

Democratizing ancient Mesopotamian research through digital scholarship

Raquel Alegre
Research IT Services
University College London
 London, United Kingdom
 r.alegre@ucl.ac.uk

Anastasis Georgoulas
Research IT Services
University College London
 London, United Kingdom
 a.georgoulas@ucl.ac.uk

Stuart Grieve
School of Geography
Queen Mary University of London
 London, United Kingdom
 s.grieve@qmul.ac.uk

Eleanor Robson
Department of History
University College London
 London, United Kingdom
 e.robson@ucl.ac.uk

Abstract—Since the 19th century, historians and archaeologists have compiled transliterations and translations of surviving cuneiform texts from the Middle East area, documenting the ancient history of the region, c. 3000 BC–75 AD. The Open Richly Annotated Cuneiform Corpus (Oracc)¹ is an international collaborative effort to gather and digitise a complete collection of cuneiform texts and their translations, with the goal of making them available to researchers and students worldwide. Oracc was developed ten years ago around the core value of ensuring accessibility to a broad audience, rather than a select group of experts. This principle presented new technological challenges, but has equally offered important benefits.

Initial transliteration of cuneiform tablets into the ASCII Transliteration Format (ATF) was performed using an Emacs plugin, the use of which was challenging for novice and experienced users alike. This precipitated the development of Nammu [1], a dedicated editor for files written in ATF, to provide a consistent environment for users to contribute to Oracc projects. This is an important step in the democratization of this research as it lowers the technological expertise required to join the platform, and reduces the amount of time needed to train new users, which was previously a large drain on Principal Investigators' time and resources. Nammu in turn takes advantage of pyORACC [2], a bespoke library developed for parsing ATF files and a key enabler of automation in the project.

Separately to the editing considerations, the Oracc website hosts the body of information editions and translations that researchers from different groups have accumulated during their work. An important aspect of this is the search capability it offers, allowing a user to retrieve information about a subject or term of their choice. A new version of this functionality is being developed, using the Elasticsearch platform to index and efficiently search large bodies of text. Users can choose to query the compiled glossaries, looking for words with a particular meaning, or for the meaning and appearances of a transliterated cuneiform term. Alternatively, they will be able to search through the information pages for a topic of their choice, effectively using the website as a domain-specific search engine. This dual functionality has been chosen so as to make the search of interest to both domain experts and the general public.

Early versions of Nammu focused on the transliteration and translation of cuneiform into English and other European languages. Meanwhile, decades of war and political instability across the Middle East have prevented researchers from Iraq, Syria and neighbouring countries from contributing to the ancient history

of their region, and excluded local communities from benefiting socially, economically or intellectually from that research. To address this pressing issue, the latest developments of Nammu have focused on the introduction of support for right to left languages such as Arabic, Kurdish and Farsi. This has required the redesign of the software to allow the interleaving of the left-to-right ATF transliterations and right-to-left language translations. Similarly, the new website search is being developed with an international audience in mind, particularly from the Middle East. These concerns are central to the Nahrein Network², which is driving the next step in Oracc's development. The Network's core mission is to foster the sustainable development of history, heritage and the humanities in post-conflict Iraq and its neighbours through collaborative, capacity-building research. Naturally, this involves establishing a dialogue with local scholarly communities in order to identify requirements particular to the region (such as the means of accessing digital content and any related challenges). These are then taken into account in all aspects, including software development. The digital outputs of Oracc play a crucial role in the Nahrein Network's effort, by enabling access to data and tools for communities internationally.

Work on the project involves professional software developers with scientific experience (*Research Software Engineers*) collaborating with academics from the domain. This collaboration grew organically as the scale of the project increased beyond what the original contributors could support, making the need for automation and more sophisticated technical solutions clearer. This practice continues to the present, with academics describing what features are required, advising on future developments, and being informed by the developers on their technological choices, while also providing domain knowledge when deeper comprehension is required or beneficial. Maintaining this open dialogue and understanding between the two sides has been key to the project's success and sustainability. In keeping with the spirit of openness, one of the core decisions has been to use existing standards (such as XML and JSON) as much as possible, release the source code of developed tools³, and provide detailed documentation for interested parties. These practices have led to users from other groups not only adopting the software, but also contributing to its development.

REFERENCES

- [1] Nammu 1.0.0 doi:10.5281/zenodo.1414437.
- [2] pyORACC 0.1.0 doi:10.5281/zenodo.1414454.

²<http://www.ucl.ac.uk/nahrein/>

³Code is available at <https://github.com/oracc/>

Programming work on Oracc is funded by UCL's School of Social and Historical Sciences, and through the Nahrein Network's grant from the UK Arts and Humanities Council's Global Challenges Research Fund.

¹<http://oracc.org>