

GOEDOC – Dokumenten- und Publikationsserver der Georg-August-Universität Göttingen

2015

Die Erstellung eines TEI-Metadatenschemas für die Auszeichnung von Texten des Klassischen Maya

Petra Maier (ULB Düsseldorf)

DARIAH-DE Working Papers

Nr. 8

Maier, Petra: Die Erstellung eines TEI-Metadatenschemas für die Auszeichnung von Texten des
Klassischen Maya

Göttingen : GOEDOC, Dokumenten- und Publikationsserver der Georg-August-Universität, 2014
(DARIAH-DE working papers 8)

Verfügbar:

PURL: <http://resolver.sub.uni-goettingen.de/purl/?dariah-2015-1>

URN: <http://nbn-resolving.de/urn:nbn:de:gbv:7-dariah-2015-1-6>

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)



Bibliographische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

Erschienen in der Reihe

DARIAH-DE working papers

ISSN: 2198-4670

Herausgeber der Reihe

DARIAH-DE, Niedersächsische Staats- und Universitätsbibliothek

Mirjam Blümm, Thomas Kollatz, Stefan Schmunk und Christof Schöch

Abstract: Im Rahmen des 2014 angelaufenen Projektes ‚Interdisciplinary Dictionary of Classic Mayan‘ (TWKM) der Universität Bonn, das die Erforschung und vollständige Entzifferung der Schrift und Sprache des Klassischen Maya zum Ziel hat, werden alle überlieferten Schriftenträger des Klassischen Maya erschlossen und systematisch dokumentiert. Hierfür wurden in einem Teil-Projekt die Inschriftentexte mittels des Datenformats der Text Encoding Initiative (TEI) erfasst: Wesentliche Informationen, die durch die Metadaten ausgezeichnet werden mussten, waren die unterschiedlichen Formen und Ausgestaltungen der Inschriftentexte, sowie die Zeichenschrift selbst. Das TEI-Metadatenchema dieses Teilprojekts stellte eine Arbeitsgrundlage dar, die im weiteren Projektverlauf angepasst und ergänzt werden konnte.

Keywords: Digital Humanities, Epigrafik, Klassisches Maya, Textkodierung, TEI, Metadaten
Digital Humanities, epigraphy, classic Maya, text encoding, TEI, metadata

Die Erstellung eines TEI-Metadatenschemas für die Auszeichnung von Texten des Klassischen Maya

Petra Maier (ULB Düsseldorf)



Petra Maier: „Die Erstellung eines TEI-Metadatenschemas für die Auszeichnung von Texten des Klassischen Maya“. *DARIAH-DE Working Papers* Nr. 8. Göttingen: DARIAH-DE, 2015.
URN: urn:nbn:de:gbv:7-dariah-2015-1-6.

Dieser Beitrag erscheint unter der
Lizenz [Creative-Commons Attribution 4.0](https://creativecommons.org/licenses/by/4.0/) (CC-BY).

Die *DARIAH-DE Working Papers* werden von Mirjam Blümm,
Thomas Kollatz, Stefan Schmunk und Christof Schöch
herausgegeben.



Abstract

Im Rahmen des 2014 angelaufenen Projektes ‚Interdisciplinary Dictionary of Classic Mayan‘ (TWKM) der Universität Bonn, das die Erforschung und vollständige Entzifferung der Schrift und Sprache des Klassischen Maya zum Ziel hat, werden alle überlieferten Schriftenträger des Klassischen Maya erschlossen und systematisch dokumentiert. Hierfür wurden in einem Teilprojekt die Inschriftentexte mittels des Datenformats der Text Encoding Initiative (TEI) erfasst. Wesentliche Informationen, die durch die Metadaten ausgezeichnet werden mussten, waren die unterschiedlichen Formen und Ausgestaltungen der Inschriftentexte, sowie die Zeichenschrift selbst. Das TEI-Metadatenchema dieses Teilprojekts stellte eine Arbeitsgrundlage dar, die im weiteren Projektverlauf angepasst und ergänzt werden konnte.

Schlagwörter

Digital Humanities, Epigrafik, Klassisches Maya, Textkodierung, TEI, Metadaten

Keywords

Digital Humanities, epigraphy, classic Maya, text encoding, TEI, metadata

Vorbemerkung

Der vorliegende Bericht beruht auf einer Projektarbeit, die im Rahmen des berufsbegleitenden Weiterbildungsstudiengangs der Bibliotheks- und Informationswissenschaft (Master in Library and Information Science - MALIS) an der Fachhochschule Köln erfolgte.

Inhaltsverzeichnis

1. Ausgangslage.....	1
2. Kurze Darstellung des Klassischen Maya.....	1
3. Forschungsstand.....	3
4. Das Format TEI.....	4
5. Projektdefinition und -planung.....	5
5.1 Zielsetzung.....	5
5.2 Vorgehensweise allgemein.....	5
5.3 Vorgehensweise bei der Erschließung der Inschriftentexte.....	6
5.4 Anforderungen seitens der Wissenschaft.....	7
6. Metadatenschema für die Erfassung der Inschriftentexte.....	7
6.1 Wiedergabe der Struktur des Inschriftentextes.....	9
6.2 TEI-Elemente.....	11
6.3 Fehlende und nicht lesbare Textstellen, Hieroglyphen und Zeichen.....	19
6.4 Kritische Betrachtung des Metadatenschemas.....	19
7. Fazit.....	20
Abbildungsverzeichnis.....	22
Tabellenverzeichnis.....	22
Literatur- und Quellenverzeichnis.....	22

1. Ausgangslage

Anfang 2014 startete das durch die Nordrhein-Westfälische Akademie der Wissenschaften und Künste geförderte Projekt ‚Textdatenbank und Wörterbuch des Klassischen Maya‘ (TWKM) unter der Leitung von Prof. Dr. Nikolai Grube (Abteilung für Altamerikanistik an der Philosophischen Fakultät der Universität Bonn). Das Projekt, das in Kooperation mit dem Forschungsverbund TextGrid (unter der Leitung der Niedersächsischen Staats- und Universitätsbibliothek Göttingen) und der Universitäts- und Landesbibliothek Bonn durchgeführt wird, hat eine geplante Gesamtlaufzeit von 15 Jahren; der Gesamtprojektplan gliedert sich in fünf Abschnitte zu jeweils drei Jahren. Gesamtziel des Projektes ist die Erschließung sämtlicher bekannter Maya-Hieroglyphentexte in einem digitalen Textkorpus, das die Grundlage für epigrafische und linguistische Analysen bildet. Mit Ablauf des TWKM-Projektes soll ein Wörterbuch – in digitaler und gedruckter Form – erstellt werden, das den gesamten bekannten Wortschatz enthält, und dessen Verwendung in der Schrift widerspiegelt (vgl. Grube 2011, 13).

Als Teilziel des ersten TWKM-Projektabschnitts ist die Erstellung einer Arbeitsversion des Wörterbuches in elektronischer Form vorgesehen. Um dies zu ermöglichen, war u. a. die Konzipierung eines Datenmodells in einer elektronischen Forschungsumgebung notwendig. Solch ein umfassendes Forschungsvorhaben, das die Erfassung aller bekannter Inschriftenobjekte und deren Inschriften beinhaltet, sowie die spätere weitere Erforschung noch unbekannter bzw. mehrdeutiger Schriftzeichen ermöglichen soll, setzt ein komplexes Metadatenkonzept voraus. Bereits mit dem Projektantrag war vorgesehen, dass die Erschließung der Inschriftentexte mittels des Standards TEI des gleichnamigen Konsortiums TEI (Text Encoding Initiative) erfolgen soll (vgl. Grube 2011, 13).

2. Kurze Darstellung des Klassischen Maya

Um das Verständnis der Projektdokumentation zu erleichtern und in das Thema einzuführen, werden im Folgenden das Klassische Maya und dessen zeitlich-geografischer Kontext kurz dargestellt.

Geografisch erstreckt sich das Gebiet der Maya über die heutigen Gebiete der mexikanischen Bundesstaaten Chiapas, Tabasco, Campeche, Quintana Roo und Yukatan, den Staaten Belize, Guatemala, sowie westliche Abschnitte von Honduras und El Salvador (s. Abb. 1) (vgl. Grube; Gaida 2006, 23).

