Classic Maya Text Repository Project: An open-access collaborative platform for research and annotation of encoded hieroglyphic texts

White Paper
National Endowment for the Humanities
Office of Digital Humanities
HAA-268887-20
Awarded to the Unicode Consortium
Gabrielle Vail, PI

Contributing authors: Deborah Anderson, William Giltinan, Andrew Glass, Christine Hernández, Holly Maxwell, Maximillian Mermell, and Stephen White

April 12, 2022
Project Overview

The Classic Maya Text Repository Project is a collaboration between the University of California at Berkeley, the Script Encoding Initiative (SEI), the Unicode Consortium, the Latin American Library at Tulane University, the Florida Institute for Hieroglyphic Research, and other partner institutions. Funding was generously provided by a Digital Humanities Advancement Grant from the National Endowment for the Humanities.

The grant period extended from February 2020 through July 2021; post-grant activities (preparation of material for online publication) occurred from August 2021 through March 2022.

Conceptualization

To make Maya texts accessible to both expert and non-expert user communities, our proposal was funded with the goal of creating an annotated digital archive, in conjunction with several thousand images of the inscriptions, a sign list and glyph catalog based on the “font” being developed for Unicode, and an advanced open-source publication platform designed specifically for the Maya script (Mayan-READ). Our Level 2 project was based on previous work undertaken with funding from NEH (PR-253360-17) and the Unicode Consortium (through several Adopt-a-Character grants) to digitally render and annotate texts from the three Postclassic hieroglyphic codices, as well as selected Classic period sites (conceptualized and initiated by research collaborator Carlos Pallán, with later work done by Gabrielle Vail and Holly Maxwell). Our proposal called for incorporating additional texts from Classic period sites that were not included in the corpus at the time the grant was funded.

Background

Thousands of texts written in a hieroglyphic script by prehispanic Maya cultures have been preserved at sites throughout the Maya lowlands (i.e., in present-day Guatemala, Belize, southern Mexico, and parts of Honduras and El Salvador), in museums throughout the world, and in private collections. Texts may be painted, carved, or incised and are found on large-scale stone monuments (stelae, lintels, altars, ballcourt markers, etc.), the walls of buildings, painted polychrome pottery vessels, small objects made from stone, bone, wood, and other materials, and in screenfold codices made from fig-bark paper. While texts from Classic period contexts (ca. 250–900 CE) are recorded on most of these materials, fewer texts were produced in the Terminal Classic and Early Postclassic periods (ca. 900–1250 CE). Legible codices are known only from the Late Postclassic (ca. 1250–1519 CE), when few of the other forms were produced. Most researchers treat the Classic period and codical corpuses separately, not only because they are from distinct time periods, but because they focus on different subject matter: Postclassic codices largely emphasize astronomical, ritual, and divinatory subjects, whereas Classic period texts are more often of a dynastic or political nature.
The goals of the project included making a subset of Classic period and codical texts available online in a tool that links visual images of the texts with transcriptions and other information interactively, and to further the development of a font that can be used to represent such texts in native-character (“glyph”) format on computers and in databases. While other online resources exist that provide images and transcription data (a representation of the text in Latin characters), such tools have lacked an interactive capability to identify which portion of the text corresponds to which portion of the transcription; moreover, the underlying texts are stored (and displayed) as images and transcriptions, as opposed to native characters. By leveraging new technologies and a digital humanities approach, this project promises to revolutionize the study of—and accessibility to—prehispanic Maya text corpora.

**Project Funding**

The project was funded in part by a Digital Humanities Advancement Grant (HAA-268887-20) from the National Endowment for the Humanities; two Adopt-a-Character grants from the Unicode Consortium (AAC-Mayan-2019 and -21), and a Dan C. Hazen Fellowship from SALALM (Seminar for the Acquisition of Latin American Library Materials). SEI work on Maya hieroglyphs has received support from NEH grant PR-253360-17 and PR-268710-20 through the Universal Scripts Project at UC Berkeley. Any views, findings, conclusions, or recommendations expressed in this report do not necessarily represent those of the National Endowment for the Humanities, the Unicode Consortium, SALALM, or any of the partner institutions.

**Project Team**

Our international team included researchers, technical consultants/developers, graphics specialists, and advisors.

Principal Investigator: Gabrielle Vail (University of North Carolina, Chapel Hill)

Project Collaborator/Research Associate: Carlos Pallán (University of Bonn)

Technical Consultants: Andrew Glass (Microsoft); Stephen White (Università Ca’ Foscari di Venezia); William Giltinan (independent consultant)

Research Assistants: Holly Maxwell (SUNY Albany); Céline Tamignaux (University of Bonn)

Research Associate (Hazen fellowship): Ashley Burke (independent consultant)

Graphics Artists: Maximillian Mermell (independent consultant)

---

1 Formerly Holly Neville.
Advisors: Deborah Anderson (UC Berkeley); Christine Hernández (Tulane University); Lisa Moore (Unicode Consortium)

Concurrent work was undertaken at Tulane University’s Latin American Library to digitize rubbings of Maya monuments and sculpture from the Merle Greene Robertson Collection for inclusion in a digital collection hosted on the Tulane University Digital Library and the Classic Maya Text Repository. This initiative was directed by Christine Hernández, Curator of Special Collections at the Latin American Library, in collaboration with members of the Digital Production Group at the Howard-Tilton Memorial Library, contract photographer Jason Kruppa, and project assistant Diego Matadamas. Matadamas also assisted with translation of project materials into Spanish.

