‚EDV-Philologie‘ (mit Abschluss M.A.) als betreutes TextGrid-Praktikum im Umfang von jeweils mindestens vier Wochen bei dem TextGrid-Projekt absolvieren zu können [11 Studierende].

3. Durch die Vergabe von Studien- bzw. Diplomarbeiten für den Diplomstudiengang ‚Informatik’ an der Universität Würzburg zu TextGrid-Nutzungsthemen (betreut zusammen mit den Kollegen Wolf von Gudenberg und Seipel). [5 Studierende].

4. Durch die Einbeziehung der TextGrid-Nutzungsthematik in Seminare des Hauptstudiums ‚Deutsche Sprachwissenschaft‘ bzw. von Postgraduate-Studiengängen für den Bereich Kodierungsverfahren, Editionsphilologie und Korpuslinguistik.

5. Aufbereiten der Campe-Wörterbuchdaten zur gemeinsamen Nutzung für das Semantic Grid.

Milestones und Reports

M 4.1 Erstellung der relevanten Corpora (Digitalisierung) als Testbed

[25] http://www.textgrid.de/fileadmin/TextGrid/reports/TextGrid_R4_1.pdf

Das Testkorpus für den Datentyp ‚Bilddaten', Modell für hohe Anforderungen an Qualität und Metadatenverwaltung wurde erstellt aus:

- den Bilddigitalisaten der Campe-Wörterbucher (5698 Bilddateien im Tif-Format, Speichervolumen 537 GB) erweitert durch Bilddigitalisate von Campes Vorarbeiten und Vorstufen zum ‚Wörterbuch zur Erklärung und Verdeutschung der unserer Sprache aufgedrungenen fremden Ausdrücke' (1372 Bilddateien, Speichervolumen 10,2 GB) und Bilddigitalisaten des Jean-Paul-Nachlasses (8648 Bilddateien im Tif-Format, Speichervolumen 681 GB).

Als Testkorpus für den Datentyp ‚Wörterbuch', Modell für stark strukturierte Textdaten, wurden die Textdaten der Wörterbucher von Joachim Heinrich Campe aufbereitet. Form und Nutzung müssen noch an die weiterentwickelten Werkzeuge des TextGridLab und an das TextGridRep angepasst werden. Außerdem ist noch über Lizensierungsmodell und Nutzungsrechte zu entscheiden.

M 4.2: Prototypischer Durchlauf der Corpora durch die verschiedenen Tools mit Evaluation (Empfehlungen für weitere Anpassungen/Entwicklungen) Mathe + Geisteswissenschaften

Report 4.1 beschreibt detailliert den Digitalisierungsworkflow von der Bilddigitalisierung über die verschiedenen Stufen der Texterfassung und Textbearbeitung bis hin zur Textkodierung valider TEI-Dateien und demonstriert das Resultat prototypisch an Beispielen aus dem Buchstabenbereich A.

R4.1: Workflow

„Zur Digitalisierung von Primärquellen für die TextGrid-Umgebung: Modellfall Campe-Wörterbuch"[26]

f) AP5: Semantic Web und TextGrid = Semantic TextGrid

Dieses Arbeitspaket war teilweise durchaus als Erkundungsmodul konzipiert, in dessen Rahmen semantic web-Technologien und vorhandene hochstrukturierte Datenbestände (insb. fein ausgezeichnete Wörterbücher) unter dem Leitaspekt einer Metalemmaliste des Deutschen zusammengeführt werden sollten. Sie sollte als Basis-Ontologie für die Erschließung geisteswissenschaftlicher Primärquellen dienen.

Das Trierer Wörterbuchnetz wurde als Service in das TextGridLab integriert und kann einerseits als zentrales Erschließungs- und Annotierungstool für Primärquellen eingesetzt werden, andererseits selbst zum Gegenstand weiterer semantischer Arbeit und Erschließung werden (s. z.B. Wörterbuch-Linkeditor M5.1 und 5.2).

[26] http://www.textgrid.de/fileadmin/TextGrid/reports/TextGrid_Report_4_1.pdf

Schon früh zeichneten sich darüber hinaus zwei Entwicklungen ab, denen gegen Mitte des Projekts analog zu AP2/3 mit einer leichten Anpassung der Arbeitspläne Rechnung getragen wurde.

Erstens stellte sich heraus, dass die geplante Integration von GermaNet aus lizenztechnischen Gründen nicht möglich sein würde. Zweitens wurden von verschiedenen Communities Wünsche hinsichtlich der Metalemmaliste an TextGrid herangetragen, die ein hohes Maß an händischen Arbeiten erforderten und die Kapazitäten von AP5 bei weitem überschritten.

Deshalb wurde Mitte 2007 ein flankierender Antrag zum Projekt „Wechselwirkungen zwischen linguistischen und bioinformatischen Verfahren, Methoden und Algorithmen: Modellierung und Abbildung von Varianz in Sprache und Genomen“ beim BMBF eingereicht, das im Rahmen des Förderschwerpunktes "Wechselwirkungen zwischen Natur- und Geisteswissenschaften" Anfang 2008 bewilligt wurde.

Von dieser Ergänzung bzw. Kooperation profitieren beide Projekte maßgeblich: einerseits konnten in AP5 die grundlegenden Konzepte, Strukturen und Datenmodelle entwickelt und getestet werden, die zur Planung und Durchführung des neuen Projekts notwendig sind, andererseits steuert „Wechselwirkungen“ die notwendigen Ressourcen zur Erzeugung philologisch gesicherter Ergebnisse und Entwicklung völlig neuer Ansätze bei, die dann wiederum in TextGrid zurückfließen werden.

Vor diesem Hintergrund waren in AP5 einige Arbeitsschritte anzupassen; so wurden allgemein aufwendige Einzelaspekte, die mit einem hohen manuellen Aufwand verbunden gewesen wären – etwa die Lemmatisierung von Belegstellen in ausgewählten Wörterbüchern – durch die Entwicklung generischer Methoden und Tools ersetzt. Diese zielen, unter Verwendung von semantic web-Technologien, auf eine aktive Einbindung der wissenschaftlichen Community in die Entwicklung einer reich annotierten Metalemmaliste bzw. deren Ausbau zu einer Basis-Ontologie.

Milestones und Deliverables

R5.1 Ontologiereport[27]

R5.1 untersucht zusammen mit R1.6 die Möglichkeiten zur Integration von Ontologien und Wortnetzen in TextGrid, wobei R.1.6 stärker auf technisch-infrastrukturelle Aspekte abzielt, R5.1 auf inhaltlich-lexikalische. Im Zentrum stand die Frage, wie vorhandene Ressourcen zur Analyse, Annotation und Recherche innerhalb von TextGrid eingesetzt werden können. Als Alternative zum lizenzrechtlich nicht verwendbaren GermaNet wurde frei zugängliche Wortnetze/Thesauri wie z.B. OpenThesaurus vorgeschlagen.

[27] Vgl. http://www.textgrid.de/fileadmin/TextGrid/reports/TextGrid_Report_5_1.pdf

M5.1 u. 5.2 (Linkeditor Wörterbuch bzw. Linkeditor Wörterbuch-Primärtext)

Auch diese Milestones waren vom Ausfall von GermaNet betroffen. Es erwies sich auch vor diesem Hintergrund als sinnvoll, sie gemeinsam neu zu strukturieren und zusammenzufassen. Der Linkeditor besteht aus zwei Komponenten: Die Grundlage bildet ein performantes Tool zur Wörterbuchrecherche, das auf einem Webservice basiert und analog zum AP2-Recherchetool gezielte Suchanfragen an das Datenbanksystem des Wörterbuchnetzes ermöglicht. Diese Komponente kann separat benutzt werden und stand schon in den ersten Demo-Versionen des TextGridLabs zur Verfügung.

Die zweite Komponente ermöglicht die Verknüpfung und semantische Annotation von Artikelstichwörtern; hier wird eine Grundlage zur interaktiven Erstellung eines Wortnetzes gelegt, das nicht auf den nhd. Standard beschränkt ist. Die Programmierung dieser Komponente wurde aufgrund von Abhängigkeiten zu anderen Tools (Authentifizierung) zunächst zurückgestellt, steht aber kurz vor ihrer Fertigstellung und soll noch in die offizielle Betaversion mit eingehen. Darüber hinaus wird dieser Ansatz im Rahmen des „Wechselwirkungs“-Projekts weiter auszubauen sein, wobei u.a. auch bestehende lexikalische Ressourcen wie z.B. OpenThesaurus mit eingebunden werden sollen. Die Wörterbuch lässt sich einerseits über eine Suchmaske, andererseits aber auch über einen Mausklick auf ein gegebenes Wort im Texteditor des Labs starten, womit eine schnelle und komfortable Möglichkeit zum Verknüpfen von Wörterbucheinträgen und Primärtexten zur Verfügung steht.

M5.3 u. R5.2 (Metalemmaliste bzw. Report über Verfahren zur Erstellung)

Wie oben dargestellt, wurde im Rahmen des BMBF-Programms "Wechselwirkungen" ein eigener Antrag zur Erstellung einer philologisch geprüften, umfassenden Metalemmaliste. gestellt, dessen Ergebnisse in TextGrid zurückfließen werden. Die umfangreichen Tests und Vorbereitungen wurden im Report 5.2 zusammengefasst.

M5.4 (Lemmatisierung von Belegstellen im Deutschen Wörterbuch der Brüder Grimm und im Mittelhochdeutschen Wörterbuchverbund)

Hier wurde, wie oben bereits dargelegt, ein generisches Tool zum interaktiven Lemmatisieren programmiert. Wortformen können aus dem TextGridLab heraus angewählt und mit einem datenbankgestützten, dynamisch erweiterbaren Lexikon verglichen werden. Eine Anreicherung dieses Lexikons mit Lemmatisierungsdaten der Arbeitsstelle des Mittelhochdeutschen Wörterbuchs der Akademie der Wissenschaften und der Literatur Mainz ist geplant.

