Milestone Report 2014-2016

Nikolai Grube¹, Christian Prager¹, Katja Diederichs¹, Sven Gronemeyer¹², Elisabeth Wagner¹, Maximilian Brodhun¹ & Franziska Diehr¹

¹ Rheinische Friedrich-Wilhelms-Universität, Bonn
² La Trobe University, Melbourne
³ Niedersächsische Staats- und Universitätsbibliothek, Göttingen

I. Project Description and Objectives

The Classic Maya hieroglyphic script, which consists of both logographic and syllabic signs, is one of the most significant writing traditions of the ancient world. As a graphic manifestation of language, writing mediates human thought, communication, and cultural knowledge in the form of texts. Deciphering a script allows ideas, values, conceptions, and beliefs to be reconstructed, and thus permits insight into the memory of past communities. In order to achieve this, the writing system and the spoken language that underlies it must be known. For Classic Mayan, this breakthrough in decipherment has already been achieved; however, in spite of great progress made in recent decades, some 40% of the script’s more than 800 signs remain unreadable even today. One reason for this situation is their lack of systematic attestation. Even in cases in which the signs are legible, texts may still elude understanding, because the Classic Mayan language itself has not survived; instead, it can only be reconstructed through comparison of the 30 Mayan languages documented since European conquest and still spoken today. However, much pre-Hispanic Mayan cultural vocabulary has been lost in the aftermath of European colonization. Consequently, comprehensive documentation and decipherment of the approximately 10,000 extant hieroglyphic texts, reconstruction of the language that they record, and documentation of that language in a dictionary are necessary prerequisites for acquiring a deeper understanding of Classic Maya culture, history, religion, and society.

The subject of the project is an incompletely deciphered, complex writing system, which the project aims to decipher with the aid of digital tools and to describe its underlying language in a dictionary. For these purposes, the hieroglyphic texts are being made machine-readable and saved in a text database with analysis and commentary. In addition, the Classic Mayan language is being represented in its original orthography in a web-based dictionary, allowing users to compare the content with its analysis. Until now, no project in the realm of digital writing systems research has demonstrated comparable standards, goals, and qualifications, or could serve as a model for conceptualizing and
developing our database. When developing their databases, research projects in Greek,\(^1\) Latin,\(^2\) or ancient Egyptian\(^3\) epigraphy are not faced with the same challenge of their respective writing systems and the corresponding languages being only partially or not at all deciphered. Our goal is to use the digital tools currently under development to compile and register newly classified signs in sign lists, make the texts machine-readable, discern readings, and document the vocabulary in its original orthography. The innovative character of our project requires flexible management when developing the digital infrastructure and deploying financial resources, as well as methodical ground work to develop working concepts that will be useful over the long term. It consequentially affects the project’s work, timeline, and personnel planning in the first two work packages. The project’s outcomes will ultimately include the development of new tools, methods, and standards in digital research on ancient writing systems and even the digital humanities as a whole, in addition to the content it produces about the Maya script.

The project’s emphases on digital epigraphy, database development, and long-term and interoperable storage of research data in particular underscore the great significance of the digital humanities for such an innovative undertaking. Yet the project will also contribute pioneering work to computer-based research on writing systems, and will develop methods and standards that will benefit other areas of research. To this end, it is co-operating with the Göttingen State and University Library (Niedersächsischen Staats- und Universitätsbibliothek; SUB), which represents the project's the informatics and information technological capacities. The project is assuming an interdisciplinary stance and operates at the intersection between applied informatics and the humanities. In this respect, the project does not merely provide an interface between these disciplines in the realm of digital epigraphy. In addition, it is also giving our research and informatics personnel interdisciplinary skills that they can integrate not only into their work in the project, but also into their teaching.

The collective tasks of the project include documenting existing research on Maya culture in a bibliographic database; collecting, archiving, and organizing hieroglyphic texts; and developing metadata models for describing and annotating the texts. Epigraphic and linguistic analysis of the texts, as well as collecting research data in a database of texts and text-bearing objects, will provide the basis for the aforementioned dictionary, which will comment on the texts and reference all existing bibliographic literature. The database will include original spellings, from which variants and developments in Classic Mayan vocabulary, grammar, and the script can be reconstructed.

Methods from the digital humanities are being employed, and in some cases innovated, in order to complete these tasks and to optimally and sustainably process the large quantity of data. The project is utilizing text mining to recognize textual patterns; identify signs, sequences of signs, or passages; test decipherments in their respective contexts of application; and propose new phonemic values. All of this work is ultimately directed towards the goal of generating a corpus linguistics record of the vocabulary of Classic Mayan in its temporal, spatial, and social dimensions.

The digital working environment is being implemented in the virtual research environment (VRE) TextGrid, where our adaptations and epigraphic tools operate under the designation Interdisciplinary Database of Classic Mayan (IDIOM). The research data will be stored long-term in the TextGrid Repository, and will be made freely available in their entirety via a web portal. In cooperation with the University and Regional Library of Bonn (ULB), select contents from the TextGrid Repository will be published in a virtual inscription archive in the ULB’s Digital Collections, including for instance images of the inscriptions, basic information about them, and analyses and translations.