Previous work

The project was first conceptualized in 2015 during discussions between UC Berkeley’s Script Encoding Initiative (SEI) and members of the Unicode Consortium. These discussions focused on developing a Unicode proposal for the Maya hieroglyphic script. Such a font would allow Maya hieroglyphs to be displayed in a native format on computers and in databases, as opposed to being represented as either images or transcriptions. In 2017, the Unicode Consortium awarded Pallán an Adopt-a-Character grant to work on a catalog of glyphs from the Maya codices and create a preliminary script proposal. Beginning in 2017, discussions centered around determining the best approach for representing Maya glyphs in fonts on computers and mobile devices. The monumental inscriptions became the focus of study so that new glyph block patterns, or “quadrats,” could be identified. Additional work undertaken prior to the start of the Classic Maya Text Repository (CMTR) Project was funded through a 2019-20 Adopt-a-Character grant from Unicode to Vail and Maxwell (2019-2020), and a 2017-2019 NEH Preservation and Access grant awarded to Anderson (PR-253360-17) for the “Universal Scripts Project,” which included funding for Pallán. The project under discussion builds on the work funded by these prior grants.

Project Objectives

The long-term goals of our project include:

- enabling visual and annotated collections of ancient texts to be effectively managed, accessed, disseminated, researched, and preserved over the long term;
- pushing the boundaries of rendering this complex script on computers; and
- expanding the Unicode Standard repertoire to include the Maya script.

---

2 Unicode is the universal computing standard specifying the representation of text in all modern software. The SEI, established in the UC Berkeley Department of Linguistics in April 2002, was founded to help prepare formal proposals for the encoding of scripts and script elements not yet supported in Unicode (see Anderson 2021).
To achieve these goals, our level 2 award had the following objectives:

(1) to provide an open-access, online collaborative platform and repository of Maya hieroglyphic texts (Mayan-READ) from both Classic and codical sources, with an emphasis on Late Classic texts from monumental contexts, dating from c. 600-900 CE, and the painted hieroglyphic codices from the Late Postclassic period, likely painted in the 15th or early 16th century. The READ platform is an open source project that has been used successfully for researching a variety of ancient texts. It was selected for its versatility in offering researchers the ability to annotate and preserve text corpora, thereby maximizing accessibility for expert and non-expert user communities alike;

(2) to create online editions of the Dresden Codex and selected hieroglyphic inscriptions from the Western Maya region, along with catalogs with full listings of signs from these sources, to be published on the Mayan-READ open access platform; and

(3) to further development of a “font” of Maya glyphs comparable to those for Latin or Chinese characters with exemplars of each sign from both the Classic and codical corpora, organized in terms of a “glyphary,” or glyph dictionary.

**READ Development**

Through our ongoing collaboration with Andrew Glass and Stephen White, our project sought to expand, enhance, and modify the existing READ (Research Environment for Ancient Documents; see [https://github.com/readsoftware/read](https://github.com/readsoftware/read)) infrastructure to support Maya epigraphic research in a virtual environment and create a repository of Maya hieroglyphic texts for study and analysis (Mayan-READ).3 One of READ’s defining features lies in its flexibility to support editorial, paleographic, and lexicographic work on ancient documents from any part of the ancient world.4 Its innovative features include an advanced visual interface able to perform markup and direct linking of images, transcription, and analysis; the ability to handle parallel analyses in the form of editions of the same document; concurrent collaboration using the same platform; and semi-automatic generation of syllabary charts, among others (see Appendix A for more details).

While the READ platform has been successfully used with several ancient languages, the structure of the Maya hieroglyphic script is such that creation of a customized version of READ (Mayan-READ) was necessary. One major innovation of the Mayan-READ tool set is the incorporation of an advanced text-input method that enables real-time typing and accurate rendering of the Maya script. This system is being developed using a prototype Unicode-compliant OpenType font and a novel complex keyboard that supports and informs work undertaken by Pallán on a Unicode proposal for the Maya script, as well as font logic developed by Glass to support Egyptian hieroglyphic quadrats. The prototype font will adapt the technique

---

3 READ is a software system for the study of ancient texts on an open-access platform.
4 [https://ephilolog.hypotheses.org/401](https://ephilolog.hypotheses.org/401)
for Maya glyph blocks and will be fully Unicode conformant, so that it can be used in major
software applications once the Mayan encoding has been added to the Unicode standard. At that
time, it will be made available in other tools used by scholars and the public to research and learn
about Maya hieroglyphics.

Modifications to READ will include extending its image-annotation features to include glyphary
lookup at the individual symbol level (including variants), thus matching current workflows for
Maya decipherment, and will allow generation of paleographic charts. A more specific hierarchy
for our project workflow will be developed, where individual signs are mapped ordinally into
glyph blocks, glyph blocks into frames, and frames into registers, thus allowing for automatically
generated transcriptions and computer-assisted transliterations and translations of the Maya
hieroglyphs, including synchronized visualization between highlighted portions of text with the
images and vice versa (Fig. 1). Once the glyph boundaries are defined and ordered into glyph-
blocks, algorithms can be used to identify the sign-arrangement configurations (quadrats) that
most closely match each context (Fig. 2).

