



NATIONAL ENDOWMENT FOR THE

Humanities

OFFICE OF DIGITAL HUMANITIES

## **Narrative Section of a Successful Application**

The attached document contains the grant narrative and selected portions of a previously funded grant application. It is not intended to serve as a model, but to give you a sense of how a successful application may be crafted. Every successful application is different, and each applicant is urged to prepare a proposal that reflects its unique project and aspirations. Program guidelines also change and the samples may not match exactly what is now required. Please use the current set of application instructions to prepare your application.

Prospective applicants should consult the current Office of Digital Humanities program application guidelines at <https://www.neh.gov/grants/odh/digital-humanities-advancement-grants> for instructions.

Applicants are also strongly encouraged to consult with the NEH Office of Digital Humanities staff well before a grant deadline.

Note: The attachment only contains the grant narrative and selected portions, not the entire funded application. In addition, certain portions may have been redacted to protect the privacy interests of an individual and/or to protect confidential commercial and financial information and/or to protect copyrighted materials.

Project Title: Cuneiform Digital Library Initiative Framework Update

Institution: University of California, Los Angeles

Project Director: Robert Englund

Grant Program: Digital Humanities Advancement Grant, Level II

## 1. Table of contents

1. Table of contents	1-1
2. List of participants	2-1
3. Abstract	3-1
4. Narrative	4-1
4.1 Enhancing the humanities	4-1
4.2 Environmental scan	4-2
4.3 History of the project	4-4
4.4 Work plan	4-5
4.5 Staff	4-6
4.6 Final product and dissemination	4-6
5. Biographies	5-1
6. Budget	6-1
6.1 Budget	6.1-1
6.2 Budget notes	6.2-1
7. Appendices	7-1
7.1 Work plan	7-1
7.2 Work plan diagram	7-2
7.3 New features details	7-3
8. Letters of commitment and support	8-1
8.1 Commitment: Gregory Crane	8-1
8.2 Commitment: Stephen Tinney	8-3
8.3 Support: Jacob Dahl (University of Oxford, Oriental Institute)	8-5
8.4 Support: Jürgen Renn (Max Planck Institute for the History of Science, Berlin)	8-6
9. Data management plan	9-1

## 2. List of participants

### *Participants*

Robert K. Englund	PI	UCLA, CDLI Director
Émilie Pagé-Perron	Coordinator	University of Toronto, CDLI Co-PI
Saurabh Trikande	Volunteer	Amazon Inc.
Prashant Rajput	Lead Programmer	UCLA
NN1	Programmer	UCLA
NN2	Web Designer	University of Toronto
Gregory Crane	Support	University of Leipzig; Perseus Director
Stephen Tinney	Support	University of Pennsylvania; ORACC Co-Director
Jürgen Renn	Advisor	Max Planck Institute for the History of Science, Berlin; CDLI Co-Director
Jacob Dahl	Advisor	University of Oxford; CDLI Co-PI
Bertrand Lafont	Advisor	CNRS; CDLI Co-PI

### *User Testers*

46 self-selected prospective user testers from the users base of the Cuneiform Digital Library Initiative services.

### **3. Abstract**

The Cuneiform Digital Library Initiative (CDLI <<http://cdli.ucla.edu>>) is a 20-year strong international digital humanities project curating data and maintaining the largest database of artifacts inscribed with cuneiform writing from ancient Iraq and adjacent regions. The CDLI Framework Update is a consolidation project aimed at ensuring both the longevity of the CDLI data and interface, and at increasing access, usability, and accessibility to the information it curates. As part of numerous sub-projects, a wide array of technologies to provide software support have been used through the years. The CDLI FU will consolidate actual features into a framework structure and prepare new data displays, including machine readable outputs, to enhance information diffusion. This update will strengthen CDLI's digital structure, facilitating maintenance and future developments, and increasing access to information about ancient cultures to actual and prospective audiences, including the disabled.

