Jamie Novotny and Karen Radner

11 Official Inscriptions of the Middle East in Antiquity: Online Text Corpora and Map Interface

Abstract: The LMU Munich-based Official Inscriptions of the Middle East in Antiquity (OIMEA) project is one of the two principal, digital text corpora of the Munich Open-access Cuneiform Corpus Initiative (MOCCI), which is a freely accessible digital humanities umbrella project established by Karen Radner and Jamie Novotny in the fall of 2015. This international project – which includes research partners in Philadelphia, Barcelona, and Rome – aims to edit all available official inscriptions of ancient Middle Eastern polities, recorded in the cuneiform script and contemporary writing systems, in a freely accessible, fully lemmatized (lexical and grammatical data tagging), and completely searchable format via the Open Richly Annotated Cuneiform Corpus (Oracc) project. In addition, OIMEA plans to make geo-referenced text editions available through its Ancient Records of Middle Eastern Polities (ARMEP) map interface, which is developed in collaboration with LMU’s Center for Digital Humanities.

Keywords: Assyria, Babylonia, geo-referencing, lemmatization, map interface

11.1 Introduction

In September 2015, the present authors founded Official Inscriptions of the Middle East in Antiquity (OIMEA) as part of the newly established Chair for the Ancient History of the Near and Middle East at LMU Munich. The aim was to widely disseminate, facilitate, and promote the active use and understanding of royally-composed texts of ancient Middle Eastern polities in academia and beyond, and to begin creating new and innovative ways for users to access the important and varied contents of these geo-referenced and linguistically-annotated (lemmatized) ancient records.¹

¹ MOCCI [http://www.en.ag.geschichte.uni-muenchen.de/research/mocci/index.html] and OIMEA [http://oracc.museum.upenn.edu/oimea/index.html]. MOCCI’s other digital text corpus, Archival Texts of the Middle East in Antiquity (ATMEA), which presently consists only of State Archives of Assyria online – SAAo [http://oracc.museum.upenn.edu/saao/index.html], is not discussed here. A third LMU cuneiform text corpus, Electronic Babylonian Literature (eBL), will be developed by Enrique Jiménez starting in 2018.

Jamie Novotny, Karen Radner, Ludwig-Maximilians-Universität, München
Given our research interests and past project affiliations, our initial focus was naturally the Akkadian and Sumerian inscriptions of Mesopotamia, primarily the self-aggrandizing texts of the kings of Assyria and Babylonia from ca. 1157–539 BCE. Thus, we initiated the following three projects: Royal Inscriptions of Assyria online, Royal Inscriptions of Babylonia online, and Inscriptions of Suhu online. In late 2016, in order to expand the dataset beyond Mesopotamia proper, work began on two other inscription-focused projects: the Electronic Corpus of Urartian Texts – eCUT, which contains cuneiform texts written in the Urartian language from Eastern Turkey, Armenia and Northwestern Iran; and Achaemenid Royal Inscriptions online – ARIo, which includes cuneiform texts written not only in Old Persian, but also in Elamite and Akkadian, chiefly from Iran.

All five projects include on the one hand, informational portal pages with details about the rulers in whose names these texts are written, their polities and the texts themselves; and on the other hand, the linguistically-annotated (lemmatized) editions with translations into English and, depending on the heritage data, also other European languages (German for RIBo), as well as the glossaries created from these editions. The text corpora are either (retro)digitized or newly created using software developed by Steve Tinney (Philadelphia) and are hosted on the Open Richly Annotated Cuneiform Corpus (Oracc) platform.

Lastly, in December 2016, the present authors, together with staff of the LMU’s Center for Digital Humanities directed by Christian Riepl and Stephan Lücke, began developing a map interface designed to display places where ancient texts were discovered or that are mentioned in those ancient sources. The main purpose of this interface is to allow users access to Oracc-hosted texts directly from the map. Ancient Records of Middle Eastern Polities 1.0 was made public in December 2017. ARMEP’s gazetteer feature, which displays cities mentioned in ancient sources, has not yet been implemented and will be part of version 2.0, which is to be developed in 2018.