Implementation

The workflows described below allowed us to achieve the objectives of our Level 2 award. To
assist other project teams with planning and implementing their own digital humanities projects,
we assess what worked well and what strategies might have saved us time or helped with
integrating the research and technology aspects of the project.

Research workflow

The PI and members of the research team were responsible for acquiring and preparing images
of texts; selecting texts for segmentation and annotation; researching the sites from which the
texts were selected, as well as the architectural structures with which they were associated;
developing and inputting metadata; “processing” the texts (Fig. 3); adding to the already existing
concordance of signs (i.e., preparing images from existing catalogs or inputting them directly
from the database; updating records and/or developing new ones); selecting exemplars of each
individual sign to be drawn; and preparing vector images of the signs (Fig. 4). This work enabled
us to reach both our research and technical objectives, as outlined above in “Project Objectives.”

We tried several different workflows to segment and annotate the texts selected for this phase of
the project, initially working within the NcodeX system developed by Pallán for a previous
phase of the project; this was originally done in FileMaker Pro (FM) and later transitioned to a
web interface. The workflow that proved most reliable and efficient, however, proved to be
importing images of the texts directly into READ and doing the segmentation and annotation
there (see Fig. 3)—the method that was used with the almanacs that were included from the
Madrid Codex.
Over the years, scholars have developed a number of different systems to catalog Maya hieroglyphs. For the Classic period corpus, we concentrated on those developed by Eric Thompson (1962), Martha Macri and Matthew Looper (2009), and Alexandre Tokovinine (2019). Each unique sign was entered into a “concordance” table, where it was assigned a “photo catalog” number based on the system implemented by Pallán, and linked to the corresponding images and codes/numbers from the catalogs mentioned above. Within the FM and READ interfaces, phonetic values were assigned programmatically through a concordance lookup based on the photocat number. READ additionally offers the researcher the ability to select the appropriate option when more than one possibility exists. This feature was not available in the NcodeX web interface, as it had not yet been put in place.

The NcodeX system, used in the early stages of the project, presented certain challenges, stemming largely from the fact that it was still under development and undergoing design modifications while it was being populated. As a result, the process of making corrections and updates to the data input into the system could be problematic at times, and design modifications led to delays in data migration to READ. In the end, the data entered into FM was exported into Excel tables that could be more easily accessed and updated by the research team, and then imported into READ. We found as we moved into the final stage of our analysis that a productive workflow involved segmenting the glyphs and assigning their phonetic values in READ, and using the glyphary tool (described below; see Fig. 5) to link them to a particular sign. This also allowed the option of recording an identifier (catalog code in a system known to the researcher) for glyphs not yet in the glyphary, which could later serve as a means of expanding our sign list.

Moving forward, we have decided to move away from using photocat numbers as unique identifiers (since they are arbitrarily assigned based on input order) and instead use the robust catalog published earlier this year by Matthew Looper and colleagues (Looper et al. 2022), which contains drawings of all known signs and variants occurring in Classic and Postclassic contexts. This coding system is based on visual distinctions and includes both codical and Classic drawings of each sign, when both exist. As a result, whereas a photocat number is an arbitrary identifier, the Looper et al. system is organized based on a visual logic that makes the identifiers more meaningful and easier to locate.

**Mayan-READ**

Researchers studying different ancient languages have adopted different conventions for discussing and presenting transcriptions and transliterations of texts. The Gandhari version of

---

5 Glyphs in the Maya script may have multiple values. By way of example, Thompson’s sign T528 has the syllabic value /ku/, the logographic value TU:N ‘stone’, and is the nineteenth of the twenty named days—pronounced as Kawak in the early colonial period.

6 The phonetic information was not included in the FM export. It therefore had to be input by the researcher, or entered programmatically through a lookup on the associated photocat number.
READ, for example, was developed with an integrated transcription-transliteration used in Gandhari research with a specific embedding practice that uses brackets to signal information about the state of the glyphs on the artifact (called text critical markup, or TCM). This allows the transcription to include not only the transcription of the particular glyph, but also additional information about the specific example of the glyph in the photo or drawing of the text selected for inclusion. In recent years, a number of different standards for encoding editions of ancient documents have been developed (e.g., Leiden, Turfan, and TEI Epidoc), and variations of these are typically used by a particular research community. In an effort to account for these variations, READ translates various traditions into an internal code that can be expressed in many of the various markup practices. Because epigraphers working with Maya texts have their own standards of practice, inputting transcriptions from the Maya research community requires either preprocessing the pre-existing transcription data into a form that READ already understands, and/or changes to the parser code. Similarly, exporting transcriptions from READ in a form that can be used by the Maya research community requires changes to the presentation code and export functions. Early in the project, the READ team revised the parser to accept TCM input used in Mayan transcriptions. For some TCM, the input transcriptions were preprocessed by scripts and/or by the researchers to import into READ. After a couple of test imports, the team used READ’s image editor, without modification, to annotate glyphs on images for the imported text, followed by linking the annotations to the transcriptions.

