

Textdatenbank und Wörterbuch des Klassischen Maya

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



ISSN 2366-5556

PROJECT REPORT 3 | EN



Published 29 Mar 2016

DOI: 10.20376/IDIOM-23665556.16.pr003.en

Annual Report for 2015

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

¹ Rheinische Friedrich-Wilhelms-Universität, Bonn

² La Trobe University, Melbourne

³ Niedersächsische Staats- und Universitätsbibliothek, Göttingen

Project Description, Goals, and Methods of IDIOM “Interdisciplinary Database of Classic Mayan”

The only partially deciphered Maya hieroglyphic script and language constitute the primary focus of this research project, which has been carried out cooperatively by the University of Bonn and the Göttingen State and University Library (Niedersächsische Staats- und Universitätsbibliothek Göttingen, SUB)¹ since mid-2014 and funded by the North Rhine-Westphalian Academy of Sciences, Humanities and the Arts. The project’s goal is to compile a text database and, on this basis, a dictionary of Classic Mayan, the language of a civilization which reached its florescence between A.D. 250 and 950. When the project was inaugurated in 2014, the approximately 10,000 extant text and image carriers that permit a detailed and precise investigation of the Classic Mayan literary language had yet to be systematically documented or comprehensively analyzed. In the context of the project, the text and image carriers will be systematically described according to uniform standards. This source material will be made machine-readable with XML (Extensible Markup Language) and integrated into the Virtual Research Environment (VRE) of the research association TextGrid², thereby creating the foundation for compiling the dictionary. The project draws on methods and technologies from the Digital Humanities that are already available in the VRE or are being developed and implemented in the context of the project. To this end, we are adapting digital tools and services provided by TextGrid to the project’s needs, and these project-specific adjustments are summarized under the acronym IDIOM “Interdisciplinary Database of Classic Mayan”. IDIOM permits systematic and networked analysis of Classic Maya text, image, and information carriers using the technology of TextGrid. The structure and functions of IDIOM thus orient themselves toward the epigraphic workflow “documentation – description – analysis – publication”, thereby supporting our work with the inscriptions. The project is taking into account and collecting descriptive data, or metadata, about the text and image carriers in

¹ <http://www.sub.uni-goettingen.de/>

² <https://textgrid.de/>

the VRE, in addition to data about the contents of the inscriptions. The VRE not only contains descriptions of the text carriers and information about the texts; in addition, with the help of the bibliographic database, it also provides the database user an overview of which authors have studied or published a monument, discussed a text passage, or were the first to propose a reading for a hieroglyph that is still valid today. The text carrier thereby receives a “biography” that is intimately tied to the text contents and is taken into account when analyzing the meaning of words.

The focus of the 2015 phase of the project was designing and implementing a metadata schema developed by the experts in Bonn and Göttingen for compiling non-textual information about the text carriers themselves, and which can serve as a model for others. Contextual information about the text carrier as an object must be compiled for epigraphic analysis, because the physical qualities, associated archaeological finds, or context of discovery may be relevant to the text’s contents and presentation. The processes of modeling an object biography and embedding texts in their cultural context must initially be addressed from an information technology perspective in order to guarantee that the texts’ contents are epigraphically analyzed in their extra-linguistic, cultural context. The project is reusing and adapting CIDOC CRM (CIDOC Conceptual Reference Model)³ to its needs so that it serves the basis for describing artifacts and relevant events in their individual histories, such as their production, discovery, acquisition, and storage, as well as actors associated with the artifact. CIDOC CRM offers an expandable ontology for terms and information in the field of cultural heritage. The resultant application profile is being augmented with elements from other standards and schemata. The reused schemata and built metadata will be stored within a Linked (Open) Data structure (LOD)⁴. LOD is based on a standard model for data interchange on the Web, RDF (Resource Description Framework)⁵. For this type of data, logical statements about resources are saved in the linguistically familiar form subject-predicate-object (for example, Teobert Maler – discovered – Stela A), which means, the storage of relationships between objects and their properties is an important design issue here. Thus, very complex queries may be configured according to the database with the standard SPARQL⁶, a graph-based query language for RDF. The representation in XML format and use of HTTP-URIs facilitates the data’s basic compatibility and citability. This conversion allows object metadata to be stored separately from and to be linked by means of the HTTP-URI to the recorded text. The RDF objects are compiled using an RDF-input mask that has been specially adapted to the project’s needs. With the help of an HTML form, the entry mask allows the project members in Bonn to quickly and easily compile metadata concerning the text carriers, as well as information about actors or places. A detailed description is provided below.

In 2015, the project devoted much attention to categorizing and schematizing the contents, a necessary step in light of the diversity and heterogeneity of the data. In addition, development efforts can be reduced by reusing already existing standards. Controlled vocabularies are defined and structured collections of terms that are applied according to special rules to reduce the ambiguity of natural language. They comprise defined terms and fixed relations according to specific and ontological concepts, and thus allow for conceptual control of information. Consistent indexation, improved queries, and precision in both search and output facilitate content-based analysis. The heterogeneous data is categorized with controlled vocabularies. Development of controlled vocabularies to support markup of textual and non-textual objects is critical to data entry and analysis. In this respect, controlled vocabularies are an essential aid when using search- and analysis tools. They are being structured by SKOS⁷ (Simple Knowledge Organisation System), so that they can be represented in

³ <http://www.cidoc-crm.org/>

⁴ <https://www.w3.org/standards/semanticweb/data>

⁵ <https://www.w3.org/RDF/>

⁶ <https://www.w3.org/TR/rdf-sparql-query/>

⁷ <https://www.w3.org/2004/02/skos/>

machine-readable format and integrated into a metadata schema. Terminological controls are created by defining and marking up preferred, alternative, or non-valid terms; the latter allows, for example, the correct output to be obtained even if a misspelled term is entered. Language tags also permit searches in a specific language. Control is achieved by creating concept schemas or knowledge organization systems, which are hierarchically organized into “broader” and “narrower” systems or are placed in a synonymous, antonymic, or causal relation with one another.

Developing controlled vocabularies and conceptual schemata is essential for the project. Researchers studying the Maya script have used numerous various terms, vocabularies, and descriptive schemas to document text carriers and the texts themselves. As a result, there exists a wide range of differentially documented text carriers. The concepts employed for documentation demonstrate little overlap in some respects and are often incomplete, erroneous, imprecise, or highly simplified. Because of inconsistencies that arise from a lack of standardized terminology in Maya studies, it is often difficult, if not impossible, to compare text carriers documented according to different standards. In order to develop the controlled vocabularies, the project is compiling all terms and taxonomies that have been used in the literature or in standard vocabularies, evaluating them with respect to their plausibility, comparability, and usability, and integrating them into the vocabularies.

Workplace Members, Tasks, and Communication in the Project

As in 2014, the three researchers Dr. Christian Prager (project coordination), Elisabeth Wagner, M.A., and Dr. Sven Gronemeyer serve as experts in Bonn on Maya writing, language, and iconography. Katja Diederichs is responsible for the information technology in Bonn. Together with Jan Kenter of the Bonn University Library (ULB), she coordinates among other remits the project’s collaboration with the ULB. Over the course of 2017, the project will be making the virtual inscription archive available online on the presentation platform Visual Library, with its data originating from the TextGrid repository. Preparations for this move were initiated as part of the development of the metadata schema. Together with the Göttingen team the mapping of the project internal metadata schema to the data structure of the Visual Library has been conceptualized, to enable data exchange between the two different systems. Franziska Diehr continued the work that she had initiated in 2014 and developed the project internal metadata schema on the basis of the scientific guidelines established in Bonn. Maximilian Brodhun, in turn, programmed and implemented the epigraphic workflow based on the metadata schema developed by Franziska Diehr, and he completed the first working version of an entry mask for RDF objects before the end of the year. The entry mask allows the Bonn scientists to easily and straight-forwardly enter the metadata. It is based on an HTML formula with which the data are converted into RDF/XML and stored in TextGrid. The entry mask was tested in November and December, and the first results were presented on December 14 during the project’s public workshop.

In 2015, the researchers in Bonn were supported in their work by the research assistants Laura Burzywoda (digitalization), Nicoletta Chanis (digitalization), Antje Grothe (bibliography), Laura Heise (digitalization), Leonie Heine (digitalization), Jana Karsch (epigraphy, digitalization), Nikolai Kiel (epigraphy, digitalization), Catherine Letcher Lazo (website editing), Lisa Mannhardt (digitalization), Mallory Matsumoto (linguistics, digitalization), and Nadine Müller (digitalization). The more than 100,000 photos from the archives that Karl Herbert Mayer, Berthold Riese, Daniel Graña-Behrens, and current project members have made available to the project are being digitalized in association with the VRE TextGrid, into which digital versions of the photographs will be integrated in 2016 and 2017.

As in 2014, written collaboration between Bonn and Göttingen is conducted by means of a jointly edited Wiki that employs the collaborative software Confluence, which is a component of the research

infrastructure for digital work in the humanities from DARIAH-DE⁸. The Wiki contains the project's complete documentation with the technical and scientific contents that have been aggregated up to this point. These will be available as reference materials over the long term. Documents such as lectures, presentations, or working lists are additionally archived in Google cloud storage and can thus be collaboratively edited. This approach has proven to be particularly productive for preparing joint presentations. For planning monthly goals and intermediate-term milestones, or to address urgent and short-term inquiries and problems, project members also participate in weekly teleconferences and monthly project meetings in Bonn or Göttingen.

The 8th project meeting on January 15-16, 2015 primarily consisted of a detailed discussion of the metadata schema for non-textual objects and the use of XML/TEI and EpiDoc for marking up textual data. The 9th meeting took place on February 9-10 in Bonn and Düsseldorf and also addressed questions about the metadata schema, the status of the RDF entry mask, and the creation of a conference poster for the DH Summit in Berlin on March 3-4. On February 11, the project members participated in a digital humanities workshop in Düsseldorf, which was organized by the North Rhine-Westphalian Academy of Sciences, Humanities and the Arts. The topic of the project's two-day workshop in Göttingen on April 27-29, 2015 was the Maya calendar and its functions and representations in the inscriptions. The Bonn project members explained to their Göttingen colleagues the functions of the Maya calendar, in order that these functions could be integrated into the VRE. Particular points of focus were the elements and structure of the Maya calendrical systems, the individual components of the calendar, and the conventions used to Maya data, as illustrated with examples. The project aims to implement the functions and structure of the Maya calendar in TextGrid and to use this work to develop a tool to calculate Maya dates that could be utilized for textual analysis. Maximilian Brodhun presented an initial version of the entry mask for non-textual RDF objects at the 10th project meeting in Bonn on May 18-19, 2015. The topics of discussion at this meeting were the organization of the objects in TextGrid, the embedding of the bibliographic data from the Zotero database, and the design of the project website, which finally went live in December 2015. Text analysis with XML/TEI and EpiDoc was discussed at the 11th project meeting in Göttingen on June 23-24. At this event, a new version of the entry mask and help texts for data entry were presented, and members jointly discussed plans for testing the entry mask in October and November 2015. At the 12th project meeting in Bonn on August 17-18, the Bonn experts hosted a workshop on the function and use of the vocabulary tools and the entry mask for compiling non-textual metadata. The entry mask was extensively tested by the Team in Bonn after this meeting, and it was made fully operational in November 2015. Franziska Diehr and Maximilian Brodhun transferred these data to Jan Kenter (ULB) in December 2015, following meetings with him in early 2015 about the project's ongoing collaboration with the ULB and a mapping of the data for the Visual Library planned for autumn 2015.

In addition to this work, the project continued to support the ongoing professionalization of its members in 2015. Members attended national and international conferences and courses focused on the digital humanities in Essen, Graz, London, and Lyon. These experiences introduced them to methods from the digital humanities and enhanced their knowledge of the XML markup language. In addition, the project organized a workshop on October 5-6 that addressed the application of XML/TEI for marking up texts written in complex writing systems, including Egyptian hieroglyphs, cuneiform, and Maya hieroglyphic writing. At this workshop, the project members responsible for the computer science aspects of implementing the research questions obtained an overview of the functions and structures of complex writing systems, including that of the Classic Maya.