M5.5 (Umkehrlexikographie)

Auch in diesem Punkt konnte wichtige Erfahrungen im Hinblick auf das Projekt Metalemmaliste gewonnen werden. Anhand eines mhd. Wörterbuchs, des BMZ[28] , wurde ein prototypisches Umkehrlexikon erstellt, bei dem Stichwortansatz und Definitionsteil umgekehrt wurden. Die Ergebnisse bestätigten die Einschätzung, dass ein sinnvoll einsetzbares Umkehrlexikon ein hohes Maß an Handarbeit voraussetzt, etwa um direkte Übersetzungen von umschreibenden Erläuterungen abzugrenzen. Solch ein philologisch geprüftes Umkehrlexikon liegt mit dem "Neuhochdeutschen Index zum mhd. Wortschatz"[29] jedoch bereits vor. Deshalb wurde auf der Datengrundlage des BMZ die Integration von Umkehrwörterbüchern in die oben skizzierte Wörterbuchsuche vorbereitet. Sobald die lizenzrechtlichen Bedingungen geklärt sind, soll der gesamte „Neuhochdeutsche Index“ im Rahmen des Tools angeboten werden, was die Recherchemöglichkeiten entscheidend erweitern wird. Inzwischen sind die Daten der Druckfassung von 1990 für die weitere Arbeit an dem Projekt auf dem Opus-Server der Universitätsbibliothek Würzburg öffentlich zur Verfügung gestellt worden.[30]

M5.6 (Metadata Application Profiles)

Da TextGrid eine Plattform für Daten unterschiedlichster Provenienz bieten soll, war der Umgang mit verschiedenen Metadatensets von Anfang an eine zentrale Anforderung. TextGrid verwendet hier ein eigenes Kernmetadatenset, das vergleichsweise wenig Elemente enthält, bei deren Auswahl aber auch auf die grundsätzliche Austauschbarkeit mit bestehenden kanonischen Sets geachtet wurde. Die genauen Relationen, die die Grundlage für ein automatisches Mapping darstellen, sind in einer separaten Dokumentation festgehalten.

g) AP6: Projektmanagement und Öffentlichkeitsarbeit

Dieses Arbeitspaket bildete zwei Schwerpunkte: Projektmanagement und Öffentlichkeitsarbeit. Ziel und Aufgabe des Projektmanagements war, sicherzustellen, dass die wesentlichen Ziele des Gesamtvorhabens erreicht werden. Ziel und Aufgabe der Öffentlichkeitsarbeit war, das Projekt und dessen Arbeitsergebnisse auf nationaler und internationaler Ebene bekannt zu machen und für einen wechselseitigen Austausch mit der Community zu sorgen.

Diese Ziele wurden erreicht.

Projektmanagement

[28] Mittelhochdeutsches Wörterbuch. Mit Benutzung des Nachlasses von Georg Friedrich Benecke ausgearbeitet von Wilhelm Müller und Friedrich Zarncke. Nachdruck der Ausgabe Leipzig 1854-1866 mit einem Vorwort und einem zusammengefaßten Quellenverzeichnis von Eberhard Nellmann sowie einem alphabetischen Index von Erwin Koller, Werner Wegstein und Norbert Richard Wolf. 4 Bde. u. Indexbd. Stuttgart: S. Hirzel 1990.

[29] Wegstein Werner, Wolf Norbert R., Koller Erwin: Neuhochdeutscher Index zum Mittelhochdeutschen Wortschatz. . Stuttgart 1990.

[30] URN: urn:nbn:de:bvb:20-opus-35530; URL: http://www.opus-bayern.de/uni-wuerzburg/volltexte/2009/3553/

Das Management des Projektes umfasste die administrativen und organisatorischen Aufgaben, die bei der Durchführung des Projektes anfielen. Dazu zählen bspw. die finanzielle Verwaltung der zentral eingesetzten Mittel, wie die Reisemittel aus der Aufstockung im zweiten Halbjahr 2008, sowie die Mittel zur Durchführung von Veranstaltungen und Konferenzen.

Die Sicherstellung der termingerechten Abgabe und Einhaltung von Meilensteinen und Ergebnisberichten lagen ebenfalls im Verantwortungsbereich des Arbeitspaketes. Dazu gehörte auch die Moderierung der Modifikation von Arbeitsplänen, wie z.B. in den Arbeitspaketen 2 und 3.

Öffentlichkeitsarbeit

Im Rahmen des Projekts wurden intensive PR-Maßnahmen für das Projekt mit nationaler und internationaler Wirkung in Zusammenarbeit mit dem Konsortium durchgeführt. Dazu zählte neben der Entwicklung einer Corporate Identity in Anlehnung an D-Grid (Logo, Website, Flyer) die Propagierung der Projektergebnisse auf ausgewählten Kongressen und Tagungen, die regelmäßige Veröffentlichung von Publikationen (siehe dazu Kapitel 4), die Bereitstellung einer regelmäßig aktualisierten, zweisprachigen Projekthomepage, die periodische Veröffentlichung eines zweisprachigen Newsletters sowie die Durchführung verschiedener Konferenzen.

Milestones und Deliverables

M6.1 Freischaltung der Projekthomepage

Die Projekthomepage konnte wie geplant veröffentlicht werden und stellt neben der Schnittstelle zur Öffentlichkeit als externe Kommunikationsplattform auch die Arbeitsplattform für die interne Kommunikation dar. Hier wurden alle Informationen rund um das Projekt zusammengeführt.

Im extern zugänglichen Bereich: Neuigkeiten, Informationen zu Projektplanung, -durchführung und -zielen, Informationen über Partner, Veranstaltungen, Veröffentlichungen, Berichte.

Im internen Bereich alle projektbezogenen Dokumente, Arbeitsverläufe in den einzelnen Arbeitspaketen und Arbeitsgruppen, Planung von Konsortialtreffen und Telefonkonferenzen, Informationen zum Fachbeirat.

R6.1 Konzept für die Öffentlichkeitsarbeit

Der Report spiegelt die Planungen, Aktivitäten und Ziele im Bereich der Öffentlichkeitsarbeit wider.

M6.2 Newsletter Version 1.0 (danach alle 3 Monate)

An dem Plan, einen zweisprachigen Newsletter in regelmäßigen Abständen zu publizieren, wurde nicht festgehalten. Anlass für die Publikation sollte vielmehr ein Hauptthema sein, das die Newsletterabonnenten interessiert und animiert, auf der TextGrid-Homepage zu verwei-

len. Daher wurden insgesamt nicht 6, sondern nur 5 der geplanten Newsletter publiziert bzw. der sechste Newsletter wird zusammen mit dem ersten Newsletter des ab Juni laufenden TextGrid-Projekts veröffentlicht.

Die Zugriffe auf die TextGrid-Homepage erhöhten sich nach jeder Newsletter-Publikation für einige Tage signifikant.

Veranstaltungen

TextGrid-Konferenzen

21.-22.1.2009 Göttingen	TextGrid Summit
16.1.2009 Trier	TextGrid-Praxis für Editoren: Praktische Erprobung des TextGridLab
13.9.2007 Göttingen	II. Göttinger Grid Seminar
23.11.2006 Göttingen	e-Science und Grid-Aktivitäten in Göttingen

Konferenzen mit TextGrid Beteiligung

22.1.2009 Göttingen	e-Humanities-Abschluss-Workshop
13.-14.10.2008 Trier	<philtag n="7">
26.-29.2.2008 Cairns (AUS)	IEEE DEST 2008
12.-13.10.2007 Würzburg	<philtag n="6"> TEI-Workshop
10.-12-9.2007	1. D-Grid All Hands Meeting
21.-23.2.2007 Cairns (AUS)	IEEE DEST 2007

11.-13.4.2007 Tübingen	GLDV-FRÜHJAHRSTAGUNG 2007
6.-7.10.2006 Würzburg	<philtag n="5">-Tagung 2006

TextGrid-Präsentationen auf Veranstaltungen

15.-16.5.2009 Mannheim	D-SPIN Sprachressourcen-Gipfel
7.-8.5.2009 Siegen	Jahrestagung des Instituts für Medienforschung Siegen: "Enhancing Humanities. Potentials of media and ICT in the Humanities."
16.-18.4.2009 Providence, Rhode Island (USA)	Bamboo-Workshop 4
10.-13.3.2009 Pisa (I)	Interedition Bootcamp 2009-03
10.-12.3.2009 Mannheim	45. Jahrestagung des Instituts für Deutsche Sprache: "Sprache intermedial: Stimme und Schrift, Bild und Ton"
2.-5.2.2009 Lyon (F)	iRODS workshop @ CC-IN2P3
3.-5.2.2009 Bielefeld	9th International Bielefeld Konferenz: Upgrading the eLibrary
12.-14.1.2009 Tucson, Arizona (USA)	Bamboo-Workshop 3
7.-12.12.2008 Indianapolis, Indiana (USA)	4th IEEE International Conference on e-Science
6.-8.11.2008 London (UK)	2008 TEI Members Meeting London

24.-26.11.2008 Edinburgh (UK)	eSI Workshop: Building Communities in the Digital Arts and Humanities
3.10.2008 Melbourne, (AUS)	eResearch Australasia: Workshop 9: e-Research in the Arts, Humanities and Cultural Heritage
22.-26.9.2008 Berlin	International Conference on Dublin Core and Metadata Applications
18.-21.9.2008 Heidelberg	ITUG Jahrestagung 2008
12.-13.9.2008 Ispra (I)	2008 Seventh International Workshop on Finite State Methods and Natural Language Processing
8.-11.9.2008 Edinburgh (UK)	UK e-Science AHM2008
6.-8.10.2008 Mannheim	Workshop Redaktionssysteme
25.-29.6.2008 Oulu (FIN)	Digital Humanities 2008
2.-6.6.2008 Barcelona (E)	23rd Open Grid Forum
7.-11.4.2008 Taipei (RC)	International Symposium on Grid Computing (ISGC) 2008
26.-28.3.2008 Perth (AUS)	19th Australian Software Engineering Conference
20.3.2008 Curtin University (AUS)	DEBII Centre Public Seminar
11.-13.3.2008 Mannheim	44. Jahrestagung des Instituts für Deutsche Sprache (IDS)
21.-22.2.2008 München	Mediävistische Editionen im digitalen Zeitalter