## 4. Narrative

The mission of the Cuneiform Digital Library Initiative (CDLI) is to collect, preserve and make available images, text and metadata of all artifacts inscribed with cuneiform and related ancient Near Eastern scripts. It is the sole project with this mission and we estimate that our 331,000 catalogue entries cover some two-thirds of all sources in collections around the world. Our data are available publicly at <http://cdli.ucla.edu> and our audiences comprise primarily scholars and students, but with growing numbers of informal learners. The current proposal exclusively addresses the need to increase the sustainability and access of our digital assets, updating and enhancing the current software structure and interface of the project. Major software component updates are crucial to the persistence of the CDLI, and the redesign of the interface will increase access, usability and accessibility for humans and machines.

### 4.1 *Enhancing the humanities*

The CDLI, through its long history, is now part of the Assyriological discipline fabric itself. According to Google Analytics, the CDLI website is visited on average by 3,000 monthly users through 10,000 sessions and 100,000 pageviews. 78% of these users are recurring visitors. The CDLI collection and its related tools are used in varied ways, but a majority of users is seeking information about a specific text or group of texts; the CDLI informs them about where the physical document is currently located, when and where it was originally created and deposited in ancient times, what is inscribed on the artifact and where it has been published. Search results display artifact images and associated annotation when available.

Through this level II digital humanities advancement grant, we will update and enhance the current infrastructure of the CDLI, with the objectives of increasing sustainability and access of the initiative. At the end of the project, CDLI software will be easy to maintain, update, extend and modify by a wider array of contributors. The interface will be easier to use by more people and on more devices. It will be possible to access the data in more ways, by machines and humans, through the interface and downloads in different formats. Our data will be more compatible with other projects (such as the Open Richly Annotated Cuneiform Corpus<sup>1</sup>). Current users will see their interaction with the CDLI improve and we will be able to cater to new audiences. We will also make available all the code base of the CDLI to the community using GitHub and all new code will be released to the Public Domain.

Sustainability tasks concern both the backend and frontend of the initiative, comprising primarily the update of the project's code and interface. The CDLI is a rare long-lived humanities digital project amid initiatives that too often dead-end with incomplete or discontinued web sites. Over the years, it has seen additions and modifications to its code base, applied by numerous student programmers with different coding specializations. Both the server side scripts and the website framework follow the standards of previous versions of programming languages. As such, a code rewrite is essential for the sustainability of CDLI's software infrastructure. On the actual server, scripts written in various versions of Perl, Python and Java, coexist. These scripts are divided into two categories, "helper tools" and "automated tasks." Helper tools are, for example, file renaming tools and text transcription processing tools. The automated task scripts are responsible for extracting data from the database in specific formats for processing, backup, or sharing. CDLI's website comprises a robust model layer, and a logic and display layer. The

---

<sup>1</sup> <http://oracc.museum.upenn.edu/>

model layer employs Doctrine ORM's entity management<sup>2</sup> to communicate with the database, representing a powerful tool but one with a steep learning curve for new programming staff. The logic and display layer needs to be reorganized in classes with methods that are clearly delineated to facilitate adding or modifying the code without the risk of damaging other parts of the site. To address this, a PHP framework, CakePHP,<sup>3</sup> will be used for this rewrite. In some instances, it will be possible to slightly modify existing code and integrate it in well organized classes. In other instances, full rewrites will be necessary, but the logic and principle can be reused since the code will be written in the same programming language as the older version.

CakePHP was chosen in part because of this possibility to recycle parts of the current code, but also because back-end tasks will be executed in Python. Thus, robust actions can be supported by a language that is very popular in the scientific world and that permits text manipulation with libraries such as the Natural Language Toolkit, resources that are absent from PHP's potential. However, PHP is the most common and versatile language for server side execution for web page rendering, and CakePHP is a mature and popular framework born 11 years ago that has a strong community base with excellent support and is easy to learn by anyone with basic programming skills, including PhD-level humanists. By updating the actual content management system (CMS) used, Drupal, and by choosing a framework that can integrate parts of the code already in use, we will fully exploit existing project software.

Accompanying these updates, we will also move our main servers (web and database servers) to the Center for Digital Humanities (CDH) virtual servers with Docker technology<sup>4</sup> so we can easily "transplant" part of our operations; this is useful for instance in development, to recover from a software or hardware failure by easily reinstating saved copies, or when maintaining or creating new mirrors. A move to virtual will also dramatically increase sustainability, will diminish our operation costs and will facilitate maintenance both for the CDLI and the CDH.