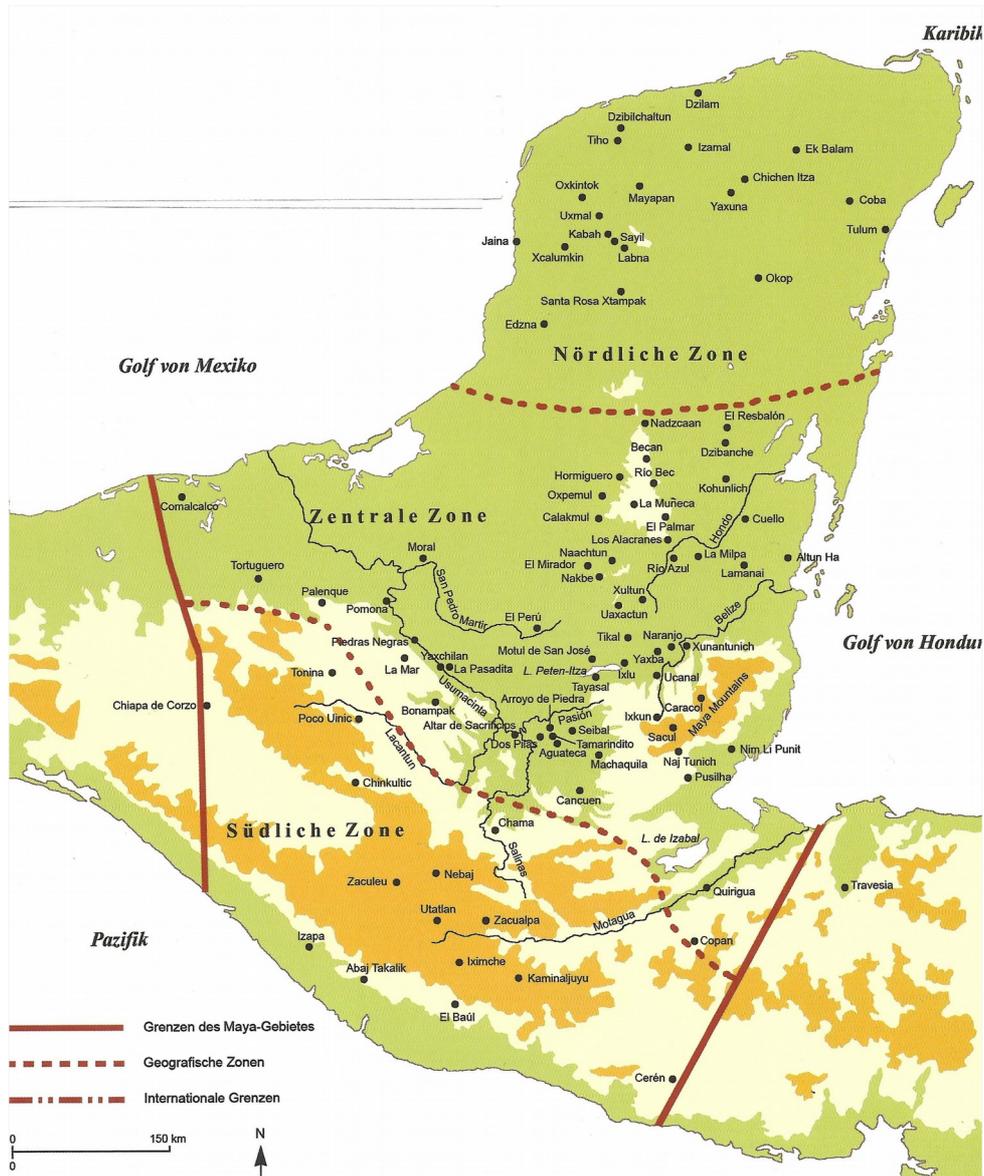


Abb. 1: Geografische Lage des Maya-Gebietes (Grube; Gaida 2006, 23)

Die Schriftlichkeit diente den Maya der vorspanischen Zeit zur Repräsentation der Herrscher und ihrer Familien: Oftmals werden Ereignisse wie die Geburt, Inthronisierung usw. in den Inschriften beschrieben, wobei diese Anlässe i. d. R. mit Kalenderdaten versehen sind, sodass sich die Inschriften bzw. die dort beschriebenen Ereignisse auf den Tag genau datieren lassen. Durch eine in der Maya-Forschung anerkannte Korrelation können diese Datumsangaben in den gregorianischen Kalender übertragen werden (vgl. Grube; Gaida 2006, 22-24).

Die Maya-Schrift ist eine Hieroglyphenschrift, deren frühesten erhaltenen Schriftzeugnisse aus dem 3. Jh. v. Chr. stammen. Mit Beginn der Klassik (250 bis 900 n. Chr.) „breitet sich die Schriftkultur im gesamten Maya-Gebiet aus.“ (vgl. Grube 1993, 222-225) Im Laufe ihrer Geschichte „hat sich die Maya-Schrift immer wieder verändert und sich den wechselnden Bedürfnissen ihrer Verfasser und Auftraggeber angepasst. Immer wieder wurden neue Zeichen erfunden, alte nicht weiter verwendet, manche veränderten ihre Lesung“ (Grube 1993, 225 f).

Mit Eroberung des Maya-Gebietes durch die Spanier zu Beginn des 16. Jahrhunderts verschwand die Hieroglyphenschrift aus dem Gebrauch, und damit auch die Kenntnis dieser Schrift (vgl. Grube 1993, 215 f).

Die Schrift der Maya ist eine sogenannte logosyllabische Schrift, d. h. sie besteht aus zwei Arten von Schriftzeichen, den Logogrammen (Wortzeichen) und den Syllabogrammen (Silbenzeichen) (Vgl. Gronemeyer 1999, Kap. 2,1). Ein Hieroglyphenblock entspricht zumeist einem Wort und setzt sich durchschnittlich aus drei bis vier Zeichen zusammen, i. d. R. in einer Kombination aus Logogrammen und Silbenzeichen. Insgesamt werden aktuell in der Mayaforschung 650 Zeichen unterschieden. Für Silben, die oft verwendet wurden, gibt es unterschiedliche Zeichen, sodass der ‚Schreiber‘ Wiederholungen vermeiden konnte. Überwiegend sind die Inschriftentexte lesbar bzw. interpretierbar, aber noch sind nicht alle Zeichen entziffert. Manche Zeichenfolgen können zwar gelesen werden, aber deren Sinn ist (noch) nicht erschlossen (vgl. Grube 2011, 6 und 11). Die Sprache des Klassischen Maya weist eine Verwandtschaft zu den heute gesprochenen Chol-Sprachen auf, die v. a. in den mexikanischen Bundesstaaten des ehemaligen Maya-Gebietes verbreitet sind, und den yukatekischen Sprachen, d. h. des Sprachgebiets der Halbinsel Yukatan (vgl. Grube 1993, 222). Dadurch tragen Korrelationen zwischen dem Klassischen Maya und den heutigen Maya-Sprachen zur Entzifferung bei.

Es gibt unterschiedliche Objekttypen, auf denen Inschriften bzw. Bilder erhalten sind; dabei handelt es sich aufgrund des feucht-warmen Klimas in der Mayaregion um unvergängliche Materialien wie Stein und Keramik: beispielsweise freistehende Monumente, architektonische Schriftträger (z. B. Türsturz oder auch eine Hieroglyphentreppe), Schmuck, Keramik, Kleinplastiken, aber auch in Höhlen wurden Inschriften gefunden in Form von Malereien oder Felsreliefs (z. B. in den Höhlen von Naj Tunich). Einen seltenen Fall stellen Kodizes dar, von denen nur drei bekannt sind.

3. Forschungsstand

In der Erforschung der Schrift und Sprache des Klassischen Maya fehlten bislang umfassende Dokumentationen. Es gibt einige Wörterverzeichnisse, die sich auf die Untersuchung punktueller Fragestellungen oder ausgewählter Hieroglyphen beschränken. Seit Ende der 1990er Jahre sind lexikografische Zusammenstellungen erschienen, die zwar kommentiert sind (also über eine rein alphabetische Auflistung hinausreichen), aber eine Dokumentation der räumlichen Verbreitung und zeitlichen Veränderungen der Schrift vermissen lassen. Dadurch können mit diesen Wortzusammenstellungen (heutige) Forschungsfragen etwa zu der Entwicklung der Schrift nicht untersucht werden.

In der Maya-Forschung beruhen die Defizite und Lücken auf der bislang unvollständigen Dokumentation und dem Mangel an digitaler Erschließung des Quellenmaterials. In anderen Sprachbereichen gibt es Projekte, die den Wissenschaftlern umfassende digitale Inschriften-Corpora zur Verfügung stellen: z. B. kann in dem digitalen Corpus ‚Thesaurus Linguae Aegyptiae‘ (TLA)¹ in ägyptischen Textmaterialien recherchiert werden, und durch bestimmte Analyseanfragen

¹ *Thesaurus Linguae Aegyptiae*. Arbeitsstelle Altägyptisches Wörterbuch. Berlin-Brandenburgische Akademie der Wissenschaften. <http://aew.bbaw.de/tla/index.html> (04.08.2014).

(z. B. nach Worthäufigkeiten) lassen sich unterschiedliche forschungsrelevante Fragestellungen beantworten. Zudem liegt für alle Texte eine Übersetzung vor. Ein ähnliches Vorhaben stellt das Projekt ‚Pennsylvania Sumerian Dictionary‘ (PSD)² der Universität Pennsylvania dar, in dessen Rahmen ein umfassendes Wörterbuch des Sumerischen erstellt wurde. Eine Besonderheit dieses Projektes ist, dass die für die Erstellung des Corpus und für die Arbeit mit der Sumerischen Sprache entwickelten Tools frei zur Verfügung gestellt werden. Dadurch ist eine Nachnutzung für andere Projekte möglich.

4. Das Format TEI

Gemäß den Vorgaben des Projektantrages erfolgt die Erschließung der Hieroglyphentexte mittels eines TEI-Metadatenschemas. Metadaten sind in diesem Kontext allgemein als strukturierte Informationen zu den Maya-Texten als Ganzes sowie zur Auszeichnung einzelner Besonderheiten in den Texten zu verstehen. Das Metadatenschema für die Texte schließt somit auch lokale Annotationen des Textes selbst mit ein.

Die Text Encoding Initiative (TEI) ist eine internationale Organisation, die 1987 gegründet wurde, um Richtlinien zur Kodierung maschinenlesbarer Texte insbesondere aus den Bereichen der Geistes- und Sozialwissenschaften zu entwickeln.³ Zugleich wird die Abkürzung TEI auch für das Metadaten-set selbst verwendet, so auch in der folgenden Projektdokumentation.⁴

TEI verwendet die Auszeichnungssprache ‚Extensible Markup Language‘ (XML), die sich heute als Standard in den geisteswissenschaftlichen Fachgebieten zur digitalen Beschreibung von Quellmaterialien durchgesetzt hat und damit gezielte Abfragen und Weiterverarbeitungen ermöglicht. Aufgrund des standardisierten Elementesets hat TEI den Vorteil einer langfristigen und eindeutigen Interpretierbarkeit der Datensätze. Des Weiteren wird durch den Einsatz von TEI in einem solchen Projekt wie TWKM die Anerkennung des Formats als Standard vorangetrieben und damit ein Datenaustausch erleichtert (Rouché & Flanders 2007-2014; vgl. Werning 2013, 3).