---

1 http://stephanus.tlg.uci.edu/
2 http://www.thesaurus.badw.de/
3 http://aaew.bbaw.de/tla/
Text databases and dictionaries are not only being used to document Classic Mayan; in addition, they are also our tools for deciphering the Maya script. New phonemic values or proposed readings for logographic and syllabic signs are being identified and continuously entered into the digital sign catalog. They can be directly tested for feasibility using the texts available in the system. Furthermore, the architecture and implementation of the database and dictionary always take into account the current state of decipherment and thus represent the dynamics of Maya epigraphic research. The writing system is only partially deciphered and many texts cannot be completely read. As such, it is necessary to document all available information about the text-bearing objects, because the meaning of words is their use in context. For this reason, we are gathering information about each text-bearing object and its context (so-called non-textual information) and linking this information to the analysis of the script and language. Maya texts often concern the inscribed object, its position, and its commissioner. Consequently, the materiality and cultural and social context of a text-bearing object provide important insights. We are accounting for these complex relationships between sources of information when designing and building the VRE, and are thus paying particular attention to the VRE’s development and adaptation to the demands of our project.

2. Project Phase I (2014 - 2016) – Documenting and Recording Text-Bearing Objects

The foundation of the project’s textual analysis, which will be the central element of the work package from 2017 through 2028, is created by documenting and recording text-bearing objects. Consequently, the project began by designing and subsequently constructing its information technology infrastructure. When compiling the Classic Mayan dictionary or analyzing the meaning of words, texts should not be considered independent of the object on which they are recorded, nor of their temporal or spatial context. The object and its content provide non-textual information, or metadata, about the text-bearing object itself, its location, neighboring monuments and associated finds, commissioner, and historical context as a whole. These data are highly significant for deciphering and interpreting the inscriptions, and carefully documenting them in the database is a prerequisite for decipherment and text interpretation.

Another essential component of the object database, in addition to its link with the text database, is its capacity to establish and query relationships between multiple texts and text-bearing objects. This feature is significant because a text can extend across multiple text-bearing objects. Such work becomes possible and relations between people, events, and places can be made visible only after very detailed metadata have been recorded. However, the database contains more than just descriptions of the text-bearing objects and their textual contents. With the literature database, the user can access an overview of which authors have studied or published a monument, discussed a text passage, or first proposed to the public a linguistic reading of a hieroglyph or sign that remains valid to this day. By closely cross-linking these data, the project’s work on the text and object database will yield an additional benefit by producing the source material for the project’s ultimate goal: a Classic Mayan dictionary.

2.1. Constructing the Technical Environment for Recording Text-Bearing Objects

Close exchange between the disciplines represented in the project is necessary to create a capture environment that fulfills the scientific needs of the Bonn workplace. Mutually communicating knowledge (epigraphy, linguistics, and ancient American studies on the one hand and data modelling, vocabulary development, and informatics on the other), collaboratively defining and specifying demands, and developing information technology capacities are processes that enhance the transfer
of knowledge and information in the project and, above all, significantly contribute to reaching the desired goal.

2.1.1. Analyzing the Project’s Scientific Needs and Developing the Metadata Schema

The Göttingen team modeled the metadata schema for recording the text-bearing objects and historical context based on a collectively specified and professionally evaluated catalog of requirements. This development was an iterative process through which the technical requirements could be continuously redefined. In order to guarantee that our own conceptions were of high quality, the schema was constructed on the basis of established international standards. Reusing pre-existing and professionally recognized terms furthermore contributes to its high degree of interoperability with other data sets and information systems. The structure of resultant schema is ontologically cross-linked, representing complex relationships and correlations.4

2.1.2. Developing Controlled Vocabularies

A total of 10 multilingual thesauruses were developed through interdisciplinary collaboration to scientifically record the text-bearing objects. The Bonn team’s task was to research, select, and define scientifically appropriate terms, whereas development of the vocabularies was supported in Göttingen by the application of methods from documentation science.

In choosing appropriate terminology, all terms that had been previously employed in the literature were checked for plausibility, comparability, and utility. The resultant collection of terms was ordered according to terminological principles and modeled in the SKOS (Simple Knowledge Organization System) format,5 in order that they could be represented in machine-readable format and integrated into the metadata schema. The terms could also thus be simultaneously mapped onto the normed data from the Getty Thesaurus,6 which allows the reused terms to be referenced.

Developing the vocabularies is extremely relevant not only for the project’s own work, but also for the discipline more broadly. Until now, a multitude of terms, vocabularies, and descriptive schemas have existed in Maya epigraphy, resulting in a wide range of differentially documented text-bearing objects. The terminology being applied demonstrates relatively little agreement at times and is often incomplete, erroneous, imprecise, or dramatically simplified.

In developing these vocabularies, the project is making a significant contribution to standardization in Maya epigraphy, because it is reusing terms that are already established in scientific research, clearly defining them for the first time, and also situating them in a terminological relationship to one another.

The vocabularies will be published in both machine- and human-readable form on the project’s website by the end of 2016, according to the project’s timeline.