---

2 RIAo [http://oracc.museum.upenn.edu/riao/index.html].
3 RIBo [http://oracc.museum.upenn.edu/ribo/index.html].
4 Suhu [http://oracc.museum.upenn.edu/suhu/index.html].
5 No URLs are provided for ARIo and eCUT as neither are yet publically accessible. Both are planned for release in 2018.
6 The count is six, when one includes the University of Pennsylvania-based Royal Inscriptions of the Neo-Assyrian Period (RINAP) Project (directed by Professor Grant Frame). Radner has been a member of that project’s editorial board since its inception in 2008 and Novotny has been a principal content contributor to both its printed books and its freely accessible online content since 2009; he is currently preparing new, online editions of the royal inscriptions of Assyria’s last great king Ashurbanipal and his lesser-known successors (668–612 BCE).
7 ARMEP [https://www.armep.gwi.uni-muenchen.de].
This paper will briefly discuss OIMEA and ARMEP, as well as address some methodological problems and technical issues in their creation and the future prospects of these two projects.

### 11.2 Overview of OIMEA and Its Sub-Projects

As is obvious from its name, the scope of OIMEA is official inscriptions, primarily from Middle Eastern polities of the first millennium BCE. The idea was inspired by the now-defunct, Toronto-based Royal Inscriptions of Mesopotamia – RIM Project. This project had attempted to edit, in a single place, royal inscriptions written in the Akkadian and Sumerian languages. Our Oracc-hosted umbrella project and search tool, which intends to go beyond the scope of the RIM Project, currently comprises six projects:

- Corpus of Kassite Sumerian Texts,\(^8\)
- Electronic Text Corpus of Sumerian Royal Inscriptions,\(^9\)
- Royal Inscriptions of Assyria online (LMU Munich),\(^10\)
- Royal Inscriptions of Babylonia online (LMU Munich),\(^10\)
- Royal Inscriptions of the Neo-Assyrian Period online,\(^11\) and
- Suhu (LMU Munich)\(^11\)

The presently available texts on OIMEA are all written in the Akkadian and Sumerian languages and in cuneiform. Starting in 2018, the project will include corpora of texts written in other languages; for example, monolingual Old Persian and trilingual Persian, Elamite and Akkadian inscriptions will be included on ARIo,\(^12\) and monolingual Urartian and bilingual Urartian and Assyrian inscriptions will be accessible via eCUT.\(^13\)

---

8 CKST [http://oracc.museum.upenn.edu/ckst/index.html], University of California Berkeley.
9 ETCSRI [http://oracc.museum.upenn.edu/etcsri/index.html], Eötvös Loránd University Budapest.
10 RINAPo [http://oracc.museum.upenn.edu/rinap/index.html], University of Pennsylvania and LMU Munich.
11 Suhu contains retro-digitized and lemmatized editions of the officially commissioned texts of the extant, first-millennium-BCE inscriptions of the rulers of Suhu; these texts were published in Frame (1995, pp. 275–331). The open-access transliterations and translations were lemmatized and updated by Alexa Bartelmus.
12 The contents of ARIo are based primarily on Schmitt, 2009, as well as data provided by Matt Stolper (Chicago) from his now-defunct Achaemenid Royal Inscriptions project [https://oi.uchicago.edu/research/projects/achaemenid-royal-inscriptions-project]. ARIo is currently managed by Henry Heitmann-Gordon.
13 The contents of eCUT are based on Salvini (2008–12). Birgit Christiansen is currently retro-digitizing, updating, and lemmatizing, as well as translating into English, that corpus of inscriptions.
The OIMEA hub on Oracc primarily serves as a multi-project search engine that enables anyone interested in the genre of official inscriptions to simultaneously search the translations, transliterations, catalogues, and portal pages of every available project on which ancient inscriptions are edited. As an informational and search hub, OIMEA strives to make the vast and varied corpora of inscriptions easily and freely accessible to every scholar, student, and interested member of the general public. Moreover, it enables its users to efficiently search that rich genre of ancient records, allowing them, for instance, to perform searches both on the transliterations and on the translations.

To give the readers of this volume a better idea of some of the content produced by OIMEA, two of the LMU-based projects, RIAo and RIBo, will be briefly described here.