The access part of the project is concerned with better catering to our actual audiences, but also opening up to *potential* ones. This can be achieved a responsive interface that conforms to web design standards<sup>5</sup> and accessibility guidelines<sup>6</sup> but also by providing features to display our data in new, more accessible ways, and through extensive documentation, and by sharing the code of the project and data bundles in more standardized and non-proprietary formats.

#### **4.2 Environmental scan**

No digital project is quite like the CDLI. The Perseus Digital Library<sup>7</sup> in the Classics is the project with the most similar goals. In cuneiform studies, the Open Richly Annotated Cuneiform Corpus (ORACC)<sup>8</sup> and the Madrid Database of Neo-Sumerian Texts<sup>9</sup> come to mind, neither of which is as accessible as the CDLI, utilizing frames and buttons for navigation, among others. Our technological approach is different principally because in our case, development and maintenance can be fully achieved by student researchers who are challenged with real life

---

<sup>2</sup> <http://symfony.com/doc/current/doctrine.html>

<sup>3</sup> <https://cakephp.org/>

<sup>4</sup> <https://www.docker.com/what-docker>

<sup>5</sup> <https://www.w3.org/standards/>

<sup>6</sup> <https://www.w3.org/WAI/intro/wcag.php>

<sup>7</sup> <http://www.perseus.tufts.edu/hopper/>

<sup>8</sup> <http://oracc.museum.upenn.edu/>

<sup>9</sup> <http://bdtms.filol.csic.es/>

problem solving, digital humanities practice, teamwork and leadership. Drupal has been our front end CMS for years and is as popular as Omeka for cultural heritage projects. The main reason we adopt the strategies proposed here is to be able to reuse, as much as possible, parts of the actual code core of the project, and to facilitate future maintenance and updates by staff with varied competencies.

### **4.3 History of the project**

CDLI's history reaches back to electronic efforts to edit and decipher the proto-cuneiform corpus of ancient Uruk dating to the last third of the 4th millennium BC funded by the German Research Association in the 1980's and 90's. Data from that project flowed into an expanded, US-led collaboration funded by DLI II, a joint program of the NSF/NEH, that in the period 2000-2003 completed the digital capture of major cuneiform collections in Europe and the US, established two online journals in cuneiform studies, and built persistent repositories of digital facsimiles of all accessible cuneiform text artifacts. Subsequent funding from the NEH and the IMLS allowed the CDLI to continue capture efforts, and to implement scalable access systems for a wide array of users, including researchers, museum staff, informal learners, and even international law enforcement. Support of the Mellon Foundation facilitated a dramatic expansion of CDLI's digitization partnerships and data management systems in the period 2009-2016. The CDLI-managed "Creating a Sustainable Cuneiform Digital Library" achieved the full capture of the Nineveh library housed in the British Museum, the Nippur collection of the University of Pennsylvania, and the Semitic Museum collection at Harvard, and many other collections worldwide. The CDLI now curates 331,000 artifacts, of which nearly half are associated with digital images and a good third with annotated transcriptions and translations.

The CDLI is hosted on physical servers maintained by the Center for Digital Humanities (CDH) at UCLA. Backups run daily on off server-center location on campus, to the University of Oxford, and to the Max Planck Institute for the History of Science in Berlin (MPIWG). The CDH offered to host our servers on their virtual machines, and a migration to this service is part of this Framework Update. The MPIWG, Oxford and UCLA all have promised long term support to the CDLI.

As part of the Framework Update project, a survey has been sent out to the community of users. Preliminary results in the form of 124 detailed answers guided us in preparing interface redesign objectives, new search options and result display. As part of this survey, we have secured beta testers for each design stage. The CDLI also has an issues board where we compile tweaks and small features to implement to increase usability. This means that the changes we will implement are explicitly based on user feedback and will clearly ease the research activities of actual users on the CDLI.

The Framework Update Project will secure the future of the Initiative by first making the data and interface sustainable and by reinventing its offer to facilitate the work of actual collaborators and users, and by renewing its usership by offering services catered to a wider audience, including the visually and physically impaired.