Because sign segmentation and annotation was initially done in NcodeX, the READ team’s early efforts focused on testing and development of READ’s input processes for the new workflows and the identification of how to best present transcriptions in formats familiar to Mayanists. Once the data were available for import, a process had to be developed for migrating the data from NcodeX to READ. After creating a process for database format migration, White was able to create the migration script to import the image annotations that had been developed in NcodeX. A complication arose, however, due to the fact that metadata were stored both in the NcodeX database model and in manifest files on the IIIF server. This necessitated further revisions to the migration code, meaning that the import and consolidation of data in Mayan-READ ended up being more challenging than was originally anticipated. Ultimately, however, a large set of transcribed text was imported into Mayan-READ, with glyphs on the images annotated by boundaries labeled with location numbering of each glyph.

After the import was complete, the data required further enhancement as the source data (glyph annotations) did not include phonetic information needed to create the linked transcription for the Mayan-READ base system (those data were stored in another location and not integrated into the NcodeX export). At this point, of the two corpora (codical and Classic) intended for inclusion

---

7 The International Image Interoperability Framework, or IIIF, is a framework that allows image data to be combined with metadata for display over the internet. During the course of the project, the NcodeX developer elected to integrate IIIF with NcodeX, which resulted in certain data needed for Mayan-READ being stored in the NcodeX MySQL database and other data being stored in IIIF “manifest” files.
in READ, the project had a partial Classic corpus that contained glyph boundaries with location markers for approximately 150 texts that could be viewed in READ but needed curation. The project team members developed strategies to add the transcription phonetics to the Classic Mayan corpus outside of NcodeX, and to create a separate codical corpus leveraging source data previously developed by the Principal Investigator.

This phase of work focused on a subset of almanacs from the Madrid Codex housed in the mayacodices.org database developed by the PI and colleagues with funding from NEH grants RZ-20724-01 (July 2002 to December 2002) and RZ-50311-04 (July 2004 to December 2006). An accelerated workflow was established that included uploading almanacs (text and images) from mayacodices.org into READ, and necessitated the addition of an input field for image annotations—called segments in READ terminology—to READ’s image editor where the researcher could create the glyph boundary and add location information and phonetic values in a single input field. Once the glyphs for each almanac or “text” were checked by team members for both location and phonetic value, a script was created to convert the annotation labels into a linked transcription. The effort was highly successful and led to the creation of the Madrid corpus within the timeframe of several months.

Building on this success, the project team adopted a similar strategy for the Classic corpus, augmented by a “concordance” of Classic glyphs that was exported from work done in NcodeX earlier in the grant period to a spreadsheet that was curated and adapted for use in Mayan-READ. The texts brought into READ from NcodeX included information identifying individual glyphs with photocat numbers, which were included in the separate concordance table, along with their attested or proposed phonetic values. The team used this information to pre-fill glyph annotation labels with location information, as well as phonetic readings. The development team also made enhancements to READ by repurposing the Codical Glyphary READ plugin to create a Classic Glyphary that was attached interactively with the glyph selection in READ’s image editor (Fig. 6). That enhancement automated the lookup of the sign in the glyphary. Currently, the research team is in the process of curating the phonetic labels for the Classic Mayan corpus. After final checks, both the Madrid and the Classic corpora will be converted to a published version for public consumption, as envisioned in our project proposal, and READ will be available as a research platform that will allow for additional Maya texts to be added and enhanced over time.

Work on this project identified several philology workflows not present in other language domains for which READ has been used previously. That discovery has spawned model changes and presentation code changes that should improve the robustness of READ for the Mayanist community, and that may assist in future language projects. These include multiple varied transcription presentations in the form of content and structure to better match the Maya hieroglyphic writing system research output, particularly with respect to the glyph block–frame structure of Maya texts, and the physical layout of words, diplomatic and reconstructed transcription, and so forth. In the Gandhari system, for example, the identification of a glyph has a one-to-one mapping with its phonetic representation, whereas in Maya texts, multiple
mappings can arise. Additionally, analysis of Maya texts benefits from recording a code to uniquely identify each sign, in addition to the phonetic value, which allows a more fine-grained interpretation. Incorporating this information required additions to the data model and extra functionality for edition editing. Modifications required to facilitate these aspects of Maya text analysis created challenges for the project team, but ultimately resulted in a version of Mayan-READ that is more useful for scholars, and that provides a platform for future research efforts.

Font

During the grant period, the project team also collaborated on the development of three fonts. The first of these involved updating the codical font developed through a previous NEH grant to Anderson (PR-253360-17) to include a slightly expanded sign repertoire. The codical font was then enlightened with OpenType logic originally developed for Egyptian hieroglyphs so that glyphs could be arranged in the traditional blocks consisting of clusters of individual glyphs. This font contains size variants of the basic signs that were generated geometrically adapting the vector outlines. It is intended to be a basic functional font that can be used to show Mayan specialists the capabilities of OpenType fonts as a complement to their scholarly presentations (Pallán, Glass, and Tamignaux 2021).

We also partnered with Alexandre Bassi and Thomas Huot-Marchand of the Atelier National de Recherche Typographique (ANRT) in Nancy, France to investigate cutting-edge typographic techniques and ways they might be applied to a Mayan font. This work led to a proof of concept font that used standardized glyph outlines and a precise glyph grid that allowed glyphs to compose into glyph blocks so that the outlines would neatly align and overlap. This technique better approximates the script grammar of the Mayan writing system and contrasts with the empty space style used in Egyptian hieroglyphic clusters. Sign blocks were designed at different rotations and aspect ratios to reflect transformations and distortions that are attested in the Maya codices, reflecting aesthetic choices the Maya scribes would make as they composed glyph blocks.