Das Metadaten-schema TEI in der aktuellen Version P 5 stellt eine definierte Menge an XML-Elementen dar. Das Schema ist unterteilt in unterschiedliche Module, die jeweils bestimmte Elemente und Attribute auszeichnen. Bspw. werden in dem Modul ‚dictionaries‘ die Elemente zur Kodierung digitaler Lexika definiert. Ein Element kann andere Elemente oder auch reinen Text enthalten. Jeder TEI-konformen Text wird durch das Element <teiHeader> eingeleitet. Dieser bildet gewissermaßen das Titelblatt der elektronischen Textdatei und beinhaltet u.a. die Dateibeschriftung (obligat) oder Angaben zur Textrevision (optional). Innerhalb einer TEI-Datei kann der Header wiederholt verwendet werden. Auf den Header folgt der Textkörper, der je nach zu beschreibendem Text sehr unterschiedlich erscheinen kann.

TEI verfolgt zwei Ziele: Zum einen soll den Wissenschaftlern ermöglicht werden, ihre Quellmaterialien mittels einer Beschreibungssprache digital darzustellen, zum anderen diese

² *Pennsylvania Sumerian Dictionary*. University of Pennsylvania. <http://psd.museum.upenn.edu/epsd1/index.html> (04.08.2014).

³ Vgl. „TEI: Frequently Asked Questions“. TEI Consortium. <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/TitlePageVerso.html> (04.08.2014).

⁴ Um die Unterscheidung der beiden Projekte zu erleichtern, wird im Folgenden das Rahmenprojekt als TWKM-Projekt bezeichnet.

digitalen Informationen durch die Nutzung einer gemeinsamen Kodierung allgemein verständlich zu repräsentieren. TEI kann sehr detailliert und spezialisiert auf unterschiedliche Quellmaterialien, d.h. mit einer umfassenden Kodierung, verwendet werden. Ebenso ist es möglich, die Kodierung auf die wesentlichen Informationen zu beschränken ohne Spezialisierung auf besondere Phänomene. Die detaillierte Kodierung hat den Vorteil, dass der beschriebene Text mehr Verwendungsmöglichkeiten bietet, wie z. B. gezielte Abfragen; zu berücksichtigen ist allerdings, dass sich dadurch auch der Eingabeaufwand erhöht und eine größere Sachkenntnis notwendig ist. Der Einsatz von TEI in unterschiedlichen Bereichen wird zudem dadurch unterstützt, dass die Auszeichnungssprache durch projektspezifische Anpassungen für die eigenen Zwecke definiert werden kann. Dies fördert zusätzlich die Nachnutzung und Verbreitung des TEI-Standards, und kann zu gegenseitigen Anregungen der unterschiedlichen Wissenschaftsbereiche führen und diese zugleich voneinander differenzieren (vgl. Rouché & Flanders 2007-2014). Das Metadatenschema für die Texte des Klassischen Maya stellt somit ein für diesen Zweck zusammengestelltes Metadaten-set dar, das die spezifischen Angaben zu beschreiben vermag.

In zahlreichen Projekten, in denen digitale Texte unterschiedlichster Gattungen erschlossen werden, wird das Metadatenschema TEI herangezogen; auf der Homepage der Initiative kann eine Liste mit einer Auswahl eingesehen werden. Darunter auch Projekte zur Erschließung digitaler Texte epigrafischen Quellenmaterials, z. B. *Inscriptions of Aphrodisias* des King's College London.⁵

5. Projektdefinition und -planung

5.1 Zielsetzung

Ziel dieses Teilprojektes war es, die ersten Grundlagen für das TEI-Metadatenschema zur Erfassung sämtlicher Klassischer Maya-Texte zu erarbeiten; das TEI-Metadatenschema bildet hierbei einen Bestandteil des gesamten Metadatenkonzepts. Da das TWKM-Projekt erst am Anfang stand und noch viele Fragen bezüglich der Dateninhalte offen waren, sollte dieses TEI-Metadatenschema eine Grundlage sein, die im weiteren Verlauf des TWKM-Projektes angepasst werden konnte. Daher wurden in den Ergebnissen auch kritische Punkte und Probleme thematisiert. Ein fertiges Metadatenschema war nicht Ziel dieses Teilprojekts.

5.2 Vorgehensweise allgemein

Die Zuständigkeiten innerhalb des TWKM-Projektes teilen sich in zwei Bereiche: Fachwissenschaftlichen Aufgaben, die das Klassische Maya betreffen, sowie die technische und informationswissenschaftliche Betreuung.

Für die Erfassung der Inschriftentexte des Klassischen Maya ist es notwendig, die grundlegende Struktur der Sprache zu kennen. Dies ist in zweierlei Hinsicht von Bedeutung: Zum einen ist es die Grundvoraussetzung für die Erfassung der relevanten Daten, zum anderen ist es für die Verständigung mit den Wissenschaftlern notwendig, um die Bedürfnisse besser nachvollziehen zu

⁵ Vgl. "Projects Using the TEI." TEI Consortium. <http://www.tei-c.org/Activities/Projects/> (04.08.2014) und Reynolds, Roueché & Godard 2007, <http://insaph.kcl.ac.uk/iaph2007/>.

können. Daher war es notwendig, sich in die Sprache des Klassischen Maya einzuarbeiten, um den Aufbau und die Fachbegriffe zu kennen.

Um die für die Wissenschaftler wichtigen Informationen zu erfassen und unterschiedliche Aspekte der Forschung abzudecken, wurden für das Metadatenkonzept verschiedene Ebenen berücksichtigt:

- Materielles Objekt: hierzu zählen neben den Artefakten der Maya auch moderne Dokumentationen wie Abriebzeichnungen, Fundberichte etc.
- Inschrift: die Erfassung der Hieroglyphentexte an sich und aller zu den Inschriften gehörenden Informationen
- Ort: sowohl der Fundort als auch Aufbewahrungsorte (z. B. Museen) sind relevant
- ‚Akteur‘: sowohl die auf den Inschriften genannten Akteure (z. B. Herrscher, Gottheiten) der Inschriftentexte bzw. –Abbildungen als auch moderne Akteure wie an der Ausgrabung beteiligte Forscher oder auch das aufbewahrende Museum werden hierunter gefasst
- Zeit: hierunter fallen die Datierung des Objektes (mit der wichtigen Anforderung der Umsetzung des Mayakalenders in den Gregorianischen Kalender), der Zeitpunkt der Entdeckung usw.

Für die Erfassung aller notwendigen Daten und Informationen wurden unterschiedliche Metadatenstandards herangezogen, um den vielfältigen Facetten gerecht zu werden. So werden die Inschriftenträger überwiegend durch CIDOC CRM⁶ beschrieben. Zur Erschließung der Inschrift selbst wurde das Metadatenschema der TEI herangezogen, das später auch die Basis für die Analyse der Mayaschrift und für die Erstellung des Mayawörterbuches darstellt. Da sich dieses Teilprojekt auf die Erarbeitung der relevanten Metadaten für die Inschriftentexte bezieht, wird im Folgenden dieser Part beschrieben. Die Feldbezeichnungen der Elemente sowie die Begriffe der Textstruktur sind Englisch, die bevorzugte Sprache des TWKM-Projektes bzw. der späteren TWKM-Datenbank.

5.3 Vorgehensweise bei der Erschließung der Inschriftentexte

Ausgehend von den Zielen und Vorstellungen der Fachwissenschaftler, welche sich aus dem Projektantrag an die Akademie der Wissenschaften und Künste und den Gesprächen ergaben, wurden die Anforderungen an das Metadatenschema bezüglich der Inschriftentexte formuliert. Um das sehr umfangreiche TEI-Metadatenset einzugrenzen, wurde eine Auswahl relevanter Module getroffen, die für die Erschließung der Inschriftentexte infrage kommen. Da TEI bereits in anderen epigrafischen Erschließungsprojekten als Grundlage dient, wurden zudem vergleichbare Projekte ermittelt, um Aufschluss über deren Metadatenstruktur zu erhalten.

⁶ Das CIDOC Conceptual Reference Model (CRM) stellt ein Dokumentationsformat für den Bereich des Kulturellen Erbes dar und ist seit 2006 offizieller ISO-Standard (ISO 21127:2006). Dieses Format wurde gewählt, um die zahlreichen Aspekte, wie bspw. Fundhistorie, Aufbewahrungshistorie, Personen wie Ausgräber, Kuratoren etc., die sich auf das Objekt selbst beziehen, adäquat abbilden zu können.

5.4 Anforderungen seitens der Wissenschaft

Die von den Wissenschaftlern gestellten Anforderungen lassen sich in zwei Bereiche gliedern: der erste Teil beinhaltet diejenigen, die sich auf das gesamte Metadatenschema beziehen; der zweite Teil listet diejenigen auf, die speziell bei der Beschreibung der Inschrift berücksichtigt werden müssen.

1. Allgemeingültige Anforderungen

- Metadatenelemente für die Erschließung aller bekannten und künftig gefundenen Inschriftentexte, d. h. unterschiedliche Darstellungen müssen berücksichtigt werden können
- Einbindung zeitlicher und räumlicher Angaben, d. h. Fundort und Datierung müssen immer abrufbar sein
- Schriftvarietäten in Korrelation mit der jeweiligen Zeit (Datierung) muss ablesbar sein, d. h. die genaue Schreibweise der Hieroglyphe resp. der Schriftzeichen muss mit dem jeweiligen (datierten) Text in Verbindung gesetzt werden
- Ermöglichen einer sprach- und schriftbasierten Suchfunktion in der Datenbank, d. h. originale Schreibung, Umschrift und Übersetzung müssen erschlossen sein
- Berücksichtigung nicht entzifferter Textstellen mit einer Abbildung der Originalschreibung
- Verweise auf Sekundärliteratur (Kurzzitate mit einer URN zu einem Literaturverzeichnis)
- Nachnutzbarkeit für andere Projekte sollte gewährleistet sein, d. h. das Metadatenschema sollte möglichst flexibel sein

2. Textspezifische Anforderungen

- Text-Bildbezug abbilden
- Die Anzahl der Textfelder sowie der Hieroglyphenblöcke bzw. Zeichen je Textfeld auf einem Inschriftenträger müssen berechnet werden können
- Form/Darstellung der Texte (Einzel-, Doppelkolumne, rechtwinklig etc.) muss ersichtlich sein
- Farbige Textbereiche definieren können
- Beschreibung unterschiedlich großer Blöcke, d. h. ‚Großbuchstaben‘ und kleiner dargestellte Blöcke müssen unterschieden werden können
- Erschließung der Inschrift und Interpretation des Textes sind zu trennen
- Lesefolge und Ausrichtung der einzelnen Schriftzeichen müssen ausgezeichnet werden

6. Metadatenschema für die Erfassung der Inschriftentexte

Da die Beschreibungssprache TEI aufgrund ihrer ursprünglichen Entwicklungsgedanken für möglichst viele Bereiche der Geisteswissenschaften geeignet sein soll und ein sehr umfangreiches Set an Elementen bereithält, ist die erste Durchsicht nach geeigneten Elementen zeitaufwendig.