In the near future, depending on funding and resources allocation, the CDLI is pursuing three distinct projects in addition to its usual data service offer. First, we are, in a coordinated international effort among UCLA, the University of Toronto and the University of Frankfurt, developing a methodology for automated translation and information retrieval for transliterated

cuneiform texts, initially for the 67,000+ annotated Sumerian administrative documents of the 21st century BC. This project will result in a full set of translations for those texts, and substrate linguistic information will be linked and served through linguistic linked open data ontologies, and displayed with specialized views to users. The second project is concerned with optical text recognition of the cuneiform script. At this stage, we are using machine learning algorithms to recognize individual signs from each other. The next steps will align signs in the images with the transcription of the text in the display of such artifacts on the CDLI website, and will use this information to develop algorithms that will make it possible to actually recognize specific signs in pictures of cuneiform artifacts. Third, the CDLI seeks to initiate the digital capture of physical seals and clay sealings worldwide, and to develop and implement electronic tool sets dedicated to complex analytical forensics in glyptic, art historical and administrative history research. The CDLI Framework Update will make it possible to install these and subsequent projects on a solid, sustainable base.

#### **4.4 Work plan**

Annex 7.1 is a detailed work plan with individual description of work packages (WP) and annex 7.2 a timeline with associated tasks with leads and staff.

*Management.* Team management: Pagé-Perron and Rajput will both be assigned tasks to manage. Their responsibilities include segmenting work into manageable units and assigning those units to appropriate participants; ensuring that participants have the required tools and information to be able to work at all time, be available to discuss tasks with the team, assist as needed, and oversee quality of the technical aspects. Coordination: Pagé-Perron will coordinate all efforts, ensuring that deliverables are met on time and ensuring communication between teams. Direction: Englund will monitor the overall quality of the deliverables. He will also ensure that the project keeps in line with the global expectations of the CDLI, for instance, contributors to the project must be able to collaborate the same way as, or with greater ease than before the update; the interface aesthetic appeal must be maintained or enhanced, etc.

*Database update (WP-A).* Work on the database will involve creating an alternate relational optimized model of the database to facilitate leveraging the interrelation between information that is stored. This operation will necessitate the revamping of the way the Filemaker interface of the project works. Currently, a FileMaker server is used as a primary hub to update the artifacts catalogue (transcriptions are only in the MySQL database). With this update, contributors will see their FileMaker interface connect with the MySQL server instead so we can fully leverage its relational capabilities which are necessary for some of the new features (see WP-C), and to enhance search capabilities and display.

*Update of the code base and of the Drupal CMS (WP-C,D,F,G).* After preparing development instances of Apache-PHP-MySQL with a fresh install of CakePHP, Drupal 8 and the new database structure, comes the update of the code base, which is the most important part of this project, essential for the sustainability of the CDLI. Under PHP7, most of the code will be migrated to the CakePHP framework to decouple functionalities and layers of processing in order to increase modularity and make the code more stable. Git versioning will be used from the start of the process and will continue to be used after the framework update. Helper tools and automated task scripts will be migrated to Python in a toolset. All the code from these work packages will be thoroughly protocolled and documentation will be written at the same time directly in a Jekyll Github website to avoid rework.

*Update of the interface (WP-B,E).* With the prime objective of rendering the project data more accessible, the current interface will be revamped. To this end, a consultant designer will work on a set of wireframes for each view of the CDLI website, taking into account the needs of our current and prospective audiences. We already surveyed our users in order to get general feedback concerning the current interface and their answers will shape aspects of the new layout. After thorough review, this new design will be integrated to both the Drupal frontend and the general CDLI website to provide a seamless experience, a sense of identity, and to help users more easily find what they are looking for. As a basis for future developments in linked open data, the main search result views will also have alternate XML output for machine readability.

*New features (WP-C).* The new features for this project are varied and described more in detail in annex 7.3. They comprise mostly new displays for data that have already been gathered but are difficult to apprehend as is, for researchers and an interested public alike. Our search engine will also be enhanced along with the display and export of search results.

*Risks.* Changes in the team could necessitate a redo of the work plan and responsibilities attribution, but the fact that the project will run for just 15 months and that there is strong leadership by three confirmed participants will mitigate the effects of such problems. The geographic distance between project team members could bring communication problems, but Pagé-Perron has been collaborating successfully with the CDLI from a distance for the past six years, as have CDLI's European partners throughout its existence. As in all projects, there is a risk of facing an unexpected challenge or participants found wanting in efficiency; our experience heretofore with CS grad students, however, gives us little cause for concern. We do recognize that, despite best efforts, personnel turnover can cause delays, so that we have added a three-month "no-cost extension" to the work plan schedule (1/1-3/30/2019) to ensure completion of any unfinished tasks.

