

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

WORKING PAPER I | DE



Published 8 Apr 2015

DOI: 10.20376/IDIOM-23665556.15.wp001.de

Die “Open Science”-Strategie im Projekt “Textdatenbank und Wörterbuch des Klassischen Maya”

Katja Diederichs¹

¹ Rheinische Friedrich-Wilhelms-Universität, Bonn

Im folgenden Beitrag wird die Veröffentlichungs- und Wissensvermittlungsstrategie des Forschungsprojektes Textdatenbank und Wörterbuch des Klassischen Maya vorgestellt und erläutert. Projektziel ist es, mittels digitaler Technik die epigraphischen und objektgeschichtlichen Inhalte sämtlicher gegebener Hieroglyphentexte datenbankgestützt zu erschließen. Auf Grundlage der dadurch entstehenden Objekt- und Textdatenbank soll gegen Ende der Projektlaufzeit ein umfassendes Wörterbuch der Klassischen Mayasprache erstellt werden. Das von öffentlichen Geldern geförderte Forschungsprojekt sieht es hierbei als Aufgabe an, die Wissensinhalte und Ergebnisse seiner Forschung der Öffentlichkeit frei zugänglich zu machen. Diese Öffnung wird allumfassend, d.h., den gesamten Forschungsprozess in Bezug auf Methodik, Inhalte und Ergebnisse betreffend, betrachtet. Die dabei vertretene Veröffentlichungspolitik wird im Sinne einer modernen, den Zeitgeist der Open Science widerspiegelnden, digitalen Geisteswissenschaft verstanden.

Das Projekt: “Textdatenbank und Wörterbuch des Klassischen Maya”

Das Projekt “Textdatenbank und Wörterbuch des Klassischen Maya” (kurz: “TWKM”) zur Erforschung von Schrift und Sprache der Mayakultur wurde 2014 von der Nordrhein-Westfälischen Akademie der Wissenschaften und der Künste als Arbeitsstelle an der Philosophischen Fakultät der Universität Bonn eingerichtet. Ziel ist es, mittels digitaler Technik in den kommenden 15 Jahren die epigraphischen und objektgeschichtlichen Inhalte sämtlicher Hieroglyphentexte datenbankgestützt zu erschließen. Auf Grundlage dieser Daten wird gegen Ende der Projektlaufzeit ein umfassendes Wörterbuch der Klassischen Mayasprache kompiliert und herausgegeben (cf. Uni Bonn n.d.).

Dieses wird auf der Basis der verfügbaren Textträger der vorspanischen Mayakultur, die in der Zeit zwischen dem 3. Jhdt. v. Chr. – 1500 n. Chr. auf dem Gebiet der heutigen Staaten Mexiko, Guatemala, Belize und Honduras vorherrschte, und deren Sprache auf Tausenden von Inschriftenträgern erhalten ist, erstellt. Diese bilden den in Hieroglyphenschrift abgefassten Sprachschatz ab und dokumentieren dadurch in Originalschreibung und Umschrift die Sprache, ihre Sprachstadien und Varietäten. Hierzu ist die Anlage einer Textdatenbank erforderlich. Diese ist in zwei Instanzen verfügbar: einerseits in

einer nicht öffentlichen Expertenversion mit Zugangsberechtigung und zweitens in einer öffentlichen Form, deren Inhalte weltweit frei über das Internet abrufbar sein werden.

Neben dem Vorhaben, ein Lexikon des Klassischen Maya zu kompilieren und Wörter auf ihre Bedeutungen hin zu analysieren, soll für jeden Textträger auch eine Objektbeschreibung erstellt werden. Nutzer der Textdatenbank können dadurch Informationen zum jeweiligen Kontext einer Inschrift abrufen. Diese enthält nicht nur Beschreibungen über die Textträger und über den Inhalt der Texte, sondern die Datenbanknutzung mit Hilfe der Literaturdatenbank verschafft auch einen Überblick darüber, welche Autorinnen und 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 Textinhalt verwoben ist und bei der Bedeutungsanalyse von Wörtern berücksichtigt wird. Die derzeit im Aufbau befindliche digitale Arbeitsumgebung innerhalb einer Virtuellen Forschungsumgebung (kurz: VFU) orientiert sich stark am Workflow der epigraphischen Arbeit, der üblicherweise mit der Dokumentation der Textträger und der Erfassung von Beschreibungsdaten beginnt, bei der epigraphischen Analyse mit der Zeichenklassifikation, Transliteration und Transkription der Texte fortfährt, bei der morphologischen Segmentierung und linguistischen Interpretation weiterfährt und optimaler Weise bei der Übersetzung und Publikation einer Inschrift endet. Neben der Korpusanalyse als Grundlagenarbeit der Lexikographie werden für das später zu erstellende Wörterbuch Phonologie, Morphologie, Syntax, Semantik und Pragmatik des Klassischen Maya untersucht, aber auch Fragen der historischen Linguistik, Untersuchungen zur Schriftgeschichte oder dem Schriftgebrauch werden in das Blickfeld der Forschung gerückt. Das digitale Sprachwörterbuch wird das Klassische Maya nicht nur in Umschrift, sondern auch die Originalschreibung abbilden. Auf diese Weise wird die Erforschung und Dokumentation der Sprach- und Schriftgeschichte gleichzeitig in Angriff genommen (cf. Uni Bonn n.d.).

Angesichts der Datenmenge kann dieses Unterfangen nur mit Mitteln und Methoden aus den digitalen Geisteswissenschaften durchgeführt werden. Methodisch ist das Langzeitvorhaben daher in diesem Feld verortet und wird in Kooperation mit dem Forschungsverbund für eHumanities TextGrid (der Abteilung Forschung und Entwicklung der Niedersächsischen Staats- und Universitätsbibliothek Göttingen) und mit der Universitäts- und Landesbibliothek Bonn verwirklicht. Mit der Ausrichtung auf die digitalen Geisteswissenschaften bildet das Wörterbuch-Projekt eine wichtige Schnittstelle zwischen Humanities und Informatik an der Universität Bonn (cf. SUB n.d.).

