

Textdatenbank und Wörterbuch des Klassischen Maya

Arbeitsstelle der Nordrhein-Westfälischen Akademie der Wissenschaften und der Künste
an der Rheinischen Friedrich-Wilhelms-Universität Bonn



ISSN 2366-5556

PROJECT REPORT 2 | DE



Published 19 Dec 2014

DOI: 10.20376/IDIOM-23665556.14.pr002.de

Jahresabschlussbericht 2014

Nikolai Grube¹, Christian Prager¹, Katja Diederichs¹, Sven Gronemeyer^{1,2}, Elisabeth Wagner¹, Maximilian Brodhun³, Franziska Diehr³ & Petra Maier⁴

¹) Rheinische Friedrich-Wilhelms-Universität, Bonn

²) La Trobe University, Melbourne

³) Niedersächsische Staats- und Universitätsbibliothek, Göttingen

⁴) ULB Heinrich-Heine-Universität, Düsseldorf

Projektbeschreibung, Zielsetzung und Methoden

Die nur teilweise entzifferte Hieroglyphenschrift und Sprache der Mayakultur steht im Mittelpunkt des hier zur Diskussion stehenden Forschungsprojekts, das in Kooperation zwischen den Universitäten Bonn und Göttingen durchgeführt wird. Ziel ist die Erstellung einer Textdatenbank und ein darauf basierendes Wörterbuch des Klassischen Maya (Blütezeit von 250 bis 950 n. Chr.). Aus dieser Zeit sind rund zehntausend Text- und Bildträger überliefert, die mit ihren schriftlichen und bildlichen Informationen einzigartige Perspektiven auf Sprache, Kultur und Geschichte dieser vorspanischen Gesellschaft aufzeigen. Bis heute fehlen allerdings eine systematische Dokumentation und eine umfassende Analyse der Bild- und Textträger, die eine eingehende und präzise Untersuchung der Klassischen Schriftsprache erlauben, indem etwa Textpassagen mit Hilfe einer Kontext- und Kookkurrenzanalyse verglichen werden, Bildinhalte mit Textpassagen korreliert oder die Beschaffenheit oder Funktion eines Textträgers in der Inschrift erfasst und damit unklare Textpassagen möglicherweise verständlich werden. Bislang war ein derartig systematisches und vernetztes Arbeiten mit Text, Bild und Informationsträgern nicht möglich, da die dafür notwendige Technologie in diesem Fach noch nicht existierte. Im Rahmen des Projekts werden die Text- und Bildträger systematisch und nach einheitlichen Standards beschrieben, das Ausgangsmaterial auf der Basis von XML maschinenlesbar gemacht und auf diese Weise die Grundlagen für die Kompilation des Wörterbuchs geschaffen. Dieses Unterfangen kann nur mit Methoden und Technologien aus den digitalen Geisteswissenschaften in Angriff genommen werden, wobei das Projekt auf Tools und Technologien zurückgreift, die in der Virtuellen Forschungsumgebung (Virtual Research Environment, VRE) TextGrid bereits vorhanden sind bzw. im Rahmen des Projektes entwickelt und implementiert werden. Wesentliche Voraussetzung ist dabei, dass nicht nur der sprachliche Inhalt der Inschriften und die ikonischen Informationen der Bilddarstellungen, sondern auch Daten über den Inschriften- und Bildträger (Beschreibungs- oder Metadaten) berücksichtigt und in einer Datenbank angelegt werden. Zu diesem Zweck werden in TextGrid Tools und Workflows entwickelt, welche I. die Dokumentation

der Schrift- und Bildträger mit Aufarbeitung des Forschungsstandes, II. die epigraphisch-linguistische Auswertung der Hieroglyphentexte sowie III. die Edition der Texte mit Transliteration, Transkription und Übersetzung in einem einzigen System ermöglichen. Die VRE enthält nicht nur Beschreibungen der Textträger oder Informationen über die Texte, sondern der Datenbanknutzer bekommt mit Hilfe der Literaturdatenbank auch einen Überblick darüber, welche Autoren sich mit einem Monument befasst oder es publiziert haben, eine Textpassage diskutiert oder erstmals eine bis heute gültige sprachliche Lesung einer Hieroglyphe präsentiert haben. Der Textträger erhält dadurch eine ‚Biographie‘, die eng mit den Textinhalten verwoben ist und bei der Bedeutungsanalyse von Wörtern berücksichtigt wird. Die derzeit im Aufbau befindliche virtuelle Arbeitsumgebung orientiert sich stark am epigraphischen Workflow, der mit der Dokumentation der Textträger und der Erfassung von Beschreibungsdaten beginnt, sich in der epigraphischen Analyse mit der Zeichenklassifikation, Transliteration und Transkription der Texte fortsetzt, mit der morphologischen Segmentierung und linguistischen Interpretation weitergeht und optimaler Weise mit der Übersetzung und Publikation einer Inschrift endet.

Projektbeginn und Aufnahme der Tätigkeit

Projektbeginn war der 1. Januar 2014. Im Zeitraum bis zur Einstellung der beiden Fachwissenschaftler Dr. Christian Prager und Elisabeth Wagner, M.A. am 01.03.2014 bzw. 03.03.2014 sowie zwei Hilfskräften wurde das Projekt an der Universität Bonn (ULB) vertraglich und administrativ etabliert (Universität Bonn Projektnummer: 98050020) und die Kooperation zwischen ULB und der Staats- und Universitätsbibliothek Göttingen (SUB) in die Wege geleitet. Hierzu wurde ein Weiterleitungsvertrag zwischen ULB und SUB abgeschlossen, der bis in das erste Quartal 2017 den Mittelabfluss von Bonn nach Göttingen für die Finanzierung der zwei Mitarbeiterstellen (150%) an der Niedersächsischen Staats- und Universitätsbibliothek Göttingen reguliert. Die Bonner Informatikerstelle und die Metadaten- sowie Entwicklerstellen in Göttingen wurden im ersten Quartal 2014 ausgeschrieben und Anfang Mai und Juni 2015 mit Katja Diederichs in Bonn, Franziska Diehr und Maximilian Brodhun (beide Göttingen) besetzt. Als weiterer Fachwissenschaftler wurde am 01.05.2014 Sven Gronemeyer, M.A. eingestellt. Die im 1. Arbeitspaket (siehe Antrag) festgelegte Konzeption des Metadatenschemas sowie die Entwicklung und Programmierung der virtuellen epigraphischen Forschungsumgebung in TextGrid wurden somit ab dem zweiten Halbjahr in Angriff genommen. Die Fortschritte im Projekt und der vorliegende Arbeitsbericht sind vor diesem Hintergrund zu evaluieren.