*Evaluation of results.* Quality checks will be performed throughout the life of the project, at three levels: tasks leads, coordinator and director. As such, deliverables will not be deemed complete if not meeting expected quality levels in terms of technique, documentation and performance / intended results. Additionally, all the public views of the interface will be verified for conformity with the W3C HTML and CSS standards, and a set of accessibility WCAG guidelines. Beta testers will be invited to try the updated interface and features at the end of the quarter 2 milestone and one month before the formal end date of the project so we can take into account their feedback throughout the project duration.

#### **4.5 Staff**

Robert K. Englund, Project director 10%

Englund is the director of the CDLI and has directed many digital humanities research projects as part of the initiative over the years, all with success and often exceeding expected outcomes. His deep knowledge of the data at hand and the processes involved in their management and display make him the best candidate for the role of director on this project.

Émilie Pagé-Perron, Project manager 25%

Pagé-Perron has been developing digital strategies for the CDLI for more than two years. She has experience in managing web projects (since 2009) and humanities projects (since 2012)

and is familiar with both the data and the software used at the CDLI. She will coordinate all efforts, lead on tasks and participate in software development.

Prashant Rajput, Lead programmer 25%

Rajput has been part of the CDLI team since September 2016. His successful experience with the initiative coupled with his expertise in both web application development and Python make him an ideal candidate as lead programmer.

Saurabh Trikande, External consultant 10%

Trikande will work pro bono for this project. Since he earlier worked with the CDLI as a UCLA computer science graduate student, he has good knowledge of the components and structure of the software. He will serve as backup programmer if the need arises, but more importantly will help us overcome unexpected challenges.

Graduate student researcher from CS 25%

Graduate student in computer science familiar with Python and PHP, Model-View-Controller (MVC) frameworks, and interested in the humanities.

Digital humanities graduate student consultant 532 hrs. total

Individual familiar with interface ergonomic design, W3C standards, and WCAG guidelines.

Their role will start with developing wireframes for the project views and will span through the full development of the interface.

#### **4.6 Final product and dissemination**

The final product of the Framework Update is a completely revamped CDLI software structure, presenting itself with a simple but efficient, usable and accessible unified interface to access the data of the project, coupled with the addition of specialized views and enhancements to make it easier to work with the project's data. Deliverables comprise: a restructured and optimized database where FileMaker can connect directly, a new responsive and intuitive interface conforming to web and accessibility standards, a fully rewritten code base using CakePHP with an updated, more flexible search function. Further, specialized views for different types of data, better automated data consistency and quality control and a more complete bibliography management system will emerge. Special attention will be given to follow W3C's HTML and CSS standards to increase accessibility. But going one step further, WCAG guidelines will also be adopted. For a full list of deliverables, see annex 7.1, and annex 7.3 for a description of new features.

Employing the usual channels of news diffusion in the field, we will explain the details of our update so that current and prospective users are well informed of the changes, why they were implemented and what impact they have on usability, access and accessibility. All code, and a complete documentation solution website built with Jekyll will be hosted on GitHub making this project effectively open source and fully documented.

To conclude, this consolidation project, by solidifying the software infrastructure of the CDLI and by increasing the accessibility and usability of the initiative's interface, will enable both the sustainability and preservation of the initiative for the years to come and offer better access to the data it curates to actual audiences and open up access to new audiences, including the visually and physically impaired.