13.-16.2.2008 Berlin	12. internationale Tagung der Arbeitsgemeinschaft für germanistische Edition
30.-31.1.2008 London (UK)	Workshop: Epistemic Networks and Grid + Web 2.0 for Arts and Humanities
6.-8.12.2007 Paderborn	Edirom Musikedition 2007: "Digitale Edition zwischen Experiment und Standardisierung"
11.-13.12.2007 Washington DC (USA)	3rd Digital Curation Conference
31.10.-03.11.2007 Maryland (USA)	TEI@20: 20 Years of Supporting the Digital Humanities - 20th Anniversary Text Encoding Initiative Consortium Members' Meeting
13.-16.9.2007 Zürich (CH)	ITUG-Jahrestagung
7.-11.5.2007 Manchester (UK)	The 20th Open Grid Forum - OGF20
2.-4.5.2007 Baden-Baden	German e-Science-Konferenz 2007
4.-6.12.2006 Amsterdam (NL)	2nd IEEE International Conference on e-Science and Grid Computing
17.-22.9.2006 Alicante (E)	ECDL Workshop: DLSci 2006
11.-15.9.2006 Karlsruhe	GridKa Schule 2006
6.-8.9.2006 Münster	Inetbib-Tagung
21.-22.7.2006 München	Vernetzungsstrukturen-Workshop
5.-8.7.2006	Digital Humanities 2006

27.-30.6.2006 Dresden	ISC2006, 21st International Supercomputer Conference
26.-27.5.2006 Berlin	Berliner Editionstagung 2006
6.4.2006	e-Science und vernetztes Wissensmanagement, Kick Off
13.-16.2.2006 Athen (GR)	16th Global Grid Forum - GGF16
20.-22.1.2006	Digitale Philologie - Probleme und Perspektiven
7.10.2004	Thementag: e-Science, Wissensmanagement im virtuellen Labor? - Managementlösungen gesucht

An den thematisch für das Projekt relevanten D-Grid-Veranstaltungen war TextGrid aktiv beteiligt.

2. des voraussichtlichen Nutzens

In den drei Jahren seit Projektbeginn wurde im Rahmen von TextGrid nicht nur der Grundstein für eine solide und modulare Grid-basierte Infrastruktur für die Geisteswissenschaften gelegt, sondern das Projekt entwickelte sich aufgrund seines fachwissenschaftlichen Fokus zu einem der Kristallisationspunkte für die e-Humanities in Deutschland (s.u.). Darüber hinaus sorgte das Projekt auch international für Aufmerksamkeit – sowohl fachwissenschaftlich (z.B. TEI, Digital Humanities) als auch technisch (z.B. IEEE, OGF).

a) Wissenschaftlicher Nutzen

Wenn Textwissenschaftler das Beziehungsgeflecht zwischen Sprache und Diskurs oder die komplexen Prozesse des Entstehens literarischer Texte untersuchen, arbeiten sie oft noch isoliert oder innerhalb in sich geschlossener Projekte. Aktuelle Forschungsaktivitäten im Bereich der Textwissenschaften lassen häufig die Verknüpfung mit existierenden Textkorpora, Wörterbüchern, Lexika oder Sekundärliteratur und die Anbindung an bereits verfügbare Textwerkzeuge vermissen, obwohl dies einen beträchtlichen Mehrwert darstellt und eine Vielzahl an Möglichkeiten der weiteren Datenverarbeitung schafft. TextGrid stellt mit seiner Architektur und seinen Werkzeugen eine Forschungsinfrastruktur zur Verfügung, die eine solche Integration ermöglicht und deshalb die Arbeitsweise von Geisteswissenschaftlern nachhaltig verändern kann. Die Notwendigkeit, hierbei auf Grid-Technologien zu setzen, liegt auf der Hand: Durch die Digitalisierungsinitiativen der letzten Jahre sind erhebliche Datenmengen entstanden, die inzwischen einige hundert Terabyte umfassen – Grids können mit solchen großen, ständig wachsenden Datenvolumen umgehen (Ausfallsicherheit, hohe Verfügbarkeit, schneller Zugriff). Die Grid- und WebService-basierte TextGrid-Infrastruktur erlaubt es, Wissenschaftler miteinander zu vernetzen, die derzeit in räumlicher Distanz an vergleichbaren Projekten arbeiten, sowie vorhandene Tools global verfügbar zu machen, die bislang lediglich lokal benutzbar sind. Auf diese Weise wird eine eHumanities-Plattform errichtet, auf der Experten verschiedener Fachgebiete eine virtuelle Arbeitsgemeinschaft bilden können. Während der Projektphase geschah dies mit einem Schwerpunkt auf der germanistischen Editionsphilologie – und ansatzweise für die Linguistik. So ermöglicht und unterstützt TextGrid derzeit die gemeinschaftliche philologische Bearbeitung, Analyse, Annotation, Edition und Publikation von Textdaten.

Dass TextGrid hiermit den Nerv der Zeit getroffen hat, zeigt die Resonanz, die bislang alleine die Präsentation des Projekts auf zahlreichen Veranstaltungen im In- und Ausland hervorgerufen hat, sowie eine Reihe von Kooperationsanfragen, bspw. seitens der Staats- und Universitätsbibliothek Bremen oder der Mainzer Akademie der Wissenschaften.

Für die Nachhaltigkeit des Erreichten ist es erforderlich, TextGrid auf eine möglichst breite fachliche Basis zu stellen. Nur eine große und aktive Community kann eine nachhaltige Weiterentwicklung von TextGrid gewährleisten. Um dieses Ziel zu erreichen, wird eine Doppelstrategie verfolgt: Zum einen müssen die im Kontext der ersten Projektphase entwickelten Prototypen zur Produktionsreife gebracht werden, denn nur robuste und leistungsfähige Soft-

ware wird die Anwender langfristig zufriedenstellen und so an TextGrid binden. Das umfasst die Implementierung neuer Features, vor allem aber die Fehlerbeseitigung und Einarbeitung der Rückmeldungen aus allgemeinen Nutzertests und spezifischen Projektkooperationen. Außerdem werden neue Werkzeuge entwickelt, die auf die Bedürfnisse weiterer geisteswissenschaftlicher Communities ausgerichtet sind. Diese Werkzeuge sollen entweder durch die Weiterentwicklung bereits bestehender Komponenten oder die Integration unabhängig von TextGrid entstandener Module einen ähnlichen Entwicklungsstand wie die bereits in TextGrid vorhandenen Module erreichen. Neben Ausbau und Pflege der in der ersten Projektphase vorrangig im Schwerpunkt der Editionsphilologie entwickelten Werkzeuge sollen nun weitere geisteswissenschaftliche Communities eingebunden werden. Im Rahmen des TextGrid-Projekts, Phase II, das für den Zeitraum vom 01.06.2009 – 31.05.2012 bewilligt wurde, sind die Musikwissenschaften, die Kunstgeschichte, die Klassische Philologie und die Linguistik als Projektpartner involviert.

b) Technischer Nutzen

Um die TextGrid-Architektur nachhaltig nutzbar zu machen und für ihre stetige Weiterentwicklung zu sorgen – idealerweise im Rahmen eines Open Source Projekts – bedarf es weiterer Schritte (eine ausführliche Darstellung des technischen Nutzens wurde im Kapitel „eingehende Darstellung des erzielten Ergebnisses“ (=> TextGrid-Architektur) beschrieben.). Zunächst gilt es, wie oben erwähnt, eine möglichst breite Nutzer-Community zu gewinnen. Hierfür muss die Software in den Produktionsbetrieb überführt und Feedback aus den neuen und alten Communities eingearbeitet werden. Soweit es die fachspezifischen Komponenten angeht, wird dies vom oben erwähnten TextGrid Verstetigungsantrag abgedeckt. Bezüglich der Kern-Infrastruktur, insbesondere des TextGridRep, wird dies im Rahmen des WissGrid-Projekts (ebenfalls D-Grid III) erfolgen. Hierbei liegt ein besonderer Schwerpunkt auf dem Aufbau eines Grid-basierten Langzeit-Repositorys für Forschungsdaten. Auch die mittelfristige (10 Jahre) Archivierung von Forschungsdaten stellt mit den Richtlinien der DFG zur Sicherung guter wissenschaftlicher Praxis eine wichtige Anforderung dar, der TextGrid entsprechen muss, wenn es von DFG-geförderten Projekten genutzt werden soll.

c) Wirtschaftlicher Nutzen

Neben den kontinuierlichen Aktivitäten von TextGrid zum Aufbau einer breiten Nutzerbasis und den oben beschriebenen Maßnahmen zur wissenschaftlichen und technischen Nachhaltigkeit von TextGrid, wird die Verwertung von TextGrid vor allem im Rahmen der in D-Grid geführten Diskussion zu Nachhaltigkeit und Geschäftsmodellen weiterentwickelt. Für eine ausführliche Darstellung sei auf den TextGrid Verstetigungsantrag (Call D-Grid III, Kapitel 4) hingewiesen. Zu den diesbezüglichen Planungen siehe dort (Kapitel 4, „Konzept eines Geschäftsmodells und Verwertung") und unten, III.3 zur Fortschreibung des Verwertungsplanes. Darüber hinaus wird das Thema im Rahmen des WissGrid-Projekts als eigenes Arbeitspaket "Geschäftsmodelle für wissenschaftsnahe Community-Grids" verfolgt.

3. des während der Durchführung des Vorhabens dem ZE bekannt gewordenen Fortschritts auf dem Gebiet des Vorhabens bei anderen Stellen

Während der Projektlaufzeit hat sich die internationale Forschung grundlegend neu ausgerichtet und der Aufbau von "eHumanities"-Infrastruktur, wie TextGrid es auch in seinem Antrag formuliert hat, ist zu einem gemeinsamen Anliegen geworden. Das Konzept und der technologische Vorsprung von TextGrid sind international anerkannt und TextGrid wird bei verwandten Aktivitäten in aller Welt konsultiert.