“Open Access” als Publikationsweg für die Wissenschaft im digitalen Zeitalter

Die Zeichen des digitalen Zeitalters sind sowohl im gesellschaftlichen Wandel als auch in einem Wandel der Arbeitsweise in der modernen Wissenschaft zu erkennen. Beide sind heutzutage zu einem großen Teil bestimmt durch computergestütztes Arbeiten und digitale Kommunikation. Die Wissenschaften widmen sich seit einiger Zeit verstärkt den Bemühungen, die neuen digitalen Werkzeuge sinnvoll in den Wissenschaftsbetrieb einzubinden. Zu diesem Zwecke wurden von vielen, sich der Wissenschaft verschreibenden Vereinigungen zunächst selbstverpflichtende Empfehlungen für Richtlinien erarbeitet, die zeitgenössische Wege der Publikation sowie Wissensvermittlung in der Wissenschaft gewährleisten sollen. Ein Beispiel dafür sind die 2002 von der “Budapest Open Access Initiative” erstellten Grundsätze, in welchen der Begriff des “Open Access” für die freie Publikation wissenschaftlicher Inhalte erstmalig geprägt wurde (cf. BOAI 2002). Weitere, u.a. im deutschsprachigen Raum ausgearbeitete Empfehlungen wie etwa die “Berliner Erklärung über den offenen Zugang zu wissenschaftlichem Wissen” aus dem Jahre 2003, in der ein offener Zugang zu wissenschaftlichem Wissen angestrebt wird, folgten (cf. Berliner Erklärung 2003). Auch in der Berliner

Erklärung wurde ein Paradigmenwechsel zur "Open Access"-Veröffentlichung hin angestoßen und der offene Zugang als erstrebenswertes Verfahren betrachtet, was idealerweise die aktive Mitwirkung eines jeden Urhebers wissenschaftlichen Wissens und eines jeden Verwalters von kulturellem Erbe voraussetzt. Hierbei wurde die Beschreibung von "Open Access" etwas weiter gefasst. Sie beschränkte sich nicht nur auf die Inhalte in Form von Publikationen wissenschaftlicher Forschungsergebnisse sondern ebenso auf die Ursprungsdaten, Metadaten, Quellenmaterialien, digitale Darstellungen von Bild- und Graphik-Material sowie wissenschaftliches Material in multimedialer Form. Der offene Zugang zu Wissen und dessen freie Verfügbarkeit sollen hierbei gewährleistet werden durch das Internet, das als "umfassende Quelle menschlichen Wissens und kulturellen Erbes" (Berliner Erklärung 2003) betrachtet wird. Die Forderung griff weiterhin das zu bewerkstelligende Problem der strukturellen Herausforderungen an die Wissenschaftsgemeinde der Zukunft auf, indem eingeräumt wurde, dass die Vision von einer umfassenden und frei zugänglichen Repräsentation des Wissens sich nur realisieren lasse, wenn sich das Internet der Zukunft durch Nachhaltigkeit, Interaktivität und Transparenz auszeichnet. Somit wurde ein Prozess in Gang gesetzt, der zur Bildung einer offeneren Wissenschaftsumgebung mithilfe der Informationstechnologie und der neuen Werkzeuge des Mediums Internet führte, welche über "Open Access" hinaus den Zugang zum gesamten wissenschaftlichen Wissen -in anderen Worten "Open Content"- ermöglichen sollen (cf. Hilf & Severiens 2013). Dieser Prozess dauert an, da sich innerhalb der fortschreitenden digitalen Welt auch die zeitgemäße digitale Wissenschaft weiterentwickelt. Innerhalb dieser wird die Entwicklung "zu einer neuen Qualität wissenschaftlichen Informationsmanagements und damit wissenschaftlicher Forschungsarbeit" führen (Hilf & Severiens 2013).

Digital Humanities und eHumanities

Um diese richtungsweisenden, jedoch im Einzelnen nicht konkret ausformulierten oder verpflichtend geltenden Standards zur Gestaltung einer modernen, offenen Wissenschaft zu schaffen, müssen bestimmte technische und methodische Maßnahmen ergriffen werden, welche in jeder wissenschaftlichen Fachdisziplin unterschiedlicher Art sind und somit verschiedenartige Expertisen erfordern.

Innerhalb der geisteswissenschaftlichen Disziplinen muss sich den neuen Herausforderungen gestellt werden, Wege zu finden, wie die Methodik und die technische Realisierung einer digitalen Wissenschaft der jeweiligen Disziplin entsprechend zu gestalten und umzusetzen sind. Die Grenze zwischen den traditionellen Geisteswissenschaften und den Computerwissenschaften bzw. der Informationstechnologie wird infolgedessen zunehmend überschritten. Die sogenannten Digital Humanities, welche sich – sehr grob formuliert – mit sowohl informationswissenschaftlichen als auch geisteswissenschaftlichen Methoden und Fragestellungen befassen, sind hierdurch entstanden. Einerseits werden mit ihrer Hilfe durch digitale Mittel die eigenen Forschungsstrukturen und -methoden einer jeweiligen Fachdisziplin der Geisteswissenschaften erweitert. Andererseits setzte die Notwendigkeit, methodische Probleme in Bezug auf geisteswissenschaftliche Fragestellungen zu lösen, Impulse in der Informatik und Informationswissenschaft, welche die Umsetzung von technologischen Lösungen antreiben (cf. Thaller 2012). Diese Wechselwirkung zwischen den Geisteswissenschaften und der Informationstechnologie entwickelt sich fort. Ein weiteres Feld, das der eHumanities ist – teils synonym verwendet – unter sehr ähnlicher Thematik in diesem Spannungsfeld einzuordnen. Die eHumanities können verstanden werden "als Summe aller Ansätze, die durch die Erforschung, Entwicklung und Anwendung moderner Informationstechnologien die Arbeit in den Geisteswissenschaften erleichtern oder verbessern wollen" (BMBF 2013). Die Digital Humanities sowie die eHumanities tragen ihren Teil dazu bei, um die Geisteswissenschaften methodisch und strukturell in die digitale Wissenschaft hinein zu bewegen.

Digitale Epigraphie des Klassischen Maya

Im geisteswissenschaftlichen Bereich, in dem sich das hier beschriebene TWKM-Projekt einordnen lässt, wird der Fokus auf geistiges Gut und kulturelles Erbe gerichtet. Im Bereich der epigraphischen Analyse der Klassischen Maya-Schrift steht die Diskussion über die teilweise noch zu entschlüsselnden Bestandteile der diesem Schriftsystem zugrundeliegenden Sprache im Vordergrund. Hierbei spiegelt sich der wissenschaftliche Diskurs zu den textuellen Inhalten einer Inschrift mit einem zu entschlüsselnden Text in verschiedenen Veröffentlichungen dazu wider. Die Beschreibung etwa einer Hieroglyphe für einen Wörterbucheintrag ist demnach nicht denkbar ohne die verschiedensten Hypothesen zu Lesarten, welche im Laufe der Jahre gesichert oder aber kontrovers diskutiert werden, mit dem betreffenden Lemma zu verbinden. In der digitalen Epigraphie werden all diese Informationen computergestützt erfasst und als Daten und Metadaten in digitaler Form maschinenlesbar gestaltet, um somit in einer Datenbank miteinander sinnvoll verknüpft zu werden.

Eine Datenbank als Grundlage für ein Wörterbuch

Für das Projekt wird nicht nur die umfassende Bibliographie an Forschungsliteratur der letzten Jahrzehnte zusammengestellt, sondern die Einträge werden auch inhaltlich in einer Datenbank mit den betreffenden Maya-Monumenten und -Textträgern verknüpft. Neben der allgemeinen Forschungsgenese zu Fragestellungen bezüglich der Sprache spielt die individuelle Objektbiographie eines Textträgers für die epigraphische Analyse einer Inschrift eine große Rolle, um ihren Inhalt zu verstehen. Dazu werden neben den textuellen auch die bildlichen Inhalte und die Objektdaten wie z.B. der archäologische Grabungskontext, historisch-kulturelle Informationen über damalige politische, soziale, dynastische etc. Verhältnisse als Metadaten annotiert, so dass sowohl eine Korpus- als auch Objektdatenbank aufgebaut werden kann. Diese bildet neben der epigraphischen Analyse sämtlicher Inschriften auch eine Art Ontologie zum kulturhistorischen Bereich "Klassik der Mayakultur" mitsamt Forschungsliteratur ab. Somit kann über rein epigraphische Fragestellungen hinaus eine semantische Suche an der Datenbank ermöglicht werden, welche je nach Fragestellung neben einer inhaltlichen Antwort auch die zutreffende Forschungsliteratur zum Diskurs liefern kann.