11.2.1 Royal Inscriptions of Assyria Online

RIAo is intended to present up-to-date editions of officially commissioned texts of the rulers of Assur and later Assyria from the end of the third millennium BCE to the fall of Nineveh in 612 BCE; it also includes numerous informational portal pages that provide the historical and cultural contexts of these important ancient sources. The project started in September 2015 and Phase 1 of the website was made available to the public in early 2016, when the project’s initial goal – the retro-digitization of 866 Assyrian inscriptions published in three discipline-standard monographs (Grayson, 1987, 1991, 1996) – had been realized; these texts date from the end of the third millennium BCE to 745 BCE. Phase 2 of RIAo was completed in February 2018. That stage included the full lemmatization (lexical and grammatical data tagging) of every available text, as well as the completion of glossaries of the Akkadian words and proper names (gods, people, places, and temples) appearing in those 866 inscriptions and the writing of numerous informational pages on Assyria’s many rulers; this work was principally carried out by Nathan Morello. The project is now entering Phase 3, which will consist of:

---

14 A new “pager” interface is being developed for Oracc and its inspiration is based on the conceptual and easy-to-search functionality of OIMEA. Oracc’s “Neo” interface will allow users to easily and efficiently navigate and search all of the publically available texts on Oracc.

15 For example, if one searches for “lion”, 91 matches are found in inscriptions from Early Dynastic times to the Neo-Babylonian Period; and if one searches for “scribe”, 209 matches are displayed for Sumerian and Akkadian texts written in the third, second, and first millennia BCE.

16 For further details, see [http://oracc.museum.upenn.edu/riao/abouttheproject/index.html]. RIAo’s focus is restricted to texts written in the Akkadian language, in the cuneiform script. It does not include information or access to documents pertaining to Modern Assyrians. For such a project, see the Modern Assyrian Research Archive Project [http://assyrianarchive.org/database/home/].
incorporating the material published by the Royal Inscriptions of the Neo-Assyrian Period (RINAP) Project (directed by Professor Grant Frame and based at the University of Pennsylvania in Philadelphia) into the dataset;

– adding score transliterations of inscriptions known from more than one exemplar;

– supplementing composite editions with individual object transliterations when these are accessible for study (in the form of photographs, hand-drawn facsimiles, or in a museum or private collection);

– writing additional portal pages on rulers and their inscriptions; and

– preparing a comprehensive bibliography of Assyrian royal inscriptions.

When RIAo is finished, it will contain fully lemmatized and completely searchable editions of the approximately 1,800 Assyrian inscriptions. By 2020, the complete corpus of Assyrian inscriptions will be easily accessible to scholars, students, and the general public. Anyone interested in Assyrian culture, history, language, religion, and texts will be able to efficiently search any Akkadian and Sumerian words appearing in the inscriptions and any English word used in the translations.

Content undergoes strict scientific control. Unlike community-built sites such as Wikipedia, Wikidata, Pelagios, and Pleiades, RIAo’s contents cannot be created or edited by anyone. This is the sole responsibility of the core OIMEA team (presently Morello and Novotny), with input from OIMEA’s international editorial and advisory boards. The present authors, as the directors of MOCCI, assume content and editorial oversight of the project. We do welcome/encourage feedback from our community of users.

11.2.2 Royal Inscriptions of Babylonia Online

RIBo, which was also founded by the authors in September 2015, intends to publish in a single place fully searchable, lemmatized editions of all of the known Akkadian and Sumerian royal inscriptions from Babylonia that were composed between 1157 and 64 BCE, together with informational portal pages and complete glossaries of Akkadian and Sumerian words and the names of gods, people, places, and temples. The scope, when compared to RIAo, is much smaller. By the time RIBo is completed,


18 However, neither the scholarly community nor general public make much use of the possibility to contact the project by email (via the “About the Project” pages).
which is anticipated to be in 2022, that open-access project will contain about 400 inscriptions.\footnote{The dataset will include Frame, 1995; Weiershäuser & Novotny, 2019, 2020, 2022. Weiershäuser & Novotny, 2019, 2020, and 2022 will appear in the newly-established Royal Inscriptions of the Neo-Babylonian Empire (RINBE) series, which is co-edited by Radner and Frame, managed by Novotny and published by Eisenbrauns. Some of the inscriptions to appear in those three volumes have already been published in Schaudig, 2001; and Da Riva, 2009, 2012, 2013.}