Eine Eingrenzung für die Epigrafik bietet EpiDoc (Epigraphic Documents). EpiDoc ist eine internationale Gemeinschaft von Wissenschaftlern mit dem Forschungsschwerpunkt antike Inschriften. Diese Gemeinschaft hat Empfehlungen für die XML-Kodierung der Inschriftentexte erarbeitet, die ein Subset der *TEI-P5-Guidelines* darstellen und speziell auf die Arbeit mit antiken

und mittelalterlichen Texten ausgerichtet sind. Die Empfehlungen sind inzwischen von altgriechischen und lateinischen Inschriften auch auf die Beschreibung Papyri und Manuskripten ausgeweitet worden (vgl. Elliott, Bodard & Cayless et al. 2006-2013). Der Vorteil dieser Empfehlungen liegt darin, die für die Beschreibung von Inschriftentexten ungeeigneten TEI-Elemente von vornherein auszuschließen und zugleich durch eigene Ergänzungen an Definitionen die Beschreibung epigrafischen Quellmaterials optimal zu unterstützen (vgl. Roueché & Flanders 2007-2014).

Um eine erste Auswahl an Elementen zu treffen, welche für die fachgerechte Beschreibung der Inschriftentexte infrage kommen könnten, wurden zunächst die Module der *TEI-P5-Guidelines* sondiert, die auf eine Relevanz schließen ließen (vgl. TEI Consortium 2014, 2). Folgende Bereiche wurden identifiziert:

- header: Jeder TEI-konforme Text muss bestimmte Beschreibungen zu der Datei selbst angeben, sodass das Modul für jede TEI-Datei relevant ist.
- core: Das Modul enthält Elemente, die in allen zu beschreibenden Textgattungen vorkommen können. Viele dieser Kern-Elemente sind flexibel einsetzbar und können an jeder Textstelle erscheinen.
- textstructure: Die Elemente dieses Moduls dienen der Beschreibung der äußeren Textstruktur. Da die Inschriftentexte in der Anordnung der Hieroglyphen strukturiert sind, können Elemente dieses Moduls für die Beschreibung relevant sein.
- gajji: Dieses Modul beinhaltet Elemente für die Beschreibung von ungebräuchlichen Schrifttypen, Symbolen und Hieroglyphen. Da die Maya-Schrift eine Hieroglyphenschrift ist, die sich aus einzelnen Zeichen zusammensetzt, wird dieses Modul in Betracht gezogen.
- figure: Für die Wiedergabe von Abbildungen, Tabellen usw., die in einem Text erscheinen, werden in diesem Modul die Elemente definiert. Auf den Inschriftenobjekten des Klassischen Maya sind oftmals Abbildungen vorhanden, die in Bezug zu dem Text stehen und daher – neben dem Text selbst – ausreichend dargestellt werden müssen.
- transcr: Das Modul definiert die Elemente zur Darstellung der Primärquelle, also der Inschriftentexte selbst. Da in dem TWKM-Projekt die Abbildungen der Quellen (z. B. Digitalfotografien der Inschriftenträger) einbezogen werden müssen, wurde dieses Modul in die Überlegungen einbezogen.

Die Module, die sich augenscheinlich auf analytische Aspekte beziehen bzw. sehr speziell auf einzelne Textgattungen ausgerichtet sind, wurden für die erste Einarbeitung nicht berücksichtigt.

Bei EpiDoc werden die beschreibenden Elemente in verschiedene Bereiche unterteilt, die für eine epigrafische Publikation infrage kommen. Wie aus den TEI-Modulen wurden auch hier die für das Projekt geeigneten Bereiche sondiert (vgl. Roueché & Flanders 2007-2014):

- the edition of the epigraphic text itself: Es werden Anleitungen für die Beschreibung der Textstruktur, der Darstellungsform sowie der Texttranskription gegeben.
- history of the discovery, documentation, and interpretation: Hier wird die Kodierung der bibliografischen Verweise erläutert. Eine Anforderung innerhalb des TWKM-Projektes ist es, bei den jeweiligen Inschriftenlesarten auf die Fachliteratur zu verweisen, in welcher diese Lesart genannt ist.

Die übrigen in EpiDoc angeführten Bereiche beziehen sich zum einen auf Informationen über die Inschriftenträger selbst (Fundhistorie etc.), zum anderen auf Elemente, die die Textanalyse betreffen. Da in dem Gesamtkonzept des Metadatenschemas die Daten zu den Inschriftenträgern in separaten Datencontainern geführt werden, wären sie an dieser Stelle redundant.

Diese Auswahl an Elementen wurde im Anschluss hinsichtlich der Anforderungen der Wissenschaftler geprüft: Welche Elemente gibt es für die Beschreibung der Inschriften-Textstruktur? Welche Elemente eignen sich für die Beschreibung der Hieroglyphen?

Bei der Erarbeitung des TEI-Schemas für die Wiedergabe der Struktur der Hieroglyphentexte zeigte sich Klärungsbedarf bezüglich der Fachtermini und der Forschungsrelevanz bestimmter Angaben. Wie ist die gebräuchlichste Bezeichnung bspw. der Seiten eines Inschriftenträgers, gibt es eine Vorder- und Rückseite? Wie kann der Bezug von Text und Bild hergestellt werden? Welche dieser Angaben gehören zu der sachlichen Wiedergabe des Inschriftentextes, welche sind bereits Interpretation? Und: Wie werden die einzelnen Hieroglyphen eindeutig angesprochen, ohne bereits eine Interpretation vorwegzunehmen?

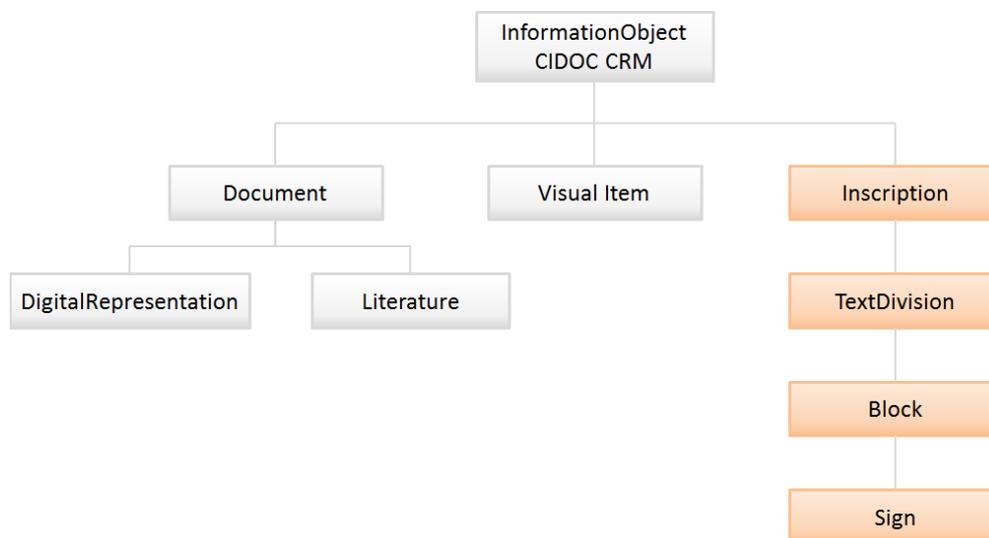
6.1 Wiedergabe der Struktur des Inschriftentextes

Eine Herausforderung bei der Wiedergabe der Textstruktur ist die Vielzahl an Ausgestaltungsformen der Inschriftentexte, die es durch die Metadaten darzustellen gilt: die Anordnung der Hieroglyphenblöcke sowie die Form des Schriftfeldes an sich variiert (Tab. 1).

Anordnung der Hieroglyphenblöcke	Einzelkolumne
	Doppelkolumne
	Kombination von Einzel- und Doppelkolumne
	Zeile
	Kombination von Kolumne und Zeile
	Initiale
Form der Schriftfelder	rechteckig
	Schriftband
	Winkelförmig
	Kartusche (d. h. mit Zierrahmen)
	„Schriftbilder“ (Hieroglyphen als Bestandteil einer Abbildung)
	Sprechblase

Tab. 1: Übersicht der formalen Gestaltungsmöglichkeiten der Inschriften

Um all diese Facetten durch die beschreibenden Daten abzudecken, wurde das Metadatenschema in aufeinander aufbauende Abschnitte gegliedert (s. Abb. 2). Diese Aufteilung sollte die Auswahl der relevanten Metadatenelemente erleichtern und das Vorgehen transparenter für die weitere Nutzung gestalten. Im Folgenden werden die Elemente für die Beschreibung der ‚Inscription‘ sowie der drei Unterabschnitte ‚TextDivision‘, ‚Block‘ und ‚Sign‘ diskutiert und beschrieben.