2.1.3. Developing and Designing the Technical Infrastructure

As part of the project’s work, data of various types are being created and stored. Image data, metadata, and text analysis files have to be managed in relation to one another in a single infrastructure with respect to their creation, storage, processing, and access regulation. The VRE TextGrid is being used

4 Complete documentation may be accessed at http://idiom-projekt.de/idiommask/schema.html
5 https://www.w3.org/TR/2009/REC-skos-reference-20090818/
6 http://www.getty.edu/research/tools/vocabularies/
for these tasks. The frontend TextGrid Laboratory (TG Lab) allows files to be created and processed, as well as fine-grained management of the rights thereto. The backend provides access to the repository TG Rep, which stores the data in a secure environment. The repository is made available by the GWDG and guarantees data storage using long-term digital preservation methods.

The densely networked structure of the files with the metadata for the text-bearing objects requires appropriate storage. This need is met by the format RDF (Resource Description Framework)\(^7\) and by storing them in a graph database in the form of a triplestore.

An entry mask is used to record the metadata in a user-friendly manner. This tool is HTML- and JavaScript-based and provides the user with multiple entry aids. When utilized as a plug-in for the TG Lab, the entry mask can be installed and directly used from the TG Lab. Examples of its supporting functions include searches in internal and external databases (to establish object relations), validating entry fields, and automatically converting data formats. Another important support in data entry is navigating within the vocabulary hierarchy, whereby the user does not initially have to search using concrete keywords. The triplestores for storing the metadata and vocabularies, the entry mask, and the project website are being made available on the project server.

2.2. Scientifically Recording Text- and Image-Bearing Objects and Their Historical Context

Since November 2015, the team in Bonn has been documenting text- and image-bearing objects using TG Lab and the entry mask. As part of this work, the research team, with support from the student assistants, conducted a systematic search for all available publications, databases, and websites of collections, museums, and other research institutions. These data form the foundation for recording the objects, a task which can be broadly divided into three components:

1. Documenting the artifacts (names, identifiers, object type and shape, measurements, state of preservation, and contexts of production, discovery, acquisition, and storage);
2. Archaeologically relevant places (with archaeological and geographic coordinates, normed and alternative place names) and their classification in a complex hierarchy of places (the place of an object’s discovery and its situation within the architectural context of a structure, the latter’s location in an architectural group in the site, and the site’s situation in the hierarchy of modern political and administrative regional authorities);
3. Situation in the relevant historical context: events and persons (e.g., wars, monument dedications, accession, ruler biographies, sociopolitical relations).

Recording is additionally supported by controlled vocabularies and references to normed data.\(^8\) Furthermore, the information can be supplemented with citations of sources that are supplied by the project’s bibliographic database in Zotero.

The ontologically networked data structure allows complex questions to be asked of the material, for example, which text-bearing objects are situated on Tikal’s Great Plaza? When were they dedicated, who commissioned them, and what historical events are mentioned there, for instance compared to those mentioned on the altars (text-object relation)?

The contents of the inscriptions from the site of Tikal and their associated persons and events have already been registered. Tikal is one of the largest and most significant Classic Maya cities. Tikal

\(^7\) [https://www.w3.org/TR/rdf-schema/](https://www.w3.org/TR/rdf-schema/)

(located in the Department of Petén, Guatemala) was selected as our initial test case for object recording, instead of the state of Campeche, Mexico as stated in the proposal. Because Tikal is well-documented in the existing literature, it is well-suited for testing the information technology environment (including the metadata schema, entry mask, and data querying) against the project’s research demands. The text-bearing objects from Campeche, in contrast, are scattered between multiple sites and have in some cases been only superficially described in publications.

We began documenting texts from the state of Campeche as per the proposal in the second half of 2016, and have recorded the metadata for the texts from the site of Calakmul. In addition, we have just started to record artifacts from the sites whose inscriptions have been documented in the over 20 fascicles published by the Corpus of Maya Hieroglyphic Inscriptions Project since 1975. The database currently includes 597 artifacts connected with 76 discovery events, 88 production events, 2 acquisition events, and 56 storage events, and 439 places, all in various stages of recording. We have also entered an additional 25 persons and groups (including researchers, curators, and museums) relevant to the artifacts’ research histories and provenience, and 302 individuals (mostly rulers) and 91 epigraphically attested events related to their historical contexts. The total number of entries in the database thus amounts to 1706.

2.3. Documenting and Digitalizing Research Materials

Texts and the objects on which they are recorded together constitute the project’s research subject. Thus, the text-bearing objects themselves must be researched as well. Because they are geographically dispersed and in many cases cannot be moved, they must be digitized in high quality in order for research to be conducted effectively and transparently. This approach facilitates data exchange and permits collaborative research from multiple sites.

In order to conduct epigraphic analysis in a VRE, the unit of text plus text-bearing object absolutely must be digitally available, and thus first digitalized if necessary. Unpublished research should be accounted for in the same manner. Descriptions or epigraphic notes, which are essentially “analog metadata” that can be re-used, are available for many archives. These materials also must be digitalized and can thus be made available for research for the first time. The project additionally is producing its own digital research materials.

