



Bundesministerium
für Bildung
und Forschung



TextGrid – Vernetzte Forschungsumgebung in den eHumanities

*Nachtrag zu der am 27.06.2008 eingereichten
Vorhabenbeschreibung*

Einsender:

Dr. Norbert Lossau

Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB)

Platz der Göttinger Sieben 1

37070 Göttingen

Tel.: (0551) 395212

E-Mail: lossau@sub.uni-goettingen.de

Laufzeit: 3 Jahre (1. Juni 2009 bis 31. Mai 2012)

- Öffentliche Version -

Inhaltsverzeichnis

1. Einführung	3
2. Aktualisierte Arbeitspakete-Struktur	6
AP 1 - TextGrid Repository	6
AP 2 - TextGrid Laboratory	7
AP 3 - Strukturelle und organisatorische Nachhaltigkeit	8
AP 4 - Coaching / Community Bildung	9
AP 5 - Technische, semantische und organisatorische Interoperabilität	11
AP 6 - Fachwissenschaftliche Nachhaltigkeit	11
AP 7 - eHumanities-Netzwerk	13
AP 8 - Nationale/internationale Liaison, Projektmanagement, Öffentlichkeitsarbeit	13
3. TextGrid Konsortium und Kompetenzen der Partner	14
4. Optimierung der Zusammenarbeit im TextGrid Konsortium	18
Anhänge	21
I Arbeitspakete	21
AP 1 - TextGrid Repository	21
AP 2 - TextGrid Laboratory	28
AP 3 - Strukturelle und organisatorische Nachhaltigkeit	30
AP 4 – Coaching / Community Bildung	32
AP 6 - Fachwissenschaftliche Nachhaltigkeit	36
AP 8 - Projektmanagement, Öffentlichkeitsarbeit, RI-Liaison	51
II Ressourcenplanung	55
Gesamtpersonenmonate aller Partner	55
Deliverables & Milestones (in zeitlicher Abfolge)	56
III WissGrid Antrag	62
IV „Digitale Bibliothek“	62
Begründung für den Erwerb der Texte der Digitalen Bibliothek	62
Abkürzungsverzeichnis:	67

1. Einführung

Der vorliegende Nachtrag hat die Aufgabe, die aktuellen Entwicklungen seit dem Einreichen des Antrags **TextGrid - Vernetzte Forschungsumgebung in den eHumanities** im Juni 2008 aufzugreifen und zu berücksichtigen. Die aktuellen Entwicklungen ergeben sich hauptsächlich aus dem Zusammengehen der fünf wissenschaftlichen Community-Projekte AstroGrid, C3Grid, HepGrid, MediGrid und TextGrid zu dem Forschungsverbund **WissGrid - Grid für die Wissenschaft**. Der WissGrid-Antrag wurde im März 2009 (DGRID3A-08) in einer überarbeiteten Fassung eingereicht und im Rahmen des D-Grid All-Hands Meetings am 26. März positiv begutachtet. Die Überarbeitung des WissGrid-Antrags ergab sich hauptsächlich aus der Notwendigkeit, die Ziele präziser zu formulieren und eine Abgrenzung zu fachspezifischen Entwicklungen deutlicher herauszuarbeiten. Die Gutachter betonten zwar „**erneut die Notwendigkeit der Bereitstellung von Querschnittsfunktionen für die Gewinnung und Etablierung weiterer (vorwiegend wissenschaftlich orientierter) Communities**“, zugleich merkten die Gutachter aber auch an, dass der fachspezifische Anteil im WissGrid-Vorhaben überwiegt und eine gemeinsame, generische Zielsetzung nicht ausreichend herausgearbeitet wurde.

Die Anforderungen der Gutachter führten zu einer Aktualisierung des WissGrid-Antrags. Das WissGrid-Vorhaben hat drei miteinander verbundene Ziele, die alle einen Community-übergreifenden Charakter besitzen:

1. Aufbau einer nachhaltigen Organisationsstruktur für die Grid-Nutzer im akademischen Bereich. Die bisherigen und neuen fachwissenschaftlichen Community-Grids organisieren sich in einer gemeinsamen Vertretungsstruktur innerhalb der deutschen Grid-Initiative und bündeln so die Interessen der Nutzer gegenüber verschiedenen anderen Interessenvertretungen wie Ressourcen-Providern (Gauß-Allianz) oder Förderinstitutionen.
2. Weitergabe und Ausbau der Erfahrungen beim Aufbau neuer Community-Grids im akademischen Bereich zur Erweiterung des Nutzerkreises von D-Grid. Dazu werden die vorhandenen Kompetenzen der bisherigen Community-Grids konzentriert und fachübergreifend zur Verfügung gestellt.
3. Erstellung und Umsetzung eines Konzepts zur Langzeitarchivierung von Forschungsdaten. Dieses Konzept soll so flexibel sein, dass auch den Anforderungen neuer Communities Rechnung getragen werden kann.

Das WissGrid-Vorhaben verfolgt nun verstärkt die Entwicklung von Blaupausen und Konzepten als Grundlage zum Aufbau neuer Community-Grids sowie die Nachnutzung von bereits erworbenen Erfahrungen und Know-How.

Da die Communities ihre jeweils eigenen fachspezifischen Dienste und Werkzeuge entwickeln (gefördert durch das jeweilige Fachreferat), die dann wegen der speziellen Ausrichtung auf die Bedürfnisse der jeweiligen Community nicht vom DGI oder der DGSE

übernommen werden, sind die Communities auch für die Pflege dieser Werkzeuge verantwortlich.

Für das hier skizzierte Vorhaben TextGrid ergeben sich dadurch eine Reihe von positiven Entwicklungen:

- TextGrid als eigenständige Fach-Community wird zusammen mit anderen bereits etablierten Fach-Communities (z.B. AstroGrid, MediGrid) innerhalb eines WissGrid-Gremiums die Interessen der jeweiligen Wissenschaftlergruppen bündeln und gegenüber den Ressourcen-Anbietern vertreten. So muss nicht jeder Fachwissenschaftler oder jede Fach-Community von Neuem den Dialog mit den Ressourcen-Anbietern suchen und in Verhandlungen (z.B. Service-Level-Agreements) eintreten. Um die Interessen der Fachwissenschaftler besser vertreten und auch perspektivisch mit Förderern diskutieren zu können, wird TextGrid eine eigenständige Rechtsform anstreben (vgl. AP 3).
- Mehrere TextGrid-Partner sind auch im WissGrid-Konsortium vertreten und bringen dort ihre Expertise ein, um einerseits Grid-Technologien und Infrastrukturen für alle fachspezifischen Disziplinen nachhaltig auszubauen und andererseits als Expertenteam zur Verfügung zu stehen, um neue geisteswissenschaftliche Disziplinen an das Grid heranzuführen. Konkret wird dies mit der linguistischen Community (MPI für Psycholinguistik, MPDL, IDS Mannheim) realisiert, die sowohl über das WissGrid-Verbundvorhaben technologisch an das Grid herangeführt wird als auch im Rahmen des TextGrid-Vorhabens fachspezifische Dienste und Tools entwickelt, die in die TextGrid-Infrastruktur eingebunden werden.
- TextGrid wird seine Erfahrungen und Kompetenzen bei der Konzeption und Entwicklung eines generischen Forschungsdaten-Archivs im WissGrid-Verbund einbringen, um auch sicher zu stellen, dass die ganze Breite an komplexen und heterogenen Inhalten und Forschungsdaten berücksichtigt wird (in den Naturwissenschaften sind die Datenmengen zwar größer, aber wesentlich homogener). Damit ist eine langfristige Speicherung *und* Nachnutzung dieser Daten und Inhalte gewährleistet.

Durch die Fokussierung der Ziele von WissGrid auf die Übertragbarkeit der bisher erreichten Technologien auf neue fachspezifische Communities (AP 2 WissGrid), den Aufbau und die Umsetzung einer nachhaltigen organisatorischen Infrastruktur (AP 1 WissGrid) und die Konzeption und prototypische Entwicklung eines Forschungsdatenarchivs (AP 3 WissGrid) sind einige relevante und zentrale Bausteine, die zur Förderung im Rahmen des D-Grid BMBF Referates vorgesehen waren (ehemals Teil I, der an Referat 524 gerichteten Vorhabenbeschreibung vom 27.06.2008), "weggebrochen". Dies betrifft v.a. die generischen fachspezifischen Dienste und Services, die zwar für die Textwissenschaften im Allgemeinen von Belang sind, aber dem Anspruch, für ALLE Wissenschaften relevant zu sein, nicht genügen.

Dies hat zur Konsequenz, dass die im Juni 2008 eingereichte Vorhabenbeschreibung um vier inhaltliche Komponenten ergänzt werden muss, die zwar generisch textwissenschaftlich übergreifend, aber zu fachspezifisch für das D-Grid Referat sind, dessen Fokus die Grundlagenförderung von Basis-Grid-Technologien ist:

- fachspezifische Anpassungen und Erweiterungen des in WissGrid entwickelten Forschungsdatenarchivs (vgl. AP 1)
- allgemeine textwissenschaftliche Tool-Entwicklungen (vgl. AP 2, z.B. Metadaten-Editor, Recherche-Tool)
- Einrichten von TextGrid als eigenständige Rechtsform, um im WissGrid-Kontext eine (geistes-)wissenschaftliche Interessensvertretungsstruktur fachübergreifend z.B. gegenüber den Ressourcen-Anbietern dauerhaft etablieren zu können (AP 3)
- TextGrid-spezifische Coaching- sowie Schulungsmaßnahmen und Community-Bildung (AP 4)

Der hier vorgelegte Nachtrag berücksichtigt diese aktuellen Entwicklungen und konzentriert sich auf die Ergänzungen und Anpassungen, basiert aber selbstverständlich auf dem im Juni 2008 eingereichten TextGrid-Antrag.

2. Aktualisierte Arbeitspakete-Struktur

In diesem Kapitel werden die aus wissenschaftlicher Sicht notwendigen Ziele des Verbundvorhabens skizziert und in Anlage I die AP-Struktur dargestellt, soweit sich Ergänzungen und Korrekturen ergeben haben. Die im Juni 2008 eingereichte TextGrid-Vorhabenbeschreibung behält ihre Gültigkeit in den Bereichen, in denen sich keine Aktualisierungen ergeben haben und beschreibt in aller Ausführlichkeit die Arbeitspakete mit Zielen, Vorgehen und Arbeitsplan.

Eine Veränderung besteht auch darin, dass die Personenmonate-Verteilung in den Arbeitspaketen nicht mehr alle Kollaborationen abzubilden versucht, sondern durch eine stärkere Konzentration an Personenmonaten den Aufbau von spezifischen Kompetenzen ermöglicht. Diese derart aufgebauten Kompetenzen werden durch interne Schulungen an die jeweils relevanten TextGrid-Partner weitergegeben. Es versteht sich von selbst, dass alle Projektpartner über Arbeitsgruppen (z.B. AG Architektur, AG Fachwissenschaftler etc.) an zentralen Koordinations- und Entwicklungsaufgaben beteiligt sind. Dieses Vorgehen spiegelt auch die Erfahrungen und Ergebnisse wider, die in einer internen Selbstevaluation gesammelt werden konnten (vgl. Kap. IV).

AP 1 - TextGrid Repository

"Alle Wissenschaftsorganisationen sehen einen dringenden Handlungsbedarf für die systematische Sicherung von im Forschungsprozess erzeugten Daten." (Allianz der deutschen Wissenschaftsorganisationen). Dazu zählen Forschungsdaten, abgeleitete Daten, und Publikationen. Im Rahmen des WissGrid-Vorhabens ("Grid für die Wissenschaft") wird ein generisches Forschungsdatenarchiv konzipiert und prototypisch entwickelt, das von Community-Grids als Basis für die Umsetzung ihrer unterschiedlichen fachlichen Anforderungen verwendet werden kann. Die vorhandenen Archivierungstechnologien in den wissenschaftlichen Communities (AstroGrid, C3Grid, HepGrid, MediGrid und TextGrid) dienen dabei als Ausgangsbasis für übergreifende, Grid-fähige Langzeitarchivierungsstrategien.

Das hier beschriebene Arbeitspaket wird in AP 1.1 (mit null Personen-Monaten) die Diskussionen, Konzeptionen und prototypischen Entwicklungen in WissGrid unter Berücksichtigung der geisteswissenschaftlichen Anforderungen (z.B. Sammlungskonzepte, fachspezifische Guidelines, Definition relevanter Formate etc.) eng begleiten, um sicher zu stellen, dass die in den Geisteswissenschaften anfallenden Forschungsdaten und Inhalte dauerhaft gespeichert und nachgenutzt werden können.

In AP 1.2 wird das in WissGrid entwickelte generische Forschungsdatenarchiv für TextGrid übernommen und angepasst, so dass ein TextGrid Repository (TextGridRep) entsteht, in dem geisteswissenschaftliche Inhalte dauerhaft nachgewiesen und für die Öffentlichkeit recherchierbar sind. Wesentlich ist dabei die Verknüpfung der Vorteile des WissGrid-Forschungsdatenarchivs zur Datenverwaltung und Langzeitarchivierung mit den erprobten Entwicklungen aus TextGrid für eine offene virtuelle Forschungsumgebung. Dies kann durch eine Abbildung der TextGrid Utilities (TG-search, TG-crud, TG-auth) auf das

Forschungsdatenarchiv erreicht werden. Die Implementierung und Integration von Schnittstellen zu anderen geisteswissenschaftlich relevanten Datenzentren und Archiven bildet einen weiteren Schwerpunkt, der auch für AP 1.3 eine wesentliche Voraussetzung ist.

In AP 1.3 werden Lösungen für die Fragestellung skizziert, auf welcher technischen und inhaltlichen (semantischen) Ebene eine Kooperation zwischen eSciDoc und TextGrid realisiert werden kann. Ansätze sind sowohl in der nachhaltigen technischen Wartung des Repositoriums denkbar, also auch im kontinuierlichen Austausch bzw. in der gemeinsamen Entwicklung von fachwissenschaftlichen Daten und Diensten.

AP 2 - TextGrid Laboratory

Neben dem TextGridRep stellt das sog. TextGridLab(oratory) die zweite wichtige Einheit innerhalb der TextGrid-Architektur dar, die weiterentwickelt und in den nachhaltigen Produktivbetrieb überführt werden muss. Diese Architektur wurde ausführlich im Antrag vom Juni 2008 beschrieben. Kurz gesagt umfasst das TextGridLab die *Services*, die mit der Middleware interagieren, ferner die Schicht der Benutzeroberfläche mit interaktiven Tools sowie Zugriff auf die einzelnen *Services* und *Utilities*. Damit wird dem nicht technisch versierten Benutzer ohne Grid-Erfahrung ein intuitiver und interaktiver Zugang zu Grid-Technologien ermöglicht.

Aufgabe von AP 2 ist die Stabilisierung und Optimierung des TextGridLab in den Bereichen, die generische textwissenschaftliche Arbeits- und Forschungszusammenhänge abbilden und unterstützen. Diese Tools sind einerseits so spezifisch textwissenschaftlich, dass sie fachwissenschaftliche Anforderungen berücksichtigen müssen, andererseits so generisch, dass sie im Rahmen der wissenschaftlichen Zugänge von praktisch allen Textwissenschaften verwendet werden. Als erster "Einstiegspunkt" in TextGrid muss auch und gerade in diesem Bereich ein besonderes Augenmerk auf die Beteiligung der Community an der Entwicklung und die Berücksichtigung von Nutzerfeedback gelegt werden.

In AP 2.1 'Allgemeine textwissenschaftliche Werkzeuge' werden die übergreifenden Funktionalitäten und Werkzeuge des TextGridLab – immer in enger Abstimmung mit AP 6, in dem die spezifisch fachwissenschaftlichen Tools entwickelt werden – (weiter)entwickelt und in den Produktivbetrieb überführt. Zu diesen Werkzeugen gehören

- Grund-GUI mit Startbildschirm: Dies ist der erste Kontakt des Nutzers mit TextGrid, er muss daher klar und übersichtlich gegliedert und intuitiv benutzbar sein.
- der Projektbrowser zum Betrachten und Verwalten der Objekte in den Projekten des Benutzers: Die Möglichkeiten zur Strukturierung und Organisation von Projekten sind in TextGrid sehr ausdifferenziert, weil dies die Arbeitsprozesse geisteswissenschaftlicher Projektzusammenarbeit abbildet; eine enge Abstimmung mit den Communitybedürfnissen ist daher notwendig.
- die Oberfläche zur Organisation von Objekten in Sammlungen: Die Entwicklung des Sammlungskonzeptes ist Gegenstand von AP 1 (s.o.), diese Konzepte müssen in der GUI abgebildet und handhabbar gemacht werden.

- das Administrationstool zur Rechteverwaltung: Die Rechte und Rollen sind ebenfalls sehr ausdifferenziert, was an den spezifischen Daten, Arbeitsweisen und Publikationsstrategien der geisteswissenschaftlichen Communities liegt. Eine nutzerfreundliche Administrationsumgebung für ihre Organisation unterstützt den Einsatz kollaborativer Arbeitsweisen in den Geisteswissenschaften.
- der Metadaten-Editor: Die Erfassung und Verwaltung der spezifischen Metadaten ermöglicht die effektive Speicherung, die Recherche, den Zugriff und die Organisation der Daten.
- der Workflow-Editor: Er macht bestimmte Arbeitsabläufe und die empirisch erzielten Ergebnisse reproduzierbar.
- das Recherchetool: Es ermöglicht innerhalb von TextGrid neben einer Projekt- bzw. Dokument-spezifischen Suche eine übergreifende Suche auch in heterogenen Textkorpora bzw. über alle TextGrid-Dokumente.
- die technischen Komponenten der Hilfe

In AP 2.2 'Evaluierung der Möglichkeit einer Migration des TextGridLab zu einer Browser-basierten Lösung' wird im Rahmen von wenigen PMs geprüft, welchen Aufwand eine solche Migration tatsächlich erfordert. Um den Fachwissenschaftlern den Zugang zu TextGrid zu erleichtern, wäre es wünschenswert, ihnen Herunterladen und Installation eines eigenen Client zu ersparen. Hier soll daher überprüft werden, ob und in welchem Umfang es möglich ist, diesen Client als Browser-basierte Lösung zu implementieren. Dadurch würde die Mobilität des Zugriffs auf die Arbeitsumgebung deutlich gesteigert werden.

AP 3 - Strukturelle und organisatorische Nachhaltigkeit

WissGrid verfolgt das Ziel, für alle Fachcommunity-Projekte eine nationale organisatorische und förderpolitische Struktur aufzubauen. Langfristiges Ziel ist es, ein neues Modell für die übergreifende Wissenschaftsförderung zu etablieren, bei dem die Förderung von Einzelwissenschaftlern und Ressourcen-Anbietern um eine dritte Förderlinie für fachwissenschaftliche Communities ergänzt wird. Dazu wird in WissGrid eine geeignete Organisationsstruktur bzw. ein kooperatives Trägermodell entwickelt, in der sich die bisherigen und neuen fachwissenschaftlichen Community-Grids in einer gemeinsamen Vertretungsstruktur innerhalb der deutschen Grid-Initiative organisieren¹. Weitere Aufgaben einer solchen Community-Vertretung bestehen darin,

- die Interessen der Fachwissenschaftler gegenüber anderen Interessensvertretungen wie IT-Ressourcen-Providern (Gauß-Allianz) und Förderinstitutionen zu bündeln, sowie
- neue Community-Grids zu beraten und zu „zertifizieren“.