Mitglieder der Arbeitsstelle, Aufgaben und Kommunikation im Projekt

Per dato sind 14 Personen im Projekt unter Vertrag. Die drei Fachwissenschaftler Dr. Christian Prager (Projektkoordination), Elisabeth Wagner, M.A. und Sven Gronemeyer, M.A. sind als Experten für Mayaschrift, -sprache und -ikonologie und Katja Diederichs ist für den informationswissenschaftlichen Bereich an der Bonner Arbeitsstelle tätig. Letztere koordiniert die Zusammenarbeit zwischen dem Projekt und der ULB bezüglich der Präsentationsplattform Visual Library, wo ab 2016 ein virtuelles Inschriftenarchiv online gestellt wird. Bei der SUB kümmert sich Franziska Diehr um die Konzeption des Metadatenschemas auf der Basis der fachwissenschaftlichen Vorgaben aus Bonn; Maximilian Brodhun entwickelt, programmiert und implementiert auf Grundlage von Franziska Diehrs Metadatenschema den epigraphischen Workflow mit dem Arbeitstitel „Text-Dokumentation · Text-Analyse · Text-Editition“ in der virtuellen Forschungsumgebung TextGrid. Unterstützt werden die Fachwissenschaftler in Bonn durch die Hilfskräfte Christiane Bahr (Digitalisierung), Laura Burzywoda (Digitalisierung), Antje Grothe (Bibliographie), Lena Heise (Digitalisierung), Leonie Heine (Digitalisierung), Jana Karsch (Epigraphik,

Digitalisierung) Nikolai Kiel (Epigraphik, Digitalisierung), Mallory Matsumoto (Linguistik, Digitalisierung).

Mit dem Projekt assoziiert sind der eHumanities Mitarbeiter der ULB, Jan Kenter, der zusammen mit Katja Diederichs die Kooperation zwischen der ULB und dem Projekt organisiert, die Bibliothekarin Petra Maier von der Universitäts- und Landesbibliothek Düsseldorf, die als freie Mitarbeiterin im Projekt unter der Vermittlung von Dr. Heike Neuroth (SUB) eine Studie über die Anwendung von XML zur Auszeichnung von Maya-Texten angefertigt hat und in diesem Bericht einen Gastbeitrag eingebracht hat. Seit November 2014 steht auch Dr. Thomas Kollatz vom Salomon Ludwig Steinheim-Institut in Essen als Experte für die XML-basierten Auszeichnungsstandards TEI und Epidoc dem Projekt als Ansprechpartner zur Seite.

Die schriftliche Kommunikation zwischen Bonn und Göttingen erfolgt über ein nicht-öffentliches Wiki der Kollaborationssoftware Confluence, welches in die Forschungsinfrastruktur DARIAH-DE eingebunden ist, worüber alle beteiligten Wissenschaftler über das Metadatenschema, die kontrollierten Vokabulare, Normdaten, epigraphische Normen und Standards u.v.m. diskutieren und Daten austauschen. Das Wiki enthält die vollständige Projektdokumentation mit den bisher aggregierten technischen und fachwissenschaftlichen Inhalten, die als Referenzmaterial langfristig verfügbar sein müssen. Es handelt sich um Dokumente zur Projektorganisation, zu den verschiedenen Arbeitsbereichen sowie Materialien, die dort abgelegt und von den Mitarbeitern abgerufen und gemeinsam bearbeitet werden. Dokumente wie etwa Vorträge, Präsentationen oder Arbeitslisten werden in einem Cloudspeicher abgelegt und können auf diese Weise kollaborativ bearbeitet werden.

Mündliche Kommunikation ist zentral für die Projektarbeit: In wöchentlichen Telefonkonferenzen erörtern die Projektbeteiligten dringende und kurzfristige Aufgaben und Arbeiten, die dann dort besprochen werden. Bei den monatlichen Projekttreffen in Bonn und Göttingen werden mittelfristige Meilensteine und Zielsetzungen festgelegt und die Arbeiten für den kommenden Monat geplant. Das 1. Projekttreffen mit allen beteiligten Akteuren fand am 14.1.2014 in Bonn statt und diente als Abstimmung zwischen den Projektpartnern. Am 19. und 20.5.2014 leiteten die Bonner Fachwissenschaftler einen Workshop zu Mayaschrift und -kalender in Göttingen. Auf diese Weise konnten sich die Mitarbeiter an der SUB mit der Materie vertraut machen. Am 19.6.2014 fand das erste Treffen zum Thema Metadaten und Vokabulare in Bonn statt. Die Entwicklung der virtuellen Forschungsumgebung war im Fokus des Meetings am 28.8.2014 und bei zwei weiteren Projekttreffen am 15.10.2014 sowie am 27.11.2014 wurde das Metadatenschema behandelt. Das für den 10.12.2014 angesetzte Treffen zwischen ULB, SUB und der Arbeitsstelle thematisiert den Import und die Präsentation der epigraphischen Forschungsdaten auf der Präsentationsplattform Visual Library bei der ULB.

Neben der Projektarbeit hat die berufliche Qualifikation der Mitarbeiter einen hohen Stellenwert. Um sich in den Bereich Digital Humanities einzuarbeiten, bildeten sich die Mitglieder der Bonner Arbeitsstelle im Jahr 2014 bei verschiedenen nationalen und internationalen Tagungen und Kursen weiter (Passau, Lausanne, London). Hier lag der Fokus besonders im Bereich Digital Humanities, wo Methoden der digitalen Geisteswissenschaften kennengelernt und angewendet wurden, insbesondere das Erlernen der Auszeichnungssprache XML stand im Vordergrund. Darüber hinaus informierten sich die Mitarbeiter bei den regelmäßigen TextGrid- Nutzertreffen über die Anwendung von TextGrid in den Geisteswissenschaften.

