



Bundesministerium
für Bildung
und Forschung



TextGrid: Institutionalisierung einer Virtuellen Forschungsumgebung in den Geisteswissenschaften

Konsortialleitung:

Prof. Dr. Norbert Lossau
Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB)
Platz der Göttinger Sieben 1
37073 Göttingen
Tel.: (0551) 39-5212, E-Mail: lossau@sub.uni-goettingen.de

Laufzeit: 3 Jahre (1. Juni 2012 bis 31. Mai 2015)

- Öffentliche Version -

Verbundpartner:

- Berlin-Brandenburgische Akademie der Wissenschaften (BBAW), Berlin
- DAASI International GmbH, Tübingen
- Fachhochschule Worms, Fakultät für Informatik und Telekommunikation
- Georg-August-Universität Göttingen, Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB) (Projektleitung)
- Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG)
- Institut für Deutsche Sprache (IDS), Mannheim
- Julius-Maximilians-Universität Würzburg, Institut für Deutsche Philologie
- Max-Planck-Institut für Wissenschaftsgeschichte (MPI WG), Berlin
- Technische Universität Berlin, Zentrum Technik und Gesellschaft (ZTG)
- Technische Universität Darmstadt, Institut für Sprach- und Literaturwissenschaft

Inhaltsverzeichnis

Verzeichnis der Akronyme	4
1. Ziele	6
2. Stand der Wissenschaft und Technik; bisherige Arbeiten	7
2.1 Stand der Forschung	7
2.2 Bisherige Arbeiten des TextGrid-Konsortiums	10
2.2.1 Nutzer und Nutzungsmodelle	11
2.2.2 Etablierung eines nachhaltigen Betriebs- und Geschäftsmodells.....	13
3. Ausführliche Beschreibung des Arbeitsplans	15
4. Verwertungsplan	17
5. Kooperation / Zusammenarbeit mit Dritten	18
7. Beschreibung der Arbeitspakete	20
AP 1: Betriebsmodell	20
AP 1.1: Rechtsform, Kostenkategorien	21
AP 1.2: Verein, Geschäftsstelle.....	22
AP 1.3: (Technische) Integration und zukünftiges Weiterentwicklungspotential, Zukunftsfähigkeit	22
AP 2: Prozessorganisation	23
AP 2.1: Umgang mit großen Nutzerzahlen und heterogenen Nutzergruppen sowie Nutzerverwaltung (Praxis und Theorie)	24
AP 2.2: Verteilte / Zentrale Entwicklungen	25
AP 2.3: Finanzierungsmodelle und -wege	26
AP 3: Nutzer-Kommunikation	27
AP 3.1: Dokumentation, Nutzer-Information	28
AP 3.2: eHumanities und TextGrid in universitären Curricula.....	28
AP 3.3: Workshops, Schulungen, Nutzer-Interaktion	29
AP 4: Betrieb und Softwarepflege Repository	30
AP 4.1: Humanities Data Centre	31
AP 4.2: Performanz, Skalierbarkeit, Lasttests.....	31
AP 4.3: Qualitätsmanagement	33
AP 4.4: Ingest Datensammlung, Publikation, Präsentation	34
AP 4.5: Technischer Regelbetrieb, Architektur, innovatives Evolutionspotential.....	35

AP 5: Betrieb und Softwarepflege Laboratory	36
AP 5.1: Sicherheit, Bugfixing, Technische Dokumentation.....	37
AP 5.2: Funktionale Tests	38
AP 5.3: Usability	38
AP 5.4: Koordination von Weiterentwicklungen und Satellitenprojekten	39
AP 5.5: Weiterentwicklung der Digitalen Bibliothek	40
AP 6: Integriertes Monitoring und Qualitätssicherung	41
AP 6.1: Integriertes Monitoring aller Arbeitspakete.....	42
AP 6.2: Nutzerakzeptanz (inklusive Befragungen).....	42
AP 7: Projektmanagement und Öffentlichkeitsarbeit.....	43
AP 7.1: Projektmanagement i.e.S. und Koordinierung potentieller Kooperationen	44
AP 7.2: Abstimmung mit Fachgesellschaften.....	44
AP 7.3: Marketing und Öffentlichkeitsarbeit	45
8. Verbundpartner-Beschreibungen	46

Verzeichnis der Akronyme

AdWG	Akademie der Wissenschaften zu Göttingen
AP	Arbeitspaket
API	Application Programming Interface
BBAW	Berlin-Brandenburgische Akademie der Wissenschaften, Berlin
BMBF	Bundesministerium für Bildung und Forschung, Berlin
C3-INAD	Infrastructure for General Access to Climate Data
CEN/ISSS	European Committee for Standardisation / Information Society Standardisation System
CLARIN	Common Language Resources and Technology Infrastructure
DAASI	Directory Applications for Advanced Security and Management, Tübingen
DARIAH	Digital Research Infrastructure for the Arts and Humanities
DFG	Deutsche Forschungsgemeinschaft, Bonn
DFN	Deutsches Forschungsnetz e.V., Berlin
DIW	Deutsches Institut für Wissenschaftsforschung, Berlin
D-SPIN	Deutsche Sprachressourcen-Infrastruktur
eAQUA	Extraktion von strukturiertem Wissen aus antiken Quellen für die Altertumswissenschaft
ERIC	European Research Infrastructure Consortium
ESFRI	European Strategy Forum on Research Infrastructures
EU	Europäische Union
GWDG	Gesellschaft für wissenschaftliche Datenverarbeitung mbH, Göttingen
GWK	Gemeinsame Wissenschaftskonferenz
HDC	Humanities Data Centre
IDS	Institut für Deutsche Sprache, Mannheim
IEEE	Institute of Electrical and Electronics Engineers
JISC	Joint Information Systems Committee, Bristol
KoZe	Kompetenzzentrum für elektronische Erschließungs- und Publikations- verfahren in den Geisteswissenschaften an der Universität Trier
LIS	Library and Information Science
LOEWE	Landes-Offensive zur Entwicklung wissenschaftlich-ökonomischer Exzellenz
MPG	Max-Planck-Gesellschaft, Berlin
MPI	Max-Planck-Institut
MPI MMG	Max-Planck-Institut zur Erforschung multireligiöser und multiethnischer Gesellschaften
MPI WG	Max-Planck-Institut für Wissenschaftsgeschichte, Berlin
ÖAW	Österreichische Akademie der Wissenschaften, Wien
PID	Persistent Identifier
PKI	Public Key Infrastructure
SADE	Scalable Architecture for Digital Editions, Berlin
SOEB	Berichterstattung zur sozioökonomischen Entwicklung in Deutschland
SOEP	Socio-Economic Panel, Berlin
SUB	Staats- und Universitätsbibliothek, Göttingen
TEI	Text Encoding Initiative
TMF	Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V.
TU	Technische Universität
VRE	Virtual Research Environment / Virtuelle Forschungsumgebung

WR
ZIB
ZTG

Wissenschaftsrat, Köln
Zuse-Institut Berlin
Zentrum Technik und Gesellschaft, Berlin

1. Ziele

Ziel ist es, in der beantragten Projektlaufzeit (01. Juni 2012 bis 31. Mai 2015) unter den Bedingungen des Echtbetriebs, der mit der Freischaltung der TextGrid-Version 1.0 im Juli 2011 gestartet ist, die entwickelten Strukturen, Verfahren und Instrumente einschließlich der verschiedenen Finanzierungsoptionen operativ zu erproben, den erweiterten und steigenden fachwissenschaftlichen Anforderungen gemäß konzeptionell zu optimieren und damit die Voraussetzungen dafür zu schaffen, dass TextGrid am Ende der dreijährigen Übergangsphase den institutionell verankerten und finanziell abgesicherten Dauerbetrieb aufnehmen kann.

Demgemäß gilt es jetzt, nicht nur die wissenschaftliche Nutzbarkeit und Attraktivität von TextGrid weiterhin unter Beweis zu stellen, sondern auch die Leistungs- und Belastungsfähigkeit des gewählten Betriebsmodells, d.h. die institutionelle Ausgestaltung von TextGrid in rechtlicher, strukturell-organisatorischer und finanzieller Hinsicht. Die Festschreibung eines bestimmten Betriebsmodells ohne eine solche Erprobungsphase wäre sowohl aus der Sicht der „TextGrid-Architekten“ als auch insbesondere für gegenwärtige und künftige TextGrid-Nutzer und ihre Forschungsprojekte problematisch und extrem risikobehaftet. Vielmehr muss die Möglichkeit bleiben, auf der Basis der im Echtbetrieb gesammelten Erkenntnisse und Erfahrungen gegebenenfalls erforderliche Justierungen und Korrekturen vorzunehmen. Dazu erscheint der vorgesehene Förderzeitraum von drei Jahren zwingend nötig, aber auch hinreichend.

Der Forschungsverbund leistet einen entscheidenden Beitrag zur modellhaften Institutionalisierung der Virtuellen Forschungsumgebung TextGrid, indem er die Herausforderungen auf fachwissenschaftlicher, sozialer, organisatorischer, finanzieller, rechtlicher und technischer Ebene untersucht, um belastbare Lösungsvorschläge zu entwickeln. Mit dem Übergang vom projektfinanzierten in den nachhaltigen Betrieb wird ein zuverlässiger und vertrauenswürdiger virtueller Forschungsraum geschaffen. Auf diese Weise wird die Entwicklung der Digital Humanities in Deutschland nachhaltig befördert.

Es steht nicht zu erwarten, dass TextGrid sich ausschließlich aus eigenen Einnahmen – seien es Mitgliedsbeiträge, seien es Gebühren und/oder Leistungsentgelte – wird finanzieren lassen und in diesem ökonomischen Sinne gänzlich autark sein können. Vielmehr wird TextGrid seine Leistungen selbst bei boomender Nachfrage aller Voraussicht nach ohne eine gesicherte Grundfinanzierung (durch Bund und Länder) und einen in Verbindung mit zweckgebunden vergebenen Projektmitteln jeweils zu erhebenden institutionellen Finanzierungsanteil auf Dauer kaum erbringen können. Dies bedeutet, dass eine substanzielle Erweiterung und Verbesserung der Forschungsmöglichkeiten, wie sie TextGrid eröffnet, nicht zum Nulltarif zu haben ist – selbst wenn man davon ausgehen darf, dass das Leistungsangebot einer VRE wie TextGrid bei entsprechender Nachfrage und Nutzung auch Einsparungspotentiale erschließen wird, z.B. durch Verzicht auf ansonsten unvermeidbare Doppel- und Mehrfacharbeit etwa im Bereich der Software-Entwicklung, durch Verzicht auf ebenso ineffiziente wie per Saldo kostenträchtigere Insellösungen und durch Wegfall daraus resultierender aufwändiger Abstimmungs- und Koordinationsprozesse. Entscheidend ist jedoch letztlich nicht dieser ökonomische Aspekt, sondern dass TextGrid bei stabilem und nachhaltigem Betrieb der geisteswissenschaftlichen Forschung die Chance zu einem wissenschaftlichen quantitativen und qualitativen Sprung eröffnet und Deutschland damit auf diesem zukunftssträchtigen Gebiet seinen Vorsprung und seine internationale Führungsrolle zu wahren hilft. Die dazu notwendige dreijährige Übergangsphase, die TextGrid den Weg vom Projekt zur Institution ebnen soll, verdient zumal vor diesem Hintergrund besondere Unterstützung.

2. Stand der Wissenschaft und Technik; bisherige Arbeiten

2.1 Stand der Forschung

Wissenschaftlerinnen und Wissenschaftler zahlreicher Fachgebiete nutzen digitale Forschungsdaten, IT-gestützte Forschungsmethoden und Ressourcen zunehmend standortübergreifend und kollaborativ. Ein entsprechender Bedarf ist dabei nicht länger auf die naturwissenschaftlichen Disziplinen begrenzt: In wachsender Anzahl profitieren heute auch geisteswissenschaftliche Forschungsverbände von den Vorteilen Virtueller Forschungsumgebungen, die neue Möglichkeiten des wissenschaftlichen Arbeitens, neue Fragestellungen, Perspektiven und Methoden eröffnen. Es wird aber auch immer klarer, dass angesichts des stetigen technischen Fortschritts und der zunehmend disziplinenübergreifend angewandten Methoden die Sicherung des nachhaltigen Betriebs einer Virtuellen Forschungsumgebung zu einer zentralen Herausforderung wird. Resultat dieser Erkenntnis ist zum Beispiel die Gründung zahlreicher Digital Humanities-Zentren unterschiedlicher Ausprägung.¹

In den letzten Jahren wurden durch die D-Grid-Initiative² des Bundesministeriums für Bildung und Forschung (BMBF) und dezidierte Ausschreibungen³ der Deutschen Forschungsgemeinschaft (DFG) zahlreiche virtuelle Forschungsumgebungen entwickelt. Allerdings konnten die VREs aus unterschiedlichen Gründen einen zuverlässigen Dauerbetrieb bisher nicht realisieren.

Die Herausforderungen liegen dabei insbesondere im Übergang aus der projektfinanzierten Förderung in einen nachhaltigen Betrieb: Vor allem waren es organisatorische und rechtliche Probleme, die sich hier als gravierendes Hindernis erwiesen. Das gilt nicht zuletzt für die Organisation der Zusammenarbeit über bundesstaatliche bzw. administrative Grenzen hinweg, beispielsweise bei der gemeinsamen Nutzung von Hard- oder Software-Ressourcen. Wie Untersuchungen im Rahmen der Arbeitsgruppe „Virtuelle Forschungsumgebungen“ der Allianz-Schwerpunktinitiative⁴ zeigen, haben sich bisher nur sehr wenige VREs Perspektiven für einen langfristigen Betrieb und eine dauerhafte Weiterentwicklung erarbeiten können. Untersucht wurden verschiedene VREs aus unterschiedlichen Disziplinen, unterschiedlichen Förderkontexten (Bundesländer, DFG, BMBF, EU), mit unterschiedlich langen Laufzeiten und unterschiedlichen Graden der Internationalisierung/Standardisierung. Im Fall der Klimawissenschaften wurde den beteiligten Partnern beim Wechsel vom ersten Projekt zum Folgevorhaben sehr schnell klar, dass die Nachhaltigkeit von Forschungsumgebungen nicht durch eine Kette von Projektförderungen erreicht werden kann⁵. Vielmehr muss sich die Weiterführung einer Forschungsumgebung auf ein

¹ Siehe z.B. centerNet (<http://digitalhumanities.org/centernet/centers/>), das 47 Digital Humanities Zentren innerhalb Europas nachweist, darunter in Deutschland das Zentrum für digitale Edition in Würzburg (<http://www.zde.uni-wuerzburg.de/>), die Initiative Hamburg Digital Humanities (HDH) (<http://www.hdh.uni-hamburg.de/>); das Zentrum Geisteswissenschaften in der digitalen Welt (<http://www.g-dig.uni-hamburg.de/index.htm>), das Göttingen Centre for Digital Humanities (GCDH) (<http://www.uni-goettingen.de/de/136017.html>), das Cologne Center for eHumanities (CCeH) (<http://www.cceh.uni-koeln.de/>) sowie die Akademie der Wissenschaften und der Literatur Mainz – Digitale Akademie (<http://www.digitale-akademie.de/>).

² <http://www.d-grid.de/>

³ http://www.dfg.de/formulare/12_12/12_12.pdf

⁴ http://www.allianzinitiative.de/de/handlungsfelder/virtuelle_forschungsumgebungen

⁵ Die nationale, aber auch internationale Förderlandschaft ist grundsätzlich auf begrenzte Projektlaufzeiten ausgelegt und fungiert als Impulsgeber, jedoch nicht als dauerhafter Geldgeber.

eigenes dauerhaft tragfähiges organisatorisches Betriebsmodell stützen, das im Zuge des laufenden Projekts zu entwickeln ist.

Auf internationaler Ebene wird die Entwicklung zurzeit hauptsächlich durch die EU und ihr Engagement im Bereich „Research Infrastructure“ vorangetrieben. In Deutschland ist der Begriff „Forschungsinfrastrukturen“ allerdings noch nicht klar definiert und derzeit im Wesentlichen durch das „European Strategy Forum on Research Infrastructures (ESFRI)“⁶ und somit durch den EU-Kontext geprägt. Eine wichtige Erkenntnis der ESFRI-Initiative ist u.a., dass große, internationale Kooperationen auf Dauer nur funktionieren, wenn ihnen ein einheitlicher organisatorischer und rechtlicher Rahmen gegeben wird. Daher implementierte die EU das Instrument ERIC (European Research Infrastructure Consortium)⁷, mit dessen Hilfe sich Konsortien zu einer Vereinigung nach europäischem Recht zusammenschließen können. Ähnlich wie bei einer deutschen Vereinssatzung sind die Managementstrukturen und Mitgliedsbeiträge klar definiert, wobei der nachhaltige Aufbau und Betrieb durch Beteiligung der Ministerien der betreffenden Länder zumindest für 10 Jahre gesichert ist. Jedes ERIC wird von der EU evaluiert, danach von den Forschungsministerien der beteiligten Mitgliedsländer unterschrieben und schließlich implementiert. Erwartet wird, dass bei künftigen Projekten bereits entwickelte, passfähige ERIC-Strukturen genutzt und weiterentwickelt werden. Der ESFRI-Prozess dürfte weltweit einmalig sein. Eine ähnliche Initiative in den USA, DataNet⁸, ist auf Grund der dortigen Wirtschafts- und Finanzkrise wesentlich kleiner ausgefallen als ursprünglich geplant und umfasst auch wesentlich weniger Fachdisziplinen. Durch das ERIC-Instrument erhalten die Forschungsinfrastrukturen einen organisatorischen und vor allem rechtlich sicheren Rahmen mit festen Verantwortlichkeiten und Managementstrukturen. Darüber hinaus werden von Beginn an die Förderorganisationen der EU-Mitgliedsländer einbezogen, was mittel- bis langfristig einer wesentlich höheren finanziellen Absicherung dient.