Um in diesem Rahmen die Belange geisteswissenschaftlicher Communities nachhaltig vertreten zu können, soll in AP 3.1 eine geeignete Rechtsform für die Organisationsstruktur

¹ Siehe Anhang III, WissGrid-Antrag, AP1, S.9-13.

von TextGrid gefunden und gegründet werden. Ein zu entwerfendes Modell wird dabei verschiedene mögliche Rollen identifizieren und anhand ihrer Rechte und Pflichten definieren. Eine Evaluation verschiedener nationaler und internationaler Rechtsformen wird der Auswahl vorausgehen. Vor der tatsächlichen Gründung müssen abhängig von der Auswahl gewisse Voraussetzungen erfüllt (z.B. eine Satzung verfasst) und eine Prüfungsphase durchlaufen werden.

Neben der strukturell-organisatorischen Nachhaltigkeit müssen gemeinsam mit den Fachcommunities Nutzungsversprechen für den nachhaltigen Betrieb der TextGrid-Infrastruktur entwickelt werden, u.a. in Form von *Service Level Agreements*. Weitere wichtige Aspekte sind Nutzerzertifizierung, VO-Management, Security.

AP 3.2 entwickelt auf Basis dieser Rechtsform Musterlizenzvereinbarungen mit unterschiedlichen Content-Providern, da Forschungsprimärdaten der TextGrid-Community in der Regel nicht immer frei von Rechten Dritter sind. Die sich daraus ergebenden Arten von Nutzungsbedingungen für den Endnutzer werden formuliert und in der AAI bzw. dem Management der Virtual Organizations technisch umgesetzt.

AP 4 - Coaching / Community Bildung

In Arbeitspaket 4 geht es nicht um technologische Aspekte bzw. technische Entwicklungen, sondern um die vom Gutachten deutlich angemahnten Maßnahmen zur Heranführung bzw. Einbindung der Communities an Grid-Technologien bzw. in TextGrid. Anders als im Projekt WissGrid, in dem generische Konzepte mit Hilfe von Blaupausen für die Einbindung neuer Communities entwickelt werden, stehen in AP 4 die spezifisch fachwissenschaftlichen Anwendungs- und Entwicklungsmöglichkeiten im Fokus. Für alle wissenschaftlichen Communities gilt nach den Erfahrungen aus den bisherigen D-Grid-Phasen, dass es gezielter Maßnahmen bedarf, um die Wissenschaften an die neuen Grid-Technologien heranzuführen. Aus diesem Grund wird in AP 4 ein intensiver wechselseitiger Dialog zwischen TextGrid-Partnern und textwissenschaftlichen Anwendern, Entwicklern und Content-Providern initiiert bzw. mit den bereits etablierten Kontakten aus der ersten Förderphase (weiter)gepflegt. Um nicht am Bedarf der reich ausdifferenzierten textwissenschaftlichen Communities vorbei zu entwickeln, werden verschiedene Maßnahmen ergriffen, um sie in den gesamten Entwicklungsprozess einzubinden und dadurch auch die allgemeine Akzeptanz zu erhöhen (AP 4.1). Im Rahmen gezielter Schulungen und Workshops, u.a. auch im Programm großer fachwissenschaftlicher Tagungen, z.B. der IEEE DEST 2009, der Jahrestagung der Arbeitsgemeinschaft germanistischer Editionen 2010 in Frankfurt a.M., der Language Resources and Evaluation Conference uam., werden Anwender, Entwickler und Content-Provider aus- und weitergebildet (AP 4.2). Der Aufbau eines TextGrid-Netzwerkes ist z.T. aus dem AP 7 des alten TextGrid II-Antrags hierher transferiert worden (AP 4.3).

In AP 4.1 soll eine community-spezifische Bedarfserhebung die Grundlagen schaffen für die gezielte Ansprache unterschiedlicher textwissenschaftlicher Communities. Obwohl es durchaus eine als "geisteswissenschaftlich" zu bezeichnende Wissenschafts- und Forschungskultur gibt, lassen sich im Einzelnen doch auch große Unterschiede konstatieren.

Sog. "große Fächer" haben nicht nur andere Kommunikations- und Ausbildungsstrukturen als sog. "Kleine Fächer", auch das Verhältnis, die "Nähe" oder "Ferne" zum Technikeinsatz ist unterschiedlich ausgeprägt. Die Ergebnisse und laufenden Arbeiten des BMBF-geförderten HRK-Projekts "Kleine Fächer" (Arbeitsstelle Universität Potsdam, Prof. Dr. Norbert Franz) bilden eine ausgezeichnete Grundlage, um die diversifizierten Communities identifizieren und ansprechen zu können.

Der jeweilige Bedarf wird mit unterschiedlichen Methoden erhoben, die gesamte Bedarfserhebung wird durch eine am Trierer Kompetenzzentrum angesiedelte Dissertation wissenschaftlich begleitet. Zum einen haben sich Nutzerszenarien, die von Fachwissenschaftlern entworfen wurden, als zentrale Konzeptionshilfe erwiesen. Neben der wertvollen Kooperation mit dem Infrastruktur-Projekt Bamboo², das eine zweijährige Bedarfsanalyse-Phase durchgeführt hat bzw. durchführt (s. auch unter AP 6.1), werden Verfahren weitergeführt, die im Laufe von TextGrid I, insbesondere in der Abschlussphase, entwickelt wurden. Am Beispiel des TextGrid-Fachwissenschaftler-Workshops lässt sich demonstrieren, wie die intensive Nutzereinbindung gestaltet werden kann: Zum einen wurden alle Aktivitäten der Anwender während des gesamten Workshops per Screencam aufgezeichnet; zum anderen wurde ein unmittelbares Feedback im Anschluss an den Workshop durch einen sehr kurzen Fragebogen eingeholt. Kernstück der Begleitung waren jedoch ausführliche Leitfadeninterviews, die mit allen Nutzern mit einem gewissen zeitlichen Abstand in einer face-to-face-Befragungssituation geführt wurden. Die Einschätzungen aus den Fragebögen sowie die Aktivitäten während der Schulung dienen als Korrektiv der reflektierteren Aussagen in den Interviews und geben so ein umfassendes Bild der spezifischen Nutzung. Bereits jetzt hat sich gezeigt, dass diese intensive Nachbereitung der Schulung die Bereitschaft deutlich erhöht, sich längerfristig mit den Möglichkeiten von TextGrid zu beschäftigen und das Interesse an der Weiterentwicklung wach zu halten.

Diese aufwändige Methode kann nicht bei allen Schulungsmaßnahmen flächendeckend eingesetzt werden, soll jedoch bei besonders viel versprechenden, ausgewählten Nutzergruppen fortgesetzt werden; neben weniger aufwändigen wie direktes Feedback durch kurze Fragebögen, Emailfeedback etc.

Nachdem die Software mit TextGrid I einen gewissen Entwicklungsstatus erreicht hat, kann der iterative Prozess von Einholung des Nutzerfeedbacks, Weiterentwicklung der Software, erneute Schulung und Information, Feedback etc. intensiviert werden.

Neben diesen sehr gezielten und gelenkten Verfahren der Nutzereinbindung bleibt das TextGrid-Konsortium wie auch in der ersten Phase prinzipiell offen für Kontakte, Entwicklungspartnerschaften und Nutzerszenarien aus den Communities, die im vorgegebenen Zielrahmen sinnvoll erscheinen. Ein Modell aus der ersten Phase wäre bspw. die Kooperation mit Interedition³.

AP 4.2 konzipiert auf der Basis von Projektdokumentation und auf den Bedarf der jeweiligen Communities zugeschnittenen Tutorials (auch als eTutorials) Schulungsmaßnahmen für unterschiedliche Adressatengruppen und führt diese durch. Die Schulungsmaßnahmen sind

² <http://www.projectbamboo.org>

³ <http://www.interedition.eu>

differenziert nach Nutzer/Anwender, Entwickler und Content-Provider. An den beteiligten Universitäten Göttingen, Trier und Würzburg sind derzeit ferner Bestrebungen im Gange, die eHumanities in den universitären Curricula zu etablieren, dabei finden auch die für TextGrid relevanten Aspekte Berücksichtigung.

Das durch die Maßnahmen in den APs 4.1 und 4.2 aufgebaute Community-Netzwerk muss systematisch weitergepflegt und ausgebaut werden. Neben den Anwendern und Entwicklern müssen Kontakte zu Entscheidungsträgern, (Fach-)Verbänden oder Fachgesellschaften hergestellt und institutionalisiert werden. Hier werden sich insbesondere die TextGrid-Partner engagieren, die als Institutionen die Nachhaltigkeit der Forschungsinfrastruktur gewährleisten können. IDS, SUB und Kompetenzzentrum sind auch aufgrund der Vielzahl der von ihnen betreuten Projekte aus ganz unterschiedlichen Fächern und mit unterschiedlichen Kooperationspartnern (wie Max Planck Institute, Wissenschaftsakademien, Forschungsinstitutionen, Verlagen u.a.m.) gut aufgestellt, um die notwendigen Kontakte zu knüpfen und zu pflegen.

AP 5 - Technische, semantische und organisatorische Interoperabilität

Die Arbeitsinhalte, die hier beschrieben waren, werden auf Grund der jetzigen, modifizierten Prioritätenstruktur zunächst zurückgestellt.

AP 6 - Fachwissenschaftliche Nachhaltigkeit

Das Hauptziel von AP 6 ist die fachwissenschaftliche Nachhaltigkeit. Erreicht werden soll dieses Ziel durch

- die in TextGrid I und II involvierten Fachwissenschaftler, die ihre Anforderungen an die in TextGrid verwendeten Werkzeuge aufgrund ihrer eingehenden Kenntnis typischer Arbeitsabläufe und auch aufgrund der Bedürfnisse ihrer laufenden Forschungsprojekte formulieren.
- die Weiterentwicklung der vorhandenen Werkzeuge zur Produktionsreife (siehe die Tabelle zum AP 6.1 im Anhang I)
- Generalisierung der vorhandenen Werkzeuge, um sie an die Anforderungen in den eng verwandten Arbeitskontexten anderer Textwissenschaften anzupassen und vielseitig verwendbar zu machen
- Neuentwicklung von Werkzeugen, die eine breite Nutzerschaft innerhalb der Textwissenschaften versprechen

Um sicherzustellen, dass die Wahl der Tools und ihre Leistungsmerkmale für eine möglichst große Anzahl von Projekten in den Textwissenschaften von Nutzen sind, haben wir eine Reihe von Maßnahmen getroffen:

- Die Wahl der Tools und die Beschreibung ihrer Leistungsmerkmale basiert auf Nutzerszenarien, die von Fachwissenschaftlern entworfen und über eine weite Streuung

- Die einbezogenen Partner verbinden Entwicklungskompetenz mit einer Position in ihrem jeweiligen Arbeitsgebiet, die es ihnen ermöglicht, die Anforderungen und Bedürfnisse ihrer jeweiligen spezifischen Communities sehr genau einzuschätzen. Möglich ist dies dadurch, dass sie häufig nicht nur ein einzelwissenschaftliches Projekt vertreten, sondern eine umfassendere Organisation.

So ist etwa

- das Kompetenzzentrum Trier auch Kooperationspartner für eine sehr große Reihe von geisteswissenschaftlichen digitalen Editionsprojekten.
- das Münchener Zentrum für Editionswissenschaften ein Zusammenschluss einer ganzen Reihe von Münchner editionswissenschaftlichen Projekten mit z.T. erheblichem Umfang.
- die Universität Paderborn, die sich im Forum Digitale Musikedition, Akademie der Wissenschaften Mainz, engagiert, und eine ganze Reihe musikwissenschaftlicher Editionen zusammenbringt.
- Eine wichtige Quelle für Anforderungen an TextGrid sind auch die laufenden Projekte der TextGrid-Partner:
 - Das Faust-Projekt (Universität Würzburg)
 - Die ca. 40 derzeit betreuten Projekte des Kompetenzzentrums Trier, z.B. das Heinrich-Heine-Portal oder das Cusanus-Portal, ferner die geplanten Projekte zu Friedrich Hölderlin, Christian Dietrich Grabbe, Arthur Schnitzler oder Ludwig Börne, um einige Editionsprojekte herauszugreifen
 - Das Münchener Zentrum für Editionswissenschaft MüZe bündelt die Aktivitäten von sechs Institutionen, zu deren Aufgabenbereich die geisteswissenschaftliche Grundlagenforschung in Form von kritischen Editionen gehört. Sie erweitern das Fächerspektrum insbesondere um die Geschichtswissenschaften und die Klassische Philologie und damit auch die entsprechenden editionsphilologischen Anforderungen. Im Bereich der Glossen-Edition ergeben sich ferner Anknüpfungspunkte zu einer germanistischen Glossenforschergruppe, so dass mit einer raschen Verbreiterung der Community zu rechnen ist.⁴

Eine ergänzende Aufstellung aller Partner und ihrer Kompetenzen und Projekte, in denen sich Anforderungen für TextGrid ergeben, findet sich in Kapitel III.

⁴ Zum germanistischen Glossenprojekt vgl. z.B. http://www.hkfz.uni-trier.de/?site_id=108&proj_id=7ae43f90473a6f90bd9729ba9475d217&sitename=Startseite.

- Auswertung von fachwissenschaftlichen Workshops im Rahmen einer Doktorarbeit, um die Probleme bei der Verwendung von TextGrid empirisch zu ermitteln.
- Internationale Gespräche über die Weiterentwicklung philologischer Infrastrukturen und Software. Neben dem bereits erwähnten Bamboo sind hier auch CLARIN⁵, DARIAH⁶ und Interedition⁷ zu nennen.

In AP 6 haben sich gegenüber dem Hauptantrag nur kleinere Änderungen ergeben:

- Im Sinne der stärkeren Blockbildung, um den Kompetenzaufbau in TextGrid zu fördern, sind die vergebenen Personenmonate auf eine kleinere Anzahl von Partnern verteilt worden.
- Durch das Ausscheiden von Saphor wurde es notwendig, einen Teil der so entfallenen Personenmonate neu zu vergeben.
- Aufgrund von wiederholten Benutzeranfragen, ob der technisch verwandte und in den computerphilologischen Communities verbreitete kommerzielle Editor OXYgen⁸ in TextGrid als Alternative zum vorhandenen Editor verwendet werden könnte, wurde in 6.1 nun die Prüfung einer entsprechenden Kooperation geplant, die es OXYgen-Lizenznehmern ermöglichen würde, ihren gewohnten Editor im TextGrid-Umfeld zu verwenden.

AP 7 - eHumanities-Netzwerk

Dieses AP ist in AP 4 und AP 8 aufgegangen.

AP 8 - Nationale/internationale Liaison, Projektmanagement, Öffentlichkeitsarbeit

TextGrid hat in der ersten Förderphase bereits wichtige Ergebnisse im Bereich der Community-Bildung erzielen können. Das Konzept zur Öffentlichkeitsarbeit wird in der zweiten Förderphase diesen Bereich unterstützen und damit intensivieren. Ein weiterer Arbeitsschwerpunkt liegt nun, da TextGrid nach einer Entwicklungsphase bereits erste Ergebnisse vorweisen kann, auch in einer allgemeinen, populärwissenschaftlichen Bekanntmachung dieser geisteswissenschaftlichen Verbundforschung, die auf innovativen Technologien wie dem Grid basiert (z.B. FAZ, Die Zeit, Spiegel, Süddeutsche, Spektrum der Wissenschaft, Goethe-Institut Portal).

Ein zusätzlicher Aspekt dieses APs ist auch die enge Vernetzung mit anderen Forschungsinfrastrukturinitiativen und -projekten im nationalen (z.B. eAqua, D-Spin) und internationalen Kontexten (DARIAH, CLARIN, Bamboo). Ziel ist dabei die Unterstützung des Aufbaus bzw. der Weiterentwicklung einer übergreifenden eHumanities-Infrastruktur.

⁵ <http://www.clarin.eu>

⁶ <http://www.dariah.eu>

⁷ <http://www.interedition.eu>

⁸ <http://www.oxygenxml.com>

3. TextGrid Konsortium und Kompetenzen der Partner

DAASI International GmbH, Tübingen (DAASI)

Arbeitsschwerpunkt in TextGrid:

Weiterentwicklung von TextGridRep und dessen Integration mit Fedora (AP 1) und technische Abbildung von Lizenzvereinbarungen und Nutzungsbedingungen (AP 3) sowie Coaching und Schulungsmaßnahmen (AP 4).

Kompetenz:

Aufbau und Betrieb von IT-Infrastrukturen u.a. in den Bereichen Grid und IdentityManagement; Verteilte Authentifizierung und Autorisierung auf Grundlage von Shibboleth, PKI, Abbildung von Zugriffsrechten mit RBAC und XACML, Erstellung von Geschäftsplänen, Durchführung von Technik-Workshops und Schulungen.

Laufende Projekte (mit TextGrid-Bezug):

SLC-GAP Integration von Short lived Credentials in D-Grid-Infrastrukturen, insb. in TextGrid-Architektur.

Fachhochschule Worms (FHW) – University of Applied Sciences Worms

Arbeitsschwerpunkt in TextGrid:

Toolentwicklung, Infrastruktur / TextGridLab, Betreuung und Nachhaltigkeit der Entwicklercommunity

Kompetenz:

Theorie und Praxis verteilter Architekturen mit Anwendungen in den eHumanities und im eGovernment, Entwicklung komplexer Softwaresysteme, Verwaltung und Syndikation großer verteilter Registries, Entwicklung von Datenmodellen.

Laufende Projekte (mit TextGrid-Bezug):

CEN/ISSS eGov Share: Spezifikationen für Metadatensyndikation in heterogenen Umgebungen; Isidorus: Open Source Topic Map Engine, gleichzeitig eine der Referenzimplementierungen für die eGov Share Spezifikationen; Ausrichtung von eHumanities Session und Tutorials in der IEEE DEST; Regelmäßige studentische Praxisprojekte im Bereich verteilter Anwendungen.

Institut für Deutsche Sprache Mannheim (IDS)

Arbeitsschwerpunkt in TextGrid:

Organisatorische Nachhaltigkeit (AP 3), Fachsäule Linguistik (AP 6.3)

Kompetenz:

Dauerhafte Bereitstellung von Forschungsinfrastrukturen, -daten und Diensten

Laufende Projekte (mit TextGrid-Bezug):

Aufbau eines Zentrums „Digitale Forschungsressourcen für die germanistische Sprachwissenschaft“ (BMBF); Deutsche Sprachressourcen-Infrastruktur D-SPIN (BMBF); WissGrid, Wechselwirkungen zwischen linguistischen und bioinformatischen Verfahren, Methoden und Algorithmen: Modellierung und Abbildung von Varianz in Sprache und Genomen.

***Ludwig-Maximilians-Universität München (LMU), Münchener Zentrum für Editions-
wissenschaft*****Arbeitsschwerpunkt in TextGrid:**

Weiterentwicklung der TEI-Codierung; fachspezifische Tool-Entwicklung (AP 6).

Kompetenz:

Editionswissenschaft, insbesondere digitale Editionen kommentierender Texte im Verhältnis zu den Referenztexten; Textauszeichnung, vernetzte Textdatenbanken

Laufende Projekte (mit TextGrid-Bezug):

„Editing glosses/Glossenedition“ (LMUexcellent); textkorpusbezogene Edition (Persius-Scholien) und textgenetische Edition (Martianus Capella)

***Max Planck Digital Library (MPDL), vertritt dabei auch das Max Planck Institute for
Psycholinguistics (Nijmegen, Holland) und das Max-Planck-Institut für Kunst-
geschichte (Florenz, Italien)*****Arbeitsschwerpunkt in TextGrid:**

Konzepte für die Integration mit eSciDoc (AP 1) und Koordinierung der Entwicklungen im Tool Bereich für die beteiligten Max-Planck-Institute (AP 6).