The materials being documented originate from archives both in Germany and overseas (e.g., Ibero-American Institute in Berlin, Carnegie Institution of Washington via Artstor). The collections of Prof. Karl Herbert Mayer and Emeritus Prof. Berthold Riese provide particularly important private archives.

In parallel to this work, the Bonn team is gathering detailed information in working lists, which include a concordance of all existing sign catalogs and classifications; lists of sites, museums and collections, text-bearing objects, grammatical morphemes, and lemmata; and an inscription archive that comprises all text-bearing objects with comments about the inscription and chronology and is also kept on file in paper.

---

9 State: 1st July 2016
10 This list provides the foundation for generating the new sign catalog.
11 This list will serve as a guideline for transliterating, transcribing, and morphologically analyzing/glossing texts.
2.3.1. Digitizing the Archival Materials

Berthold Riese’s inscription archive consists of 135 binders containing photographs, drawings, and epigraphic notes; as of the time of writing, just about one-third of the materials have been digitized in over 14,000 files. At present, over 20,000 of the ca. 40,000 images (slides, negatives, prints) from Karl Herbert Mayer’s photographic archive have already been digitized as well. The project’s research materials are additionally supplemented by the team members’ private holdings, as well as donations from colleagues such as Dr. Daniel Graña-Behrens and Stephan Merk. The materials are being digitalized with scanners by undergraduate assistants in Bonn and supplied with metadata by a graduate assistant. We expect a total of over 70,000 documents to be scanned and integrated into the digital archive over the coming years.

2.3.2. Documenting and Digitalizing Inscriptions

Traditional methods, such as photography and drawing, as well as newer techniques like photogrammetry and 3D structured light scanning, are being used to document the inscriptions. Compared to existing imaging procedures and measuring techniques, documenting and measuring artifacts with a 3D scanner prove to be particularly advantageous for epigraphic research. 3D scanning permits thorough examination of the numerous eroded inscriptions that are no longer readable to the naked eye and cannot be rendered legible using photography and subsequent image processing. Detailed features can be more easily recognized with virtual manipulations (e.g., simulating lighting from various angles). In addition, fragmented text-bearing objects can be virtually reconstructed, and data acquired from application of earlier methods can be supplemented and refined. Documentation and measurement procedures also require notably less time for each individual object than do pre-existing techniques. Furthermore, 3D scanning contributes to archiving and storing existing cultural heritage in a form that is true to the original. 3D scanning is thus an essential component of the documentation trips to archaeological sites, museums, and archives that are undertaken by the project, as specified in the proposal.

The project places particular emphasis on collaboration with museums and collections whose inventories include artifacts with texts that can be scanned for epigraphic analysis and published as a 3D object. The 3D objects are made available to the museums after they have been processed, compiled, and rendered with software. As early as 2015, the project used these methods to document Maya artifacts from the exhibition “Relief Collections from Great Ages” [“Reliefsammlung der großen Epochen”] at the Knauf Museum in Iphofen, as well as the wooden lintels from Tikal held at the Museum of Cultures in Basel. The trips during which this documentation was conducted occupied 3 work days in 2014 and 68 work days in 2015, and are expected to occupy 46 work days in 2016.

2.3.3. Compiling the Bibliography

The free and open-source application Zotero is being used to collect, manage, and cite diverse online and offline sources, thereby documenting the existing scientific literature. This application supports editing and processing of bibliographic citations and lists, as well as collaborative work from multiple locations. The bibliographic database currently contains almost 16,000 entries and is being continuously expanded, with an anticipated eventual total of 70,000 entries. The contents also being regularly revised and tagged by theme.
3. Presentations and Publications

The project is working to promote timely, open, interdisciplinary, collaborative science. To the greatest extent possible, the project strives for completely open access to its scientific publications (“Open Access”), documentation of its methodology and work procedures (“Open Methodology”), as well as its research data (“Open Data”) and the software used by the project (“Open Source”), in accordance with “Open Science”. Unimpeded access to the project’s research, as well as a guarantee of productive re-use, must be achieved using free licenses.

For this reason, all data compiled in the context of the project are published online under the internationally valid Creative Commons copyright license CC BY 4.0 “Open Access”.12

The project strives to make accessible all digitalized images and drawings of Maya monuments from various publications for which the project has received or purchased the rights for worldwide publication. These images, together with annotated metadata, will be made available online in the TextGrid Repository and Portal, along with citations of the original sources and authorship. Digital images and publications whose rights are restricted will at a minimum be made available for research purposes within the TG Lab to a registered group of scientific users.

The individual “Open” strategies feed into or necessitate each other in the project’s work process. The project thus uses several different platforms on which methods, research results, data, and metadata developed over the course of the project can be shared, published, and made available for continued use.

3.1. Website, Social Media, and Portal

The project’s website,13 which was released in 2015 and is based on WordPress, serves as the project’s principal platform for online publications and guarantees rapid dissemination of its research results at no cost. The German National Library has registered our internet presence as a publication platform.14 Current work on lexicography, decipherment, and linguistics are published in the sections “Working Papers,” “Research Reports”, and “Project Reports”, along with working papers and concept papers about the use of tools from the digital humanities in epigraphy. In collaboration with the SUB, the project established a workflow for registering a unique digital object identifier (DOI) to ensure the long-term referencing and citeability of these digital objects. The DOI references the object itself, not the storage location (URL).