Unlike RIAo, which comprises a single text corpus, the contents of RIBo are divided into several sub-corpora, generally grouped by “dynasty” or period.\footnote{The “dynastic” numbering follows that of the Royal Inscriptions of Mesopotamia, Babylonian Periods (RIMB) publications of the now-defunct Royal Inscriptions of Mesopotamia (RIM) Project directed by A. Kirk Grayson at the University of Toronto. “Babylon 1” = Kassite Period (1595–1155 BCE); “Babylon 2” = Second Dynasty of Isin (1157–1026 BCE); “Babylon 3” = Second Dynasty of the Sealand (1025–1005 BCE); “Babylon 4” = Bazi Dynasty (1004–985 BCE); “Babylon 5” = Elamite Dynasty (984–979 BCE); “Babylon 6” = Uncertain Dynasties (978–625 BCE); “Babylon 7” = Neo-Babylonian Dynasty (625–539 BCE); “Babylon 8” = Akkadian inscriptions of the Persian Period (538–330 BCE); “Babylon 9” = Macedonian rulers of Mesopotamia (currently no inscriptions known); and “Babylon 10” = Seleucid era (305–64 BCE).} However, all of these sub-corpora will be accessible from one interface.\footnote{[http://oracc.museum.upenn.edu/ribo/corpus].}

Phase 1 was first made public in early 2016 and was completed in early 2018. The content created includes:
- lemmatized editions of the inscriptions published in Frame 1995 and Da Riva 2013, as well as the famous “Cyrus Cylinder” and the “Antiochus (Borsippa) Cylinder”; and
- numerous informational portal pages on Babylonian rulers and their inscriptions, as well as on the various Babylonian King Lists.

During Phase 2, scheduled to run from 2018–2022, RIBo will produce fully lemmatized and searchable editions of the complete corpus of royal inscriptions of the six rulers of the Neo-Babylonian Empire (625–539 BCE): Nabopolassar, Nebuchadnezzar II, Amel-Marduk, Neriglissar, Labaši-Marduk, and Nabonidus. The transliterations, translations (English, as well as German), and glossaries (Akkadian, Sumerian, and proper names) will be fully searchable.

Eventually, RIBo will also include official inscriptions from the second millennium BCE, namely of the First Dynasty of Babylon, and the Kassite Period, but these corpora are being assembled/prepared by project partners in Munich and Philadelphia.\footnote{Frans van Koppen of the Institut für Assyriologie und Hethitologie at LMU Munich is currently overseeing the creation of editions of the inscriptions of the First Dynasty of Babylon, including the famous Law Code of Hammurabi stele. Grant Frame is working on the Kassite material.}
11.3 The Map Interface Ancient Records of Middle Eastern Polities

To make the content of MOCCI and its sub-projects more readily accessible to non-specialists that might shy away from the traditional corpus organization according to text IDs, a map-based interface that would have allowed access to geo-referenced text editions hosted on the Oracc platform has been created. ARMEP 1.0 was officially released in December 2017.

The interface’s core concept was inspired by the LMU-designed VerbaAlpina platform, a map-based interface devoted to exploring the diverse languages of the Alps and their interactions within this culturally rich linguistic region. Because the VerbaAlpina and Oracc data formats were not readily compatible, a more general map-based interface for accessing text-based data, that is also designed to be compatible on mobile devices, has been developed.

Figure 11.1: ARMEP base map showing the find-spots of ancient texts

ARMEP 1.0 presently displays roughly 6,700 ancient texts according to their find-spots (Figure 11.1) and allows users to filter displayed information by catalogue data (date,

---

23 David and Tobias Englmeier developed the interface with feedback from the present authors, as well as from Oracc’s creator Tinney. Riepl and Lücke oversaw software architecture and development, while the Englmeiers were the principal designers of ARMEP. The data included was curated by Novotny.

24 [https://www.verba-alpina.gwi.uni-muenchen.de/].
genre, language, material support, object type, period, provenience, ruler, and script) and content (translations, transliterations, lemma, and cuneiform signs). The following is an example of metadata filtering: when a user selects “Neo-Assyrian” for the period, “Royal Inscription” as the genre, “clay” as the material, and “prism” as the object, the map (using the current dataset) displays fifty-five (composite) texts originating from five different sites (Ashur, Babylon, Nimrud, Nineveh, and Sippar (Figure 11.2).