Kompetenz:

Entwicklung von Forschungsinfrastrukturen und fachspezifische Tool-Entwicklung.

Laufende Projekte (mit TextGrid-Bezug):

eSciDoc als eResearch-Infrastruktur, fachspezifische Tool-Entwicklung (ViRR für MPI for European History of Law, Faces für MPI for Human Development)

***Georg-August Universität Göttingen: Niedersächsische Staats- und Universitäts-
bibliothek Göttingen (SUB)*****Arbeitsschwerpunkt in TextGrid:**

Entwicklung von digitalen Forschungsinfrastrukturen (TextGrid Repository, AP 1) und Konsortialleitung (AP 8).

Kompetenz:

Nachhaltige Organisation mit Schwerpunkt Langzeitarchivierung analoger und digitaler (fach)wissenschaftlich relevanter Objekte, dauerhafte Bereitstellung von Forschungsinfrastrukturen und Diensten, Erfahrung in der Konsortialführung großer Verbundprojekte.

Laufende Projekte (mit möglichem TextGrid-Bezug):

Entwicklung von Forschungsdatenarchiven (WissGrid, BMBF), Entwicklung von geisteswissenschaftlichen Forschungsinfrastrukturen (DARIAH, EU), Bereitstellung von Publikations-Infrastrukturen (Driver, EU)

Universität Trier: Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften an der Universität Trier (Trier)**Arbeitsschwerpunkt in TextGrid:**

Tool-Entwicklung (AP 2 und AP 6); Coaching und Schulungsmaßnahmen (AP 4)

Kompetenz:

Digitalisierung geisteswissenschaftlicher Grundlagenwerke (insbesondere Wörterbücher und Editionen); Entwicklung verteilter Arbeitsumgebungen (Tares - Artikelredaktionssystem für das Mittelhochdeutsche Wörterbuch, Forschungsnetzwerk und Datenbanksystem FuD, INF-Projekt für den SFB 600; Europäische Geschichte Online - EGO), dauerhafte Bereitstellung von Forschungsinfrastrukturen und Diensten.

Laufende Projekte (mit TextGrid-Bezug):

WissGrid; Wechselwirkungen zwischen linguistischen und bioinformatischen Verfahren, Methoden und Algorithmen: Modellierung und Abbildung von Varianz in Sprache und Genomen

Technische Universität Kaiserslautern (TUKL)**Arbeitsschwerpunkt in TextGrid:**

Entwicklung eines OCR Service für Frakturerkennung und Integration in TextGrid; Bereitstellung von Werkzeugen zum Arbeiten mit OCR Ergebnissen. (AP 6.8)

Kompetenz:

Bildverstehen und Mustererkennung; Dokumentanalyse; Layoutanalyse; statistische Sprachmodelle.

Laufende Projekte (mit TextGrid-Bezug):

OCROpus Open Source OCR System (Google Projekt; www.ocropus.org); Decapod Book Capture Project (Mellon Foundation Projekt; www.decapod-project.org).

Julius-Maximilians-Universität Würzburg, Institut für Deutsche Philologie (Würzburg)

Arbeitsschwerpunkt in TextGrid:

Koordination der Tool-Entwicklung (AP 6); Werkzeug-Entwicklung für Philologen.

Kompetenz:

Digitale Editionen, Textauszeichnung, Software für Philologen

Laufende Projekte (mit TextGrid-Bezug):

Faust-Edition (DFG), deren Software-Anforderungen in die Entwicklung der TextGrid-Werkzeuge als Nutzeranforderung eingehen. Aufbau eines Zentrums für digitale Editionen und Kooperation mit Würzburger Projekten, z.B. dem Langzeitvorhaben Jean-Paul-Edition. Wechselwirkungen zwischen linguistischen und bioinformatischen Verfahren, Methoden und Algorithmen: Modellierung und Abbildung von Varianz in Sprache und Genomen.

Universität Paderborn/Detmold (UPB)

Arbeitsschwerpunkt in TextGrid:

Entwicklung und Integration fachspezifischer Lösungen für Musikeditionen (AP 6); Coaching und Schulungsmaßnahmen (AP 4)

Kompetenz:

Konzeption und Erprobung von Modellen digitaler Musikeditionen; Betreuung externer Editionsprojekte, darunter: OPERA-Projekt Bayreuth, Reger-Werkausgabe Karlsruhe, Weber-Gesamtausgabe Berlin/Detmold; Mitwirkung in den TEI SIGs Correspondence und Music. Aufbau projektspezifischer Datenbanken, u.a. zur Vorbereitung von Brief- und Tagebucheditionen.

Laufende Projekte (mit TextGrid-Bezug):

DFG-Projekt Entwicklung von Werkzeugen für digitale Formen wissenschaftlich-kritischer Musikeditionen; NEH/DFG-Bilateral Digital Humanities Program: Digital Music Notation Data Model and Prototype Delivery System.

4. Optimierung der Zusammenarbeit im TextGrid Konsortium

Auf dem letzten internen TextGrid-Strategietreffen Ende Februar 2009 in Trier haben die Projektpartner eine interne Evaluierung der Arbeitsprozesse und Kommunikation durchgeführt, die zur Optimierung der Zusammenarbeit in dem größer werdenden Verbundprojekt beitragen soll. Damit wird auch ein wesentlicher Beitrag zum "Risk-Management" geleistet.

Folgende Punkte konnten identifiziert werden:

1. Intensivierung der Kommunikation zwischen Fachwissenschaftlern und IT-Experten / Entwicklern

Problem: Die IT-Experten/Entwickler mussten oftmals zu lange auf den Input von Fachwissenschaftlern warten, da die Fachwissenschaftler (FaWi) den (organisatorischen) Aufwand unterschätzt hatten. Darüber hinaus handelt es sich bei den fachwissenschaftlichen Arbeiten auch vielfach um für sie neue Arbeitsweisen – das Produkt (z.B. das Szenariopapier) ist oftmals nur ein (zu überwindender) Zwischenschritt, der für die technische Entwicklung erforderlich ist, mündet aber nicht in einer wissenschaftlichen Publikation.

Lösungsstrategien:

- Zusammenarbeit zwischen FaWi und Entwicklern verbessern, Entwickler müssen klare (auch zeitliche) Vorgaben definieren
- Rapid prototyping – damit nicht am Bedarf vorbeientwickelt wird, FaWi rechtzeitig einen Eindruck vom Produkt gewinnen und entsprechend Feedback geben können
- Der iterative Prozess (Prototyp - Testen/Rückmeldung - Einarbeitung) muss von Beginn an erfolgen, auch wenn der Prototyp an sich selber noch nicht aussagekräftig ist
- Regelmäßige Treffen / Skype calls (inklusive der Mitglieder der TextGrid-Steuerungsgruppe)

2. Tool-Entwicklung ohne frühzeitige Einbindung aller relevanten Partner

Problem: Es werden Spezifika bei TextGridTools entwickelt, die nicht von Beginn an mit allen Partnern abgestimmt sind. Dadurch wird zu spät entdeckt, dass diese Spezifika am Ende zu einem z.T. erheblichen zeitlichen Mehraufwand bei der Integration in das TextGrid Laboratory führen.

Lösungsstrategien:

- Tool-Prototypen von Beginn an in die TextGrid Umgebung integrieren
- Relevante APIs (*application programming interface*) rechtzeitig mit allen Partnern und besonders den Entwicklern der jeweiligen Partner abstimmen

- Mittlerweile existiert eine funktionierende Infrastruktur, so dass die beiden vorgenannten Punkte leichter umzusetzen sind – insbesondere im Rahmen von TextGrid-Programmiersprints

3. Einheitliches Framework für Software-Tests

Problem: Bislang existiert kein einheitliches Konzept für Tests neu entwickelter bzw. überarbeiteter Software, da Services, GUI-Komponenten und Grid-Komponenten der Middleware diesbezüglich sehr unterschiedliche Anforderungen stellen.

Lösungsstrategien:

- (Unit-)Tests sollten im Rahmen des Builds automatisiert durchgeführt werden – hierbei bietet sich an, auf die entsprechenden Erfahrungen bei eSciDoc und der SUB zurückzugreifen

4. Neubewertung der grafischen Komponenten

Problem: Die GUI-Entwicklung wurde zunächst unterschätzt und mit zu wenig Ressourcen angegangen, so dass in der aktuellen Beta-Version noch nicht für jeden implementierten Service ein entsprechendes GUI-Element (Plugin) existiert.

Lösungsstrategien:

- die Ressourcenplanung wurde für TextGrid II entsprechend geändert
- Nutzerführung / Usability stärker gewichten: dies ist in TextGrid II deutlicher betont und entsprechende Ressourcen stehen zur Verfügung
- Prototypen sofort integrieren (s.o.)

5. Kompetenzschwerpunkte im Verbundprojekt festlegen (Schwerpunktbildung)

Problem: Verantwortlichkeiten für allgemeine Aufgaben (Betreuung und Pflege des bug tracking Systems, SVN, Wiki etc.) sind erst im Laufe des Projekts gewachsen. Dies hat "hier und da" zu Verzögerungen bei Entscheidungsprozessen geführt. Der zeitliche Aufwand für diese Schwerpunktsetzung war nicht eingeplant.

Lösungsstrategien:

- In TextGrid II werden, basierend auf den langjährigen Erfahrungen in TextGrid I, von Beginn an Kompetenzschwerpunkte bei den Partnern gebildet, die dann die Verantwortung haben, die entsprechenden Inhalte bzw. Entwicklungen zu koordinieren
- Von Projektbeginn an werden die Kompetenzen auf die beteiligten Partner festgelegt, Lücken dabei identifiziert und durch Absprachen geschlossen (z.B. bei den Kern-Metadaten)
- Durch eine Konzentration der Personen-Monate bei den entsprechenden Partnern stehen auch die Ressourcen dafür zur Verfügung
- Schaffung von Anreizsystemen bei diesen "Kümmerern" durch die Motivation, diese Ergebnisse in Form von Publikationen, Vorträgen vorzustellen

6. Anforderungen an die potentiellen MitarbeiterInnen in TextGrid sind sehr speziell und z.T. hoch

Problem: Optimale MitarbeiterInnen in TextGrid bringen sowohl geisteswissenschaftliche als auch ausgewiesene IT-Kompetenz mit. Darüber hinaus müssen sie sich idealerweise sprachlich und terminologisch in der Welt zwischen Geisteswissenschaft und IT bewegen und als Mittler auftreten, was hohe Kommunikationsfähigkeiten erfordert.

Lösungsstrategien:

- Gemeinsame Ausschreibungen und gemeinsam durchzuführende Bewerbungsgespräche
- Auf entsprechende Kommunikationsfähigkeiten und Erfahrungen in größeren Projekten bei potentiellen Bewerbern achten
- Entsprechende zeitliche Ressourcen einkalkulieren für das zwar eigenverantwortliche, aber auch kooperative Arbeiten

7. Allgemeine Anregungen

- „Neue Ehrlichkeit“ was Zeitplanung / Einhaltung von Deadlines angeht (insbesondere bzgl. der Toolentwicklung), auf (eventuelle) Probleme rechtzeitig hinweisen und nicht aus "falscher Scham" verschweigen
- realistischen Aufwand für übergreifende und allgemeine Konzeption sowie Entscheidungsfindung einkalkulieren; Entscheidungswege verkürzen
- die Arbeitsgruppen waren in dieser Form nicht geplant, haben sich aber als äußerst nützlich erwiesen und werden in TextGrid II entsprechend eingeplant. Dabei sollte darauf geachtet werden, dass diese AGs auch nur temporär besetzt sein können, um damit auf aktuelle Anforderungen flexibel zu reagieren
- Programmiersprints haben sich als sehr fruchtbar erwiesen und sollten regelmäßig stattfinden, die Gruppengröße ist dabei flexibel und wird dem jeweiligen Thema angepasst
- es wird jeweils ein Haupt-Koordinator in Persona benannt für TextGridLab und TextGridRep
- besonderes Augenmerk wird auf die interne Schulung/Ausbildung bei den neu hinzukommenden Entwicklern gelegt – damit ließe sich das bisherige Konzept der Entwicklerworkshops auch intern testen
- für die eSciDoc-Koordination muss seitens der MPDL ein (technischer) Ansprechpartner benannt werden (ist bereits geschehen)
- Verantwortlichkeiten für "Kärnerarbeiten" festlegen, gerecht verteilen, Asymmetrie der Beteiligung vermeiden

Das Konsortium wird regelmäßig überprüfen, dass die Lösungsstrategien umgesetzt sowie für neu auftauchende Probleme entsprechende Strategien entwickelt werden.

Anhänge

I Arbeitspakete

AP 1 - TextGrid Repository

Nummer des Arbeitspaketes:	AP 1
Titel:	TextGrid Repository
AP-Leiter:	SUB
Start-Monat:	Juni 2009 (M1) bis Mai 2012 (M36)
Gesamtanzahl PM:	71

Personenmonate pro Partner gesamt und pro Unterarbeitspaket (UAP)⁹

Partner (Akronym)	DAASI	FHW	IDS	LMU	MPDL	SUB	Trier	TUKL	Würz- burg	UPB	Gesamt
Gesamt:	18		7		4	9 (+27) 10	6				44 (+27)
AP 1 Rep											
AP 1.1: Entwicklung eines generi- schen For- schungsda- tenreposito- riums: Integ- ration mit WissGrid ¹¹											0
AP 1.2: TextGrid spezifische Anpassun- gen	18		6			8* (+27)	6				65
AP 1.3:			1		4*	1					6

⁹ Die Verantwortlichen für das jeweilige Unterarbeitspaket sind mit einem * gekennzeichnet.

¹⁰ Bei diesen PM handelt es sich um 27 PM TV-L E 11.

¹¹ Für TextGrid ergeben sich hier null PM, da über WissGrid entsprechende PM finanziert werden.

Schnittstellen und Sicherheitskonzepten), organisatorische Einbindung in D-Grid (z.B. Definition von D-Grid Service Level Agreements) sowie offene Schnittstellen und Konzepte zur Interoperabilität mit anderen D-Grid-Initiativen.

Für die Entwicklung eines generischen D-Grid-Forschungsdatenrepositoriums sind in WissGrid für alle CommunityGrids insgesamt 168 PM vorgesehen. Da drei TextGrid-Partner an dem WissGrid-Konsortium maßgeblich beteiligt sind, werden hierfür keine PM im vorliegenden Antrag vorgesehen.

Repositorien-Systeme gehen über bisherige Lösungen zur langfristigen Archivierung hinaus. Sie sind gezielt für Archivierungszwecke geschaffen worden ("trusted digital repositories", [RLG/OCLC]) und sichern die langfristige Datenhaltung von digitalen Objekten wie Publikationen, Texten, Bildern, Simulationen, multimedialen Inhalten u.a. In der internationalen Grid-Community gibt es derzeit intensive Diskussionen [OGF23, IEEE08] über die Integration von Repository-Technologien, vor allem wegen (a) ihrer Funktionalitäten zur Verwaltung dynamischer Dateien in heterogenen Formaten, (b) ihrer Mechanismen zur Langzeitarchivierung und (c) zur semantischen Verknüpfung von Objekten, Objektteilen und Objektkontext.

Für die Verbindung zweier so fortgeschrittener Technologien wie Grid und Repositorien gibt es unterschiedliche Szenarien:

1. die Nutzung von Repositorien in einer eigenen, separaten Installation (z.B. lizenz- bzw. urheberrechtliche Einschränkungen oder gesetzliche Vorschriften für Krankengeschichten und personenbezogene Daten, wie sie in MediGRID vorliegen, können dies nahe legen);
2. physisch getrennte Repositorien, deren Inhalte untereinander verknüpft sind (auf diese Weise "förderierte Repositorien" bieten neue Möglichkeiten zum wissenschaftlichen Austausch zwischen Communities); oder
3. eine gemeinsame Repositoryinstallation, in der Communities den Wartungsaufwand teilen und durch gemeinsame Zugänge und interoperable Datenmodelle den Nutzen weiter erhöhen (dies bringt potenziell auch Vorteile für Communities aus sehr unterschiedlichen Bereichen, wie dies vom Projekt eSciDoc demonstriert wird [eSciDoc]).

Für TextGrid bietet sich vor allem das zweite Szenario einer Repositorien-Föderation an, sowohl im Hinblick auf die in AP 1.2 zu leistenden TextGrid-spezifischen Anpassungen als auch mit dem Ziel der Interoperabilität nicht nur in D-Grid sondern auch mit anderen geisteswissenschaftlichen Datenquellen. Auch dies sind Anforderungen und Standpunkte, die im Rahmen von WissGrid noch detailliert diskutiert und getestet werden.

Obwohl das parallel zu TextGrid laufende WissGrid-Projekt erst für den 24. Monat eine Umsetzung des generischen Forschungsdatenrepositoriums geplant hat, wird bereits im 9. Monat eine erste Architekturskizze öffentlich diskutiert. Diese offene Diskussion sowie die ständige Begleitung der Arbeiten durch die drei TextGrid-Projektpartner in WissGrid (Trier, IDS, SUB) ermöglichen eine frühe Planung der Installation und Migration des TextGrid-Forschungsdatenrepositoriums. Im Prinzip können die Arbeiten an der Anpassung des

Repositoriums für TextGrid (AP 1.2) sofort mit der Festlegung auf Schnittstellen beginnen. Es ist realistisch – und die TextGrid-Partner in WissGrid werden darauf hinarbeiten –, dass diese Spezifikation noch im ersten Projektjahr abgeschlossen ist.

AP 1.2: TextGrid-spezifische Anpassungen

Die enge Mitarbeit von TextGrid Partnern in WissGrid (AP 1.1) stellt sicher, dass das WissGrid-Forschungsdatenrepositorium in seiner grundlegenden Architektur auch für TextGrid verwendet werden kann. In diesem Arbeitspaket werden die Einbindung des Repositoriums in die TextGrid-Infrastruktur und die nötigen fachspezifischen Anpassungen vorgenommen. Dazu zählt vor allem die Migration der TextGrid-Utilities – m.a.W. der TextGrid Schnittstellen TG-crud, TG-search, und TG-auth*, auf die alle TextGrid-Dienste und -Tools zugreifen – wie auch die Einbettung der Adaptern zur Analyse von TEI-codierten Eingabedaten, die Strukturdatensuche und spezifisch textwissenschaftliche Funktionalitäten. Neben den Arbeiten zur Integration des Repositoriums wird es zum einen mit neuen Funktionalitäten erweitert und zum anderen seine Skalierbarkeit vergrößert.

Obwohl die Ergebnisse aus AP 1.1 und WissGrid erst im Laufe des Projektes in dieses AP 1.2 rückgeführt werden können (erwarteter Termin nach aktueller Planung WissGrid: Monat 12), müssen von Beginn an TextGrid-spezifische Arbeiten am Repositorium vorgenommen werden. Eine möglichst lückenlose Überführung der existierenden in die auf dem WissGrid Repositorium aufbauende Infrastruktur muss garantiert werden. Darüber hinaus sind einige Arbeiten in der Infrastruktur – wie z.B. die Verbesserung der Skalierbarkeit der TextGrid Suche – weitgehend unabhängig von dem zukünftigen Forschungsdatenrepositorium. Daher werden diese Arbeiten von Beginn an pro-aktiv durchgeführt.