⁸ <https://de.dariah.eu/>

Presentation of the Project

In 2015, the project was introduced in the context of presentations at scientific conferences and workshops, as well as in publications online and in print.

Presentations

In 2015, the project presented its work at 13 national and international conferences. The goal of these activities was not only to introduce the project, but also to network with other digital humanities projects pursuing similar research goals. As part of the SUB (Niedersächsische Staats- und Universitätsbibliothek Göttingen, Göttingen State and University Library) and DARIAH-DE panel “The Knowledge of Images: Varieties of Digital Annotation” [Das Wissen der Bilder. Spielarten des digitalen Annotierens], the project presented together with Frauke Sachse and Michael Dürr a workshop report entitled “Digital Epigraphy – Researching the Hieroglyphic Texts and Pictorial Messages of the Maya in the Virtual Research Environment TextGrid” [Digitale Epigraphik – Die Erforschung der Hieroglyphentexte und Bildbotschaften der Maya in der Virtuellen Forschungsumgebung TextGrid]. This report addressed the subject of “Text – Image – Inscription: Annotating the Hieroglyphic Writing and Languages of the Maya” [Text - Bild - Inschrift: Hieroglyphenschrift und Sprachen der Maya annotieren].

At the DH Summit in Berlin on March 3-5, the project presented a poster and contributed to the poster slam. The event was particularly oriented towards the DH project networks funded by the Federal Ministry of Education and Research [Bundesministerium für Bildung und Forschung (BMBF)], two DH centers, and junior teams, as well as associated project partners of *TextGrid* and DARIAH-DE. As part of the workshop “Semantic Web Applications in the Humanities” organized by the Digital Humanities Research Collaboration [Digital Humanities Forschungsverbund (DHFV)] on March 10 in Göttingen, the project presented its metadata concept and the technical foundations underlying its description of the inscriptions and analysis of the hieroglyphic texts. In her presentation on April 22 at the annual EpiDoc workshop in London, Katja Diederichs addressed the markup of hieroglyphic texts using EpiDoc. The EpiDoc guidelines describe standards for the structured Markup and representation of epigraphic texts in digital form using XML/TEI. Christian Prager presented on copyright and open access in the digital humanities on May 11 in Göttingen in a talk entitled “Open Access vs. Copyright in the Project ‘Text Database and Dictionary of Classic Mayan’”. This conference, which was entitled “Store it, Share it, Use it” and was organized by DARIAH-DE, SUB, and DAI (German Archaeological Institute) [Deutsches Archäologisches Institut]⁹, focused on licensing digital research data in linguistics and literary studies.

Nikolai Grube, Sven Gronemeyer, and Elisabeth Wagner represented the project with a presentation and a booth at the Academies’ Day 2015 event on May 11 in Berlin. Nikolai Grube, Katja Diederichs, and Elisabeth Wagner also presented the project at the closing workshop of the collaborative research project MayaArch3D. At this meeting, the project addressed controlled vocabularies, the metadata schema, and its Open Science strategy. On behalf of the project, Katja Diederichs and Christian Prager presented a talk entitled “Text Database and Dictionary of Classic Mayan” on July 21 in the ULB as part of a digital humanities information and networking event held by Jan Kenter in the ULB. The goal of these regular events is to establish links between various digital humanities projects in Bonn.

Christian Prager represented the project in Heidelberg on September 16, giving the paper “Documentation, Analysis, and Editing of Classic Maya Hieroglyphic Texts in the Virtual Research Environment TextGrid” at the workshop “Historical Semantics and Semantic Web.” This event, which

⁹ <https://www.dainst.org/>

was organized by the Union of the German Academies of Sciences and Humanities, focused on electronic and digital publishing. This same subject was the focus of another conference in Göttingen on September 29 that the German Initiative for Network Information [Deutsche Initiative für Netzwerkinformation (DINI)] and DARIAH-DE organized jointly. At this event, Christian Prager and Katja Diederichs presented the project to representatives from various libraries and media and computing centers. The goal of this workshop was to highlight the potential that the DARIAH-DE infrastructure and services offer for digital work. As part of the workshop “Digital Humanities Images” organized by the Max Weber Foundation, Christian Prager introduced the project on October 2 in a talk entitled “Digital Epigraphy as Exemplified by the Dictionary of Classic Mayan.” Franziska Diehr represented the project at another conference in Wolfenbüttel on November 2. This event, “Digital Metamorphosis: Digital Humanities and Editorial Theory,” was hosted by the Herzog August Library in Wolfenbüttel. Project members also delivered a talks in Bonn at its own event on October 5 (presentation title: “Maya Script”, see Workshops), at the 20th European Maya Conference on December 13 (“Digital Epigraphy – The Text Database and Dictionary of Classic Mayan Project”), and at the project’s annual workshop on December 14.

Conferences at which the project intends to present its work in 2016 include the 18th Conference of Mesoamericanists in Berlin and the EAGLE 2016 conference on digital and traditional epigraphy. Other presentations about the project and the current status of its work are planned among others for conferences in Basel (March), Mainz (April) and Cambridge (July).

Workshops

The project organized two workshops in 2015, which took place on October 5-6 and December 14-15, respectively. On October 5-6, the project hosted an international workshop entitled “Digital Epigraphy: XML/TEI and EpiDoc for Epigraphic Research on Non-Alphabetic Writing Systems.” The focus of the workshop, which was attended by over 30 participants, was using the markup language XML (Extensible Markup Language), and particularly XML/TEI and EpiDoc, for epigraphic research. TEI is a standard for coding printed works or marking up linguistic information in texts. Special guidelines for structured markup of epigraphic documents in XML/TEI are summarized in EpiDoc. The Bonn research project will use these guidelines to digitally compile Maya inscriptions and, on this basis, compile a dictionary of Classic Mayan. Until now, EpiDoc has only been used to digitally code and annotate texts written in alphabetic scripts.

The Bonn workshop focused on the question of whether texts in syllabic and logo-syllabic scripts could also be documented and analyzed with XML/TEI and EpiDoc. The discussion also addressed the question of whether existing standards may be applied to linguistically annotate hieroglyphic and cuneiform writing systems, some of which are only partially or barely deciphered. The project invited both domestic and international experts on writing systems and IT specialists to the English-language event. In order to incorporate a wide range of examples into the discussion, the invitees included writing systems experts who specialize in hieroglyphic Luwian (Annick Payne, Basel), Aegean scripts (Miguel Valerio, Valencia), hieroglyphic Egyptian (Daniel Werning, Berlin), cuneiform writing systems (Hubert Mara, Heidelberg), and Aztec hieroglyphs (Gordon Whittaker, Göttingen). As a scholar experienced in the digital markup of inscriptions, Thomas Kollatz (Essen) also reported on his use of XML/TEI in his epigraphic research, documentation, and editing of Jewish epitaphs. In addition, Gabriel Bodard (London), an expert on Greek and Latin texts on stone and papyrus who has been involved in the development of EpiDoc, discussed guidelines for epigraphic and digital annotation, referencing existing research projects that have employed EpiDoc.

On the first day of the workshop, participants presented on the graphemics and graphetics of the various writing systems and their respective states of decipherment, their transliteration and transcription conventions, their states of digital investigation, and their idiosyncrasies and special characteristics. More specifically, the presentations addressed the following topics and research questions: 1a) graphemics and graphotactics of the respective writing systems: which sign types are present and which general principles operate in complex writing systems (e.g., affixation, infixation, ligatures, superposition, elision and abbreviation, diacritics, phonetic complementation, semantic markup, etc.). Do projects already exist in which the markup of these scribal practices has been coded? 1b) Grapheme inventory: sign classification and sign catalogs: what meaning is being ascribed to sign classification in epigraphic research? Where are they used and what role do they play in transliterating and transcribing texts? 2) Graphetics (script use: formal arrangement of linguistic elements that do not differ in meaning, the configuration of texts). Allography and orthographic variants: what meaning does this orthographic principle have for the writing system in question—is it being accounted for when annotating and analyzing texts, and if so, how? What is the reading order of signs in a text (linear, columnar spellings)? Are there word separators, or how can independent lemmas be identified? What graphic aids do the scribes use for this purpose? Are there deviations in reading order—and are they perhaps even systematic? Is there an association between image and text—how are texts integrated into images or vice versa? How are text fields arranged, and what influence does the size of the text field or the type of writing material have on the graphemics? 3) State of decipherment and readability of the texts: what is the state of decipherment for the script in question? How do researchers treat individual, undeciphered signs; are they being identified as such in the transliteration and transcription, and if so, according to what rules? How does one deal with hypothetical decipherments of individual signs or text passages? Are (physical) lacunae in the text annotated and commented and accordingly marked when editing? These problems, research questions, and challenges considered on the first day were then discussed in a round table on the second day.

One of the main goals of a structured markup of syllabic and logo-syllabic writing systems is to represent the structure of the original spelling and the order of the signs in the text as accurately as possible using XML/TEI for queries. In the case of partially deciphered or undeciphered scripts, another issue is that undeciphered signs or passages must be reproduced with codes and nomenclature, instead of transliterating and transcribing them. This procedure is necessary for documenting unclear signs requiring further discussion in their original contexts of use. Basically, the texts marked up with XML/TEI should be able to function as the basis for (corpus) linguistic analysis, in addition to epigraphic analysis. Yet documentation of the original spellings is also fundamental to epigraphic study of syllabic and logo-syllabic hieroglyphic and cuneiform writing systems. Many epigraphic documentation and database projects use transliterations and transcriptions from which the original spelling can no longer be discerned; consequentially, systematic graphemic and graphetic investigations are not possible. Text markup should also facilitate analysis of the inscriptions, including their linguistic investigation. Our project has identified a desideratum here and wanted to discuss possible XML/TEI- and EpiDoc based solutions in collaboration with experts on writing systems. An important result of this event was the establishment of the research group EnCoWs "Encoding Complex Writing Systems". The goal of this collective of writing systems specialists, linguists, and experts in the digital humanities is to work over the coming months to develop common standards for marking up texts composed with a complex writing system. Communication among members of this group is facilitated by a Google Group account, to which 41 writing systems specialists currently belong. This Wiki allows documents to be collectively edited, discussions to be conducted, and materials to be exchanged. The results of this discussion group will be incorporated into the project's current and future work, with the goal of defining the standards formulated there and recording them in a joint working paper for publication.

The project's annual conference took place on December 14-15 in Bonn following the 20th European Maya Conference. The Bonn project contributed financial support and personnel to this latter event, Europe's largest conference on Maya civilization. It also took the opportunity to present its work to conference participants in a follow-up workshop, in addition to a presentation at the European Maya Conference itself. The annual workshop, a follow-up on the kick-off event in Düsseldorf on October 14, 2014, generally provides the project with a forum for reporting on the state of its current and completed project activities. Standards and conventions for transliterating, transcribing, and analyzing Maya texts were discussed and established at the subsequent workshop with members of the scientific advisory board. The results of this workshop will be published on the project website in a working paper.