**Abb. 2: Auszug aus dem Gesamtkonzept des Metadatenschemas
(Farbige Markierung: TEI als Grundlage)**

6.2 TEI-Elemente

Das Basispaar eines TEI-Elementes bilden der TEI-Header und ein Textelement. Der Header enthält Metadaten, die das Dokument als Ganzes beschreiben und kann sehr umfassend oder auch recht ‚schmal‘ gehalten werden. Das Textelement enthält die Metadaten des Dokuments selbst. Das Element <teiHeader> bildet mit seinen beschreibenden bzw. erklärenden Informationen sozusagen das elektronische Titelblatt, während das <text>-Element den Textinhalt des Objekts mit Annotationen, die dessen Struktur und weitere Eigenschaften deutlich machen, enthält.

teiHeader

Das <teiHeader>-Element muss nach den Vorgaben der *TEI-P5-Guidelines* mindestens das Element <fileDesc> (file description), das die elektronische Datei beschreibt, enthalten. Diesem Element sind wiederum drei obligate Elemente zugewiesen: <titleStmt>, <publicationStmt> und <sourceDesc>.

Das <title>-Subelement @type, mit dem auf alternative Namensformen verwiesen werden kann, ist an dieser Stelle redundant; in der Fachliteratur alternativ genannte Bezeichnungen für die Inschriftenträger werden in einem sogenannten Vokabular⁷ hinterlegt, sodass hier eine Beschränkung auf die gebräuchliche Bezeichnung ausreichend scheint.

⁷ Die Vokabulare, die für das TWKM-Projekt erstellt werden, werden durch nach dem ‚Simple Knowledge Organisation System‘ (SKOS) kodiert.

Ebenso wird auf die Wiedergabe von Personen, die in Verbindung mit dem Objekt stehen, an dieser Stelle verzichtet. Diese Angaben werden in der CIDOC CRM-Kategorie ‚Actor‘ bzw. ‚Appellation‘ hinterlegt, die Verbindung zu dem Objekt wird in dem Metadatenschema über die eindeutige URI der TWKM-ID gewährleistet. Das bietet den Vorteil, dass Daten, die an anderer Stelle vorgehalten werden, nicht nochmals erstellt werden müssen. Ebenso verhält es sich mit den Objektdaten: Maße, Fundkontext, Datierung etc. können durch das Metadaten-set von CIDOC CRM ausführlich und adäquat ausgezeichnet werden. Das bedeutet, dass für den `teiHeader` nur wenige Elemente genutzt werden, beispielsweise entfallen die Angaben `<extent>`, `<notesStmnt>`, `<author>` oder auch `<geoDecl>` für die Fundkoordinaten - die Dateneingabe ist hier demnach gering und wenig aufwendig.

Für das TWKM-Projekt könnten das Element `<teiHeader>` demnach auf folgende Angaben reduziert werden:

```
<teiHeader>
  <fileDesc>
    <titleStmnt>
      <title>[TWKM-ID]</title>
    </titleStmnt>
    <publicationStmnt>
      <authority>[name]</authority>
      <idno type="URI">[Verlinkung zu der Objekt-ID]</idno>
    </publicationStmnt>
    <sourceDesc>
      <p>[z. B. Copan, Stele D]</p>
    </sourceDesc>
  </fileDesc>
</teiHeader>
```

Die Identifikationsnummer innerhalb des `<publicationStmnt>`-Elements führt über einen Hyperlink zu dem jeweiligen Objekt selbst und damit zu allen den Inschriftenträger betreffenden Metadaten.

In den TextGrid-Empfehlungen werden zusätzlich `<encodingDesc>` (Beschreibung der Kodierung) und `<editorialDecl>` (Beschreibung der Editionsprinzipien) mit dem Element `<normalization>`, das den Grad der Vereinheitlichung und Normalisierung wiedergibt, angegeben (Vgl. Blümm & Wegstein 2008, 22f). Ob diese Elemente für das TWKM-Projekt an dieser Stelle praktikabel sind, ist zu prüfen.

Inscription

Für die Beschreibung der Inschriftentexte muss darstellbar sein, dass es mehrere Inschriften auf einem Objektträger geben kann und dass sich einzelne Texte auf Abbildungen beziehen können. Die Beschreibung der Inschriftentexte muss das Gesamtbild, also die Anordnung der Texte und zugehöriger Abbildungen, widerspiegeln.

Vor der Beschreibung des Inschriftentextes wird – analog dem Beispiel von EpiDoc (vgl. Bodard 2007-2014) – auf das digitale Faksimile (Digitalisat eines Abriebes, einer Zeichnung oder eine

Digitalfotografie) mit dem Element <facsimile> und der entsprechenden URI des Digitalisats verwiesen.

Die Inschrift wird durch das Tag <text> gekennzeichnet. Dieses Element enthält entweder einen einzelnen, eigenständigen oder einen aus mehreren Teilen bestehenden Text. Wenn mehrere Texte zusammengehören, wird das <text>-Element durch <group> umschlossen, um die Einheit darzustellen (vgl. TEI Consortium 2014, 150 und 1445). Dies könnte eventuell für die Beschreibung zweier zusammengehörender Fragmente für die Maya-Inschriften sinnvoll sein. Der Text selbst wird in dem Element <body> wiedergegeben, allerdings enthält dieses Element jeweils nur die eigenständigen Texte. D. h., dass ab dieser Beschreibungsebene ausschließlich die einzelnen Texte angesprochen werden.

Zwei weitere Elemente des Körpers sind <front> und <back>: <front> dient der Beschreibung aller Inhalte die vor dem eigentlichen Text stehen (z. B. Titelseite, Vorwort, Widmung), <back> alle Teile, die diesem angehängt sind. Es wäre allerdings denkbar, dass Einleitungsformeln oder auch Schlussformeln (z. B. die Nennung des Künstlers, der die Inschrift schuf), durch diese Elemente von der eigentlichen Textbeschreibung differenziert würde. Da dies bereits Interpretation der Inhalte darstellt, soll auf die Verwendung der Tags <front> und <back> verzichtet werden; für die Maya-Inschriften ist die Verwendung des <body>-Elements ausreichend.

Da die Maya-Texte an unterschiedlichen Stellen auf dem Objekt auftreten, wurde zunächst die Seite definiert. Für die Wissenschaftler ist es üblich, von Vorder- und Rückseite der Inschriftenträger zu sprechen. Als Vorderseite wird – soweit vorhanden – diejenige mit dem Herrscherbildnis bezeichnet, ansonsten die mit der Datumsangabe. Daraus ergeben sich die Seitenbezeichnungen: front, right, left, back. Allerdings sind diese Bezeichnungen nicht mit den TEI-Elementen, die bereits von der Verwendung ausgeschlossen wurden, zu verwechseln. Diese Spezifizierung ist Bestandteil des <body>-Elements, nicht des <text>-Elements. Wenn es sich um einen zusammenhängenden Text handelt, der über mehrere Seiten läuft, ist die Spezifizierung Bestandteil der Textdivision (s. u.).

Im TWKM-Projekt wurden bereits Abkürzungen für die Beschreibung der Bilder festgelegt, die analog zur Beschreibung der Textfelder verwendet werden können, sodass eine einheitliche Ansprache entsteht:

Abkürzung für		Erläuterung
f bzw. b	front bzw. back	Die Seite mit der Herrscherabbildung oder der Angabe des Datums gilt im Allgemeinen als Vorderseite. Geklärt werden muss hier, wie mit den Objekten zu verfahren ist, bei denen diese Angaben nicht bekannt/ersichtlich sind.
l bzw. r	left bzw. right	Die Seiten links und rechts ausgehend von der Vorderseite.
t bzw. u	top bzw. underside	Bezeichnung für die Ober- und Unterseite des Inschriftenträgers. Inschriften auf der Unterseite sind bspw. bei Türsturzen und der Standfläche von Keramik vorhanden.
g	girth	Wird für einen umlaufenden Text verwendet, bspw. bei zirkulären Altären.

Tab. 2: Bezeichnungen für die Spezifizierung @type von <body>

Ob die Bezeichnung ‚girth‘ auch bei Vasenabrollungen bzw. umlaufenden Texten verwendet werden soll, ist noch offen. Ebenso ist die Umsetzung für unregelmäßige Objekte, wie z. B. Inschriften auf Zoomorphen (Skulpturen in Tierform) oder in Höhlen, ungeklärt.

Eine besondere Herausforderung stellte die Umsetzung der variantenreichen Anordnungsmöglichkeiten der Hieroglyphenblöcke wie auch die der Schriftfeldformen dar; bei einer Kolumne bspw. muss eindeutig nachvollziehbar sein, an welcher Stelle eine neue Zeile beginnt, an welcher Stelle eine Kolumne endet und wo die Lesefolge in der nächsten Spalte beginnt. Vergleichbar ist dies mit dem Lesen einer Zeitung. Wie lassen sich Einzelkolumnen und Doppelkolumnen umsetzen? Ausgehend von diesen ‚einfachen‘ Beispielen wurde insgesamt nach einer Möglichkeit der Strukturwiedergabe gesucht. Diese Grundlage könnte dann für weitere Erscheinungsformen z. B. einen rechtwinkligen Text geprüft und erweitert werden.