Zur Adaption des generischen Forschungsdatenrepositoriums für TextGrid müssen in erster Linie die TextGrid-Utilities TG-crud, TG-search und TG-auth* [TextGrid Manual] mit dem Fedora-Repository aus WissGrid verknüpft werden. Die Schnittstellen basieren auf REST/SOAP, wodurch Offenheit und Nutzerfreundlichkeit gewährleistet sind. Die grundlegenden Konzepte, die bei der Entwicklung der TextGrid-Utilities angewendet wurden, sind stark an die Konzepte von Repositorien-Technologien angelehnt (z.B. Verknüpfung Daten und Metadaten, an REST orientierte Methoden für Create, Read, Update, Delete). Die Verknüpfung der jeweiligen Mechanismen für Autorisierung (XACML in Fedora vs. RBAC in TextGrid vs. GSI in D-Grid) und Authentifizierung (Grid-Shib, Projekt Short-Lived-Credentials) wird daher die größte Hürde für die Integration sein, deren prototypische Umsetzung aber vom Projekt WissGrid erwartet wird.

Für die Umsetzung der TextGrid Strukturdatensuche in einer integrierten Grid/Repository-Infrastruktur, bei der dem Nutzer direkt eine Maske zur Suche in XQuery angeboten wird, können die Konzepte des Oxford Research Archive [ORA] hilfreich sein. ORA entwickelt eine Fedora-basierte XQuery-Suche für das Oxford Text Archive [OTA], mit einer der TextGrid Strukturdatensuche ähnlichen Architektur. Dabei werden Daten – ähnlich den TextGrid Adaptern – beim Ingest nicht nur ins Archiv, sondern auch in eine XML-Datenbank zur Indizierung weitergeleitet. Die Erweiterung der Skalierbarkeit der TextGrid-Strukturdatensuche stellt ein weiteres dringendes Desiderat dar. Im Rahmen dieses Unterarbeitspaketes werden daher die Skalierbarkeit der aktuellen eXist-Datenbank getestet,

und Mechanismen zur Verteilung und zum Aufbau von Clustern von XML-Datenbanken erprobt und umgesetzt.

Eine wesentliche Erweiterung, ermöglicht durch semantische Repositorien-Technologien, ist die Erstellung eines erweiterten TextGrid-Projekt- und Sammlungskonzeptes. Während Daten in TextGrid-Projekten bisher nur über ihre Metadaten strukturierbar sind, sollen sie in Zukunft auch innerhalb eines Projektes hierarchisch untergliedert werden können. Vor allem bei größeren fachwissenschaftlichen Projekten hilft eine solche Untergliederung einzelner Module (z.B. einzelne Bänder, Abschnitte innerhalb eines Bandes etc.). Die Struktur kann in verschiedenen Tools zur Verbesserung der Darstellung und Erhöhung der Nutzerfreundlichkeit mit übernommen werden, Beispiele hierfür sind der Navigator und der Web-Publisher.

Anders als die Strukturierung innerhalb von Projekten ist das Sammlungskonzept projektübergreifend angelegt. Sammlungen ermöglichen die Verknüpfung von einzelnen Datensätzen, wobei die Datensätzen aus jedem beliebigen Projekt kommen können (solange der Nutzer Leserechte hat und ein Persistent Identifier existiert). Ein typischer Anwendungsfall wäre eine wissenschaftliche Analyse auf einem ausgewählten Korpus. Bisher ist es zwar in TG-search möglich, Suchen bzw. XQuery-basierte Analysen auf bestimmte Metadatenelemente einzugrenzen (z.B. finde alle Vorkommnisse von xy in Werken von „Goethe“ bzw. Werken aus dem 18. Jh.). Durch das Sammlungskonzept werden Nutzer in Zukunft ihren Korpus präzise selektieren und auf dieser Sammlung arbeiten können (z.B. finde alle Vorkommnisse von „xy“ in Sammlung „yz“). Neben der besseren Verwaltbarkeit erhöht dies die Wissenschaftlichkeit und Nachvollziehbarkeit der Ergebnisse in TextGrid. Ohne ein Sammlungskonzept ist es unmöglich, die genauen Datensätze zu dokumentieren, auf denen ein Wissenschaftler eine Analyse durchgeführt hat – ein anderer Wissenschaftler könnte in der Zwischenzeit z.B. ein neues Werk aus dem 18. Jh. publiziert und damit die Menge aller Werke aus dem 18. Jh. erweitert haben.

Die Implementierung technischer Schnittstellen zu anderen Repositorien stellt einen weiteren wesentlichen Aspekt des Arbeitsprogramms dar. Repositorien-Technologien bringen zumeist bereits diverse Standard-Schnittstellen mit, darunter OAI-PMH (Protocol for Metadata Harvesting¹²), z39.50/zing bzw. SRU/SRW¹³ oder das neue Protokoll zur Repository-Föderation OAI-ORE (Object Reuse and Exchange¹⁴). Neben diesen Repository-spezifischen Standards werden ggf. in WissGrid für D-Grid-Communities neue Schnittstellen entwickelt. Für eine mögliche Verknüpfung von TextGrid mit weiteren geisteswissenschaftlichen Repositorien wie Perseus [Perseus] oder der dMGH ist evt. eine Anpassung dieser Schnittstellen bzw. Erweiterung mit neuen Schnittstellen nötig.

AP 1.3: Kooperations-Szenarien mit eSciDoc

Das BMBF-Projekt eSciDoc [eSciDoc] hat in den letzten 5 Jahren eine produktive digitale Forschungsinfrastruktur für die Max-Planck-Gesellschaft aufgebaut. In dieser Infrastruktur

¹² <http://www.openarchives.org/pmh/>

¹³ <http://www.loc.gov/standards/sru/>

¹⁴ <http://www.openarchives.org/ore/>

werden auch die Daten aus den geisteswissenschaftlichen Max-Planck-Instituten archiviert und virtuellen Forschungsumgebungen zugänglich gemacht. TextGrid hat schon früh mit eSciDoc und auch einigen Max-Planck-Instituten Kontakt aufgenommen, da Daten und Funktionalitäten wechselseitig von Interesse sind. Darüber hinaus basiert der technische Kern von eSciDoc auf Repository-Technologien.

Eine Zusammenarbeit mit eSciDoc ist daher auf verschiedenen Ebenen viel versprechend:

- **technisch, Funktionalitäten:** Viele Dienste oder Tools können kollaborativ entwickelt werden. So hat es z.B. mit dem Kunsthistorischen Institut in Florenz (MPI) bereits intensive wissenschaftliche Diskussionen gegeben. Dennoch werden derzeit noch Dienste wie Named Entity Recognition, Verknüpfung mit Normdaten und anderen parallel und mit unterschiedlichen Schnittstellen entwickelt. Durch Interoperabilität der Dienste und einen gemeinsamen Dienstekatalog sind hier wesentliche Synergien möglich.
- **semantisch, Daten:** Die Daten vieler MPIs sind von großem Interesse für TextGrid Nutzer und umgekehrt. Schnittstellen und Adaptern zwischen den beiden Systemen könnten in einem ersten Schritt eine übergreifende Recherche in den Daten beider Repositorien ermöglichen. Weitergehend würde durch die Diskussion und Entwicklung gemeinsamer Daten- (Sammlungen/Kollektionen) und Objektmodelle (Files, Metadaten) die Wiederverwendbarkeit der Forschungsdaten über Systemgrenzen hinweg gefördert werden.
- **organisatorisch:** Entwicklung eines gemeinsamen Zukunftsszenarios, um die Aufgabenverteilung in einer nationalen Forschungsinfrastruktur festzulegen.

Im Rahmen dieses Unterarbeitspaketes werden diese unterschiedlichen Ebenen evaluiert und mögliche Kooperationswege definiert. Die bereits im Jahr 2007 getroffene Kooperationsvereinbarung zwischen der MPDL und der SUB könnte den politischen Rahmen für diese Ziele schaffen. (<http://tinyurl.com/dcrnqp>).

Reports (R) & Milestones (M)

Bezeichnung	Monat	Beschreibung
M 1.1.1	9	WissGrid-Workshop mit Vorstellung der generischen Architektur des WissGrid Forschungsdatenrepositoriums
M 1.1.2	36	Integration TextGrid / Forschungsdatenrepositorium WissGrid
R 1.2.1	16	Roadmap zur Integration Grid/Repository; mit Prototypen für verteilte Suche und das Projekt-/Sammlungskonzept
R 1.2.2	36	TextGridRep: Manual zum TextGrid-Forschungsdatenarchiv (Architektur, Installation, Erweiterbarkeit)
R 1.3.1	14	Aufbau einer nationalen föderierten Repositorien-Infrastruktur für die Geisteswissenschaften (Visionspapier)
R 1.3.2	28	Wege zur Verknüpfung eSciDoc / TextGrid (erstes Konzept)

Literatur

- [allianz] Wissenschaftsorganisationen starten Schwerpunktinitiative zur "Digitalen Information". Presseinformation der Max-Planck-Gesellschaft, FP/2008 (126), 12.Juni 2008.
- [eSciDoc] BMBF-Projekt: eSciDoc. <http://www.escidoc-project.de>
- [DCC] Digital Curation Centre. <http://www.dcc.ac.uk>
- [IEEE08] IEEE e-Science Konferenz. Special Session on Digital Repositories. Indianapolis, 7-12. Dezember 2008. <http://escience2008.iu.edu>
- [OAIS] Reference Model for an Open Archival Information System (OAIS). ISO Standard, 14721:2003. <http://nost.gsfc.nasa.gov/isoas>
- [OGF23] Repository Workshop auf dem Open Grid Forum 23. Barcelona, 5.Juni. <http://www.ogf.org>
- [nestor-daten] Severiens, Hilf: Langzeitarchivierung von Rohdaten. nestor-materialien 6. <http://edoc.hu-berlin.de/series/nestor-materialien/6/PDF/6.pdf>
- [Perseus] Perseus Digital Library. <http://www.perseus.tufts.edu>
- [RLG] RLG/OCLC Report: Trusted Digital Repositories: Attributes and Responsibilities. May 2002. <http://www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf>
- [TextGrid Manual] R3.5: TextGrid Manual - Tool Development. http://www.textgrid.de/fileadmin/TextGrid/reports/R3_5-manual-tools.pdf
- [UNESCO] UNESCO Guidelines for the Preservation of Digital Heritage. March 2003. <http://unesdoc.unesco.org/images/0013/001300/130071e.pdf>

AP 2 - TextGrid Laboratory

Nummer des Arbeitspaketes:	AP 2
Titel:	TextGridLab
AP-Leiter:	Trier
Start-Monat:	Juni 2009 (M1) bis Mai 2012 (M36)
Gesamtanzahl PM:	48

Personenmonate pro Partner gesamt und pro Unterarbeitspaket (UAP)¹⁵

Partner (Akronym)	DAASI	FHW	IDS	LMU	MPDL	SUB	Trier	TUKL	Würzburg	UPB	Gesamt
Gesamt:		8					36		4		48
AP 2											
AP 2.1: Tools		8					36*				44
AP 2.2: Migration									4*		4

Ziele (Kurzbeschreibung)

Ziel dieses Arbeitspakets ist die Sicherung der technischen Nachhaltigkeit im Rahmen der Gewährleistung einer stabilen Produktionsumgebung. Dazu gehören die Produktionsreife und Pflege der grundlegenden Werkzeuge des TextGridLab sowie die Evaluierung der Möglichkeiten der perspektivischen Weiterentwicklung hin zur Browser-basierten Architektur.

¹⁵ Die Verantwortlichen für das jeweilige Unterarbeitspaket sind mit einem * gekennzeichnet.

Beschreibung des Arbeitspaketes

AP 2.1: Allgemeine textwissenschaftliche Werkzeuge

In diesem UAP werden die übergreifenden Funktionalitäten und Werkzeuge des TextGridLab (weiter)entwickelt und in den Produktivbetrieb überführt. Zu diesen Werkzeugen gehören

- Grund-GUI mit Startbildschirm
- der Projektbrowser zum Betrachten und Verwalten der Objekte in den Projekten des Benutzers
- die Oberfläche zur Organisation von Objekten in Sammlungen
- das Administrationstool zur Rechteverwaltung
- der Metadaten-Editor
- der Workfloweditor
- das Recherchetool
- die technischen Komponenten der Hilfe

Reports (R) & Milestones (M)

Bezeichnung	Monat	Beschreibung
M 2.1.1	18	TextGridLab 1.5
M 2.1.2	30	TextGridLab 1.9
M 2.1.3	36	TextGridLab 2.0

laufend: Teilnahme an Entwicklerworkshops, Einarbeitung von Nutzer-Feedback

Abhängigkeiten zu anderen APs:

- AP 6.1 Generalisierung von Tools bzw. generell AP 6
- AP 4: Coaching (Einholung und Berücksichtigung von Nutzer-Feedback)

AP 2.2: Evaluierung der Möglichkeit einer Migration des TextGridLab zu einer Browser-basierten Lösung

Um den Fachwissenschaftlern den Zugang zu TextGrid zu erleichtern, wäre es wünschenswert, dass sie nicht einen eigenen Client für die Arbeit herunterladen müssen. Hier soll daher überprüft werden, ob es möglich ist, diesen Client als Browser-basierte Lösung zu implementieren. Dadurch würde die Mobilität des Zugriffs auf die Arbeitsumgebung deutlich gesteigert werden. In diesem AP soll es deshalb darum gehen, die technischen Möglichkeiten eines solchen Schrittes auszuloten. Es gilt zu prüfen, ob und wie Funktionalitäten des aktuellen TextGridLab in einem Web-basierten Gridportal angeboten werden können und wie eine halbautomatische Portierung von der aktuellen Eclipse-Rich-Client-Technologie hin zu einer Webanwendung organisiert werden kann.

Reports (R) & Milestones (M)		
Bezeichnung	Monat	Beschreibung
M 2.2.1	30	Evaluierung von technologischen Möglichkeiten zur Portierung

AP 3 - Strukturelle und organisatorische Nachhaltigkeit

Nummer des Arbeitspaketes:	AP 3
Titel:	Strukturelle und organisatorische Nachhaltigkeit
AP-Leiter:	IDS
Start-Monat:	Juni 2009 (M1) bis Mai 2012 (M36)
Gesamtanzahl PM:	21

Personenmonate pro Partner gesamt und pro Unterarbeitspaket (UAP)¹⁶

Partner (Akronym)	DAASI	FHW	IDS	LMU	MPDL	SUB	Trier	TUKL	Würzburg	UPB	Gesamt
Gesamt:	12		9*								21
AP 3											
AP 3.1 Rechtsform			6*								6
AP 3.2 Rechtliche Grundlagen	12		3*								15

Ziele (Kurzbeschreibung)

Ziel des Arbeitspaketes ist die strukturelle und organisatorische Nachhaltigkeit von TextGrid zu sichern. Dazu muss insbesondere eine Community-spezifische Rechtsform zur Integration in das D-Grid-Betriebsmodell implementiert werden.

¹⁶ Die Verantwortlichen für das jeweilige Unterarbeitspaket sind mit einem * gekennzeichnet.

Beschreibung des Arbeitspaketes

AP 3.1: Rechtsform für TextGrid

Voraussetzung für die organisatorische Nachhaltigkeit von TextGrid und die Nutzung des im WissGrid-Projekt entwickelten Betriebsmodells ist die Findung und Gründung einer geeigneten Rechtsform. Dazu muss zunächst ein Organisationsmodell entworfen werden, das Community-spezifisch verschiedene mögliche Rollen mit ihren Rechten und Pflichten identifiziert und definiert. Dabei werden Ergebnisse verwandter Initiativen, wie sie z.B. in dem europäischen Projekt DARIAH entwickelt werden, und die Arbeiten der anderen Community-Grids berücksichtigt. Außerdem müssen die verschiedenen in Frage kommenden deutschen und europäischen Rechtsformen im Hinblick auf das Organisationsmodell und das WissGrid-Betriebsmodell evaluiert werden und die am besten geeignete ausgewählt werden. Das Organisationsmodell muss dann nach einer Prüfungsphase, je nach gewählter Rechtsform z.B. in eine Satzung transformiert werden.

AP 3.2: Rechtliche Grundlagen

Im Unterschied zu den anderen Community-Grids sind die Forschungsprimärdaten der TextGrid-Community häufig nicht frei von Rechten Dritter. Um Content-Providern und Endnutzern einen rechtssicheren Umgang mit ihren Forschungsdaten in TextGrid zu ermöglichen, müssen daher Musterlizenzvereinbarungen zwischen unterschiedlichen Parteien (siehe Antrag) entwickelt werden und die sich daraus ergebenden Typen von Nutzungsbedingungen in der AAI / dem VO-Management technisch umgesetzt werden.

Reports (R) & Milestones (M)

Bezeichnung	Monat	Beschreibung
M 3.1.2	36	Implementierung des überarbeiteten WissGrid-Betriebsmodells. Gründung einer Organisation zur nachhaltigen Umsetzung des Betriebsmodells.
R 3.1.1	18	Draft-Version Organisationsmodell inkl. Rollen, Rechten und Nutzungsversprechen
R 3.2.1	24	Musterverträge (AP 3.2) und technische Umsetzung
R 3.1.2	30	Finales Organisationsmodell und Satzung/Statuten der zu gründenden TextGrid-Organisation

Beziehungen zu und Abhängigkeiten von anderen APs

- AP 4 – Einbeziehung der Zielgruppen in die Entwicklung eines tragfähigen Geschäfts- und Organisationsmodells
- AP 6 – Input zu fachwissenschaftsspezifischen Bedingungen und Anforderungen

AP 4 – Coaching / Community Bildung

Nummer des Arbeitspaketes:	AP 4
Titel:	Coaching & Community Building
AP-Leiter:	Trier
Start-Monat:	Juni 2009 (M1) bis Mai 2012 (M36)
Gesamtanzahl PM:	61

Personenmonate pro Partner gesamt und pro Unterarbeitspaket (UAP)¹⁷

Partner (Akronym)	DAASI	FHW	IDS	LMU	MPDL	SUB	Trier	TUKL	Würzburg	UPB	Gesamt
Gesamt:	6	6	7			1	34			7	61
AP 4											
AP 4.1: Bedarfsanalyse							6*				6
AP 4.2: Schulung	6	6	6				24*			6	48
AP 4.3: Community-Netzwerk			1			1*	4			1	7

Ziele (Kurzbeschreibung)

Ziel dieses Arbeitspaketes ist die Heranführung der wissenschaftlichen Fach-Communities an die virtuelle Forschungsinfrastruktur auf der Grundlage spezifischer Bedarfserhebung. In diesem Arbeitspaket wird daher die konkrete Ansprache der TextGrid-relevanten Adressaten und Träger konzipiert und umgesetzt, um so eine aktive Community-Bildung vorantreiben zu können:

¹⁷ Die Verantwortlichen für das jeweilige Unterarbeitspaket sind mit einem * gekennzeichnet.

Beschreibung des Arbeitspaketes

AP 4.1: Bedarfsanalyse: Die in der Abschlussphase von TextGrid I entwickelten Maßnahmen zur Bedarfserhebung und Usability-Analyse werden konsequent weiterentwickelt. Dazu zählen, wie oben beschrieben

- Beschreibung von Nutzerszenarien,
- intensive und persönliche Workshopevaluierung,
- Kontrastierung der Nutzeraktivitäten mit den Interviews während der Schulung,
- iterativer Prozess der Einarbeitung des Nutzerfeedbacks,
- wiederholte Befragung der Nutzer

AP 4.2: Schulung: Auf der Basis von Projektdokumentation und nutzeradäquaten Tutorials (auch als eTutorials) werden Schulungsmaßnahmen für unterschiedliche Adressatengruppen konzipiert und durchgeführt.