Obwohl die "Digital Humanities" - die Nutzung von Computertechnologien in den Geisteswissenschaften - schon seit mehreren Jahren etabliert werden, ist die kollaborative Entwicklung von gemeinsamen Werkzeugen und Arbeitsumgebungen ein relativ neues Phänomen. "eHumanities" bezeichnet Aufbau und Nutzung der für kollaborative Arbeitsumgebungen nötigen Mittel (gemeinsame Technologien, Konzepte und Organisation). So geläufig der Begriff "eHumanities" inzwischen klingen mag, er kam erst im Laufe der TextGrid-Projektlaufzeit auf und wurde durch TextGrid konzeptuell und technologisch entscheidend mitgeprägt und schließlich von diversen Initiativen international aufgegriffen.

Exkurs: eHumanities - nationale und internationale Entwicklungen

Forschung und Lehre in den Humanities machen zunehmend Gebrauch von digitalen Werkzeugen. Die technologische Unterstützung wird kontinuierlich ausgebaut und steht gerade heute vor einem weiteren großen Entwicklungsschritt. Ein enormes Potenzial liegt – gerade für die Geisteswissenschaften – in der Anwendung moderner Internet-Technologien, dem Einsatz z.B. von „Social Software“ und der konsequenten Vernetzung von Informationen und Werkzeugen. Umfangreicher Zugriff auf wissenschaftliche Daten, generische Werkzeuge zur Analyse und zur Unterstützung der Arbeitsprozesse, verbesserte Zusammenarbeit zwischen Forschergruppen und über Disziplinen hinweg – diese Stichwörter zählen zu den Triebfedern auf dem Weg in die vernetzten „enhanced“ Humanities, die „eHumanities“.

Der Schlüssel zum „e“ in den eHumanities liegt in der breiten technologischen Unterstützung der geisteswissenschaftlichen Forschung. Im Zeitalter von Web 2.0, Semantic Web, Grid-Technologien und vielem mehr gilt es, Forschern und Wissenschaftlern in den geisteswissenschaftlichen Disziplinen einen vernetzten, kollaborativ nutzbaren Virtuellen Forschungsrahmen anzubieten, der – eingebettet in eine standardisierte technologische Infrastruktur – dem Wissenschaftler an jedem Ort Zugang zu Forschungsdaten, Diensten und Publikationen bietet. Forschungsdaten und Publikationen können in dieser Umgebung aus verschiedenen geisteswissenschaftlichen Blickwinkeln betrachtet werden und führen zu individuellen und unterschiedlichen Ansätzen sowie Ergebnissen. Erst dieser gemeinsame Zugriff auf forschungsrelevante Information gepaart mit geeigneten (fach-)spezifischen Diensten erlaubt die Generierung völlig neuer Forschungsfragestellungen und führt somit zu einer Veränderung der gesamten Forschungslandschaft. Beispielhaft sei hier nur die empirische Textforschung (Text Mining) genannt.

Die Aktivitäten im Bereich der Geisteswissenschaften erstrecken sich über verschiedene Gebiete: Forschung, Datenpflege, Dienste-Standardisierung, Schulung, Unterstützung, Standards, Richtlinien, Verknüpfung von Forschungsinfrastrukturen, Langzeitarchivierung, Ressourcen-Allokation und Erhalt der Infrastruktur.

In diversen Disziplinen in Deutschland konnten z.B. spezialisierte Archive aufgebaut werden, die einzelne der oben beschriebenen Charakteristika bereits umgesetzt haben. Projekte wie das Heinrich-Heine-Portal oder das Portal 'Monumenta Germaniae Historica' zielen darauf ab, die Werke aus einem Spezialbereich (nämlich die Werke von Heinrich Heine bzw. Quellen zur mittelalterlichen Geschichte) frei über das Internet verfügbar vorzuhalten.

'Collate' war ein europäisches Projekt zum Aufbau eines "Collaboratories", eines Online-Portals zur Zusammenarbeit im Bereich von Filmstudien. Nach Auslaufen der Projektförderung ist der Status dieser Initiative aus fünf europäischen Ländern zurzeit ungewiss. Projekte wie 'Hyper Nietzsche' bzw. 'discovery – Philosophy in the digital Era' sind hier ebenfalls als richtungsweisende Initiativen zu nennen. Auch die Arbeiten des Sonderforschungsbereichs 'Linguistische Datenstrukturen' in Tübingen und noch einige andere sind in diesem Kontext relevant, um nur einige wenige zu nennen.

Im Rahmen der BMBF-Förderausschreibung „Wechselwirkungen zwischen Natur- und Geisteswissenschaften“ wurden im Jahr 2008 eine Reihe von Projekten gefördert, die optimal in ein deutsches eHumanities-Netzwerk eingebunden werden können:

Fächerschwerpunkt Archäologie

- 3D-Sutren – Interaktive Analysewerkzeuge für einen Web-Atlas gescannter Sutratexte in China

Fächerschwerpunkt Sprachwissenschaft

- Extraktion von strukturiertem Wissen aus Antiken Quellen für die Altertumswissenschaften (eAQUA)
- Wechselwirkungen zwischen linguistischen und bioinformatischen Verfahren, Methoden und Algorithmen: Modellierung und Abbildung von Varianz in Sprachen und Genomen

Die Altertumswissenschaften sind an textbasierten Werkzeugen und Diensten besonders interessiert, da auf diese Weise Informationen und Daten zu Grabungsobjekten mit entsprechender Literatur und Annotationen angereichert werden können.

Am Forschungszentrum „Deutscher Sprachatlas“ in Marburg liefen zwei Vorhaben, die in diesem Kontext ebenfalls erwähnt werden sollen:

- „Digitaler Wenker-Atlas“ (DiWA): Im diesem Projekt, das Bestandteil des Programms „Retrospektive Digitalisierung von Bibliotheksbeständen“ war, wurden großformatige Sprachkarten im Internet (ca. 2 GB/Karte) publiziert. Inzwischen ist DiWA zu einem Informationssystem gewachsen, das neben Bildern auch Ton und zukünftig auch Text verfügbar halten wird.

- „regionalsprache.de“: Das Projekt läuft seit 2008, ist auf 19 Jahre angelegt und hat ein Volumen von 14 Mio. Euro. Ziel des Vorhabens ist die Verfügbarmachung aller sprachraumbezogener Großprojekte (Sprachatlanten) der Bundesrepublik an einem Ort sowie die Verfügbarmachung unveröffentlichter Wissenschaftsmaterialien (Ton, Text, Bild/Karte). Ferner wird erstmals eine Erhebung der modernen deutschen Regionalsprachen unternommen. Die Daten werden ausschließlich im Internet zugänglich sein.

Diese Projekte sind forschungs- und nutzerorientiert, aber eben auch inhaltlich spezialisiert und oftmals technologisch und/oder zeitlich begrenzt. Damit solche spezialisierten Projekte nachhaltig betrieben und die bereits beschriebenen Vorteile der Vernetzung realisiert werden können, ist eine Einbindung in eine eHumanities-Infrastruktur erforderlich.

Europaweit ist die Entstehung nationaler Kompetenzzentren zu beobachten, deren Aufgabe der nachhaltige Aufbau einer Forschungsinfrastruktur ist. Besonders hervorzuheben sind hier das AHDS in Großbritannien und DANS in den Niederlanden, die auf nationaler Ebene sehr erfolgreich arbeiten und international anerkannt sind.

Neue Initiativen zum Aufbau von Infrastruktur für die Geisteswissenschaften entstehen an einzelnen Institutionen, national und international sowie gezielt für einzelne geisteswissenschaftliche Disziplinen bzw. disziplinübergreifend. Dabei lässt sich beobachten, dass disziplin-orientierte Initiativen oft besser den Nutzer integrieren, also "mit" dem Wissenschaftler gemeinsam arbeiten, und nicht "für" den Wissenschaftler eine neuartige Arbeitsumgebung definieren.

TextGrid befindet sich in gewisser Weise zwischen diesen Dimensionen. Als nationales Projekt gestartet, hat es auch international enge Kontakte geknüpft und versucht, internationale Entwicklungen mitzugestalten. Das Projekt ist primär spezifisch auf die Textwissenschaften ausgerichtet und wissenschaftliche Nutzer sind direkte TextGrid-Partner. Dennoch versucht TextGrid jetzt und in Zukunft, auch andere Disziplinen anzusprechen und mit einzubeziehen.

a) national

In einigen Ländern haben sich aus der Initiative von Wissenschaftlern bereits nationale Zentren für die Geisteswissenschaften gebildet. Oftmals geschah dies aus der Notwendigkeit eines Datenarchivdienstes, der schließlich gewachsen ist und nun auch weiterführende Dienste anbietet. Das Arts and Humanities Data Service (AHDS)[31] in Großbritannien war der erste diesbezügliche Dienst, und er wurde nach kurzzeitiger Ungewissheit nun zu einem Network of Centers in the Humanities[32] ausgebaut. Erst in jüngster Zeit entstehen auch in anderen Ländern vergleichbare Dienste; darunter DANS in den Niederlanden[33], das Digital Humanities Observatory in Irland (DHO)[34], der Australian National Data Service (ANDS)[35], der SND

[31] http://ahds.ac.uk/

[32] http://www.arts-humanities.net/noc

[33] http://www.dans.knaw.nl/

[34] http://www.dho.ie/

– Swedish National Data Service in Schweden[36] und der Schweizerischer Informations- und Datenarchivdienst für die Sozialwissenschaften – SIDOS in der Schweiz[37].

Auch in Deutschland gibt es erste Bestrebungen für eine nationale eHumanities-Strategie, die auch in wesentlichen Teilen durch TextGrid-Partner mitinitiiert wurde.[38] Begleitet werden diese Bestrebungen durch wissenschaftliche Vertreter aus den verschiedensten geisteswissenschaftlichen Disziplinen, Förderer (BMBF, DFG) und internationale Partner (siehe nachfolgend).

AHDS / CeRch

Das Arts and Humanities Data Service (AHDS) ist das erste bekannte nationale Archiv und Dienstleistungszentrum für die Geisteswissenschaften. Das AHDS wurde 1995 als nationales Datenzentrum gegründet, das Empfehlungen für die richtige Datenerstellung und deren langfristige Pflege gab, und auch selbst aktiv Daten sammelte und archivierte. Dabei konzentrierte sich das AHDS auf fünf Schwerpunkte in den Geisteswissenschaften (Archäologie, Geschichte, Literatur und Linguistik) und der Kunst (Visuelle Kunst, Darstellende Kunst), die auch jeweils durch einzelne Institutionen besonders gefördert wurden. Im Bereich der Literatur und Linguistik war dies das Oxford Text Archive (OTA), das 1976 von Lou Burnard gegründet wurde.