Another aspect of the website, in addition to special case studies, is rapid and broad dissemination of research data in the section “Documentation” (e.g., site list, sign concordance, 3D meshes), as well as announcements and general information about the project.

Furthermore, communication with users is critical: all research data that result from the working lists allow the potential for feedback that is intended above all for quality control purposes. An additional component is discussions about the online publications, which can be publically commented to foster scientific discourse.

---

12 http://creativecommons.org/licenses/by/4.0/
13 http://www.mayawoertebuch.de
14 ISSN 2366-5556
In addition to our website, social media channels play a significant role for the project. Channels on Facebook\(^\text{15}\) and Twitter\(^\text{16}\) went online at the same time as the website. Sketchfab\(^\text{17}\) offers not just a platform for 3D models, but also its own models, and other users’ models can be shared and linked as well.

Furthermore, beginning in the second quarter of 2018, the project intends to develop a portal that not only presents the data from TextGrid, but also enables fine-grained searches. The portal will be accessible via the website, but technologically independent from it. Eventually, all of the research data created in the project and current working versions of the dictionary will be made available on the portal. In addition, the portal will offer targeted search functions that permit complex queries of the TextGrid Repository containing all of the textual and non-textual metadata mentioned in Section 3.2.

3.2. Using and Adapting the ConedaKOR Image Database for Registering Digital Materials

An essential component of the project’s work is publishing the data it has recorded and compiled. In particular, the research material donated to the project that has been documented and digitalized and concerns Maya non-textual objects and materials (see Section 3.3.1) needs to be managed, archived long-term, and made publically accessible for research. For this task, the project will employ its own image database in ConedaKOR\(^\text{18}\) as an additional publication platform. In ConedaKOR, the digitalized archival materials, such as uninscribed utilitarian objects, artifacts, and architecture from the Classic Maya, are arranged and represented in relationship to one another. Beginning in late 2016, the project has been developing a metadata schema and testing the process of recording stored contents in its own ConedaKOR database, using selected digital images. The plan is to develop and represent in the given graph database a metadata model oriented not towards the digital images, but towards the represented entities (e.g., artifacts, places, people). The contents should be linked with metadata to bibliographic references, in addition to descriptions of the object’s history. For this, the project will again use its Zotero bibliography, which will be technically linked with the data structure. In the future, access to our ConedaKOR database in the DARIAH infrastructure should be possible via a web service.

3.3. The Inscription Archive in the Digital Collections of the ULB Bonn

In collaboration with the ULB Bonn, the project intends to develop an inscription archive for Classic Mayan in ULB’s Digital Collections.\(^\text{19}\) The archive will present the text-bearing objects documented by the project with their digital images and object-related metadata, as well as their inscriptions with transcriptions and translations.

Filling the archive with content depends on previously finished workpackages and milestones in the project: these concern the building of technical structures, which are necessary for the creation of content, which consecutively will be done therein. Hereto the following prerequisites have to be fulfilled:

1. Complete and documented metadata schemas for texts and text-bearing objects, as well as a productive recording environment;
2. Recording data created under real-world conditions;

\(^{15}\) https://www.facebook.com/idiom.project/
\(^{16}\) https://twitter.com/idiom_project/
\(^{17}\) https://sketchfab.com/idiom-project/
\(^{18}\) https://kor.uni-frankfurt.de/
\(^{19}\) http://digitale-sammlungen.ulb.uni-bonn.de/
3. Completing a mapping from the internal data format to the target format used by the Digital Collections (METS/MODS);
4. Publishing these data in TextGrid Rep so that they can be retrieved via TextGrid’s OAI-PMH interface.

Meaningful test data already has been created for the text-bearing objects, because the project fulfilled prerequisites 1, 2, and 3 by the end of 2015. Prerequisite 4 will not be fulfilled until the inscriptions have been processed with the annotation tools for text mark-up and analysis that will be developed in 2017-18, and have been inputted into the TextGrid Repository and can thus be imported into the inscription archive using the interface. With these test data, the company Semantics, the service provider for the ULB, will be able to prepare an initial projection of the inscription archive for the text-bearing objects. As soon as the technology for text mark-up and analysis have been developed in the second project phase, prerequisite 4 will have been fulfilled for both text-bearing objects and text, and complete test data can be delivered to Semantics. At that time, we expect to be able to implement the inscription archive in the ULB’s Digital Collections and to successively publish the inscriptions from Campeche, as well as the work that has already been done on Tikal.

3.4. Presentations, Public Relations, and Networking

From 2014 through 2016, the project has presented at numerous national and international events with papers, workshops, presentations, or posters (see overview in appendix).20

The goal of these activities has been to establish interdisciplinary ties to other projects in the digital humanities that have similar research questions and goals, as well as to present the project within its own discipline and to the general public. Of the 26 events, 12 were (co-)organized by the project, its backers, or its collaborating partners. Additionally, the Academy in Düsseldorf created an exhibition column created that has been set up at multiple open events and was always occupied by project members, for example during the opening press conference at the state parliament.21

3.5. Publications

Documentation of the development of the completed object metadata schema and its version-controlled source code are being made freely available on the web through Git, a service for software development projects.