Schulungsmaßnahmen

Grundlage und gemeinsame Bezugsbasis für alle Schulungsmaßnahmen ist die ausführliche Dokumentation der TextGrid-Komponenten (Gesamtarchitektur, Baseline-Encoding, Tools etc.), die zweisprachig vorgenommen werden soll (Deutsch und Englisch). Ergänzend zur Dokumentation werden didaktisch aufbereitete Materialien für den Face-to-Face-Unterricht, aber auch ausgearbeitete eTutorials zur Verfügung gestellt. Unter Berücksichtigung der in AP 4.1 erhobenen spezifischen Community-Struktur und des ermittelten Bedarfs werden Schulungsmaßnahmen konzipiert und durchgeführt. Dabei wird differenziert in Seminare für Entwickler, Nutzer, Content-Provider und ggf. weitere Trägergruppen. Durch Modularisierung der Schulungseinheiten können flexible, individuelle und bedarfsgerechte Pakete geschnürt werden, bspw.:

- Grundlagen / Einführung in TextGrid
- Entwickler- / Programmierer-Schulungen, z.B. Eclipse, Webservice etc.
- Community-spezifische Anwendungen und Tools bzw. Community-spezifische Technologie-Entwicklung, z.B. Grid-Anwendungen für Altphilologen, TextGrid für Linguisten, Grid-Anwendungen für Editionsphilologen etc.
- Toolspezifische Schulungen, z.B. Kollationierer, Lemmatisierer. Einführung in das Tool zur Codierung von Noten etc.
- Train-the-Trainer-Schulungen für TextGrid-Vermittler

Für die Umsetzung des Konzepts kann auf die Mitarbeit des Würzburger Kompetenzzentrums für EDV-Philologie zurückgegriffen werden, das mit der Ausbildung von Postgraduate-Studierenden im Bereich der EDV-Philologie und der in Würzburg seit 2003 entwickelten <philtag>-Workshops im Bereich "Kodierungsverfahren" mit internationalen

Teilnehmern aus Fachwissenschaft, EDV-Philologie und Informatik über langjährige Erfahrungen auf diesem Gebiet verfügt. Das Trierer Kompetenzzentrum ist ferner in der einschlägigen TEI Special Interest Group "Education" aktiv. Insgesamt ist folgendes übergreifendes Schulungskonzept vorgesehen: In Monat 18 (mit TextGrid-Release 1.5) sowie zum Abschluss in Monat 30 (TextGrid-Release 1.9) sind gemeinsame TextGrid-weite öffentliche mehrtägige Maßnahmen geplant, die jeweils gebündelt an einem Tag bzw. in Parallel-Sessions die Schwerpunkte für Content-Provider, Fachwissenschaftler und Entwickler setzen. Ferner werden große Community-Tagungen genutzt, um TextGrid-Workshops zu platzieren, wie etwa die TEI-Jahrestagung, die Jahrestagung der Arbeitsgemeinschaft germanistischer Editionen u.a.m. Dazwischen werden in halbjährlichem Turnus kleinere, zweitägige Hands-on-Meetings für die spezifischen Experten angeboten, dies in Monat 6, Monat 12 und Monat 24, die von einem oder mehreren der zuständigen TextGrid-Partner ausgerichtet werden können. In Monat 36 findet schließlich ein großes TextGrid-Summit statt. Ferner soll flexibel auf Anfragen reagiert werden, spezifische Schulungen bei interessanten Communities bzw. Community-Events (Fachverbandtagungen etc.) durchzuführen. Zwei interne Schulungen in den Jahren 2010 und 2011 dienen dazu, alle beteiligten Mitarbeiter auf denselben Stand zu bringen und insbesondere die Kommunikation zwischen Entwicklern und Fachwissenschaftlern zu pflegen und damit die Arbeitsprozesse zu optimieren.

AP 4.3: Community-Netzwerk: Ziel dieses Arbeitspaketes ist der Aufbau und der Betrieb eines nachhaltigen nationalen eHumanities-Netzwerkes, das mit internationalen Initiativen und Netzwerken eng kooperiert. Dazu werden geeignete Institutionen, Fächer, Projekte und Keyplayer (inkl. Entscheidungsträger, Verbände, Fachgesellschaften etc.) identifiziert und dokumentiert; Kontakte werden systematisch und gezielt hergestellt und gepflegt.

Für die Content-Provider gilt es, langfristig attraktive Angebote zu erarbeiten, so dass wissenschaftlich relevanter Content über das TextGridRep verfügbar gemacht und nachhaltig gespeichert werden kann.

Für die fachwissenschaftlichen Angebote gilt es, auf vorhandene Fächerstrukturen und -kulturen einzugehen und auf der Basis von State-of-the-Art- und Bedarfsanalyse maßgeschneiderte Angebote zu entwickeln.

Für Entwickler müssen zum einen Anreize geschaffen werden, Dienste und Tools kooperativ mit und in der TextGrid-Community zu entwickeln, zum anderen Workshops und Schulungen aus dem Bereich der Architektur und Technologie angeboten werden. Ein wichtiges Ziel wird es sein, eine Roadmap für die Bildung einer aktiven Open Source Community zu entwickeln, die die technische Entwicklung auch nach Projektende weiter betreibt.

Bei den Fachkollegen, den Fachgesellschaften, den Universitätsleitungen und den politischen Entscheidungsträgern muss ferner ein Bewusstsein für die Relevanz nicht nur des Infrastrukturaufbaus und -betriebs (vgl. auch AP 3), sondern auch der Infrastrukturforschung geschaffen werden. Die Einbindung von eHumanities-Themen in curriculare Standards sowie gezielte Nachwuchsförderung gehören ebenfalls zu den wichtigen Maßnahmen einer erfolgreichen Community-Bildung.

Reports (R) & Milestones (M)

Bezeichnung	Monat	Beschreibung
M 4.1.1	6 (und danach laufende Begleitung)	Erstellung bzw. Einholung und laufende Aktualisierung von Nutzerszenarien und Nutzerfeedback; Vermittlung in die Entwicklercommunity in geeigneter Form
M 4.2.1	6 (und danach laufende Aktualisierung)	Dokumentation TextGrid, Erstellung von Tutorials (laufende Erweiterung unter Berücksichtigung des sich aus der Community artikulierenden Bedarfs).
R 4.3.1	12	Bericht zu eHumanities-Themen mit Relevanz für die allgemeine Informatik.
R 4.2.1	18	Bericht zu notwendigen Veränderungen und Ergänzungen bei den Issue-Trackern etc. für Entwickler.

Schulungsmaßnahmen

Bezeichnung	Monat	Beschreibung
S 1.1	6	Hands-on-Meeting Content-Provider
S 1.2	6	Hands-on-Meeting Fachwissenschaftler
S 1.3	6	Hands-on-Meeting Entwickler
S 2.1	12	Hands-on-Meeting Content-Provider
S 2.2	12	Hands-on-Meeting Fachwissenschaftler
S 2.3	12	Hands-on-Meeting Entwickler
S 3	18	TextGrid-weiter Workshop Content-Provider, Fachwissenschaftler, Entwickler
S 4.1	24	Hands-on-Meeting Content-Provider
S 4.2	24	Hands-on-Meeting Fachwissenschaftler
S 4.3	24	Hands-on-Meeting Entwickler
S 5	30	TextGrid-weiter Workshop Content-Provider, Fachwissenschaftler, Entwickler
S 6	36	TextGrid Summit

Weitere Workshops und Schulungen auf großen Fachkonferenzen orientieren sich am internationalen Tagungskalender. Weitere spezifische Schulungen werden flexibel auf Anfrage und nach Kapazität entweder bei einem der TextGrid-Partner oder bei der

anfragenden Institution durchgeführt. Solche Schulungen auf Anfrage können z.T. von diesen Institutionen (mit)finanziert werden.

Abhängigkeiten zu anderen APs

- Das AP zur Community-Bildung steht in enger Abhängigkeit zu allen anderen APs.

AP 6 - Fachwissenschaftliche Nachhaltigkeit

Nummer des Arbeitspaketes:	AP 6
Titel:	Fachwissenschaftliche Nachhaltigkeit
AP-Leiter:	Universität Würzburg
Start-Monat:	Juni 2009 (M1) bis Mai 2012 (M36)
Gesamtanzahl PM:	264 PM (entspricht 7,3 FTE)

Personenmonate pro Partner gesamt und pro Unterarbeitspaket (UAP)¹⁸

Partner (Akronym)	DAASI	FHW	IDS	LMU	MPDL	SUB	Trier	TUKL	Würzburg	UPB	Gesamt
Gesamt:	0	22	18	18	24	25	38	18	67 (+18)¹⁹	20	268
AP 6											
AP 6.1: Generalisierung, Koordination und Produktionsreife		22	2			25	18		57*		124
AP 6.2: Literaturwissenschaftliche Editionsphilologie							20*		10		30

¹⁸ Die Verantwortlichen für das jeweilige Unterarbeitspaket sind mit einem * gekennzeichnet.

¹⁹ 18 PM für die »leere« Fachsäule, vgl. die Beschreibung zu AP 6.7.

AP 6.3: Linguistik			16*		12						28
AP 6.4: Kunstgeschichte					12*						12
AP 6.5: Klassische Philologien, Geschichte				18*							18
AP 6.6: Musikwissenschaftliche Edition									20*		20
AP 6.7: Freie Fachsäule								(18)			18
AP 6.8: OCR und Dokumentbildverarbeitung							18*				18

Ziele (Kurzbeschreibung)

Das Projekt TextGrid ist in der kurzen Zeit seines Bestehens auf eine ausgesprochen positive und breite Resonanz gestoßen. Für die Nachhaltigkeit des Erreichten ist es unbedingt erforderlich, an diesen Erfolg anzuknüpfen und TextGrid auf eine möglichst breite fachliche Basis zu stellen. Um dieses Ziel zu erreichen, wird in diesem AP eine Doppelstrategie verfolgt: Zum einen werden die im Kontext von TextGrid I entwickelten Prototypen zur Produktionsreife gebracht, denn nur robuste und leistungsfähige Software 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 I entstandener Module in TextGrid einen ähnlichen Entwicklungsstand wie die aus TextGrid I übernommenen Module erreichen. Neben Ausbau und Pflege der in der ersten Projektphase vorrangig im Schwerpunkt der Editionsphilologie entwickelten Werkzeuge werden in TextGrid II also weitere geisteswissenschaftliche Communities eingebunden. Die Auswahl der Communities erfolgte aufgrund ihrer Kompetenz im Umgang mit Virtuellen Forschungsumgebungen,

ihrer Größe und Bedeutung für die Aufbereitung und Sicherung des kulturellen Erbes und mit der Absicht, neue Datentypen zu integrieren.

Beschreibung des Arbeitspaketes

Ziel dieses Arbeitspaketes ist es einerseits, die bereits in Form von Prototypen vorhandenen Werkzeuge zur Produktionsreife weiterzuentwickeln und sie dabei so generalisieren, dass sie auch in verwandten geisteswissenschaftlichen Communities eingesetzt werden können. Außerdem sollen einige neue Werkzeuge die Palette der in TextGrid vorhandenen Werkzeuge so erweitern, dass viele typische Arbeitsabläufe in den Textwissenschaften in dieser Umgebung durchführbar sind.

Die Struktur der Unterarbeitspakete (UAP) sieht folgendermaßen aus (ausführliche Beschreibung findet sich im TextGrid II Antrag in der Version Juni 2008):

Aufgrund des Umfangs dieses Arbeitspaketes findet sich am Ende eine komplette Aufstellung aller Milestones und Reports des AP 6.

AP 6.1: Generalisierung, Koordination und Produktionsreife

Dieses UAP hat die Aufgabe, die Prototypen der allgemeinen Werkzeuge aus der TextGrid I Entwicklungsphase zur Produktionsreife zu führen, die Arbeit der anderen UAP zu koordinieren und die entstandenen und entstehenden Werkzeuge auf ihre Generalisierbarkeit zu untersuchen, um auf diese Weise ihre Verwendbarkeit in anderen Kontexten zu gewährleisten.

Milestones (M) & Reports (R)

Vgl. auch die Gesamtmilestones für AP 6 am Ende der AP 6-Beschreibung

Bezeichnung	Monat	Beschreibung
M 6.1.0	2	Planung und Durchführung eines internen Entwickler-Koordinationsworkshops am Beginn des Projekts, erste gemeinsame Leitlinien, Planung für die Koordination
M 6.1.1	6	Release einer ersten integrierten Version von TextGrid-Tools des APs. Dies umfasst die Koordination des gemeinsamen Milestones M 6.1g sowie die Integration von Weiterentwicklungen auf der Basis von Erkenntnissen aus TextGrid-Summit und Entwicklerworkshop im Januar 2009, sowie dem Tutorial im Rahmen der IEEE DEST
M 6.1.2	18	Koordination des gemeinsamen Zwischenreleases »TextGrid-Lab 1.5« (vgl. M 6.2g unten) incl. weiterer Desiderate aus <ul style="list-style-type: none"> ▪ TextGrid-Summit und Entwicklerworkshop ▪ dem IEEE DEST Tutorial

		<ul style="list-style-type: none"> ▪ der ersten Runde der Bedarfserhebung und Usability-Evaluation, vgl. AP 4 ▪ dem Feedback aus dem Faust-Editionsprojekt ▪ laufend über die sonstigen Kanäle wie die eingehenden Rückmeldungen aus dem Feedbackformular
M 6.1.3	30	<p>Release 1.9 Aus AP 6.1 enthalten sein werden</p> <ul style="list-style-type: none"> ▪ Feedback aus dem Projekt <i>Digitalisierung eines Briefkorpus (Staatsbibliothek Berlin)</i> ▪ Feedback aus der weiteren Bedarfserhebung und Usability-Evaluation (vgl. AP 4) ▪ Ergebnisse weiterer Anwenderworkshops (vgl. AP 4) ▪ Weiterentwicklungen der o.g. Tools hin zu Produktionsreife
M 6.1.4	36	<p>produktionsreifes Abschlussrelease: TextGrid 2.0, das die eingegangenen Desiderate umsetzt. Mit AP 4 / S 4.3: Abschlussworkshops für Anwender und Entwickler.</p>
laufend		Einarbeitung des über die diversen Kanäle eingegangenen Feedbacks (AP 4)

AP 6.2: Literaturwissenschaftliche Editionsphilologie

In diesem UAP wird die Entwicklung von zwei editionsphilologischen Werkzeugen, einem Kollationierungsprogramm und einem Text-Bild-Editor, auf der Grundlage des in TextGrid I Geleisteten durchgeführt.

Milestones (M) & Reports (R)

Bezeichnung	Monat	Beschreibung
M 6.2.1	6	Erster Integrationsschritt mit den anderen UAPs
M 6.2.2	30	Fertigstellung des Linkeditors Text-Bild und des Kollationierungsprogramms
M 6.2.3	36	Release 2.0: Einarbeitung von Feedback in beide Tools, Produktionsreife

AP 6.3: Linguistik

Dieses UAP wird die Integration der umfangreichen Ressourcen des IDS und des MPI für Psycholinguistik, sowohl auf der Ebene der Daten wie der Dienste, in die TextGrid-Umgebung durchführen.

Milestones (M) & Reports (R)

Bezeichnung	Monat	Beschreibung
M 6.3.1	6	Webservice-Integration (Funktionsebene): Föderierte Suche über COSMAS-II bzw. OWID in TextGrid Requirements-Analyse und Spezifikation der Schnittstellen zu LEXUS und ANNEX/IMEX
M 6.3.2	18	Fertigstellung der Webservice-Integration auf Metadatenebene Fertigstellung des LEXUS-Interface 1. Version der Konvertierungsroutinen
M 6.3.3	30	Fertigstellung der gesamten Webservice-Integration für COSMAS II und OWID Fertigstellung des ANNEX/IMEX-Interface

AP 6.4: Kunstgeschichte

Dieses UAP wird Digilib, ein bereits bestehende Werkzeug zur Bereitstellung und Annotation von Bilddaten, in die TextGrid-Umgebung integrieren. Hinzukommt die Integration einer TEI-orientierten Publikationsplattform auf der Basis von eSciDoc in TextGrid.

Milestones (M) & Reports (R)

Bezeichnung	Monat	Beschreibung
M 6.4.1	6	Definition des Architekturmodells zur Anbindung von Digilib Definition von Repräsentationsformaten zur Darstellung von TEI-Dokumenten innerhalb von strukturellen Editionen
M 6.4.2	18	Unterstützung von zunächst 2 Repräsentationsformaten für strukturelle Editionen Fertigstellung der ersten integrierten Version von Digilib. Erhebung des Anwenderfeedbacks
M 6.4.3	30	Zweite Version der Anbindung von Digilib unter Einbeziehung des Anwenderfeedbacks und mit Bereitstellung der erweiterten Funktionalitäten (Messfunktionen, Metadatenintegration)

		Integration der Repräsentationsformate für eine wissenschaftliche strukturelle Edition Revision der erstellen strukturellen Edition
M 6.4.4	36	Vollständige Integration der entwickelten Tools in die TextGrid-Plattform und Schulungen für die Anwendenden

AP 6.5: Klassische Philologien, Geschichte

Weiterentwicklung des XML-Editors und der Text-Bild-Annotationssoftware zur Edition von Glossen.

Milestones (M) & Reports (R)

Bezeichnung	Monat	Beschreibung
M 6.5.1	6	Erarbeitung eines adäquaten Auszeichnungssystems für glossierte Texte und Integration in TEI (Text Encoding Initiative)
M 6.5.2	18	Weiterentwicklung des Text-Bild-Linkeditors (Image-Mapping) für die Anforderungen glossierter Texte; Erstellung des Präsentationstools;
M 6.5.3	30	Fertigstellung des Editionstools für glossierte Texte
M 6.5.4	36	Einarbeitung von Feedback in Präsentations- und Editionstool; Produktionsreife; Integration in die TextGrid-Plattform (XML-Editor)

AP 6.6: Musikwissenschaftliche Edition

Weiterentwicklung des XML-Editors zur Eingabe und Darstellung von Noten.

Milestones (M) & Reports (R)

Bezeichnung	Monat	Titel/Beschreibung
M 6.6.1	6	Erste TextGrid-Integration und Dokumentation: Integration der erstellten musikalischer Sonderzeichen in den XML-Editor (vgl. AP 6.6.1); Dokumentation des neuen Tools (vgl. AP 4.)
M 6.6.2	18	Anzeigetool:

		Integration des aktuellen Zwischenstands der Entwicklung des Noten-Anzeigetools in die TextGrid-Infrastruktur (vgl. AP 6.6.3*)
M 6.6.3	30	Anzeige- und Eingabetool: Fertigstellung des Anzeige- und Eingabetools (vgl. AP 6.6.3*)

* AP 6.6.2 aus der Version des Antrags von 2008 entfällt aufgrund des reduzierten Projektumfangs. Die notwendigen Vorarbeiten für AP 6.6.3 werden auf ein Minimum begrenzt und in AP 6.6.3 übernommen; auf die Bereitstellung von Masken für den Metadaten-Editor muss ebenso verzichtet werden wie auf Im-/Exportfunktionen zu MusicXML (AP 6.6.3).

Ansonsten bleibt die Struktur des Arbeitspakets aus der Version des Antrags von 2008 erhalten.