Bei der Projektarbeit ergibt sich gewissermaßen ein Mehrwert in der Nutzung der Text- und Objektdatenbank, welche im eigentlichen Sinne das Quellenmaterial für das Projektziel zur Verfügung stellt: Für das Wörterbuch des Klassischen Maya. Dieses am Ende der Projektlaufzeit aus dem umfassenden Korpus an Inschriften aus der Textdatenbank kompilierte Wörterbuch des Klassischen Maya bildet in seiner Entstehung hierbei gewissermaßen den epigraphischen Workflow in der digitalen Epigraphie ab.

Zielpublikum des Projektes

Als ein von öffentlichen Geldern gefördertes Forschungsprojekt, welches in den digitalen Geisteswissenschaften einzuordnen ist, sieht es das Projekt als Aufgabe an, die Wissensinhalte und Ergebnisse seiner Forschung nicht nur einem Fachpublikum, sondern auch der breiteren Öffentlichkeit zugänglich zu machen.

Im Jahre 2004 erstellten die Wissenschaftsministerien verschiedener OECD-Länder einen Forderungskatalog, die "Declaration on Access to Research Data from Public Funding" (cf. OECD 2004), aufgrund dessen 2007 von der OECD die Empfehlungen und Prinzipien für den Zugang zu aus öffentlichen Geldern finanzierten Forschungsprojekten ausgearbeitet wurden, die sogenannten "Principles and Guidelines for Access to Research Data from Public Funding" (cf. OECD 2007). Angelehnt an diese Empfehlungen ist es Ziel, die größtmögliche Zugangsmöglichkeit zu den im Projekt erstellten und zusammengetragenen Daten zu gewährleisten. Hierbei müssen jedoch die

verschiedenen nationalen gesetzlichen und lizenzrechtlichen, die Urheberschaft währenden Beschränkungen, die einigen Daten auferliegen, berücksichtigt werden.

Im Sinne dieser und ähnlicher Appelle sowie selbstverpflichtenden Vereinbarungen zur Verwirklichung einer offenen Wissenschaft wird erhofft, dass der optimale Zugriff zu den im Laufe des Projektes entstehenden Informationen die Kooperation innerhalb der Fachwissenschaften fördert. Da im Projekt dem Grundsatz gefolgt wird, dass der Wert von Daten in ihrer Nachnutzung liegt (cf. OECD 2007: 11), wird Zusammenarbeit als Möglichkeit gesehen, gemeinsam die entstehenden Wissensinhalte zu verifizieren. Somit kann der wissenschaftliche Diskurs und Fortschritt, der u.a. zum Entstehen eines das Klassische Maya vollständig umfassenden Wörterbuchs beiträgt, in der gegebenen Disziplin bestmöglich vorangetrieben werden. Das entstehende Wörterbuch soll nicht nur der Wissenschaft sondern auch der interessierten Öffentlichkeit vollständig und finanziell, gesetzlich sowie technisch barrierefrei zur Verfügung stehen. Dabei wird die Auffassung vertreten, dass die Ergebnisse des Projektes über den wissenschaftlichen Bereich hinaus von Nutzen sind, da sie das kulturelle Erbe der Mayakultur der dafür aufgeschlossenen Öffentlichkeit näher bringen können, was außerdem die Wertschätzung für aus öffentlichen Geldern finanzierte Datensammlungen in der Gesellschaft steigern kann (cf. OECD 2004).

“Open Science”-Strategie

Bezogen auf die Auffassung von moderner Wissenschaft als “Open Science” (cf. OpenScienceASAP n.d.) ist nicht nur der Zugang zu Forschungsergebnissen sondern die Öffnung des gesamten wissenschaftlichen Prozesses wichtig.

Das bedeutet, das Projekt arbeitet daran, der interessierten Öffentlichkeit sowie der Wissenschaft einen allumfassenden offenen Zugang zu ermöglichen zu sowohl seinen wissenschaftlichen Publikationen (“Open Access”) und zur freien Dokumentation seiner Methodik und Arbeitsprozesse (“Open Methodology”) als auch zu seinen Forschungsdaten (“Open Data”) und zu der vom Projekt genutzten Software (“Open Source”). Hierbei wird erwartet, im Laufe des Projektes durch eine transparente und zeitnahe Dokumentation von Zwischenergebnissen sowie durch die Bereitstellung einer von der Allgemeinheit gebräuchlichen Infrastruktur eine gemeinschaftliche Beteiligung am themenbezogenen Diskurs zu fördern und innerhalb der Fachdisziplin sowie interdisziplinär einen Austausch in Bezug auf etwa Methodik, Technik sowie Inhalte anzuregen, die das Projekt in seiner Entstehung mitgestalten können (cf. OpenScienceASAP n.d.). Ein ungehinderter Zugang zu den Forschungsinhalten des Projektes und ebenso die Gewährleistung deren produktiver Nachnutzung muss durch freie Lizenzen verwirklicht werden (cf. DFG 2014). Die Strategie einer “Open Science” in diesem Sinne fordert damit einhergehend eine den gesamten wissenschaftlichen Prozess abdeckende Bandbreite an gängigen freien Zugangsberechtigungen, welche durch das Schlagwort “Open” gekennzeichnet werden und folgendermaßen beschrieben werden können:

- “Open Access” beschreibt das offene Publizieren, was die betreffenden Inhalte für alle zugänglich und nutzbar macht.
- “Open Methodology” betrifft die nachvollziehbare Dokumentation und Publikation der Anwendung von Methoden und Werkzeugen, die zur Datenerstellung und Analyse genutzt werden, sowie des gesamten damit verbundenen Arbeitsprozesses.
- “Open Source” bezieht sich auf das Nutzen sowie das Gestalten und Dokumentieren eigener quelloffener Technologien mitsamt der Schaffung einer Möglichkeit, diese ohne Barrieren nachzunutzen.

- “Open Data” umfasst das Veröffentlichen von erstellten Rohdaten und Metadaten in standardisierten, offenen Formaten sowie das Schaffen von datenübergreifenden Strukturen, welche es ermöglichen, auf die Daten zuzugreifen und diese frei und auf jegliche Weise nachzunutzen (cf. Kraker et al. 2011: 645).

Die verschiedenen Plattformen des Projektes

Innerhalb der Arbeitsprozesse des Projektes fließen die einzelnen “Open”-Strategien ineinander über oder bedingen einander. Dennoch sollen nun die einzelnen Prinzipien sowie die Projekt-Strategien in Bezug darauf im Folgenden näher erörtert werden. Hierbei werden zunächst die verschiedenen Plattformen vorgestellt, auf denen die im Laufe des Projektes erstellten Forschungsergebnisse, Daten und Metadaten publiziert und zur Verfügung gestellt werden.