## 5. Biographies

Robert K. Englund

Englund teaches at the Department of Near Eastern Languages and Cultures, Humanities Division, at the University of California, Los Angeles. He has conducted his major research on the proto-cuneiform texts from late 4th millennium BC Mesopotamia, and, as principal investigator of the project Cuneiform Digital Library Initiative (CDLI), Los Angeles / Oxford / Berlin, on the electronic documentation and edition of cuneiform generally. He is editor of and contributor to the online journals Cuneiform Digital Library Journal and Bulletin (CDLJ&B), and has collaborated with UCLA CS graduate student Sai Deep Tetali in the creation of an iPad app “cdli tablet” dedicated to informal daily presentations of cuneiform artifacts. Englund finished his BA at the University of California at Berkeley in 1977, and following a year of graduate work at the University of Chicago, moved to Munich, where he wrote his dissertation entitled *Verwaltung und Organisation der Ur III-Fischerei (The Administration and Organization of Ur III Fisheries)*. The thesis is concerned above all with the administration of Babylonian fisheries, focusing on an analysis of the accounting terminology in the Ur III period (ca. 2100-2000 BC) archives as a tool for understanding the organization and social position of fishermen and comparable state-dependent workers and supervisors of household economic units. He conducted post-doctoral research and taught at the Free University of Berlin in the 1980's and 90's, and moved to Los Angeles in 1996.

Émilie Pagé-Perron

Pagé-Perron is co-principal investigator at the CDLI and PhD candidate in Assyriology at the University of Toronto. Her research focuses on corpus and network analysis of large corpora of cuneiform texts to answer questions concerning social identity and history. She has worked as a web manager in the industry both as freelance and with FM Solutions in Geneva where she gained experience in leading teams. In 2010, she designed and managed the implementation of the intranet of Louis Dreyfus Company using Drupal 6 following a public request for proposals. Since 2011, she has collaborated with the CDLI and knows the ins and outs of its database and code, including the Drupal front end. She possesses advanced knowledge of Drupal, CakePHP, PHP, MySQL, HTML, CSS and the W3C standards.

Prashant Rajput

Rajput is a graduate student in the computer science department at UCLA and currently works as graduate student researcher for the CDLI where he is responsible for updating and maintaining CDLI's website which is critical to its proper functioning. He has also worked with website design and development in the past in which he built a website to analyze the spatial distribution of different categories of crimes in his native India. In this project he worked with HTML and CSS for website design, JavaScript and PHP for scripting, and MySQL as the database technology. His website helped users understand the types of crimes committed in their neighborhood, and also to know their distribution. He has worked in various team projects with different labs such as NYU Aerodynamics Lab and Laboratory for Advanced System Research (LASR).

Saurabh Trikande

Trikande is a machine learning engineer working in Catalogue selection and Search platforms at Amazon. He completed his MA in computer science with a focus on machine learning. During his tenure at UCLA he worked as graduate student researcher at CDLI and has since

been actively involved in the project. Prior to this, he worked at Smart Grid research Center at UCLA, UAV research in Defence R&D in India, Institute of Technology Management and NCL Innovation Labs. He has broad experience in building scalable and sustainable software solutions for research projects. His expertise and experience in software engineering will help push the CDLI to reinvent itself to more customer focused and industry standards.