Nach einer Reorientierung der Forschungsförderung durch AHRC und JISC im Jahr 2007 werden die Arbeiten des AHDS in das neue Centre for e-Research (CeRch) am King's College London sowie ein nationales Netzwerk von geisteswissenschaftlichen Zentren[39] übergehen. Diese finanziell und strukturell solide Konstruktion ermöglicht es, näher beim eigentlichen Wissenschaftler zu sein, als das dem zentralisierten AHDS möglich war.

DANS

Der Data Archiving und Networked Service, kurz DANS[40], ist eine zum AHDS vergleichbare Institution der Royal Netherlands Academy of Arts and Sciences (KNAW) und der Netherlands Organisation for Scientific Research (NWO). Wissenschaftler aus Fachrichtungen von der Archäologie bis hin zur Wissenschaftsgeschichte können im Archiv EASY von DANS ihre Daten archivieren, sie mit Metadaten beschreiben und Lizenzrechte zur Zugänglichkeit und auch zur Wiederverwendung der Daten definieren.

Bamboo

[35] http://ands.org.au/

[36] http://www.snd.gu.se/

[37] http://www.sidos.ch/

[38] Vgl. Heike Neuroth, Andreas Aschenbrenner, Felix Lohmeier: e-Humanities - eine virtuelle Forschungsumgebung für die Geistes-, Kultur- und Sozialwissenschaften. In: Bibliothek. Forschung und Praxis, 3 (2007), S. 272-279. http://www.bibliothek-saur.de/2007_3/272-279.pdf

[39] http://www.arts-humanities.net/noc

[40] http://www.dans.knaw.nl/

Das Bamboo-Projekt[41] ist eine Vorbereitungsstudie der National Science Foundation zum Aufbau einer eHumanities-Infrastruktur, die durch ein nationales Strategiepapier der ACLS[42] und vielfältige institutionelle Aktivitäten inspiriert wurde. Obwohl ursprünglich von Technologen initiiert, versucht Bamboo dezidiert alle (Wissenschaftler, Techniker, Bibliothekare, Universitäten) in die Zieldefinition für die zukünftigen eHumanities zu integrieren. Durch die intensive Diskussion mit Nutzern verändert das Projekt sukzessive seine Ausrichtung: weg von technischen Services hin zu kollaborativen und sozialen Umgebungen. Obwohl Bamboo primär US-weit ausgerichtet ist, versucht das Projektteam von internationalen Erfahrungen zu lernen und auch mögliche zukünftige internationale Vernetzungen vorzubereiten.

b) international

Auch international haben sich während der TextGrid-Projektlaufzeit relevante Initiativen entwickelt, teils aufgrund vorhandener Förderung in Infrastruktur (z.B. EC e-Infrastructure[43], ESFRI[44]), teils aufgrund des dringenden Bedarfs in wissenschaftlichen Communities. Im Folgenden werden internationale Netzwerke beschrieben, die Infrastruktur für die geisteswissenschaftliche Forschung errichten. Obwohl in Bezug auf bestimmte Aspekte auch relevant, werden Infrastruktur für Publikationsserver wie DRIVER[45]; Langzeitarchivierungsprojekte wie DPE[46], PLANETS[47] oder CASPAR[48]; generell technologische Entwicklungsprojekte wie BRICKS[49], DILIGENT[50], D4Science[51] und andere verwandte Initiativen an dieser Stelle nicht beschrieben.

Interedition

Edition und Analyse von wissenschaftlichen Texten ist bisher hauptsächlich auf institutioneller Ebene durchgeführt worden. Kooperationen über Institutsgrenzen hinweg haben sich vornehmlich auf den Gedankenaustausch im Rahmen von Konferenzen und Publikationen beschränkt. Interedition ergreift darüber hinaus konkrete Maßnahmen zum Austausch von Werkzeugen zur Edition, Publikation und Analyse von wissenschaftlichen Texten. Dieser europäische Verbund lebt hauptsächlich durch den Beitrag der beteiligten Partner und durch

[41] http://projectbamboo.uchicago.edu/

[42] ACLS (American Council of Learned Societies): Our Digital Commonwealth. Cyberinfrastructure for the Humanities and Social Sciences. December 2006.
http://www.acls.org/programs/Default.aspx?id=644&linkidentifier=id&itemid=644

[43] http://cordis.europa.eu/fp7/ict/e-infrastructure/

[44] http://cordis.europa.eu/esfri/

[45] http://www.driver-repository.eu/

[46] http://www.digitalpreservationeurope.eu/

[47] http://www.planets-project.eu/

[48] http://www.casparpreserves.eu/

[49] http://www.brickscommunity.org/

[50] http://www.diligentproject.org/

[51] http://www.d4science.eu/

die Synergien zwischen den Partnern; Reisegelder werden durch das Förderungsschema COST[52] zur Verfügung gestellt.

CLARIN

Wie auch Interedition ist CLARIN ein Beispiel für ein aus einer spezifischen Disziplin initiiertes internationales Infrastrukturprojekt. CLARIN (Common Language Resources and Technology Infrastructure) vernetzt existierende linguistische Datenzentren. Das deutsche Teilprojekt von CLARIN heißt D-SPIN.

Als ESFRI-Projekt[53] ist CLARIN langfristig ausgerichtet: In einem zweijährigen Vorprojekt wird die zukünftige Infrastruktur konzeptioniert (preparation phase), anschließend in einem bis zu zehnjährigen Zeitraum aufgebaut (construction phase) und schließlich permanent gewartet (operational phase). Um den Übergang in permanente Strukturen sicherzustellen, ist das Vorprojekt zu einem guten Teil auf organisatorisch-politische Themen ausgerichtet: Organisationsstruktur, rechtliche Rahmenbedingungen, Finanzierung, Rahmenstrategie und Ausrichtung auf Zielgruppen. Die Finanzierung wird nach anfänglicher Förderung durch ESFRI schrittweise an die nationalen Ministerien der jeweiligen Partner übergeben.

DARIAH

Anders als die beiden bereits beschriebenen Netzwerke Interedition und CLARIN stellt DARIAH ein thematisch sehr breites Netzwerk aus europäischen eHumanities-Zentren und errichtet eine europäische Forschungsinfrastruktur für die Geisteswissenschaften. Die verteilte Infrastruktur wird für die geisteswissenschaftliche Forschung in ihren Kernpunkten die Langzeitarchivierung von Forschungsdaten ausbauen sowie ihren Austausch und gemeinsamen Nutzen fördern.

DARIAH wurde ursprünglich von den Kernpartnern AHDS, DANS, dem CNRS[54] in Frankreich und der Max Planck Gesellschaft in Deutschland initiiert. Bei der Antragseinreichung vereinte das Projekt auch Partner aus Irland, Dänemark, Griechenland, Zypern, Kroatien und Slowenien. Projektstart für das zweijährige Vorprojekt[55] war September 2008; die Kooperation mit CLARIN, Bamboo und einer Vielzahl anderer Forschungsverbünde wird aktiv gesucht.

c) technisch

Neben den aus den Geisteswissenschaften heraus gestarteten Initiativen gab es während der TextGrid-Projektlaufzeit auch einige relevante technologische Entwicklungen. Hervorzuheben ist hier vor allem, dass der Abstand zwischen Grid- und Repositorien-Technologien deutlich kleiner geworden ist, und damit auch ein Hauptaugenmerk von TextGrid deutlich gewonnen hat.

---

[52] http://www.cost.esf.org/

[53] European Strategy Forum on Research Infrastructures - ESFRI. http://cordis.europa.eu/esfri/

[54] Centre National de la Recherche Scientifique - CNRS. http://www.cnrs.fr/

[55] siehe die Erklärung von ESFRI im Absatz zu CLARIN

Da Grid-Technologien ursprünglich aus den Naturwissenschaften - speziell der theoretischen Physik - stammen, ist die Hauptanforderung an Grid-Technologien die Virtualisierung von Hardware (Storage, Compute). Geisteswissenschaftliche Anwendungsfälle arbeiten eine konzeptuelle Ebene darüber mit virtualisierten Informationen und Diensten. Repository-Technologien bieten Mechanismen zur Virtualisierung und Verknüpfung von Informationen und Diensten, Langzeitarchivierung, semantischen Beschreibung des Kontextes von Informationsobjekten und vielen mehr. Aus technologischer Sicht existiert TextGrid somit genau in dem Zwischenraum zwischen Grid- und Repositorientechnologien. Entwicklungen zueinander sowohl aus dem Bereich der Repositorien hin zu (Grid-)Infrastruktur als auch umgekehrt aus der Infrastruktur hin zu Repositorien-Technologien sind daher für TextGrid von größtem Interesse.

Aus der Infrastruktur ist die Neuentwicklung von **iRODS** zu nennen. Zu TextGrid-Projektbeginn gab es zwar bereits Gerüchte um eine mögliche Neuentwicklung des Storage Resource Brokers (SRB) durch das San Diego Supercomputing Center (SDSC), die erste Version der neuen Software wurde aber erst Anfang 2008 fertiggestellt - zu spät, um für TextGrid als Speicherinfrastruktur in Frage zu kommen und zum damaligen Zeitpunkt aufgrund der Lizenzfrage noch zu unsicher. Die Konzepte der "Rules" und "Microservices" sind aber jedenfalls viel versprechend für die Community, und iRODS wird auch im weiteren Bestehen von TextGrid beobachtet.

Das auf die "europäische" Grid-Software gLite aufbauende **gCube**[56], das die technologische Basis für die Europäischen Projekte DILIGENT und D4Science darstellt, könnte wie iRODS in Zukunft interessant werden. gCube wird von einer Digital Libraries Community entwickelt.