AP 6.7: Freie Fachsäule

Dieses Modul soll im Laufe von TextGrid II zugeordnet werden, um eine sich eröffnende Kooperation mit einem existierenden Projekt realisieren zu können, das bereits fachspezifische Software entwickelt hat, die in TextGrid integriert werden soll.

AP 6.8: OCR und Dokumentbildverarbeitung

Diese Gruppe von Diensten beseitigt ein Defizit aller textbasierten Digital Humanities, indem es die optische Zeichenerkennung von Frakturschriften unter Berücksichtigung großer historisch varianter Textmengen ermöglicht.

Milestones (M) & Reports (R)

Bezeichnung	Monat	Beschreibung
M 6.8.1	6	Erste Integration Das vom TUKL/DFKI entwickelte Open-Source System OCRopus wird in die TextGrid-Architektur integriert und so allen Partnern verfügbar gemacht. Gleichzeitig werden eine separate Webschnittstelle und Webservices entwickelt, die die Benutzung des OCR Systems ermöglichen.
M 6.8.2	18	Release 1.5 Zum Zwischenrelease nach der Hälfte der Projektlaufzeit wird die Erkennung von gebrochener Schrift zur Verfügung stehen inklusive einer Reihe von Varianten und eine Schätzung der Fontgröße des erkannten Textes. Im Dialog mit den Projektpartnern werden weitere Elemente für die Ausgabe identifiziert.

<p>M 6.8.3</p>	<p>30</p>	<p>Release 1.9 Sechs Monate vor Projektende sind alle im Vorfeld definierten Funktionen verfügbar. Insbesondere erlaubt die Frakturerkennung die Verarbeitung von Texten mit mehreren Schriftarten und hält entsprechende Informationen im Ausgabeformat vor.</p>
<p>M 6.8.4</p>	<p>36</p>	<p>Release 2.0 Während der finalen Integration und Evaluation werden Stabilisierung und Dokumentation der Software im Vordergrund stehen.</p>

Die folgende Tabelle zeigt den Stand der Entwicklung an den Werkzeugen:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z																								
Status TextGrid 23.2.2009																																																		
	allgemein																																																	
	Editionsphilologie					Linguistik					Kunstgeschichte					klass. Philologie					Musikwissenschaft					seit Projektphase																								
	1					2					3					4					5					1					2					3					4					5				
	Kern-features					Zusatz-features																																												
3	Tools																																																	
6	TextGridLab (GUI, allgemeine Tools)																																																	
7	Projektverwaltung																																																	
8	Projektbrowser / Navigator (GUI)																																																	
9	Metadaten-Annotation																																																	
10	XML-Editor																																																	
11	Glossen-Editor																																																	
12	Noten-Editor																																																	
13	Workflow-Editor																																																	
14	Recherchetool (Frontend TG-search)																																																	
15	Text-Bild-Editor																																																	
16	Glossen-Annotation																																																	
17	Integration Digilib																																																	
18	Kollationierung																																																	
19	Link-Editor Text																																																	
20	Lemmatisierung																																																	
21	Sortieren *																																																	
22	Streaming-Editor (XSLT, für XML) *																																																	
23	Streaming-Editor (für -XML)																																																	
24	Tokenizer *																																																	
25	Upload-Tool																																																	
26	Wörterbuchnetz-Integration																																																	
27	Integration der COSMAS und LEXUS Services																																																	
28	Bibliophietool																																																	
29	Text Publisher (Print)																																																	
30	Text Publisher (Web)																																																	
31	OCR																																																	
32																																																		
33	Infrastruktur																																																	
34																																																		
35	TG-auth*: AAI																																																	
36	TG-crud: Datenverwaltung																																																	
37	TG-search: Suche																																																	
38	TG-log: Logging-Framework																																																	
39	PyTGC: WSGI-basierter Web Service Container für Services in Python																																																	
40	eSciDoc-Integration																																																	
41	(LZA-Repository)																																																	
42																																																		
43																																																		
44																																																		

Legende:
 0 nicht begonnen / nur Planung
 1 Entwicklung prototypisch
 2 Entwicklung unfertig, nicht getestet
 3 Entwicklung (alpha) fertig, internes Feedback vorhanden
 4 Entwicklung (beta) fertig, Community-Feedback vorhanden
 5 Community-Feedback eingearbeitet

(ausgelagert in separaten DFG Antrag)

* Dienste ursprünglich für die Einbindung in Workflows, nicht zur Steuerung per GUI gedacht. Da Workflows erst seit kurzem verfügbar sind, gibt es bislang kein nennenswertes (internes) Feedback

Die Planung der Milestones und Reports hat sich gegenüber dem Hauptantrag nicht verändert.

Übergreifende Reports (R) & Milestones (M) für AP 6

Aufgrund der Größe von AP 6 entwirft jedes Unterarbeitspaket von AP 6 eine eigene, auf die individuelle Tätigkeit im UAP zugeschnittene Milestoneplanung. Um die Entwicklung innerhalb des Arbeitspaketes zu koordinieren, gibt es dennoch eine Reihe von gemeinsamen Milestones. Die einzelnen Unterarbeitspakete planen dabei, welchen Beitrag aus ihrem Arbeitsfeld sie zu jedem der gemeinsamen Milestones liefern. Die Koordination erfolgt durch AP 6.1.

Die Releaseplanungen werden mit den anderen relevanten Arbeitspaketen (etwa AP 2, AP 4) koordiniert.

Bezeichnung	Monat	Beschreibung
M 6.0g	2	Entwickler-Koordinationstreffen Ein Treffen (gemeinsam mit AP 4), bei dem mit allen Entwicklern, insbesondere auch der neuen Partner bzw. der Neueinstellungen, über den Stand von TextGrid und die Planungen für die neue Projektphase gesprochen wird, außerdem grundlegende Vereinbarungen für Entwicklungsprozesse etc. Näher beschrieben in AP 6.1
M 6.1g	6	erste AP 6-übergreifende Integration Zu diesem Milestone sollte jeder Partner / jedes UAP eine Grundversion eines ersten Tools vollständig in die TextGrid-Architektur integrieren. Ziel ist es, frühzeitig für eine Integration zu sorgen und zu verhindern, dass Insellösungen entstehen, die sich später nur schwer integrieren lassen. Die künftige Entwicklung sollte auf dieser Integration aufbauen, es sollte also möglichst weiter gemeinsam in einer integrierten Lösung entwickelt werden.
M 6.2g	18	Release 1.5 Nach einer konzentrierten Entwicklungsphase mit UAP-spezifischer Detailplanung sollte hier eine Synchronisation auf ein gemeinsames stabiles Zwischenrelease erfolgen.
M 6.3g	30	Release 1.9 Sechs Monate vor Projektende sollten zumindest die Grundgerüste aller Tools stehen und zu einem Release vereinigt werden, sodass die restliche Projektlaufzeit für

		Tests, die Einarbeitung/Berücksichtigung der Testergebnisse und des Feedbacks aus der Community, die weitere Abstimmung mit anderen Komponenten, die Fertigstellung der Dokumentation und die Publikation als Teil der TextGrid-Open-Source-Software zur Verfügung steht.
M 6.4g	36	Release 2.0 Abschlussrelease mit allen Entwicklungen.
Laufend		Während der gesamten Projektlaufzeit werden eingehende Rückmeldungen zu Fehlern, Unzulänglichkeiten und neuen Anforderungen in die Software integriert

Reports

Zu den Monaten 12, 24 und 36 wird jeweils in einem Report (Bezeichnung: R 6.0.1, R 6.0.2 sowie R 6.0.3) über die Entwicklung des vergangenen Jahres sowie den Stand der Planungen (Pflichtenhefte) berichtet. Die Berichte über die entwickelten Tools sollten dabei als Endnutzerdokumentation geschrieben (und in AP 4 eingebracht) werden.

Alle Reports (R) & Milestones (M) für AP 6

Bezeichnung	Monat	Beschreibung
M 6.1.0	2	Planung und Durchführung eines internen Entwickler-Koordinationsworkshops am Beginn des Projekts, erste gemeinsame Leitlinien, Planung für die Koordination
M 6.1.1	6	Release einer ersten integrierten Version von TextGrid-Tools des APs. Dies umfasst die Koordination des gemeinsamen Milestones M 6.1g sowie die Integration von Weiterentwicklungen auf der Basis von Erkenntnissen aus TextGrid-Summit und Entwicklerworkshop im Januar 2009, sowie dem Tutorial im Rahmen der IEEE DEST
M 6.1.2	18	Koordination des gemeinsamen Zwischenreleases »TextGridLab 1.5« (vgl. M 6.2g unten) incl. weiterer Desiderate aus <ul style="list-style-type: none"> ▪ TextGrid-Summit und Entwicklerworkshop ▪ dem IEEE DEST Tutorial ▪ der ersten Runde der Bedarfserhebung und Usability-Evaluation, vgl. AP 4 ▪ dem Feedback aus dem Faust-Editionsprojekt ▪ laufend über die sonstigen Kanäle wie die eingehenden

M 6.1.3	30	<p>Release 1.9</p> <p>Aus AP 6.1 enthalten sein werden</p> <ul style="list-style-type: none"> ▪ Feedback aus dem Projekt <i>Digitalisierung eines Briefkorpus (Staatsbibliothek Berlin)</i> ▪ Feedback aus der weiteren Bedarfserhebung und Usability-Evaluation (vgl. AP 4) ▪ Ergebnisse weiterer Anwenderworkshops (vgl. AP 4) ▪ Weiterentwicklungen der o.g. Tools hin zu Produktionsreife
M 6.1.4	36	<p>produktionsreifes Abschlussrelease: TextGrid 2.0, das die eingegangenen Desiderate und die Anforderungen umsetzt. Mit AP 4 / S 4.3: Abschlussworkshops für Anwender und Entwickler.</p>
M 6.2.1	6	Erster Integrationsschritt mit den anderen UAPs
M 6.2.2	30	Fertigstellung des Linkeditors Text-Bild und des Kollationierungsprogramms
M 6.2.3	36	Release 2.0: Einarbeitung von Feedback in beide Tools, Produktionsreife
M 6.3.1	6	<p>Webservice-Integration (Funktionsebene): Föderierte Suche über COSMAS-II bzw. OWID in TextGrid</p> <p>Requirements-Analyse und Spezifikation der Schnittstellen zu LEXUS und ANNEX/IMEX</p>
M 6.3.2	18	<p>Fertigstellung der Webservice-Integration auf Metadatenebene</p> <p>Fertigstellung des LEXUS-Interface</p> <p>1. Version der Konvertierungsroutinen</p>
M 6.3.3	30	<p>Fertigstellung der gesamten Webservice-Integration für COSMAS II und OWID</p> <p>Fertigstellung des ANNEX/IMEX-Interface</p>
M 6.4.1	6	<p>Definition des Architekturmodells zur Anbindung von Digilib</p> <p>Definition von Repräsentationsformaten zur Darstellung von TEI-Dokumenten innerhalb von strukturellen Editionen</p>
M 6.4.2	18	<p>Unterstützung von zunächst 2 Repräsentationsformaten für strukturelle Editionen</p> <p>Fertigstellung der ersten integrierten Version von Digilib.</p>

		Erhebung des Anwenderfeedbacks
M 6.4.3	30	Zweite Version der Anbindung von Digilib unter Einbeziehung des Anwenderfeedbacks und mit Bereitstellung der erweiterten Funktionalitäten (Messfunktionen, Metadatenintegration) Integration der Repräsentationsformate für eine wissenschaftliche strukturelle Edition Revision der erstellten strukturellen Edition
M 6.4.4	36	Vollständige Integration der entwickelten Tools in die TextGrid-Plattform und Schulungen für die Anwendenden
M 6.5.1	6	Erarbeitung eines adäquaten Auszeichnungssystems für glossierte Texte und Integration in TEI (Text Encoding Initiative)
M 6.5.2	18	Weiterentwicklung des Text-Bild-Linkeditors (Image-Mapping) für die Anforderungen glossierter Texte; Erstellung des Präsentationstools;
M 6.5.3	30	Fertigstellung des Editionstools für glossierte Texte
M 6.5.4	36	Einarbeitung von Feedback in Präsentations- und Editionstool; Produktionsreife; Integration in die TextGrid-Plattform (XML-Editor)
M 6.6.1	6	Erste TextGrid-Integration und Dokumentation: Integration der erstellten musikalischer Sonderzeichen in den XML-Editor (vgl. AP 6.6.1); Dokumentation des neuen Tools (vgl. AP 4.)
M 6.6.2	18	Anzeigetool: Integration des aktuellen Zwischenstands der Entwicklung des Noten-Anzeigetools in die TextGrid-Infrastruktur (vgl. AP 6.6.3)
M 6.6.3	30	Anzeige- und Eingabetool: Fertigstellung des Anzeige- und Eingabetools (vgl. AP 6.6.3)
M 6.8.1	6	Erste Integration Das vom TUKL/DFKI entwickelte Open-Source System OCRopus wird in die TextGrid-Architektur integriert und so allen Partnern verfügbar gemacht. Gleichzeitig werden eine separate Webschnittstelle und Webservices entwickelt, die die Benutzung des OCR Systems ermöglichen.

M 6.8.2	18	Release 1.5 Zum Zwischenrelease nach der Hälfte der Projektlaufzeit wird die Erkennung von gebrochener Schrift zur Verfügung stehen inklusive einer Reihe von Varianten und eine Schätzung der Fontgröße des erkannten Textes. Im Dialog mit den Projektpartnern werden weitere Elemente für die Ausgabe identifiziert.
M 6.8.3	30	Release 1.9 Sechs Monate vor Projektende sind alle im Vorfeld definierten Funktionen verfügbar. Insbesondere erlaubt die Frakturerkennung die Verarbeitung von Texten mit mehreren Schriftarten und hält entsprechende Informationen im Ausgabeformat vor.
M 6.8.4	36	Release 2.0 Während der finalen Integration und Evaluation werden Stabilisierung und Dokumentation der Software im Vordergrund stehen.
M 6.0g	2	Entwickler-Koordinationstreffen Ein Treffen (gemeinsam mit AP 4), bei dem mit allen Entwicklern, insbesondere auch der neuen Partner bzw. der Neueinstellungen, über den Stand von TextGrid und die Planungen für die neue Projektphase gesprochen wird, außerdem grundlegende Vereinbarungen für Entwicklungsprozesse etc. Näher beschrieben in AP 6.1
M 6.1g	6	erste AP 6-übergreifende Integration Zu diesem Milestone sollte jeder Partner / jedes UAP eine Grundversion eines ersten Tools vollständig in die TextGrid-Architektur integrieren. Ziel ist es, frühzeitig für eine Integration zu sorgen und zu verhindern, dass Insellösungen entstehen, die sich später nur schwer integrieren lassen. Die künftige Entwicklung sollte auf dieser Integration aufbauen, es sollte also möglichst weiter gemeinsam in einer integrierten Lösung entwickelt werden.
M 6.2g	18	Release 1.5 Nach einer konzentrierten Entwicklungsphase mit UAP-spezifischer Detailplanung sollte hier eine Synchronisation auf ein gemeinsames stabiles Zwischenrelease erfolgen.
M 6.3g	30	Release 1.9 Sechs Monate vor Projektende sollten zumindest die

		Grundgerüste aller Tools stehen und zu einem Release vereinigt werden, sodass die restliche Projektlaufzeit für Tests, die Einarbeitung/Berücksichtigung der Testergebnisse und des Feedbacks aus der Community, die weitere Abstimmung mit anderen Komponenten, die Fertigstellung der Dokumentation und die Publikation als Teil der TextGrid- <i>Open-Source</i> -Software zur Verfügung steht.
R 6.0.1	12	Report über die Entwicklung des vergangenen Jahres sowie den Stand der Planungen (Pflichtenhefte). Die Berichte über die entwickelten Tools sollten dabei als Endnutzerdokumentation geschrieben (und in AP4 eingebracht) werden.
R 6.0.2	24	Report über die Entwicklung des vergangenen Jahres sowie den Stand der Planungen (Pflichtenhefte). Die Berichte über die entwickelten Tools sollten dabei als Endnutzerdokumentation geschrieben (und in AP4 eingebracht) werden.
R 6.0.3	36	Report über die Entwicklung des vergangenen Jahres sowie den Stand der Planungen (Pflichtenhefte). Die Berichte über die entwickelten Tools sollten dabei als Endnutzerdokumentation geschrieben (und in AP4 eingebracht) werden.
M 6.4g	36	Release 2.0 Abschlussrelease mit allen Entwicklungen.
Laufend		Während der gesamten Projektlaufzeit werden eingehende Rückmeldungen zu Fehlern, Unzulänglichkeiten und neuen Anforderungen in die Software integriert

Eine Übersicht aller Reports und Milestones findet sich zudem im Anhang II, Ressourcenplanung.

AP 8 - Projektmanagement, Öffentlichkeitsarbeit, RI-Liaison

Nummer des Arbeitspaketes:	AP 8
Titel:	Projektmanagement, Öffentlichkeitsarbeit, RI-Liaison
AP-Leiter:	SUB
Start-Monat:	Juni 2009 (M1) bis Mai 2012 (M36)
Gesamtanzahl PM:	38 (plus 50 % E 5, SHK)

Personenmonate pro Partner gesamt

Partner (Akronym)	DAASI	FHW	IDS	LMU	MPDL	SUB	Trier	TUKL	Würzburg	UPB	Gesamt
Gesamt:			1			41			1		43
AP 8			1			41			1		43

Ziele (Kurzbeschreibung)

Ziel des Arbeitspaketes ist, sicher zu stellen, dass das Projekt alle Ziele innerhalb des vorgegebenen Zeitrahmens und des Budgetrahmens erreicht. Dazu gehört die fachliche wie auch die formale und administrative Leitung des Projektes. Weitere Schwerpunkte dieses Arbeitspaketes sind die koordinierte Öffentlichkeitsarbeit für das Verbundprojekt sowie die Vernetzung mit anderen "Research Infrastructure" (RI) Initiativen und Projekten (z.B. D-Spin, eAqua, CLARIN, DARIAH, Bamboo etc.). Darüber hinaus fällt die Koordinierung der Arbeitsgruppen, die sich in TextGrid etabliert haben, in diesen Bereich.

Beschreibung des Arbeitspaketes

Die Gesamt-Leitung des Projekts dient neben der zielorientierten Führung des Verbundvorhabens vor allem der koordinierten Vermittlung von Ergebnissen in die Fachcommunities sowie in die D-Grid-Community. Darüber hinaus gehört die fachliche Gesamt-Koordination von Arbeitsgruppen, die Koordination und Verknüpfung von Ergebnissen dazu. Die Abstimmung der einzelnen Arbeitspakete miteinander und deren Verknüpfung untereinander ist eine weitere Kernaufgabe. Ein wichtiger Schwerpunkt ist die Vorbereitung und Durchführung von Kooperationsvorhaben: Die in „TextGrid - Vernetzte Forschungsumgebung in den eHumanities“ entwickelten Lösungen sollen größtmögliche

Interoperabilität zu Projekten und Initiativen aus dem nationalen und internationalen Umfeld aufweisen, um eine Vernetzung über Projekt-, Länder- und Staatsgrenzen hinweg zu ermöglichen. Dieser Aspekt spielt eine große Rolle in Bezug auf die Nachhaltigkeit von TextGrid und die Verbreiterung der Nutzerbasis.