Most significantly for the grant project, we created a Classic font as a proof of concept to investigate how the more complex Classic glyph outlines would participate in the cluster logic developed for the codical font. Sign drawings were done by Mermell and Tamignaux, in collaboration with the PI (see Appendix B for a discussion of the process used).

Glyphary

Work funded in part by NEH grant PR-253360-17 (June 2016-present; Deborah Anderson, PI) resulted in the development of a glyphary (illustrated glyph dictionary) for the codices through the collaborative efforts of Glass and Pallán. Under the auspices of our recent award, we extended the original codical glyphary to support a parallel database of Classic Maya glyphs. This Classic glyph set is based on the coding conventions developed by Macri and Looper (2003;
see also Macri and Vail 2009) but is also extensively cross-referenced to other cataloging schemes such as that developed by Eric Thompson (1962) and the more recent sign list of Alexandre Tokovinine (2017), as well as the “photocat” numbers generated for the NcodeX project. The glyphary is integrated into Mayan-READ so it can be hosted as a pane and generates an automatic look-up when a specific sign is selected in the baseline image in the READ interface (see Fig. 5).

Project Contributions and Next Steps

Accomplishments

Significant steps were made towards developing Mayan-READ as a virtual research environment (VRE) for Mayanists, resulting in the upcoming publication of selected almanacs from the Postclassic Madrid Codex and of Classic period texts from the Western Maya region, with transcriptions, translations, and annotations (see Fig. 6 for an example).

Additionally, the following were developed as part of this project:

- A repository of several thousand Maya texts from Classic and Postclassic sources, including 170 that are segmented to the sign or glyph block level, with most also including an identification of signs by a unique identifier; these will be made available on the READ platform for open-access use by researchers, students, Mayan language speakers, and other interested parties. This was done in collaboration with the Latin American Library at Tulane University, which is digitizing its large corpus of rubbings of Maya monuments (numbering almost 2000) created by Mayanist Merle Greene Robertson (see Vail and Hernández 2021).
- A concordance of Classic period glyphs with each glyph (sign) identified by its code or catalog number based on the catalogs most commonly used by researchers, as well as its phonetic, logographic, and/or calendrical value, and, for logographs, its English meaning.
- A Classic glyphary, or illustrated dictionary of glyphs, and syllabary incorporating approximately 40% of Classic Maya signs.
- A font with Classic signs from the Central, Western, and Southern regions (450 signs drawn). This font builds on work done for the codical font, which in turn had adapted successful methods used for Egyptian hieroglyphs.
- A quadrat tool that uses algorithms to discern the composition of glyph blocks (the number, arrangement, and size of the signs) so that the process of identifying quadrat type can be automated instead of being determined by the individual researchers (see Maxwell and White 2021).

Dissemination

The project’s results were made publicly available through a poster session at the Rocky Mountain Council for Latin American Studies annual conference in March 2021 (posters are
available in our Gather space at https://app.gather.town/app/SxCqb3WBBUhsXvMa/classicmayatexts and will soon be available at http://hieroglyphicresearch.org/cmtr.htm); a workshop with Kaqchikel Maya speakers presented virtually in July 2021; a virtual presentation at the annual meeting of SALALM (Seminar for the Acquisition of Latin American Library Materials) in July 2021; and demonstrations to colleagues and students at various institutions (including UNC-Chapel Hill and SUNY-Albany). Over the next several months, additional materials will be posted on the project website at http://hieroglyphicresearch.org and/or uploaded to the Unicode document register, with backups to be available on the UC Berkeley Department of Linguistics server.

Future work

Future enhancements of the READ system are contemplated to enable Mayan-READ to serve as a more productive virtual research environment for Mayanists. For example, using the two curated corpora—those from the Madrid Codex and the Western Maya region—will allow the transcription presentation code work to be finalized. READ has paleography tools that enable the researcher to view, compare, and categorize glyphs, which use sort codes and grouping codes to drive the paleography chart tool. The codes currently in use were chosen randomly to simplify code adjustments to test linked Mayan transcriptions with the paleography editor. These codes will need to be reviewed and updated as the research community decides how best to enable paleographic tools. The glyph tagging tool is designed to use a taxonomy agreed upon by the research community to customize the categorization of glyphs into groups specific to targeted research or to enable searching by special features. Developing this taxonomy is a priority for future work so it can be input into each Mayan-READ project database to replace the current generic set.

A new glyph catalog authored by the Maya Hieroglyphic Database (MHD) team (Looper et al. 2022) was published in the Glyph Dwellers series in January 2022. Its catalog numbers will need to be incorporated into the project data so they can become the unique identifiers for each sign. Collaboration with and participation from other Maya script projects, such as the MHD project, will be critical to the creation of a joint Unicode proposal for Maya hieroglyphs. We also intend to explore further collaboration with ANRT to create a Unicode proposal font, which would take advantage of the strides in glyph creation demonstrated by Alexandre Bassi.

Challenges and Evaluation

Several changes to the original work plan and budget were made and approved by NEH’s Office of Digital Humanities in April 2020 and May 2021 in order to deal with issues resulting from the pandemic, including the cancellation of conferences; the necessity of adapting to a purely virtual work environment and meeting format; and the difficulty of obtaining certain resources, including print publications and archival materials, because the institutions housing them were
closed or had restricted access. We nevertheless were able to accomplish the goals outlined in our initial grant proposal, with minor modifications.