Zusammenfassend lässt sich festhalten, dass es verschiedene, z.T. vielversprechende Ansätze gibt, wie eine virtuelle Forschungsumgebung nachhaltig betrieben, finanziert und weiterentwickelt werden kann, dass bisher aber noch keine VRE in Deutschland dieses Stadium erreicht hat. Es gibt auch bisher keine systematische Untersuchung darüber, welche Schritte in einer bestimmten Entwicklungsphase einer VRE unternommen werden müssen, um die Voraussetzungen für einen nachhaltigen Betrieb zu schaffen.

Nach bisheriger, noch vorläufiger Auswertung von sieben detailliert befragten VREs in Deutschland durch die Allianz AG „Virtuelle Forschungsumgebungen“ sind es vor allem die im Folgenden skizzierten Faktoren, die direkten Einfluss auf die Nachhaltigkeit nehmen (können):

- **Technologie:** Viele VREs geben an, dass ihre Technologie noch im Beta-Stadium ist und noch nicht verlässlich arbeitet, z.B. fehlen Performanz- und Belastungstests. Zusätzlich gibt es schlecht kalkulierbare Abhängigkeiten von externen Technologien.
- **Handhabung:** Die Handhabung einer VRE durch die Nutzer wird überwiegend als zu komplex empfunden. Das bedeutet, dass aller Wahrscheinlichkeit nach wesentlich mehr Aufwand in die Verbesserung der Nutzerfreundlichkeit inklusive Nutzerbefragungen und Nutzerschulungen investiert werden muss (Stichwort „usability“ etc.).

⁶ http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri

⁷ http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=eric

⁸ <http://www.nsf.gov/pubs/2007/nsf07601/nsf07601.htm>

- **Forschungspraxis:** Hier stellen sich keine besonderen Probleme, da sich die VREs offensichtlich gut in den Forscheralltag und in die Forschungspraxis eingefügt haben.
- **Technische Unterstützung:** Auch hier gibt es keine nennenswerten Schwierigkeiten, da die technische Unterstützung bei den meisten VREs angemessen und ausreichend zu sein scheint.
- **Trainings- und Schulungsunterstützung:** Hier geben die meisten VREs an, dass dies noch ausgebaut werden muss und nicht genügend Projektmittel zur Verfügung stehen bzw. dass dieser Faktor zu Beginn nicht hinreichend berücksichtigt wurde.
- **Größe der Nutzergruppe:** Die meisten VREs geben an, dass die Nutzergruppe zurzeit noch zu klein ist, angesichts des Potentials aber noch erweitert werden kann und soll.
- **Sicherheit/Vertrauen:** Einige VREs geben an, dass noch Nachholbedarf im technischen Bereich und bei der Schaffung von Vertrauen besteht. Bei den meisten VREs gibt es hierzu jedoch keine konkreten Hinweise.
- **Akzeptanz:** Bei einigen VREs ist die Nutzerakzeptanz noch relativ gering, da in den betreffenden Fachdisziplinen der gesamte Forschungsprozess bzw. große Teile davon noch nicht standardmäßig IT-gestützt ablaufen.

Folgende Publikationen und Ergebnisse sind in diesem Zusammenhang von Bedeutung:

- **Definition VRE:**
http://www.allianzinitiative.de/fileadmin/user_upload/2011_VRE_Definition.pdf
- Virtuelle Forschungsumgebungen - Ein **Leitfaden** für WissenschaftlerInnen:
http://www.allianzinitiative.de/fileadmin/user_upload/Leitfaden_VRE_de.pdf
- **Fragenkatalog** für die Analyse von VREs:
http://www.allianzinitiative.de/fileadmin/user_upload/Fragenkatalog_VRE_de_01.pdf
- **Ergebnisse** einer Expertenrunde „Förderung Virtueller Forschungsumgebungen“ mit Vertretern der verschiedenen Fördereinrichtungen (Sep. 2010):
<http://www.wissgrid.de/veranstaltungen/abgeschlosseneveranstaltungen/expertenrunde1.html>
- Aspekte der **Implementierung eines Betriebsmodells** für Virtuelle Forschungsumgebungen (Mai 2011):
http://www.wissgrid.de/publikationen/deliverables/wp1/WissGrid_AP1_D15_v101_final.pdf

- **Expertise** zur Errichtung einer Virtuellen Arbeitsumgebung für die Sozial-Oekonomische Berichterstattung (SOEB), soeb-Arbeitspapier 1-2010⁹: <http://www.soeb.de/virtaug.php>
- Schlünzen F., Agapov I., Grimme C., Rathmann T. & Enke H. (2011): **Gridarchitekturen** im Überblick, Evaluation existierender Architekturkonzepte: <http://www.wissgrid.de/publikationen/deliverables/wp2/Gridarchitekturen-im-Ueberblick-V2.0.pdf>
- Leitfaden zum **Forschungsdaten-Management** und Checkliste zum Forschungsdaten-Management: <http://www.wissgrid.de/publikationen/deliverables/wp3/WissGrid-oeffentlicher-Entwurf-Leitfaden-Forschungsdaten-Management.pdf> und <http://www.wissgrid.de/publikationen/deliverables/wp3/WissGrid-oeffentlicher-Entwurf-Checkliste-Forschungsdaten-Management.pdf>

2.2 Bisherige Arbeiten des TextGrid-Konsortiums

Das Projekt TextGrid startete am 1. Februar 2006 und wird seither vom BMBF im Rahmen der D-Grid-Initiative gefördert. Die erste Projektphase endete im April 2009. Die Förderung der zweiten Phase läuft noch bis Mai 2012. Deren Ziel ist es, neben der Ausdehnung auf weitere geisteswissenschaftliche Disziplinen im Zuge der Weiterentwicklung, die Konsolidierung und Verbesserung von TextGrid-Grundlagen für den Übergang des Projekts in den Dauerbetrieb zu schaffen. Bearbeitet wurden/werden insgesamt sechs Arbeitspakete:

- AP 1: TextGrid Repository
- AP 2: TextGrid Laboratory
- AP 3: Strukturelle und organisatorische Nachhaltigkeit
- AP 4: Coaching/Community Building
- AP 6: Fachwissenschaftliche Nachhaltigkeit
- AP 8: Nationale/internationale Liaison, Projektmanagement, Öffentlichkeitsarbeit

Mit der Freischaltung der TextGrid-Version 1.0 im Juli 2011 ist eine der Grundvoraussetzungen für den nachhaltigen Betrieb von TextGrid geschaffen und ein wesentliches Projektziel noch vor dem ursprünglich für Mai 2012 vorgesehenen Termin erreicht worden. Mit TextGrid 1.0 steht nunmehr eine Virtuelle Forschungsumgebung für die standortungebundene, kooperative Forschung namentlich in den textbezogenen Geisteswissenschaften zur Verfügung. Mit ihr wird national wie international gesehen Neuland betreten und die Vorrangstellung Deutschlands auf diesem Gebiet gesichert und weiter ausgebaut.

Mit seinen beiden Hauptkomponenten, dem „**TextGrid Laboratory**“ (TextGridLab) und dem „**TextGrid Repository**“ (TextGridRep), richtet sich TextGrid an wissenschaftliche Nutzerinnen und Nutzer, die Werkzeuge und Dienste für die Verzeichnung, Beschreibung, Annotation, Erschließung, Auswertung und Publikation von kulturellen Artefakten, insbesondere von Texten, Bildern, Handschriften, Noten und weiteren Objekten benötigen und die Wert auf die vielfältige Nachnutzbarkeit von Forschungsdaten legen. Bis zum Ende der laufenden Förderphase werden

⁹ Dickmann F., Enke H., Harms P. (2010)

die Werkzeuge des TextGridLab kontinuierlich optimiert und in Zukunft zusätzliche Komponenten wie etwa das extern entwickelte Werkzeug „XML Print“ stufenweise integriert. Dabei werden sowohl die Angebote für die Kerndisziplinen Editionswissenschaft und Linguistik ausgebaut als auch für die Fächer Musikwissenschaft, Kunstgeschichte und Klassische Philologie neue Spezialwerkzeuge entwickelt. Komplementär dazu ermöglicht das TextGridRep als zweite Hauptkomponente der Virtuellen Forschungsumgebung bereits jetzt die Speicherung und langfristige Aufbewahrung sowie Nachnutzung geistes- und kulturwissenschaftlicher Forschungsdaten. Dabei gewährleistet TextGrid derzeit die Verfügbarkeit und Sicherung dieser Daten für einen Zeitraum von zehn Jahren.

2.2.1 Nutzer und Nutzungsmodelle

Die aktive Nutzung von TextGrid hat sich bereits zu diesem frühen Zeitpunkt außerordentlich erfreulich entwickelt, was den Bedarf in den Fachdisziplinen unterstreicht. Seit dem nach fünfjähriger Entwicklungs- und Testphase im Juli 2011 erfolgten Release von TextGrid 1.0 wird die Forschungsumgebung sowohl in Projekten als auch in der Lehre genutzt; außerdem gibt es verschiedene Formen der bereits realisierten oder geplanten Weiterentwicklung. In der nachfolgenden Aufstellung wird unterschieden nach Fachwissenschaftlerinnen und Fachwissenschaftlern, die bereits die Version 1.0 produktiv einsetzen, weiteren Nutzungsanfragen, dem Einsatz von TextGrid in der Lehre, Kooperationspartnern, mit denen TextGrid gemeinsam weiterentwickelt wird, strategischen und inhaltlichen Allianzen sowie der sozialwissenschaftlichen Begleitforschung. Im Einzelnen handelt es sich um folgende Nutzungsmodelle:

- Fachwissenschaftler (Nutzer der v1.0):	ca. 35 in 8 Projekten
- Konkrete Nutzungsanfragen:	18 Projekte ¹⁰
- Perspektivische Nutzung:	4 Projekte
- Einsatz in der Lehre:	10 Seminare
- Konkrete gemeinsame Entwicklungen (Tools u. Repository):	4 Projekte
- Geplante gemeinsame Entwicklungen (Tools u. Repository):	3 Projekte
- Strategische Allianzen:	5 Projekte
- Inhaltliche Allianzen:	5 Projekte
- Sozialwissenschaftliche Begleitforschung:	1 Projekt

Damit nutzen heute bereits acht z.T. langfristig (über 10 Jahre) angelegte Forschungsverbünde die Virtuelle Forschungsumgebung TextGrid aktiv für die Erstellung wissenschaftlicher Editionen, für die Analyse ihrer Forschungsdaten, für die Entwicklung projektspezifischer Analyse- sowie Visualisierungswerkzeuge und schließlich für die langfristige digitale Archivierung und Bereitstellung der Forschungsdaten zur weltweiten Nachnutzung durch die Scientific Community. Zu nennen sind hier beispielhaft:

- „Theodor Fontanes Notizbücher“ (Hybrid-Edition) (Theodor-Fontane-Arbeitsstelle, Universität Göttingen, 2011-2014)
- „Blumenbach online“ (Akademie der Wissenschaften zu Göttingen, 2010-2025)
- „Edition der Briefe von Ernst Kantorowicz“ (Deutsches Literaturarchiv Marbach, 2010-2012)
- „Fiedellieder von Theodor Storm“ (Hochschule für Musik Detmold/Paderborn, 2011)
- „Briefwechsel Grimm – Savigny – Bang“ (TU Darmstadt, seit 2011)

¹⁰ Davon 12 im BMBF-eHumanities-Call, <http://www.bmbf.de/foerderungen/16466.php>

- „Virtuelles Skriptorium St. Matthias“ (Stadtbibliothek und Stadtarchiv Trier/KoZe, TU Darmstadt, 2010-2013)
- „Deutsch-jüdische Publizistik des 19. Jhdts.“ (Steinheim-Institut, Essen, seit 2011)
- „Georg Greflinger – Digitale Archiv-Edition“ (FU Berlin, seit 2011)

Bis Oktober 2011 – nur drei Monate nach dem Release der Virtuellen Forschungsumgebung TextGrid – lagen darüber hinaus bereits 18 weitere Anfragen von Forschungsverbänden vor, die nach eingehenden Konzeptionsgesprächen mit dem TextGrid-Konsortium gegenwärtig in der konkreten Planungsphase stehen. Darunter befindet sich mit dem bei der Akademie der Wissenschaften NRW beantragten Forschungsvorhaben „Wörterbuch des Klassischen Maya“ der Universität Bonn ein wissenschaftlich äußerst innovatives altamerikanistisches Projekt, das – ebenso wie das bereits seit 2010 mit TextGrid arbeitende Projekt „Blumenbach online“ (Akademie der Wissenschaften zu Göttingen) – auf eine 15-jährige Laufzeit angelegt ist.

Auf die dezidierte Nutzungsempfehlung von TextGrid im Rahmen des BMBF-eHumanities-Calls sind zwölf Forschungsvorhaben zurückzuführen, in deren Rahmen eine Kontaktaufnahme mit TextGrid-Partnern sowie erste Planungsgespräche erfolgt sind. Weitere vier Forschungsverbände haben sich aufgrund von Schulungen bereits für eine Nutzung von TextGrid im Rahmen von Editionsprojekten ab dem Jahr 2012 entschieden. Zum gegenwärtigen Zeitpunkt liegen dem Konsortium damit 22 Anfragen von Forschungsverbänden vor, die TextGrid nutzen wollen. Ein Vergleich mit den computergestützten Angeboten der infrastrukturell ungleich stärkeren Klimaforschung, die nach zehnjähriger Entwicklungszeit heute rund 100 Nutzer verzeichnen, zeigt, dass die wissenschaftlichen Nutzerzahlen in Kürze in einer ähnlichen Größenordnung liegen werden. Langfristig wird für den geistes- und kulturwissenschaftlichen Bereich eine wesentlich höhere Nutzerzahl erwartet.

Besonders hervorzuheben ist dabei der wissenschaftlich innovative Charakter der beteiligten Forschungsverbände. Dieser gründet zum einen auf der Vielgestaltigkeit des Materials und der Objekte (Beispiel: „Blumenbach online“), zum anderen auf den Dimensionen der bereitgestellten oder bereits verfügbaren Corpora und schließlich auf dem oft interdisziplinären Forschungsansatz (u.a. „Relationen im Raum“). Nutzungsanfragen liegen aus ganz unterschiedlichen Fachdisziplinen wie Baugeschichte, Judaistik, Kunstgeschichte, Latinistik u.a. vor, die das Spektrum von TextGrid weit über die anfangs adressierte Editionsphilologie und Linguistik hinaus erweitern. Viele Projekte formulieren darüber hinaus eine auf die eHumanities reflektierende Forschungsfrage: Sie loten, z.T. dezidiert, die Möglichkeiten der neuartigen Editionsmethoden ebenso wie die damit verbundenen neuen Fragestellungen aus (vgl. u.a. „ePoetics“, „eCodicology“, „Freischütz digital“). Dezidiert nachgefragt wird außerdem die Nutzung der bereits heute in der „Digitalen Bibliothek“ des TextGridRep verfügbaren, texttechnologisch erschlossenen literarischen Primärtexte. Das gilt auch für mehrere der im Rahmen des BMBF-Calls eingereichten Forschungsanträge.

Gerade das Repository ist dazu angelegt, einen nachhaltigen Beitrag zur Forschung zu liefern, der die oben formulierte Frage nach den Möglichkeiten der eHumanities jeweils neu, aus unterschiedlichen fachwissenschaftlichen Perspektiven und z.T. in interdisziplinärer Ausrichtung beantwortet. Hierfür liefert die Tatsache, dass bereits heute rund 35 Wissenschaftler aus so unterschiedlichen Disziplinen wie Sprach- und Literaturwissenschaft, Archäologie, Musikwissenschaft, Sozialforschung, Publizistik sowie Informationswissenschaft und Informatik TextGrid für ihre Vorhaben nutzen, einen eindrucksvollen Beleg. Die dem TextGrid-Konsortium vorliegenden Nutzungsanfragen aus altamerikanistischen, ägyptologischen, geologischen, bioinformatischen und historischen Fächern verweisen zudem auf den hohen Bedarf an

computergestützter, interdisziplinärer und fachübergreifender Forschung, wie sie TextGrid ermöglicht.

In der Lehre wird TextGrid bereits an der Humboldt-Universität Berlin, der Technischen Universität Darmstadt, der Hochschule für Musik Detmold, der Universität Köln, der Universität Würzburg und der Universität Göttingen eingesetzt.