Das Management des Projektes umfasst die administrative und organisatorische Durchführung des Projektes. Dazu gehören unter anderem die finanzielle Verwaltung der zentral eingesetzten Mittel, die Pflege und die Beantragung von Mitgliedschaften des Konsortiums in nationalen und internationalen Standardisierungsinitiativen und Gremien (z.B. TEI, Unicode, DCMI etc.). Die Sicherstellung der termingerechten Abgabe und Einhaltung von Meilensteinen und Ergebnisberichten liegt ebenfalls im Verantwortungsbereich dieses Arbeitspaketes. Das Projektmanagement gewährt den reibungslosen Ablauf der internen und externen Kommunikation. Dies beinhaltet auch die Kommunikation zum Projektträger und die damit verbundenen formalen und inhaltlichen Aspekte des Projektes sowie die Kommunikation mit dem neu einzuberufenden Fachbeirat. Das Management wird darüber hinaus für die formal-rechtliche Etablierung des Konsortiums und für die Aufnahme z.B. weiterer assoziierter Partner im Verlauf des Projektes und danach sorgen. Die Organisation und Durchführung der Meetings (Agenda, relevante Dokumente, Protokoll etc.) der Partner für die Projektbesprechungen werden von dem Projektmanagement geleistet, so z.B. regelmäßige „face-to-face“ Meetings sowie Video- und/oder Telefonkonferenzen.

Die koordinierte Öffentlichkeitsarbeit ist eine Kernaufgabe, die einen hohen Grad an Integration innerhalb der D-Grid-Community und einen hohen Grad an Akzeptanz bei der Zielgruppe, den Fach-Communities erwirkt. Eine den gesamten Projektverlauf umfassende PR-Strategie und die daraus folgenden PR-Aktivitäten stellen die Basis für die Öffentlichkeitsarbeit dar und unterstützen damit maßgeblich die Aktivitäten in AP 4 (Community Bildung) und den Aufbau eines nationalen eHumanities-Netzwerkes. Es ist vorgesehen, die PR-Aktivitäten des Projekts durch den Einsatz einer Agentur professionell zu unterstützen. Dabei stehen folgende Ziele im Fokus:

- kontinuierliche öffentlichkeitswirksame Darstellung des Projektes und seiner Ziele in den Medien
- öffentlichkeitswirksame Gestaltung und Verbreitung von Projektergebnissen
- Kampagnenarbeit
- Beratung bei der wirksamen Wegbereitung der Ergebnisse zu den definierten Zielgruppen (bei Bedarf)
- Unterstützung bei Vorbereitung, Durchführung und Nachbereitung von Pressekonferenzen

Die PR-Arbeit lässt sich sinnvoll in zielgruppenorientierte, multiplikatorenorientierte und gruppenübergreifende, ereignisgebundene Aktivitäten untergliedern. Folgende Aktivitäten sind vorgesehen: Zur Unterstützung der einzelnen Arbeitspakete und als zielgruppenorientierte Aktivität ist die Organisation und Durchführung von Workshops und Konferenzen geplant (vgl. AP 3, AP 4, AP 6):

- ein internationaler Experten-Workshop

- sechs Community Experten-Workshops
- zwei Strategie-Workshops
- sechs Schulungsveranstaltungen (AP 4)

Ereignisorientierte Aktivitäten machen die erreichten Meilensteine des Verbundvorhabens innerhalb und außerhalb der Zielgruppe sichtbar:

- Offizieller Start des erweiterten Web-Auftritts. Einladung an die Communities, sich aktiv an den vorgesehenen Entwicklungen zu beteiligen, Pressemeldung
- Bekanntmachen des Erreichten über Newsletter, Pressemitteilungen
- Auftritte auf Fachmessen und wichtigen Internationalen Veranstaltungen

Zur Unterstützung der o.g. Aktivitäten und Maßnahmen werden folgende begleitende Materialien erstellt und veröffentlicht:

- Printmedien:
 - Flyer
 - Poster
 - Info-Broschüren
- Elektronische Medien:
 - erweiterter Web-Auftritt
 - Newsletter
 - E-Mail (Pflege und Erweiterung der im TextGrid-Integrationsprojekt erstellten Verteiler)

Eine Reihe von Arbeitsgruppen (AGs) wird zur Behandlung AP-übergreifender Spezialthemen und zur AP- und institutionsübergreifenden Kommunikation eingerichtet. Die AGs haben dazu eigene Mailinglisten und führen regelmäßig Besprechungen (Telefon-/Videokonferenz oder persönliche Treffen) durch.

AG Architektur (SUB)

Diese AG leistet technische Konzeptions- und Koordinierungsarbeit, die TextGrid insgesamt betrifft. Hier werden die Gesamtarchitektur, Schnittstellen, Programmierkonventionen und ähnliche Themen konzeptionell vorbereitet. Damit diese Themen effektiv an alle Betroffenen kommuniziert werden können, sollte aus jeder beteiligten Institution wenigstens ein technisch orientierter Mitarbeiter (Informatiker, Programmierer) teilnehmen. Die AG dient auch der Koordination weiterer technisch orientierter (Sub-)AGs.

AG Fachwissenschaftler – (Uni Würzburg)

Diese Arbeitsgruppe koordiniert TextGrid-weit die Diskussionen zu Metadaten- und Datenformaten sowie Abläufen und Anforderungen an Tools. Teilnehmen werden

Fachwissenschaftler (wenigstens ein Vertreter pro Teildisziplin) sowie einzelne Vertreter der AG Architektur.

Weitere Sub-AGs

Bei Bedarf werden kurzfristig aus den o.A. AGs oder aus dem Plenum weitere Sub-AGs zur fokussierten Behandlung spezieller in der praktischen Arbeit auftretender Probleme gebildet.

Reports (R) & Milestones (M)

Bezeichnung	Monat	Beschreibung
R 8.1	3	PR-Konzept: Eine erste Version des PR-Konzepts mit einer Strategie zur Unterstützung des Verbundvorhabens in Bezug auf Bekanntmachung, Verbreitung etc. der Ziele und Ergebnisse des Projekts liegt vor. Dieses Konzept wird halbjährlich aktualisiert.
M 8.1	3	(Re-)Launch der Homepage: Die Homepage des Verbundvorhabens ist umfassend aktualisiert sowie erweitert und erlaubt eine kooperativ verteilte Eingabe und Pflege.
M 8.2	1 - 36	Organisation und Durchführung verschiedener Konferenzen und (Experten-) Workshops
R 8.2	1 - 6	PR Materialien: Flyer, Poster, PPT-Vorlagen etc. sind erstellt und liegen zur Nutzung bereit. Sie werden regelmäßig aktualisiert und an aktuelle Entwicklungen und Bedürfnisse angepasst.
R 8.3	6, 12, 18, 24, 30 und 36	Newsletter: Ein regelmäßig erscheinender Newsletter informiert in deutscher und englischer Sprache über laufende Aktivitäten und Ergebnisse des Verbundprojektes. Der Newsletter kann über den Web-Auftritt des Verbundprojektes abonniert werden und wird in relevanten Newsgroups, Communities und Newslettern beworben.

Wird ein wichtiger Milestone erreicht, so wird dies bekannt gegeben. E-Mail-Verteiler und Presseverteiler erlauben eine gezielte Ansprache und Information der jeweiligen Interessentengruppe.

Abhängigkeiten zu anderen APs

Die Abhängigkeiten zu anderen Arbeitspaketen sind grundsätzlich stark ausgeprägt, da das Erreichen wichtiger Milestones der Anlass für die Veröffentlichung von News oder Pressemeldungen ist.

II Ressourcenplanung

Im Folgenden wird eine Aufstellung der Gesamtpersonenmonate aller Projektpartner gegeben, wodurch abschließend nochmals deutlich wird, welche Partner in welchen Arbeitspaketen ihre Schwerpunkte haben.

Daran anschließend wird eine Gesamtübersicht über alle Milestones und Reports des Projektes gegeben. Hierbei wird in die beantragten drei Projektjahre differenziert um eine leichtere Übersicht und Einordnung der zu erledigenden Arbeiten zu ermöglichen.

Gesamtpersonenmonate aller Partner

Im Folgenden werden alle Personenmonate (PM) aller Projektpartner in einer Tabelle zusammen gefasst. Die Leitung des jeweiligen APs (nicht der UAPs) ist mit einem *L* markiert.

Partner (Akronym)	DAASI	FHW	IDS	LMU	MPDL	SUB	Trier	TUKL	Würz- burg	UPB	Gesamt
Gesamt:	36	36	42	18	28	103	114	18	90	27	512
AP 1 Rep						<i>L</i>					0
AP 1.1 WissGrid											0
AP 1.2 Fachwiss	18		6			35	6				65
AP 1.3 ESciDoc			1		4	1					6
AP 2 Lab							<i>L</i>				0
AP 2.1 Tools		8					36				44
AP 2.2 Migration									4		4
AP 3 Nachhaltig- keit			<i>L</i>								0
AP 3.1 Rechtsform			6								6

AP 3.2 Rechtliche Grundlagen	12		3								15
AP 4 Community							<i>L</i>				0
AP 4.1 Bedarf							6				6
AP 4.2 Coaching	6	6	6				24			6	48
AP 4.3 Netzwerk			1			1	4			1	7
AP 6 Fachwiss.									<i>L</i>		0
AP 6.1 Generalis.		22	2			25	18		57		124
AP 6.2 Lit-Phil							20		10		30
AP 6.3 Linguistik			16		12						28
AP 6.4 Kunst- geschichte					12						12
AP 6.5 Geschichte				18							18
AP 6.6 Musik										20	20
AP 6.7 freie Säule									18		18
AP 6.8 OCR								18			18
AP 8			1			41	<i>L</i>			1	43

Deliverables & Milestones (in zeitlicher Abfolge)

Im Folgenden werden alle Deliverables und Milestones zusammengefasst. Diese werden nach Jahren untergliedert dargestellt. Bei mehrfach auftretenden Deliverables und Milestones wurden diese auch mehrfach aufgeführt.

Deliverables & Milestones: Jahr 1

Bezeichnung	Monat	Beschreibung
M 6.0g	2	Entwickler-Koordinationsstreffen Ein Treffen (gemeinsam mit AP 4), bei dem mit allen Entwicklern, insbesondere auch der neuen Partner bzw. der Neueinstellungen, über den Stand von TextGrid und die Planungen für die neue Projektphase gesprochen wird, außerdem grundlegende Vereinbarungen für Entwicklungsprozesse etc. Näher beschrieben in AP 6.1
M 6.1.0	2	Planung und Durchführung eines internen Entwickler-Koordinationsworkshops am Beginn des Projekts, erste gemeinsame Leitlinien, Planung für die Koordination
R 8.1	3	PR-Konzept: Eine erste Version des PR-Konzepts mit einer Strategie zur Unterstützung des Verbundvorhabens in Bezug auf Bekanntmachung, Verbreitung etc. der Ziele und Ergebnisse des Projekts liegt vor. Dieses Konzept wird halbjährlich aktualisiert.
M 8.1	3	(Re-)Launch der Homepage: Die Homepage des Verbundvorhabens ist umfassend aktualisiert sowie erweitert und erlaubt eine kooperativ verteilte Eingabe und Pflege.
R 8.2	6	PR Materialien: Flyer, Poster, PPT-Vorlagen etc. sind erstellt und liegen zur Nutzung bereit. Sie werden regelmäßig aktualisiert und an aktuelle Entwicklungen und Bedürfnisse angepasst.
S 1.1	6	Hands-on-Meeting Content-Provider
S 1.2	6	Hands-on-Meeting Fachwissenschaftler
S 1.3	6	Hands-on-Meeting Entwickler
M 4.1.1	6	Erstellung bzw. Einholung und laufende Aktualisierung von Nutzerszenarien und Nutzerfeedback; Vermittlung in die Entwicklercommunity in geeigneter Form
M 6.1.1	6	Release einer ersten integrierten Version von TextGrid-Tools des APs. Dies umfasst die Koordination des gemeinsamen Milestones M 6.1g sowie die Integration von Weiterentwicklungen auf der Basis von Erkenntnissen aus TextGrid-Summit und Entwicklerworkshop im Januar 2009, sowie dem Tutorial im Rahmen der IEEE DEST
M 6.2.1	6	Erster Integrationsschritt mit den anderen UAPs
M 6.3.1	6	Webservice-Integration (Funktionsebene): Föderierte Suche über COSMAS -II bzw. OWID in TextGrid Requirements-Analyse und Spezifikation der Schnittstellen zu LEXUS und ANNEX/IMEX
M 6.4.1	6	Definition des Architekturmodells zur Anbindung von Digilib Definition von Repräsentationsformaten zur Darstellung von TEI-Dokumenten innerhalb von strukturellen Editionen
M 6.5.1	6	Erarbeitung eines adäquaten Auszeichnungssystems für glossierte Texte und Integration in TEI (Text Encoding Initiative)
M 6.6.1	6	Erste TextGrid-Integration und Dokumentation: Integration der erstellten musikalischer Sonderzeichen in den XML-Editor (vgl. AP 6.6.1); Dokumentation des neuen Tools (vgl. AP 4)

M 6.8.1	6	Erste Integration Das vom TUKL/DFKI entwickelte Open-Source System OC-Ropus wird in die TextGrid Architektur integriert und so allen Partnern verfügbar gemacht. Gleichzeitig werden eine separate Webschnittstelle und Webservices entwickelt, die die Benutzung des OCR Systems ermöglichen.
M 6.1g	6	erste AP 6-übergreifende Integration Zu diesem Milestone sollte jeder Partner / jedes UAP eine Grundversion eines ersten Tools vollständig in die TextGrid-Architektur integrieren. Ziel ist es, frühzeitig für eine Integration zu sorgen und zu verhindern, dass Insellösungen entstehen, die sich später nur schwer integrieren lassen. Die künftige Entwicklung sollte auf dieser Integration aufbauen, es sollte also möglichst weiter gemeinsam in einer integrierten Lösung entwickelt werden.
M 4.2.1	6 (danach laufende Aktualisierung)	Dokumentation TextGrid, Erstellung von Tutorials (laufende Erweiterung unter Berücksichtigung des sich aus der Community artikulierenden Bedarfs).
R 8.3	6, 12, 18, 24, 30 und 36	Newsletter: Ein regelmäßig erscheinender Newsletter informiert in deutscher und englischer Sprache über laufende Aktivitäten und Ergebnisse des Verbundprojektes. Der Newsletter kann über den Web-Auftritt des Verbundprojektes abonniert werden und wird in relevanten Newsgroups, Communities und Newslettern beworben.
M 1.1.1	9	WissGrid Workshop mit Vorstellung der generischen Architektur des WissGrid Forschungsdatenrepositoriums
S 2.1	12	Hands-on-Meeting Content-Provider
S 2.2	12	Hands-on-Meeting Fachwissenschaftler
S 2.3	12	Hands-on-Meeting Entwickler
R 4.3.1	12	Bericht zu eHumanities-Themen mit Relevanz für die allgemeine Informatik.
R 6.0.1	12	Report über die Entwicklung des vergangenen Jahres sowie den Stand der Planungen (Pflichtenhefte). Die Berichte über die entwickelten Tools sollten dabei als Endnutzerdokumentation geschrieben (und in AP 4 eingebracht) werden.

Deliverables & Milestones: Jahr 2

Bezeichnung	Monat	Beschreibung
R 1.3.1	14	Aufbau einer nationalen föderierten Repositorien-Infrastruktur für die Geisteswissenschaften (Visionspapier)
R 1.2.1	16	Roadmap zur Integration Grid/Repository; mit Prototypen für verteilte Suche und das Projekt-/Sammlungskonzept
S 3	18	TextGrid-weiter Workshop Content-Provider, Fachwissenschaftler, Entwickler
M 2.1.1	18	TextGridLab 1.5.

R 8.3	6, 12, 18, 24, 30 und 36	Newsletter: Ein regelmäßig erscheinender Newsletter informiert in deutscher und englischer Sprache über laufende Aktivitäten und Ergebnisse des Verbundprojektes. Der Newsletter kann über den Web-Auftritt des Verbundprojektes abonniert werden und wird in relevanten Newsgroups, Communities und Newslettern beworben.
M 6.1.2	18	Koordination des gemeinsamen Zwischenreleases »TextGridLab 1.5« (vgl. M 6.2g unten) incl. weiterer Desiderate aus <ul style="list-style-type: none"> ▪ TextGrid-Summit und Entwicklerworkshop ▪ dem IEEE DEST Tutorial ▪ der ersten Runde der Bedarfserhebung und Usability-Evaluation, vgl. AP 4 ▪ dem Feedback aus dem Faust-Editionsprojekt
M 6.3.2	18	Fertigstellung der Webservice-Integration auf Metadaten-ebene Fertigstellung des LEXUS-Interface 1. Version der Konvertierungsroutinen
M 6.4.2	18	Unterstützung von zunächst 2 Repräsentationsformaten für strukturelle Editionen Fertigstellung der ersten integrierten Version von Digilib. Erhebung des Anwenderfeedbacks
M 6.5.2	18	Weiterentwicklung des Text-Bild-Linkeditors (Image-Mapping) für die Anforderungen glossierter Texte; Erstellung des Präsentationstools;
M 6.6.2	18	Anzeigetool: Integration des aktuellen Zwischenstands der Entwicklung des Noten-Anzeigetools in die TextGrid-Infrastruktur (vgl. AP 6.6.3)
M 6.8.2	18	Release 1.5. Zum Zwischenrelease nach der Hälfte der Projektlaufzeit wird die Erkennung von gebrochener Schrift zur Verfügung stehen inklusive einer Reihe von Varianten und eine Schätzung der Fontgröße des erkannten Textes. Im Dialog mit den Projektpartnern werden weitere Elemente für die Ausgabe identifiziert.
M 6.2g	18	Release 1.5 Nach einer konzentrierten Entwicklungsphase mit UAP-spezifischer Detailplanung sollte hier eine Synchronisation auf ein gemeinsames stabiles Zwischenrelease erfolgen.
R 3.1.1	18	Draft-Version Organisationsmodell inkl. Rollen, Rechten und Nutzungsversprechen
R 4.2.1	18	Bericht zu notwendigen Veränderungen und Ergänzungen bei den Issue-Trackern etc. für Entwickler.
R 8.3	6, 12, 18, 24, 30 und 36	Newsletter: Ein regelmäßig erscheinender Newsletter informiert in deutscher und englischer Sprache über laufende Aktivitäten und Ergebnisse des Verbundprojektes. Der Newsletter kann über den Web-Auftritt des Verbundprojektes abonniert werden und wird in relevanten Newsgroups, Communities und Newslettern beworben.
S 4.1	24	Hands-on-Meeting Content-Provider

S 4.2	24	Hands-on-Meeting Fachwissenschaftler
S 4.3	24	Hands-on-Meeting Entwickler
R 6.0.2	24	Report über die Entwicklung des vergangenen Jahres sowie den Stand der Planungen (Pflichtenhefte). Die Berichte über die entwickelten Tools sollten dabei als Endnutzerdokumentation geschrieben (und in AP 4 eingebracht) werden.
R 3.2.1	24	Musterverträge (AP 3.2) und technische Umsetzung