Workshops, Tagungen und Auftaktveranstaltung

2014 wurde das Projekt an fünf nationalen und internationalen Tagungen und Veranstaltungen dem Fachpublikum und der Öffentlichkeit vorgestellt. Vom 31.1. – 2.2.2014 stellten Nikolai Grube, Christian Prager und Elisabeth Wagner das Projekt im Rahmen der 17. Mesoamerikanistentagung in Basel einem Fachpublikum vor. Vorgestellt wurden dabei die Ziele und Aufgaben des Wörterbuchprojekts, und dessen technische Umsetzung. Am 4. und 5.6.2014 stellten die Mitglieder der Forschungsstelle das Projekt im Rahmen einer öffentlichen Ausstellung der Akademie der Wissenschaften im Düsseldorfer Landtag vor. Hierzu wurde im Vorfeld eine Ausstellungssäule mit Bild- und Textmaterial entworfen. Beim Tag des offenen Denkmals am 14. September konnten die Mitarbeiter erneut das Projekt einer breiten Öffentlichkeit vorstellen. Kollegen und Fachleute konnten sich am 14. Oktober bei der Auftaktveranstaltung in Düsseldorf über das Projekt informieren. Im Rahmen des internationalen Workshops “Words in Context: Perspectives and Strategies for the Lexicography of Classic Mayan” im Karl-Arnold-Haus der Wissenschaften, Düsseldorf, stellten alle Mitarbeiter der Forschungsstelle in Bonn und aus Göttingen ihren jeweiligen Arbeitsschwerpunkt einem internationalen Fachpublikum vor. Kollegen aus den USA (David Stuart und Marc Zender), Australien (Peter Mathews) und Deutschland (Gordon Whittaker) wurden eingeladen, an diesem Workshop teilzunehmen und aktuelle Forschungsergebnisse vorzutragen. Bei einem anschließenden Workshop zu Standards und Konventionen bei der Transliteration, Transkription und Analyse der Mayatexte einigten sich die Forscher auf gemeinsame Normen und Standards, die im Projekt zur Anwendung kommen werden. Es ist geplant 2015 den Workshop zu wiederholen um über Fortschritte und Ergebnisse zu diskutieren. Zuletzt wurde das Projekt Fachleuten aus dem Bereich der Digital Humanities in Essen vorgestellt. Im Rahmen des 5. TextGrid-Nutzertreffens am 25. und 26.11.2014 stellten die Mitglieder der Forschungsstelle das Projekt Kollegen und Fachleuten aus den eHumanities vor. Für Februar 2015 ist geplant, dass Projekt international bei einer Digital Humanities Tagung in Graz vorzustellen, so dass hier weitere Kontakte zu anderen DH Projekten geknüpft werden können. Im Dezember 2015 tritt das Akademieprojekt im Rahmen der 20. Europäischen Mayakonferenz als Mitorganisator einer Tagung zu digitalen Methoden in der Mayaforschung auf.

Wissenschaftlicher Beirat und Kooperationen

Im Rahmen unseres Workshop zu Mayaschrift am 14. Oktober konnten wir renommierte internationale Forscher als wissenschaftliche Beiräte unseres Projektes gewinnen: Prof. Dr. David Stuart (Austin), Dr. Marc Zender (New Orleans), Prof. Dr. Peter Mathews (Melbourne) und Prof. Dr. Gordon Whittaker (Göttingen). Es ist geplant, dass sich der Beirat im kommenden Jahr im Rahmen der 20. Europäischen Mayakonferenz in Bonn trifft, wo die Arbeitsstelle ein eigenes Panel zum Wörterbuchprojekt organisieren wird.

Aufgaben und laufende Arbeiten in Bonn

In dieser Sektion des Arbeitsberichts berichten die Mitglieder der Bonner Forschungsstelle über ihre Tätigkeiten und den Stand der Arbeit.

Textarchiv

In der Bonner Arbeitsstelle werden derzeit das **Textarchiv** und verschiedene Datensammlungen und Arbeitslisten erarbeitet bzw. zusammengestellt, die nicht nur zeitnah auf der Projektwebsite

www.mayawoerterbuch.de veröffentlicht, sondern auch in die virtuelle Forschungsumgebung in TextGrid eingepflegt werden. Das in Aktenschränken und digital anzulegende Inschriftenarchiv befindet sich im Aufbau und wird Anfang 2015 eingerichtet. Sobald die notwendigen Räumlichkeiten im Frühjahr 2015 von der Universität Bonn zur Verfügung gestellt werden, wird das analoge Archiv in Aktenform angelegt. Eingegliedert in das Bonner Maya-Inschriftenarchiv wird Berthold Rieses Maya-Inschriftendokumentation, die 135 Aktenorder mit Fotografien, Zeichnungen und Notizen umfasst und im Sommer 2014 der Arbeitsstelle übergeben wurden. Dessen Indexierung und Digitalisierung ist im Gang und wird Ende 2015 abgeschlossen sein. Das 40.000 Bilder umfassende Fotoarchiv von Karl Herbert Mayer (Graz) wird ebenfalls in den kommenden Jahren in das Inschriftenarchiv der Forschungsstelle eingearbeitet. Zu diesem Zweck wurde das Archiv vor Ort inspiziert und eine Bestandsaufnahme durchgeführt, die in einem Projektbericht zusammengefasst ist und auf der Projektwebsite als Arbeitspapier (Working Note) veröffentlicht wird. Das Archiv von Karl Herbert Mayer setzt sich auf Diapositiven, Filmnegativen sowie Abzügen zusammen und wird sukzessive nach Bonn überführt, wo es von Hilfskräften indexiert, digitalisiert und archiviert wird. Sobald die Räumlichkeiten einzugsbereit sind, soll gleichzeitig mit der Anlage des analogen Inschriftenarchivs begonnen werden. Aufgrund der großen Datenmenge wird die Digitalisierung und Indexierung der Fotos mindestens bis Ende 2016 andauern. Das analoge und digitale Inschriftenarchiv der Arbeitsstelle wird nicht nur durch die eigenen Bestände des Projektleiters und der Mitarbeiter angereichert, sondern auch durch Schenkungen von Kollegen im In- und Ausland, die ihre Bildbestände für die Digitalisierung und open access Veröffentlichung zur Verfügung stellen werden. Als Präsentationsplattform dieser Daten dient die Digitale Sammlung der ULB, mit der diesbezüglich eine Kooperation vereinbart ist. Vorbereitungen zur Nutzung der Visual Library Software wurden in diesem Jahr mit allen Beteiligten erörtert und die detaillierte Auseinandersetzung beginnt im Laufe des Jahres 2015.

Fundstättenliste

Im direkten Zusammenhang mit der Anlage des Inschriftenarchivs begannen Christian Prager und eine Hilfskraft ein **Inventar** sämtlicher Inschriften- und Bildträger zusammenzustellen. Grundlage hierzu ist eine **Fundstättenliste**, die derzeit 515 Einträge mit Angaben zum Fundort (Koordinaten, Bibliographie, usw.) umfasst. Hierzu wurden und werden alle einschlägigen Publikationen, Datenbanken und Webseiten von Sammlungen, Museen und anderen Forschungseinrichtungen systematisch bearbeitet und nach Inschriftenträgern der Mayakultur durchsucht um auf dieser Grundlage eine Übersicht über sämtliche publizierten oder dokumentierten Inschriften mit entsprechenden Metadaten (wie etwa Fundort, Abmessungen oder Literaturangaben) zu erhalten. Die Arbeit ist noch im Gange und wird Mitte 2015 soweit abgeschlossen sein, dass sie sämtliche publizierten und dokumentierten Inschriftenträger enthalten wird. Fundort- und Inschriftenliste bilden die Grundlage einer Datensammlung, die nicht nur in die virtuelle Forschungsumgebung eingepflegt, sondern auch über die Webpräsenz mit Hilfe einer SQL-Datenbank veröffentlicht wird. Beide Listen werden ständig gepflegt und sollen über die Onlinepräsenz Fachkollegen und Interessierten zur Verfügung stehen. Die datenbanktechnischen Grundlagen dazu wurden von Sven Gronemeyer vorbereitet und mit Hilfe eines externen Programmierers umgesetzt (siehe unten). Die Arbeiten zum Inventar und zur Datenbank sind noch nicht abgeschlossen. Grundlegend für diese dokumentatorische Arbeit war die Entwicklung eines detaillierten Metadaten- oder Beschreibungsschemas der Inschriftenträger auf der Basis bisheriger Dokumentationen in unserem Fachbereich, das bereits in Grundzügen im Forschungsantrag festgelegt wurde und die Basis für die Arbeit der Metadatenspezialistin Franziska Diehr (SUB) bildet. Details dazu und Stand der Bearbeitung finden sich im Abschnitt über das Metadatenchema.