TextDivision

‚TextDivision‘ bildet die Unterklasse zu ‚Inscription‘ und beschreibt einen ganz bestimmten Textabschnitt bzw. ein Textfeld auf einem Inschriftenträger. Für die Beschreibung eignet sich das Element <div> des TEI-Standards. Dieses Element kann entweder gezählt oder ungezählt verwendet werden. Die gezählte Variante spiegelt eine Hierarchie einzelner Textabschnitte wider, wobei <div1> die oberste Ebene, <div2> die nächstuntergeordnete usw. beschreibt. Da bei den Inschriftentexten keine Hierarchie der einzelnen Textabschnitte vorliegt und diese jeweils gleichwertig nebeneinander betrachtet werden, wird die Variante ohne Zählung gewählt. Klassifizierungen des Textes sind durch die Attribute @type bzw. @subtype möglich. So können beispielsweise einzelne Textteile separat beschreiben werden; analog zu dem Element <body> können die ‚Passagen‘ durch @n genauer definiert werden. Für die Klassifizierung ist die Unterscheidung nach Anordnungstypen sinnvoll (s. Tab. 2). Allerdings wäre hierfür ein eindeutiges Vokabular zu erstellen, z. B. die Anordnungsmöglichkeiten der Hieroglyphenblöcke als Wert des Attributes @type und die Formenbeschreibung als Wert zu @subtype:

```
<div type="combination-column-line" subtype="right-angled">
```

Durch die Ergänzung einer Zählung, kann das entsprechende Textfeld innerhalb der Inschriftenseite genauer beschrieben werden:

```
<div n="B1-D3" type="combination-column-line" subtype="right-angled">
```



Abb. 3: Maya-Inschrift mit bildlichen Darstellungen und Beschriftung der Hieroglyphen (Rasterung); Yaxchilan, Türsturz 8⁸

In der Archäologie sind häufig auch nur Bruchstücke von Inschriftentexten vorhanden; die EpiDoc-Empfehlungen sehen hierfür den @type ‚fragment‘ vor, der der entsprechenden Beschreibung des Textabschnitts vorangestellt wird:

```
<div type="fragment">
```

Das Ende einer Kolumne wird durch <cb> (column break) getaggt. Zur Berücksichtigung der drei erhaltenen Inschriften-Kodizes ist zusätzlich die Beschreibung eines Seitenwechsels notwendig. Der Beginn einer neuen Seite wird durch <pb> (page break) ausgezeichnet.

In der Wissenschaft werden die einzelnen Hieroglyphenblöcke durch eine Rasterung vergleichbar der Aufteilung eines Schachbretts angesprochen; diese Bezeichnung muss sich in den TEI-Elementen wiederfinden. Unter Inscription wird die Gesamtrasterung der Inschrift wiedergegeben, sodass jede einzelne Hieroglyphe gezielt angesprochen werden kann, z. B. ist der Block D3 der Abb. 3 eindeutig festgelegt. Bei dieser Rasterung kann es jedoch vorkommen, dass ein Block nicht eindeutig den beiden Koordinaten zuzuordnen ist bzw. dass auf einem Koordinatenpunkt zwei Blöcke stehen. In diesem Fall wird eine weitere Untergliederung gemacht, sodass die ‚Unterblöcke‘ bspw. als ‚A2a‘ und ‚A2b‘ bezeichnet werden. Mithilfe der ‚Koordinaten‘ kann die Grundstruktur der Inschrift beschrieben werden.

Der Bezug von Text und Bild ist sowohl auf Ebene des Textabschnitts als auf Ebene einzelner Blöcke relevant. Es gibt unterschiedliche Konstellationen: ein Textabschnitt kann sich insgesamt auf eine bildliche Darstellung beziehen, der Textabschnitt ist als ‚Sprechblase‘ zu einem Akteur zu betrachten oder die Blöcke bzw. ein Block befindet sich auf einem Akteur oder einem Gegenstand. Für die eindeutige Bezeichnung wird ein kontrolliertes Vokabular erstellt.

Bei der Überlegung, wie ein Text-Bild-Bezug beschrieben werden kann, kam der Vergleich mit der Gattung ‚Comic‘ auf. Eine Recherche ergab, dass es die auf TEI basierende *Comic Book Markup*

⁸ Graham & von Euw 1977, Taf. 3:27. URL: <https://www.peabody.harvard.edu/CMHI/flash/detailview.swf?num=8&site=Yaxchilan&type=Lintel> (05.08.2014).

Language (CBML; Walsh 2012) gibt. Für die Auszeichnung von ‚Sprechblasen‘ wird als Tag <balloon>⁹ für einen zum Bild gehörenden Text <caption>¹⁰ in einem eigenen CBML-Modul eingeführt. Es wurde diskutiert, ob für die Beschreibung der Inschriftentexte und Abbildungen analog verfahren werden könnte; hierfür wäre entweder das cbml-Modul zu integrieren oder eine eigene Typisierung zu definieren. Allerdings ist <caption> auch in TEI definiert, sodass die TEI-Elemente evtl. ausreichen.

Nach den TEI-P5-Guidelines wurde die Wiedergabe von Text-Bild-Bezügen durch <figure> umgesetzt. Die bildliche Darstellung wird durch <graphic> und einer URL definiert. Eine Beschreibung des Bildes durch das Element <figDesc> ist nicht notwendig, da dies in den CIDOC CRM abgedeckt wird.

Bsp.:

```
<figure>
  <graphic url="..."/>
  <ab type="caption">[Schriftzeichen, die zu einem Bild gehören]</ab>
</figure>
```

Schriftzeichen auf Figurendarstellungen (Personen, Götter, Tiere) oder auf Gegenständen sind häufig. In einer Diskussion zeigte sich, dass es in der Darstellung wichtig ist, an welcher Stelle der Schriftpart steht: Auf dem Kopfschmuck ist er ein Zeichen für den Herrscher, Schriftzeichen auf dem Oberschenkel sind ausschließlich bei den Untergebenen (vgl. Abb. 4) zu finden. Die Schrift ist hier demnach Ausdruck des soziokulturellen Gefüges und damit eine wichtige Information für die Forschung. Um die Unterscheidungen durch die Metadaten beschreiben zu können, wurde durch die Bonner Wissenschaftler ein weiteres Vokabular erstellt, das zur Spezifizierung durch das type-Attribut dient.

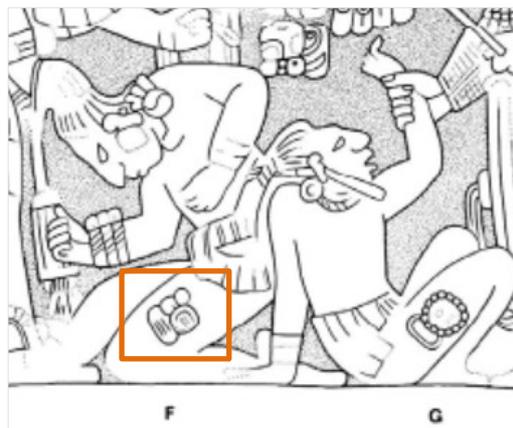


Abb. 4: Hieroglyphen auf Personen zum Ausdruck der gesellschaftlichen Stellung (Auszug aus Yaxchilan, Türsturz 8)

Block

Eine Lösungsversuch für die Beschreibung der Blöcke sah eine Unterteilung des <div>-Elements durch ein definiertes Attribut mit Angaben der genauen Blockkoordinaten (z. B. A1) vor, sodass ein Block folgendermaßen beschreiben wäre: <div type="block" n="coordinates". Nach dem

⁹ „<balloon>“. In: Walsh 2012, <http://dcl.slis.indiana.edu/cbml/schema/cbml.html#TEI.balloon> (10.08.2014).

¹⁰ „<caption>“. In: Walsh 2012, <http://dcl.slis.indiana.edu/cbml/schema/cbml.html#TEI.caption> (10.08.2014).

gleichen Schema wäre das einzelne Logogramm oder Silbenzeichen als @subtype=sign definiert worden. Bereits beim Erarbeiten weiterer, relevanter Beschreibungskriterien wie bspw. die Hervorhebung einzelner Zeichen erwies sich dieser Ansatz als unbrauchbar. Nach den *TEI-P5-Guidelines* sind innerhalb des <div>-Elements sehr wenige Core-Elemente wie bspw. <gap> (Lücke) zulässig. Das benötigt Tag <hi> (highlighted) jedoch zur Kennzeichnung farbiger Blöcke ist nicht erlaubt. Daher musste ein anderer Lösungsweg gefunden werden.

Nach Durchsicht der Elemente und der Suche nach vergleichbaren Fällen in den Richtlinien von EpiDoc schien die Lösung, ein Element dazwischenschalten. Infrage kommen hierfür <l> (line) oder <ab> (anonymous block), wobei <l> gemäß den *TEI-P5-Guidelines* zur Beschreibung von Versen dient. Im Gegensatz zu <l> kann <ab> freier gestaltet werden, sodass dieses Element gewählt wird (vgl. TEI Consortium 2014, 508):

```
<div n=A type="column">
  <ab type="Block" n=A1>
    T1:257.1:624:178
  </ab>
  ...
</div>
```

Eine alternative Darstellung der Blöcke ermöglicht das Element <milestone>:

```
<milestone unit="block" n=A1>T1:257.1:624:178
<milestone unit="block" n=A2>...
```

Die Verwendung des <milestone>-Tags ist allerdings vor der Verwendung zu diskutieren. "Since it is not structural, validation of a reference system based on milestones cannot readily be checked by an XML parser, so it will be the responsibility of the encoder or the application software to ensure that they are given in the correct order" (TEI Consortium 2014, 114 f).

Für eine klarere Beschreibung der Struktur war es sinnvoll, Zeilenumbrüche auszuzeichnen. Hierfür wird das Element <lb> (line break) an der Stelle des ‚Zeileneendes‘ gesetzt, d. h. bei einer Doppelkolumne nach dem zweiten Hieroglyphenblock.