Publications

1. DIEDERICH, K. 2015. Unsere „Open Science“-Strategie. Electronic Document *Textdatenbank und Wörterbuch des Klassischen Maya Working Paper 1*. <http://mayawoerterbuch.de/?p=4334>. DE | EN
2. DIEDERICH, K., C.M. PRAGER., E. WAGNER., S. GRONEMEYER. & N. GRUBE. 2015. Textdatenbank und Wörterbuch des Klassischen Maya. Electronic Document *Wikipedia*. https://de.wikipedia.org/wiki/Textdatenbank_und_Wörterbuch_des_Klassischen_Maya.
3. DIEDERICH, K., C.M. PRAGER., E. WAGNER., S. GRONEMEYER. & N. GRUBE. 2015. Text Database and Dictionary of Classic Mayan. Electronic Document *Wikipedia*. https://en.wikipedia.org/wiki/Text_Database_and_Dictionary_of_Classic_Mayan.
4. GRONEMEYER, S. Class Struggle: Towards a Better Understanding of Maya Writing Using Comparative Graphematics, in H. Kettunen & C. Helmke (ed.) *On Methods: How We Know What We Think We Know about the Maya. Proceedings of the 17th European Maya Conference, 2012* (Acta Mesoamericana 28): 101–17. Markt Schwaben: Verlag Anton Saurwein.
5. GRONEMEYER, S., C.M. PRAGER., E. WAGNER., K. DIEDERICH., N. GRUBE., F. DIEHR. & M. BRODHUN. 2015. Website. Electronic Document *Textdatenbank und Wörterbuch des Klassischen Maya*. <http://mayawoerterbuch.de/>. DE | EN | ES
6. GRONEMEYER, S., C.M. PRAGER. & E. WAGNER. 2015. Evaluating the Digital Documentation Process from 3D Scan to Drawing. Electronic Document *Textdatenbank und Wörterbuch des Klassischen Maya Working Paper 2*. <http://mayawoerterbuch.de/?p=3405>.
7. GRUBE, N. & C.M. PRAGER. 2015. Textdatenbank und Wörterbuch des Klassischen Maya, in N.-W. Akademie der Wissenschaften und der Künste (ed.) *Jahrbuch 2015*: 160–64. Düsseldorf: Nordrhein-Westfälische Akademie der Wissenschaften und der Künste.
8. GRUBE, N., C.M. PRAGER., K. DIEDERICH., S. GRONEMEYER., E. WAGNER., M. BRODHUN., F. DIEHR. & P. MAIER. 2015. Jahresabschlussbericht 2014. Electronic Document *Textdatenbank und Wörterbuch des Klassischen Maya Project Report 2*. <http://mayawoerterbuch.de/?p=4477>. DE | EN | ES
9. MAIER, P. 2015. Ein TEI-Metadatenchema für die Auszeichnung des Klassischen Maya. Electronic Document *Textdatenbank und Wörterbuch des Klassischen Maya Working Paper 3*. <http://mayawoerterbuch.de/?p=5067>. DE | EN
10. PRAGER, C.M. 2015. Das Textdatenbank- und Wörterbuchprojekt des Klassischen Maya: Möglichkeiten und Herausforderungen digitaler Epigraphik, in H. Neuroth, A. Rapp, & S.

- Söring (ed.) *TextGrid: Von der Community – für die Community: Eine Virtuelle Forschungsumgebung für die Geisteswissenschaften*: 105–24. Glückstadt: Werner Hülsbusch.
11. PRAGER, C.M., S. GRONEMEYER. & E. WAGNER. 2015. A Ceramic Vessel of Unknown Provenance in Bonn. Electronic Document *Textdatenbank und Wörterbuch des Klassischen Maya Research Note 1*. <http://mayawoerterbuch.de/?p=2221>.
 12. WAGNER, E., S. GRONEMEYER. & C.M. PRAGER. 2015. Tz'atz' Nah, a "New" Term in the Classic Mayan Lexicon. Electronic Document *Textdatenbank und Wörterbuch des Klassischen Maya Research Note 2*. <http://mayawoerterbuch.de/?p=3827>.

Scientific Advisory Board, Collaboration, and Networking

The scientific advisory board is composed of Prof. David Stuart (Austin), Prof. Marc Zender (New Orleans), Prof. em. Peter Mathews (Melbourne), and Prof. Gordon Whittaker (Göttingen). The board met in Bonn on December 15, and another meeting is planned for November 2016 as part of a conference in Bonn or Düsseldorf on Classic Mayan lexicography.

The project participated in a workshop in Essen on January 28-29, which was organized by Thomas Kollatz for the two eHumanities projects "Relations in Space" and "Inscriptions in their Spatial Context" that are funded by the Federal Ministry of Education and Research [Bundesministerium für Bildung und Forschung (BMBF)]. The workshop presented an opportunity to test the transferability of the software components developed in both projects, which the Bonn project also intends to reuse. Both projects concentrate on identifying relationships between content- and space-related features of culturally relevant objects. They link epigraphic studies and objects in space in order to draw conclusions about the meaning and function of text carriers. The Bonn Maya project also follows this approach and thus seeks professional exchange with these projects.

At the conference "Digital Humanities Images" in Bad Godesberg organized by the Max Weber Foundation [Max Weber Stiftung], the project engaged in productive exchange with the research project "Altägyptische Kursivschriften: Digitale Paläographie und systematische Analyse des Hieratischen und der Kursivhieroglyphen"¹⁰. This digital humanities project, which is funded by the Academy of Sciences and Literature in Mainz, is compiling a digital paleography of the cursive hieratic script and uses digital presentation and interpretation methods that are also being employed by the Bonn Maya project. Since both projects are conducting research on complex, hieroglyphic writing systems and pursuing common goals, they are considering a closer collaboration. Members of the Mainz project attended the workshop in Bonn on December 14. At the conference "Forms and Functions of the Editing and Paleography of Ancient Egyptian Cursive Scripts", the Bonn Maya project will also present its work and its approaches to representing the versatile Maya hieroglyphs with information technology. Close collaboration is also being pursued as part of the group "Encoding Complex Writing Systems", which will bring together over 40 researchers to develop common guidelines for using TEI to encode complex writing systems.

Tasks and Current Work in Bonn

¹⁰ <http://www.adwmainz.de/projekte/altaegyptische-kursivschriften/informationen.html>

In this section of the report, the project members of the Bonn workplace introduce their activities and the status of their work.

Text Archive

The work on the text archive initiated in 2014 was continued in 2015 (see 2014 project report). Since 2014, the text archive, various data collections, and working lists have been developed and compiled. They will be soon published on the project website www.mayawoerterbuch.de and also entered into the virtual research environment in TextGrid. The project plans to publish digital images of the photographs online on the website under the CC BY label. It has access to the archives of Prof. Karl Herbert Mayer, Prof. Berthold Riese, and Dr. Daniel Graña-Behrens, as well as those of the project director and members. As of December, student research assistants have digitized and indexed over 21,000 of the 40,000-plus analog photographs from the inscription archive of Prof. Karl Herbert Mayer (Graz). These digital images will be integrated into the virtual research environment over the course of the coming year and partially published on the website. Out of the 135 folders of photographs, drawings, and notes from the archive of Prof. Berthold Riese, 34 folders had been scanned and 11,000 digital images had been produced and inventoried as of December. We anticipate that a total volume of ca. 45,000 objects will be integrated into the virtual research environment. Due to the large quantity of data, digitalization and indexation of the photos from the archives of Berthold Riese and Karl Herbert Mayer will continue at least until mid-2017, parallel to processing the hieroglyphic inscriptions. In 2015, the project also obtained access to the extensive photo archive of the Maya specialist Daniel Graña-Behrens. Over the course of many field seasons, Graña-Behrens documented inscriptions from Maya sites in the Northern Yucatan Peninsula. In addition to 5,000 slides, his data collection includes drawings that will be digitized and indexed in 2016. The project plans to make these photos available to the public on the website in the near future.

Documenting Maya Inscriptions

In early 2015, the Bonn Maya project acquired a Breuckmann smartScan 3D fringe production scanner. According to the company website¹¹, the manufacturer describes the system as follows:

“The fringe projection system works on the basis of the so-called miniaturized projection technique (MPT). This procedure is known as active triangulation, which in the Anglo-Saxon terminology is often referred to as ‚white light scanner‘. A white light scanner consists of various hardware and software components which are one or two digital cameras (measuring field and resolution depend on the respective application), a projection unit (comparable with a slide projector), plus a computer installed with a data acquisition and evaluation software. The 3D data acquisition of an object with a white light scanner or a fringe projection system takes place in the following sequence: After the 3D scanner, the measuring object, and as the case may be a turntable or a robot for an automated acquisition, have been set up, the sensor components (cameras, projector) are calibrated. This is followed by the digital measurement of the object, in that the projector projects a sequence of fringe patterns (here: gray code procedure combined with phase shift technique) onto the measuring object, whereby each object displays an individual fringe pattern sequence in accordance with its contour. The system cameras capture this projected fringe pattern at a predefined viewing angle, while the individual projection varies for each object, and even for each view of the object. Depending on the application, either one

¹¹ This paragraph’s content is an excerpt from the manufacturer’s website:
<http://aicon3d.com/products/breuckmann-scanner/smartscan/at-a-glance/functional-principle.html>.

or two cameras are used. [...] An individual measuring sequence is completed within approximately one second; depending on the complexity of the measuring object, the entire measuring process takes between a few seconds and several minutes. The process is completed once the object has been completely digitized, i.e. individual images of all sides of the object have been captured and automatically saved. Within a matter of a few seconds, the computer then calculates the 3D data of the measuring object by aligning the individual captures by reference to the respective object geometry or by aid of index marks which prior to the data acquisition have been attached to the object. The result initially is a so-called point cloud which is then – depending on the measuring task in hand [...] and the aid of respective specialized programs - converted into a thinned net of triangles, referred to as the ‘mesh’. It is in this step of the process that for instance the surfaces are smoothed, or that the measurement noise is reduced and the data volume is decreased, without the overall data quality being compromised. The exact three-dimensional replica of the object is available as a data set in various formats and can be used for any type of further processing [...].”

The first tests of the scanner were conducted in the Bonn Ancient Americas Collection [Bonner Altamerika Sammlung (BASA)] on an original piece and three replicas made of plaster and fiberglass: 1) the original of a tobacco flask with an inscription and image of a figure; 2) the cast of a large Aztec altar with the image of the goddess Coyolxauhqui that is attached to the wall of the lobby in the Department of Anthropology of the Americas; 3) a plaster cast of block K1 from Step LXIII of the Hieroglyphic Stairway at Copan; and 4) the fiberglass replica of the left panel from the sanctuary in the Temple of the Sun at Palenque, on the basis of which Sven Gronemeyer completed a drawing on a digital drawing tablet. The text on the latter object was reinterpreted in the publication “*Tz’atz’ Nah*, a ‘New’ Term in the Classic Mayan Lexicon” on the basis of the 3D scan, which helped to make an important textual passage legible (see Publication 12). Another project publication concerns the aforementioned documentation procedures and their advantages for epigraphic research, especially when dealing with eroded passages (see Publication 6). The scanned objects from the Bonn collection were published by the project’s account on the internet platform Sketchfab¹², specialized for 3D objects (more in the section “Social Media”).

Furthermore, the project aims to document Maya texts in collaboration with museums and collections that have objects with inscriptions in their possession. In 2015, we successfully collaborated with the Knauf Museum – Relief Collections from Great Ages [“Reliefsammlung der großen Epochen”] in Iphofen, the Museum of Cultures in Basel, Switzerland, and the Ethnological Museum of Berlin. Another collaboration was agreed upon in March during Christian Prager’s official trip to Cambridge, England. The Museum of Archaeology and Anthropology in Cambridge possesses a large collection of plaster casts of Maya monuments that were produced over a century ago and have since been kept in the museum’s storage facilities. The casts depict the monuments’ state at the time of discovery; many of the monuments have since been destroyed or have badly eroded. The project will document these casts with the 3D scanner, integrate these into the inscription archive, and make them available for research on the website. These objects will be documented during a trip to Cambridge in the summer of 2016.

Iphofen

¹² <https://sketchfab.com/>



Iphofen: Palenque, Oval Tablet

Between July 29-August 4, 2015, project members Sven Gronemeyer, Antje Grothe, Christian Prager, and Elisabeth Wagner used the 3D fringe production scanner to document plaster casts of Maya monuments in the Knauf Museum in Iphofen. The museum was opened in 1983 in the 17th-century bursary of a former prince-bishop under the patronage of the family-owned company Knauf. It possesses 205 original plaster casts of monuments that were produced all over the world in the great museums and *in situ* at archaeological sites. The collection, which is distributed between 20 rooms, includes gypseous alabaster replicas from Mesopotamia, Egypt, Rome, Greek, the ancient Americas, India, Cambodia, and Easter Islands that span the great cultural periods since 3500 B.C.E. The more famous pieces on display include molds of the boundary stela of Senusret III, the Code of Hammurabi,

and the Rosetta Stone that is now in the British Museum.