Deliverables & Milestones: Jahr 3

Bezeichnung	Monat	Beschreibung
R 1.3.2	28	Wege zur Verknüpfung eSciDoc /TextGrid (erstes Konzept)
R 8.3	6, 12, 18, 24, 30 und 36	Newsletter: Ein regelmäßig erscheinender Newsletter informiert in deutscher und englischer Sprache über laufende Aktivitäten und Ergebnisse des Verbundprojektes. Der Newsletter kann über den Web-Auftritt des Verbundprojektes abonniert werden und wird in relevanten Newsgroups, Communities und Newslettern beworben.
S 5	30	TextGrid-weiter Workshop Content-Provider, Fachwissenschaftler, Entwickler
M 2.1.2	30	TextGridLab 1.9.
M 2.2.1	30	Evaluierung von technologischen Möglichkeiten zur Portierung
M 6.1.3	30	Release 1.9 Aus AP 6.1 enthalten sein werden <ul style="list-style-type: none"> ▪ Feedback aus dem Projekt <i>Digitalisierung eines Briefkorpus (Staatsbibliothek Berlin)</i> ▪ Feedback aus der weiteren Bedarfserhebung und Usability-Evaluation (vgl. AP 4) ▪ Ergebnisse weiterer Anwenderworkshops (vgl. AP 4) Weiterentwicklungen der o.g. Tools hin zu Produktionsreife
M 6.2.2	30	Fertigstellung des Linkeditors Text-Bild und des Kollationsierungsprogramms
M 6.3.3	30	Fertigstellung der gesamten Webservice-Integration für COSMAS II und OWID Fertigstellung des ANNEX/IMEX-Interface
M 6.4.3	30	Zweite Version der Anbindung von Digilib unter Einbeziehung des Anwenderfeedbacks und mit Bereitstellung der erweiterten Funktionalitäten (Messfunktionen, Metadatenintegration) Integration der Repräsentationsformate für eine wissenschaftliche strukturelle Edition Revision der erstellten strukturellen Edition
M 6.5.3	30	Fertigstellung des Editionstools für glossierte Texte
M 6.6.3	30	Anzeige- und Eingabetool: Fertigstellung des Anzeige- und Eingabetools (vgl. AP 6.6.3)

M 6.8.3	30	Release 1.9 Sechs Monate vor Projektende sind alle im Vorfeld definierten Funktionen verfügbar. Insbesondere erlaubt die Frakturerkennung die Verarbeitung von Texten mit mehreren Schriftarten und hält entsprechende Informationen im Ausgabeformat vor.
M 6.3g	30	Release 1.9 Sechs Monate vor Projektende sollten zumindest die Grundgerüste aller Tools stehen und zu einem Release vereinigt werden, sodass die restliche Projektlaufzeit für Tests, die Einarbeitung/Berücksichtigung der Testergebnisse und des Feedbacks aus der Community, die weitere Abstimmung mit anderen Komponenten, die Fertigstellung der Dokumentation und die Publikation als Teil der TextGrid-Open-Source-Software zur Verfügung steht.
R 3.1.2	30	Finales Organisationsmodell und Satzung/Statuten der zu gründenden TextGrid-Organisation
S 6	36	TextGrid Summit
M 1.1.2	36	Integration TextGrid / Forschungsdatenrepositorium WissGrid
M 2.1.3	36	TextGridLab 2.0.
M 3.1.2	36	Implementierung des überarbeiteten WissGrid-Betriebsmodells. Gründung einer Organisation zur nachhaltigen Umsetzung des Betriebsmodells.
M 6.1.4	36	produktionsreifes Abschlussrelease: TextGrid 2.0, das die eingegangenen Desiderate und die Anforderungen umsetzt. Mit AP 4 / S 4.3: Abschlussworkshops für Anwender und Entwickler.
M 6.2.3	36	Release 2.0: Einarbeitung von Feedback in beide Tools, Produktionsreife
M 6.4.4	36	Vollständige Integration der entwickelten Tools in die TextGrid-Plattform und Schulungen für die Anwendenden
M 6.5.4	36	Einarbeitung von Feedback in Präsentations- und Editions-tool; Produktionsreife; Integration in die TextGrid-Plattform (XML-Editor)
M 6.8.4	36	Release 2.0: Während der finalen Integration und Evaluation werden Stabilisierung und Dokumentation der Software im Vordergrund stehen.
M 6.4g	36	Release 2.0 Abschlussrelease mit allen Entwicklungen.
R 8.3	6, 12, 18, 24, 30 und 36	Newsletter: Ein regelmäßig erscheinender Newsletter informiert in deutscher und englischer Sprache über laufende Aktivitäten und Ergebnisse des Verbundprojektes. Der Newsletter kann über den Web-Auftritt des Verbundprojektes abonniert werden und wird in relevanten Newsgroups, Communities und Newslettern beworben.
R 6.0.3	36	Report über die Entwicklung des vergangenen Jahres sowie den Stand der Planungen (Pflichtenhefte). Die Berichte über die entwickelten Tools sollten dabei als Endnutzerdokumentation geschrieben (und in AP 4 eingebracht) werden.
R 1.2.2	36	TextGridRep: Manual zum TextGrid Forschungsdatenarchiv (Architektur, Installation, Erweiterbarkeit)

Permanente Arbeiten:

Laufend (AP 6)	Während der gesamten Projektlaufzeit werden eingehende Rückmeldungen zu Fehlern, Unzulänglichkeiten und neuen Anforderungen in die Software integriert
Laufend (AP 8)	M 8.2: Organisation und Durchführung verschiedener Konferenzen und (Experten-) Workshops

III WissGrid Antrag

Der Antrag des Projektes „WissGrid“ wurde eingereicht und positiv begutachtet. Das Projekt wird zum 1. Mai 2009 bewilligt.

IV „Digitale Bibliothek“

Begründung für den Erwerb der Texte der Digitalen Bibliothek

Die Digitale Bibliothek stellt eine umfangreiche Sammlung von Texten vom Anfang des Buchdrucks bis zu den ersten Jahrzehnten des 20. Jahrhunderts in digitaler Form zur Verfügung. Für die Germanistische Literaturwissenschaft ist die Sammlung von besonderem Interesse, da sie nahezu alle wichtigen kanonisierten Texte und zahlreiche weitere literarhistorisch relevante Texte enthält. Ähnliches gilt für die Philosophie und die Kulturwissenschaften insgesamt. Die Texte stammen zum größten Teil aus Studienausgaben und sind daher zitierfähig; das gilt auch für den Rest, der weitgehend auf die Digitalisierung von Erstdrucken zurückgeht.

Die Texte waren lange Zeit nur als CDs verfügbar, sind jedoch zurzeit über das Portal Zeno.org zugänglich. Dort stehen sie unter einer Lizenz, die der Verlag als 'gemeinfrei' bezeichnet. Diese Lizenz erlaubt die freie Verwendung der Werke: "Dieser Inhalt darf *als Einzelwerk oder Werkbestandteil* – auch zu kommerziellen und gewerblichen Zwecken - kopiert, verbreitet, öffentlich wiedergegeben und Dritten zugänglich gemacht werden."²⁰ Allerdings behält sich der Verlag die Gesamtpublikation, auch auszugsweise, ausdrücklich vor: "*Die Übernahme der Gesamtheit der verfügbaren Inhalte wie auch von wesentlichen Teilen in eine andere Datenbank ist nicht gestattet.*"²¹

²⁰ <http://www.zeno.org/Zeno/-/Lizenz%3A+Gemeinfrei>

²¹ Ebda. Hervorhebungen im Original.

Bisherige Nachnutzungsmöglichkeiten

Es stellt sich die Frage, ob die auf Zeno.org angebotenen Texte unter den bestehenden juristischen und technischen Bedingungen für eine wissenschaftliche Benutzung geeignet sind.

Dazu muss erst einmal geklärt sein, was 'wissenschaftliche Benutzung' heute genau heißt. Drei Formen der Verwendung sind unter dieser Bezeichnung zusammengefasst:

- 1) Lesen,
- 2) Suchen,
- 3) Weiterverwenden.

1) Lesen

Im Fall von Forschungsliteratur ist die wichtigste Form der Nutzung sicherlich die Lektüre, sei es am Bildschirm oder auf dem Papier. Das gilt aber nicht für die kanonischen Texte, von denen hier die Rede ist, da viele in wohlfeilen Ausgaben für eine bequeme Offline-Lektüre zur Verfügung stehen. Etwas anders stellt sich die Situation bei den Wörterbüchern und Lexika da, die auf diese Weise bequem zur Hand sind. Insgesamt aber ist Lesen sicherlich nicht die wichtigste Form der Verwendung dieser Texte.

2) Suchen

Suchen ist für die meisten Geisteswissenschaftler sicherlich die wichtigste Form des Zugriffs auf die Texte. Leider bleibt Zeno.org hierbei weit hinter den Möglichkeiten zurück,²² die die CD-ROM-Editionen bei DirectMedia aufweisen - und auch diese sind nicht wirklich zufriedenstellend.²³ Das größte Problem ist das Fehlen einer Segmentierung der Daten nach relevanten Kategorien. Zeno.org erlaubt es nicht, die Suche auf einzelne Werke, Autoren,²⁴ bestimmte Zeitabschnitte oder spezifische Gattungen einzuschränken. Das macht die Suche für viele Zwecke nur umständlich oder gar nicht brauchbar. Die Datenaufbereitung der CD-ROMs macht deutlich, dass in den Rohdaten viele Metadaten enthalten sind, die sich auf der Website gar nicht finden oder nur so, dass sie für die Suche nicht verwendet werden können. Hier gibt es also einen deutlichen Unterschied zwischen einer populären Verwendung dieser Texte, für die Zeno.org durchaus ausreichend ist, und einer wissenschaftlichen Verwendung, die einen genaueren Zugriff auf die verschiedenen Metadaten voraussetzt.

3) Weiterverwenden

Das Weiterverwenden der Daten in eigenen Forschungskontexten wird in den kommenden Jahren eine selbstverständliche Verwendung sein. In den letzten Jahren sind weltweit

²² Beschreibung der Suchfunktionen in Zeno.org: <http://wiki.zeno.org/Zeno:Suchfunktion>.

²³ Eine Beschreibung der Suchfunktionen in den DirectMedia-Editionen findet sich hier: <http://computerphilologie.uni-muenchen.de/jahrbuch/jb1/jannidis-2.html>

²⁴ Die CD-ROMs erlauben es zumindest, die Suche auf spezifische Autoren und einzelne Werke einzuschränken.

zahlreiche E-Humanities-Zentren²⁵ entstanden, die inzwischen zunehmend miteinander kommunizieren und eine Infrastruktur von Diensten anbieten, deren Ziel es ist, das Erstellen und Verwenden digitaler Korpora möglichst zu vereinfachen.²⁶ Schon jetzt kann ein Forscher einen Text in diesem Netzwerk statistisch auswerten, andere Dienste sind in Planung oder befinden sich bereits in der Betaphase. In diesem Kontext sind u.a. folgende Formen der Weiterverwendung zu erwarten:

- Zusammenstellung eigener Korpora für quantitative Analysen.
- Integration spezifischer Werke oder Werkblöcke in eigene Editionen oder Textverbünde, z.B. in ein Wörterbuch-Netzwerk.
- Verwendung von Texten als Grundlage einer eigenen Edition, z.B. Erstellung einer Studienausgabe durch Kommentierung von Texten.
- Kollationierung mit anderen Textzeugen / Ausgaben.
- Weitere Erschließung der vorliegenden Texte durch Eintragen von Metadaten, z.B. Geschlecht, Konfession oder soziale Herkunft der Autoren.
- Eintrag von forschungsspezifischem Markup, z.B. narratologische Auszeichnung von Erzähltexten.
- Integration aller Daten in möglichst umfassende Sammlungen, um alle Belegstellen zu finden.

Nur für die wenigsten dieser Verwendungsweisen ist die augenblickliche Situation, also das Angebot der Texte auf Zeno.org, befriedigend: die Texte sind zerstückelt und ihre Metadaten sind nicht trivial erschließbar, wenn überhaupt greifbar. Allerdings enthalten die Texte in der ursprünglichen Auszeichnung, die auf den CD-ROMs sichtbar ist und um die es bei dem Kaufangebot geht, diese und weitere Metadaten.

Typische Interessenten an diesen Texten sind:

- Historische Linguisten, die diachrone Korpora aufbauen
- Editoren digitaler Editionen (Literaturwissenschaftler, Philosophen, Historiker, Musikwissenschaftler usw.), die auf diese Weise wichtige Kontexte für ihre editierten Texte bereitstellen und ihre Texte damit verlinken können.
- Textwissenschaftler, die mit quantitativen Verfahren Texte untersuchen wollen.
- Fachwissenschaftler, die nur in einem oder wenigen Texten intensiv suchen wollen, diese aber dafür vorher noch genauer aufbereiten möchten, z.B. durch die Auszeichnung von Namen und Daten.
- Alle Wissenschaftler, die größere Textmengen zur Entwicklung von Texttechnologien benötigen

²⁵ Umfassend, wenn sicherlich noch nicht vollständig:
<http://digitalhumanities.pbwiki.com/centerNet+Digital+Humanities+Centers+>

²⁶ Mustergültig ist hier etwa TAPOR <http://tapor.ualberta.ca/>.

Fachwissenschaftliche Anwendungsszenarien

Im Folgenden ist eine Reihe von Anwendungen für die "Digitale Bibliothek" skizziert, die sich aus Diskussionen mit Fachwissenschaftlern entwickelt haben:

1) Die Wörterbücher in der Textsammlung könnten umgehend in das vorhandene Wörterbuchnetz eingebunden werden. Mit den Lexika und Enzyklopädien würde der Anfang eines vergleichbaren Netzes für Nachschlagewerke gelegt werden. Beides kann dann wiederum als Service zur Verfügung gestellt werden - wie es für Wörterbücher ja schon der Fall ist -, so dass Editoren in TextGrid Verweise einschließlich von Links auf Einträge in den Wörterbüchern und Enzyklopädien sehr leicht in ihren Kommentar einfügen könnten.

2) Aufbau von Textsammlungen für spezifische Zwecke.

Alle Texte würden über persistente URLs zur Verfügung gestellt werden, so dass dann stabile Verweise auf einen Großteil der kanonisierten Literatur und deutlich darüber hinaus vor dem 20. Jahrhundert für alle in TextGrid verfügbaren Korpora möglich wären. Viele Editoren scheuen wg. der Instabilität der Links z.Zt. vor solchen Verweisen noch zurück. Editoren könnten dann in TextGrid ohne großen Mehraufwand Kontexte für ihre Texte anbieten und das sehr viel schneller als früher. Hier ein Beispiel (noch aus der Zeit vor TextGrid - aber ein Import nach TextGrid ist in Arbeit), wie eine solche Kontextualisierung dann aussehen könnte:

<http://linglit193.linglit.tu-darmstadt.de:8080/exist/jgoethe/edition.xql?c=/db/jgoethe>

Bei dieser Edition des Jungen Goethe mussten die Herausgeber noch alle Kontexte, z.B. das mythologische Wörterbuch oder die Texte von Goethes Vorbildern, selbst digitalisieren.

Ähnliche Bedürfnisse haben Sprachwissenschaftler bei der Zusammenstellung ihrer Korpora. Textwissenschaftler würden also deutlich schneller solche Editionen und Korpora abschließen bzw. sehr viel mehr Material integrieren können. Schon diese Möglichkeit würde die Attraktivität von TextGrid als Arbeitsumgebung für alle Textwissenschaftler außerordentlich steigern.

3) Quantitative Analysen.

Es gibt erste Kontakte zwischen TextGrid und Projekten wie philomine, eAqua oder Monk, die mit Textmining-Verfahren auf großen Korpora arbeiten. Geplant ist eine Schnittstelle von TextGrid zu mindestens einem solchen Projekt, das die Software für philologisches Textmining entwickelt, um solche innovativen Strategien der nicht-hermeneutischen Textforschung zu ermöglichen. Hier ist ein Link zu einer Diskussion solcher Möglichkeiten; mit dem dort erwähnten Martin Mueller stehen wir wg. einer Kooperation in Kontakt:

<http://www3.isrl.uiuc.edu/~unsworth/hownot2read.html>

All das haben wir nicht beschleunigt vorangetrieben, da wir ja erst im Laufe der kommenden Jahre mit einer nennenswerten Menge von Texten rechneten. Das würde sich durch den Kauf der digitalen Bibliothek natürlich schlagartig ändern. Textwissenschaftler könnten dann in TextGrid ihre Texte für solche Untersuchungen zusammenstellen und vorbereiten (etwa

mit dem TextGrid-Tokenizer, -Lemmatizer und dem Streaming-Editor), um sie dann von den spezialisierten Werkzeugen auswerten zu lassen und die Ergebnisse in TextGrid weiterzuverarbeiten.

Idealerweise können bald einige Verfahren zur statistischen Textauswertung direkt als Service an TextGrid angegebunden werden, z.B.

die offenen Services von Tapor <http://portal.tapor.ca/portal/portal>

4) Besseres OCR

Die Erkennungsqualität des vom TUKL/DFKI in TGII entwickelten OCR-Services steigt, je mehr Text ihm für statistische Auswertungen zur Verfügung steht. Unserer Einschätzung nach ist dieser Service für viele Textwissenschaftler von größtem Interesse und seine Qualitätssteigerung ebenso.

Insgesamt wäre neben den skizzierten technischen Möglichkeiten, die sich durch den Ankauf der digitalen Bibliothek auftun, der größte Gewinn, Textwissenschaftler sehr viel schneller und von Anfang an in deutlich größerer Zahl für TextGrid interessieren zu können; es wäre also eine der wirksamsten Maßnahmen für das fachwissenschaftliche Communitybuilding.

Vereinbarte Lizenzbedingungen mit TextGrid

Das Lizenzmodell wird zur Zeit zum größtmöglichen Nutzen der Öffentlichkeit von Rechtsexperten ausgearbeitet.

V Abkürzungsverzeichnis:

AAI	Authentifizierungs- und Autorisierungs-Infrastruktur
AG	Arbeitsgruppe
ANNEX/IMEX	ANNotation EXploration tool for web-based exploration of annotated corpus resources (incl. video, audio and text)
AP	Arbeitspaket
APIs	Application Programming Interface
CEN/ISSS	Comité Européen de Normalisation / Information Society Standardization System
COSMAS II	Corpus Search, Management and Analysis System
DCMI	Dublin Core Metadata Initiative
DFG	Deutsche Forschungsgemeinschaft
DFKI	Deutsches Forschungszentrum für Künstliche Intelligenz
DGI	D-Grid Integrationsprojekt
DGSE	Deutsche Grid Support Einrichtung
dMGH	digitale Monumenta Germaniae Historica
EDV	Elektronische Datenverarbeitung
FaWi	Fachwissenschaftler
GSI	Grid Security Infrastructure
GUI	Graphical User Interface
HRK	Hochschulrektorenkonferenz
IEEE	Institute of Electrical and Electronics Engineers, Inc.
IEEE DEST	International Conference on Digital Ecosystems and Business Intelligence
LZA	Langzeitarchivierung
NEH/DFG	National Endowment for the Humanities (NEH) / Deutsche Forschungsgemeinschaft (DFG)
OAI-PMH	Open Archives Initiative - Protocol for Metadata Harvesting
OCR	Optical Character Recognition

OGF23	23rd Open Grid Forum
ORA	Oxford Research Archive
OTA	Oxford Text Archive
OWID	Online-Wortschatz-Informationssystem Deutsch
PKI	Public Key Infrastruktur
PM	Personenmonate
PPT	Power-Point
RBAC	Role Based Access Control
REST	REpresentational State Transfer Architektur
SIGs	Special Interest Groups
SOAP	Simple Object Access Protocol
SRU	Search/Retrieve via URL
SRW	Search/Retrieve for the Web
SVN	SubVersionN (Open-Source-Software zur Versionsverwaltung)
TEI	Text Encoding Initiative
TV-L	Tarifvertrag der Länder
UAP	Unterarbeitspaket
ViRR	Virtueller Raum Reichsrecht
VO- Management	Virtual Organization Management
XACML	eXtensible Access Control Markup Language
XML	Extensible Markup Language