Literatur

Bisherige Forschungsergebnisse müssen in der virtuellen Forschungsumgebung abgebildet werden und sollen mit Hilfe einer Literaturdatenbank eingebunden werden. Ziel ist es, für jeden Textträger eine Objektbeschreibung und -biographie zu erstellen. Nutzer der Textdatenbank können dadurch Informationen zum jeweiligen Kontext einer Inschrift abrufen. Die virtuelle Arbeitsumgebung enthält daher nicht nur Beschreibungen über die Textträger oder Informationen über den Inhalt der Texte, sondern der Datenbanknutzer bekommt mit Hilfe der Literaturdatenbank auch einen Überblick darüber, welche Autoren sich mit einem Monument befasst oder es publiziert haben, eine Textpassage diskutiert oder erstmals eine bis heute gültige sprachliche Lesung einer Hieroglyphe bzw. eines Zeichens der Öffentlichkeit präsentiert haben. Der Textträger erhält dadurch eine Biographie, die eng mit dem Textinhalten verwoben ist und bei der Bedeutungsanalyse von Wörtern berücksichtigt werden muss. Die Bearbeitung der relevanten Fachliteratur ist daher ein zentraler Punkt, der mit Beginn des Projekt in Angriff genommen wurde. Ziel ist es, grundsätzlich den gesamten Literaturbestand vor Ort für die Anlage des Inschriftenarchivs und die Auszeichnung der Metadaten zu durchsuchen und darauf in der virtuellen Forschungsumgebung sowie auf der Internetpräsenz des Projektes zu verweisen. Für die Erfassung der Literatur wird die freie und quelloffene Anwendung Zotero verwendet um unterschiedliche Online- und Offlinequellen zu sammeln, zu verwalten und zu zitieren. Die Anwendung unterstützt die Bearbeitung und Edition von Literaturangaben und Literaturlisten und erlaubt kollaboratives Arbeiten an unterschiedlichen Standorten. Über eine API können Literaturdaten direkt auf die Website eingespielt werden, so dass über die Onlinepräsenz Literatur gesucht, dargestellt und heruntergeladen werden kann. Ziel ist es, eine vollständige Literaturdatenbank zu Mayakultur aufzubauen, die in die virtuelle Forschungsumgebung bei TextGrid eingespielt und dann mit den Datensätzen verknüpft werden kann. Derzeit wird die Literatur bis 1960 eingearbeitet; Grundlage ist die Bibliographie zur Archäologie, Linguistik und Anthropologie Mesoamerikas von Ignacio Bernal mit über 10.000 Literaturstellen. Jeder Eintrag wird mit Hilfe von Datenbanken und der Originalliteratur auf Vollständigkeit und Richtigkeit überprüft und – wenn vorhanden – mit einer URL verknüpft, wodurch auf eine Onlineversion der entsprechenden Monographie oder Artikels zugegriffen werden kann. Bis dato sind 4.836 Datensätze erfasst und sobald die Einarbeitung von Bernals Bibliographie beendet ist, folgt die Bearbeitung weiterer Fachbibliographien, von Websites u.v.m. Für das Projekt wurde in Kooperation mit der Universitäts- und Landesbibliothek für zwei Jahre die bibliographische Datenbank "Anthropology Plus" erworben, mit deren Hilfe die Datensätze überprüft und weitere Literatur erfasst und eingepflegt werden kann. Die Kosten hierfür teilen sich die Arbeitsstelle und die ULB. Wir rechnen mit einem Bestand von insgesamt 30.000 – 40.000 thematisch markierten Datensätzen (tags), auf die später verwiesen werden kann. Der Datenbestand kann über die Projektwebsite www.mayawoerterbuch.de nach bibliothekarischen Standards recherchiert und open access heruntergeladen werden. Zu diesem Zweck entwickeln derzeit zwei externe Programmierer eine Zotero API, die bis Jahresende fertig gestellt wird, so dass die Literaturdatenbank im Laufe des kommenden Jahres online gestellt und recherchierbar sein wird (siehe Abschnitt über Onlinepräsenz).

Grapheminventar und -konkordanz

Für die epigraphische Arbeit im Projekt werden derzeit verschiedene Arbeitslisten, Diskussionspapiere und Daten zusammengetragen, die im Laufe des Projekts in die virtuelle Forschungsumgebung eingepflegt und über die Projektwebsite direkt abrufbar sein werden. Ein Projekt ist die Erstellung einer Konkordanz sämtlicher Zeichenklassifikationen, die seit 1931 publiziert wurden. Ziel ist es, auf dieser Grundlage einen vollständigen Zeichenkatalog der Mayaschrift zu entwickeln, der die bisherigen Kataloge integriert und als Onlineversion jederzeit revidiert und modifiziert werden kann. Zu diesem

Zweck wurden in den vergangenen Monaten sämtliche Einzelzeichen aus den bisher bekannten Katalogen als Einzeldateien extrahiert und in einer tabellarischen Übersicht mit ihrer jeweiligen Katalognummer zusammengestellt. Mit Hilfe dieser Übersicht lassen sich die Kataloge besser vergleichen und darauf aufbauend ein aktualisierter Zeichenkatalog entwickeln, der die Unzulänglichkeiten der bisher bekannten Klassifikationen korrigiert. Neben der Zeichenklassifikation dient die Tabelle auch als Hauptreferenz für die sprachliche Lesung. In der Tabelle wird festgehalten, wie die phonetische bzw. logographische Lesung des Zeichens lautet, der Grad der jeweiligen Lesungssicherheit angegeben, die Herleitung sowie eine Quellenangabe. Die Liste wird in Teamsitzungen besprochen und kritisch diskutiert. Die Inhalte der Tabelle sind Normdaten und werden in einem Gazetteer gespeichert. Auf diese Weise wird sichergestellt, dass alle Projektmitarbeiter bei der Bearbeitung der Texte dieselbe Lesung eines Zeichens verwenden. Die fast zweihundertseitige Tabelle mit über 10.000 Einträgen wird derzeit bearbeitet und online gestellt, sobald die dafür notwendige Datenbankstruktur für WordPress programmiert und eingerichtet ist. Dies soll im ersten Quartal 2015 erfolgen.