7.1 Work Plan				
Group	Name	Description	Work Packages	Deliverables
A	Data & Database	Create an optimized relational schema for the database and connect with the PHP Framework and Filemaker	WPA1: DB schema and integration (+ bibliographic model)	DA1: Optimized relational schema for the data
			WPA2: Research method to connect FM with MySQL	DA2: Procedure to connect Filemaker directly to the MySQL database
			WPA3: Apply method designed in WPA2	DA3: Filemaker interface connectivity with the new database
B	Interface concept	Conceptualization of all views of the project	WPB1: Assess the needs of each audience and draft preliminary wireframe	DB1: Wireframes draft for the theme
			WPB2: Evaluate draft by CDLI associate and intended audiences and correct wireframes	DB2.1: Wireframes draft (2nd version) for the theme DB2.2: Set of recommendations and rules for ergonomics and accessibility in building the underlying views
C	Code core	Migration of all search and tools functionalities to the PHP Framework and unification of server scripts that run outside of the framework	WPC1: Model layer of the framework	DC1: Creation of the model layer of the PHP Framework
			WPC2: Search engine logic and result display update	DC2: Controller + views for search
			WPC3: Migrate user functionalities	DC3: User controller + views
			WPC4: Migrate admin functionalities	DC4: Admin controller + views
			WPC5: Calendar management	DC5: Calendar automatic generation and management logic and view
			WPC6: Sorting and searching tablets by date information	DC6: Advanced time search and sort feature
			WPC7: List of unknown signs	DC7: User management and view of unknown signs
			WPC8: Data processing, all backups, exports and file bundles and automations revised and unified	DC8: Unified backend scripts for backup and data processing (including catalogue csv automated daily export)
			WPC9: Glosaries management and display	DC9: Glosaries management and display
			WPC10: Sign variants and preferred readings management	DC10: Signs and their readings management, preferred readings view and search with variants permutation
			WPC11: ATF structure and contents checkers	DC11: Check ATF checker on demand and on upload
			WPC12: ATF template generators	DC12: ATF template generator
			WPC13: Bibliography management	DC13: Bibliography management system + display
			WPC14: Enhance views structures to conform to W3C	DC14: All views responsive with conform HTML & CSS, following major WCAG guidelines
D	Drupal core	Install Drupal 8, migrate all content from the Drupal 7 site and create a unified login scheme for the PHP Framework and Drupal	WPD1: Migrate to Drupal 8 including Drupal views	DD1: Front site migrated to Drupal 8
			WPD2: Migrate collections pages to Drupal	DD2: Collections pages in Drupal
			WPD3: Create node type and views for CDLP	DD3: Integrate CDLP in Drupal
			WPD4: Create a unified login for the whole CDLI site (Drupal + PHP Framework)	DD4: Unified login for all functionalities
E	Interface design & implementation	Creation of a unified theme for Drupal 8, search and tools, and integrate it in the site	WPE1: create template	DE1: Custom responsive and accessible theme
			WPE2: integrate in Drupal	DE2: Drupal template ready
			WPE3: integrate in the framework views	DE3: CDLI new interface template
F	Doc	Documentation and dissemination	WPF1: All code commented (inline and in documentation)	DF1: Commented code and online documentation for PHP code base (GitHub)
			WPF2: Documentation for the automated systems	DF2: Online documentation for backend code (GitHub)
			WPF3: Diffusion strategy and testing	DF3: Thoroughly tested interface and actual and prospective audiences advised of state of the project throughout
G	Servers	Set-up, tweaks and optimization	WPG1: Dev server installation	DG1: Docker installations on dev stations and a central dev server
			WPG2: Implementing portability tweaks, https and retired P# redirects	DG2 : containers more portable
			WPG3: Setting up a replication process for database and all files at Compute Canada, database and webfiles at Berlin and Oxford sites	DG3: Canadian mirror (website and backups), Berlin mirror updated, Oxford mirror



7.3 New features description		Deliverables	
Search	search fields	DC2	Implementation of new search fields that span across database fields (e.g. for artifact number, publication number (s), museum number, etc.) and with other options than exact string search, such as fuzzy search and rule based "forgiving" search. Additional features will include sign reading permutation when searching in the transliteration field.
	results display	DC6	Implementation of sort capabilities based on extended temporal information. It is currently not possible to sort texts chronologically by month, ruler and year, only alphabetically by ruler.
		DC2	Cross-field sorting of results: Sorting will also be updated so the user can sort search results according to more than one criteria.
Displays and management	calendars	DC5	A dynamic construction of the calendars of the 3rd millennium based on the individual analysis of texts made by specialists. The new calendar feature will dynamically extract the data associated with the different texts (month names, month number, ruler, year, period and place) and construct on the fly the calendar system based on this information. Managers will be able to modify the results. At this time, the calendar display is static. Display will be sortable.
	list data	DC7, DC9, DC10	The CDLI manages lists of signs, sign readings and preferred readings, unknown signs, glossaries, a two consecutive words dictionary and also lists of people and places. Management and display views will be prepared to facilitate working with those lists.
	bibliography management	DC13	The CakePHP add-on "PhpBibliography" will be adapted to manage bibliographic records related to artifacts.
Automated tasks	redirect retired entries to their counterpart in use		Using our database of outdated artifacts IDs, we will prepare a redirection system so citations using retired artifact IDs automatically lead to their actual entry.
	textual data consistency and quality check	DC11	A textual information checker will be implemented for uploads of transliterations by collaborators. We will also prepare our own transcriptions checker that can be used before submitting the text for addition to the database. The current template generator will be updated.