TextGrid Laboratory und Repository

Die Speicherung der textuellen und graphischen Korpusdaten, insbesondere zur nachhaltigen und dauerhaften Verfügbarkeit und Nachnutzbarkeit wird operativ durch den Projekt-Partner, die Abteilung für Forschung und Entwicklung der SUB Göttingen, in Kooperation mit der Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen (kurz: GWDG) sichergestellt. Die dort seit 2006 entwickelte Virtuelle Forschungsumgebung für die Geisteswissenschaften TextGrid ermöglicht hierbei die Archivierung der Materialien, welche mittels persistenter Identifikatoren (kurz: PIDs) zitierbar werden, auf einem durchsuchbaren Online-Repository, dem TextGrid Repository, auf dem sie aufbewahrt werden. In diesem Langzeitarchiv werden die Daten weltweit unter der “Open Access”-Lizenzregelung entgeltfrei durch das Internet zugänglich gemacht (cf. DARIAH.de n.d.).

Zur Erstellung der Daten und Digitalisate für die Projekt-Datenbank wird zunächst das TextGrid Laboratory (kurz: TextGrid Lab) genutzt. Hierzu werden alle im Projekt zu erstellenden Daten wie Digitalisate und Metadaten, d.h. Objektdaten und linguistische Analysen durch ein projekteigenes Frontend eingegeben und erstellt. Das Frontend befindet sich innerhalb der VFU, im TextGrid Lab, worin autorisierte Projektmitarbeiter die Daten und Metadaten mittels einer für das Projekt eigens entworfenen Eingabemaske erstellen und bearbeiten können. Das Laboratory beherbergt einen dynamischen Speicher. Das bedeutet, die Daten sind zunächst änderbar, editierbar und können mit anderen Daten logisch verbunden und aggregiert werden. Zum TextGrid Lab wird nach Registrierung Zugang gewährt, dieser ist kostenfrei. Um jedoch den vollen Zugang zu den für die Arbeit freigegebenen Projektdaten zu erhalten und diese gegebenenfalls mitzugestalten, ist über die Registrierung hinaus eine Zulassung durch den Projektmanager zur Teilnahme am Projekt vonnöten.

Nachdem die Daten innerhalb des TextGrid Labs auf Konsistenz und formale Validität geprüft wurden, können sie schließlich in einem weiteren Schritt veröffentlicht werden. Hierzu werden sie auf das öffentliche TextGrid Repository gespielt, welches einen statischen Speicher beinhaltet. Das heißt, die hier befindlichen Inhalte können nicht mehr ohne weiteres verändert oder gelöscht werden. Das Repository ist durch das World Wide Web frei zugänglich, und die aus dem Laboratory übertragenen Inhalte sind dort durchsuchbar. Darüber hinaus sind die Daten und Metadaten downloadbar.

Virtuelles Inschriftenarchiv

Eine weitere Präsentation ausgewählter Projekt-Daten wird in Kooperation mit der Universitäts- und Landesbibliothek Bonn (kurz: ULB) angeboten. In den Digitalen Sammlungen der ULB-Webpräsenz wird ein Virtuelles Inschriftenarchiv eingerichtet. Dieses präsentiert ausschließlich "Open Access"-Inhalte frei zugänglich durch das Web. Hierzu werden ausgewählte Daten und Metadaten zu Textträgern mitsamt ihrer linguistischen Analysen aus dem TextGrid Repository an den ULB-Speicher übermittelt. Die zu übertragenden Inhalte müssen hierzu auf ein anderes Metadatenformat "gemappt" werden, d.h., die Inhalte des einen Formats müssen in ihrer Semantik konsistent dem anderen Format zugeordnet werden, damit sie später in den Digitalen Sammlungen entsprechend abgebildet werden können. Das dort genutzte Format besteht aus den für Digitale Bibliotheken bewährten, offenen XML-basierten Standards METS und MODS, deren Schemadefinitionen darüber hinaus die Inhalte zum DFG-Viewer kompatibel machen (cf. DFG-Viewer 2015). Die zugeordneten Daten werden dazu aus dem TextGrid-Speicher mittels einer Schnittstelle der ULB-Software übermittelt, welche die Inhalte in der vom Projekt zuvor definierten Strukturierung im Virtuellen Inschriftenarchiv der Digitalen Sammlungen abbildet. Die auf dem OAI-PMH (The Open Archives Initiative Protocol for Metadata Harvesting) basierende Schnittstelle ermöglicht hierbei weiteren Diensten und Anwendungen durch den freien Zugang zu den Projekt-Metadaten des Repositoriums, deren Nachnutzung in einem anderen Kontext, was für die Metadaten darüber hinaus einen Mehrwert ergibt.

Projekteigene Website

Weitere projekteigene Informationen aber vor allem Publikationen, die im Laufe des Projektes entstehen, werden auf der projekteigenen Website, die eine eigene ISSN (International Standard Serial Number) besitzt, kostenlos in digitaler Form zur Verfügung gestellt. Hier werden Abhandlungen, Working Papers sowie weitere, in den Bereich der grauen Literatur fallenden Inhalte wie z.B. Technical Reports und Dokumentationen der Arbeit des Projektes veröffentlicht werden. "Open Access"-Erstpublikationen werden nach dem goldenen sowie Zweitveröffentlichungen über den grünen Weg veröffentlicht (cf. Internet Portal OA n.d.). Hierbei erfüllt die Website die Kriterien zur Veröffentlichung wissenschaftlicher Publikationen in fortlaufender Weise unter der gegebenen ISSN. Diese entspricht den Sammelrichtlinien der Deutschen Nationalbibliothek zur Zuteilung einer ISSN (cf. Matthias & Wiechmann 2014); außerdem werden die zu veröffentlichenden Inhalte wie regelmäßig erscheinende Einträge eines Blogs mit wissenschaftlicher Ausrichtung behandelt.

Mit der Zuteilung einer ISSN wird die Website als Blog im Katalog der Deutschen Nationalbibliothek und im internationalen ISSN-Portal nachgewiesen (cf. Schütz 2015). Damit ist etwa die Zitierfähigkeit der Inhalte des Blogs als Netzpublikation mit offenem Adressatenkreis gewährleistet (cf. Matthias & Wiechmann 2014).

"Open Access"-Strategie

Der Begriff "Open Access" beschreibt ein mittlerweile weltweit bekanntes Paradigma zur freien Zugänglichkeit und Veröffentlichung von Inhalten verschiedener Art und Herkunft. Dem Projekt ist nicht nur die Bereitstellung von Forschungsergebnissen und ähnlicher Literatur über "Open Access"-Publikationen wichtig sondern auch die Veröffentlichung der Daten bzw. Inhalte, über die Forschung betrieben wird. Diese betreffen sowohl digitalisierte Inhalte zu rein textuellen Inhalten als auch solche zu musealen sowie archäologischen Grabungsobjekten. Die technischen Mittel, die eine Digitalisierung dieser Inhalte ermöglichen, sind hierbei ebenso unterschiedlichen Regularien unterworfen wie die Rechte, die an diese gebunden sind.