Neben diesen Initiativen aus der Infrastruktur-Community gibt es aus der Repository-Community eine Reihe von Aktivitäten zur Virtualisierung von Speicherinfrastruktur. Zu nennen sind hierbei vor allem das **Akubra**-Speicherprojekt des **Fedora** Repository-Systems[57], Smart Storage[58] von EPrints, und die Kooperation von DSpace und Fedora im **DuraSpace** Projekt[59].

All diese Initiativen sind erst während der TextGrid-Projektlaufzeit entstanden und sind im Zwischenraum zwischen Infrastruktur und semantischer Verwaltung von Informationsobjekten angesiedelt. Das rasante Erschließen dieser neuen Technologien durch eine Vielzahl von Projekten bedeutet natürlich eine gewisse Unsicherheit in Bezug auf Nachhaltigkeit und generische Anwendbarkeit von Einzellösungen. TextGrid hat sich daher auch in Abstimmungsprozessen und möglicher Standardisierung stark gemacht. Diese Aktivitäten befinden sich weit-

[56] http://www.gcube-system.org/

[57] Fedora Commons Technology Roadmap, V0.9. Viewed July 2008. http://www.fedora-commons.org/-pdfs/FedoraCommonsRoadmapDraft.pdf

[58] David Tarrant: From open storage to smart storage: enabling EPrints repository preservation. In: The Sun Preservation and Archiving Special Interest Group (PASIG) Spring Meeting (http://sun-pasig.org), May 27-29, 2008, San Francisco. http://eprints.ecs.soton.ac.uk/15818/

[59] DuraSpace. http://expertvoices.nsdl.org/hatcheck/2008/11/11/dspace-foundation-and-fedora-commons-receive-grant-from-the-mellon-foundation-for-duraspace/

gehend erst in ihren Anfängen und können nur durch entschlossene Weiterführung des Einsatzes zu einem Ziel geführt werden.

- Open Grid Forum: Repositories Track. http://www.isgtw.org/?pid=1001110
- IEEE e-Science: Digital Repositories. http://escience2008.iu.edu/
- Digital Curation Conference: RECURSE Repositories. http://www.dcc.ac.uk/events/dcc-2008/
- DReSNeT, Digital Repositories e-Science Network. http://www.dresnet.net/

4. der erfolgten oder geplanten Veröffentlichungen des Ergebnisses nach Nr. 6

a) Bereits erfolgte Veröffentlichungen

- Neuroth, Heike / Jannidis, Fotis / Rapp, Andrea / Lohmeier, Felix: Virtuelle Forschungsumgebungen für e-Humanities. Maßnahmen zur optimalen Unterstützung von Forschungsprozessen in den Geisteswissenschaften. In: Bibliothek. Forschung und Praxis, 2/2009. S. 161-169.
- Marc Wilhelm Küster, Christoph Ludwig und Andreas Aschenbrenner: TextGrid: eScholarship und vernetzte Angebote. In: it - Information Technology, Heft 4 (August) 2009, Themenheft Informatik in den Philologien (Gastherausgeber: Moulin, Claudine / Burch, Thomas / Rapp, Andrea).
- TextGrid - Ein Community-Grid für die Geisteswissenschaften. In: Lajos Herpay, Sonja Neweling, Uwe Schwiegelshohn (Hg.): D-Grid. Die Deutsche Grid-Initiative. Vorstellung der Projekte. Veröffentlichung im Rahmen des D-Grid All Hands Meeting, März 2009. S. 36-37.
- Andreas Aschenbrenner, Tobias Blanke, Neil P Chue Hong, Nicholas Ferguson, Mark Hedges: A Workshop Series for Grid/Repository Integration. In: D-Lib Magazine, January/February 2009, Volume 15 Number 1/2.
- Tobias Blanke, Andreas Aschenbrenner, Marc Küster, Christoph Ludwig: 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.
- Aschenbrenner, Andreas / Meffert, Katja: Wissenschaftliche Infrastruktur in den Geisteswissenschaften? – Eine Wegbeschreibung. In: Braungart, Georg / Gendolla, Peter / Jannidis, Fotis (Hg.): Jahrbuch für Computerphilologie – online, 9.8.2008 (Hg.). http://computerphilologie.tu-darmstadt.de/jg07/aschmeff.html
- Aschenbrenner, Andreas: Feature – Editing, analyzing, annotating, publishing: TextGrid takes the a, b, c to D-Grid. In: iSGTW 30 January 2008, Jg. 54. http://www.isgtw.org/?pid=1000828

- Rapp, Andrea: Das Projekt "TextGrid. Modulare Plattform für verteilte und kooperative wissenschaftliche Textdatenverarbeitung – ein Community-Grid für die Geisteswissenschaften". Chancen und Perspektiven für eine neue Wissenschaftskultur in den Geisteswissenschaften. In: Jahrbuch der historischen Forschung in der Bundesrepublik Deutschland: Berichtsjahr 2006 / hrsg. von der Arbeitsgemeinschaft historischer Forschungseinrichtungen in der Bundesrepublik Deutschland. München: Oldenbourg, 2007, S. 61-68. http://www.ahf-muenchen.de/Forschungsberichte/Jahrbuch2006/AHF_Jb2006_FB_B1_Rapp.pdf
- Neuroth, Heike / Aschenbrenner, Andreas / Lohmeier, Felix: e-Humanities - eine virtuelle Forschungsumgebung für die Geistes-, Kultur- und Sozialwissenschaften. In: Bibliothek. Forschung und Praxis, 3 (2007), S. 272-279. http://www.bibliothek-saur.de/2007_3/272-279.pdf
- "TextGrid: Ein Community Grid für die Geisteswissenschaften". In: Neuroth, Heike / Kerzel, Martina / Gentzsch, Wolfgang (Hg.): Die D-Grid Initiative. Göttingen: Universitätsverlag, 2007. S. 64-66. http://www.univerlag.uni-goettingen.de/content/-list.php?notback=1&details=isbn-978-3-940344-01-4
- Küster, Marc Wilhelm / Ludwig, Christoph / Aschenbrenner, Andreas: TextGrid as a Digital Ecosystem. IEEE DEST 2007, 21.-23. Februar, Cairns, Australien, SPECIAL SESSION 3: e-Humanities for Digital Eco-systems: A Social, Cultural, Economic and Political Agenda. http://www.textgrid.de/fileadmin/TextGrid/veroeffentlichungen/-DigitalEcosystem07_CameraReady-1.pdf
- Aschenbrenner, Andreas / Blanke, Tobias / Dunn, Stuart / Kerzel, Martina / Rapp, Andrea / Zielinski, Andrea: Von e-Science zu e-Humanities – Digital vernetzte Wissenschaft als neuer Arbeits- und Kreativbereich für Kunst und Kultur. In: Bibliothek. Forschung und Praxis, 1 (2007), S. 11-21.http://www.bibliothek-saur.de/2007_1/011-021.pdf
- Aschenbrenner, Andreas / Gietz, Peter / Küster, Marc Wilhelm / Ludwig, Christoph / Neuroth, Heike: 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. http://www.textgrid.de/fileadmin/TextGrid/veroeffentlichungen/ecdl06_textgrid.pdf
- Büdenbender, Stefan / Leuk, Michael: Daten als Dienste: Wörterbücher als Erschließungsinstrumente in der virtuellen Arbeitsumgebung „TextGrid“. Data as services: dictionaries as a means of accessing texts in the virtual research environment “TextGrid”, in: it – Information Technology, Heft 4 (August) 2009, S. 191-196, Themenheft Informatik in den Geisteswissenschaften (Gastherausgeber: Moulin, Claudine / Burch, Thomas / Rapp, Andrea).
- Küster, Marc Wilhelm / Ludwig, Christoph / Aschenbrenner, Andreas: TextGrid: eScholarship und vernetzte Angebote, in: it – Information Technology, Heft 4 (Au-

gust) 2009, S. 183-190, Themenheft Informatik in den Geisteswissenschaften (Gastherausgeber: Moulin, Claudine / Burch, Thomas / Rapp, Andrea).

- Peter Gietz, Andreas Aschenbrenner, Stefan Büdenbender, Fotis Jannidis, Marc Wilhelm Küster, Christoph Ludwig, Wolfgang Pempe, Thorsten Vitt, Werner Wegstein, Andrea Zielinski: TextGrid and eHumanities. In: Proceedings of the Second IEEE International Conference on e-Science and Grid Computing E-SCIENCE '06 . IEEE Computer Society 2006. Amsterdam 2006.
- Werner Wegstein, Vorüberlegungen zu einem parallelen polnisch-deutschen Textkorpus, in: Waldemar Czachur, Marta Czyżewska (Hgg.): Vom Wort zum Text. Studien zur deutschen Sprache und Kultur. Festschrift für Professor Józef Wiktorowicz zum 65. Geburtstag, Warszawa 2008, S. 145 – 152.
- Christian Schneiker, Dietmar Seipel, Werner Wegstein, Klaus Prätor: Declarative Parsing and Annotation of Electronic Dictionaries, in: Bernadette Sharp. Michael Zock (Eds.), Natural Language Processing and Cognitive Science. Prodeedings of the 6th International Workshop on Natural Language Processing and Cognitive Science – NLPCS 2009 (in conjunction with ICEIS 2009), Milan, Italy, May 2009, S. 122 – 132.
- Christian Schneiker, Dietmar Seipel, Werner Wegstein: Schema and Variation: Retro-Digitizing Printed Dictionaries, in: Proceedings of the Third Linguistic Annotation Workshop, sponsored by the Association for Computational Linguistics Special Interest Group for Annotation (ACL-SIGANN), (in conjunction with the Joint conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing), 2. – 7. August, Singapore 2009, S. 82 – 89.
- [url = http://www.aclweb.org/anthology/W09-3014].

b) Geplante Veröffentlichungen

- Pempe, Wolfgang / Zielinski, Andrea / Gietz Peter / Haase, Martin / Funk, Stefan / Simon, Christian: TEI-Documents in the Grid. In: Literary and Linguistic Computing, 2009, 24(3), S. 267-279.
- Kerzel, Martina / Mittelbach, Jens / Vitt, Thorsten: TextGrid – Virtuelle Arbeitsumgebung für die Geisteswissenschaften, in: Künstliche Intelligenz, Themenheft „Kulturerbe und Künstliche Intelligenz“, Heft 4.2009 (im Druck).
- Christian Schneiker, Dietmar Seipel, Werner Wegstein (University of Wuerzburg): Declaratively Creating and Processing XML/TEI Data. Paper accepted 10 July2009 for the TEI Members Meeting, November 2009, Ann Arbor (USA).
- Stanislava Grigorova, Studien zur Digitalisierung von Campes ‘Verdeutschungswörterbuch. Phil. Diss. Universität Würzburg, Juni 2009 (im Druck).