Objects on display from the pre-Hispanic Americas include stone monuments from Cerro Sechin and Chavin de Huantar (Peru), as well as a collection of Aztec, Totonac, and Zapotec artifacts. A separate room houses a series of Maya monuments that were fringe production-documented during the Bonn team's four-day stay. In particular, the project documented 1) the Oval Tablet from the wall of the west corridor of House E in the palace at Palenque, 2) the mid-section of the relief in the Temple of the Sun (back wall of the cella), 3) the right side of the central tablet in the Temple of the Foliated Cross (back wall of the cella), 4) the sculpted Tablet of the Temple of the Cross on the west side of the entrance into the cella, 5) Altar L from Quirigua, 6) a fragment of the hieroglyphic stairway from Seibal, and 7) details of Lintel 23 from Yaxchilan. The 3D models are currently being analyzed and prepared for publication on the project's website in 2016.

Basel



Basel: Tikal, Temple IV, Lintel 3

Christian Prager, Elisabeth Wagner, and Sven Gronemeyer's two research trips to Basel from July 9-18 and August 26-September 4 concentrated on documenting the wooden Lintel 3 from Temple IV and Lintels 2 and 3 from Temple I at the Maya site of Tikal with the 3D fringe production scanner. The so-

called panels of Tikal are among the few extant historical wooden artifacts from the Maya civilization, which were long able to survive the tropical climate of the Maya region and were brought to Basel in 1878. Their texts record key events in the later history of the Maya lowlands and mention, among other topics, the end of Calakmul's hegemony over the site of Tikal, which had endured for over a century. The panels have been described as the "Mona Lisa" of the Basel collection thanks to their singularity, and numerous publications have underscored their importance.

In cooperation with the curator of the Department of the Americas, Alexander Brust, the individual wooden planks were completely scanned in the museum storage facility. From Lintel 3 from Temple 4, only the text and imagery on the front side could be scanned, as the object is currently on display in the museum. It is planned to scan the backside as soon as the exhibition is reconfigured next year. Using the 3D model, the project will prepare reproductions of the lintels, and has also planned a joint publication. The 3D models will be presented to the museum and will also be made available to interested researchers upon inquiry. The museum's current holdings additionally include plaster casts of monuments from Yaxchilan, some of which were also scanned. The project plans to scan more of the museum's objects with Maya inscriptions in the future and to make these publicly available on the website.

Berlin

Elisabeth Wagner undertook a visit to the Ethnological Museum of Berlin on March 15-21 to photograph Maya objects with inscriptions and figural imagery in the museum's storage facility. These photographs will also be made available to the museum. The project's photo documentation of objects with Maya inscriptions from the museum's storage currently includes complete ceramic vessels, ceramic sherds, and figurines, as well as a fragment of a stone inscription. We plan to continue this documentary work with the remaining text carriers (stone monuments, ceramic vessels, small sculptures, and jewelry) in the museum's public exhibition and to also document them with the fringe production scanner and photogrammetry in the near future.

Archaeological Site List

The list of archaeological sites described in the 2014 annual report is now online in the form of a database accessible via the link <http://mayawoerterbuch.de/?cat=201&lang=en>. As of now, 296 of the 526 documented sites with Maya inscriptions have been published, and the remaining 229 sites will soon be processed and put online. The working list registers ruins in Mesoamerica, sorted primarily by name, where Maya hieroglyphic texts have been found during archaeological exploration and excavation and have been verifiably documented. To the extent possible, the archaeological sites are shown in Google Maps using the researched geographic coordinates. This list constitutes the basis of the inscription inventory that is currently being prepared by the Bonn team and will also be published and curated on the project website as a separate, dynamic working list. With these materials, the project offers an overview of the sites that have previously been documented, including references to inscription carriers that have been discovered there. This database-based publication form is dynamic and has the advantage over printed site lists that newly discovered ruins with inscriptions can be immediately entered into the site database. Thus, the working list can be continuously updated. To ensure its long-term use and storage, printed versions of this liquid document at its current state of work will be published at regular intervals in the project's annual report. For each ruin, essential metadata like the preferred name, alternative designations or spellings, abbreviations, geographic coordinates, and further references will be documented, in addition to bibliographic citations

(personal communications, citations of literature, maps, internet references, etc.). Our metadata concept will be introduced below. The research data themselves are in the public domain and are being made available under the license CC BY.

The site database is two-tiered. It is accessed by means of a three-column table that contains all of the sites sorted in ascending alphabetical order. All three columns of the table can also be resorted (alphabetically descending). In the first column on the left, the preferred name of the corresponding site (in accordance with the conventions of the Corpus of Maya Hieroglyphic Inscriptions [CMHI]) appears in blue color. Alternative names, if applicable, appear below in black print. If multiple alternative names exist, they are separated with an equals sign “=”.

The middle column “Acronym” records the three-letter code that is used as an abbreviation for the site name. At present, the acronyms that appear here have been verified in published sources (Graham 1975, 1982; Graham & Mathews 1999; Riese 2004; Mathews 2005; Fash & Corpus of Maya Hieroglyphic Inscriptions Program 2012). The sort function allows the user to obtain an overview of which sites do not yet have an acronym. The third column contains the country and state or *departamento*/district, as applicable, in which the site is located. This column can only be sorted by country. The user can use the selection field at the head of the site list to display by default sites located in Mexico, Belize, Costa Rica, El Salvador, Guatemala, and Honduras. The instant search function also permits queries by full or partial site name.

The preferred name of a site also functions as a link to the second level of the database, which presents detailed information about each site. The general information about each site includes:

- 1) Designations
- 2) Identification numbers
- 3) Geographic location
- 4) Number of inscriptions from earlier inventories
- 5) Online presence
- 6) Additional comments and references

The first category “Designations” features additional alternative abbreviations and names attested in the literature. The corresponding sources may be found under the category “Sources” in the “Comments” field. It is essential to note alternative names, because the literature often contains various names for one and the same site. Another possibility for clearly identifying a site is to cite ID numbers from authority files and thesauri and to link them with entries in the site list. Authority lists are registers of standardized designations that are used as key words (descriptions) in documentation. They permit the unambiguous association and identification of persons, personal names, geographic features, subject headings, entities, work titles, etc.

We also link sites with the Getty Thesaurus of Geographic Names Online (TGN)¹³, a database with normed data that was developed for use in museum documentation. The thesaurus includes current and historical place names in various languages, as well as information such as geocoordinates and types of places. Its authority files are published under a free license (Open Data Commons Attribution License (ODC-By) 1.0) and are expanded and modified by selected experts in a special editing and review process. Sites in our database are thus linked not only with the normed data from TGN; the project will enter normed data into the TGN as well, for instance if existing data are incorrect or entries are missing.

¹³ <http://www.getty.edu/research/tools/vocabularies/tgn/index.html>

If a TGN entry exists, it appears in the database under the category “ID Number” in the field “Getty ID”. When possible, sites are also linked with ID numbers from the GeoNames vocabulary. This database contains over 10 million geographic names, each of which is categorized into one of nine classes and one of over 645 codes. The database contains longitude and latitude, height above sea level, population, and administrative classifications, in addition to place names in various languages. All coordinates employ WGS84 (World Geodetic System 1984), which is also used in GPS. Sites also receive a subject-specific ID number in the field “Atlas ID”, if one exists. For Mexican sites, the official ID assigned by the Instituto Nacional de Arqueología e Historia (INAH) is also cited; similarly, Guatemalan sites are listed with the site codes by the Instituto de Arqueología e Historia (IDAEH), and sites in Belize are designated with the registration number by the National Institute of Culture and History (NICH).

The exact geographic location is well-documented for most sites. Under the category “Position”, we present information concerning a) country and political-territorial subdivision in which the site is located, and b) geographic coordinates (WGS84). An ISO 3166 code is recorded in parentheses after the country and political-territorial subdivision. This international, standardized country code consists of three parts: the first represents the current country, the second the political-territorial subdivision, and the third is for names of countries that are no longer in use (since 1977). Geographic coordinates are given in decimal format and are taken from diverse sources, which are listed under “Sources/Comments”. Furthermore, this category provides a link to a tool for converting the cited coordinates.

The reliability of each entry is checked using Google Maps or other portals. Is the cited geographic location plausible, and how exact are the measurements? In many cases, the specifications from the literature can be verified using Google Maps. Uncertain or problematic measurements are noted under “Sources/Comments”. Each location is documented with a bibliographic citation, and the user can check it with the embedded maps from Google Maps.

The number of hieroglyphic texts that have been documented to date in the two published inscription inventories is noted under the category “Number of Inscriptions”. The first inventory was compiled by Sylvanus Morley in 1948 (see above). The second inventory is Berthold Riese’s (1980) overview that contains all sites with inscriptions known at that time, as well as the number of inscriptions per site. This area will be expanded in the future with data from the CMHI and our own research.

Moreover, under the category “Online”, we link the sites to the websites of research projects, refer to Google Street Views, for example, or link to Facebook and Twitter pages if the research projects have their own social media channels. Commentaries about the site, the reliability of the geographic coordinates, or literature, etc. are located under “Comments” and “References”. The references include not only bibliographic sources for the entries themselves, but also references to further literature that has been entered into the bibliographic database Zotero and can thus be continuously updated. Literature searches will be made available as a function in the future.

List of Museums and Collections

In addition to the aforementioned archaeological site database, the museum list that was previously described in 2014 is currently available for search via the URL <http://mayawoerterbuch.de/?cat=204&lang=en>. The list contains all collections of Maya objects in public museums and archaeological sites that are accessible to researchers. Currently, it contains over 200 entries and is being continually expanded. Together with the site list, the museum list constitutes a further basis for the inscription inventory. In addition to numerous Maya objects of unknown

provenience that often end up in museum collections through donations, bequests, or purchases, some museums and institutions also acquire Maya artifacts when conducting their own research. Thus, information about museums and collections are essential for compiling object biographies, which also include information about the object's documentation, history of acquisition, and location. In order to document a given object's history of acquisition and documentation as thoroughly as possible, we are documenting existing collections and their earlier designations, in the case that their name has previously been changed. In addition, we are taking into account museums and collections that have since closed, which are denoted in the database published online with a cross enclosed in parentheses after its name.

Like the site list, the museum database is two-tiered. The entry page features a four-columned table, and each column can be sorted alphabetically in ascending or descending order. The columns contain the name, acronym, city, country, and corresponding state or *departamento*/district. The second level of the database consists of the data record of the selected museum or collection. Each dataset includes the online presence of the relevant museum with links to its website and its social media channels. It also contains the collection's current name and, where applicable, its acronym, location, and contact information. The list also includes direct links to catalogs and databases with inventories of Maya objects as an aid for further research. When available, geographic coordinates, Geoname IDs, TGN and ULAN (Union List of Artist Names)¹⁴ are provided for precise georeferencing. Each dataset also features a map from Google Maps indicating the precise location of the museum. By registering the project with Geonames, a number of museums that had not yet been georeferenced were given a Geonames ID, or existing references could be corrected or specified. Like the TGN, the ULAN also is available as an authority file and can thus be similarly supplemented and modified.

Using the selection fields over the table head, users can display only museums and collections in certain countries. Additionally, the most recent entry and the last three modified entries are indicated in a separate menu on the upper right-hand side of the entry page. Furthermore, a comment function invites website users to report new museums or collections, as well as corrections, changes, or updates to museums for which entries already exist, so that the project can integrate this information into the database or update existing entries.

Literature

Another of the project's goals is to compile a complete bibliography of Maya civilization (see the 2014 project report). The project will compile an object description and bibliography for each text carrier. Text database users can thus retrieve information about an inscription's context. The VRE thus contains descriptions of the text carriers or information about the text's contents. In addition, the bibliographic database provides the user with an overview of the authors who have studied or published a monument, discussed a text passage, or were the first to publicly propose an as yet still valid reading for a hieroglyph or sign. The project is using the free and open-source application Zotero to compile the bibliography from various online and offline sources, to manage the bibliography, and to produce citations. The application supports developing and editing bibliographic references and lists. It also permits collaborative work from different locations. The bibliographic database contained 4,836 entries as of December 2014; by the end of 2015, it featured 11,441 entries. Each entry is checked for completeness and accuracy using databases and the original sources; when possible, it is also linked via a URL to an online version of the corresponding monograph or article. In cooperation with the Bonn University Library, the project acquired a license to access the bibliographic database "Anthropology

¹⁴ <http://www.getty.edu/research/tools/vocabularies/ulan/index.html>

Plus” for 2015 and 2016. The project and the ULB are sharing the costs for this license. With the aid of this resource, the project can check the data sets and register and enter additional literature. The bibliographic data can be exported into various formats and are also being successfully used in IDIOM in order that entries can also be referenced there. The literature database will be made freely available for research over the course of 2016 and available via an export function on the project website.