Modifications included changing the venue for disseminating information (to allow for conferences that adopted a virtual format) and the necessity of making new digital editions of a corpus of images (rubbings) of Maya texts that were not available in sufficient resolution to be used in the online repository. The Latin American Library at Tulane University (the repository housing the rubbings) assumed most of the cost of photographing and creating stitched images of the rubbings, although we also sought separate funding to develop the associated metadata through a Dan C. Hazen Fellowship offered by the Seminar for the Acquisition of Latin American Library Materials (SALALM), an organization to which both the PI and Project Advisor Hernández belong. This funding was used to hire an additional project member to assist with image preparation and the capture of metadata. (For a more extensive discussion of the rubbings project, see Appendix C and Vail and Hernández 2021).

The most significant project challenges arose from the need to consolidate data from multiple sources, including several that were still in development while the project was ongoing (the team transitioned from a FM Pro database to the online NcodeX system to READ). Because the goal of the project was for the data to be implemented on the open-access, publicly accessible READ platform, the research team had to learn three different workflows, and the READ developers had to handle migration of data from NcodeX to READ, including changes in data formats that were not built into the initial migration scripts. Resulting delays affected our schedule and necessitated some workarounds, but the project data will be made publicly available within the next several months through hieroglyphicresearch.org, a site maintained by the Project Director and William (Ty) Giltinan.8

**Lessons Learned**

The project successfully demonstrated that READ could be modified to handle a logosyllabic script, which provides important implications for future work on other scripts, such as Aztec hieroglyphs. It further validated the value of using a proven, open source toolset to consolidate data on Maya texts from multiple sources and display those data in an interactive manner that will enhance usability for scholars and educators alike. Finally, the project resulted in valuable data and tools that can be leveraged in future research, and furthered the development of a glyphic font for Maya texts. The project team, however, had to overcome a number of difficulties to reach this point.

Reflecting on these, we offer the following recommendations to those undertaking digital humanities projects: (1) Ensure that an experienced software developer is in charge of hosting the project tools and data, and of managing the process through which tool versions transition from development to production use by the team members; (2) Host the most current version of 8 Backup copies will be stored on the UC Berkeley Department of Linguistics server.
all project data in locations managed by the grant administrator or principal investigator; and (3) Have all team members execute clear undertakings at the outset of the project addressing the use of prior-created and newly-developed data and materials to ensure that they are made available without restriction in an open-source environment.

Another valuable lesson is to be realistic in setting project goals and not to overcommit in order to secure funding. As a corollary, we recommend that a process be put in place to reevaluate the attainability of project objectives during each project phase. Not doing so may result in the necessity of team members having to donate extensive time to the project to ensure that the objectives are met.

In terms of the project workflow, selecting materials (in our case, images) and obtaining permissions in advance, where possible, is useful for scheduling and planning purposes and optimizes project time. In thinking through the process used for creating vectorized drawings of the Classic glyphs (see Appendix B), project members reflected that an initial outlay of time and resources to research and try different processes and tools may have led to a more efficient method of vectorizing. Streamlining the technical steps, knowing that they would be repeated approximately 500 times, could have had meaningfully affected the overall process.

**Impacts of Project**

**Summary of Outcomes**

Our Level 2 award led to the creation of a repository for Classic (ca. 250-900/1000 CE) and Late Postclassic (ca. 1250-1520 CE) Maya hieroglyphic texts in the form of line drawings, rubbings, and photographs. These are stored in Mayan-READ, an online open-access virtual research environment that enables visual and annotated collections of ancient texts to be effectively managed, accessed, disseminated, researched, and preserved over the long term. Building on prior work sponsored by the Script Encoding Initiative at UC Berkeley and the Unicode Consortium and funded in part by NEH awards, the project team worked with a subset of the texts in the repository to (1) prepare an illustrated dictionary or “glyphary” of Late Classic period glyphs from the Western, Central, and Southern Maya regions; (2) prepare drawings of approximately 60% of the glyphs represented in those texts, as a preliminary step towards developing a font of Classic Maya glyphs that will serve as the foundation for a proposal to Unicode so that Maya hieroglyphs can be included in the Unicode Standard and be easily typed using a standard computer keyboard; and (3) provide a textual and visual analysis of approximately 30 texts from the Classic and Postclassic periods in Mayan-READ to highlight the functionalities of this tool for use by researchers, students, educators, and Maya speakers. From a long-term perspective, the project is a critical first step towards making Maya hieroglyphs accessible to the general public. Users from all walks of life—from Maya speakers in the Yucatan and neighboring areas to school children and interested parties throughout the world—will be able to access Maya hieroglyph materials through Mayan-READ, to learn and study the
script, appreciate the cultural contributions of prehispanic Maya cultures, and even contribute to the project by adding additional materials.

**Impacts**

Having a curated digital repository in a virtual research environment enables new digital research methodologies. Of particular importance in this regard is READ’s enhanced functionality, which enables it to handle logosyllabic scripts. This has important implications for future work on other scripts, including Aztec hieroglyphs.

The development of a quadrat feature vector (see Maxwell and White 2021) provides a computerized method for similarity analysis, thereby enhancing research by Mayanists interested in the identification of scribal hands, as well as chronological and geographical variation. Such a tool also has implications for research in the field of pattern recognition.