Alternatively, users can search the contents of the texts themselves. For example, when a user searches for the Akkadian lemma “anzû” (a mythical lion-headed eagle) and “lābu” (a word for lion), the map (using the current dataset) displays twenty-three (composite) texts found at seven sites (Ashur, Babil, Babylon, Nimrud, Nineveh, Uruk, and Zinçirli) (Figure 11.3).

25 The current dataset includes: RIAo (866 Assyrian inscriptions from the third millennium BCE to 745 BCE), RIBo (209 Babylonian inscriptions from 1157–64 BCE), RINAP (674 Assyrian inscriptions from 744–612 BCE), SAAo (4888 Neo-Assyrian archival texts published by the Helsinki-based Neo-Assyrian Text corpus project), and Suhu (33 inscriptions from the ninth century BCE).
Figure 11.3: Example of ARMEP filter by content results

The lemmatized texts on Oracc can be accessed from the “Item View” pop-up box, which is accessible through the “Cluster Overview” pop-up (Figure 11.4).

Figure 11.4: Sample “Cluster Overview” (left) and “Item View” (right)
ARMEP’s architectural structure was deliberately designed so that it can be easily cloned and adapted to meet the needs of other geo-referenced corpora. At LMU, its versatility has already been demonstrated by its adaptation for two other datasets.26

ARMEP does not intend to produce digital maps. This open-access, web-based tool is not envisioned as being just a map of Middle Eastern polities, but rather as an interface that contextualizes ancient sources geographically by their find-spots and that allows users direct access to lemmatized editions of the ancient texts found at those sites. This interface is specifically designed to break away from the boring, traditional “catalogue”-style of corpus organization, which is not particularly informative or interesting to non-specialists.

11.4 Methodological Problems and Technical Issues

For the most part, the present authors experienced relatively few methodological and/or technical issues with the development of OIMEA and ARMEP.

In terms of corpus building, most potential problems had been ironed out several years earlier as the software used for the Oracc platform has been in continual use since 2007, when it was developed on the basis of the electronic Pennsylvania Sumerian Dictionary.27

However, a few relatively minor problems still exist. For OIMEA and its sub-projects, these range from relatively straightforward technical issues to complex challenges:

1. printing easy-to-read editions from the Oracc web interface is not (yet) user-friendly nor are the results (yet) elegant;
2. multi-language support: the Oracc web interface does not properly handle multi-language translations of texts and does not (yet) support right-to-left scripts such as Aramaic, Arabic or Hebrew;
3. geo-referencing place name data in the glossaries is not yet possible;
4. disambiguation of namesakes: Oracc’s lemmatizer is currently not able to reliably disambiguate namesakes of people and places, even when additional information is provided for a name’s guide word (that is, a word’s primary meaning in the glossary);
5. extensive manual review of auto-lemmatized data is still required as the Oracc processor is unable to fully and accurately assess the citation form (the dictionary form of words), the sense (the meaning of the word in context), and the transcription (how the word was pronounced).

27 ePSD [http://psd.museum.upenn.edu/epsd1/index.html].
Most of these issues are technical rather than conceptual, especially the handling of print output (1). Tinney aims to resolve multi-language support (2) by late 2018. The ability to geo-reference place names (3) will be incorporated into the glossary generation as part of the development of ARMEP 2.0 in late 2018. The fix is to simply add a Pleiades ID field to entries in the glossary of names; the longitude and latitude coordinates will then be retrieved automatically from Oracc’s LMU Munich-based Geonames project. The disambiguation of namesakes (4) is needed if one wants to use Oracc-based data to create prosopographies. In order to achieve this, the system processor used for glossary validation needs to be fine-tuned so that the checker looks for 100% matches between lemmatized data in the source files and the corresponding glossary file. The current workaround is to add underscores to the citation form; for example, “Tukulti-apil-Ešarra[Tiglath-pileser III, king of Assyria]RN” becomes “Tukulti-apil-EšarraIII[Tiglath-pileser III, king of Assyria]RN”. The last issue, the need for the manual review of auto-lemmatized data (5), is a thorny problem that is likely to remain unresolved for some time as it requires refinement to the Oracc logic processor.

