Introduction

The Yale Babylonian Collection is one of the leading collections of Babylonian artifacts in the world, comprising over 45,000 items, including cuneiform tablets, cylinder seals, and other artifacts. However, the size and heterogeneity of the collection makes organizing it difficult. Efforts to digitize such collections have been stymied by disagreements over spelling, challenges in scanning artifacts, and more. For example, cylinder seals vary widely in size and material, making consistent 3D scans problematic. My 490 project aims to design a database schema and frontend to organize the Yale Babylonian Collection better.

Cuneiform Digital Library Initiative

The largest current project to organize Babylonian artifacts is the Cuneiform Digital Library Initiative (CDLI), a joint project of the University of California, Los Angeles, the University of Oxford, and the Max Planck Institute for the History of Science, Berlin. It is the reference database for most researchers in the field.

However, the database and frontend is outdated. An analysis of the data from the project repository reveals that it is a flat database with no relationships between tables. The server itself appears to be running on PHP 5.5, which was only supported until 10 July 2016.

Figure 1. Screenshot of CDLI full search page

---

1 https://babylonian-collection.yale.edu/
2 https://cdli.ucla.edu/
3 https://github.com/cdli-gh
As Dr. Émilie Pagé-Perron, a co-Principal Investigator of the CDLI put it:

The current CDLI schema is basically composed of three flat tables linked using the artifact id (no foreign key) and a couple of other tables containing textual analysis results (eg lists of words and the like). This simple design was put in place in order to provide a SQL copy of the FileMaker Database which is at the heart of the catalogue data, so it could be accessed through a web interface. We are currently in the midst of the full redesign of the database.  

Flat tables with no foreign keys makes searching and organizing the data especially difficult. Search on the CDLI website usually takes more than 10 seconds. Additionally, there is no straightforward way to discover relationships between objects on the frontend. For example, the “Period” attribute does not link to a page with other artifacts that originate from the same period.

Finally, the data needs to be cleaned further. Some columns are ambiguous, like “subgenre_remarks”. The “subgenre” column includes rows like “witness Archaic Vocabulary” and “Archaic Lu2 A (witness)”. The database could benefit from using enumerated types and better naming of columns.

On August 2017, the CDLI announced that it would be going through a core update to include “a unified interface for the project website, more powerful search capabilities, more internal links to navigate the catalogue, and intuitive displays for calendars, glossaries, bibliographies, etc”. As confirmed by Dr. Pagé-Perron, this involves a database design with proper relationships and foreign keys. Other projects that use more updated database schema include the Ur project at Penn (http://www.ur-online.org/) and the Diyala project at the University of Chicago (http://diyala.uchicago.edu).

Yale Babylonian Collection

The Yale Babylonian Collection currently uses FileMaker to store data of its collection. I am working with Professor Agnete Lassen to build a new database design and frontend system for the collection. The focus of the project will be on a specific subset of artifacts: seal cylinders.

Seal cylinders, as mentioned earlier, present unique challenges in 3D scanning and capturing images. At the same time, they lend themselves well to relational databases because cylinders were used to create impressions on tablets by different users, allowing for several relationships to be captured. Additionally, the contents of the cylinder seals, such as depictions of gods, iconographic elements, and more allow for finer-grained sorting of the data.

One academic analysis of the issue is “Progress Report: An Online database for the documentation of seals, sealings, and seal impressions in the Ancient Near East” by Christina Tsouparopoulou from the University of Heidelberg. It provides some historical background and describes the structure of the FileMaker database used for CDLI. It also elaborates on the frontend of the CDLI, which is based on the Zope web application framework. Finally, it gives some suggestions as to how various frontend and backend improvements could benefit research in the area.

Project

Based on this report and recommendations from the other database projects, I will design a suitable seal cylinder database for the Yale Babylonian Collection and build a working web application to access it.

Email correspondence.
Additionally, I will closely monitor the development of the new CDLI core and other projects to analyze how they designed their databases.

![Database Diagram](https://quickdatabasediagrams.com/)

**Figure 2. Preliminary design built on https://quickdatabasediagrams.com/**

Based on my analysis of the issue, a strong relational database is required, ruling out NoSQL systems like MongoDB. The backend will be a web framework like Flask. Finally, the frontend could be built in React or Vue, although JavaScript-based frontends might suffer from Search Engine Optimization issues. However, they are better suited for dynamic presentations of data.

The final goal of the project is to build a foundation for future development and use by researchers. At the same time, it must make decisions on which data points to exclude or restrict, such as spelling. These decisions will be analyzed in my final project report.