Per 13. Februar 2012 ergibt sich damit folgende Statistik der TextGrid-Accounts:

- insgesamt registrierte Nutzer: 912
- Nutzer, die sich bereits neu in TextGrid Version 1.0 eingeloggt haben: 253

2.2.2 Etablierung eines nachhaltigen Betriebs- und Geschäftsmodells

Besondere Bedeutung kommt jetzt den in den bisherigen Arbeitspaketen 3 und 6 entwickelten Überlegungen zum Betriebs- und Geschäftsmodell und damit zur wissenschaftlichen, rechtlichen, strukturell-organisatorischen und finanziellen Gestaltung von TextGrid als einer institutionell verfassten und als solche auf Dauer funktionsfähigen Institution zu. Hier gibt es jeweils unterschiedliche Optionen: Prinzipiell lässt sich TextGrid künftig strukturell-organisatorisch gestalten als gemeinnützige GmbH (mit Vollrechtsfähigkeit und eigenem Grundkapital), als eingetragener Verein (mit Vollrechtsfähigkeit) oder als Abteilung im Rahmen von D-Grid (mit indirekter Rechtsfähigkeit über die D-Grid GmbH). Dementsprechend ergeben sich auch für die Finanzierung unterschiedliche Möglichkeiten, die sich teilweise kombinieren lassen (und in der Praxis wohl auch kombiniert werden müssen). Denkbar wären:

- Beiträge von Vereinsmitgliedern (natürlichen und/oder juristischen Personen)
- Gebühren und Leistungsentgelte von Nutzern (natürlichen und/oder juristischen Personen)
- Institutionelle Grundfinanzierung (durch Bund und Länder)
- Projektbezogene Förderung, gegebenenfalls unter Nutzung von Overheads zur anteiligen institutionellen Finanzierung

Die mit den unterschiedlichen Optionen verbundenen Vor- und Nachteile, Chancen und Risiken wurden eingehend untersucht. Nach Abwägung der gegebenen Alternativen hat sich das Konsortium entschlossen, im Juni 2012 einen Verein „**TextGrid – Verein zum nachhaltigen Betrieb einer digitalen Forschungsinfrastruktur in den Geisteswissenschaften e.V.**“ zu gründen, der im Verlauf der dreijährigen Übergangsphase auf seine Eignung als institutionelle Basis für den Dauerbetrieb getestet werden soll. Die konstituierende Mitgliederversammlung wird voraussichtlich am 16.05.2012 erfolgen (Zeitplan siehe Anlage I Blatt 2). Eine Vereinssatzung liegt vor (vgl. Anlage I Blatt 1) und wird zurzeit in den verschiedenen Rechtsabteilungen der Partner geprüft. Eine inhaltliche Zusage hinsichtlich Voll-Mitgliedschaft und Mitgliedsbeitrag haben bereits folgende Institutionen abgegeben; sie haben damit zum Teil auch eine Verantwortung für die (Teil-)Institutionalisierung verschiedener Bereiche übernommen:

- Akademie der Wissenschaften zu Göttingen (AdW)
- Akademie der Wissenschaften und der Literatur, Mainz
- DAASI International GmbH, Tübingen
- Fachhochschule Worms, Fakultät für Informatik und Telekommunikation
- Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG)
- Georg-August-Universität Göttingen, Niedersächsische Staats- und Universitätsbibliothek (SUB)
- Institut für Deutsche Sprache (IDS), Mannheim (Geschäftsstelle / Vereinssitz)

- Max-Planck-Institut für Wissenschaftsgeschichte (MPI WG), Berlin
- Salomon Ludwig Steinheim-Institut für deutsch-jüdische Geschichte an der Universität Duisburg-Essen
- Technische Universität Darmstadt, Institut für Sprach- und Literaturwissenschaft
- Universität Paderborn, Musikwissenschaftliches Seminar Detmold/Paderborn
- Universität Würzburg, Lehrstuhl für Computerphilologie und Neuere Deutsche Literaturgeschichte

Mit dem nun erreichten Entwicklungsstand unter Einschluss der bis Mai 2012 zu erwartenden Verbesserung und Erweiterung seines Angebots als VRE hat TextGrid die angestrebten Ziele – z.T. sogar vorfristig – erreicht und präsentiert sich als Erfolgsgeschichte im ebenso anspruchsvollen wie schwierigen Feld der Digital Humanities. TextGrid steht hier an der Spitze der Forschung und beweist mit der wachsenden Nutzerzahl seine Funktionsfähigkeit und Attraktivität für die Geisteswissenschaften. Es profiliert sich dabei nicht nur als neues Werkzeug zur besseren Bewältigung herkömmlicher Forschungsthemen, -fragen und -probleme, sondern auch als Impulsgeber zur Erschließung neuer Forschungsfelder, zur Entwicklung neuer Forschungsfragen und –verbünde sowie zur Erprobung neuer methodischer Ansätze. Solche innovativen Methoden, etwa zur empirischen Prüfung mittels corpusbasierter Verfahren, erweitern und ergänzen in fruchtbarer Weise die bestehenden geisteswissenschaftlichen Instrumente und sichern den Geisteswissenschaften in Deutschland weiterhin eine Spitzenstellung in der internationalen Forschungslandschaft. Es steht allerdings außer Frage, dass der mit der Entwicklung und Unterhaltung einer solchen VRE verbundene personelle, organisatorische und finanzielle Aufwand nur zu rechtfertigen ist, wenn TextGrid einen entsprechenden Forschungsschub dieser Art und Größenordnung auszulösen vermag. Auch die mit dem nachhaltigen Betrieb verbundenen Kosten wären nur so zu legitimieren. Anders gesagt: Als Institution kann und muss sich TextGrid langfristig über eine hinreichend große Zahl gegebenenfalls auch zahlungsbereiter Nutzergemeinschaften und ein breites Spektrum ihrer fachlichen Provenienzen legitimieren. Einen entsprechenden Nachweis der wissenschaftlichen Attraktivität von TextGrid zu erbringen, war in der bis Juli 2011 laufenden Beta-Phase des Projekts naturgemäß nicht möglich, und der Nachweis wird sich auch in der bis Mai 2012 für den Echtbetrieb verbleibenden Zeitspanne nicht in der Weise erbringen lassen, dass im unmittelbaren Anschluss daran bereits an eine selbsttragende Finanzierung zu denken wäre. Vielmehr bedarf es dazu eines längerfristigen Betriebs unter wissenschaftlichen Alltagsbedingungen und einer erst jetzt sinnvoll leistbaren wesentlichen Erweiterung des Nutzerspektrums. Eben darauf zielt die beantragte Weiterführung der Förderung für drei Jahre. Das Konsortium ist überzeugt, dass sich TextGrid danach nicht nur als erfolgreiches Projekt, sondern auch als lebens- und funktionsfähige, institutionalisierte Forschungsinfrastruktur erweisen wird, die der Wissenschaft in Forschung und Lehre neue Möglichkeiten und Perspektiven zu eröffnen vermag. Dass sich TextGrid hier auf einem höchst zukunftssträchtigen Feld bewegt, lässt sich u.a. auch daraus ersehen, dass in dem 2013 startenden 8. EU-Rahmenprogramm Forschungsinfrastrukturen und Virtuelle Forschungsumgebungen eine prominente Rolle spielen werden. Die Ausschreibung des BMBF im eHumanities-Call weist in die gleiche Richtung.

3. Ausführliche Beschreibung des Arbeitsplans

Eine besondere Herausforderung für den funktions- und leistungsfähigen Dauerbetrieb wird darin liegen, für TextGrid einerseits eine stabile und für Nutzer ebenso wie für Anbieter eine hinreichend Sicherheit und Berechenbarkeit garantierende Architektur zu finden, diese Architektur andererseits aber so flexibel zu gestalten, dass auf Veränderungen sowohl auf der Nutzer- als auch auf der Anbieterseite mit gebotener Zügigkeit erfolgreich reagiert werden kann. Gerade hier gilt es, in der beantragten Übergangsphase geeignete Lösungen zu entwickeln, unter Alltagsbedingungen zu testen und gegebenenfalls die Architektur zu optimieren.

Besondere Anforderungen ergeben sich für TextGrid durch die Klientel, für die und mit der das Angebot dieser VRE entwickelt wird. Denn anders als bestehende VREs etwa im Bereich der Astronomie, der Medizin oder der Klimaforschung adressiert TextGrid mit den Geisteswissenschaften eine Community von großer Heterogenität in den Orientierungen, Quellen, Themen, Fragestellungen und Methoden, bei einer nach wie vor dominierenden Tendenz zu individueller Forschung und vergleichsweise begrenzten Erfahrungen mit kooperativen Arbeitsformen. Dem gilt es im Zuge der weiteren Arbeit an und mit TextGrid Rechnung zu tragen. Dabei bietet die Übergangsphase nicht nur die Chance, generell für die Möglichkeiten von TextGrid zu werben, sondern auch die beabsichtigte Einbeziehung von Fächern wie Kunstgeschichte, Klassische Philologie und Musikwissenschaft mit der gebotenen Sorgfalt vorzunehmen und überzeugend zu dokumentieren.

Schließlich soll die Übergangsphase auch dazu genutzt werden, die nötigen Instrumente und Verfahren der Qualitätssicherung und Evaluation in den unterschiedlichen Arbeitsbereichen für den Dauerbetrieb mit seinen Alltagsanforderungen und -belastungen gezielt weiterzuentwickeln und auf Praktikabilität und Validität zu prüfen.

Von daher stellen sich in den kommenden drei Jahren sowohl operative als auch konzeptionelle Aufgaben: Zum einen gilt es, die laufende Arbeit von TextGrid aufrecht zu erhalten; zum anderen geht es um die konzeptionelle Optimierung, Erprobung und institutionelle Implementierung der Strukturen, Verfahren und Instrumente für dessen künftigen nachhaltigen Betrieb. Mit einer Einstellung der Förderung am Ende der laufenden Projektphase würde demgegenüber nicht nur die Chance zur Etablierung einer dauerhaft funktionsfähigen VRE vertan; ohne kontinuierliche Pflege würden auch die derzeit schon verfügbaren und genutzten Tools binnen Kurzem zu einer Software-Ruine ohne Gebrauchs- und Nachnutzungswert veralten. Das wiederum hätte nicht nur für TextGrid selbst fatale Folgen; vor allem die zahlreichen Nutzer von TextGrid gerieten mit ihren eigenen Forschungsprojekten dann in arge Bedrängnis – ein kaum wettzumachender Vertrauensverlust für die Verwendung von Forschungsinfrastrukturen und darüber hinaus auch ein schwerer Rückschlag für die eHumanities selbst sowie für neue Entwicklungen und sich abzeichnende Innovationen digitaler Methodik in den Geistes- und Kulturwissenschaften.

Im Einzelnen ergeben sich für die dreijährige Übergangsphase in den Dauerbetrieb folgende sieben Arbeitspakete (AP) (siehe Abb. 1):

- AP 1 – Betriebsmodell
- AP 2 – Prozessorganisation
- AP 3 – Nutzer-Kommunikation
- AP 4 – Betrieb und Softwarepflege Repository
- AP 5 – Betrieb und Softwarepflege Laboratory
- AP 6 – Integriertes Monitoring und Qualitätssicherung
- AP 7 – Projektmanagement und Öffentlichkeitsarbeit

Die Institutionalisierung von TextGrid ist dabei Gegenstand jedes einzelnen Arbeitspakets; dazu gehört die Unterstützung der verschiedenen Nutzungsszenarien ebenso wie die Entwicklung von Strategien zur institutionellen, fachwissenschaftlichen, strukturell-organisatorischen und technischen Nachhaltigkeit (siehe Abb. 1).



Abb. 1: Institutionalisierung von TextGrid und deren Umsetzung in den einzelnen Arbeitspaketen.

4. Verwertungsplan

Mit der Freischaltung der Version 1.0 im Juli 2011 ist TextGrid bereits in seine 'Verwertungsphase' eingetreten – und dies, wie die steigende Zahl der Nutzer belegt, sehr erfolgreich. So erweist sich TextGrid schon jetzt als wichtiger Motor für die Entwicklung von VREs in den Geisteswissenschaften. Mit der vorgesehenen Ausweitung seiner Nutzergemeinschaft – über die (editions-)philologischen und linguistischen Disziplinen hinaus – auf Fächer wie Musikwissenschaft, Kunstgeschichte und Archäologie wird TextGrid seine Vorreiterrolle auf diesem ebenso anspruchsvollen wie schwierigen Gebiet der Digital Humanities in den nächsten Jahren weiter ausbauen können. Im enger werdenden Austausch zwischen Anbietern und Nutzern sind dabei auf technologischer Ebene weitere Verbesserungen und Verfeinerungen des Systems zu erwarten. Allerdings liegt die Bedeutung von TextGrid als VRE nicht nur darin, neue Werkzeuge zur besseren Bewältigung herkömmlicher Forschungsthemen, -fragen und -probleme zu entwickeln und bereitzustellen, sondern zugleich auch Impulse zu geben zur Erschließung neuer Forschungsfelder, zur Erarbeitung wegweisender neuer Forschungsfragen, zur Erprobung innovativer methodischer Ansätze und nicht zuletzt zur Bildung und Gestaltung neuer Formen fächerübergreifender wissenschaftlicher Vernetzung und Kooperation. Da Geisteswissenschaften an fast jedem Universitäts- und Forschungsstandort in Deutschland vertreten sind, bietet sich für TextGrid hier schon rein quantitativ ein außerordentlich großes Nutzer-Potential. Um es nachhaltig ausschöpfen zu können, bedarf es angesichts der hier nach wie vor noch zu verzeichnenden Zurückhaltung gegenüber 'neuen Technologien' in den kommenden Jahren allerdings verstärkter, gezielter und sorgfältig strukturierter Überzeugungsarbeit. Dabei wird man insbesondere auf die Sogwirkung setzen können (und müssen), die von Vorhaben ausgeht, die ihren wissenschaftlichen Erfolg nicht zuletzt der Nutzung von TextGrid verdanken. Für die Geisteswissenschaften selbst wird dies in vielen Bereichen zu Um- und Neuorientierungen führen. Dies betrifft nicht nur die weitere Verbreitung von VREs in diesen Fächern und ihre wachsende Bedeutung für die geisteswissenschaftliche Arbeit in Forschung und Lehre; vielmehr wird sich darüber hinaus auch das Zusammenspiel der am Wissenschaftsprozess direkt oder indirekt beteiligten Akteure und 'Stakeholder' – Wissenschaftlerinnen und Wissenschaftler als Nutzer entsprechender Dienste, Software-Entwickler, Software-Produzenten und Software-Anbieter, Daten- und Ressourcenlieferanten, Wissenschaftsmanager, Wissenschaftsförderer und Wissenschaftspolitiker, Lehrende und Lernende – entsprechend der wachsenden Bedeutung virtueller Forschungsumgebungen neu konfigurieren (müssen). Dabei werden Forschungsinfrastrukturen generell größeres Gewicht erhalten und die Dienstleistungsstruktur in der Wissenschaft zunehmend prägen. TextGrid wird auch hier seine Wirkung entfalten. Nicht zuletzt gilt dies mit Blick auf die europäische Ebene und die Möglichkeiten grenzüberschreitender Kooperationen. TextGrid ist hier bereits international sehr gut vernetzt und wird auch künftig sowohl davon profitieren als auch seinerseits wichtige Impulse geben. Vor allem aber kann und wird TextGrid dazu beitragen, dass Deutschland auf diesem zukunftsweisenden Gebiet seine internationale Führungsposition wahrt.

5. Kooperation / Zusammenarbeit mit Dritten

Damit sich TextGrid als Institution am 'Markt' durchsetzen und auf Dauer behaupten kann, bedarf es nicht nur beharrlicher Anstrengungen aller an diesem Vorhaben unmittelbar Beteiligten, sondern auch geeigneter äußerer Rahmenbedingungen im wissenschaftlichen und wissenschafts-politischen Umfeld. Hier kann – und muss – TextGrid selbst sicherlich wichtige Anregungen und Impulse geben; doch braucht derartige 'Überzeugungsarbeit' schon angesichts der Breite und Vielfalt der damit ins Spiel kommenden 'Player' Geduld, Zeit und einen langen Atem. Zu den wichtigsten Akteuren, die es hier anzusprechen und für TextGrid (noch) zu gewinnen gilt, gehören:

Die Wissenschaftlerinnen und Wissenschaftler

Als tatsächliche oder potentielle Nutzer gilt es nicht nur, sie von der Leistungsfähigkeit der mit TextGrid verfügbaren Instrumente und Verfahren zu überzeugen, sondern auch dazu anzuregen, auf dieser Basis eigene neue Forschungsfragestellungen zu generieren, damit verbundene neue Anforderungen an das Programm-Angebot zu entwickeln, zu artikulieren und kooperativ und aktiv an seiner weiteren Ausgestaltung mitzuwirken.

Die Hochschulen, ihre Leitungen und die Hochschulrektorenkonferenz (HRK)

Diejenigen, die hier Verantwortung tragen, müssen den mit VREs wie TextGrid hochschulintern verbundenen veränderten Anforderungen personell, strukturell-organisatorisch und in Bezug auf Ressourcen Rechnung tragen und entsprechende Bedarfe in ihren eigenen Struktur- und Finanzplanungen angemessen berücksichtigen. Dazu gehört auch die Integration VRE-bezogener Qualifikationsangebote in die Studiengänge und Curricula. Nur wenn einschlägig ausgebildetes Personal vor allem im Nachwuchsbereich zur Verfügung steht, werden sich VREs auf Dauer institutionell konsolidieren, behaupten und das damit vorhandene Innovationspotential ausschöpfen können.¹¹ Das setzt auch voraus, dass entsprechende Forschungs- und Entwicklungsleistungen als solche stärker als bisher gewichtet und gewürdigt werden und damit auch karrierefördernd wirken können. Andernfalls ist abzusehen, dass gerade hochqualifiziertes Personal auf diesem Gebiet der Wissenschaft verloren geht und auf anderweitige, lukrativere Positionen außerhalb des Hochschulsektors ausweicht.

Die wissenschaftlichen Fachgesellschaften

Da die Leistungsfähigkeit von VREs und dementsprechend auch von TextGrid wesentlich von ihrer fachspezifischen Ausgestaltung abhängt, kommt hier den Fachgesellschaften als Vermittlungsinstanzen besondere Bedeutung und eine neue, attraktive Aufgabenstellung zu. Sie könnten und sollten beispielsweise im Rahmen ihrer (Jahres-)Tagungen Foren anbieten, auf denen sich die neuen Möglichkeiten standortunabhängiger kooperativer Forschung präsentieren und kritisch diskutieren lassen.

¹¹ Der Wissenschaftsrat empfiehlt die Verknüpfung von Forschungsinfrastrukturen mit Förderkonzepten für den wissenschaftlichen Nachwuchs in seinen Empfehlungen zu „Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften“ vom 28.1.2011.

Die Wissenschaftsorganisationen

Auch sie sind gefordert, hier werbend und unterstützend zu wirken, Möglichkeiten struktureller und finanzieller Unterstützung – z.B. als etwaige Trägerinstitutionen für VREs – zu prüfen, auszuschöpfen und ihren Einfluss als Akteure von Gewicht und Reichweite sowohl gegenüber ihren Mitgliedern und Mitgliedsinstitutionen als auch gegenüber der politischen Ebene geltend zu machen. Das betrifft mit Blick auf TextGrid insbesondere Organisationen mit (zumindest teilweise) geisteswissenschaftlicher Ausrichtung.

Die Wissenschaftspolitik

Insbesondere in Gestalt des BMBF, aber auch einer Reihe von Landesministerien, z.B. in Niedersachsen, gehört die Wissenschaftspolitik zu den treibenden Kräften bei der Propagierung, Einrichtung und Förderung von VREs. Dabei muss es jetzt darum gehen, dass die bereits verfolgten Initiativen verstärkt und verbreitet Sogwirkung entfalten und andernorts dazu motivieren, entsprechende eigene Möglichkeiten zu sondieren und auszuschöpfen. Das betrifft nicht zuletzt die Überwindung und Beseitigung föderaler Schranken und Hemmnisse, die die nötige Kooperation über Ländergrenzen hinweg nach wie vor stark behindern. Der Wissenschaft sind nationale ebenso wie internationale Grenzziehungen generell wesensfremd. Das gilt auch für VREs und namentlich für TextGrid. Was das Verhältnis von Bund und Ländern auf diesem Feld betrifft, so scheinen sich hier erfreulicherweise neue Möglichkeiten des Zusammenwirkens abzuzeichnen.