28 Oracc’s “Print text” function is now set up to handle projects with multi-language translations. On screen, the print option displays the editions in a readable format, but when the text is printed on paper (or to PDF) the results are less than desirable, as the multi-column format is not properly handled.

29 The only outstanding problem is that translation languages are displayed alphabetically by ISO 639-1 language code. Thus, English, which is the primary language of OIMEA projects, may not always be the default translation language in the Oracc pager when other translation languages are used.

30 Pleiades [http://pleiades.stoa.org/]. Various members of the OIMEA team have been volunteer content contributors of Pleiades since 2015 and they have added over 1,100 place resources for Assyrian and Babylonian cities, city gates, city walls, palaces, and temples. Over the course of 2018, we plan to add many more place resources so that ARMEP’s and Pleiades’ geo-referenced data are fully compatible.

32 Heather D. Baker (University of Toronto) is currently working on such a project, the Prosopography of the Neo-Assyrian Empire online (PNAo) [http://oracc.museum.upenn.edu/pnao/index.html].


34 The Oracc processor cannot properly analyze citation forms (CF) with marginally different guide words (GW). For example, Tukulti-apil-Ešarra[Tiglath-pileser I, king of Assyria]RN, Tukulti-apil-Ešarra[Tiglath-pileser II, king of Assyria]RN, and Tukulti-apil-Ešarra[Tiglath-pileser III, king of Assyria]RN are too similar to be disambiguated by the lemmatizer. The tolerance setting needs to be readjusted. For some information on Oracc CFs and GWs, see [http://oracc.museum.upenn.edu/doc/help/lemmatising/primer/index.html].

35 Oracc’s auto-lemmatizer function is not publically (or privately) documented. The criteria by which lemmatizer selects GWs (or senses) for CFs is not known to the authors. It is usually the GWs/sences, not the CFs (part of speech [POS] or normalization [NORM]), that require manual correction.
As for the ARMEP map interface, there were some issues at the beginning of development because the VerbaAlpina and Oracc data formats were not readily compatible. In addition, the catalogues, glossaries, transliterations, translations, and lists of signs could not be exported from Oracc. The solution[^36] was simply to make Oracc catalogue, glossary, transliteration, and translation data available in JavaScript Object Notation (JSON), under a CC0 or public domain license,[^37] a feature that Tinney implemented in early 2017. For ARMEP 1.0, no further compatibility issues were encountered.

### 11.5 Future Prospects

Over the next five years (2018–2022), OIMEA’s content will be expanded to incorporate inscriptions written in scripts other than cuneiform: Aramaic and Luwian are on top of the list. In addition, its lemmatized contents and glossaries will be improved, especially by standardizing the information in glossaries across its numerous sub-projects.

During 2018, ARMEP 2.0 is being developed and will be released at the end of the year (or in 2019). The new version of that open-access web interface will feature a gazetteer mode that will display places (including cities, villages, temples, mountains, and bodies of water) mentioned in ancient sources whose coordinates are known with a reasonable degree of certainty. This will substantially expand the information displayed in ARMEP 1.0, which shows texts according to their find-spots. This gazetteer function will display all geo-referenced places named in the defined text corpus (currently about 6,700 texts). For example, if a user clicks on the “View Places in Text” link of a 7th century BCE Assyrian inscription (e.g., the “Final Edition” of the Annals of Sennacherib), the map will show all of the cities that that king claims to have conquered and destroyed, as well as cities from whose rulers tribute was received. This innovative and dynamic geo-referenced rendering, which visualizes data in an easy-to-digest manner never before used in ancient Near Eastern studies, will further enhance the accessibility and usability of geographical information mentioned in cuneiform sources beyond specialist academics to casual or inexperienced users, including beginner students and members of the general public.

[^36]: The Englmeier brothers (with input from Lücke) and Tinney easily resolved the compatibility issue, as well as Oracc’s lack of exportability.

[^37]: For further information, see [http://oracc.museum.upenn.edu/doc/opendata/index.html].
Bibliography