Taken together, the newly enhanced tools will make it possible within the next few years to pair documentation of ancient monuments with entire digital renderings of glyphic texts, based on text input of Maya characters. Having these digital renderings online in an open-access platform will make the Maya corpus fully machine readable for the first time, while allowing researchers to study and discuss ancient texts, even when access to high quality images may be restricted due to institutional policies.

The work done by the project also has significant implications for Unicode encoding and lays the foundation for an eventual proposal to Unicode. The key to achieving our goal of being able to send, receive, and search text written in Maya hieroglyphs is to get the script into the Unicode Standard. Ideally, Unicode should support both codical (Postclassic) and Classic signs in order to represent a wide variety of Maya hieroglyph script materials. In building on earlier work on codical signs, our project added components for the Classic signs, namely, a font and a list of glyphs and information that identifies them. While additional Classic signs need to be incorporated, the work done by the project has helped lay the groundwork for an eventual proposal to Unicode. Once approved and published in Unicode, a working font will be required. The groundbreaking font prototype done for Mayan-READ has clearly demonstrated that such a font can be implemented in current operating systems, another key element for Unicode approval.

In conclusion, successful implementation of our project promises to bring searchability and usability of ancient texts to the next level, and has implications for user communities in a global context.
References Cited

Anderson, Deborah

Looper, Matthew, Martha Macri, Yuriy Polyukhovych, and Gabrielle Vail

Macri, Martha J., and Matthew G. Looper

Macri, Martha J., and Gabrielle Vail

Maxwell, Holly, and Stephen White

Pallán Gayol, Carlos, Andrew Glass, and Céline Tamignaux

Thompson, J. Eric S.

Tokovinine, Alexandre

Vail, Gabrielle, and Christine Hernández
Appendix A. Technical approaches envisioned for grant proposal.\textsuperscript{9}

Successful implementation of our project called for modifications to the existing READ infrastructure to make it capable of supporting epigraphic research of Maya hieroglyphic texts.\textsuperscript{10} Originally created to support languages written with alpha-syllabic writing systems, such as Gāndhārī (written with akṣara-based Kharoṣṭhī script), Pali, and Sanskrit, READ was later expanded to accommodate both fully alphabetic languages (epigraphic Latin) and more recently, logosyllabic ones, such as Egyptian and Maya hieroglyphs.

The database architecture at the core of READ uses a linked data model that represents individual entities such as syllables, words, and segments (e.g. the location of a glyph on the writing surface), along with the layers of interpretation identified by the researcher, such as physical lines of glyphs, grammatical structures, and lemma with attested forms. The system also manages the complexity of the linked data using software engineering techniques such as constraint systems, state machines, and state tables. This architecture effectively models the abstract structures used by manuscript and epigraphic scholars and gives READ the flexibility to accommodate the varying needs and standards of different research communities working with various ancient languages and scripts.

One of READ’s primary strengths lies in its ability to organize and recall robust image collections (e.g. different textual corpora), while facilitating mapping and annotating images and metadata. Once these collections are integrated on a single, comprehensive resource, scholars can access them and generate their own analyses and digital editions of the hieroglyphic texts on several levels, and connect them with the array of additional resources that we envision for the Mayan version of READ.

Through a series of applications and plugins (including the “catalog,” “dictionary,” and “lexicographer” tools), READ supports multiple workflows for text transcription, translation, and paleographic, phonological, morphological, and lexicographic analyses. Together these enable more nuanced consensual vs. interpretive text translations, as well as the ability to enable multiple interpretations (different editions) of a single document to coexist and be presented side-by-side, while carefully preserving attribution and authorship of each.

The Mayan-READ tool set is being developed to enable real-time typing and accurate rendering of the Maya script using a prototype Unicode-compliant OpenType font and a novel complex keyboard. Users will be able to connect to the platform and capture Maya texts by typing a sequence and pressing a conversion key to display the associated hieroglyph in the font. This

\textsuperscript{9} Contributed by Stephen White and Andrew Glass.
\textsuperscript{10} READ currently supports TEI-based workflows by exporting files to TEI compliant XML. Extending external integration is contemplated during Phase 2 (including the ability to import from TEI into READ).
system will enable arranging individual hieroglyphs to create and render arbitrary glyph blocks including very complex arrangements with as many as twelve signs in a single block. The dynamic block structuring leverages font logic invented by Glass to support Egyptian hieroglyphic quadrats.

Modeling successes in past READ projects, the creation of Mayan-READ was planned to occur in stages, the first of which includes the development of a prototype to understand the issues. The next stage involves altering the code and data tables to import diplomatic transcriptions, or direct encodings with critical markup which closely match the glyphs and layout on the artifact, done via READ’s parser. This results in the creation of a base edition, which is linked with the boundary annotations, described above, created on the facsimile of the artifact. Since the parser uses the encoding map to validate the input, this stage iterates between correcting the encoding map and correcting the input transcription. This repetitive process eventually results in a complete encoding map and curated transcriptions for the script—in this case, Maya hieroglyphs. The next stage comprises adjusting the various READ editors for any domain-specific practices pertaining to transliteration, lexicography, paleography, grammatical analysis, syntactic analysis, and presentation preferences. The final stage involves creating export formats and processes to allow data that has been verified and enhanced in READ to be exported for use in other projects. The overall process is intended to create a platform that supports future imports of textual data into READ, allows scholars to use READ’s editing and display tools to work with text corpora from multiple sources, and further allows for either linking of the interactive texts with other educational and research tools, or exporting of those data so that they can be used in other projects.
Appendix B. Process for creating vectorized drawings of Classic period glyphs.\footnote{Contributed by Maximillian Mermell.}