Like the website, the digital documents are being regularly edited for compilation in a volume that is being released through print on demand. Thus, the project’s publications also are available to the public in libraries in print form and can be purchased worldwide online. The first of these yearbooks for the reporting year 2014-201522 contains English translations of the articles that had already appeared online, as well as selected website contents.

A complete list of publications for 2014-2016 can be found in the appendix.

---


The epigraphic workflow must be represented as exactly as possible in the VRE for the project’s corpus-oriented work. In this respect, the essential components of epigraphic work are: documentation of the text-bearing object; epigraphic analysis, including sign classification; transliteration and transcription; morphological segmentation and glossing; and linguistic interpretation. In addition to corpus analysis as the foundational work for lexicography, we are also studying the phonology, morphology, syntax, semantics, and pragmatics of Classic Mayan. Furthermore, our research agenda encompasses questions concerning historical linguistics, script history and use, as well as writing conventions, which we will address comparatively to the greatest extent possible. Use of this integrated approach in a VRE has already elicited positive reactions from within the scholarly community.23

In addition to the ongoing work from the first phase, the second phase of the project focuses on three work packages: the sign catalog, epigraphic mark-up of texts, and linguistic analysis.

4.1. Sign Catalog

A new, digital sign catalog will be compiled in TextGrid based on the working lists of pre-existing sign inventories and will re-use the entry mask programmed for non-textual objects. The catalog will represent a consolidation of and improvement on existing lists; moreover, it will be continually supplemented as previously uncatalogued signs or graphic variants are identified.

Beginning in the second half of 2015, the project has been modeling the metadata schema for the sign catalogue and establishing the criteria for re-classifying signs and their variants. This work will be completed by the third quarter of 2016 so that the sign catalog can be populated with actual data following completion of a test phase at the end of the project’s first phase.

The catalog’s ontological structure will differentiate between signs as information-bearing units and graphs. Each sign can receive multiple transliteration values, for instance to account for polyphonic signs. More importantly, the catalog will take into account alternative reading proposals that inevitably arise when dealing with a partially deciphered writing system. In order to assess these proposals, criteria tailored to logograms, syllabic signs, and diacritics will be established that will allow arguments for specific readings to be qualitatively defined and established in linguistic contexts. The criteria are linked using propositional logic,24 so that different permutations of the defined characteristics are automatically assigned a corresponding confidence rating in the catalog.

Comparative investigation of signs in the catalog and later in the context of the corpus can thus reference the reading confidences and contribute to confirming or rejecting existing hypotheses, or, ideally, facilitate new decipherments. Readings with the highest confidence ratings can be filtered and will provide the basis for the dictionary entries once the linguistic reading has been confirmed. All readings with lower plausibility ratings can be appropriately marked in the dictionary entries, as with different lemmas judged to be equal alternatives. Similarly, the semantic domain of undeciphered signs can be measured according to level of confidence whenever possible. Consequentially, the

---

23 Stephen Houston and Simon Martin, for instance, see great potential in the project due to its work “collecting and cross-referencing data.” (S. HOUSTON & S. MARTIN. 2016. Through seeing stones: Maya epigraphy as a mature discipline. Antiquity 90(350): 443-455.

24 The criteria thus adhere to Boolean algebra and can be true or false, for example, “can be found in the Landa manuscript”, “occurs as a homophonous substitution of a logogram”, “the semantic field is known”, etc.
dictionary will be the first that represents and accommodates the situation of a writing system that has only been partially deciphered.

Each sign is linked to at least one graph, which is associated with a standardized image as a reference and is named as a variant form. When designing the catalog, an innovative classification schema was developed that records sign variants using subgraphemic segmentation and variation. Concordances to existing catalogs are registered at the level of the grapheme, as is information about the graph-icon (by means of a controlled vocabulary) and its relations to the graphs of other signs. This process allows studies of the genesis of signs (icon – sound relationship).

4.2. Epigraphic Text Recording and Mark-Up

A text is linked with its original orthography with the Text Image Link Editor in TG Lab, which can mark-up areas in a digital image. Documenting the original orthography is fundamental to epigraphic work with syllabic and logo-syllabic hieroglyphic and cuneiform writing systems, because the original written form is no longer visible in transcriptions of a text. Mark-up is conducted on the level of the graph. From a linguistic perspective, the graph, as the smallest graphic unit in a writing system, has not yet been associated with a grapheme as a sign. In contrast, a grapheme, the smallest meaningful contrastive unit in a writing system, has been appropriately assigned. Each area identified as a graph is linked to the sign catalog, which is consequently a prerequisite and starting point for epigraphic text recording and mark-up (work package “Text Mark-Up I”). A graph number and transliteration value can thus be semi-automatically assigned to each marked-up area. Therefore, this value should be seen more as a sign name (pending further analysis), and it should also be able to indicate the sign class (traditionally indicated using capital/lower-case spellings).