Die Förderorganisationen, insbesondere die privaten Stiftungen

Auch hier lässt sich auf wichtige Initiativen verweisen, die insbesondere von der DFG, aber auch von großen privaten Förderorganisationen gestartet und getragen wurden/werden. Dessen ungeachtet werden Drittmittelgeber ihre Unterstützungsmöglichkeiten, -richtlinien und -verfahren mit Blick auf neue, spezifische Ressourcenbedarfe und Finanzierungsnotwendigkeiten von VREs überdenken und etwaige rigide Beschränkungen auf kurzfristige Förderzeiträume zugunsten längerfristiger Finanzierungsangebote künftig soweit als möglich aufbrechen müssen. Erfreulicherweise gibt es in dieser Hinsicht inzwischen ebenfalls verstärkte Bemühungen. Auch die Grenzen zwischen institutioneller Grundfinanzierung und projektbezogener Zusatzförderung bedürfen in diesem Kontext einer den neuen Anforderungen besser Rechnung tragenden Bestimmung. Das alles ändert allerdings nichts daran, dass Drittmittelgeber bereits angesichts ihrer begrenzten finanziellen Möglichkeiten und ihrer generellen Ausrichtung auf befristete, impulsgebende Hilfen als Dauerfinanziers von VREs prinzipiell nicht in Betracht kommen. Hier bleiben die Haushalte der öffentlichen Hände gefordert.

TextGrid und die an seiner Entwicklung unmittelbar Beteiligten werden diese im wissenschaftlichen und wissenschaftspolitischen Umfeld notwendigen Entwicklungen, Veränderungen und Anpassungen auch weiterhin nach Kräften befördern. Dazu sind auf der Basis des angelaufenen Echtbetriebs von TextGrid 1.0 bereits Kontakte, Sondierungsgespräche und (Vor-)Verhandlungen mit verschiedenen Akteuren geplant und zum Teil auch schon aufgenommen worden (so u.a. mit dem Niedersächsischen Ministerium für Wissenschaft und Kultur zu Fördermöglichkeiten im Rahmen des „Niedersächsischen Vorab“ der VolkswagenStiftung). Allerdings lassen sich diese Gespräche und Verhandlungen bis zum Abschluss der laufenden Förderphase Ende Mai 2012 naturgemäß weder flächendeckend durchführen, noch sind bis dahin bereits substantielle Verbesserungen des Umfeldes zu erwarten – dies umso weniger, als unangemessener Zeitdruck Gespräche und Verhandlungen auf diesem komplexen Feld eher belasten und den Erfolg

gefährden würde: Schnelle, aber allenfalls 'zweitbeste' Lösungen sollten hier keinesfalls angestrebt, sondern die an dieser Stelle zweifellos besonders 'harten Bretter' in bewährter Manier stark und langsam, mit Leidenschaft und Augenmaß gebohrt werden.

7. Beschreibung der Arbeitspakete

Im Folgenden werden die Arbeitspakete im Einzelnen inklusive der Nennung von Meilensteinen (M) und Reports (R) beschrieben. Eine zeitliche Gesamtübersicht aller Meilensteine und Reports findet sich als Gantt-Diagramm in Anlage II.

AP 1: Betriebsmodell

Nummer des Arbeitspakets	AP 1
Titel	Betriebsmodell
AP-Leitung	Institut für Deutsche Sprache (IDS), Mannheim
Start-Monat	Juni 2012 (M1) bis Mai 2015 (M36)
Gesamtanzahl PM	45

Ziele (Kurzbeschreibung)
Die grundlegenden rechtlichen, organisatorischen, ökonomischen und technischen Rahmenbedingungen für den nachhaltigen Betrieb der Virtuellen Forschungsumgebung TextGrid stehen im Zentrum dieses Arbeitspakets. Neben dem Geschäftsbetrieb des zu gründenden, zunächst befristet angelegten Vereins wird in diesem AP auch eine Rechtsform für den Dauerbetrieb der VRE gesucht. Darüber hinaus geht es um die kontinuierliche Bearbeitung der juristischen Fragen, die im Projektverbund auftreten, sowie um die präzise Berechnung von Kostenszenarien für den Produktivbetrieb. Auf der technischen Ebene kann TextGrid nur weiterbestehen, wenn Entwicklungsperspektiven vom jetzigen Stand der Technik zu neuen sich abzeichnenden Standards webbasierter Computertechnik exploriert und durch technologische Machbarkeitsstudien und Blaupausen vorbereitet werden.

Beschreibung des Arbeitspakets

Der nachhaltige Betrieb einer VRE wie TextGrid muss juristisch fundiert und wirtschaftlich abgesichert erfolgen. Generalisierbare Lösungen für die sich dabei stellenden Probleme werden von den am AP 1 maßgeblich beteiligten Partnern, dem IDS in Mannheim und der Universität Würzburg, erarbeitet.

Die zentralen Aufgaben dieses Arbeitspakets bestehen im Betrieb einer Geschäftsstelle für den zu Beginn der Förderphase zu gründenden eingetragenen und gemeinnützigen Verein „**TextGrid – Verein zum nachhaltigen Betrieb einer digitalen Forschungsinfrastruktur in den Geisteswissenschaften e.V.**“, in der Erarbeitung eines Vorschlags für eine tragfähige Rechtsform, die für einen Dauerbetrieb von TextGrid die besten rechtlich-organisatorischen Rahmenbedingungen

schafft, sowie in der Kalkulation der Kosten, die bei einem kontinuierlichen und nachhaltigen Betrieb der Virtuellen Forschungsumgebung TextGrid zu erwarten sind. Die Kostenkalkulation wird betriebswirtschaftlich fundiert durchgeführt werden, wobei insbesondere auch die unterschiedliche Intensität der Infrastruktur-Nutzung in entsprechenden Kostenszenarien durchgerechnet werden soll.

Die Geschäftsstelle des Vereins wird auch mit den rechtlichen Fragen befasst sein, die sich bei einem praktischen Betrieb stellen werden; denn die bisherigen Erfahrungen zeigen, dass eine Vielzahl juristischer Probleme identifiziert werden muss. Für textbasierte Forschungen – auch an historischen Corpora – gibt es einen hohen Klärungsbedarf im Hinblick auf Lizenzmodelle sowie auch den Umgang mit Urheber-, Leistungsschutz- und Datenschutzrechten.

Technisch gesehen kann TextGrid nur dann nachhaltig arbeiten, wenn Entwicklungsperspektiven für die Werkzeuge vom jetzigen Stand der Technik zu neuen sich abzeichnenden Web-Standards exploriert werden und durch Machbarkeitsstudien und Blaupausen vorbereitet werden.

AP 1.1: Rechtsform, Kostenkategorien

Leitung	Institut für Deutsche Sprache (IDS), Mannheim
Gesamtanzahl PM	18

Basierend auf den Arbeiten im bisherigen AP 3 „Strukturelle und organisatorische Nachhaltigkeit“ von TextGrid wurde für die Übergangsphase in den Dauerbetrieb die Einrichtung eines Vereins zu Beginn der neuen Förderphase vorgeschlagen. In diesem Zusammenhang wurden entsprechende Statuten und die rechtliche Reglementierung für den technischen Betrieb erarbeitet, die zu Beginn der Übergangsphase implementiert werden können. Diese Maßnahme hat zum Ziel, dem Betrieb zunächst eine andernorts bereits bewährte, flexible, demokratisch legitimierte Struktur und eine rechtliche Grundlage als gemeinnützige Organisation zu geben¹². Diese Rechtsform soll den technischen Betrieb in seiner Erprobung in der Übergangsphase begleiten und in der Praxisanwendung ihrerseits auf ihre Akzeptanz beim Nutzer getestet werden. AP 1.1 wird daher die Erfahrungen und Prozesse der täglichen Vereinsarbeit dokumentieren und auswerten sowie die Tragfähigkeit dieser Rechtsform und die Akzeptanzkriterien bei potentiellen Nutzern evaluieren.

Darüber hinaus sollen anhand dieser Ergebnisse verschiedene alternative rechtliche Gestaltungen untersucht werden. Ziel ist die Institutionalisierung von TextGrid, die den Gegebenheiten des technischen Betriebs gerecht wird und den Forderungen nach Rechtssicherheit und demokratischer Einflussnahme seitens der Nutzer Rechnung trägt. Konkret gilt es zu prüfen, ob die jetzt gewählte Rechtsform des Vereins verstetigt oder eine Alternative gefunden werden muss.

Schließlich sind vor allem detaillierte Kostenkalkulationen kontinuierlich zu erstellen, da zu erwarten steht, dass mit einer größer werdenden Nutzergemeinschaft für TextGrid neue, bisher noch nicht definitiv abschätzbare Anforderungen und entsprechend größerer Aufwand verbunden sein werden. AP 1.1 wird daher anfallende Kosten des täglichen Betriebs erfassen, um zum Ende

¹² Vgl. z.B. das „Deutsche Forschungsnetz e.V.“ (DFN) oder die „Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V.“ (TMF)

der Förderphase empirisch gestützte Kostenkalkulationen und Kostenmodelle für eine nachhaltige Finanzierung bereitzustellen.

AP 1.2: Verein, Geschäftsstelle

Leitung	Institut für Deutsche Sprache (IDS), Mannheim
Gesamtanzahl PM	9

Der TextGrid-Verein benötigt eine Geschäftsstelle, die alle den Verein betreffenden organisatorischen Tätigkeiten übernimmt. Im Unterschied zu dem in AP 7 konzentrierten Projektmanagement beschränkt sich die Geschäftsstelle ausschließlich auf die unmittelbar den Verein selbst betreffenden organisatorischen Aufgaben, arbeitet dabei aber eng mit dem Projektmanagement zusammen. Die Geschäftsstelle untersteht satzungsgemäß dem Vorstand des Vereins und unterstützt diesen bei der Erfüllung seiner Aufgaben. Zu den wesentlichen Tätigkeitsfeldern der Geschäftsstelle gehören insbesondere die administrative Vereinsarbeit, die Bearbeitung rechtlicher Fragestellungen, Hilfestellungen bei der Einberufung der Mitgliederversammlungen und die Unterstützung des Vorstandes bei der Abfassung der Jahresberichte etc.

AP 1.3: (Technische) Integration und zukünftiges Weiterentwicklungspotential, Zukunftsfähigkeit

Leitung	Universität Würzburg
Gesamtanzahl PM	18 (zusätzl. stud. Hilfskraft)

Die technische Entwicklung innerhalb wie außerhalb der eHumanities schreitet fort, und damit verändern sich auch die Anforderungen der Nutzer. Auf technologischer Ebene kann das Weiterbestehen von TextGrid daher nur gesichert werden, wenn ein Plan existiert, der entsprechende Entwicklungen berücksichtigt und einbindet. Dies bezieht sich insbesondere auf die Entwicklung hin zu mehr webbasierten Schnittstellen, einschließlich einer besseren Einbindung externer Tools, und zu besseren Möglichkeiten der internationalen Kollaboration auf technischer Ebene.

AP 1.3 wird daher ausgewählte aktuelle und künftige Techniken, Frameworks, einschlägige Initiativen sowie Entwicklungen in Bibliotheken und andernorts auf ihre Anwendbarkeit in TextGrid hin evaluieren. Die Ergebnisse werden fortlaufend im öffentlichen TextGrid-Wiki dokumentiert und, wo dies sinnvoll und angemessen erscheint, werden kleinere Demonstratoren oder Strategiepapiere zu einer Umsetzung in TextGrid entwickelt.

Reports (R) & Meilensteine (M):

<i>Bezeichnung</i>	<i>Monat</i>	<i>Beschreibung</i>
M 1.2.1	1-36	Geschäftsstelle Verein (laufend)
R 1.1.1	18	Identifikation von Kostenfaktoren und Erarbeitung von Kostenmodellen
R 1.1.2	24	Erfahrungsbericht Rechtsform: Praxisbewährung und Nutzeranforderungen

R 1.3.1	24 und 36	Machbarkeitsstudien zu neuen technischen Entwicklungen für TextGridLab
R 1.1.3	30	Finalisierung Kosten- und Finanzierungsmodelle
M 1.3.2	30	Blaupausen zur Integration zukunftsfähiger Entwicklungen

Beziehungen bestehen insbesondere zu folgenden APs:

- AP 2.1: Umgang mit großen Nutzerzahlen und heterogenen Nutzergruppen sowie Nutzerverwaltung (Theorie und Praxis)
- AP 2.3: Finanzierungsmodelle und -wege
- AP 3: Nutzer-Kommunikation
- AP 7: Projektmanagement und Öffentlichkeitsarbeit

AP 2: Prozessorganisation

Nummer des Arbeitspakets	AP 2
Titel	Prozessorganisation
AP-Leitung	Max-Planck-Institut für Wissenschaftsgeschichte (MPI WG), Berlin
Start-Monat	Juni 2012 (M1) bis Mai 2015 (M36)
Gesamtanzahl PM	27 (und Unterauftrag ProWiss)

Ziele (Kurzbeschreibung)
Das Arbeitspaket hat die Etablierung von Prozessen zur Verwaltung großer heterogener Benutzergruppen zum Gegenstand. Dazu gehört die Entwicklung eines Beratungskonzeptes für unterschiedliche Nutzergruppen anhand von exemplarischen Workflows aus der wissenschaftlichen Nutzung ebenso wie der Aufbau eines elektronischen Market Space für die Toolentwicklung, die Entwicklung von Richtlinien und Qualitätsstandards für Plugins bzw. ergänzende Tools für TextGridLab und TextGridRep, und schließlich die Entwicklung und Etablierung eines Finanzierungsmodells.

Beschreibung des Arbeitspakets

Es werden Verfahren entwickelt und implementiert, die es ermöglichen, dass TextGridLab und TextGridRep eine deutlich größere Zahl von Nutzern als bisher wirksam unterstützen können. Ein weiterer Schwerpunkt ist die Ausweitung der Anwendergruppen von philologisch-editorisch arbeitenden Wissenschaftlerinnen und Wissenschaftlern auf interdisziplinär und kollaborativ arbeitende Nutzergruppen aus unterschiedlichen geisteswissenschaftlichen Disziplinen einschließlich der Entwicklung der dazu erforderlichen Werkzeuge.

Während TextGrid die Anforderungen, die sich aus der philologischen und editionswissenschaftlichen Arbeit ergeben, bereits jetzt in weiten Bereichen erfüllt, muss es für die Betriebsphase zu einem System ausgeweitet werden, das von allen geisteswissenschaftlichen Disziplinen für die Analyse und Annotation von Schriftquellen insgesamt verwendet werden kann.

Daraus ergeben sich neue organisatorische und technische Anforderungen. Es müssen größere Gruppen von Nutzern verwaltet und gemäß den spezifischen Anforderungen aus den unterschiedlichen disziplinären Kontexten entsprechende Nutzungsszenarien entwickelt und realisiert werden. Das erfordert Veränderungen der bestehenden Workflows, die Weiter- und Neuentwicklung von Tools als Ergänzung der bestehenden Umgebung sowie die Entwicklung entsprechender Schulungskonzepte. Neuentwicklungen sollten hierbei mittel- bis langfristig weitgehend von den Nutzergruppen selbst durchgeführt werden – mit Unterstützung durch den Projekt-eigenen Support. Doppelentwicklungen sind dabei unbedingt zu vermeiden. Auch muss die Qualität der Entwicklungen geprüft und beurteilt werden. Analog zu den Modellen von OpenSource-Projekten wie Drupal oder Eclipse wird daher ein Market Space für Tools aufgebaut, in dem je nach qualitativer Bewertung der Tools diese als geprüfte Erweiterungen des Kerns von TextGrid oder als optionale bzw. experimentelle Ergänzungen angeboten werden. Die im Market Space verfügbaren Erweiterungen werden darüber hinaus nach Anwendungsszenarien gruppiert angeboten, z.B. in Form von Beispiel-Workflows und Nutzungsbeispielen, um den Anwendern die Orientierung zu erleichtern.

AP 2.1: Umgang mit großen Nutzerzahlen und heterogenen Nutzergruppen sowie Nutzerverwaltung (Praxis und Theorie)

Leitung	Technische Universität Darmstadt
Gesamtanzahl PM	9

Im Zusammenhang mit der boomenden Entwicklung der eHumanities ist mit einem raschen Anwachsen der Nutzerzahlen bei gleichzeitiger Diversifizierung der Anforderungen und Bedarfe zu rechnen. Eine zielführende Nutzung neuer eHumanities-Ansätze und die Integration heterogener Nutzergruppen kann nur dann gelingen, wenn die Angebote und Perspektiven von TextGrid transparent gemacht und – angepasst an die spezifischen Bedürfnisse der Fachdisziplinen – den Communities vermittelt werden. Die Schaffung einer entsprechenden Informationsbasis, die Identifizierung spezifischer Anforderungen größerer Nutzergruppen und in der Folge die Bildung und aktive Gestaltung von Communities sind der Schlüssel für eine nutzerzentrierte Weiterentwicklung des Angebotes. Daran wird AP 2.1 anknüpfen und entsprechende Maßnahmen erarbeiten.

Die Förderung der eHumanities und die Schaffung von Kommunikationsstrukturen müssen auf der Grundlage ermittelter Anforderungen von Nutzergruppen erfolgen und zugleich zu deren Formierung bzw. Erweiterung beitragen. Die Identifizierung und Bündelung der wachsenden Communities und ihrer spezifischen Bedürfnisse ist eine Aufgabe, die im AP 2.1 als erste bearbeitet werden muss, da sie die notwendige Grundlage für den weiteren Auf- und Ausbau bildet. Die Ermittlung von Anforderungen und – auf dieser Basis – die Optimierung der Angebote und Strukturen sind weitere zentrale Aspekte der Community-Bildung. Ziel ist es, vorhandene Dienste und Services transparent und bekannt zu machen und dabei spezifischen Bedürfnissen der Fachdisziplinen Rechnung zu tragen. Dadurch sollen zum einen die Nutzerbasis verbreitert und zum anderen die Nutzungsmöglichkeiten optimiert werden (Bündelung von Ressourcen). Die in AP 3.3 adressierten spezifischen Coaching- und Schulungsmaßnahmen professionalisieren diesen Prozess.