1. Starting with a rasterized image, open the glyph in Adobe Illustrator
2. Use the Image Trace tool with the following parameters set:
   1. View: Tracing Result
   2. Mode: Black and White
   3. Threshold: use this tool to determine how much black the Image Trace includes or
      excludes from the drawing
   4. Paths: use this tool to determine whether edges will have sharp edges (low) or
      smooth curves and complex angles (high)
   5. Corners: set closer to Less
   6. Noise: set closer to 1
   7. Method: Abutting
   8. Create: Fills
   9. Options: Snap Curves to Lines checked
   10. Options: Ignore White checked
3. Expand the traced image by going to Object > Expand…
4. Use the Eraser tool to remove unwanted peripheral shapes and noise
5. Using a combination of the Pencil, Eraser, and Smooth tools, refine the traced image and
   work to capture the original drawing as accurately as possible
6. Cross-hatching isn’t captured well by Image Trace - erase cross-hatching and recreate the
   lines using the Pen tool
7. Move the image to a standardized size artboard
8. Section the artboard in sixths - based on the glyph’s relative height in the text, scale the
   drawing accordingly
9. Export the glyph as an SVG and AI file
Appendix C. The Merle Greene Robertson Maya Rubbings Collection.¹²

The Merle Greene Robertson Maya Rubbings Collection provides digital images of almost 2,000 rubbings of carved monuments and architecture of the ancient Maya. The Latin American Library (LAL) of Tulane University is the repository for the corpus of original ink on rice paper rubbings of Maya relief sculpture and carved hieroglyphic texts created by Merle Greene Robertson (MGR). Merle trained as an artist and began her study of the ancient Maya as a teacher in northern California during the 1960s. With her husband, Lawrence Robertson, the couple began an odyssey that took them to Maya archaeological sites and museum collections across southern Mexico, Guatemala, Honduras, El Salvador, and Belize where they conducted archaeological research and documented the ancient Maya cultural fluorescence spanning the centuries of the Classic period (200 to 1000 CE). Amassed over a 40-year long career, Merle Greene Robertson’s corpus of rubbings records and preserves Maya monument and architectural art and hieroglyphic texts that were being lost at an alarming rate through systematic pillaging, the ravages of the tropical environment, and natural disasters.

Merle had always intended that the images of her Maya rubbings be available to the global community of researchers and scholars, and the MGR Maya Rubbings digital collection is created in alignment with that same goal. Moreover, this digital collection forms a crucial component of a larger project funded by a Digital Humanities Advancement Grant (HAA-268887-20) through the National Endowment for the Humanities. With funding from NEH and the Unicode Consortium, the Classic Maya Text Repository Project seeks to make Maya hieroglyphic texts accessible to both expert and non-expert user communities through the creation of an annotated digital archive. The digital archive will reside on an advanced open-source web platform (Mayan-READ) designed specifically for the Maya hieroglyphic script (described above).

The first phase will populate the collection with images of 50 rubbings and 62 line drawings of monuments from settlements located in the western and central Maya lowlands, which include the sites of Palenque, Caracol, Copán, Tikal, Naranjo, and Yaxchilan. Imaging and metadata work were supported by a Dan C. Hazen SALALM Fellowship (2020).

---

¹² Contributed by Christine Hernández.
Figure 1. Synchronized visualization between highlighted portions of the transcription of the text with the image and vice versa in Mayan-READ (rubbing of Yaxchilan Lintel 3 by Merle Greene Robertson; courtesy of the Latin American Library, Tulane University).

b. Glyph block D5 from the Tablet of the 96 Glyphs, Palenque (rubbing by Merle Greene Robertson). It includes four signs, the second one infixed in the first. Image courtesy of the Latin American Library, Tulane University.

c. Glyph blocks at E3-F6 from the Tablet of the 96 Glyphs, Palenque (rubbing by Merle Greene Robertson), showing a variety of different quadrat structures. Image courtesy of the Latin American Library, Tulane University.

Figure 2. Quadrat structures.
a. Tablet of the 96 Glyphs, with segmentation and location information imported from NcodeX by Stephen White.

b. Annotating the text with phonetic values (undertaken by Holly Maxwell and Gabrielle Vail).
c. Macri and Looper (2009) codes associated with the individual glyphs.

d. Phonetic values associated with the individual signs, following the annotation process.

Figure 3. Segmentation and annotation of texts in Mayan-READ.
Figure 4. Vectorized glyph drawings by Maximillian Mermell and Céline Tamignaux added to the glyph dictionary created by Andrew Glass.
Figure 5. Interactive glyphary tool, used to identify the highlighted sign in Lintel 25 from the site of Yaxchilan (rubbing by Merle Greene Robertson; courtesy of the Latin American Library, Tulane University).
Figure 6. Synchronized visualization between highlighted portions of the transcription of the text with the image and glyph in Mayan-READ (rubbing of Yaxchilan Lintel 3 by Merle Greene Robertson; courtesy of the Latin American Library, Tulane University).