## 9. Data management plan

### Documentation and Metadata

As a companion endeavor to the Cuneiform Digital Library Initiative, cdliwiki<sup>1</sup> documents all aspects of the CDLI in a range of articles on history, specific inscribed artifacts, and genres, as well as discussions of processes and data acquisition. On the CDLI website itself, there are also articles discussing the museum collections holding the physical artifacts<sup>2</sup> as well as the terms of use of the data<sup>3</sup>. These tools will be used to help document the project and its outcomes. Awaiting the results of a French research group<sup>4</sup> working on the alignment of our texts' metadata with the CIDOC-CRM ontology, we will prepare an XML output for each search result view. It will be machine readable and easily modifiable to include CIDOC-CRM attributes. The software created through the project will be thoroughly commented, and a GitHub Jekyll website will serve as a code documentation hub directly in the same GitHub location as the code.

### Textual Data

The transcriptions and some text structural information of cuneiform inscriptions of artifacts are preserved using the "Canonical ASCII Transliteration Format" (C-ATF).<sup>5</sup> The text itself is encoded in UTF8 and the original language transliterations are restricted to simple ASCII characters. This notation system has been in use for 15 years and because of its simplicity and high level of standardization, many research projects base their work on the CDLI or will use a derivative of C-ATF. The ATF notation created by the CDLI is the widest-used standard in the field. With this project, we will continue to offer the input and output of transcriptions in this format but we will also store and offer to view and download data in an updated format where the transliterations lines of the text will also be UTF-8, and appropriate IPA characters will be used instead of derived ASCII transcription. This use of two standards makes the data usable by more people and thus will enhance its preservation. These textual data are currently viewable online and downloadable in text format. Because the CDLI is a long-lasting initiative, there are already quality checks and versioning systems in place. These checkers will be enhanced as part of the work plan. Each time a change is saved in one of the texts, a backup copy of the previous version is saved in the database.

### Licensing

New software and documentation generated by the project will be released to the public domain by using the Creative Commons license "Public Domain Dedication" (CC01.0).<sup>6</sup>

### Storage and Backup

During the research, GitHub will be used as a versioning system for the code base of the project. The Center for Digital Humanities (CDH) at the University of California, Los Angeles gives us technical support and external backups that increase the security and recoverability of the data. We also have a mirrors of the servers at the Max Planck Institute for the History of Science, Berlin (MPIWG); and through them at the Max Planck Society's persistent storage hub in Göttingen) and at the University of Oxford. Additionally, we expect news shortly from

---

<sup>1</sup> <<http://cdli.ox.ac.uk/wiki/>>

<sup>2</sup> See, for example, the page of the British Museum <<http://cdli.ucla.edu/collections/bm/bm.html>>

<sup>3</sup> <<http://cdli.ucla.edu/?q=terms-of-use>>

<sup>4</sup> <<http://triplestore.modyco.fr:8080/ModRef/>>

<sup>5</sup> <<http://oracc.museum.upenn.edu/doc/help/editinginatf/cdliatf/index.html>>

<sup>6</sup> <<https://creativecommons.org/publicdomain/zero/1.0/>>

Compute Canada concerning an application for resources allocation in order to set up a Canadian web site mirror and backups. These services come at no cost when allocated.

#### Preservation

By renewing periodically our agreements with the CDH, the MPIWG-Berlin and the University of Oxford, we are convinced that the CDLI offers optimal storage security and web server longevity; CDLI is in fact a model of data persistence—the longest lived digital humanities project in the field of Assyriology, with its predecessor the Uruk Project at the Free University of Berlin now 26 years in existence. Since the framework update project will increase the access of the data and interface of the CDLI, its usage will increase. For any eventual risk to the preservation of the software or the data, we will put copies of our work in official repositories to maximize their preservation.

#### Data Sharing

The code produced by this project will be released in the public domain and we will encourage anyone to use, modify and reuse any of its components. It will be available on GitHub with its accompanying full documentation.

#### Responsibilities and Resources

Because the CDLI has been running for many years, our lab is equipped with the needed material to undertake the framework update. Since some of our operations will be transferred to virtual servers at the Center for Digital Humanities at UCLA, costs will be impacted negatively where we will be able to repurpose two of our physical servers and retire another one, and we will no longer be required to maintain the retired physical server, or the material update of the servers now virtual. Our mirrors and backups are hosted for free at the Oriental Institute in Oxford and at the MPIWG. Each of these services are responsible for the maintenance and backup of their own servers.