Ein weiterer Aspekt ist der Ausbau der Strukturen, die einen Austausch zwischen unterschiedlichen Nutzergruppen ermöglichen. Die fachwissenschaftliche Herangehensweise an Editionsprojekte zum Beispiel mit dem Schwerpunkt einer wissenschaftshistorischen Analyse eines

Dokumentes führt zu anderen Auszeichnungen eines Textes als dessen Bearbeitung aus philologischer Sicht. Fachwissenschaftlich gesehen sind die semantische Analyse eines Textes, die Herausarbeitung seiner Struktur, die Analyse von technischen Terminologien und der Aufweis von inhaltlichen Beziehungen zu andere Quellen von zentraler Bedeutung, während der editionswissenschaftliche Blickwinkel auf eine möglichst genaue Rekonstruktion eines Textes und seiner Veränderungen im Laufe seiner Entstehungs- und Rezeptionsgeschichte abzielt. Beide Herangehensweisen sind bisher zwar bereits eng miteinander verbunden, eine kollaborative elektronische Edition kann jedoch die Zusammenarbeit zwischen diesen beiden Gruppen auf eine neue, wesentlich dynamischere Ebene heben. Die bestehende TextGrid-Infrastruktur bietet technisch gesehen für diese Verbindung eine gute Grundlage, die Beteiligung von Fachwissenschaftlern bei der Textanalyse lässt jedoch noch zu wünschen übrig. Hierzu müssen zwei Bereiche ausgebaut werden: auf der technischen Ebene müssen die bisherigen Auszeichnungsmöglichkeiten von Texten um Komponenten ergänzt werden, die eine weitergehende semantische Auszeichnung, etwa durch Nutzung fachspezifischer XML-Standards sowie die Anbindung an Ontologien für die Beschreibung der Relationen von Textteilen und Texten untereinander, ermöglichen; organisatorisch gesehen bedarf es der Erweiterung um eine Kommunikationsplattform, mit deren Hilfe nicht nur die Analyse von Texten systematisch koordiniert werden kann, sondern sich auch alle kollaborativen Prozesse wirksam unterstützen lassen.

AP 2.2: Verteilte / Zentrale Entwicklungen

Leitung	Max-Planck-Institut für Wissenschaftsgeschichte (MPI WG), Berlin
Gesamtanzahl PM	18

Sowohl in der Übergangsphase als auch im späteren produktiven Dauerbetrieb können von TextGrid selbst keine oder in nur sehr begrenztem Maße eigene Entwicklungen durchgeführt werden, die neue Features ausgehend von entsprechenden disziplinären Anforderungen in TextGrid einbetten. Entwicklungen müssen daher im Wesentlichen durch Dritte erfolgen, die über das dafür notwendige domänenspezifische Wissen verfügen. TextGrid selbst kann dabei nur unterstützend wirken. Umso dringlicher ist dann allerdings eine zentrale Qualitätssicherung auch von externen Plugins sowie die systematische Dokumentation von Erweiterungen, die Doppelentwicklungen vermeiden hilft. Dafür müssen geeignete Strukturen erarbeitet und in der Entwicklergemeinschaft etabliert werden. Im Zentrum stehen hierbei die Entwicklung eines Market-Space für Tools und Plugins sowie die Erarbeitung von Qualitätskriterien und entsprechenden Richtlinien.

Dazu ist vorgesehen, die sich aus der Analyse der Nutzergruppen (AP 2.1) ergebenden Anforderungen zu sammeln und daraus Feature-Listen zu entwickeln, die mit bestehenden Tools verglichen werden sowie Vorschläge für Anpassungen bzw. Neuentwicklungen zu erarbeiten und diese dann an die Entwicklergemeinschaft zurückzugeben. Auch hier gilt es, neue Entwicklungen anzuregen und Dopplungen zu vermeiden. In begrenztem Umfang müssen allerdings auch Entwicklungen, insbesondere verallgemeinernde Erweiterungen, sowie Anpassungen an neue Rechnergenerationen und Anbindungen an zukünftige Infrastrukturen für Storage und Dissemination zentral übernommen werden. Dafür sind wiederum praktikable Finanzierungsmodelle zu erarbeiten (AP 2.3).

AP 2.3: Finanzierungsmodelle und -wege

Leitung	ProWiss / Prof. A. Horstmann
Gesamtanzahl PM	via Unterauftrag Staats- und Universitätsbibliothek Göttingen (SUB)

Für TextGrid kommt – das war bereits mehrfach betont worden – kein genuin kommerzielles Ertragsmodell in Frage. So wird TextGrid seine Leistungen selbst bei boomender Nachfrage ohne gesicherte Grundfinanzierung kaum erbringen können. Von daher bleibt auch für den Dauerbetrieb nur, alle verfügbaren Finanzierungsquellen soweit als möglich zu kombinieren. Dabei hängen die betreffenden Optionen wiederum ihrerseits maßgeblich von der jeweils gewählten Organisationsform ab (Verein, gemeinnützige GmbH, unselbstständige Abteilung innerhalb einer größeren Institution etc.). Prinzipiell denkbar sind mithin neben einer gesicherten Grundfinanzierung (durch Bund und Länder) Beiträge von Vereinsmitgliedern, Gebühren und Leistungsentgelte von Nutzern sowie eine projektbezogene Förderung, gegebenenfalls unter Nutzung von Overheads zur anteiligen institutionellen Finanzierung. Die verschiedenen Varianten sind, ausgehend von der für die dreijährige Übergangsphase vorgesehenen Organisationsform eines gemeinnützigen Vereins, im Einzelnen durchzurechnen, durchzuspielen und auf Praxistauglichkeit zu testen. Dazu gehört es nicht nur festzulegen, welche Leistungen künftig in welcher Höhe kostenpflichtig bzw. ohne Entgelt verfügbar sein sollen und können, sondern auch, auf der Ebene von Wissenschaftspolitik, -administration und -förderung zu sondieren, ob und inwieweit sich hier gegenüber der bisherigen Praxis deutlich weitergehende Finanzierungsspielräume eröffnen könnten – nicht zuletzt durch Überwindung föderaler Schranken und Hemmnisse sowie insbesondere ein engeres finanzielles Zusammenwirken von Bund und Ländern. Die Ergebnisse werden in den Report des Arbeitspaketes AP 1.1 (und hier in R 1.1.1 und R 1.1.3) einfließen, der abschließende Report 1.1.3 wird dabei mit entsprechenden Empfehlungen vorgelegt werden. Eine enge Beobachtung des vom BMBF eingesetzten Beratergremiums und die Umsetzung der dortigen Diskussionen und Empfehlungen sind dabei selbstverständlich.

Reports (R) & Meilensteine (M):

<i>Bezeichnung</i>	<i>Monat</i>	<i>Beschreibung</i>
M 2.1.1	1-36	(Laufende) Kartierung der heterogenen Nutzercommunity
R 2.2.2	1-36	(Laufende) Dokumentation von externen Plugins
M 2.2.1	10, 20, 30	Entwicklerworkshop mit externen Projektpartnern in Kooperation mit AP 5.4
M 2.1.2	12	Einrichtung eines Nutzerforums und dauerhafte Pflege
M 2.2.2	18	Bereitstellung des Market Space

Beziehungen bestehen insbesondere zu folgenden APs:

- AP 1.1: Rechtsform, Kostenkategorien
- AP 3: Nutzer-Kommunikation
- AP 5.4: Koordination von Weiterentwicklung und Satellitenprojekten

AP 3: Nutzer-Kommunikation

Nummer des Arbeitspakets	AP 3
Titel	Nutzer-Kommunikation
AP-Leitung	Technische Universität Darmstadt
Start-Monat	Juni 2012 (M1) bis Mai 2015 (M36)
Gesamtanzahl PM	36

Ziele (Kurzbeschreibung)
Ziel des Arbeitspakets ist der Aufbau einer informierten, stabilen Nutzergemeinschaft, die zu einem aktiven Träger der Infrastruktur wird. Die dafür erforderlichen Maßnahmen und Aktivitäten setzen auf unterschiedlichen Ebenen an: Neben klassischen Dokumentationen und Hilfetexten, Workshops, Schulungen und Seminaren sollen in diesem AP methodische Reflexionen der Anbieter-Nutzer-Interaktion sowie der Wechselwirkungen von eScience- und eLearning-Ansätzen ihren Raum erhalten.

Beschreibung des Arbeitspakets

Virtuelle Forschungsumgebungen wie TextGrid knüpfen an traditionelle Methoden und Forschungsfragen an und unterstützen auf innovative, effiziente Art und Weise diese Forschungsarbeit, darüber hinaus erschließen sie aber auch neue Dimensionen wissenschaftlichen Arbeitens. Um diese hohen Erwartungen zu erfüllen, bedarf es intensiver Nutzerkontakte und Nutzerschulungen (Studierende, Fachwissenschaftler, Entwickler). Dabei sind die Vermittlungs- und Kommunikationskanäle durchaus multidirektional angelegt. Das bedeutet, dass ein Konzept zur nachhaltigen Einbindung der wissenschaftlichen Nutzer in die Anwendung, Pflege und Weiterentwicklung installiert, evaluiert und institutionalisiert werden muss.

Die Nutzergruppen sind in vielerlei Hinsicht heterogen: Sie bringen unterschiedliche Erfahrungshorizonte in der Anwendung avancierter Technologien mit, sie basieren auf unterschiedlichen Ausbildungsniveaus und sie verfolgen unterschiedliche Ziele beim Einsatz von eHumanities-Methoden. Was die Einstellungen betrifft, reicht das Spektrum von größter Skepsis bis zu höchster Begeisterung. Erforderlich ist daher ein koordiniertes Maßnahmenbündel, das möglichst breite Nutzerschichten anspricht, fortbildet und so in das 'TextGrid-Ökosystem' integriert, sodass ausgehend von kleinen, fachspezifischen Initiativgruppen eine stetig wachsende Zahl von aktiven Trägern der Infrastruktur entsteht. Diese Maßnahmen müssen stark fachwissenschaftlich ausgerichtet und disziplinspezifisch gestaltet sein, und sie müssen vielfältige traditionelle und moderne Kommunikations- und Vermittlungstechniken bedienen. Dazu gehören u.a. (Informations-)Vorträge, am Bedarf der Nutzergruppen orientierte, spezifische Workshops und Schulungen, klassische ausführliche Dokumentationen und Einführungen, Tutorials, Sandboxes und Webinare. Die Veranstaltungen müssen ebenfalls Formate und Raum für die Nutzerintegration und für Nutzeraktionen bieten und jeweils evaluativ begleitet werden, um die Ergebnisse für die Weiterentwicklung der Nutzer-Kommunikation fruchtbar machen zu können. Eine Kooperation des Instituts für Sprach- und Literaturwissenschaft mit dem Darmstadt eLearning Center bietet sich

dabei an. Das eLearning Center verfügt über sehr gute Erfahrungen bei der Dissemination und Vermittlung neuer Technologien für wissenschaftliche Anwender und ist sowohl am Diskurs eScience / eLearning als auch daran interessiert, sein Portfolio hier weiter auszubauen und fortzuentwickeln.

AP 3.1: Dokumentation, Nutzer-Information

Leitung	Technische Universität Darmstadt
Gesamtanzahl PM	18

Neben der technischen Dokumentation für Entwickler muss eine Dokumentation für Anwender als Basis aller weiteren Disseminierungsbemühungen entstehen, die verlässliche und verständliche Informationen zu TextGrid und konkrete Anleitungen zur Nutzung der Werkzeuge gibt. Use Cases, Demomaterial und Sandboxes ergänzen diese Basisinformationen. Zu berücksichtigen ist dabei der quer durch alle Nutzerschichten hinweg geäußerte Wunsch nach einer Dokumentation in deutscher Sprache.

AP 3.2: eHumanities und TextGrid in universitären Curricula

Leitung	Technische Universität Darmstadt
Gesamtanzahl PM	Eigenanteile DARIAH-DE, AP 2

Der wissenschaftliche Nachwuchs in den geistes- und kulturwissenschaftlichen Disziplinen muss konsequent an die Möglichkeiten und Methoden computerunterstützten wissenschaftlichen Arbeitens herangeführt werden. In diesem Zusammenhang kann auf die Ergebnisse des DARIAH-DE-Projektes in Arbeitspaket 2 zurückgegriffen werden, mit dem eng zusammengearbeitet wird. Daher sind hier keine Personalmittel erforderlich. Ein erstes Ergebnis, an dem Partner von DARIAH-DE maßgeblich beteiligt waren, stellt die Studiengangsbroschüre „Digitale Geisteswissenschaften“¹³ dar.

¹³ Vgl. <http://www.cceh.uni-koeln.de/dh-degrees-2011>

AP 3.3: Workshops, Schulungen, Nutzer-Interaktion

Leitung	Technische Universität Darmstadt
Gesamtanzahl PM	18 (zusätzlich student. Hilfskraft)

Hier werden basierend auf den Erhebungen in AP 2.1 und in engem Austausch mit den Nutzerbedürfnissen und Nutzerrückmeldungen Konzepte und Formate entwickelt und erprobt, die eine nachhaltige Vermittlung der Anforderungen und Möglichkeiten von TextGrid als Baustein der eHumanities-Entwicklungen gewährleisten.

Am 22./23.2.2012 fand das erste TextGrid-Nutzertreffen an der TU Darmstadt statt, die Resonanz war mit über 50 Teilnehmerinnen und Teilnehmern sehr hoch. Vertreten waren bereits produktiv mit TextGrid arbeitende Forschungsverbünde wie auch perspektivische Nutzer mit einem insgesamt hohen Anteil an Projekten, die eine Beteiligung an der (Weiter-)Entwicklung vorhandener Tools planen. Neben der Präsentation aktueller Neuerungen und Muster-Workflows wurde den Vorstellungen von Nutzerprojekten, offenen Gesprächsrunden und kleineren Arbeitsgruppen möglichst viel Raum gegeben. Dieses Konzept hat sich zum einen in Bezug auf die Initiierung des Nutzerdialogs (insbesondere auch der Nutzer untereinander), auf die Identifizierung möglicher Synergien und Kooperationen sowie in Bezug auf die Ermittlung weiterer Bedarfe als ausgesprochen fruchtbar erwiesen. Dezidiert wurde der Wunsch nach Beibehaltung bzw. Ausweitung dieses offenen, begleiteten Konzeptes, bei dem der Nutzer im Mittelpunkt steht, geäußert. In Zukunft wird dieser Prozess auch durch Nutzerforen und entsprechende Emaillisten etc. begleitet (AP 2.1), was den persönlichen Austausch jedoch nicht ersetzen kann. Bezüge ergeben sich ferner zu AP 2.2, wo die Koordination der Entwicklergemeinschaft koordiniert und unterstützt wird, sowie insbesondere zu AP 5.4, das Schulungen für Entwickler konzipiert und durchführt. Das Gesamtkonzept und –monitoring aller Schulungen wird dagegen in diesem AP gebündelt. Neben den Nutzertreffen sollen wie bisher flexibel Schulungen für konkrete Projekte und Nutzergruppen angeboten werden. Erfahrungen aus den Seminaren des Master of Linguistic and Literary Computing an der TU Darmstadt sowie ggf. studentischen Master-Projekten und Qualifizierungsarbeiten fließen in die methodisch-theoretische Reflexion von fachwissenschaftlichen und didaktischen Aspekten sowie in die Weiterentwicklung der Nutzerpflege mit ein. Neben der Kooperation mit dem Darmstadt eLearning Center ist auf den von Prof. Nina Janich am Institut aufgebauten Schwerpunkt 'Wissenschaftskommunikation – SciCoRe' zu verweisen, der einen fruchtbaren Austausch in diesem Bereich verspricht.

Die Perspektiven zur Integration der unterschiedlichen Gruppen in die TextGrid-Community sind daher nach den bisherigen Erfahrungen aus der Sicht der Antragsteller sehr viel versprechend, die Signale aus den Communities ermutigend.

Reports (R) & Meilensteine (M)

<i>Bezeichnung</i>	<i>Monat</i>	<i>Beschreibung</i>
R 3.1.1	6-36	Einrichtung von Sandboxes, Entwicklung von Use Cases, Zusammenstellung von Demo-Materialien (laufend aktualisiert)
R 3.1.2	6-36	Dokumentation für Anwender (laufend aktualisiert)
M 3.3.1	6, 12, 18, 24, 30, 36	Nutzertreffen mit je spezifischen Schwerpunkten

R 3.3.1	20, 36	Report über die Erfahrungen aus den Nutzer- schulungen, Empfehlungen für die weitere Aus- gestaltung des Konzepts
---------	--------	---

Beziehungen bestehen zu folgenden APs:

- AP 2.1: Umgang mit großen Nutzerzahlen und heterogenen Nutzergruppen sowie Nutzerverwaltung (Theorie und Praxis)
- AP 2.2 Verteilte / Zentrale Entwicklungen
- AP 5.3: Usability
- AP 5.4: Koordination externer Weiterentwicklung
- AP 6.1: Integriertes Monitoring aller Arbeitspakete
- AP 6.2: Nutzerakzeptanz (inkl. Befragungen)
- AP 7: Organisatorischer Austausch

AP 4: Betrieb und Softwarepflege Repository

Nummer des Arbeitspakets	AP 4
Titel	Betrieb und Softwarepflege Repository
AP-Leitung	Staats- und Universitätsbibliothek Göttingen (SUB)
Start-Monat	Juni 2012 (M1) bis Mai 2015 (M36)
Gesamtanzahl PM	112,5

Ziele (Kurzbeschreibung)
Ziel des Arbeitspakets ist der Betrieb des fachwissenschaftlichen Langzeitarchivs für geisteswissenschaftliche Forschungsdaten und dessen Softwarepflege; dazu zählen auch die dem TextGridLab als Schnittstelle zum TextGridRep dienenden Middleware-Services. Das bestehende TextGridRep muss für höhere Nutzerzahlen und einen Produktivbetrieb im Hinblick auf Performanz/Skalierbarkeit und Sicherheit verbessert werden; für eine Evaluation sind regelmäßige Lasttests notwendig. Um den Produktivbetrieb effizient zu gestalten, müssen technische Prozesse optimiert werden. Dazu zählt beispielsweise die Bereitstellung großer Datensammlungen in TextGrid (Ingest), die aktuell einen hohen manuellen Aufwand erfordert und die in AP4 dokumentiert und weiter automatisiert wird, sowie die projektspezifische Präsentation dieser Daten.