Linguistik

Standardisierungen und Normierungen sind auch im Bereich der Maya-Linguistik dringend notwendig um auf gemeinsame Forschung zu betreiben. Derzeit wird ein Inventar gebundener, grammatikalischer Morpheme mit deren Funktion und Bedeutung erstellt. Es fasst den aktuellen Forschungsstand aller bisher bekannten grammatischen Elemente des Klassischen Maya mit Quellenangabe zusammen und bildet die Grundlage für die weitere Forschung in diesem Bereich. Die Liste wird ebenfalls im ersten Quartal 2015 online verfügbar sein, so dass eine öffentliche Diskussion über Morpheme stattfinden kann. Die Auszeichnung und Benennung von gebundenen ebenso wie lexikalischen Morphemen erfolgt bei der Glossierung, wobei in der Mayalinguistik bisher noch keine einheitlichen Standards gelten (wie dies bereits im Forschungsantrag angesprochen wurde). Aus diesem Grund fand vom 4. bis 6.9.2014 an der Abteilung für Altamerikanistik der Universität Bonn der erste Workshop zu Glossierungsregeln und -standards in Mayasprachen statt. Der jährlich stattfindende Workshop ist eine gemeinsame Initiative der Forschungsprojekte "Textdatenbank und Wörterbuch des Klassischen Maya" und "XML-basierte Erfassungsstandards für koloniale Lexikographien amerindischer Sprachen am Beispiel des K'iche'" (jun. Prof. Frauke Sachse und PD Dr. Michael Dürr). Die Resultate dieses Workshops werden derzeit intern diskutiert und sollen als Working Paper "The Bonn Glossing Rules" auf der Webpräsenz des Forschungsprojektes veröffentlicht werden. Ebenfalls in Arbeit ist eine Konkordanz sämtlicher Wortlisten, die bisher zum Klassischen Maya zusammengestellt wurden (z.B. Boot, Mathews). Ziel ist, nicht nur den Forschungsstand in einer Übersicht zusammenzustellen, sondern auch die unterschiedlichen Transkriptionen des Klassischen Maya zu vergleichen und in die virtuelle Forschungsumgebung einzubinden.

Präsentation der Forschungsdaten in der Digitalen Bibliothek

Die von der Software-Firma Semantics erstellte Software Visual Library der ULB Bonn erhält üblicherweise Metadaten aus dem hbz-Verbundkatalog mittels einer OAI-Schnittstelle. Diese werden innerhalb des Visual Library Managers mit dem Digitalisat verbunden und in den Digitalen Sammlungen der ULB im WWW öffentlich zugänglich präsentiert. Für das Digitale Inschriftenarchiv stammen sowohl Metadaten als auch Digitalisate aus der TextGrid-Datenbank. Daher wurden vorbereitend mehrere mögliche Wege technisch konzipiert, auf denen die Daten abgebildet und in das Visual Library fließen sollen, wie zum Beispiel die Nutzung einer OAI-PMH-Schnittstelle zur Übermittlung öffentlicher Daten des Repositoriums. Auch die Nutzung der Publikationsarchitektur SADE, welche gefilterte Daten

übertragen kann, die sowohl zur VL als auch zur projekteigenen Website fließen, wurden in Betracht gezogen. Ebenfalls wurde die Anwendung des SPARQL-Endpoints in TextGrid erwogen. In diesem Fall würden lediglich Daten übermittelt werden, zu denen im TextGrid-CRUD-Service die Zugänglichkeit als öffentlich definiert wird. Die Konzeption der dargestellten Lösungswege wurde insbesondere in verschiedenen Treffen mit den zuständigen Kollegen zwischen Katja Diederichs, Jan Kenter (ULB Bonn), Max Brodhun (SUB) und Vertretern von Semantics erarbeitet. In Zusammenarbeit mit der ULB und der Firma Semantics wird die Präsentation der Inschriften aufgrund gegebener Daten von TextGrid in den Digitalen Sammlungen der ULB umgesetzt werden. Die Umsetzung betrifft sowohl die technische Einrichtung der Inhaltspräsentation als auch die thematische Gliederung und Browsingstruktur. Ende 2015 werden die ersten Daten in den Digitalen Sammlungen der ULB abgebildet werden.

Onlinepräsentation und Social Media

Onlinepräsentation

Zur Präsentation des Projektes, seiner Fragestellungen, seiner Zielsetzungen und zur Verbreitung seiner Forschungsergebnisse wird eine Webpräsenz (www.mayawoerterbuch.de) eingerichtet. Während die Website primär an ein akademisches Publikum gerichtet ist, hat das Projekt auch eine Mission an die allgemeine Öffentlichkeit, so dass auch grundlegende Informationen zur Mayaschrift und zur Entzifferungsgeschichte gegeben werden, ebenso wie spezielle Forschungsfragen, die eng mit der Zielsetzung des Projektes verknüpft sind.

Der Auf- und Ausbau der Website findet in drei Phasen statt. Phase I ist der konzeptionelle, designerische und funktionale Entwurf auf Basis direkt gecodeter PHP-Seiten mit SQL-Abfragen. Letztere interagieren mit einem MySQL-Datenbankdesign, welches bereits vorhandene Forschungsdaten und Arbeitslisten vorhalten. Phase I dient also vor allem als Machbarkeitsstudie. Die Arbeiten hieran begannen Ende Mai 2014 durch Sven Gronemeyer und wurden größtenteils im Oktober 2014 abgeschlossen. Eine Basisversion dieses statischen Konzeptes wurde im September 2014 live gestellt. Hierauf wurde Phase II eingeleitet, die Portierung des Konzeptes in CMS (Content Management System), welche als mittelfristiger Meilenstein angesetzt war, aber aus mehreren Gründen vorgezogen wurde. Da das Websitekonzept bislang noch recht rudimentär war, gestaltet sich eine Portierung der Styles in CMS-Templates und die Contentmigration einfacher. Aufgrund seines einfachen Frontends, der PHP-nahen Templateerstellung und der Einrichtung multipler Sprachinstanzen fiel die Wahl auf WordPress als CMS, ebenso in Hinblick auf eine moderne, Interaktivität fördernde Seitenstruktur. Ebenso wichtig ist eine zweite Komponente, die Einbindung einer API zur Literaturdatenbank des Projektes auf Zotero, die ca. 30.000-40.000 Einträge umfassen wird. Da für WordPress bereits ein Zotero-Plugin besteht, ist es für Entwickler einfacher, auf diesem Core eine den Projektbedürfnissen angepasste Lösung zu entwickeln. Die Aufsetzung und Anpassung des CMS sowie die Programmierung der Zotero-API wird von der Firma Beuse Projektmanagement in Köln vorgenommen. Ein Go-Live der Website aus Phase II ist für Ende 2014 vorgesehen. Phase III der Website stellt eine zukünftige Entwicklung dar, die erst relevant wird, sobald eine ausreichende Anzahl von fertigen Objekten im TextGrid-Lab vorhanden ist. Über das TextGrid-eigene Modul SADE (Scalable Architecture for Digital Editions) ist eine Publikation von RDF/XML-Daten aus dem TextGrid-Rep auf beliebige Webinstanzen möglich. Neben Anpassungen in SADE muss vor allem im WordPress-CMS eine Schnittstelle geschaffen werden, um das RDF/XML auf der Website publizieren zu können. Voraussichtlich kann die Konzeptionierung von Phase III gegen Ende 2015 eingeleitet werden.