Projekteigene Digitalisate

Für den Umgang mit der Digitalisierung von Objekten wird hier die 2013 überarbeitete Fassung der DFG-Praxisregeln zur Digitalisierung (cf. DFG 2013), welche u.a. Empfehlungen zur Digitalisierung bildbezogener musealer Objekte beinhalten, als maßgebend betrachtet. Die DFG, deren Vertreter sich u.a. im Fachbeirat zum TextGrid-Projekt befinden (cf. TextGrid 2012), versucht jeweils den Stand der Technik in Bezug auf Digitalisierung in den Wissenschaften im deutschsprachigen Raum zu definieren und in den kommenden Jahren voraussichtlich regelmäßig Ergänzungen der Praxisregeln zu verfassen. Die DFG vertritt hierbei als mitunterzeichnende Partei der Berliner Erklärung eine der modernen digitalen wissenschaftlichen Arbeit entsprechende "Open Access"- und "Open Source"-Politik.

Gemäß den DFG-Praxisregeln ist dem Projekt wichtig, die Digitalisierung wissenschaftsrelevanter Materialien der Forschung und der interessierten Öffentlichkeit weltweit nach der "Open Access"-Regelung zugänglich zu machen. "Hierbei muss die Rechtklärung bezüglich der zu digitalisierenden Materialien erfolgt sein. Insbesondere eventuell vorliegende Urheber-, Persönlichkeits- und Leistungsschutzrechte sind dabei zu berücksichtigen" (DFG 2013: 6).

Die vom Projekt digitalisierten Objekte werden, falls sie einer lizenzrechtlichen Beschränkung unterliegen, den Embargofristen entsprechend innerhalb des nicht öffentlich zugänglichen Teiles der Datenbank vorgehalten, und nur geprüfte Inhalte, die sich unter einer freien Lizenz befinden, können ohne weitergehende Anfragen und Registrierung öffentlich zugänglich gemacht werden.

Projekteigene Publikationen

Alle im Rahmen des Projektes erstellten, eigenen digitalisierten Bild- und Textmaterialien sowie wissenschaftliche Veröffentlichungen werden unter der international geltenden Creative Commons Urheberrechtslizenz CC BY 4.0 "Open Access" online veröffentlicht. Diese Lizenzstufe ist durch ihre Offenheit eine sogenannte "Free-Culture"-Lizenz, wobei sie sich nach den Forderungen der Berliner Erklärung und der Budapester Open Access Initiative richtet (cf. Creative Commons n.d.a). CC BY 4.0 verweist darauf, dass die Inhalte ohne weiteres gelesen, heruntergeladen, erweitert, gekürzt, verändert und weitergegeben werden können. Sie dürfen sowohl für kommerzielle als auch für nicht-kommerzielle Zwecke verwendet werden, ohne finanzielle, gesetzliche oder technische Hindernisse. Es muss jedoch dafür Sorge getragen werden, dass der Urheber oder die Urheberin genannt wird und ein Link auf das Originalwerk verweist (cf. Creative Commons n.d.b.). Creative Commons Lizenzen bauen hierbei auf dem geltenden Urheberrecht auf. Urheberrechtlich geschützte Inhalte wie Texte, Bilder etc. können durch eine CC-Lizenz gemeinfrei nachgenutzt werden. Soweit es rechtlich und finanziell möglich ist, wird für Veröffentlichungen als Lizenzmodell der Goldene Weg der "Open Access"-Publikation gewählt.

Das bedeutet, Erstveröffentlichungen werden gleichzeitig frei verfügbar unabhängig von ihrer anderweitigen Publikation über die projekteigenen Veröffentlichungsplattformen durch das Internet zugänglich sein. Die Publikationen der Texte durchlaufen hierbei einen Qualitätssicherungsprozess in Form eines Peer Reviews oder Editorial Reviews.

In Bezug auf bereits an anderer Stelle veröffentlichte Forschungstexte, die im Rahmen des Projektes von Bedeutung sind, wird, falls die Bedingungen dazu erfüllt sind, vom unabdingbaren Zweitveröffentlichungsrecht Gebrauch gemacht, "für das sich die DFG gemeinsam mit ihren Partnern aus der Allianz der Wissenschaftsorganisationen über Jahre einsetzte, und das mit der Novellierung des deutschen Urheberrechtsgesetzes zum 1. Januar 2014 in Kraft trat" (Fournier 2015: 6).

Dadurch können Inhalte als Autorenmanuskripte 12 Monate nach Erstveröffentlichung online frei zur Verfügung gestellt werden (cf. Fournier 2015: 6). Des Weiteren bietet sich dem Projekt für gegebene, schwieriger zu regelnde Fälle die Nutzung weiterer verschiedener CC-Lizenzabstufungen an, welche

die Kontrolle erhöhen über z.B. die Integrität der Inhalte, die Zitierpflicht oder die nichtkommerzielle Nutzung. D.h. diese Lizenzen können im CC-Lizenzrahmen enger definiert werden als die CC BY-Lizenzen es sind, so dass gleichwohl die bestmögliche freie Verbreitung der Inhalte unter verschiedenen Bedingungen gewährleistet werden kann. Um die Verbreitung der Inhalte im Allgemeinen zu erhöhen und technisch zu ermöglichen, verfügt das CC-Lizenzmodell über eine Konstruktion von drei Lizenzschichten. Neben einer juristisch eindeutig definierten und nutzbaren Definition der verschiedenen Lizenzen gibt es jeweils eine sogenannte menschenlesbare Lizenzschicht, welche die Rechte allgemein leicht verständlich für alle Interessierten beschreibt. Des Weiteren ist die dritte Schicht technisch derart gestaltet, dass sie maschinenlesbar ist, also von Anwendungen im Web gefunden und verarbeitet werden kann (cf. Creative Commons n.d.b.).

Außerhalb des Projektes erstellte Digitalisate und Publikationen

Das Projekt versucht alle digitalisierten Bilder und Umzeichnungen von Maya-Monumenten aus verschiedenen Publikationen, deren Rechte zur weltweiten Veröffentlichung es erhalten oder erworben hat, auf dem TextGrid Repository mitsamt annotierten Metadaten der Öffentlichkeit unter Verweis auf Originalquellen und Urheberschaft zugänglich zu machen. Bei Digitalisaten und Publikationen, deren Rechte eingeschränkt sind, werden diese innerhalb der VFU, im TextGrid Laboratory nur einem ausgewählten Kreis an Fachwissenschaftlern zu Forschungszwecken zugänglich gemacht.

“Open Methodology”-Strategie

Mit der offenen Dokumentation des projekteigenen digitalen epigraphischen Workflows wird die Methodik und der Arbeitsprozess, der zur Entwicklung der Datenstruktur führt, transparent. Interessierte können diese zur Anregung bei der Entwicklung eigener Lösungen thematisch ähnlicher Sachverhalte nutzen.