Virtual Inscription Archive in the Digital Library

The ULB offers free online access to its inventory of electronically accessible media via its website in the so-called Digital Collections.¹⁵ These collections include electronic publications, as well as digitalized editions from this library’s existing inventory. The ULB employs the software Visual Library and the VL Manager of the partner software firm Semantics Kommunikationsmanagement GmbH¹⁶ to manage the digital images and media and to present them and make them available online.

The project has included in its plans the development of a Virtual Inscription Archive in the Digital Collections of the ULB, to be realized in cooperation with this library. The digital images and some object-descriptive metadata, as well as text analysis and translations from selected Maya inscriptions, will be represented in the inscription archive, with the source of these items being the project database. The project database, in turn, will be compiled and stored in the VRE TextGrid. The project’s digital images stored in the VRE with their metadata are thus not a part of the library’s holdings in the typical sense that are hosted on the server of the North Rhine-Westphalian Library Service Centre [Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen]¹⁷ according to the usual procedures and can be edited using the Visual Library Manager. In order to be able to present the project’s contents in the Digital Collections of the ULB and to make them publically available there as planned, new solutions have had to be developed in Bonn and Göttingen in collaboration with the ULB and the company Semantics.

The collaboration with Semantics and the ULB continued in 2015 to this respect. This work focuses on the goal of transferring the first test data from the metadata compiled by the projects to Semantics, in order that both parties can meet concrete measures for technically planning and implementing the necessary steps for transferring and representing the data in the Virtual Inscription Archive.

To begin with, the procedural changes necessitated by the project’s work were updated at the beginning of 2015 in a meeting in Bonn between Jan Kenter (ULB), Christian Prager, and Katja Diederichs after the discussions with Semantics in December 2014. Additionally, the workflow for data exchange with the Visual Library software was established and addressed. From among the existing options, initially a possible data transfer was suggested using a OAI-PMH¹⁸ interface. OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) is a low-barrier mechanism for repository interoperability. Subsequently, Max Brodhun (SUB) in Göttingen evaluated whether TextGrid would permit delivering data mapped onto another format via OAI-PMH. After testing the interface’s technical design with positive results, the project agreed to elect this option and discard other suggested alternatives in favor of transferring data with the OAI-PMH interface. This technical choice led to the decision to additionally communicate an update of the original offer with Semantics, since it would be possible to forego the planned training in the Semantics software Visual Library Manager. This software is used to design the representation of contents in the Visual Library. Similarly, licensing

¹⁵ <http://digitale-sammlungen.ulb.uni-bonn.de/>

¹⁶ <http://www.semantics.de/>

¹⁷ <https://www.hbz-nrw.de/>

¹⁸ <https://www.openarchives.org/>

for the software's use and management, as well as distribution of URNs, were no longer necessary, since the data transferred using the interface already had persistent identifiers generated by TextGrid.

After having completed the metadata schema for the project's entry mask, the mapping of selected contents could be realized in the fall of 2015. For this process, elements of the metadata schema were identified that would be relevant for the presentation in the Visual Library. It is technically necessary to map these onto another metadata format in order to consistently assign the semantic contents of the metadata to a format based on another syntax. The semantic contents are largely based on CIDOC-CRM and describe non-textual contents. The format that is correspondingly used in the Digital Collections to represent contents consists of the open XML-based standards METS (Metadata Encoding & Transmission Standard)¹⁹ and MODS (Metadata Object Description Schema)²⁰. These standards have been successfully tested for digital libraries and their schema definitions furthermore make the contents compatible with the DFG-Viewer.

In September 2015, an initially mapping was then tested for practicability, in another meeting in Göttingen between Katja Diederichs and Franziska Diehr it was discussed and eventually defined. Here metadata elements that are used for visual presentation in the Visual Library were defined for mapping. The project also defined elements needed only for searches in the Digital Collection. Metadata elements that, for instance, describe the essential, basic information about an inscription and are supposed to be presented here, like the artifact title, place, or measurements, were accordingly assigned to the appropriate MODS elements. For example, the content of the metadata schema for the preferred title would be represented in a mapping as follows: a connection with the element `idiom:preferredTitle`, which contains a literal with the value "Yaxchilan Lintel 8", is established via the metadata element `idiom:Artefact`. In a mapping, these contents would be represented in the MODS element `<mods:titleInfo>` (with the attribute `xml:lang` to be filled out) and the subelement `<title>` nested in it.

Thus, in the metadata schema the description of the title element connected with the artifact class would be presented as follows:

```
<idiom:preferredTitle>Yaxchilan Lintel 8</idiom:preferredTitle>
```

This presentation would correspond to the following representation in MODS:

```
<mods:titleInfo xml:lang="en">
<mods:title>Yaxchilan Lintel8</mods:title>
</mods:titleInfo>
```

In late 2015, the first test data transfer of completely mapped objects was transferred to Semantics so that their usability could be evaluated. This refers to the object-descriptive, METS/MODS-based data on the one hand and, on the other hand, to a mock-up that graphically illustrates how the representation of the data could eventually look in the Visual Library portal.

The project plans to create an XML/TEI-based metadata schema for data from textual analysis in 2016. The next step of this undertaking will require finding a way to transfer the missing transcriptions, transliterations, and translations into Visual Library. There, representations of the inscriptions will eventually be implemented for the Virtual Inscription Archive with the textual metadata (transliteration, transcription, and translation), in addition to the object-descriptive data. Over the course of 2016, the exchange between Semantics, the project in Bonn, and the SUB in Göttingen will be continued through additional procedures concerning the metadata.

¹⁹ <http://www.loc.gov/standards/mets/>

²⁰ <http://www.loc.gov/standards/mods/>

Online Presentation and Social Media

Online Presentation

2015 also saw continuing development of the project's WordPress website²¹, which had been initiated in 2014 as Phase II of the project's online presentation. The go-live that Beuse Project Management as the executing agency had projected for late 2014 could not be realized for various reasons. Based on the initial implementation of the website in a CMS (Content Management System), the website required another iterative process to synchronize the design, conception, usability, and technology as defined by all of the project's members.

The first correction loop, with a three-columned page layout along the lines of the corporate design of the North Rhine-Westphalian Academy of Sciences, Humanities and the Arts, was completed in March 2015. In April, Sven Gronemeyer received several elementary training sessions from Beuse Project Management, which consisted of multiple modules: WordPress as an editing interface; use of the backend; basic familiarity with object-oriented PHP for managing templates and changes in the core; use of git for version control between the local development environment and the remote bare repository for joint development. A second correctional loop to fine-tune functionalities, designs, and continual bug-fixing was completed in June 2015. Additional, small adjustments were made through the end of 2015 as part of continuing editorial operations.

After a stable system could thus be established, the bare repository and the MySQL database with the website's contents was cloned on the Virtual Machine of the [Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen] GWDG²², the website's Internet Service Provider. Additionally, a backup tool was implemented that a cron job (time-based process) uses to make daily backups of the contents. From this point on, continuous editorial work was conducted on the clone to test its system stability.

Besides editing diverse project publications, content work consisted primarily of quality assurance of the working lists. The site list and the museum list in particular were reviewed to guarantee a unified level of information for the go-live. Of the 526 archaeological sites on the working list, 296 entries could be revised by the time of the go-live, along with 127 of the 365 museums and collections. Catherine Letcher Lazo was hired as a research assistant in October 2015 to coordinate the editorial work, which consisted of editing the German texts, coordinating the translation of the contents into English (executed by Mallory Matsumoto), and translating the contents into Spanish.

After the test phase and additional content management, the website was introduced to the public on December 14, 2015 during the project's second annual conference, following the actual go-live the week before. Some minor bugs require additional fixes in 2016; additionally, the working lists still need to be developed, and the Spanish version remains to be completed. The plugin "Magic Fields"²³ is also being used for the sign concordance and for the morpheme list. The latter requires additional, adapted templates, whose conceptualization will be continued in 2016.

The Zotero plugin for managing citations is also still under development. It is divided into a functional branch (communication with the Zotero API) and a backend branch (embedding of the plugin in WordPress). Its development will be completed in 2016. Phase III of the website (integration of SADE (Scalable Architecture for Digital Editions) for publishing TextGrid objects) is currently not scalable and

²¹ <http://mayawoerterbuch.de>

²² <https://www.gwdg.de/de>

²³ <https://de.wordpress.org/plugins/magic-fields/>

is being suspended. Instead, starting in 2016, the website will be expanded with additional functions: a gallery of non-epigraphic photos from the digitalized archives searchable by descriptions/tags, a comparable gallery for the 3D-files, and an additional category for external publications not released on the website. These initiatives are taking place as part of the project's "Green Open Access," or self-archivation of materials.

Social Media

One aspect of the website is that research data and questions can be quickly and broadly disseminated, whenever possible in interaction with the user, i.e., the scientific community. For example, feedback can be given on all working lists to foster discussions that nonetheless are conducted outside of visible communication channels. For publications on the website, the project will offer a public comment function. This function, which is natively available in WordPress, will create a more informal, faster format with wide coverage. These comments aim primarily at facilitating scientific dialogue that engages with the contents of the publications.

In addition to the website itself, social media channels also play a major role. Channels on Facebook²⁴ and Twitter²⁵ were put online at the same time as the website go-live. Both channels allow the project to incorporate itself into the scientific community and the broader public by announcing events, new publications, new entries in the working lists, etc. At the same time, they permit a behind the scenes look into the project and thus real-time insight into the project's work. More colloquial comments are expected and even encouraged here.

The project has established another channel with the web-based service Sketchfab²⁶. This platform offers an online viewer for 3D models and also has a social component for sharing, liking, and commenting on models. It uses the WebGL JavaScript API to represent 3D files on the server side and the browser-based public domain OSG.JS JavaScript library. Consequentially, the project's freely manipulable 3D models can be viewed independent of the user's platform (including interactive features such as movable light sources or rendering options). Thanks to its status as an educational establishment, the project received a gratis pro-account that allows more complex models to be uploaded (up to 10 million faces). Other cultural institutions (e.g., British Museum, Peabody Museum) already have their own channels, making an interactive repository of Maya objects available. The 3D models in Sketchfab can also be easily shared via other social media or embedded in other websites, including that of the project.

Tasks and Current Work in Göttingen

The design and development of the project's IT components are part of the duties of the SUB Göttingen. The tasks of developing the metadata schema and programming and adapting the VRE TextGrid to the project's work, needs, and goals occupy the forefront of this work. In the following paragraphs, we summarize the team's work in 2015.

²⁴ <https://www.facebook.com/idiom.project/>

²⁵ https://twitter.com/idiom_project/

²⁶ <https://sketchfab.com/idiom-project/>

Metadata for Non-Textual Objects

The core components of the work in 2015 consisted of facilitating the recording of data for non-textual objects. Implementation essentially comprises four large areas that are closely connected and strongly influence each other in their development: formulation of scientific demands on metadata recording, design of the metadata schema, development of vocabularies, and implementation of the entry mask in the VRE. At the end of 2014, these areas were all still in design, and they were made productive and operational over the course of 2015.

The development of the entry mask is an iterative process in which all areas are dependent upon each other. Because the following steps all influence each other, they cannot be separated and require close collaboration and constant communication:

- 1) Formulating scientific requirements of the metadata schema
- 2) Developing and adapting the metadata schema
- 3) Developing vocabularies
- 4) Technical implementation
- 5) Testing the prototype

Most of the scientific requirements of the metadata schema were already formulated in 2014, allowing the project to go ahead and integrate some of them into the metadata schema. In the first four months of 2015, an initial version of the metadata schema was completed, which indicated the structure for data organization needed for its technical implementation. In the next step, the first prototype of the entry mask was developed. It was then evaluated and improved by all team members during a period of a thorough testing. During this period, the scientific requirements of the metadata schema were also made more concrete. This, in turn necessitated adjustments to and technical implementation of the metadata schema. The opportunity to directly interact with the entry mask indicated which components were missing or defective.

