

# OCHRE DATA SERVICE

MILLER PROSSER and SANDRA SCHLOEN

---

## SWIMMING IN DATA

This year the OCHRE Data Service jumped into the deep end of the pool by embarking on an ambitious new project called CEDAR (Critical Editions for Digital Analysis and Research). The CEDAR project will use the OCHRE database to analyze multiple versions of a text to determine how it evolved over time and to allow users to create a critical text edition.<sup>1</sup> The case studies for this massive undertaking are the biblical book of Genesis, the Sumerian tale of Gilgamesh, and William Shakespeare’s “The Taming of the Shrew.” While these three texts may seem to have nothing in common, each in its own right has a complex history. One of the central goals of the project is to demonstrate that the same database tools and strategies can be applied to a wide range of texts. By studying the various copies of these texts, scholars wish to elucidate the transmission history of the texts, to understand how and why they evolved over time, and even to identify observable scribal practices. To achieve these research goals in a database environment requires a radical approach to data. In the case of the book of Genesis, our researchers wish to understand where manuscripts vary in the most minute vowel or accent mark. Therefore, each letter, vowel, accent, and punctuation has to be recorded separately for every single manuscript. Only then can manuscripts be compared on this detailed level.

How do we achieve this result in a research database environment? Time to dive into the pool; the “content pool.” To serve as a common backdrop for all texts of Genesis, for example, we developed the idea of a content pool. The content pool collects all known textual content of Genesis, including all variations from all attested manuscripts. As such, the content pool of Genesis ends up looking like a strange collage of Genesis. It does not read through smoothly as a text from beginning to end. That is intentional. Apart from the content pool, each attested manuscript of Genesis exists in the database as a text. The content of these actual Genesis texts is provided from the content pool of Genesis. Readers familiar with the Hebrew text of Genesis will recognize the opening phrase as בְּרֵאשִׁית, “In the beginning.” The first consonant is a ב (bêt). This consonant lives in the content pool. However, the very same letter *bêt* also appears in every other manuscript that attests this first word. In the database, we reuse the exact same letter from the content pool to represent the first letter of every attested text. This approach may seem overly complicated on its face, but from a computational perspective, it allows the database to understand how every manuscript of Genesis aligns or diverges. If the letter *bêt* is shared between two manuscripts, then the computer knows that these two manuscripts align at that point. This same alignment strategy then applies to every letter, vowel, accent, and punctuation throughout all manuscripts. This is all possible thanks to the content pool. The same strategy has been implemented for comparing dozens of copies of the Sumerian tale of Gilgamesh and is in the process of being applied to the various preserved copies of “The Taming of the Shrew.” In terms of scope and importance, this project has the potential to radically change the way manuscript studies are performed.<sup>2</sup>

Time to hop out of the pool and dry off for our next major achievement this year. This year, the OCHRE Data Service made significant strides in implementing advanced tools for data analytics and web publication.<sup>3</sup> Many of the projects using the OCHRE database platform have been focused on the

important tasks of gathering and curating data. Archaeology projects use OCHRE to record data in the field during an excavation, sitting right there at the edge of the excavation area in the hot and dry landscape of Turkey, Israel, or Peru capturing data as it is being unearthed. Philology projects create and edit editions of texts, create dictionaries, and identify people in the texts. These are all long-term tasks. However, now with an accumulation of data and years of work, many projects are ready to perform complex analyses and to make their data available online to the world. In this past year, the OCHRE Data Service has developed and implemented a number of powerful tools to facilitate these tasks.

First, we integrated the robust open-source statistical tool called R. The R statistical environment has been adopted widely in the Humanities and Social Sciences due its easy-to-use analytical libraries. With technical support from the University of Chicago Research Computing Center, the OCHRE Data Service can now provide access to R statistical libraries, allowing users to send live OCHRE data to an R-server to run various analyses, then to receive the results locally.

Two years ago, we reported on the new Geospatially-Enabled OCHRE, or GEO.<sup>4</sup> Since then, we have added greater functionality to this new tool. This OCHRE feature now allows for the complete integration of geospatial data and all other project data. For example, an excavator is able to click a spot on a map in OCHRE to indicate the find spot of a newly excavated object. This information is stored in the database, allowing the object to be understood in the spatial context of the rest of the site. With this new tool, maps and top plans can be created from live data. These maps can be viewed right in OCHRE or printed for reference in the excavation field. The entire process becomes a powerful system for recording and interpreting archaeological data more fluidly and dynamically than previously possible.