Für die Phasen I und II sind die bereits vom Projekt erarbeiteten (und fortgeführten) Arbeitslisten als zu publizierende Forschungsdaten von Belang. Es handelt sich dabei um eine Fundstättenliste und Museumsliste mit Maya-Artefakten (die beide in Phase II mit einem Maps-Plugin erweitert werden), eine Konkordanzliste aller zehn Zeichenkataloge für die Mayaschrift sowie eine Morphemliste des Klassischen Maya. Alle Arbeitslisten sind nach verschiedenen Kriterien sortierbar, filterbar und durchsuchbar. Weiterhin wird das Projekt laufend technische Berichte (etwa zur Digitalisierung von Archiven) und Arbeitspapiere (etwa zu Glossierungsregeln) veröffentlichen.

Social Media

Ein Aspekt der Website ist die schnelle und breite Bekanntmachung von Forschungsdaten und -fragen, nach Möglichkeit in Interaktion mit dem Nutzer, d.h. der Fachwelt. So kann etwa bei allen Arbeitslisten Feedback gegeben werden, um Diskussionen anzustoßen, die allerdings abseits sichtbarer Kanäle ablaufen würden. Um ein weniger formelles und schnelles Format von großer Reichweite zu schaffen, wird das Projekt neben der eigentlichen Website Kanäle in sozialen Netzwerken nutzen, insbesondere Facebook und Twitter. Es ist geplant diese im kommenden Jahr einzurichten. Hier können etwa Neuigkeiten gepostet, Entzifferungen diskutiert und Veranstaltungen angekündigt werden – neben vielen anderen Möglichkeiten, Fachkollegen und das allgemeine Publikum an der Arbeit des Projektes quasi in Echtzeit teilhaben zu lassen.

Aufgaben und laufende Arbeiten in Göttingen

Die Konzeption und Entwicklung der komputationellen Infrastruktur des Projekts gehört zum Aufgabenbereich unseres Projektpartners SUB in Göttingen. Im Vordergrund steht die Entwicklung des Metadatenschemas sowie die darauf aufbauende Programmierung und Anpassung der virtuellen Forschungsumgebung TextGrid an die Aufgaben, Bedürfnisse und Zielsetzungen des Projekts. Im folgenden fassen wir die bisherigen Arbeiten zusammen und geben einen Ausblick auf die Arbeiten im Jahr 2015.

Metadaten: Anforderungen und Konzipierung

Die umfangreichen Anforderungen des Projekts an die virtuelle Forschungsumgebung stellen sich dementsprechend auch an das Metadatenschema. Es gilt eine Domäne mit einer äußerst heterogenen Datenmenge zu beschreiben: Die Inschriften, Text- und Bildträger, Akteure, Ereignisse, Orte, Quellen uvm. müssen in einer formal logisch strukturierten Form in einem Metadatenschema repräsentiert werden, damit die Daten maschinell les- und verarbeitbar werden. Für die Konzeption des Metadatenschemas können zwei große Ebenen unterschieden werden: I. textuelle Informationen, dazu gehören Transliteration, Transkription, Übersetzung, Textauszeichnung sowie die Komponenten zur Erstellung des Wörterbuchs; II. die nicht-textuellen Objekte, die den Kontext der Inschriften bilden: Text-, bzw. Bildträger, Akteure (sowohl in den Texten erwähnte, als auch für die Forschungsgeschichte relevante Personen), Ereignisse, Orte sowie forschungsrelevante Quellen. Das Metadatenschema wird auf Basis von international anerkannten Standards aufgebaut. Die Nachnutzung von Standards ist immer anzustreben, um eine hohe Interoperabilität mit anderen Systemen zu ermöglichen und gleichzeitig die Qualität des eigenen Schemas durch die Nachnutzung bestehender Konzepte zu verbessern.

Als Basis für das Metadatenschema dienen im Wesentlichen zwei Standards: Das CIDOC CRM zur Repräsentation der nicht-textuellen Objekte und XML-TEI / EpiDoc zur Auszeichnung der Texte. Das CIDOC CRM (CIDOC Conceptual Reference Model) ist ein ISO-Standard, der im Museumsbereich entwickelt wurde, um museale Objekte und Prozesse formal beschreibbar zu machen. Dieses Referenzmodell eignet sich daher optimal zur Beschreibung epigraphischer Prozesse und Objekte sowie der Dokumentation von Forschungsgeschichte. Der umfangreiche Standard der TEI (Text Encoding Initiative) wird seit Jahren erfolgreich zur Auszeichnung von Texten insbesondere im geisteswissenschaftlichen Bereich eingesetzt. Eine Spezialisierung von TEI zur Beschreibung von Inschriften ist EpiDoc. Dieses stellt, im Gegensatz zu TEI, Elemente bereit, die insbesondere zu der Auszeichnung von Inschriften benötigt werden, wie z.B. das Fehlen von Text aufgrund von Verwitterungen.

Jede der beiden Ebenen sind für sich genommen bereits sehr komplex, daher wird die Entwicklung des Schemas in zwei Schritten erfolgen: Im ersten Jahr (2014) wurde und wird der Teil des Schemas für die nicht-textuellen Objekte und 2015 der für die Textauszeichnung und -analyse sowie die Wörterbuchkomponenten entwickelt. Bevor ein Schema konzeptualisiert und erstellt werden kann, müssen die Anforderungen daran erhoben und erläutert werden. Dazu ist sowohl das Wissen über die Schemamodellierung sowie domänenbezogene Expertise notwendig. Für den Modellierungsprozess ist es wichtig, die zu repräsentierende Domäne zu verstehen. Der Wissensaustausch zwischen der Bonner und der Göttinger Arbeitsstelle ist unumgänglich. Dafür wurden unter anderem das Wiki sowie auch Telefonkonferenzen und vor allem persönliche Treffen genutzt (siehe oben). Die Arbeitsgrundlage bildet ein Anforderungskatalog, der alle Elemente, die zu der Repräsentation der Domäne benötigt werden, auflistet und deren Verwendung erklärt. Die Elemente werden Schritt für Schritt begutachtet und revidiert. Dieser Prozess ist zum Teil sehr zeit- und arbeitsintensiv, doch für die spätere Anwendung in der virtuellen Forschungsumgebung unumgänglich: Nur wenn alle Fragen bezüglich der Anforderungen an das Metadatenschema abschließend geklärt sind, lassen sich Probleme bei der technischen Umsetzung umgehen, die zu diesem Zeitpunkt nur noch schwer oder gar nicht mehr zu beheben wären.