Other operators for indicating grapho-tactics, which require further development, also exist for indicating the spatial relations of signs in a block. The basis of this mark-up that is created with the Text Image Link Editor is a metadata schema based on TEI and EpiDoc, which still needs to be developed. It will be able to register information about text structure, state of preservation (based on the Leiden System), and graph color.

The resultant text recording and mark-up is purely descriptive from an epigraphic perspective and thus makes the digital image machine-readable. The mark-up still does not include any linguistic analysis, but is nonetheless necessary for creating a digital corpus and the multi-step, analytical annotation that builds upon the corpus.

The mark-up criteria are currently being drafted and will be implemented beginning in 2017, along with the necessary adaptations in TextGrid.

4.3. Linguistic Text Annotation and Analysis

The work package “Text Mark-Up II” is being developed in parallel to the descriptive text mark-up, although it will be implemented later (expected to be productive in the fourth quarter of 2018). This work package technologically depends on, builds upon, and expands “Text Mark-Up I”. By the time it is completed, the prerequisites for complete analysis and compilation of the digital corpus will have

---

25 In some cases, the editor must decide which polyphonic phonetic value should be used. Alternative readings of the same sign function (logogram/syllabic sign) are automatically provided in order of decreasing plausibility.

26 For example, **CHAN** may appear in the contexts in 1) **CHAN-na** and 2) **CHAN-nu**, but in these cases can only be read as 1) *chan*, “snake” oder 2) *cha’an*, “guardian; captor”.
been fulfilled. The corpus will then be able to be mapped for use in the Digital Collections of the ULB and published on the project’s website.

Beginning in 2017, the project will, in collaboration with Dr. Cristina Vertan (Hiob Ludolf Centre for Ethiopian Studies, University of Hamburg), reuse a XML-based tool\textsuperscript{27} that Dr. Vertan has developed for the syllabic writing system Fidal and used for the ancient Ethiopian language Ge‘ez. The project will adapt this tool for epigraphic and linguistic analysis of Maya hieroglyphic texts and implement it in the VRE. In 2016, the tool’s technological potential was positively evaluated and the collaboration was arranged. The tool facilitates annotation with correction, as well as multi-level annotation. Additionally, it allows dictionary entries and lemmas to be created from the analysis.

Nonetheless, before these technological tasks can be completed, the multi-level analysis schema must be defined and the grammatical rules of Classic Maya must be compiled, the latter supported by the project’s working list of grammatical morphemes. Like the sign catalog, the annotation tool has to be so flexible in its design that it can take into account new research and insights into Classic Mayan grammar and morphosyntax and correct previous analyses.

The team in Bonn has been preparing the scientific prerequisites for this work since 2016, and these will be successively integrated into the programming beginning in 2017. In contrast to the pre-existing epigraphic tradition of having the transcription (in some cases with morphological segmentation) and translation to follow the transliteration, the analysis schema will comprise a total of eight steps. These steps will reflect in fine detail and with transparency the demands on machine-readability, and will also generally increase an analysis’ intelligibility. The analytic annotation will build upon the prior descriptive steps 1) classification and 2) graphematic transliteration from “Text Mark-up I”, and will consist of the following steps: 3) phonemic transliteration, 4) morphological transcription, 5) morphophonemic transcription, 6) morphological glossing, 7) consolidated transcription, and 8) translation.\textsuperscript{28}

4.4. Wörterbuch

According to the proposal, a continually updated working version of the dictionary was to be prepared at the same time that the digital inscription archive was being constructed. As was explained in Section 4.3, however, the technical foundation first had to be created, particularly for text analysis. Thus, lexicological functions will first be designed and developed in the annotation tool in parallel with “Text Mark-Up II”.

However, preparations have already been underway for compilation of the dictionary: from 2014–2015, student assistants digitalized 114 dictionaries and grammars from over 20 Mayan languages. Many of these have already been made searchable with OCR technology and will be gradually marked up to make Classic Mayan cognates and etyma findable in the annotation tool.

A concordance of various word lists was compiled with the working list of lemma in order to compare different authors’ transcriptions. Because of their significance for the history of research in the field, these diverse conventions will also be incorporated into the dictionary.

\textsuperscript{27} Created in the context of the EU research project TraCES: https://www.traces.uni-hamburg.de/

\textsuperscript{28} In the proposal, for instance, an example analysis in Appendix 11 shows a total of 19 analytical steps. These are still relevant, although the tools allow the analysis schema to subsume some of these points and include them elsewhere. An example of the analytical steps: 1) 74ost.683br:126bb > 2) \textit{SIH.ja:ya} > 3) \textit{siy-ya=ja} > 4) \textit{si(hly-aj} > 5) siy-aj-Ø > 6) ‘gift-INCH-3s.ABS’ > 7) siyaj > 8) “be born”.

13
List of Publications 2014 - 2016

Project-related publications by project members and external authors, as well as other publications in which project members are involved, include the following contributions (multilingual articles are marked with the relevant abbreviations DE, EN, and/or ES). (See also Section 3.5 “Publications”.)