Finally, we have implemented the first iteration of an OCHRE API that allows projects to publish OCHRE data directly to a traditional website. Thanks to support from the Digital Library Development Center at the Regenstein Library, this API provides a stable URL for every item in the database. Every letter in every text, every piece of pottery from an excavation, every bibliographic item, and every scholarly observation can be published on a website and referenced with a stable web address. In layman's terms, this allows projects to create traditional websites to present and share data that is pulled live directly from an OCHRE project. See the following examples:

- A Digital Signary of Sumero-Akkadian Signs<sup>5</sup>  
<https://projects.rcc.uchicago.edu/ochre/signary/>
- Archaic and Classical Coin Hoards of the Ancient Greek World  
[https://projects.rcc.uchicago.edu/ochre/OCHRE\\_HARP/coin-hoard-site/index.html](https://projects.rcc.uchicago.edu/ochre/OCHRE_HARP/coin-hoard-site/index.html)
- The OCHRE Coin Hoard App  
<https://datacomestolife.maps.arcgis.com/apps/webappviewer/index.html?id=48a70a51aba44dbdabbc874c06e0c9eb>

In other exciting news, Sandra Schloen (OCHRE Data Service manager and OCHRE software developer) was selected as winner of a contest to present a user case study at the ESRI developer conference. Sandra's work with ESRI GIS tools is at the cutting edge of GIS and database integration. Also, Miller Prosser began a two-year term as a member of the newly formed ORACC Advisory Committee. ORACC (Open Richly Annotated Cuneiform Corpus) provides an online publication platform for ancient textual corpora. This new relationship will provide valuable opportunities for collaboration between ORACC and OCHRE.

In addition to thanking our existing projects, the OCHRE Data Service wishes to acknowledge the various new projects that have joined the OCHRE family. This year we are very excited to be working

with a number of new projects here at the University of Chicago, including John Novembre's project studying ancient DNA, Alan Kolata's project seeking to publish materials from his excavations in South America, María Cecilia Lozada's excavations at Corral Redondo in Peru, and Paul Sereno's archaeological and paleontological project at Gobero in Niger. Other new projects this year include the excavations at Tel Dan in Israel (Hebrew Union College/Nelson Glueck School of Biblical Archaeology, Grand Rapids Theological Seminary, Cornerstone University, UCLA and Southern Baptist Theological Seminary), at Hippos-Sussita in northern Israel (University of Haifa, Zinman Institute), at Idalian in Cyprus (Lycoming College), and a project to produce a digital companion to the forthcoming printed excavation report from Tell el-Borg in Egypt (Trinity).<sup>6</sup>

From pools to publication, the OCHRE Data Service is swimming in data this year!

## NOTES

<sup>1</sup> A critical edition is a version of a text compiled from multiple attested manuscripts. It presents a version of a text that probably never existed. It represents something of a complete idea of the text. Many of the texts from the ancient world are only partially preserved, requiring reconstruction from multiple attested documents. Other texts show interesting variations that reflect the living history of the text. A critical edition attempts to account for these many factors.

<sup>2</sup> We wish to acknowledge the hard work of the CEDAR project workers: Sarah Yardney (PhD'17), Joey Cross (PhD candidate, NELC), Andrew Wilent (PhD candidate, NELC), Arianna Gass (PhD candidate, English), and Ashleigh Cassemere-Stanfield (PhD candidate, English).

<sup>3</sup> We wish to acknowledge the contributions of two talented students in the College (now alumnae): Samantha Dean (BA'18) and Julia Oran (BA'18).

<sup>4</sup> Geographic Information Systems (or GIS) are tools that help represent geographic and spatial data on maps. See [https://oi.uchicago.edu/sites/oi.uchicago.edu/files/uploads/shared/docs/ar/11-20/15-16/ar2016\\_OCHRE.pdf](https://oi.uchicago.edu/sites/oi.uchicago.edu/files/uploads/shared/docs/ar/11-20/15-16/ar2016_OCHRE.pdf).

<sup>5</sup> This site also provides a link to a digital list of Gardiner's Egyptian sign list with graphical tutorials on how to draw each sign.

<sup>6</sup> We wish to acknowledge the contribution of OCHRE specialist Andrew Wright (MA'17) whose archaeology expertise and attention to detail have made it possible to support these many archaeological field projects.