The newly implemented requirements then had to be tested, making the process a cycle that only concludes when all desired demands have been implemented in the metadata schema and the entry mask. This goal was attained in late 2015, which allowed the entry mask to pass into an operational version. Below, the individual phases and the corresponding work packages will be explained in detail.

Formulating Scientific Requirements and Developing and Adapting the Metadata Schema

The project used a catalog of specifications to formulate the scientific requirements, whose development began in 2014 and was completed in early 2015. On the basis of this work, the project could continue the work of modeling the schema that it had initiated during the previous year.

The basis for modeling the IDIOM metadata schema was CIDOC CRM (CIDOC Conceptual Reference Model), an ISO standard that was developed in museum studies for formally describing museum objects and processes. Thus, this reference model is also suited for describing epigraphic processes and objects, as well as for documenting the history of research. Additional standards and schemas were used for modeling along with CIDOC CRM. Many standard terms were not directly integrated into the IDIOM schema, however. Instead, they functioned as a super-class or super-property for a specified IDIOM term.

The scientific requirements of the metadata schema have proven to be comprehensive and complex. The project would like to document all known information about the history of a given text carrier, including statements that contradict each other and may originate from different sources. For

example, take the size of a text carrier: in many cases, the same artifact has been measured by multiple researchers with divergent results. As a result, different measurements for a single artifact circulate in the scientific community and literature. These disparate measurements are not necessarily the result of erroneous measuring or the use of different measuring techniques; they may, for example, result from the monument having eroded over time. For this reason, the correctness of such citations cannot always be determined. Thus, the project has decided to document all extant measurements of an artifact, always citing the relevant source.

The IDIOM Schema

The main features of the metadata schema will be introduced below (Figure 1).

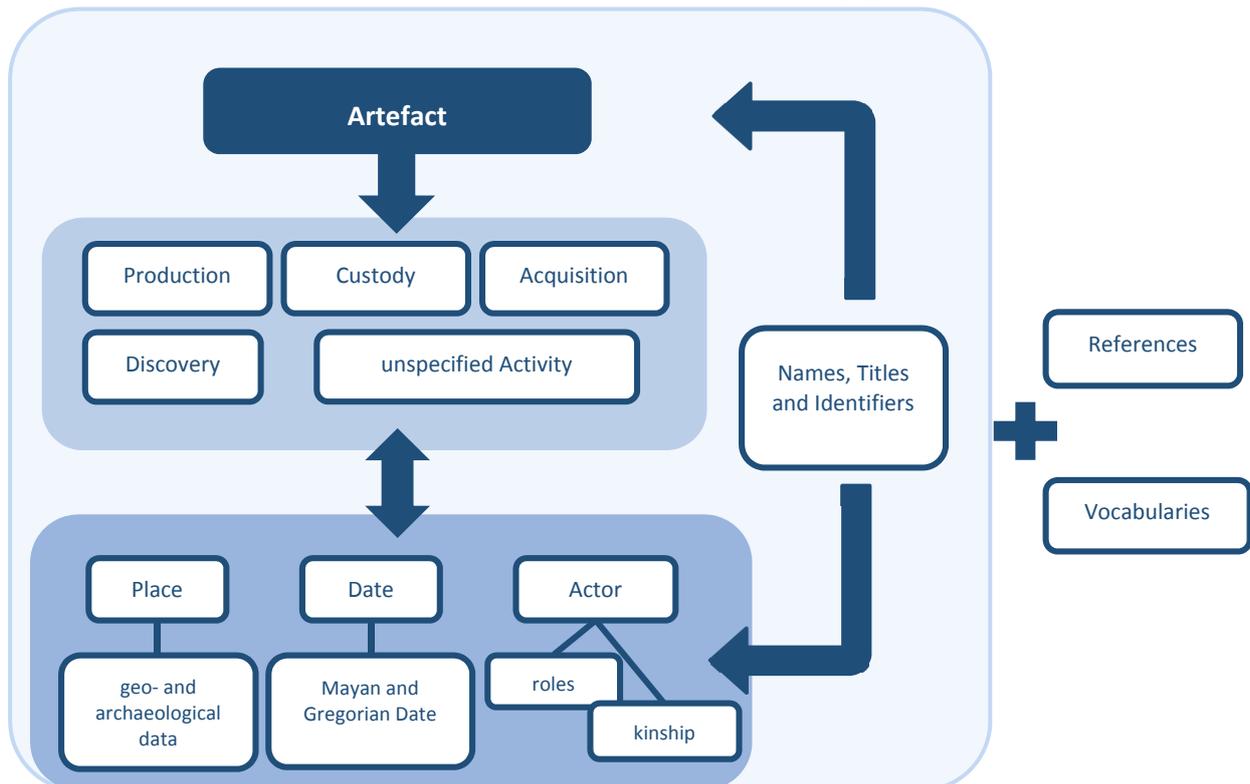


Figure 1: Simplified representation of the IDIOM schema

Physical artifacts that have been crafted or modified by hand are summarized under the class *Artefact* (idiom:Artefact). In addition to text and image carriers, this category also includes other objects that do not necessarily bear an inscription or visual image. This procedure allows the project to document objects that exist in connection with an information carrier, but may not themselves contain text or imagery, such as buildings, artifact fragments, tombs, etc.

The object's properties, such as form, size, and condition, can be described by means of the schema class *Artefact*. In order to also compile the artifact's biography in a subsequent step, this class is linked with events, which themselves are described in detail. The essential events documented for the object biography will be explained briefly here.

The event *Production* (crm:E12_Production) documents the production of an artifact, with information about the materials and techniques used (controlled vocabularies), place and date of production, producer, and commissioner.

The first scientifically documented discovery of an artifact is described with the event *Discovery* (idiom:Discovery). Here, detailed specifications can be recorded about the context of discovery, discoverer, site, and date.

For artifacts currently in the possession of a person (e.g., private collector) or institution (e.g., museum), this event can be registered with the event *Custody* (idiom:Custody). Changes in ownership of an artifact can be described using the event *Acquisition* (crm:E8_Acquisition).

It is possible to create additional events that are connected to the artifact or historically relevant for the history of Maya civilization by defining a non-specified event *Activity* (crm:E7_Activity). Using the controlled vocabulary *Activity Type* (idiom:ActivityType), the event can be categorized, which permits a search for expeditions, wars, or accessions, for example. Furthermore, with the controlled vocabulary *Role* (idiom:ActorRole), persons and groups that have participated in events can be assigned the concrete roles that they have filled in the described event. Events can be lined with *Places* (crm:E53_Place), *Dates* (idiom:Date), and *Actors* (crm:E39_Actor), for which additional information is compiled.

Archaeological sites and places of artifact discovery, among others, are recorded as places. By means of various relations, hierarchies of places can be established here to represent the organization of a site, for example. Furthermore, places are associated with geocoordinates and IDs from the normed datasets Getty TGN and GeoNames. Excavation quadrants and other referential coordinates used in archaeological excavations can also be compiled with the IDIOM schema.

A date can be given in the form of a date in the Gregorian calendar as well as of a Long Count. In addition, the entry mask provides function for automatic conversion.

The project is also registering actors related to events in the artifact biography, such as researchers, collectors, dealers, museums, archives, and other institutions. Furthermore, Classic Maya persons and groups (including those of mythological origin) are also being recorded. They either exist in connection with the history of the artifact, play a role in the content of the text, or are otherwise significant for Classic Maya civilization. Familial relationships can also be documented for persons.

In research practice, more than one designation is often used for places, actors, and artifacts. The schema thus allows those with multiple *Names* (crm:E82_Actor_Appellation, crm:E48_Place_Name), *Titles* (crm:E35_Title), and *Identifiers* (crm:E42:Identifier) to be linked.

Another essential requirement from a research perspective is that bibliographic sources be given for the cited information. *Sources* (idiom:Source) can be added for all previously given information (artifacts, events, places, dates, actors, descriptions, etc.).

Thanks to its structure, the IDIOM schema allows individual descriptive sets to be connected with each other. The advantage of this set-up is that each entity is only created once and then linked with others. For example, a site does not need to be re-registered for each new *Discovery*; instead, a link is established between the complete description of the site (with coordinates, various place names, etc.) and the complete description of the *Discovery* (with actors, date, etc.). As such, complex relationships can be established.

Adjusting the Schema during the Test Phase

As previously mentioned, changes were made to the metadata schema during the test phase. One such adaptation will be described here using the following example.

The project is documenting Classic Maya persons, which are summarized under the class *idiom:EpigraphicActor* in the IDIOM schema. However, a Classic Maya person, such as a ruler, has not

just one name, but instead may have both a birth name and a regnal name, for instance. Furthermore, multiple names are often used in the literature to refer to one and the same person. The project aims to represent the current state of research, as well as the history of research. As a result, it needs to compile all known names and name variants of each Maya person. In the IDIOM schema, this procedure was implemented in the prototype as follows:

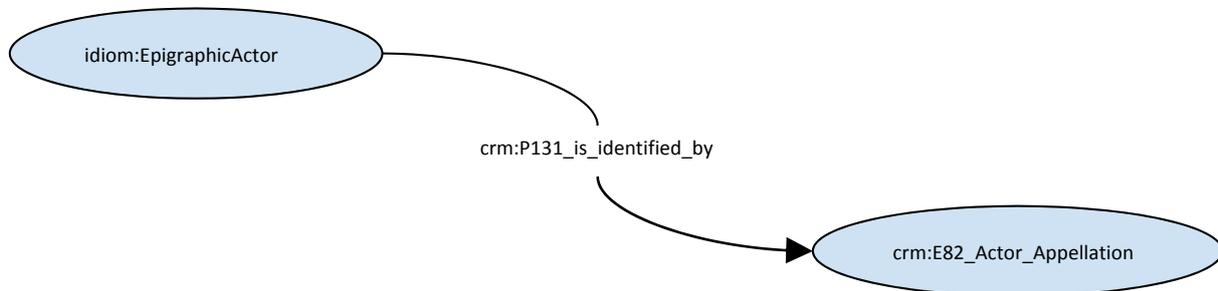


Figure 2: Initial modeling version

The class *crm:E82_Actor_Appellation* describes the name of an actor. This class is linked to *idiom:EpigraphicActor* by means of the relation *crm:P131_is_identified_by*. Since a person has at least one name (1 – n), it is possible to add an infinite number of names using the entry mask. However, during the test phase, the team realized that it needed to define a preferred name for each person. The project aspires to develop and establish standards that will be adopted by others; thus, by determining specific designations for people, it hopes to contribute to the development of naming conventions.

The requirements changed during the test phase when the project decided that each “Epigraphic Actor” should be assigned a preferred name and that additional names could also be given. These alterations, in turn, required the following changes to be made to the metadata schema:

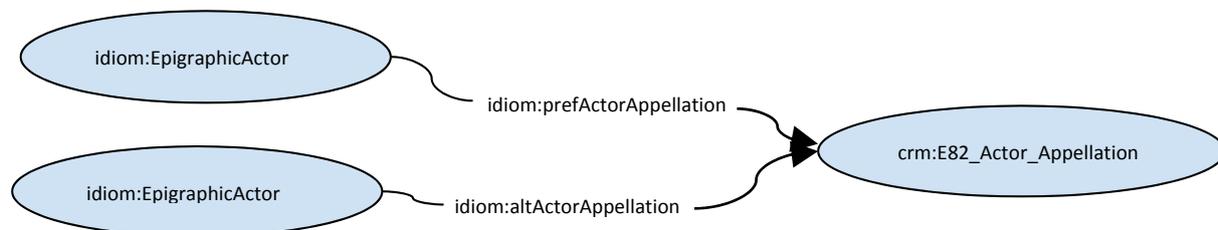


Figure 3: Modeling adjusted to the modified requirements

The class *idiom:EpigraphicActor* is now linked with the class *crm:E82_Actor_Appellation* through two relations. The relation *idiom:prefActorAppellation* allows exactly one preferred designation to be assigned to a person. As such, entering a preferred name is a mandatory step. Additional designations can be provided using the relation *idiom:altActorAppellation*. This example is representative of other modifications executed during the test phase.

Changes to the metadata schema entail additional time and effort invested in modelling, for instance to search through standards for appropriate terms, to classify these in the project’s schema, and to add new classes and associated subclasses or definitions of subproperties. Additionally, they also require that the entry mask be modified as well, which in turn requires that more time and effort be invested in programming.