Der Anforderungskatalog ist momentan zu zwei Dritteln begutachtet und die dementsprechenden Teile im Metadatenschema umgesetzt. Nach erfolgreicher Klärung der Anforderungen schließt sich eine Phase der intensiven Recherche nach weiteren Standards (z.B. für die Beschreibung von bildlichen Darstellungen) an. Darauf folgend wird eine maschinenlesbare Variante des Schemas und eine ausführliche Dokumentation erstellt.

TEI Metadatenschema für die Erfassung von Textdaten

(von Petra Maier, freie Mitarbeiterin im Projekt) Zielsetzung des Projektes, welches im Rahmen des Weiterbildungsstudienganges Master in Library and Information Science (MALIS) durchgeführt wurde, war die Erstellung eines Metadatenschemas mittels des Datenformates TEI zur Erfassung der Maya-Texte. Das erstellte TEI-Metadatenschema bildet eine Grundlage, welche im weiteren Verlauf den noch offenen Anforderungen angepasst und erweitert werden kann. Zunächst wurden die Anforderungen an das Metadatenschema bezüglich der Inschriftentexte formuliert, wie bspw. die genaue Anordnung der Textfelder (Einzel-/Doppelkolumne etc.) oder auch farbige Bereiche. In einem zweiten Schritt wurden aus den sehr umfangreichen TEI-Metadatensets relevante Module ausgewählt, die für die Erschließung der Inschriftentexte infrage kommen. Zudem wurden die EpiDoc-Empfehlungen, welche eine Eingrenzung der TEI-Guidelines zur Beschreibung von Inschriftentexten darstellen, herangezogen. Für die adäquate Umsetzung der Anforderungen wurde das Metadatenschema in aufeinander aufbauende Abschnitte gegliedert: ‚Inscription‘ sowie der drei Unterabschnitte ‚TextDivision‘, ‚Block‘

und ‚Sign‘. Mit dieser Gliederung des TEI-Metadatenschemas lässt sich die genaue Anordnung der Hieroglyphenblöcke wiedergeben. Farbige Textbereiche und Lücken im Text lassen sich durch die Metadaten beschreiben. Insgesamt ist somit die (grobe) Wiedergabe der Textstruktur mittels der Metadaten möglich. Nicht gelöst ist die Problematik, die Zeichen als Einzelbestandteile der Blöcke eindeutig zu identifizieren. Nach der aktuellen Auszeichnung werden die Zeichen hintereinander vergleichbar einem Fließtext geschrieben. Für die Lösung gibt es derzeit unterschiedliche Vorschläge und Ansätze, die im weiteren Projektverlauf zu prüfen sind. Ebenso bietet das derzeitige TEI-Schema noch keine Möglichkeit für die Auszeichnung von zum Text gehörenden Abbildungen, deren Größenverhältnis sowie die Angabe deren genauen Position auf dem Schriftträger.

Vokabulare und Normdaten

Zur Unterstützung der Auszeichnung nicht-textueller sowie textueller Objekte werden kontrollierte Vokabulare verwendet. Kontrollierte Vokabulare sind für eine ordentliche Dokumentation essentiell: sie ermöglichen begriffliche Kontrolle und Kontinuität, verhindern Ambiguität und Fehler. Wenn möglich, werden Standardvokabulare genutzt. Dafür gibt es zwei Szenarien: I. Ein vorhandenes Vokabular kann umfänglich genutzt werden oder II. es können Konzepte in das eigenentwickelte Vokabular via Matching intergriert werden. Ein Beispiel: Für die Entwicklung eines Thesaurus für Objekttypen und -formen werden teilweise Konzepte aus dem Art and Architecture Thesaurus von Getty angewendet.

Wieviele Vokabulare letztendlich benötigt werden zeigt sich erst nach Abschluss der Entwicklung beider Teile des Metadatenschemas. Die Entwicklung der Vokabulare erfolgt parallel zur Konzeption des Metadatenschemas. Für die maschinenlesbare Repräsentation der Vokabulare wird SKOS (Simple Knowledge Organisation System) verwendet. Für die Entwicklung der Vokabulare und um das Matching auf Standardvokabulare zu vereinfachen, wird ein Tool verwendet. Eine Evaluation nach dem am besten geeigneten Instrument steht noch aus; dies wird zu Beginn des kommenden Jahres erfolgen. Im Projekt werden Normdaten im Bereich der Georeferenzierung eingebunden. Um in standardisierter Weise auf Orte verweisen und diese auch in der Anwendung auf einer Karte visualisieren zu können, verwenden wir die Normdaten von Getty TGN (Thesaurus of Geographic Names) und Geonames.

Informationstechnologie

TextGrid-Lab

Das TextGrid-Lab¹ bildet das Front-End der technischen Infrastruktur. Von dieser Anwendung heraus werden alle Objekte erstellt, um später mit diesen arbeiten zu können. In diesem Zusammenhang sind verschiedene Aspekte relevant, die im Folgenden beschrieben werden.

Benutzer- und Zugriffsverwaltung

Um das kollaborative Arbeiten aller Projektmitglieder zu ermöglichen bietet TextGrid einen Autorisierungs- und Authentifizierungsdienst (TG-auth²) an. Mittels dieses Services ist es möglich, jedem Mitglied des Projektes mögliche spezifische Rollen zu zusprechen. Dabei wird zwischen Projekt-

¹ <https://dev2.dariah.eu/wiki/display/TextGrid/Main+Page#MainPage-Frontend:TextGridLaboratory>

² <https://dev2.dariah.eu/wiki/pages/viewpage.action?pageId=8131296>

Manager, Editor und Observer unterschieden. Zusätzlich ist es möglich, jedem Nutzer die Möglichkeit des Löschens von Objekten zu geben.

Objektverwaltung

Um die Objekte allen Nutzern des Projektes zur gleichen Zeit zur Verfügung stellen zu können wird eine serverseitige Bereitstellung und Speicherung des Datenbestandes notwendig. Dabei sind die Bereiche Erstellen, Lesen, Updaten und Löschen notwendige Mechanismen. Um dies zu ermöglichen bietet TextGrid den Services TG-crud³ an. Im Projekt „Textdatenbank und Wörterbuch des klassischen Maya“ sind dabei folgende Objekte relevant:

- deskriptive Metadaten-Objekte
- Text-Analysedateien
- Bild-Objekte, um auf diesen Markierungen und Referenzierungen vorzunehmen

Für diese Objekte werden die im Folgenden erklärten Anwendungen eingesetzt.