27. BRODHUN, M. 2014. Mayan Database [RDF-Inputmask in CoffeeScript], Available at http://git.projects.gwdg.de/mayandatabase.git

List of Presentations 2014 - 2016

Tabular list of presentations. (See also Section 3.4 “Presentations, Public Relations, and Networking”.)

<table>
<thead>
<tr>
<th>Date</th>
<th>Presentation</th>
<th>Event (Location)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/2/2014</td>
<td>Presentation of the project, papers: 1) Ziele und Aufgaben des Wörterbuchprojektes IDIOM, 2) Konzept, Umsetzung und Forschungsstrategie des Wörterbuchprojektes IDIOM, 3) Paläographie und Ikonologie im Rahmen des Wörterbuchprojektes IDIOM</td>
<td>Conference: XVII. Mesoamerikanisten-Conference (Basel)</td>
</tr>
<tr>
<td>4-5/6/2014</td>
<td>Exhibition column</td>
<td>Event: Die Projekte der AWK-NRW im Landtag (Düsseldorf), organized by the AWK-NRW</td>
</tr>
<tr>
<td>14/9/2014</td>
<td>Exhibition column</td>
<td>Event: Tag des offenen Denkmals at the AWK-NRW (Düsseldorf), organized by the AWK-NRW</td>
</tr>
<tr>
<td>14/10/2014</td>
<td>Presentation: Text Database and Dictionary of Classic Mayan</td>
<td>Workshop: 1st Annual Workshop of the Text Database and Dictionary of Classic Mayan Project (Düsseldorf), organized by the project</td>
</tr>
<tr>
<td>Date</td>
<td>Event</td>
<td>Conference/Workshop</td>
</tr>
<tr>
<td>------------</td>
<td>------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>11/5/2015</td>
<td>Exhibition column</td>
<td>Event: Akademientag 2015 (Berlin), organized by the Union of German Academies of Sciences and Humanities</td>
</tr>
<tr>
<td>5/6/2015</td>
<td>Paper: Textdatenbank und Wörterbuch des Klassischen Maya (presentation of project)</td>
<td>Workshop: Abschluss Verbundprojekt MayaArch3D (Bonn)</td>
</tr>
<tr>
<td>21/7/2015</td>
<td>Paper: Textdatenbank und Wörterbuch des Klassischen Maya (Projektvorstellung)</td>
<td>Event: Digital Humanities (Bonn), organized by Jan Kenter, ULB</td>
</tr>
<tr>
<td>11/9/2015</td>
<td>Exhibition column</td>
<td>Event: Tag des offenen Denkmals at the AWK-NRW (Düsseldorf), organized by the AWK-NRW</td>
</tr>
<tr>
<td>16/9/2015</td>
<td>Paper: Dokumentation, Analyse und Edition der Hieroglyphentexte der Klassischen Maya in der Virtuellen Forschungsumgebung TextGrid</td>
<td>Workshop: Historische Semantik und Semantic Web (Heidelberg), organized by the Union of German Academies of Sciences and Humanities</td>
</tr>
<tr>
<td>2/10/2015</td>
<td>Paper: Digitale Epigraphik am Beispiel des Wörterbuchs des Klassischen Maya</td>
<td>Workshop: Digital Humanities Bilder (Bonn)</td>
</tr>
<tr>
<td>5/10/2015</td>
<td>Presentation: Maya Hieroglyphic Writing</td>
<td>Workshop: Digitale Epigraphik: XML/TEI und EpiDoc für die epigraphische Forschung zu nicht-alphabetischen Schriftsystemen (Bonn), organized by the project</td>
</tr>
<tr>
<td>14/12/2015</td>
<td>Presentation: The Text Database and Dictionary of Classic Mayan Project</td>
<td>Workshop: 2nd Annual Workshop of the Text Database and Dictionary of Classic Mayan Project (Bonn), organized by the project</td>
</tr>
<tr>
<td>Date</td>
<td>Event/Conference</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>----------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>15/1/2016</td>
<td>Paper: <em>Digitale Epigraphik - Die Erforschung der Hieroglyphentexte und Bildbotschaften der Maya in der Virtuellen Forschungsumgebung TextGrid</em></td>
<td>Conference: XIX. Mesoamerikanisten- Conference (Berlin)</td>
</tr>
<tr>
<td>18/5/2016</td>
<td>Exhibition column</td>
<td>Event: Akademientag 2016 (Hamburg), organized by the Union of German Academies of Sciences and Humanities</td>
</tr>
<tr>
<td>17/11/2016</td>
<td>Edificios como reliquias ancestrales. Un caso de Copán, Honduras</td>
<td>Conference: 1o Congreso Internacional de Arquitectura e Iconografía Precolumbina (Valencia)</td>
</tr>
<tr>
<td>13/12/2016</td>
<td>Of Codes, Glyphs and Kings: Tasks, Limits and Approaches in the Encoding of Classic Maya Hieroglyphic Inscriptions</td>
<td>Seminar: Digital Classicist: Digitale Methoden in den Altertumswissenschaften (Berlin)</td>
</tr>
</tbody>
</table>

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/