Eine Variante des TEI-Metadatenschemas für eine Doppelkolumne, deren erster Block größer dargestellt ist, könnte folgendermaßen aussehen:

```

<text>
  <body type="front">
    <div type="column" n=A>
      <ab type="block" n=A1.B1>
        <hi rend="tall">[Schriftzeichen]</hi>
      </ab>
      <ab type="block" n=A2>
        </lb>[Schriftzeichen]
      </ab>
      <ab type="block" n=B2>
        </lb>[Schriftzeichen]
      </ab> ...
    </div>
  </body>
</text>

```



Sign

Ein Hieroglyphenblock besteht i. d. R. aus drei bis vier (max. fünf) Zeichen in unterschiedlichen Zusammensetzungen. Die Wiedergabe der Leserichtung ist hierbei eine wichtige Anforderung. In der Wissenschaft hat sich hierfür ein Standard durchgesetzt, nach dem bspw. nebeneinander stehende Zeichen durch einen Punkt, übereinander stehende durch einen Doppelpunkt getrennt werden. Diese Konvention spiegelt ebenfalls wider, ob ein Zeichen vertikal oder horizontal in dem Block integriert ist; für die Wiedergabe der Zeichenanordnung kann dieser Standard verwendet werden.¹¹

	T1	T257	T1	T624	T178
	T1:257.1:624:178				

Abb. 5: Darstellung der Lesefolge einzelner Zeichen eines Hieroglyphenblocks (vgl. Grube 2011, 7)

Die Darstellung der Schriftzeichen des Klassischen Maya ist vielfältig und auch nach Region und Zeit unterschiedlich bzw. einem Wandel unterlegen. Daher war es notwendig, dass die Schriftzeichen jeweils mit der Originalschreibung verknüpft werden. Nur so können die Entwicklung und die Varianten der Zeichen greifbar gemacht werden. Für die Wiedergabe des Inschriftentextes wurden nach der in der Wissenschaft gängigen Methode die Zeichen durch eine Klassifikation dargestellt: z. B. T178 wäre nach der Klassifikation von Thompson die Wiedergabe der Silbe ‚la‘. In diesem Bereich liegt bereits der Übergang zu einer Interpretation der Zeichen vor und ist daher kritisch zu sehen.

Neben der Klassifikation von Thompson gibt es noch Weitere, die für das TWKM-Projekt in einer eigenen Zeichenkonkordanz zusammengeführt und ergänzt werden sollten. Die Zeichen erhalten eine eindeutige Identifikationsnummer, die später vorrangig verwendet wird. Die Konkordanz wird im Laufe des TWKM-Projektes erstellt und nach Bedarf erweitert: Nicht interpretierbare

¹¹ Der im Antrag vorgesehenen Typisierung der Lesefolge wurde nicht weiter nachgegangen. Vgl. Grube 2011, Anlage 11.

Schriftzeichen werden nicht durch ein Fragezeichen beschrieben, sondern erhalten bereits eine eindeutige Nummer innerhalb der Konkordanz; die Lesung, Transkription etc. können je nach Kenntnisstand ergänzt werden. Da die Nummer jeweils der standardisierten Zeichenform zugewiesen wird, müssten auch die jeweiligen Schreibvarianten eine eigene ID erhalten, da nur so die Verbindung der Schreibvariante zum zeitlichen und geografischen Gebrauch hergestellt werden kann. Dies erlaubt, dass in der Beschreibung durch die Metadaten stets eine eindeutige Nummer verwendet werden kann; dadurch können – wie in den Anforderungen verlangt – nicht entzifferte Textstellen mit Verweis auf die Originalschreibung berücksichtigt werden. Für die Umsetzung der Zahlzeichen, die eine eigene Kategorie innerhalb der Maya-Schrift darstellen, lag noch keine Lösung vor, sodass sich nach dem aktuellen Stand nicht alle Datierungen in den Inschriften wiedergeben ließen.

Denkbar wäre, die Konkordanz nach TEI bspw. gemäß einer Taxonomie zu beschreiben (vgl. TEI Consortium 2014, 46 f). Die einzelnen Schriftzeichen ließen sich demnach mit einer ID ansprechen. In der Beschreibung würde dann bspw. das Attribut `xml:id="I156"` für eine TWKM-Nummer gleichgesetzt.

Für die Beschreibung der Zeichen nach der Originalvorlage können eventuell die TEI-Elemente `<g>` bzw. `<glyph>` (Referenz zu `<g>`) verwendet werden, die insbesondere für Zeichen eingesetzt werden, für die kein Unicode existiert (vgl. TEI Consortium 2014, 181). Die EpiDoc-Empfehlungen beschränken sich darauf, `<g>` nur zu verwenden „where a symbol is non meaning-bearing“, in einem folgenden `@type`-Attribut wird das Symbol beschrieben, bspw. ein Kreuzifix.¹² Denkbar wäre ein TWKM-Projekt-spezifisches Modul für die Konkordanz, das ähnlich dem XML-Schema zur Beschreibung des Tags `<glyph>` strukturiert ist.

Da in der Wiedergabe der Zeichen eine Interpretation vorliegt, ist es wichtig, jede Lesart mittels einer Verweisung auf die Sekundärliteratur zu belegen. Für die Sekundärliteratur wird ein Verzeichnis mit dem quelloffenen Literaturverwaltungsprogramm Zotero angelegt; Zotero erlaubt einen Datenexport in das TEI-Format. Die Verweisung auf einen entsprechenden Eintrag erfolgt durch das `<ref>`-Tag, das auf den entsprechenden Eintrag in dem Literaturverzeichnis verlinkt:

```
<ref target="#Stuart 2008>158-159"></ref>
```

6.3 Fehlende und nicht lesbare Textstellen, Hieroglyphen und Zeichen

Lücken im Text können auf allen drei Unterabschnitten im Inschriftentext dargestellt werden, je nach Umfang der fehlenden Textstelle, d. h. innerhalb der Beschreibungen von `<div>`, `<block>` und auch `<sign>`. Sie werden jeweils durch das Element `<gap>` eingeleitet und durch ein Attribut genauer definiert. Nach den *TEI-P5-Guidelines* sind die Attribute optional, allerdings ist es sinnvoll hier den EpiDoc-Empfehlungen zu folgen, die das Attribut `@reason` verbindlich vorgeben. Als Werte sind `,lost'`, `,illegible'`, `,omitted'` und `,elipsis'` vorgesehen.¹³ EpiDoc bietet sehr umfassende Vorgaben zur Beschreibung nicht darzustellender Textstellen. Unter anderem ist es möglich, auch die Quantität einer Lücke, soweit bekannt, anzugeben:

```
<gap reason="illegible" quantity="1" unit="block"/>
```

¹² "Symbol (Non meaning-bearing)". In: *EpiDoc-Guidelines*. <http://www.stoa.org/epidoc/gl/latest/trans-symbol.html> (22.07.2014).

¹³ "<gap>". In: *EpiDoc-Guidelines*. <http://www.stoa.org/epidoc/gl/latest/ref-gap.html> (15.08.2014).

Das in wissenschaftlichen Publikationen verwendete sogenannte Leidener Klammersystem wird auch in der Maya-Forschung zur Umsetzung der Originalinschrift verwendet, wodurch Textlücken sowie deren Quantität dargestellt werden können. Somit bietet sich eine Verwendung der EpiDoc-Implementierung für die Umsetzung des Leidener Klammersystems an.

6.4 Kritische Betrachtung des Metadatenschemas

Durch die kleinteilige Untergliederung der Inschrift besteht die Gefahr, dass die TEI-Struktur unübersichtlich wird – hier ist zu überlegen, ob Elemente weggelassen werden können, und dennoch dasselbe Ergebnis erzielt wird. Oder ob für die unterschiedlichen Anordnungsmöglichkeiten (Einzel-, Doppelkolumne etc.) eine jeweils angepasste Auswahl an Elementen definiert werden sollte (vergleichbar mit den *TEI-P5-Guidelines* mit den Unterteilungen nach Textgattung). Es ist sinnvoll, mehrere optionale Elemente zu definieren, sodass je nach Bedarf aus dem Set ausgewählt werden kann.

Ein Vergleich mit den Anforderungen der Wissenschaftler, die im Laufe des Projektes formuliert wurden, zeigte, dass das Elemente-Set diese zu einem hohen Grad abdeckt. Nicht gelöst war die Problematik, die Zeichen als Einzelbestandteile der Blöcke eindeutig zu identifizieren. Nach der aktuellen Auszeichnung werden die Zeichen hintereinander vergleichbar einem Fließtext geschrieben. Die Zeichenkonkordanz, die die einzelnen Zeichen und deren jeweiligen Varianten mittels einer `xml:id` auszeichnet, kann hier Abhilfe schaffen. Dadurch wäre auch die Berechnung der Anzahl aller in einer Inschrift enthaltenen Zeichen möglich. Für die Beschreibung der Zeichen gibt es unterschiedliche Möglichkeiten: Es konnte mit Abschluss dieses Projektes noch nicht entschieden werden, ob sich das Element `<g>` bzw. `<glyph>` oder `<milestone>` für die Auszeichnung der einzelnen Zeichen eignet. Evtl. kann hier, nach der Erstellung der Konkordanz, eine praktikable Lösung gefunden werden. Die geforderte Trennung von reiner Textbeschreibung und Interpretation war nicht möglich, da für die einzelnen Zeichen keine eindeutig kodierte Sprache vorliegt; weil die Schrift und Sprache des Klassischen Maya noch nicht gänzlich erforscht ist, waren zum Teil noch keine eindeutigen Zuweisungen erfolgt.

Fraglich war, ob die Wiedergabe der Anordnung der Schriftzeichen innerhalb eines Blockes durch den bisherigen Standard (Punkt für zwei nebeneinanderstehende Zeichen usw.) für die Forschung ausreicht oder ob hierfür eine präzise Auszeichnung, die eine gezielte Abfrage ermöglicht, notwendig ist. Evtl. könnte auch die in dem TWKM-Projektantrag genannte Typisierung der Zeichenanordnung eine gute Lösung bieten (vgl. Grube 2011, Anlage 11).

Da zu den Texten oftmals Abbildungen gehören, die für eine Deutung und damit die Interpretation des Textinhaltes äußerst relevant sind, wurden für die Auszeichnung der bildlichen Darstellungen zu wenige Elemente genutzt. Größenverhältnisse von Abbildungen bzw. die Angabe der genauen Position waren nach derzeitigem Stand nicht möglich. Um die genaue Position zu bezeichnen und auch leere Flächen aufzuzeigen, müsste für die Objekte eine weitere Rasterung definiert werden, das nicht nur die Hieroglyphen durch Koordinaten festlegt, sondern immer in gleichem Seitenverhältnis für alle Inschriften gilt. So könnten die bildlichen Darstellungen eindeutig beschrieben und evtl. daraus das Größenverhältnis ermittelt werden.