Metadaten-Eingabemaske

Die Metadatenobjekte haben einen heterogenen Charakter und die Gesamtheit der Metadatenobjekte gestaltet sich komplex. Die Beziehungen zwischen den einzelnen Metadaten-Objekten sind dabei von zentralem Interesse. Damit diese Beziehungen effektiv gespeichert werden können und damit ebenfalls die Möglichkeit besteht diese Beziehungen effizient abfragen zu können, werden die Daten im Dateityp des Resource-Description-Framework⁴ (RDF) in XML⁵ Repräsentation angelegt. Bei diesem Datentyp werden logische Aussagen über Ressourcen in der sprachlich bekannten Form Subjekt-Prädikat-Objekt gespeichert (Beispiel: Konrad Zuse - entwickelte - den Z3). Um die Eingabe der Metadaten für den Anwender möglichst komfortabel und unkompliziert zu gestalten, wurde eine spezielle Eingabemaske erstellt, die ein HTML Formular zur Verfügung stellt. Aus diesem Formular werden die jeweiligen Daten im Folgenden in RDF/XML umgewandelt und in TextGrid gespeichert. Durch die Speicherung in RDF Format wird die Möglichkeit gegeben die Daten in einem Datennetz zu verbinden (Linked Data)⁶. Durch diesen Aspekt können Zusammenhänge erkannt werden, die im Vorhinein noch nicht ersichtlich waren. Des Weiteren besteht ein hoher Grad der Austauschbarkeit mit Daten aus anderen Projekten. Durch eine offene Gestaltung der Daten und die Verbindung in einem offenen Datennetz (Linked Open Data) wird dieser Punkt weiter verstärkt.

Die Eingabemaske soll dem epigraphischen Workflow “Dokumentation - Analyse - Edition” möglichst nah nachempfunden werden. Für diesen Zweck wurden weitere Anpassungen, wie das Erstellen eines spezifischen Objekts zu einem bestimmten Zeitpunkt ermöglicht. Die korrekte Eingabe der Metadaten ist von zentraler Bedeutung. Zur Unterstützung dieses Prozesses werden der eingebenden Person Hilfsmittel zur Verfügung gestellt. Die Auswahl von bestimmten Begriffen aus einem Menü, welche aus einem kontrolliertem Vokabular generiert werden, verhindert beispielsweise Rechtschreibfehler. Bei einer Referenzierung zu einem anderen Objekt wird die Datenbank abgefragt, um den entsprechenden eindeutigen Identifier aus der Datenbank zu übernehmen, um auch an dieser Stelle Rechtschreibfehler zu vermeiden. Des Weiteren werden Funktionen zur Validierung bestimmter Inhalte verwendet. So wird zum Beispiel das Speichern einer unvollständigen URL verhindert.

³ <https://dev2.dariah.eu/wiki/display/TextGrid/TG-crud>

⁴ <http://www.w3.org/RDF/>

⁵ <http://www.w3.org/XML/>

⁶ <http://www.w3.org/wiki/LinkedData>

Datenbank und Speicher

Alle Objekte werden in erster Linie auf dem TextGrid Server gespeichert. Für die effiziente Speicherung der Objekte mit ihren Beziehungen und die Abfrage der Daten werden die Daten redundant in einer speziellen Datenbank für RDF-Daten abgelegt. Diese Datenbanken werden als Graphdatenbanken bezeichnet. Für die RDF-Tripels im speziellen spricht man von Tripelstores⁷. Im Projekt wurde dabei für den Tripelstore BigData⁸ entschieden, da dieser eine hohe Anzahl von Tripeln speichern kann und Möglichkeiten der Parallelisierbarkeit bietet, um vor eventuellen Ausfällen der Datenbank geschützt zu sein. Bei der Abfrage der Daten wird auf den Standard SPARQL Protocol and RDF Query Language⁹ (SPARQL) gesetzt. Die Kombination aus RDF – Tripelstore und SPARQL hat sich in der gängigen Praxis bei komplexen und heterogenen Datenlandschaften als effizient und effektiv herausgestellt.

Speicher

Für die erste Phase des Projektes wurden 3 TB Speicher des Projektes DARIAH-DE¹⁰ beantragt. Dieser Speicher liegt beim Göttinger Rechenzentrum GWDG¹¹ und bietet eine kostenneutrale Lösung.

Publikation der Objekte

Sowohl die erstellten RDF-Metadatenobjekte, als auch die analysierten und in TEI ausgezeichneten Volltexte sollen nach der Fertigstellung auf verschiedene Plattformen publiziert werden. Für die Veröffentlichung im TextGrid Repository sind Mechanismen in das TextGrid Lab eingebaut und für den gesamten Workflow optimiert. Weitergehend werden weitere Mechanismen für die Veröffentlichung bereitgestellt. Die Anwendung Scalable Architecture for Digital Editions¹² (SADE) bietet eine Möglichkeit alle Objekte auch an externen Stellen zu veröffentlichen. Mit dieser Möglichkeit kann auf die Veröffentlichung im TextGrid Rep verzichtet werden. Allerdings wird parallel auch auf die Mechanismen der Langzeitarchivierung verzichtet. Um die Daten bereits in einer aufbereiteten und an das Design (mittels eines bei Bedarf angepasstem CSS¹³) der Projekt-Website angepassten Form präsentieren zu können, werden die die Objekte an dieser Stelle mittels SADE veröffentlicht. Diese Publikation erfolgt parallel zum TextGrid Repository.

Als dritte Möglichkeit wird das Extrahieren der Metadaten-Objekte mittels einer OAI-PMH¹⁴ Schnittstelle angeboten.

Volltextindizierung

Die ausgezeichneten Volltexte sollen im Nachhinein mittels einer Volltextsuche durchsucht werden können. Für diese Funktion wird das bereits in der TextGrid-Infrastruktur eingebaute TG-search¹⁵ verwendet. Diese nutzt für die Suchanfragen Elasticsearch. Für diesen Zweck werden die Daten bei der Indizierung vorher in das Format JavaScript Object Notation Format¹⁶ (JSON) umgewandelt.

⁷ <http://www.w3.org/wiki/LargeTripleStores>

⁸ <http://bigdata.com/>

⁹ <http://www.w3.org/TR/rdf-sparql-query/>

¹⁰ <https://de.dariah.eu/>

¹¹ <http://www.gwdg.de/>

¹² <http://www.bbaw.de/telota/software/sade/sade-1>

¹³ <http://www.w3.org/Style/CSS/>

¹⁴ <http://www.openarchives.org/pmh/>

¹⁵ <http://www.openarchives.org/pmh/>

¹⁶ <http://www.w3schools.com/json/>

Bonn und Göttingen, 1. Dezember 2014

i.A. Dr. Christian Prager



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>