Beschreibung des Arbeitspakets

Dieses Arbeitspaket hat den Betrieb und die Pflege des TextGrid Repository zum Gegenstand, das die langfristige Verfügbarkeit und Zugänglichkeit der geisteswissenschaftlichen Forschungsdaten sowie deren optimale Vernetzung innerhalb und außerhalb der Forschungsumgebung garantiert. Dies setzt eine stabile Infrastruktur der Middleware voraus, die einen ausfallsicheren Produktivbetrieb auch bei hohen Nutzerzahlen und Datenmengen gewährleistet.

AP 4.1: Humanities Data Centre

Leitung	Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG)
Gesamtanzahl PM	Eigenanteil über das MWK Niedersachsen mit ca. 300.000 €/Jahr

Ein Forschungsverbund unter Leitung der Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen (GWDG) ist vom Niedersächsischen Ministerium für Wissenschaft und Kultur (MWK) aufgefordert worden, bis Ende Oktober 2011 einen Antrag einzureichen, der auch in engem Bezug zum Zukunftskonzept II der Universität Göttingen im Rahmen der Exzellenzinitiative steht. Dieser Antrag ist termingerecht eingereicht worden. Ziel ist der Aufbau und Betrieb eines Forschungsdatenzentrums für die Geisteswissenschaften, das die Forschungsdaten dieser Disziplinen (z.B. kritische Editionen, Multimedia-Objekte oder 3D-Modelle) entsprechend den wissenschaftlichen Anforderungen der DFG langfristig sichert und, sofern rechtlich möglich, zur Nachnutzung anbietet. Ein solches Humanities Data Centre bildet somit auch die zentrale Datenschnittstelle für VREs. Partner sind neben der GWDG (Projektleitung) und dem ZIB Berlin auch das Max Planck Institute for the Study of Religious and Ethnic Diversity (MPI MMG) und die Akademie der Wissenschaften zu Göttingen (AdW Göttingen) als fachwissenschaftliche Partner sowie die Berlin-Brandenburgische Akademie der Wissenschaften (BBAW) und die SUB Göttingen als Service-Infrastruktureinrichtungen. Im Sommer 2012 wird mit einem Bescheid seitens des Niedersächsischen Ministeriums für Wissenschaft und Kultur gerechnet.

AP 4.2: Performanz, Skalierbarkeit, Lasttests

Leitung	Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG)
Gesamtanzahl PM	18
- davon PM GWDG	9
- davon PM DAASI Intern. (Skalierbarkeit TG-crud und TG-auth*)	4,5
- davon PM FH Worms (Lasttests)	4,5

Performanz, Skalierbarkeit

Mit den Ergebnissen von TextGrid II liegt für das TextGrid Repository eine produktive und leistungsfähige Infrastruktur vor. Die zu erwartenden höheren Nutzerzahlen und ein möglichst ausfallsicherer Produktivbetrieb der gegenwärtigen Infrastruktur erfordern jedoch Verbesserungen hinsichtlich Performanz und Skalierbarkeit. Das in TextGrid II erarbeitete High Availability-Konzept dient dabei als Grundlage und soll entsprechend den Ergebnissen der Lasttests für den Regelbetrieb bei der GWDG vorbereitet und angepasst werden.

Dafür müssen zunächst mögliche Performanz-Schwachstellen identifiziert sowie ein redundanter Aufbau der einzelnen Middleware-Komponenten modelliert werden. Die vorhandenen

Schnittstellen werden hinsichtlich der Skalierung dieser Komponenten gemeinsam geprüft. Unterstützt wird die Problemanalyse durch Lasttests, die auf den Tests der bestehenden TextGrid 2.0-Installation basieren und diese erweitern. Im Wesentlichen handelt es sich dabei um folgende drei Middleware-Komponenten und die ihnen zugrunde liegenden Datenspeicher:

1. TG-auth*, das Authentifizierungs- und Autorisierungs-Framework. Da die zugrunde liegende Datenbanktechnik LDAP sowohl hoch performant als auch einfach redundant verteilbar ist, sind hier primär die entsprechenden Verteilungsszenarien zu konzipieren und zu implementieren. Zu prüfen sind die Nutzung eines Multimaster-Verbunds schreibender TG-auth*-Instanzen sowie in beliebig großer Zahl Nur-Lese-Replikate für die Beantwortung der eigentlichen Zugriffsanfragen.

2. TG-search, die Retrieval-Engine (Indizierung und Suche). Als Datenspeicher wird hier eine Kombination aus der XML-Datenbank eXist mit angeschlossener Volltextsuchmaschine (Lucene) und der RDF-Datenbank Sesame zur Abbildung von Objektrelationen eingesetzt. Dabei sind sowohl die eXist- als auch die Sesame-Datenbank derzeit nicht clusterfähig und müssen entweder entsprechend weiterentwickelt bzw. durch geeignete Produkte ersetzt werden.¹⁴

3. TG-crud, die zentrale Komponente für Create-, Retrieve-, Update- und Delete-Operationen, die sowohl die einzelnen Datenobjekte verwaltet und im TextGrid Repository mit der Möglichkeit der Replikation sichert, als auch die Retrieval-Datenbanken beliefert. Entsprechend den oben skizzierten Änderungen sind die vorhandenen Implementierungen der TG-crud Storage-Schnittstellen anzupassen bzw. zu erneuern. Dieses Vorgehen verspricht Performanz-Vorteile gegenüber der jetzigen Implementierung, da Konzepte der simultanen Befüllung von Hadoop und SOLR über HDFS (Hadoop Distributed Filesystem) zum Tragen kommen.

Im Rahmen dieses Arbeitspakets werden gemeinsam Konzepte erstellt, im Hinblick auf Ausfallsicherheit, Verteilbarkeit und Skalierung praktisch evaluiert und schließlich als Proof-Of-Concept aufgebaut. Darüber hinaus wird der Programmcode von TG-crud und TG-auth* auf Parallelisierbarkeit untersucht, getestet und gegebenenfalls erweitert. Als nur lesende Komponente bleibt TG-search hier ausgenommen, da beliebig viele voneinander unabhängige Instanzen aufgesetzt werden können. Die Ergebnisse dieses Arbeitspakets werden in die Arbeitspakete 4.3 und 4.5. einfließen, die die Erarbeitung einer produktiv einsetzbaren Version des Proof-Of-Concepts sowie den Betrieb und das Monitoring der neuen Repository-Architektur zum Gegenstand haben.

Grundlage des Verteilungskonzepts ist die bereits eingeführte Virtualisierung der TextGrid-Middleware. Durch die Arbeiten des von TextGrid unterstützten britischen Projekts TEXTvire¹⁵ steht heute ein ISO-Image für die gesamte Middleware zur Verfügung, auf das die zu leistenden Arbeiten aufsetzen können.

¹⁴ Es bietet sich z.B. an, für das Retrieval-Backend Alternativ-Technologien zu verwenden, die genuin auf Skalierbarkeit und Performanz ausgerichtet sind, wie etwa Hadoop in Kombination mit einer auf Lucene/SOLR basierten Volltextsuche. Sollte sich ein solches Szenario als gangbar erweisen, könnte die eXist-Datenbank dann, mit der entsprechenden Anpassung von TG-search, zusätzlich für reine XQuery-Abfragen eingesetzt werden. Die existierenden Alternativen zu Sesame, etwa OpenLink Virtuoso oder 4Store, werden hinsichtlich ihrer Tauglichkeit für TextGrid einzeln evaluiert. Der Austausch dieser Komponente ist, da sowohl beim Schreiben als auch beim Lesen der Objekte standardisierte Protokolle zum Einsatz kommen, ohne nennenswerten Aufwand realisierbar.

¹⁵ <http://www.jisc.ac.uk/whatwedo/programmes/vre/textvire.aspx>

Lasttests

Das TextGridRep bildet die zentrale Instanz von TextGrid, die von allen TextGridLab-Instanzen angesprochen wird. Von zentraler Bedeutung sind daher Tests, die Aufschluss darüber geben, wie sich die einzelnen Komponenten auch unter Spitzenbelastung verhalten, da Ausfälle im Regelbetrieb des TextGridRep schwerwiegende Folgen hätten; denn sie gingen nicht nur mit einer unerwünschten Unterbrechung aktueller Arbeiten, sondern möglicherweise auch mit dem Verlust noch nicht gespeicherter Daten einher.

Bereits zum Release der Version 1.0 von TextGrid hat die Fachhochschule Worms in Eigeninitiative Lasttests für die wichtigsten Komponenten des TextGridRep entwickelt und durchgeführt. Die entsprechenden Aktivitäten waren im Erstantrag nicht vorgesehen und wurden aus Eigenmitteln der Fachhochschule unterstützt. Hierfür wurde eigens für TextGrid ein separates Computernetz zur Durchführung der Tests aufgebaut. Diese Tests basieren auf dem Open-Source-Tool Apache JMeter, das sowohl über vorgefertigte, standardisierte Anfragen verfügt als auch auf einfache Weise programmtechnisch erweitert werden kann, so dass sich damit beliebige Test-Szenarien umsetzen lassen. Um die Last zu steigern, können die Tests zusätzlich auf mehrere Rechner verteilt werden. In dem für TextGrid 1.0 benutzten Testaufbau wurden 4 Server mit je 15 Threads verwendet und somit 60 parallel arbeitende Clients mit sekundlichen Anfragen simuliert. Dabei gelang es, mehrere Fehler u.a. in der Basis-Komponente eXist aufzudecken, die unter bestimmten Voraussetzungen zu einem vollständigen Ausfall des Systems hätten führen können. Auch das Release der Version 2.0 von TextGrid wird wieder durch begleitende Lasttests abgesichert werden.

AP 4.2 untergliedert sich in drei Teilbereiche: Im ersten werden in regelmäßigen Abständen Lasttests auf einem dem Produktivsystem entsprechenden System durchgeführt. Neben der Aufdeckung von Fehlern kann damit zuverlässig vorhergesagt werden, wie das System auf zusätzliche Benutzer und steigendes Datenaufkommen reagieren wird. Der zweite Teilbereich betrifft die kontinuierliche Entwicklung neuer sowie die Weiterentwicklung existierender Lasttests. Zum aktuellen Zeitpunkt decken die Tests weder alle Komponenten noch alle Einzelfunktionen des TextGridRep ab. Um verlässliche Aussagen über die Stabilität im Dauerbetrieb zu treffen, muss die Testabdeckung verbessert werden. Im dritten Teilbereich werden im Produktivbetrieb regelmäßig anfallende Logdateien und eingehende Fehlermeldungen analysiert, um mögliche Fehlerszenarien im TextGridRep durch neue Lasttests nachstellen und die Fehler selbst anschließend eliminieren zu können. Die Arbeiten in allen drei Teilbereichen werden sich über die gesamte Projektdauer erstrecken.

AP 4.3: Qualitätsmanagement

Leitung	Staats- und Universitätsbibliothek Göttingen (SUB)
Gesamtanzahl PM	49,5
- davon PM SUB	18
- davon PM GWDG	27
- davon PM DAASI	4,5

Ziel dieses Arbeitspakets ist die Erstellung eines Handbuchs, das die verschiedenen Aspekte des Qualitätsmanagements dokumentiert und als Anleitung für einen Regelbetrieb der TextGrid-Middleware dient. Das Qualitätsmanagement kann unterteilt werden in Administration, Technische Dokumentation, Bugfixing und Monitoring.

Die Administration betrifft sowohl die Pflege der Server, auf denen die Middleware-Dienste laufen, als auch deren Sicherheit, wie beispielsweise regelmäßige Betriebssystem-Updates, das Gewährleisten sicheren Zugriffs, den Schutz vor Angriffen etc. Hier kann auf die Expertise der GWDG und auf bereits in TextGrid II erbrachte Vorarbeiten zurückgegriffen werden, die weiter auszubauen sind.

Die vorhandene (in Zukunft voraussichtlich mehrsprachig zu erstellende) technische Dokumentation der Server und der Middleware-Komponenten aus TextGrid 2.0 wurde als Online-Dokumentation in einem öffentlich zugänglichen, derzeit als geisteswissenschaftlicher Service durch das Projekt DARIAH-DE gehosteten Wiki von den Entwicklern gemeinsam erstellt und wird ebenso wie die interne Dokumentation des TextGrid 2.0 Produktivsystems weitergeführt und ergänzt. Hier werden im Sinne des Qualitätsmanagements die Funktionen der Dienste sowie ihr Application Programming Interface (API) beschrieben und deren interne Struktur dokumentiert.

Ferner sollen neu auftretende – z.B. von Nutzern gemeldete – Fehler in den Middleware-Komponenten korrigiert werden. Im Zentrum stehen dabei hauptsächlich die Komponenten TG-auth*, TG-crud, TG-search und TG-publish. Dabei wird die vorhandene Infrastruktur der GWDG genutzt, um beispielsweise ein Ticketing-System für solche Bugs einzuführen, die bislang über E-Mail-Listen gesammelt und manuell in das Issue-Management-System Jira eingepflegt wurden. Dieses soll weiterhin genutzt und für eine dauerhafte Nutzung zu/zur DARIAH/GWDG migriert werden. Zudem wird ein Release-Management konzipiert, um eine regelmäßige Veröffentlichung (Deployment) verbesserter Dienste zu gewährleisten.

Das umfangreiche Monitoring der TextGrid-Middleware-Services und ihrer Funktionalitäten wird eingebunden in die existierende Monitoring-Infrastruktur der GWDG unter Nutzung bereits vorhandener Monitoring-Komponenten (Icinga, Nagios). Notwendige Reaktionen auf auftretende Service-Ausfälle werden im Handbuch Qualitätsmanagement dokumentiert, so etwa Anweisungen für Service-Neustarts, zur Verortung von Start-Skripten etc.

AP 4.4: Ingest Datensammlung, Publikation, Präsentation

Leitung	Staats- und Universitätsbibliothek Göttingen (SUB)
Gesamtanzahl PM	18

In TextGrid II wurden bereits umfangreiche Datensammlungen in das TextGrid Repository eingespielt, wie z.B. die Digitale Bibliothek. Um weiteren, auch großen Datensammlungen eine Archivierung im TextGrid Repository zu ermöglichen, wird die vorhandene technische Dokumentation des externen Import-Tools – hier wird die Java-Bibliothek koLibRI¹⁶ genutzt – unter Einbeziehung importierender Nutzer weiter verfeinert und ergänzt. Darüber hinaus wird eine inhaltliche Dokumentation erstellt, die die für einen Import erforderlichen Kenntnisse vermittelt, wie

¹⁶ koLibRI - kopal Library for Retrieval and Ingest: http://dp4lib.langzeitarchivierung.de/index_koLibRI.php.de

etwa das von TextGrid für den Import vorgegebene Metadatenschema, Metadaten-Konversion oder -Mapping etc. Fehler im Import-Modul der Java-Bibliothek koLibRI werden, falls erforderlich, ebenfalls behoben.

Weiterhin wird die SUB Göttingen Nutzern bei der Einspielung ihrer Datensammlungen technischen und inhaltlichen Support sowie Beratung bieten. Beratung bei der Strukturierung der einzuspielenden Daten sowie deren verfügbarer Metadaten ist in diesem Zusammenhang von eminenter Bedeutung ebenso wie auch die Vorbereitung auf die nachträgliche Bearbeitung der Daten im TextGridLab und die anschließende Publikation per TG-publish bzw. die direkte Publikation im TextGridRep.

Für die Präsentation der eingespielten Daten kann zunächst der existierende TextGridRep-Browser genutzt werden¹⁷, der Zugriff auf alle im TextGridRep vorhandenen Daten ermöglicht. Für die Präsentation projektspezifischer Daten kann jederzeit ein im Aussehen sowie im Datenbestand angepasster Repository-Browser aufgesetzt werden. Darüber hinaus wird eine Referenzinstanz eines Präsentations-Frameworks entwickelt, das eine feingliedrige Darstellung projektspezifischer Inhalte erlaubt. Hierzu kann möglicherweise das Framework SADE genutzt werden, beispielsweise in Verbindung mit einer projekteigenen eXist-Installation, die aus dem TextGrid Repository die benötigten Daten kopiert und entsprechend den Projektanforderungen visualisiert.

AP 4.5: Technischer Regelbetrieb, Architektur, innovatives Evolutionspotential

Leitung	Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG)
Gesamtanzahl PM	27
- davon PM GWDG	18
- davon PM SUB	9

Ziel dieses Arbeitspakets ist die technische Verstetigung der TextGrid-Infrastruktur sowie die Vorbereitung des Regelbetriebs des TextGridRep bei der GWDG. Vorbereitend wird zunächst die Architektur der TextGrid-Middleware evaluiert und eine Architekturskizze einschließlich Dokumentation erstellt. Hier kann ein großer Teil der bereits vorhandenen Dokumentation nachgenutzt werden. Je nach Ergebnis dieser Analyse werden gegebenenfalls Code- und Infrastrukturanpassungen erforderlich sein – dies möglichst unter Beibehaltung der Middleware-API, die vom TextGridLab und anderen Clients genutzt wird.

Es wird untersucht und soweit möglich implementiert, inwieweit vorhandene Konzepte und existierende Dienste der GWDG für einen Regelbetrieb des TextGridRep nachgenutzt werden können bzw. inwieweit die TextGrid-Middleware auf solche Konzepte angepasst werden oder diese Dienste nutzen kann. Zu nennen sind hier zum Beispiel aktuelle Storage- und Bitstream-Preservation-Konzepte wie etwa eine Implementierung der DARIAH Storage und Admin API und der Service zur Erzeugung von persistenten Identifiern (PIDs). Eine Replikation der Daten des TextGridRep und somit eine automatisierte Verteilung der Daten auf verschiedene Standorte ist in TG-crud bereits grundsätzlich implementiert, eine fein konfigurierbare Replikation existiert jedoch

¹⁷ <http://www.textgridrep.org>

nicht. Eine Replikation wäre zudem eher dem Bereich des Storage/der Bitstream-Preservation zuzuordnen. Es ist zu untersuchen, ob eine solche Replikation mit den existierenden Konzepten der GWDG möglich ist.