After transitioning from the prototype to the operation version of the entry mask in late 2015, the IDIOM schema for non-textual objects can be considered complete. Documentation of the schema will be published in early 2016 in machine- and human-readable format.

Developing Vocabularies

The project is developing controlled vocabularies to support the documentation of non-textual objects. Although the use, development, and management of vocabularies is an essential part of many documentation and research projects, viable software solutions from the non-commercial sector are unfortunately lacking. The project's search for open-source tools was largely unfruitful: most programs were not up-to-date, were largely if not completely undocumented, and could not be integrated into the project's VRE. Even the open source tool SKOS Vocabulary Editor²⁷ (SKOtch) is not regularly updated. Nonetheless, the project elected to use this latter program, since it is easy to operate and offers the range of functions that the project requires. SKOtch supports vocabulary modeling in RDF with the SKOS standard and stores the data in a triplestore.

When formulating scientific demands and developing the metadata schema, the project became aware that it would need 10 controlled vocabularies for indexing. Development of the vocabularies was initiated in late 2015 and will be continued in the upcoming year. For 7 of the 10 vocabularies, the terms, including preferred and alternative designations, were compiled in English, in order that they could be available in the entry mask and thus used for recording objects. The project still needs to finish the terminology, matching to normed vocabularies (especially the Getty Art and Architecture Thesaurus), and documentation of the vocabulary. After these tasks have been completed, the resultant products will be published in machine- and human-readable format.

Technical Implementation of the Metadata Schema

Technical implementation of the metadata schema is divided into the following steps:

- 1) Implementation of the metadata schema in a machine-readable format
- 2) Implementation of the epigraphic workflow in the entry mask
- 3) Test phase for the entry mask
- 4) Editing
 - a. Usability
 - b. Technical errors
- 5) Revision of the metadata schema, if necessary
- 6) (Resume process at step 1)

Compiling the Turtle

The schema must be available in machine-readable format so that the entry mask can process data that are assigned to a class. For this purpose, a file was compiled in the RDF syntax *Turtle*²⁸.

²⁷ <https://trac.nbic.nl/skosvocabularyeditor/>

²⁸ <https://www.w3.org/TR/turtle/>

```
idiom:Artefact a rdfs:Class;

rdfs:label "Artefact"@en;

form:prefix "http://idiom.uni-bonn.de/termstwo/";

tgforms:helptext "An artefact is a physical object crafted by Classic Maya. The artefact may or may not carry epigraphic content, an inscription and/or visual information. Examples: 'Temple of the Cross of Palenque', 'Yaxchilan Lintel 8.';

rdfs:subClassOf crm:E22Man-MadeObject.
```

Figure 4: Example of a class in the IDIOM schema implemented in Turtle/RDF

```
idiom:preferredTitle a rdf:Property;

rdfs:label "Preferred Title"@en;

rdfs:domain idiom:Artefact;

rdfs:range xsd:Literal;

tgforms:isMandatory true;

tgforms:isRepeatable false;

tgforms:hasInput tgforms:text.
```

Figure 5: Example of a field property belonging to the class idiom:Artefact in Turtle/RDF

This procedure must be undertaken for each class and each property of a class. As such, changes to the schema required that corresponding changes be made to the Turtle file.

The Turtle file contains all information needed to generate the form in the entry mask. In addition to the metadata themselves, information concerning their technical implementation are also provided at this point:

- Reproducibility
- Required field
- Type of entry
 - Button
 - Text field
 - Text area
 - Check box
 - Drop-down
- Priority number (position in which the field is represented in the form)
- Help text for entry

Visual Implementation of the Schema in the Entry Mask

Some aspects of the visual implementation of the metadata scheme required that changes be made to the core components TG-forms²⁹. This affects 1) sub-objects of a class in the schema, and 2) identification of special field properties.

Initially, multiple schema objects could be loaded in one form, but there was no visual delimitation and affiliation, since deeper levels were not initially taken into account. For the example *Epigraphic Actor*, see Figure 6.

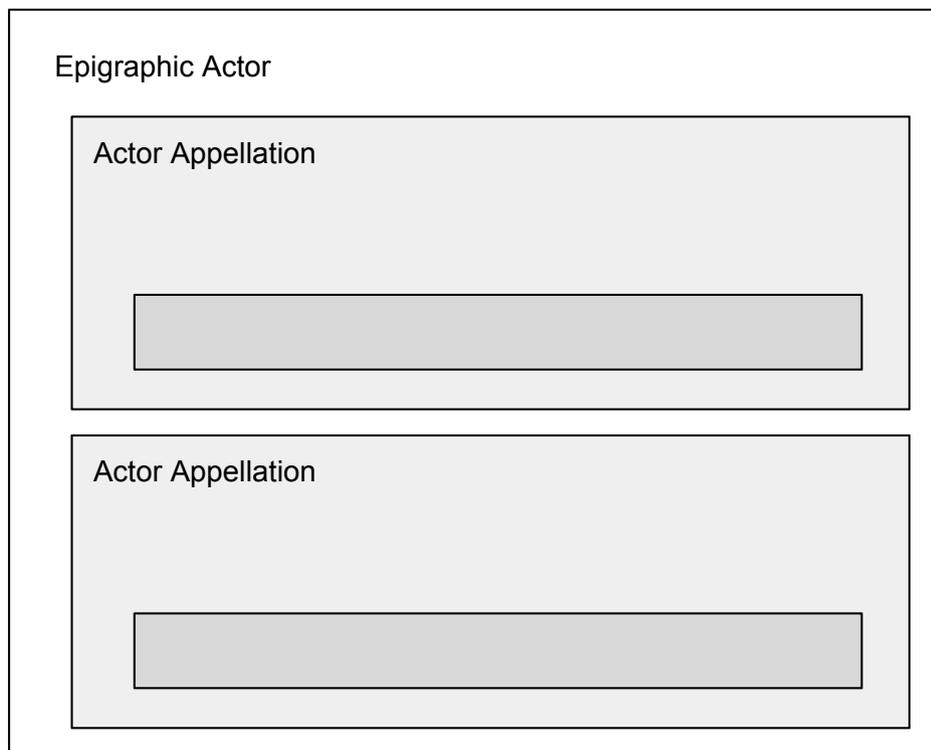


Figure 6: Schematic representation of the schema visualized in the entry mask. The source is clearly assigned to an Actor Appellation.

Thus, not every conceptual object of the metadata schema should also be conceptualized as a discrete object for the database. As an example, take the relationship between the schema classes *idiom:EpigraphicActor* and *crm:E82_Actor_Appellation*. Both of these objects are their own classes for the metadata schema. However, the class *crm:E82_Actor_Appellation* is directly linked to the object for storing and representing this relationship. Thus, on the one hand, various types of storage were necessary in such cases; on the other hand, the visualization had to be sorted out. For the visualization, a directly visible connection had to exist between the objects. One possible solution would have been to place *crm:E82_Actor_Appellation* under the existing form. However, the Actor Appellation has additional sub-objects, such as a bibliographic citation, for which a further sub-form is required; consequentially, it would no longer be clear in the case of repeated citations to which entry the citation is referring. Hence, we decided in favor of a nested representation and implemented a storage procedure using referenced IDs. For this, an ID (e.g., *crm:P102_hasTitle15259507240GP*) is generated for each sub-form by activating the button. This ID is used to store the object in triplestore and is employed to reference this sub-object when opening the super-object. This procedure allows one to

²⁹ This is a TextGrid application for generating forms for compiling metadata.

recognize at all times to what a relation is referring and prevents RDF reifications³⁰ from being used here.

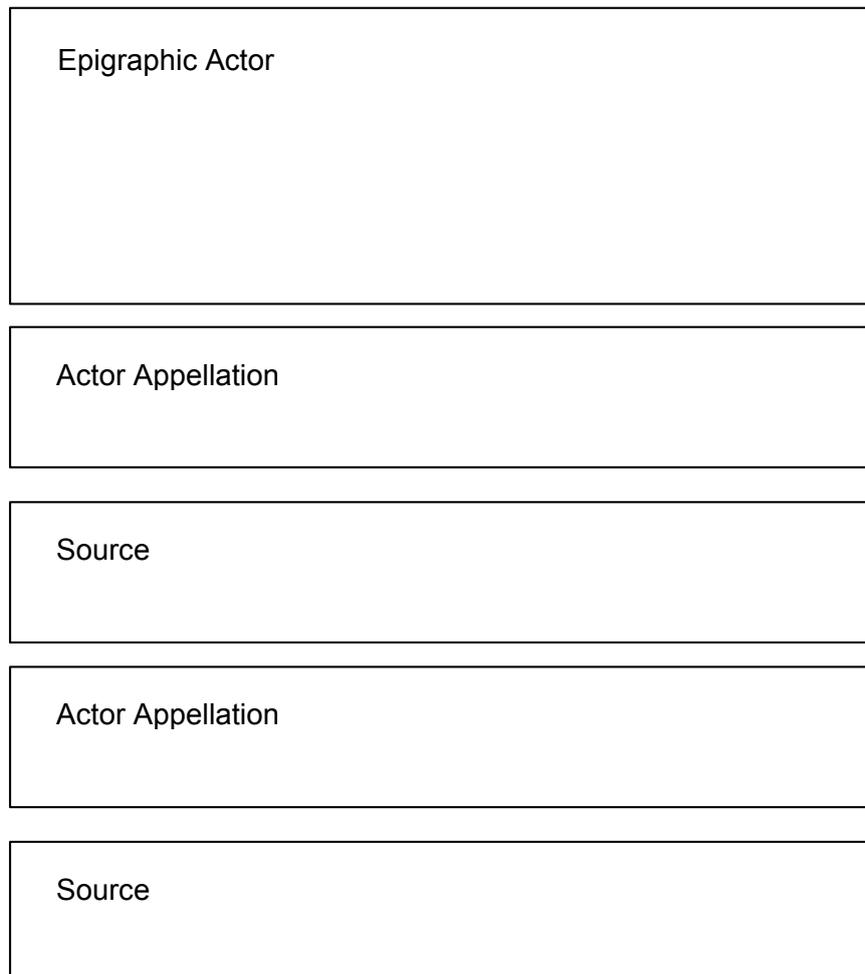


Figure 7: Schematic representation of the schema visualized in the entry mask. It is not clear what source belongs to which Actor Appellation.

Technical Implementation

The entry mask serves to compile the metadata of non-textual objects in RDF. Additional requirements beyond those made in the previous year were added to the entry mask to make data entry as intuitive and as scientifically accurate as possible. These supporting functions are described below. The IDIOM entry mask is configured based on an RDF entry mask that already exists in TextGrid, but has been adapted to meet the project's specific needs. As such, some adjustments also had to be made to the core application TG-forms. These changes were integrated into the overall application of TG-forms and can thus also be used by other projects.

³⁰ RDF reifications are statements about statements, i.e., additional information about a compiled triple. They go beyond the actual characteristics of RDF. They can be cumbersome to handle and are often confusing, for which reason the project tries to avoid using them whenever possible.

Supporting Functions in the Entry Mask

Date Correlations

The date is often recorded on text carriers in the Long Count format, which is also frequently used in research. Yet in order to be able to programmatically analyze these dates, they have to be available in Gregorian format. For this reason, the project added an automatic conversion that is performed immediately after the user enters information into the field Long Count and then clicks outside of the field.

In order to also be able to automatically calculate the Long Count date, including the Calendar Round, this function also calculates *Tzolk'in* and *Haab* values. These calculations are also executed after clicking outside of the Long Count field, whereafter the values are added to the corresponding fields.

The screenshot shows a form with the following fields and values:

- Long Count:** 9.0.12.0.1
- Tzolk'in Coefficient:** 13
- Tzolk'in:** Imix
- Haab Coefficient:** 14
- Haab:** Ch'en
- Gregorian Date:** 447-10-10

At the bottom, there is a 'Source' field with a link icon, a plus sign, and an information icon.

Figure 8: Example of calculating the appropriate *Tzolk'in* and *Haab* coefficients and the Gregorian date based on an entered Long Count date.