TextGrid stellt hierzu die geeignete Plattform bereit, die den kompletten wissenschaftlichen Arbeitsablauf in einer einzigen Umgebung ermöglicht: Vom Sammeln und Generieren der Primärdaten bis hin zum Archivieren und Publizieren (cf. TextGrid n.d.). Die VFU TextGrid stellt dem Projekt alle benötigten digitalen Werkzeuge hierzu in einem Framework zur Verfügung. Somit ist methodisch die Erstellung mit der Aufbewahrung der Objekte sowie mit der Wissensvermittlung der Inhalte verbunden. Das bedeutet, die Digitalisierung etwa von Abbildungen musealer Objekte wird hier nicht von der späteren Nutzung getrennt betrachtet, was somit schon bei der Modellierung der Daten zu beachten ist.

Das TextGrid Lab ist modular aufgebaut und kann durch verschiedene externe “Open Source“-Anwendungen erweitert werden, so dass alle Arbeitsabläufe wie gewünscht in einer einzigen Umgebung durchführbar sind (cf. TextGrid n.d.). Hierbei fließen die verschiedenen Datenerschließungsschritte innerhalb desselben Workflows ineinander. In Bezug auf die softwaretechnische Umsetzung bedeutet dies: Das Eingeben der Digitalisate, das Erschließen der Objekt-Metadaten, die Text-Bildverknüpfung, die Auszeichnung mit epigraphischen bzw. linguistischen Metadaten und das Erstellen weiterer Metadaten wird innerhalb derselben virtuellen Arbeitsumgebung bewerkstelligt. Darüber hinaus werden die Präsentation der Inhalte und die Erstellung eines Zugangs zu diesen ebenso innerhalb der VFU erledigt. Die Dokumentation der Arbeitsprozesse wird anhand von Usecases und ähnlichen schriftlichen Proceedings auf einer der projekteigenen Plattformen in digitaler Form online angeboten.

Um die Daten mit den eindeutig definierten Inhalten an Metainformationen zu versehen, wird ein auf XML-basierten Sprachen beruhendes Metadatenschema für die Objekte und eines für die linguistische

Analyse der Daten konzipiert und erstellt. Die Dokumentation der Konzeption sowie der versionskontrollierte Sourcecode des fertiggestellten Objekt-Metadaten- und des Analyse-Metadatenschemas werden angeboten, indem sie webbasiert – etwa durch frei zugängliche Hosting-Dienste für Software-Entwicklungsprojekte wie GitHub – frei zur Verfügung gestellt werden.

“Open Source”-Strategie

Die technischen Mittel zur Nutzung der Projekt-Daten und -Metadaten in Form von wissenschaftlicher Software setzen sich zusammen aus frei verfügbaren “Open Source”-Anwendungen. Wie zuvor beschrieben, stammen die im Projekt genutzten softwaretechnischen Werkzeuge zur Datenerstellung aus dem Framework der “Open Source”-Software TextGrid (cf. Gietz et al. 2011), welche dadurch keiner Nutzungsbeschränkung unterliegen. Der Source Code und die Versionierung von TextGrid sind hierbei öffentlich zugänglich; Tools und Services sind dokumentiert und als solche veränderbar und erweiterbar (cf. TextGrid n.d.). Verschiedene “Open Source”-Lizenzen – wie etwa GNU – wurden für die einzelnen Software-Derivate unter dem Gesamtpaket der Virtuellen Forschungsumgebung TextGrid festgelegt (cf. Gietz et al. 2011). Bei Modellierung des Metadatenschemas und Speicherung der Metadaten nach dessen Vorgaben werden hierbei ausschließlich quelloffene Daten- und Metadatenstandards sowie Programmiersprachen, deren Quellcodes und Dokumentationen offenliegen, genutzt.

Die TextGrid-Infrastruktur ermöglicht dank ihrer serviceorientierten und modularen Architektur und gegebener Schnittstellen, welche auf “Open Source”-Sprachen und offenen Standards beruhen, die Interoperabilität zwischen TextGrid und externen Anwendungen, Datenbanken sowie deren Inhalten.

“Open Data”-Strategie

Um der eigens erstellten Datensammlung Vernetzung und Austausch mit anderen Daten zu ermöglichen, müssen gegebene Daten verschiedener Provenienz sowohl offen verfügbar als auch miteinander vergleichbar sein. Nur durch die Vergleichbarkeit mit anderen Daten können diese anwendungsübergreifend nachhaltig genutzt und verlinkt werden und in eine offene Datenstruktur wie das sogenannte “Linked Open Data” (kurz: “LOD”) eingebunden werden oder etwa im größeren Zusammenhang eines Semantic Webs nutzbar gemacht werden.

Interoperabilität als Voraussetzung für Linked Open Data

Vergleichbarkeit bedeutet im technischen Sinne Interoperabilität. Hierbei muss die Interoperabilität für datenbeschreibende Daten, d.h. für Metadaten, technisch auf verschiedenen Ebenen umgesetzt werden: strukturell, syntaktisch und semantisch.

- Die strukturelle Interoperabilität bezieht sich darauf, dass den verschiedenen Metadatenformaten ein gemeinsames Modell zugrunde liegt - wie z. B. das RDF-Modell oder die Schnittstellendefinition des OAI-PMH-Protokolls.
- Die syntaktische Interoperabilität wird gewährleistet durch eine gemeinsame Syntax wie die des Auszeichnungsstandards XML, in der die Daten abgebildet werden und vorliegen.
- Die semantische Interoperabilität hingegen erfordert eine über die formgebende Syntax hinausgehende Nutzung derselben Metadatendefinitionen. Diese werden festgelegt in anerkannten Metadatenstandards wie der auf XML basierende Standard TEI oder mittels kontrollierter Vokabularien zur Wissensorganisation wie SKOS (Simple Knowledge Organization System).

Die Interoperabilität soll somit technisch und inhaltlich unter einem gemeinsamen Nenner definieren, wie die Daten ihre Semantik bewahrend und ohne maßgeblichen Informationsverlust austauschbar gemacht werden können (Rühle 2012: 2 ff.).

Interoperabilität umgesetzt im Projekt

Um eine Austauschbarkeit der Projekthinhalte zu ermöglichen werden die besagten drei Teilaspekte der Interoperabilität durch die Nutzung der TextGrid-Infrastruktur und durch die entsprechende Modellierung der Metadaten umgesetzt. Die zuvor erwähnte offene TextGrid-Infrastruktur erlaubt einen hohen Grad an struktureller Interoperabilität. Das heißt zum Beispiel, anderen Anwendungen wird die Entnahme und Nachnutzung der dort gespeicherten Daten ermöglicht. Dies geschieht u.a. mittels Interfaces, welche durch offene Standards wie REST (Representational State Transfer) oder OAI-PMH einen Austausch zwischen verschiedenen Anwendungen oder Datenbanken herstellen.