Reports (R) & Meilensteine (M):

Bezeichnung	Monat	Beschreibung
R 4.5.1	6	Architekturskizze inklusive Identifikation der Lücken und der Optimierungsmöglichkeiten (AP 4.5 zusammen mit AP 4.2 und 4.3)
R 4.4.1	12	Dokumentation/Leitfaden Import (technisch/inhaltlich) (AP 4.4)
R 4.4.2	15	SADE Dokumentation, Referenzinstanz (AP 4.4)
R 4.5.2	18	Update der in Monat 6 erstellten Architekturskizze, Potenzial der Optimierungen, erste Umsetzungen (AP 4.5)
M 4.4.1	30	Update SADE (Dokumentation und Referenzinstanz) (AP 4.4)
M 4.5.1	30	Architekturskizze eines produktiven Regelbetriebs (mit Umsetzung aller wichtigen Punkte)

Beziehungen bestehen zu folgenden APs:

- AP 1.3: (Technische) Integration und zukünftiges Weiterentwicklungspotential, Zukunftsfähigkeit
- AP 2.1: Umgang mit großen Nutzerzahlen
- AP 2.2: Verteilte / Zentrale Entwicklungen
- AP 3.1: Dokumentation, Nutzer-Information
- AP 5.4: Koordination von Weiterentwicklungen und Satellitenprojekten
- AP 5.5: Weiterentwicklung der Digitalen Bibliothek

AP 5: Betrieb und Softwarepflege Laboratory

Nummer des Arbeitspakets	AP 5
Titel	Betrieb und Softwarepflege Laboratory
AP-Leitung	Universität Würzburg
Start-Monat	Juni 2012 (M1) bis Mai 2015 (M36)
Gesamtanzahl PM	79,5

Ziele (Kurzbeschreibung)
Für die meisten Anwender ist das TextGridLab die sichtbare Seite von TextGrid; entsprechend wichtig ist es daher, das Lab für den Produktivbetrieb weiter zu stabilisieren und mit einer entsprechenden technischen Dokumentation die Integration weiterer (externer) Module durch interessierte Wissenschaftlerinnen und Wissenschaftler bzw. Projekte vorzubereiten.

Beschreibung des Arbeitspakets

In diesem Arbeitspaket wird die Client-Software von TextGrid, das TextGridLab, gepflegt, für die stets notwendige Weiterentwicklung vorbereitet und diese kollaborative Entwicklung koordiniert. Die Pflege betrifft zunächst das Beseitigen von Fehlern in der aktuellen TextGridLab-Version. Die weitere Pflege der umfangreichen funktionalen Tests für das Lab wird dessen Stabilität im Produktivbetrieb sicherstellen, so dass es auch bei den unvermeidlichen Anpassungen an neue Java- und Eclipse-Versionen sowie der Integration neuer Module durch externe Projekte nicht zu Instabilitäten kommt. Um die Benutzerfreundlichkeit des TextGridLabs zu steigern, sind außerdem weitere Usability-Studien vorgesehen, aus denen sich übergreifende Designrichtlinien ableiten lassen.

In den letzten Jahren ist eine Reihe webbasierter geisteswissenschaftlicher Werkzeuge entstanden, die unkompliziert und einfach über Webschnittstellen anzusprechen sind. Ein Vorteil der Architektur von TextGrid besteht darin, dass solche Werkzeuge sehr leicht einzubinden sind und damit das Leistungsangebot für diejenigen Anwender, die ihre Daten TextGrid anvertrauen, schnell anwächst. Solche Werkzeuge zu identifizieren und sie prototypisch zu integrieren, wird die Attraktivität von TextGrid für verschiedene Nutzergruppen erhöhen und zugleich Projektanträge generieren, um solche Werkzeuge nachhaltig und benutzergerecht zu adaptieren. TextGrid versteht sich als Open-Source-Projekt und setzt insofern bei der Weiterentwicklung auf die Mitwirkung vieler, auch externer und internationaler Nutzer. Durch die Beteiligung externer Kooperationspartner an der Entwicklung entsteht ein dynamisches System, das koordiniert werden muss.

Schon jetzt zeigt sich, dass die Forschungsdatensammlung „Digitale Bibliothek“ einen wichtigen Bestandteil von TextGrid ausmacht. Deren weitere Erschließung für die wissenschaftliche Nutzung wird diesen hohen Stellenwert allem Erwarten nach bestätigen. Die Finanzierung für die fachwissenschaftliche Tiefen-Erschließung erfolgt aus Eigenmitteln der Universität Würzburg.

AP 5.1: Sicherheit, Bugfixing, Technische Dokumentation

Leitung	Technische Universität Darmstadt
Gesamtanzahl PM	36

In AP 5.1 wird die konstante technologische Stabilisierung des TextGridLab gewährleistet. Eine zeitnahe und sichere Behebung von Fehlern, verbunden mit anschließenden funktionalen Tests, ist Voraussetzung für den dauerhaften Betrieb und den Aufbau von Vertrauen in die VRE. Um einerseits Kontinuität beispielsweise bei Personalwechsel zu sichern und andererseits auch Entwickler externer Kooperationsprojekte und weitere Interessenten rasch einbinden zu können, ist eine zuverlässige, gründliche und verständliche Dokumentation unabdingbar. Die Basis-Informationen der Dokumentation werden kollaborativ von den Entwicklern zusammengetragen, die Redaktion, Aufbereitung und Weitervermittlung dieser Informationen erfolgt in AP 5.1.

AP 5.2: Funktionale Tests

Leitung	Fachhochschule Worms
Gesamtanzahl PM	4,5

Automatisierte Tests des Gesamtsystems aufgrund typischer Arbeitsschritte von geisteswissenschaftlichen Anwendern erlauben es, bei Programmveränderungen und Anpassungen auftauchende Probleme schnell zu erkennen und zu lokalisieren. Das Zentrum für Technologietransfer und Telekommunikation der Fachhochschule Worms hat im Rahmen der funktionalen Tests von TextGrid 1.0 die Initiative übernommen, eine umfangreiche Testumgebung für diese Art der automatisierten funktionalen Tests aufzubauen. Die entsprechenden Aktivitäten waren im ursprünglichen Antrag nicht vorgesehen und wurden aus Eigenmitteln der Fachhochschule finanziert. In einem Labor der Hochschule wurden auf allen Arbeitsplätzen zusätzliche Festplattenpartitionen mit der Testumgebung Sikuli eingerichtet, so dass Tests mit bis zu 21 parallel arbeitenden Rechnern durchgeführt werden können. Ein kompletter Durchlauf der aktuell existierenden funktionalen Tests dauert etwa 70 Minuten, deckt jedoch erst etwa 10% des Funktionsumfangs des TextGridLab ab. Deshalb ist es von entscheidender Bedeutung, im zweiten Teil dieses Arbeitspakets weitere funktionale Testpläne, basierend auf den Anforderungen der Nutzer, zu erstellen. Da im Dauerbetrieb bereits produktiv mit TextGrid gearbeitet wird, dürfte es in dieser Phase des Projektes besonders einfach sein, ausführliche Use-Case-Beschreibungen von den Nutzern zu erhalten, um auf diese Weise Testpläne optimal an die jeweiligen Anforderungen anpassen zu können. Durch enge Abstimmung mit den Fach-Communities lässt sich auch gewährleisten, dass häufig genutzte Funktionen möglichst schnell in funktionale Testpläne aufgenommen werden. Gleichzeitig wird die Fachhochschule Worms proaktiv die im Produktivbetrieb anfallenden Logdateien analysieren und auftretende typische Fehlerszenarien in ihren Testplänen berücksichtigen.

AP 5.3: Usability

Leitung	Max-Planck-Institut für Wissenschaftsgeschichte (MPI WG), Berlin
Gesamtanzahl PM	18

Hier wird systematisch untersucht, wie TextGrid in der wissenschaftlichen Praxis eingesetzt wird. Dazu werden Testgruppen aus unterschiedlichen Disziplinen identifiziert – in dieser Phase mit dem Schwerpunkt auf bisher nicht mit TextGrid arbeitenden Gruppen, insbesondere in der textbasierten historischen Forschung – und deren Arbeitsabläufe beim Einsatz von TextGrid analysiert. Daraus werden Vorschläge für Veränderungen und Ergänzungen von TextGrid entwickelt, die dessen optimale Nutzung ermöglichen. Als Fallbeispiel soll mindestens ein Editionsprojekt systematisch von der Projektionsphase bis zur Ausführung mithilfe von TextGrid begleitet werden. Das Projekt selbst wird durch Eigenmittel des Berliner Max-Planck-Instituts für Wissenschaftsgeschichte (MPI WG) finanziert. Aus den Ergebnissen der Usability-Untersuchungen ergeben sich unmittelbar Vorschläge für Bugfixes bzw. Veränderungen der Benutzerschnittstellen (AP 5.1) sowie Use-Cases für die funktionalen Tests (AP 5.2). Zur Erstellung der Usability-Studien ist eine enge Kooperation mit AP 3 geplant. Erfahrungen aus den dort vorgesehenen Nutzerschulungen und Workshops sollen hierbei einerseits zur Definition und Modifikation der in diesem Arbeitspaket

schwerpunktmäßig begleiteten Nutzergruppen beitragen, andererseits werden sich aus den Ergebnissen der Usability-Studien Anforderungen an die Form der Dokumentation und Vorschläge für die Entwicklung spezifischer Schulungsangebote für die in diesem AP begleiteten Nutzergruppen ergeben. Ein weiterer Input wird von diesem AP für die in AP 2 zu entwickelnden Strukturen für die Organisation von Benutzergruppen erwartet.

AP 5.4: Koordination von Weiterentwicklungen und Satellitenprojekten

Leitung	Universität Würzburg
Gesamtanzahl PM	21
- davon Universität Würzburg	18 (zusätzl. stud. Hilfskraft)
- davon DAASI International	3

Mit den Entwicklungen aus TextGrid und aus den Projekten, die TextGrid nutzen wollen, entsteht ein 'Ökosystem' von Tools. Diese sind in TextGrid ineinander integriert. Dazu sollten die beteiligten Entwicklerinnen und Entwickler Architektur und Schnittstellen von TextGrid verstehen und sich je nach Integration an bestimmte Prozesse halten. Dieses Arbeitspaket unterstützt sie dabei im Zusammenspiel mit den anderen technischen Arbeitspaketen und interessierten externen Projekten.

Geeignete Maßnahmen, die dazu je nach Bedarf in diesem AP durchgeführt werden können, umfassen u.a.

- die Entwicklung von Konzepten und Workflows für eine Integration von externen Tools ins TextGridLab
- gemeinsame „Integrations-Workshops“ oder Programmiersprints auch mit Ansprechpartnern aus anderen APs
- den Entwurf von Templates für eine kollaborative Dokumentation der technischen Schnittstellen und ihrer Benutzung
- die Vermittlung der für ihre Fragestellungen relevante Dokumentation und Ansprechpartner an Entwicklerinnen und Entwickler

Übergreifendes Ziel ist es, für Entwickler die Verwendung von TextGrid und die Integration von neuen Werkzeugen in TextGrid sowie dessen Verwendung in ihren jeweiligen Arbeitskontexten so einfach wie möglich zu gestalten. Ein entsprechendes Satellitenprojekt wird exemplarisch im Rahmen dieses APs durchgeführt. Dies geschieht konkret im Rahmen einer Kooperation mit dem Langzeitprojekt „Johann Friedrich Blumenbach – online“, das bei der Akademie der Wissenschaften zu Göttingen angesiedelt ist. Der kommerzielle TextGrid-Partner DAASI International GmbH wird das noch von der Universität Würzburg in das TextGridLab zu integrierende Kollationswerkzeug CollateX im ersten Projektjahr an die besonderen Anforderungen bezüglich des synoptischen Vergleichs strukturell ähnlicher Texte (XML-Dokumente; zwei oder mehr) anpassen, insbesondere hinsichtlich der Anzeige/Ausgabe von Gemeinsamkeiten und Differenzen auf struktureller Ebene auch bei Texten mit nur wenigen textlichen Übereinstimmungen. DAASI wird in den Jahren 2 und 3 des Projekts und darüber hinaus solche Anpassungsdienste kommerziell anbieten, um Satelliten-Projekten, die über keine entsprechende Software-Entwicklungskompetenzen verfügen, die von ihnen benötigte, aber noch fehlende Funktionalität in TextGrid zur Verfügung zu stellen.

AP 5.5: Weiterentwicklung der Digitalen Bibliothek

Leitung	Universität Würzburg
Gesamtanzahl PM	18 PM als Eigenanteil

Während der Projektlaufzeit wurde der umfangreichste Ordner der Digitalen Bibliothek zeno.org, der Literaturordner, in den Markup-Standard TEI P5 (Text Encoding Initiative) umgewandelt und veröffentlicht. Derzeit ist der Philosophieordner in Bearbeitung und soll in Kürze ebenfalls veröffentlicht werden.

Die dabei durchgeführten Arbeitsschritte sollen zukünftig sukzessive auf die weiteren Ordner der Zeno-Daten angewandt werden (Geschichte, Kulturgeschichte, Kunst, Märchen, Musik, Naturwissenschaften, Soziologie, Nachschlagewerke). Hierbei werden die bereits vorhandenen Transformationen als Grundlage herangezogen und an die jeweilige spezifische Struktur der einzelnen Ordner angepasst. Daneben ist zur Aufbereitung der Wörterbücher die Entwicklung eines neuen speziellen Mappings mit dazugehörigen Transformationsroutinen geplant.

Für die Überführung der Daten nach TEI P5 sind mehrere Arbeitsschritte erforderlich: Nach der Analyse und Sichtung der Daten werden mit Hilfe einer selbstentworfenen Oberfläche die Werke eines jeden Autors manuell als solche gekennzeichnet, ein Mapping der Datenstruktur auf TEI P5 erstellt und eine entsprechende Transformationsroutine entwickelt. Diese beinhaltet auch die Extraktion von Metadaten zu den einzelnen Werken sowie die Aufbereitung der Daten zur Weiterverwendung im TextGridLab und TextGridRep (Aufsplitten der Dateien und Erstellen von Metadateien).

Reports (R) & Meilensteine (M):

<i>Bezeichnung</i>	<i>Monat</i>	<i>Beschreibung</i>
M/R 5.2.1	1 – 36	Laufende Funktionale Tests
M/R 5.3.1	6	Usability, Zusammenfassung der bisherigen Nutzerszenarien und Identifikation der Testgruppen
M 5.4.1	12	Konzepte und Workflows zur Integration von Tools in das TextGridLab
R 5.5.2	18	Usability: Anforderungen an TextGrid aus der Arbeit mit den Testgruppen
M 5.1.1	24	Technische Dokumentation
M 5.5.1	24	Vollständige Veröffentlichung der aufbereiteten Daten
M/R 5.5.3	27	Usability: Abschluss des Fallbeispiels und sich daraus ergebende Anforderungsanalyse
R 5.5.1	36	Tiefenerschließung der Wörterbücher und Enzyklopädien

Beziehungen bestehen zu folgenden APs:

- AP 1.3: (Technische) Integration und zukünftiges Weiterentwicklungspotential, Zukunftsfähigkeit
- AP 2.1: Umgang mit großen Nutzerzahlen und heterogenen Nutzergruppen sowie Nutzerverwaltung
- AP 2.2: Verteilte / Zentrale Entwicklungen
- AP 3.1: Dokumentation, Nutzer-Information
- AP 4.4: Ingest Datensammlung, Publikation, Präsentation
- AP 4.5: Technischer Regelbetrieb, Architektur, innovatives Evolutionspotential

AP 6: Integriertes Monitoring und Qualitätssicherung

Nummer des Arbeitspakets	AP 6
Titel	Integriertes Monitoring und Qualitätssicherung
AP-Leitung	Technische Universität Berlin, Zentrum Technik und Gesellschaft (ZTG)
Start-Monat	Juni 2012 (M1) bis Mai 2015 (M36)
Gesamtanzahl PM	30

Ziele (Kurzbeschreibung)
Das AP zielt auf die Entwicklung eines integrierten Monitoring von digitalen Forschungsinfrastrukturen in den Geisteswissenschaften ab. Konkret wird ein integriertes Monitoring in TextGrid implementiert. Dazu wird erstens eine Beurteilungsmethodik erarbeitet und zweitens werden Beurteilungskategorien festgelegt. Darauf aufbauend werden Governance-Werkzeuge zum steuernden Eingreifen (sofern Prozesse nicht den gewünschten Verlauf nehmen) TextGrid zur Verfügung gestellt.

Beschreibung des Arbeitspakets

In diesem AP implementiert das Zentrum für Technik und Gesellschaft (ZTG) der Technischen Universität Berlin als bisher externer Partner im TextGrid-Konsortium ein integriertes Monitoring. Dafür wird eine Beurteilungsmethodik erarbeitet und abgestimmt. Für das steuernde Eingreifen werden Governance-Werkzeuge entwickelt.

Integriertes Monitoring bedeutet ein umfassendes Monitoring, das nicht nur „harte Zahlen“ in seiner Bewertung berücksichtigt (z.B. die Anzahl der Klicks auf einer Website), sondern auch soziale (z.B. Nutzerperspektive, Nutzer-Entwickler-Interaktionen) oder ökologische Dimensionen (z.B. Energieverbrauch durch Betrieb von digitalen Forschungsinfrastrukturen) in die unmittelbare systematische Erfassung von Prozessen aufnimmt. Dabei werden verschiedene Forschungsbereiche (z.B. Science & Technology Studies, Software Engineering) geprüft und ihre für das integrierte Monitoring geeigneten Ansätze miteinander kombiniert. In Anlehnung an den Ansatz zum integrierten Geschäftsbericht (internationaler Reporting-Trend, zunehmend auch in Form von Sustainability Reports) wird auf diese Weise ein integriertes Monitoring entwickelt, das die Leistungen der Forschungsinfrastruktur TextGrid in den Feldern Wissen, Technologie, Ökonomie, Ökologie und Soziales zu erfassen und auf der Basis einer integrierenden Auswertung zu bewerten vermag.