III. Anlagen

Annex A - Partnerliste

Fachwissenschaftliche Partner

- **Technische Universität Darmstadt**

  Schwerpunkt der Arbeit im AP2: Entwicklung Community-spezifischer Werkzeuge, wie Annotations- und Analyse-Tools

- **Niedersächsische Staats- und Universitätsbibliothek Göttingen (Projektleitung)**

  Schwerpunkt der Arbeit im AP3: Anbindung der Community-Tools und Vorschläge für Entwicklungen an der Integrations-Plattform und AP6: Projektmanagement und Öffentlichkeitsarbeit

- **Institut für Deutsche Sprache Mannheim**

  Mitarbeit im AP2: Entwicklung Community-spezifischer Werkzeuge, wie Annotations- und Analyse-Tools

- **Universität Trier**

  Schwerpunkt der Arbeit im AP5: Semantic Web und TextGrid = Semantic TextGrid und Mitarbeit im AP2: Entwicklung Community-spezifischer Werkzeuge, wie Annotations- und Analyse-Tools

- **Fachhochschule Worms**

  Schwerpunkt der Arbeit im AP1: Inhaltliche Studie mit Empfehlungen über die Nachnutzbarkeit internationaler Editionstools und Mitarbeit im AP3: Anbindung der Community-Tools und Vorschläge für Entwicklungen an der Integrations-Plattform und AP4: Entwicklung der Community Muster-Applikation

- **Universität Würzburg**

  Schwerpunkt der Arbeit im AP4: Entwicklung der Community Muster-Applikation

Kommerzielle Partner

- **DAASI International GmbH**

  Aufbau des TextGridRep insbes. Grid-Anbindung (TG-Crud) und der Authentifizierungs- und Autorisierungsinfrastruktur (TG-Auth); Tools und Vorschläge für Entwicklungen an der Integrations-Plattform.

  Mitarbeit in AP5: Entwicklung des Wörterbuch-Link-Editiors und AP2 Workflowmanager.

- **Saphor GmbH**

  Mitarbeit im AP2: Entwicklung von Community-spezifischen Werkzeugen, wie Annotations- und Analyse-Tools

Annex B - Arbeitspakete und Deliverables

AP1 – Inhaltliche Studie mit Empfehlungen über die Nachnutzbarkeit internationaler Editionstools

- M1.1. Text Processing
- R1.1. Text Processing
  (http://www.textgrid.de/fileadmin/TextGrid/reports/TextGrid_Report_1_1.pdf)
- M1.2. Linking
- R1.2. Linking
  (http://www.textgrid.de/fileadmin/TextGrid/reports/TextGrid_Report_1_2.pdf)
- M1.3. Text Retrieval
- R1.3. Text Retrieval
  (http://www.textgrid.de/fileadmin/TextGrid/reports/TextGrid_Report_1_3.pdf)
- M1.4. Publishing
- R1.4. Publishing
  (http://www.textgrid.de/fileadmin/TextGrid/reports/TextGrid-R1_4_Publishing.pdf)
- M1.5. Management von Workflow, Access, Kommunikation und Nutzer
- R1.5. Management von Workflow, Access, Kommunikation und Nutzer
  (http://www.textgrid.de/fileadmin/TextGrid/reports/TextGrid_Report_1_5.pdf)
- M1.6. Ontologien
- R1.6. Ontologien
  (http://www.textgrid.de/fileadmin/TextGrid/reports/TextGrid_Report_1_6_Ontologien.pdf)

*AP2 – Entwicklung Community – spezifischer Werkzeuge (Annotations-, Analyse – Tools)

- M2.1. Tokenizer, Workflow-Editor
- M2.2. Lemmatisierung, XML Editor, Rich Client Platform (GUI)
- R2.1. TextGrid Tools I
  (http://www.textgrid.de/fileadmin/TextGrid/reports/TextGrid_Report_2_1.pdf)
- M2.3. Recherchetool, Streaming-Editor I, Datei-/Rechtemanagement, Metadaten-Annotation, grafischer Link-Editor, Bild-Segmentierung, Link-Editor Text, Bibliographietool, Sortieren
- R2.2. Dokumentation

(auf der Homepage: R2.2 TextGrid Tools II:

http://www.textgrid.de/fileadmin/TextGrid/reports/TextGrid-R2.2_ToolsII.pdf)

- M2.4. Streaming-Editor II, Kollationierung
- M2.5. Text Publisher (Print), Text Publisher (Web), OCR
- R2.3. Dokumentation (auf der Homepage: User's Manual TextGrid-Tools: http://www.textgrid.de/fileadmin/TextGrid/reports/Report_2.3_final.pdf)

AP3 – Anbindung der Community – Tools und Vorschläge für Entwicklungen an der Integrations – Plattform

- M3.1. Ermittlung der Middleware Anforderung aus den anderen APs
- R3.1. Bericht über Evaluation der vorhandenen Grid-Middleware-Standards und Software-Pakete unter Berücksichtigung der geplanten Dienste der Integrationsplattform und der in M3.1. ermittelten Anforderungen

  (http://www.textgrid.de/fileadmin/TextGrid/reports/TextGrid_Report_3_1.pdf)
- R3.2. Spezifikation der Architektur für die TextGrid-Middleware

  (auf der Homepage: R3.2. TextGrid-Architektur:

  http://www.textgrid.de/fileadmin/TextGrid/reports/TextGrid_Report_3_2.pdf)
- M3.2. Implementierung eines Prototyps der TextGrid-Middlewareplattform mit Anbindung an die IP
- R3.3. Spezifikation aller von der TextGrid-Middlewareplattform zu bedienenden Grid-Schnittstellen (Version 1)

  (wird fortlaufend aktualisiert – verfügbar auf Anfrage)
- R3.4. Middleware-Tests, unter Berücksichtigung der Werkzeuge (AP2) und Musterapplikationen (AP4)

  (wird fortlaufend aktualisiert – verfügbar auf Anfrage)
- R3.5. User-Manual zur Anbindung von Werkzeugen an die TextGrid-Middleware

  (auf der Homepage R3.5. TextGrid Manual – Tool Development:

  http://www.textgrid.de/fileadmin/TextGrid/reports/R3_5-manual-tools.pdf)
- M3.3. Unterstützung und Evaluation bei der Anwendung des Prototyps für Musterapplikationen
- M3.4. Implementierung einer Produktivversion der TextGrid-Middleware
- R3.6. User-Manual zur Installation eines Datengridknotens (für assoziierte Textarchive)

  (http://www.textgrid.de/fileadmin/TextGrid/reports/TextGrid_Report_3_6-Datengrid-Knoten.pdf)

AP4 – Entwicklung der Community Muster – Applikation

- M4.1. Erstellung der relevanten Corpora (Digitalisierung) als Testbed
- R4.1. Handbuch: Dokumentation der Tools und potentiellen Workflows

  Die Dokumentation der Tools wurde im Hinblick auf Überschneidungen mit Reports zu AP2 dorthin ausgegliedert. Titel des Reports 4.1 „Zur Digitalisierung von Primärquellen für die TextGrid-Umgebung: Modellfall Campe-Wörterbuch“: http://www.textgrid.de/fileadmin/TextGrid/reports/TextGrid_R4_1.pdf)
- M4.2. Prototypischer Durchlauf der Corpora durch die verschiedenen Tools mit Evaluation (Empfehlungen für weitere Anpassungen/Entwicklungen)

AP5 – Semantic Web und TextGrid = Semantic TextGrid

- M5.1. WB-Link für die Vernetzung von Wörterbüchern
- R5.1. Reports über Ontologie-Software (s. AP1)

  (auf der Homepage: R5.1.: Ontologien und Wortnetze:

  http://www.textgrid.de/fileadmin/TextGrid/reports/TextGrid_Report_5_1.pdf)
- M5.2. WB-Link zur Verknüpfung von Wörterbüchern und Primärquellen unter Verwendung von GermaNet
- M5.3. Meta-Lemmalisten
- M5.4. Lemmatisierung von Belegstellen-Zitaten im Verbund mhd. Wörterbücher und im Deutschen Wörterbuch
- M5.5. Erstellung von Umkehrwörterbüchern
- R5.2. Report über Tests mit philologisch-linguistischen Verfahren zur Erstellung der Meta – Lemmaliste

  (http://www.textgrid.de/fileadmin/TextGrid/reports/TextGrid_R5_2.pdf)
- M5.6. Entwicklung geeigneter Metadata Application Profiles
- R5.3. Abschlussreport