Die Auswahl der Elemente kann als Grundlage zur weiteren Ausarbeitung dienen; bis zu der endgültigen Fertigstellung werden vermutlich noch weitere Aspekte berücksichtigt werden

müssen – in den Diskussion werden stets neue Anforderungen bzw. Schwierigkeiten identifiziert. Des Weiteren ist zu erörtern, ob dieses Metadatenschema auch seltene Formen der Inschriftentexte, die nicht in den verfügbaren Beispielen thematisiert wurden, beschreiben kann. Sobald auf dieser Arbeitsgrundlage die optimale Darstellung der Zeichen entschieden ist, kann im Anschluss die Transkription und Transliteration der Zeichen, die ebenfalls in TEI beschrieben werden sollen, erfolgen.

Das Metadatenschema zeigt, dass die Attribute der *TEI-P5-Guidelines* für die Erfassung der Inschriftentexte zusätzlicher Anpassungen bspw. bezüglich der Attributwerte bedarf; häufig wurden die Präzisierungen der EpiDoc-Richtlinien herangezogen. Allerdings scheinen insbesondere bei der Beschreibung des Bild-Text-Bezuges TWKM-spezifische Anpassungen notwendig. Die Auswahl an Elementen zeigt zudem, dass das Schema aus einem Mix unterschiedlicher TEI-Module besteht; nur so konnten die verschiedenen Aspekte der Inschriftentexte berücksichtigt werden.

7. Fazit

Die Einarbeitung für dieses Projekt verfolgte zwei sehr komplexe Stränge: ein Grundverständnis des Klassischen Maya ist Voraussetzung, um den Diskussionen der Wissenschaftler dieses Bereichs folgen und die Anforderungen nachvollziehen zu können. Zum anderen ist die (Grund-)Kenntnis des Formats TEI unerlässlich. Hier zeigt sich, dass die Vorauswahl nach Modulen für einen ersten Überblick hilfreich war. Sowohl die TEI-Module als auch die Abschnitte der EpiDoc-Empfehlungen erleichtern die Auseinandersetzung mit dieser bislang fremden Materie. Die *TEI-P5-Guidelines* erlauben einen schnellen Einstieg und in der Online-Version ein schnelles Suchen nach einzelnen Elementen, die stets mit Beispielen die Verwendungsmöglichkeiten aufzeigen. Allerdings war es schwierig, Pflichtelemente eines Moduls zu identifizieren: aus der Übersicht der Einzelelemente geht nicht hervor, welches Element in der Hierarchie darunter obligat und welches optional ist. Darauf wird nur in den Ausführungen der *Guidelines* hingewiesen. Daher ist es stets notwendig, die jeweiligen Kapitel der ausgewählten Elemente zu prüfen.¹⁴ Bei Problemen, für die die EpiDoc-Empfehlungen keine Lösungen anboten, ergab auch die Recherche bei anderen Projekten, die Inschriften digital erschließen, keinen weiteren Aufschluss, wie bspw. bei der Umsetzung der einzelnen Zeichen. Dennoch zeigt das Beispiel der ‚Comic book markup language‘, dass es nicht nur bei epigrafischen Projekten evtl. Lösungsansätze gibt.

Der regelmäßige Austausch zwischen allen Projektbeteiligten war eine Grundvoraussetzung für das Gelingen eines solchen Projektes – nicht in dem Metadatenschema oder bei der technischen Infrastruktur (Layout, Suchfunktionen etc.) berücksichtigte Anforderungen sind später nur mit großem Aufwand zu korrigieren. Daher war zu Beginn eine eher feingliedrige und sorgfältige Zusammenarbeit wichtig. In den Treffen des TWKM-Projektteams kristallisierte sich schrittweise heraus, welche Daten und Informationen für die Umsetzung des TWKM-Projektes von Bedeutung sind.

¹⁴ Bei Verwendung eines XML-Editors wie z. B. von oXygen können die Daten leicht auf Validität und Wohlgeformtheit geprüft werden.

Die Erarbeitung des Metadatenkonzepts insgesamt zeigte, dass dieser Bereich große Übereinstimmungen mit der bibliothekarischen Erschließung aufweist: Die Ansetzung normierter Daten für die Namen, die Erstellung kontrollierter Vokabulare und das Erkennen von gemeinsamen Strukturen innerhalb der Daten sind aus der Formal- und Sacherschließung sowie der Ansetzung von Normdaten in Wissenschaftlichen Bibliotheken zu finden – auch wenn die Auszeichnungssprachen für das TWKM-Projekt in Wissenschaftlichen Universalbibliotheken vermutlich kaum eine Rolle spielen.¹⁵ Ein Blick über den Tellerrand lohnt sich auch für Bibliothekare, da sie durch ihre Expertise Forschungsvorhaben aus dem Bereich der sogenannten Digital Humanities sinnvoll unterstützen könnten.

¹⁵ Nach den DFG-Praxisregeln zur Digitalisierung von 2009 soll für die Erschließung mittelalterlicher Handschriften das TEI-Format verwendet werden (vgl. dort, S. 18). U. a. die Herzog-August-Bibliothek in Wolfenbüttel sowie die Universitätsbibliothek Heidelberg folgen dieser Empfehlung.

Abbildungsverzeichnis

Abb. 1: Geografische Lage des Maya-Gebietes

Abb. 2: Auszug aus dem Gesamtkonzept des Metadatenschemas (Gelbe Markierung: TEI als Grundlage)

Abb. 3: Maya-Inschrift mit bildlichen Darstellungen und Beschriftung der Hieroglyphen (Rasterung); Yaxchilan, Türsturz 8

Abb. 4: Hieroglyphen auf Personen zum Ausdruck der gesellschaftlichen Stellung (Auszug aus Yaxchilan, Türsturz 8)

Abb. 5: Darstellung der Lesefolge einzelner Zeichen eines Hieroglyphenblocks

Tabellenverzeichnis

Tab. 1: Übersicht der formalen Gestaltungsmöglichkeiten der Inschriften

Tab. 2: Bezeichnungen für die Spezifizierung @type von <body>

Literatur- und Quellenverzeichnis

Blümm, Mirjam; Wegstein, Werner. 2008. „The TEI header for texts in baseline encoding“. In: Mirjam Blümm et al. (Hrsg.): *TextGrid's baseline encoding for text data in TEI P5 (2007-2009)*, S. 19-27. <http://www.textgrid.de/fileadmin/TextGrid/reports/baseline-all-en.pdf> (25.08.2014).

Bodard, Gabriel (2007-2014). „Structure of an EpiDoc edition“. In: Tom Elliott; Bodard, Gabriel; Cayless, Hugh, et al.: *EpiDoc Guidelines: Ancient documents in TEI XML (Version 8)*. <http://www.stoa.org/epidoc/gl/latest/supp-structure.html> (21.07.2014).

Elliott, Tom; Bodard, Gabriel; Cayless, Hugh, et al. 2006-2013. *EpiDoc: Epigraphic Documents in TEI XML*. <http://epidoc.sf.net> (15.08.2014).

Graham, Ian; van Euw, Eric. 1977. *Corpus of Maya Hieroglyphic Inscriptions*. Vol. 3.1: Yaxchilán. Cambridge: Peabody Museum of Archaeology and Ethnology, Harvard University.

Gronemeyer, Sven. 1999. *Das Schriftsystem der Maya: Hausarbeit im Rahmen des Proseminars „Schriftsysteme Amerikas“*. Bonn. Online unter: <http://www.sven-gronemeyer.de/research/schrift.html> (04.08.2014).

Grube, Nikolai. 2011. *Textdatenbank und Wörterbuch des Klassischen Maya (TWKM). Antrag für ein Forschungsprojekt im Rahmen des Forschungsprogramms der Deutschen Akademien der Wissenschaften (Akademieprogramm) für 2013*. Bonn.

Grube, Nikolai. 1993. *Schrift und Sprache der Maya*. In: Die Welt der Maya. Reiss-Museum der Stadt Mannheim. 3., überarb. u. veränderte Aufl. Mainz: von Zabern.

Grube, Nikolai; Gaida, Maria. 2006. *Die Maya: Schrift und Kunst*. Berlin; Köln: SMB-DuMont. „Projects Using the TEI“. TEI Consortium. <http://www.tei-c.org/Activities/Projects/> (04.08.2014).

- Pennsylvania Sumerian Dictionary*. University of Pennsylvania.
<http://psd.museum.upenn.edu/epsd1/index.html> (04.08.2014).
- Reynolds, Joyce; Roueché, Charlotte; Bodard, Gabriel. 2007. *Inscriptions of Aphrodisias*,
<http://insaph.kcl.ac.uk/iaph2007>.
- Roueché, Charlotte; Flanders, Julia (2007-2014). "Gentle Introduction to Mark-up for Epigraphers".
In: Tom Elliott et al.: *EpiDoc Guidelines: Ancient documents in TEI XML (Version 8)*.
<http://www.stoa.org/epidoc/gl/latest/intro-eps.html> (15.08.2014).
- Thesaurus Linguae Aegyptiae*. Arbeitsstelle Altägyptisches Wörterbuch. Berlin Brandenburgische
Akademie der Wissenschaften. <http://aaew.bbaw.de/tla/index.html> (04.08.2014).
- „TEI: Frequently Asked Questions“. TEI Consortium. <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/TitlePageVerso.html> (04.08.2014).
- TEI Consortium, eds. 2014. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version
2.6.0., 20.01.2014. <http://www.tei-c.org/Guidelines/P5/> (04.08.2014).
- Walsh, John A. 2012. *Comic book markup language*. School of Library and Information Science,
Indiana University. <http://dcl.slis.indiana.edu/cbml/> (10.08.2014).
- Werning, Daniel A. 2013. *Datenkodierung in TEI XML im Rubensohn-Projekt (Arbeitsbericht)*.
11.03.2013. http://www.gwdg.de/~dwernin/drafts/Werning-TEI_im_Rubensohn_Projekt.pdf
(05.08.2014).