Disabling Unnecessary Buttons

The entry mask contains three different types of buttons:

- 1) Open a context menu to search for particular objects and enter them as a linked object
- 2) Generate a sub-form and thereby add additional metadata
- 3) Delete these two entries

The function of adding data should only be available when no link has previously been entered or when no sub-form was loaded. In this case, the delete function, in turn, should not be available (see Figure 6). Otherwise, if a link or sub-form has already been generated, the user should not be able to activate

the add button again (see Figure 8 and 9). If data need to be corrected, the contents should first have been deleted. After this button is clicked, the add button is then made active again. This procedure reduces the likelihood of errors.

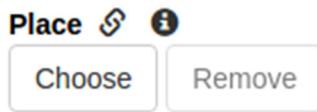


Figure 9: Field before selecting another object. “Choose” is activated and “Remove” is deactivated.



Figure 10: Field after selecting another object. “Choose” is deactivated and “Remove” is activated.

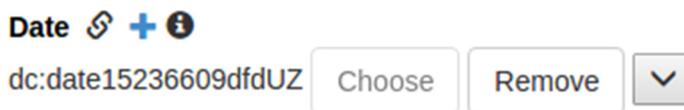


Figure 11: A sub-form has been created. “Choose” is deactivated and “Remove” is activated. In addition, a sub-form can always be opened or closed.

Search Functions

Specific queries have been established behind the respective context menus to link objects with others by means of a URI. Thus, queries for bibliographic sources, places, artifacts, actors, and activities, for example, have been stored. These SPARQL queries are sent to the triplestore. All search functions permit only parts of the search term to be considered and capitalization to be ignored; additionally, special characters do not necessarily have to be entered as such (e.g., entering ‘a’ also searches for ‘ä’).

Opening the Sub-forms

Every object can consist of multiple sub-objects that are connected with each other in the schema by a relation. However, a link is not created directly to a sub-object from another object. Instead, a given link always exists from one sub-object to another sub-object. As such, these sub-objects are not stored as a discrete database objects and do not receive their own TextGrid URI. These sub-objects are stored in triplestore. When the button is clicked, they receive an identifier so that affiliation can be guaranteed at the database level. Each sub-object from the next lower level is represented in one shade of gray darker, in order to better differentiate a sub-form from the object above it (see Figure 9).

Additional Title  

crm:P102_hasTitle15236b9b5c7x3

Title

Status * 

Title Type *

Title *

Source   

dct:isReferencedBy15236b9bb070f

Source

Confidence Level * 

Figure 12: Title is a sub-form of an artifact; consequentially, it is a bit darker than the white matrix. Source is a second-level sub-form and is thus even slightly darker.

Saving and Loading the Sub-forms

Storing the data within the created sub-form initially proved to be more difficult than in the case of the first-level data. For this reason, the algorithm had to be re-programmed to a recursive procedure. The next challenge appeared when loading the data from a sub-form. The first point of identification for assigning first-level data (including those from the first sub-forms) is the TextGrid URI. However, at the second level, the TextGrid URI is not the initial point of identification, because these data have to be recalled using the ID generated by the entry mask. As a result, the algorithm for opening, in addition to the algorithm for storing data, had to be modified to be recursive.

Link to the Bibliographic Database

An important aspect of compiling the metadata is entering the bibliographic sources. They are inserted into the triplestore via RDF export after a data dump from the bibliographic database Zotero. After this step, the sources can be searched using the entry mask and linked with the relevant data. The search function responds to author, title, year of publication, and organization. When searching, special characters must be accounted for. If the search term includes the letter < a >, for example, occurrences of the letter < ä > also have to be included in the search results.

Linking the Vocabularies

Use of vocabularies is unavoidable. They are stored in RDF/SKOS using the tool SKOth, whereas the data are stored in a triplestore. The project also had to install and configure the triplestore Sesame,

because the tool SKOth is not compatible with the triplestore Fuseki. However, in order to have the data collectively available in a database, the data are transferred from Sesame to Fuseki using a script.

An Example Research Question

In November 2015, the project began entering large quantities of metadata (751 objects with 16,857 triples) concerning the archaeological site of Tikal. These data were used as the basis for a comprehensive presentation of the project at the European Maya Conference in Bonn in December 2015. At this event, the project visually represented queries based on a research question to give our scientific colleagues a clear example of how the recorded data could be used. The sites and dedicatory dates of artifacts in Tikal occupied the forefront of this example question: which text carriers are located on the Main Plaza? When were they dedicated, and which ruler commissioned them?

The geo-browser of the project DARIAH-DE was used to visualize the artifacts on a map. For this purpose, the geocoordinates had to be available for the locations where the artifacts were found. A CSV file was generated from the data and sent to the geo-browser, which represented the data on a map and made it possible to animate the chronology of the artifacts' dedication along a timeline.

For this research question, artifacts, places (hierarchically according to an archaeological order), production and discovery events, and epigraphic actors had to be queried and be produced as a sorted result. Thanks to the links within the schema, they could be retrieved using a SPARQL query. Figure 10 schematically illustrates this query and the corresponding objects.

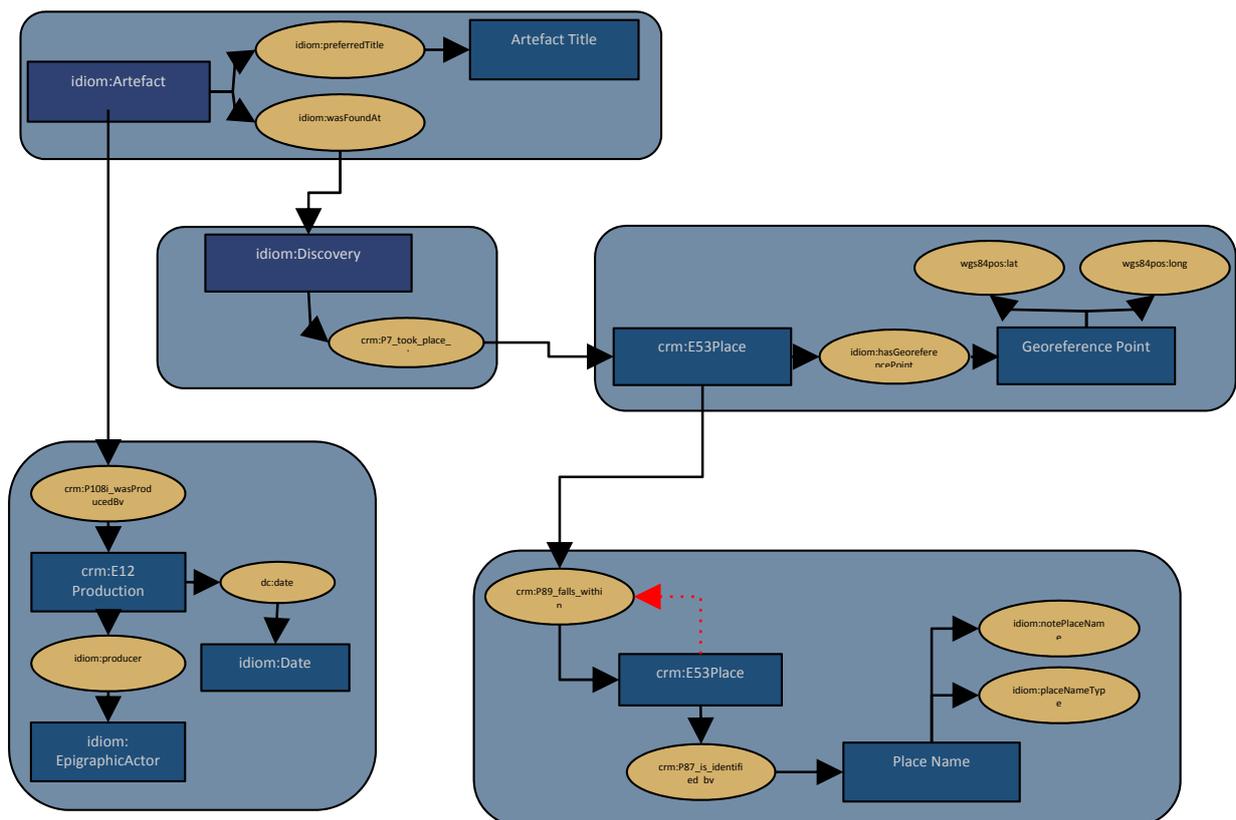


Figure 13: Schematic illustration of the SPARQL query and the links between the objects in RDF

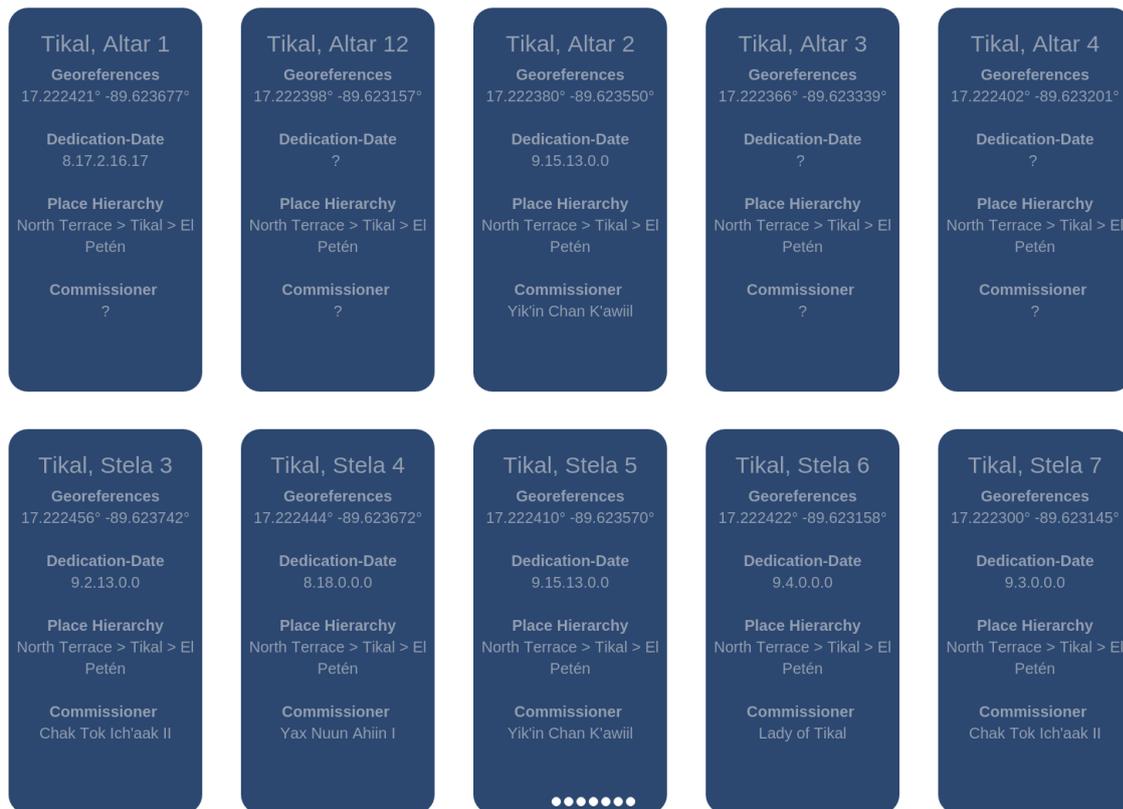


Figure 14: Representation of a SPARQL query for the research question about the site of Tikal. The results show artifacts from the North Terrace.

Triplestore

Tests of the triplestore Blazegraph, which the project had decided to use last year, indicated that it was incompatible with the basic application TG-forms. TG-forms constitutes the generic point of departure for generating the forms and fields for compiling metadata, including communication with a triplestore. At this point, TG-forms has to be adapted to the specific needs of each project. To facilitate the integration of Blazegraph, however, profound changes would have been necessary, which would not have been possible without further collaboration with other projects using TG-forms. The advantage of Blazegraph is that multiple versions can be parallelized. However, since our project has a server-controlled backup and thus does not require such high performance, the greatest advantage was given to compatibility of the source code, and thus to functionalities. Consequentially, the project decided to use the triplestore Fuseki.³¹

Bonn and Göttingen, March 1, 2016

Dr. Christian Prager, on behalf of the authors



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/>

³¹ https://jena.apache.org/documentation/serving_data/