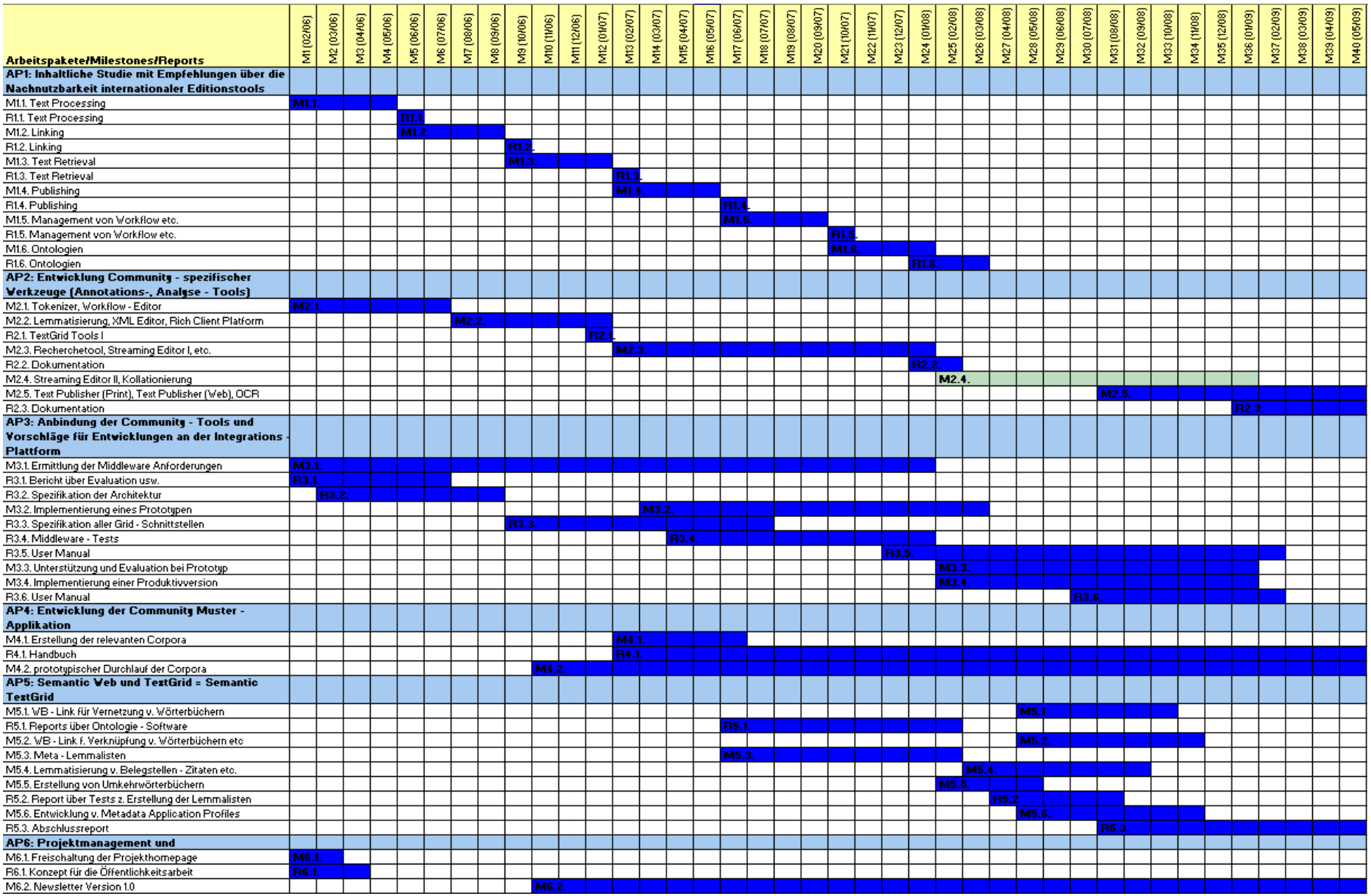
  Der AP-spezifische Abschlussreport ist nach allgemeinem Projektbeschluss im allgemeinen TextGrid-Abschlussbericht aufgegangen.

AP6 – Projektmanagement und Öffentlichkeitsarbeit

- M6.1. Freischaltung der Projekthomepage
- R6.1. Konzept für die Öffentlichkeitsarbeit

  (http://www.textgrid.de/fileadmin/TextGrid/reports/TextGrid_R6_1.pdf)
- M6.2. Newsletter Version 1.0 (danach alle drei Monate)

TextGrid Milestones/Reports – Ist – Zustand



Annex C - Arbeitsgruppen

AG1 – Architektur

Aufgaben: Die AG Architektur leistet technische Konzeptionsarbeit zur Gesamtarchitektur von TextGrid (z.B. Architekturdiagramm Schichtenmodell, Schnittstellen, Codierungsfragen). Die Arbeitsgruppe ist somit an der Schnittstelle und übergreifend zwischen Grid Infrastruktur und den Tools angesiedelt.

Ziele:

- Tool-übergreifende Koordination der Integration der TextGrid Software
- Definition der Schnittstellen zur Integrationsplattform, Replika-Management (August 2007)
- Definition generischer Konfigurationsinterfaces für die Services im TextGridLab
- Testframework, Überführung in Produktionsdienste (Juni 2008)
- Integration von existierenden Archiven: Importtool (inkl. Schnittstellen), Anleitung für Aufbau eines Text/Grid-Knotens
- Open Source gehen, inkl. Codezyklen, Bugzille, Aufbau Forum/Helpdesk, etc.

Moderatoren:

- Martin Haase, DAASI
- Christoph Ludwig, FH Worms

AG2 – Textformate

Aufgaben: Die AG Textformate spielt eine übergreifende Rolle zur AG Archive, AG Wörterbücher und AG Linguistische Korpora. Hier werden die fokussierten Diskussionen zu Meta-/Datenformaten, Abläufen und Anforderungen an Tools TextGrid-weit zusammengeführt. Ziel ist die Definition einer Kernkodierung für historisch-kritische Editionen, Dramen, Gedichte, Prosatexte, Linguistische Korpora und Wörterbücher sowie die Definition von Metadatenmodell und -formaten; außerdem sollen die Ergebnisse für künftige TextGrid-Anwender dokumentiert und mit Fachkollegen diskutiert werden.

Ziele:

- Kernkodierung …
  → über alle Textsorten definieren (bis Ende 2007)
  → endnutzergerecht dokumentieren (mit AP4, Anfang 2008)
  → mit Fachkollegen evaluieren (bis Mitte 2008)
- Betreuung Metadatenformat
- Evaluation der Recherchemöglichkeiten

- Koordination zwischen den Sub-AGs

Moderatoren

- Thorsten Vitt, TU Darmstadt
- Andrea Zielinski, IDS Mannheim

AG3 – Wörterbücher

Aufgaben: Interne wie externe Ressourcen im Bereich „Wörterbücher“ werden durch eine spezifische Schnittstelle bzw. die Überführung in eine eigene Datenbank in TextGrid eingebunden. Die AG

- definiert eine Schnittstelle, um aus TextGrid heraus Suchanfragen an extern bereits existierende Wörterbuchinstallationen zu stellen. Die Schnittstelle basiert auf einem Kernset inhaltlich relevanter Elemente und Suchkategorien, die eine interaktive Suche über alle eingebundenen Wörterbücher ermöglicht;
- definiert ein geeignetes Schema zur Kodierung von Wörterbüchern auf der Basis von TEI P5. Das Schema dient gleichermaßen als technische Basis für die Überführung externer Daten in TextGrid, als Mittel zur Validierung von Wörterbüchern und als Leitfaden zu ihrer Erstellung;
- verfasst einen Leitfaden, der die Schnittstelle und ihre Gebrauchsweise beschreibt und auch beschreibt, auf welche Ressourcen mit Hilfe der Schnittstelle zugegriffen werden kann;
- bindet externe Partner mit ein, und bringt Probematerial ein, darunter historische Lexika (z.B. Grimm, Campe) bzw. solche, die historische (Lexer, BMZ) oder dialektale (Dialektwörterbuchverbund) Varietäten dokumentieren;

Ziele:

- Spezifikation von Schnittstellen zu existierenden Wörterbüchern (Mai 2007)
- Umsetzung der Schnittstellen gemeinsam mit existierenden Wörterbüchern (Winzer-Wörterbuch, Luxemburger WB etc.), ggf. auch aus Frankreich oder anderen Sprachen
- Evaluation der Einbindung von eLexiko
- Technische Dokumentation der Schnittstelle

Moderatoren:

- Stefan Büdenbender, Uni Trier
- Christian Graiger, Uni Würzburg

AG4 – Linguistische Korpora

Aufgaben: Informationsressourcen (aus externen Quellen) im Bereich „Linguistische Korpora“ sollen in Zukunft in TextGrid integriert werden können. Diese werden mit TextGrid Tools

bearbeitet und sollen daher mit dem TextGrid Austauschformat kompatibel sein. Diese Arbeitsgruppe

- erstellt eine Karte linguistischer Ressourcen (Formate und technische Eigenschaften)
- beschreibt die Anforderungen von linguistischen Ressourcen an TextGrid Meta-/Datenformate (siehe das TextGrid Austauschformat der AG Textformate)
- beschreibt, wie linguistische Ressourcen in TextGrid eingebunden werden können, möglichst eng angelehnt an die Arbeit der AG Architektur
- bindet dabei relevante externe Initiativen mit ein
- verfasst einen Leitfaden, der die Einbindung linguistischer Ressourcen in TextGrid beschreibt

Ziele:

- Adelung als Korpus
- Input zum Recherchetool

Moderatoren:

- Andrea Zielinski, IDS Mannheim

AG5 – Print

Aufgaben: Die Arbeitsgruppe Print evaluiert Umsetzungsmöglichkeiten für das Tool „Text Publisher (Print)“ – was kann im Rahmen der TextGrid Ressourcen geschafft werden, und was muss in ein externes Projekt ausgelagert werden. Die Arbeitsgruppe hat im Februar 2008 den DFG Antrag Print eingereicht, der im Oktober 2009 bewilligt wurde.

Moderatoren:

- FH Worms

Annex D - Fachbeirat

TextGrid setzt auf inhaltliche, technische und strategische Beratung durch einen Fachbeirat. Das trägt dazu bei, dass Maßnahmen und Entwicklungen im Projekt zielgerichtet sind und dem aktuellen Stand der Forschung auf informationstechnologischem und fachwissenschaftlichem Gebiet entsprechen.

Dem TextGrid-Fachbeirat gehören an:

- Lou Burnard (Oxford University)
- Reiner Diedrichs (Gemeinsamer Bibliotheksverbund)
- Dr. Martin Doerr (Institute of Computer Science, Griechenland)
- Prof. Dr. phil. Kurt Gärtner (Universität Trier)
- Dr. phil. Roland Kamzelak (Deutsches Literaturarchiv Marbach)
- Hans-Jörg Lieder (Staatsbibliothek zu Berlin)
- Peter Robinson (University of Birmingham)
- Laurent Romary (MPDL, Max-Planck-Gesellschaft)
- Prof. Hans E. Roosendaal (Universität Twente, Niederlande)
- Abby Smith (Council on Library and Information Resources, Washington)
- Prof. Dr. Steffen Staab (Universität Koblenz)
- Dr. Klaus Ullmann (Deutsches Forschungsnetz)