Darüber hinaus werden die Projekt-Daten durch die TextGrid-Datenarchitektur technisch nachhaltig auffindbar gemacht, indem sie jeweils eine Art dauerhafte digitale Adresse, technisch formuliert, indem sie eine PID in Form einer URI (Uniform Resource Identifier) erhalten. Die Auffindbarkeit explizit freier und offener Inhalte ist ebenso notwendig für deren Interoperabilität wie ihre Verarbeitung und Verbreitung. Den Projekthinhalten werden dazu mittels der zuvor erwähnten dritten CC-Lizenzschicht die entsprechenden Nutzungsrechte in maschinenlesbarer Form beigelegt, welche in der standardisierten CC Rights Expression Language (kurz: CC REL) ausgedrückt sind. Somit können die Inhalte mit den an sie gebundenen Nutzungsbedingungen von Suchmaschinen und Softwares gefunden, erkannt und verarbeitet werden.

Die entsprechende Modellierung der Projekt-Metadaten wird in Form von maschinenlesbaren RDF-Triples durchgeführt. Das Format RDF (Resource Description Framework) ist als ein für die semantische Web-Suche bewährtes dreigliedriges Datenmodell beschrieben. Es ist ein allgemein anerkannter, vom Web-Konsortium W3C unterstützter Standard um Informationen webtauglich zu kodieren und zu präsentieren, und um eine strukturelle Interoperabilität mit anderen Inhalten zu erschaffen. Umgesetzt werden kann das Modell auf syntaktischer Ebene mittels verschiedener formaler Sprachen wie etwa der von TextGrid genutzte und unterstützte XML-basierte Metadatenstandard. Die syntaktische sowie semantische Interoperabilität der Daten und Metadaten wird demgemäß durch deren technische und inhaltliche Modellierung erzeugt – etwa beim Aufbau des Metadatenschemas und der Definition der sich darin befindenden Metadaten. Die in einer ontologieartigen Struktur definierten Inhalte sollen hierbei nicht nur auf etablierten Metadatenstandards beruhen. Sie sollen des Weiteren semantisch etwa mittels kontrollierter Vokabularien übergreifend verständlich definiert werden, um bei einer Verbindung mit Daten anderer Herkunft auf inhaltlich sinnvolle Weise vergleichbar und verbindbar zu sein. Eine semantische Interoperabilität ermöglicht hierbei ein Verweisen auf dieselben Konzepte und "Dinge", was einer der wichtigen Punkte der "Linked Open Data"-Modellierung ist (cf. Beer et al. 2014: 7).

Das vom Projekt u.a. für die Auszeichnung der Analyse der epigraphischen Inhalte genutzte, auf der Auszeichnungssprache XML basierende Metadatenformat TEI (Text Encoding Initiative) wird von TextGrid unterstützt. Hierbei hat sich TEI als internationaler Standard für die auf Texte bezogene digitale Erschließung geisteswissenschaftlicher Inhalte etabliert (cf. TextGrid n.d.). Neben TEI sind weitere für die digitale Epigraphie anerkannte Metadatenstandards wie EpiDoc ebenso hilfreich bei der Konzeption und Modellierung des Analyse-Metadatenschemas, welches hierdurch gewährleistet, dass die einzugebenden Metadaten inhaltlich in konsistenter Form abgespeichert und verständlich und somit sinnvoll verwertbar werden.

Auch die Objekt-Metadaten, welche die Textträger im weiteren Kontext umschreiben, sind in der ontologieartigen Datenbankstruktur miteinander verknüpft. Diese Strukturierung beruht auf

gemeinsamen Standards zu Konzepten aus dem Bereich der Dokumentation kulturellen Erbes. Sie wird nach dem CIDOC Conceptual Reference Model (kurz: CIDOC CRM) konzipiert und mittels übergreifend genutzter, kontrollierter Vokabularien wie SKOS u.ä. inhaltlich beschrieben.

Beispielsweise erarbeitet und definiert das Projekt Getty-Vokabularien, um diese zu erweitern. Die beigesteuerten Vokabularien für TGN (Getty Thesaurus of Geographic Names), AAT (Art & Architecture Thesaurus), CONA (Cultural Objects Name Authority) und ULAN (Union List of Artist Names) garantieren dadurch eine über das Projekt hinaus gesicherte Nutzung kontrollierter Vokabularien für den Bereich etwa der Maya-Kunst und -Architektur und in Bezug auf Maya-Persönlichkeiten und Maya-Fundstätten. Die gemeinsame Verwendung von anerkannten Standards und Vokabularien erhöht die Möglichkeit einer inhaltlich gesicherten und sinnvollen Beschreibung sowie eines Austausches mit anderen thematisch ähnlichen Inhalten.

Schlussendlich wird durch die Nutzung solcher etablierter Modelle und Terminologien für das Projekt eine sinnvolle, allgemeingültige Beschreibung der verschiedenen Metadaten auf semantischer Ebene gewährleistet werden. Neben der XML-basierten Definition der Metadaten, werden ihre Inhalte in auf der Sprache Turtle basierten RDF-Triples definiert und in sogenannten Triplestores gespeichert. Dies gewährleistet den Ausbau einer weiteren syntaktischen Ebene an Interoperabilität. Hierbei baut sich die Datenbank technisch gesehen in der Form eines gerichteten Graphen auf, dessen Kanten die Knoten und Blätter verknüpfen, die jeweils aus einem RDF-Triple, also aus Metadaten zu je einer Ressource bestehen. Um die Ressourcen miteinander zu verknüpfen, werden die ihnen zugewiesenen persistenten und eindeutigen digitalen Identifizierungen in Form von PIDs genutzt. Die im Projekt neben der XML-Auszeichnung gewählte Umsetzung der Triples in Turtle-Form ist hierbei eine vom Web-Konsortium W3C vorgeschlagene Empfehlung.

Dadurch werden die Inhalte des gerichteten RDF-Graphen in abgekürzter Textform kodiert. Turtle ist kompatibel mit der ebenfalls von W3C empfohlenen Abfragesprache SPARQL. Die in SPARQL formulierten Abfragen ermöglichen die semantische Suche innerhalb des Datennetzes des Projektes. Jeder Knoten und jedes Blatt des Graphen, der aus projekteigenen Daten aufgebaut ist, kann mittels einer SPARQL-Abfrage angesteuert und ausgelesen werden. Die in Graphenform verbundenen, über das Web offen zugänglichen und von dort aus per SPARQL-Abfragen durchsuchbaren Daten bilden dabei schließlich die als "Open Data" oder im größeren Datenkontext verstandenen Sinne als "Linked Open Data" bezeichneten Inhalte.

"LOD" ist die mittlerweile für die Offene Wissenschaft etablierte Methode um Daten frei recherchierbar, verfügbar und miteinander verbindbar zu machen. Durch eine "Linked Open Data"-Struktur wird die "Open Access"-Philosophie gewissermaßen ausgeweitet von der Veröffentlichung der Forschungsergebnisse auf die Veröffentlichung der diesen zugrundeliegenden Daten. Dadurch wird allen Personen oder Gruppen, die Interesse an den Daten haben, die Möglichkeit gegeben, z.B. eigene Hypothesen an diesen zu testen, oder sie in jeglicher anderen Art zu verwenden, als es das vorgestellte Projekt tut (cf. Kraker et al. 2011: 646). Beispielsweise können vielerlei Datenvisualisierungen oder etwa verschiedenartige inferenzbasierte korpuslinguistische Methoden o.ä. an den Inhalten vorgenommen werden. Durch die "Open Data"-Strategie des Projektes wird also den Daten über das Projekt hinaus ein potenzieller Mehrwert gegeben, der jederzeit und allseits ausgeschöpft werden kann.