AP 6.1: Integriertes Monitoring aller Arbeitspakete

Leitung	Technische Universität Berlin, Zentrum Technik und Gesellschaft (ZTG)
Gesamtanzahl PM	18

Es wird auf einer breiten und interdisziplinären Literaturbasis systematisch untersucht, welche Dimensionen aus den Bereichen Wissen, Technologie, Ökonomie, Ökologie und Soziales das integrierte Monitoring einer digitalen Forschungsinfrastruktur beinhalten muss. Hierzu werden erstens existierende Monitoring-Ansätze einer umfassenden Auswertung unterzogen. Zweitens wird eine Dokumentenanalyse und drittens eine Befragung relevanter Akteure durchgeführt. Einbezogen werden dabei die Ergebnisse der Studie zur Nutzerakzeptanz (AP 6.2). Darauf aufbauend sind in einem nächsten Schritt eine Beurteilungsmethodik zu erarbeiten und die Beurteilungskategorien zu definieren. Außerdem werden Governance-Werkzeuge zum steuernden Eingreifen entwickelt. Ziel ist die Implementierung eines integrierten Monitoring von TextGrid.

AP 6.2: Nutzerakzeptanz (inklusive Befragungen)

Leitung	Berlin-Brandenburgische Akademie der Wissenschaften (BBAW)
Gesamtanzahl PM	12

Als Projektpartner führt die Berlin-Brandenburgische Akademie der Wissenschaften (BBAW) eine Nutzerumfrage durch. Diese stellt die Bedarfe vor allem im Hinblick auf zwei unterschiedliche Nutzergruppen fest. Dies sind zum einen kurzlaufende Forschungsprojekte sowie Einzelwissenschaftlerinnen und -wissenschaftler. Für diese Gruppen stellt sich die Frage, mit welchen Tools und Ressourcen ihre Forschungsarbeit von TextGrid über das bereits zur Verfügung gestellte Instrumentarium hinaus noch besser unterstützt werden kann. Im Zentrum des APs steht die Erhebung von prototypischen Nutzungsszenarien der Vorhaben im Akademien-Programm, es bezieht aber exemplarisch auch andere Forschungsgruppen und Einzelwissenschaftler mit ein. Berücksichtigt werden sowohl die virtuelle Arbeitsumgebung (TextGridLab) als auch die Präsentations- und Storage-Ebene des TextGridRep. Die Vorhaben des Akademien-Programms bieten sowohl in ihrer Menge und Bedeutung für die deutschen Geisteswissenschaften als auch in ihrer Konzentration auf die geisteswissenschaftliche Grundlagenforschung sowie in der Vielfalt von Anforderungen, die Editionen und Wörterbuchprojekte an eine virtuelle Arbeitsumgebung stellen, einzigartige Informationschancen zur Beurteilung der Nutzerakzeptanz. Aufgrund ihrer langen Laufzeit und des damit einhergehenden hohen Interesses an einer im Durchschnitt zwischen 15 und 25 Jahren währenden Arbeit mit der Arbeitsumgebung und der Langfristverfügbarkeit ihrer Ergebnisse eignen sie sich hervorragend für eine langfristig orientierte Abschätzung der Bedarfe. Ziel des APs ist es auch, zu prüfen, inwieweit TextGrid im Akademien-Programm flächendeckend als Virtuelle Forschungsumgebung etabliert werden kann.

Reports (R) & Meilensteine (M):

<i>Bezeichnung</i>	<i>Monat</i>	<i>Beschreibung</i>
R 6.2.1	12	Report „Ergebnisse der Nutzerbefragung“
R 6.1.1	12	Report zur Analyse „Dimensionen des integrierten Monitoring einer digitalen Forschungsinfrastruktur“
R 6.1.2	18	Konzept Beurteilungsmethodik, Beurteilungskategorien und Governance-Werkzeuge
M 6.1.3	20	Workshop mit Projektpartnern zum erarbeiteten Konzept des integrierten Monitoring
M 6.1.4	24	Implementierung des integrierten Monitoring, Report „Integriertes Monitoring einer digitalen Forschungsinfrastruktur“
R 6.1.5	36	Abschließender Bericht zur Implementierung inklusive Evaluierungen

Beziehungen bestehen zu folgenden APs:

- AP 2.1: Umgang mit großen Nutzerzahlen und heterogenen Nutzergruppen
- AP 2.2: Verteilte / Zentrale Entwicklungen
- AP 3.1: Dokumentation, Nutzer-Information
- AP 4.3: Qualitätsmanagement
- AP 5.3: Usability
- AP 7: Projektmanagement und Öffentlichkeitsarbeit

AP 7: Projektmanagement und Öffentlichkeitsarbeit

Nummer des Arbeitspakets	AP 7
Titel	Projektmanagement und Öffentlichkeitsarbeit
AP-Leitung	Staats- und Universitätsbibliothek Göttingen (SUB)
Start-Monat	Juni 2012 (M1) bis Mai 2015 (M36)
Gesamtanzahl PM	36 (zusätzlich student. Hilfskraft)

Ziele (Kurzbeschreibung)
<p>Das Arbeitspaket unterstützt die Projektpartner, um sicherzustellen, dass alle Projektziele innerhalb des vorgegebenen Zeit- und Budgetrahmens erreicht werden. Dazu gehört die fachliche wie auch die formale und administrative Leitung des Projektes. Weitere Schwerpunkte dieses Arbeitspakets sind die koordinierte Öffentlichkeitsarbeit für das Verbundprojekt sowie die Vernetzung mit anderen Digital Humanities-Initiativen und Projekten. Im Zuge der Etablierung einer nachhaltigen Institution kommt der Abstimmung mit den Wissenschaftsorganisationen (Fachgesellschaften etc.) eine Schlüsselrolle zu.</p>

Beschreibung des Arbeitspakets

Die Gesamtleitung des Projekts umfasst neben der Konsortialleitung des Verbundvorhabens vor allem die koordinierte Vermittlung von Ergebnissen in die Fach-Communities, z.B. mittels geeigneter PR-Materialien, Social-Media-Strategien und durch intensiven Austausch mit den Fachgesellschaften (in Zusammenarbeit mit den anderen APs). Weitere zentrale Aufgaben sind die fachliche Gesamtkoordination der einzelnen Arbeitspakete und -gruppen sowie die Koordination und Verknüpfung von Ergebnissen. Die Abstimmung der einzelnen Arbeitspakete untereinander und deren Verknüpfung miteinander ist eine weitere Kernaufgabe.

Das Management des Projektes betrifft dessen administrative und organisatorische Durchführung. Dazu gehört die finanzielle Verwaltung der zentral eingesetzten Mittel ebenso wie die Sicherstellung der Einhaltung von Meilensteinen und der termingerechten Abgabe von Ergebnisberichten. Das Projektmanagement hat ebenfalls die Verantwortung für den reibungslosen Ablauf der internen und externen Kommunikation.

AP 7.1: Projektmanagement i.e.S. und Koordinierung potentieller Kooperationen

Leitung	Staats- und Universitätsbibliothek Göttingen (SUB)
Gesamtanzahl PM	12

AP 7.1 hat, wie auch bereits in den ersten beiden Förderphasen, die fachliche, formale und administrative Unterstützung und Koordination der Projektpartner im Hinblick auf die fristgerechte Umsetzung der in den einzelnen Arbeitspaketen formulierten Arbeitsziele zum Gegenstand. Die Koordinierung potentieller fachlicher wie auch strategischer Kooperationen bildet einen weiteren Schwerpunkt dieses APs. Bereits in der zweiten Förderphase konnte TextGrid zahlreiche Kooperationen mit Forschungsverbänden konzipieren und gestalten, die mit der dritten Förderphase vor der konkreten Umsetzung stehen. Dazu gehören u.a. die Projekte, die im Rahmen des BMBF-Calls für eHumanities als förderungswürdig eingestuft wurden, aber auch strategische Kooperationen wie die potentielle Anbindung von SADE (Scalable Architecture for Digital Editions) und die Integration des Open Source e-research environments eSciDoc.

AP 7.2: Abstimmung mit Fachgesellschaften

Leitung	Staats- und Universitätsbibliothek Göttingen (SUB)
Gesamtanzahl PM	6

Dieses AP sieht zur Unterstützung der anderen Arbeitspakete und zur Ansprache der einzelnen Zielgruppen die Planung und Durchführung regelmäßiger Experten- und Strategiegelgespräche vor. Halbjährlich werden im Wechsel ein thematisches und ein strategisches Expertengespräch stattfinden. Zudem ist eine enge Abstimmung mit den für das Projekt und dessen Nutzer-Community relevanten Fachgesellschaften zu gewährleisten. Eine weitere Aufgabe ist die enge

Vernetzung mit anderen Fachgesellschaften und eHumanities-Zentren sowie Forschungsinfrastrukturinitiativen und -projekten im nationalen (z.B. eAqua, D-Spin) und internationalen Kontext (DARIAH, CLARIN, Bamboo). Ziel ist dabei, Aufbau und Weiterentwicklung einer übergreifenden eHumanities-Forschungsinfrastruktur zu unterstützen.

AP 7.3: Marketing und Öffentlichkeitsarbeit

Leitung	Staats- und Universitätsbibliothek Göttingen (SUB)
Gesamtanzahl PM	18

Die koordinierte Öffentlichkeitsarbeit und eine professionelle Marketingstrategie sind Kernaufgaben, bei denen es darum geht, einen möglichst hohen Grad an Akzeptanz bei den Fach-Communities als Zielgruppe und den politischen Entscheidungsträgern zu erreichen. Eine den gesamten Projektverlauf umfassende PR- und Marketingstrategie und die daraus folgenden konkreten Umsetzungen stellen die Basis dar und unterstützen damit maßgeblich die Aktivitäten in AP 3 (Community-Bildung) sowie den Aufbau eines nationalen eHumanities-Netzwerkes. Ereignisorientierte Aktivitäten wie die im Februar 2012 erfolgte Produktion eines Imagefilms zum Thema „Virtuelle Forschungswelten: Neue Technologien in den Geisteswissenschaften“ mit TextGrid als Schwerpunkt ebenso wie die Einrichtung eines übergreifenden Digital Humanities-Blogs im deutschsprachigen Raum werden die erzielten Arbeitsfortschritte des Verbundvorhabens innerhalb und außerhalb der Zielgruppe sichtbar machen.

Reports (R) & Meilensteine (M):

<i>Bezeichnung</i>	<i>Monat</i>	<i>Beschreibung</i>
M 7.3.1	1-36	Webauftritt TextGrid
R 7.3.1	1-36	Öffentlichkeitsarbeit (laufend)
R 7.3.2	6 (18, 30)	Social-Media Konzept, Umsetzung & Betreuung
R 7.3.3	9 (21, 33)	Marketingkonzept & Umsetzung
R 7.2.1	12	Konzept / Organisation Experten-, Strategiegelgespräche und TextGrid Summit etc.
R 7.2.2	12 (24, 36)	Kooperation mit Fachgesellschaften

Beziehungen zu anderen Arbeitspaketen: Naturgemäß bestehen zu allen Arbeitspaketen enge Beziehungen.

8. Verbundpartner-Beschreibungen

Im Folgenden findet sich eine projektspezifische Beschreibung aller Partner des TextGrid-Konsortiums:

Berlin-Brandenburgische Akademie der Wissenschaften (BBAW), Berlin

Arbeitsschwerpunkt: Nutzerakzeptanz inkl. Befragungen (AP 6.2)

Kompetenz: Aktive geisteswissenschaftliche Grundlagenforschung in zahlreichen Fach-Communities von der Antike bis zur Gegenwart und deren digitale Unterstützung in Workflow und Präsentation von Forschungsdaten; Erfahrungen bei der Durchführung von Nutzer- und Expertenbefragungen (im Rahmen des DFG-Projekts D-Spin hat die BBAW in Form von Experteninterviews die Bedürfnisse und Anforderungen an die Gestaltung von Zugriffsmöglichkeiten auf linguistische und andere geisteswissenschaftliche Ressourcen evaluiert); Betrieb von informationstechnischer Infrastruktur zur Entwicklung und Präsentation von Forschungsdaten; vielfältige Erfahrungen im Umgang von Forschenden in der Benutzung von digitalen Arbeitsumgebungen

Laufende Projekte mit TextGrid-Bezug: Die BBAW ist Verbundpartner im BMBF-Projekt DARIAH-DE

DAASI International GmbH, Tübingen

Arbeitsschwerpunkt: Betrieb und Softwarepflege Repository: Performanz/Skalierbarkeit (AP 4.2); Exemplarische Durchführung einer Anpassung für ein Satellitenprojekt (AP 5.4)

Kompetenz: Aufbau und Betrieb von IT-Infrastrukturen u.a. in den Bereichen Grid Computing und Identity Management; 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: DAASI war an beiden TextGrid-Projekten beteiligt und entwickelt Teile der Software TG-auth in weiteren Projekten weiter. Beteiligung an DARIAH-DE, wobei TextGrid-Expertise mit eingebracht wird.

Fachhochschule Worms, Fachbereich Informatik und Telekommunikation

Arbeitsschwerpunkt: Lasttests (AP 4.2), funktionale Tests (AP 5.2)

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: DFG-Projekt „XML-Print“, Isidorus: Open Source Topic Map Engine; Ausrichtung von eHumanities Session und Tutorials in der IEEE DEST; Regelmäßige studentische Praxisprojekte im Bereich verteilter Anwendungen und der Frontendentwicklung; Teststrategien

Georg-August-Universität Göttingen, Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB) (Projektleitung)

Arbeitsschwerpunkt: Betrieb und Softwarepflege (TextGrid Repository, AP 4), Konsortialleitung und Öffentlichkeitsarbeit (AP 7)

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

Laufende Projekte mit TextGrid-Bezug: Entwicklung von Forschungsdatenarchiven (WissGrid, BMBF); Entwicklung von geisteswissenschaftlichen Forschungsinfrastrukturen (DARIAH, EU), Bereitstellung von Publikations-Infrastrukturen (Driver, EU)

Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG)

Arbeitsschwerpunkt: TextGrid Regelbetrieb und Qualitätsmanagement (AP 4.2, 4.3, 4.5)

Kompetenz: Betrieb von Infrastrukturen für Langzeitarchivierung; Entwicklung von Umgebungen für verteiltes Rechnen und Speicherung; Konzeption von Prozessabläufen, Policies und Dienstgütevereinbarungen für nachhaltige Speicherlösungen; virtualisierte Rechen- und Speicherinfrastrukturen

Laufende Projekte mit TextGrid-Bezug: Entwicklung von geisteswissenschaftlichen Forschungsinfrastrukturen (DARIAH, EU), Forschungsinfrastruktur für die Geistes- und Sozialwissenschaften (CLARIN D, BMBF)

Institut für Deutsche Sprache (IDS), Mannheim

Arbeitsschwerpunkt: Betriebsmodell (AP 1)

Kompetenz: Juristische und ethische Aspekte von digitalen Forschungsprimärdaten; Betriebsmodelle; Betriebs- und Finanzierungsmodelle für eine dauerhafte Bereitstellung von Forschungsinfrastrukturen und Diensten; technische und organisatorische Langzeitarchivierungsstrategien.

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 (BMBF)

Julius-Maximilians-Universität Würzburg, Institut für Deutsche Philologie

Arbeitsschwerpunkt: Betrieb und Softwarepflege Laboratory (AP 5)

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 (BMBF)

Max-Planck-Institut für Wissenschaftsgeschichte (MPI WG), Berlin

Arbeitsschwerpunkt: Usability-Untersuchungen mit besonderem Schwerpunkt auf der fachwissenschaftlichen Anwendung (AP 5.3), Prozessorganisation des Übergangs von Prototypen in den Dauerbetrieb (AP 2)

Kompetenz: Digitale Editionsprojekte und Arbeitsumgebungen für die historische Forschung, Bereitstellung von elektronischen Quellen (European Cultural Heritage Online)

Laufende Projekte mit TextGrid-Bezug: Entwicklung von Bildwerkzeugen (digilib) als Plugin für TextGrid, Entwicklung einer Produktionslinie für digitale Editionen (Edition Open Access), Entwicklung eines XML-Workflows und von Sprachtechnologien für die textbasierte Quellenarbeit in den Geisteswissenschaften (XML-Workflow), Digital Manuscripts to the Europeana (EU-Projekt, Beginn März 2012)

Technische Universität Berlin, Zentrum Technik und Gesellschaft (ZTG)

Arbeitsschwerpunkt: Integriertes Monitoring und Qualitätssicherung (AP 6)

Kompetenz: Interdisziplinäre Wissenschafts- und Technikforschung, Sozialwissenschaftliche Begleitforschung zu digitalen Forschungsinfrastrukturen, E-Science und Digital Humanities

Laufende Projekte mit TextGrid-Bezug: Seit Nov. 2010 Begleitforschung zu C3-INAD und TextGrid „E-SCIENCE UND IHRE NEUEN INTERFACES. Technische und institutionelle Transformationen der Wissenschaft und deren forschungspraktische Folgen am Beispiel der Geisteswissenschaften (TextGrid) und der Klimaforschung (C3-Grid)“

Technische Universität Darmstadt, Institut für Sprach- und Literaturwissenschaft

Arbeitsschwerpunkt: Nutzer-Kommunikation (AP 3); Dokumentation (3.1), Coaching und Schulungsmaßnahmen (AP 3.3), Lab Support (AP 5.1)

Kompetenz: Digital Humanities-Methoden in Forschung und Lehre, Digitale Lexikographie, Digitale Editionsphilologie, Digitalisierung und semantische Erschließung von Primär-Quellen und Forschungsdaten

Laufende Projekte mit TextGrid-Bezug: WissGrid (BMBF); Wechselwirkungen zwischen linguistischen und bioinformatischen Verfahren, Methoden und Algorithmen: Modellierung und Abbildung von Varianz in Sprache und Genomen (BMBF); DARIAH (BMBF); Entwicklung eines XML-basierten Satzprogramms für wissenschaftliche Editionen und Wörterbücher (DFG); Virtuelles Skriptorium St. Matthias / Trier (DFG)