Schlussbemerkung

Das TWKM-Projekt befindet sich zurzeit in der Startphase, was bedeutet, dass momentan begonnen wird, die in diesem Working Paper genannten Methoden und Werkzeuge zu entwickeln, die den

Arbeitsprozess innerhalb des Projektes sowie die Veröffentlichung der Forschungsergebnisse gestalten. Diese sollen dazu beitragen, die Nutzung einer Virtuellen Forschungsumgebung sowie die Erstellung einer offenen Datenbank und -Infrastruktur zur Erforschung der Sprache des Klassischen Maya zu ermöglichen. Im Laufe des Projektes werden sich einige Änderungen in der Roadmap ergeben. Doch das Ziel eines in jeder Hinsicht offenen und transparenten Projektes, das seinen Teil zu einer "Open Science" beiträgt, wird sich wie ein roter Faden durch die Projektarbeit ziehen. Somit kann mit Zuversicht daran gearbeitet werden, zum Ende der Projektlaufzeit die gewünschten Ziele zu erreichen.

Quellen und Literatur

* Anmerkung: Alle referenzierten Internetquellen wurden zuletzt abgerufen: 08.04.2015. Alle Verlinkungen sind in der Onlineversion des Papers verfügbar.

Literatur

Berliner Erklärung

2003 *Berliner Erklärung über den offenen Zugang zu wissenschaftlichem Wissen.*

Beer, Nikolaos, Kristin Herold, Wibke Kolbmann, Thomas Kollatz, Matteo Romanello, Sebastian Rose, Niels-Oliver Walkowski

2014 Interdisciplinary Interoperability. *DARIAH-DE Working Papers*, 3.

Budapest Open Access Initiative (BOAI)

2002 *Read the Budapest Open Access Initiative.*

Bundesministerium für Bildung und Forschung (BMBF)

2013 *Bekanntmachung des Bundesministeriums für Bildung und Forschung von Richtlinien zur Förderung von Forschungs- und Entwicklungsvorhaben aus dem Bereich der eHumanities.*

Creative Commons

n.d.a *Understanding Free Cultural Works.*

n.d.b *Mehr über die Lizenzen.*

DARIAH.de Digital Research Infrastructure for the Arts and Humanities

n.d. *TextGrid.*

Deutsche Forschungsgemeinschaft (DFG)

2013 *Praxisregeln "Digitalisierung" (DFG-Vordruck).*

2014 Appell zur Nutzung offener Lizenzen in der Wissenschaft. *Information für die Wissenschaft*, 68

Deutsche Forschungsgemeinschaft (DFG) – DFG-Viewer

2015 *Profil der Metadaten.*

Fournier, Johannes

2015 Open Access und Open Data. Positionen und Perspektiven der Deutschen Forschungsgemeinschaft (DFG). *Archäologische Informationen*, 38.

Gietz, Peter, Markus Widmer, Stefan Funk, Andreas Witt, Oliver Schonefeld, Norman Fiedler

2011 Musterverträge (AP 3.2) und technische Umsetzung. *TextGrid Report* 3.2.1.

Hilf, Eberhard R. & Thomas Severiens

2013 *Vom Open Access für Dokumente und Daten zu Open Content in der Wissenschaft.*

Internet Portal Open Access

n.d. Der freie Zugang zu wissenschaftlicher Information – Open Access Strategien

Kraker, Peter, Derick Leony, Wolfgang Reinhardt & Günter Beham

2011 The Case for an Open Science in Technology Enhanced Learning. *International Journal of Technology Enhanced Learning*, 6 (3): 643–654.

Matthias, Angela & Brigitte Wiechmann

2014 *Sammelrichtlinien.* Deutsche Nationalbibliothek (DNB).

Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB)

n.d. *IDIOM.*

OpenScienceASAP

n.d. *Was ist Open Science?*

Organisation for Economic Co-operation and Development (OECD)

2004 *OECD – Declaration on Access to Research Data from Public Funding.*

2007 *OECD Principles and Guidelines for Access to Research Data from Public Funding.*

Rheinische Friedrich-Wilhelms-Universität Bonn, Abteilung für Altamerikanistik (Uni Bonn)

n.d. *Textdatenbank und Wörterbuch des Klassischen Maya.*

Rühle, Stefanie

2012 *Richtlinie für die interoperable Gestaltung von Metadatenprofilen.* Kompetenzzentrum Interoperable Metadaten (KIM), SUB Göttingen.

Schütz, Christian

2015 *Nationales ISSN-Zentrum für Deutschland – häufig gestellte Fragen (FAQ).* Deutsche Nationalbibliothek (DNB).

TextGrid Konsortium

2012 *Abschlussbericht: Vernetzte Forschungsumgebung in den eHumanities.*

2014 *Das Projekt. TextGrid – Virtuelle Forschungsumgebung für die Geisteswissenschaften.*

Thaller, Manfred

2012 *Controversies around the Digital Humanities: An Agenda.* *Historical Social Research*, 37 (3): 7-23.

Internetressourcen und -quellen**Comité international pour la documentation Conceptual Reference Model (CIDOC CRM)**

2014 *What is the Cidoc CRM.* Cidoc CRM Home Page.

Creative Commons

n.d. *Attribution 4.0 International (CC BY 4.0).*

Creative Commons Wiki

2013 *CC Rel.*

Epidoc: Epigraphic Documents in TEI XML (EpiDoc)

n.d. *EpiDoc: Guidelines.*

Getty Research Institute

n.d. *Getty Vocabularies.*

GitHub, Incorporated

n.d. *GitHub Website.*

GNU General Public Licence

2014 *GNU General Public Licence.*

Hakala, Juha

2010 *Persistent identifiers – an overview.* *Technology Watch Report (TWR): Standards in Metadata and Interoperability.*

International Standard Serial Number – ISSN International Centre

n.d.a *International Identifier for Serials.*

n.d.b *The ISSN for Electronic Media.*

Library of Congress

2015a *Metadata Encoding and Transmission Standard – METS.*

2015b *Metadata Object Description Schema – MODS.*

Open Archives

n.d. *Open Archives Initiative Protocol for Metadata Harvesting.*

TEI: Text Encoding Initiative

n.d. *TEI Website.*

TextGrid Konsortium

n.d. *TextGrid Website.*

Universitäts- und Landesbibliothek Bonn (ULB)

n.d. *Digitale Sammlungen.*

World Wide Web Consortium (W3C)

2006 *W3C. Naming and Addressing: URIs, URLs, ...*

-
- 2007 *W3C. SPARQL Query Language for RDF.*
2011a *W3C. REST.*
2011b *W3C. TURTLE – Terse RDF-Triple Language.*
2012 *W3C. SKOS Simple Knowledge Organization System – Home Page.*
2014 *W3C. Resource Description Framework (RDF).*
2015a *W3C. Website.*
2015b *W3C. Linked Data.*



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/>