

An Open Access Index for the Geographical Distribution of the Cuneiform Corpus

Rune Rattenborg
University of Uppsala

Carolin Johansson
University of Uppsala

Seraina Nett
University of Uppsala

Gustav Ryberg Smidt
University of Uppsala

Jakob Andersson
University of Uppsala

§1. Introduction¹

We present here the first comprehensive version of the Cuneiform Inscriptions Geographical Site Index, an open access digital index which includes standardised locational and attribute information for more than 500 archaeological locales from across the Eastern Mediterranean and the Middle East where texts written in different varieties of the cuneiform script have been found.² This resource is derived from the research programmes of the *Memories For Life* and the *Geomapping Landscapes of Writing*³ projects of the Uppsala University Department of Linguistics and Philology, financed respectively by the Swedish Research Council (grant no. 2016-02028) and the Riksbankens Jubileumsfond (grant no. MXM19-1160:1). Through the application of a corpus-wide perspective on the cuneiform script, these projects are advancing the application of digital humanities research design to the study of writing in the ancient world, set at the interface of text, materiality, and open access data repositories. The rapid growth of digital repositories and applications in cuneiform studies of recent decades

has brought with it a more pressing need for global, standardised indices for accessing, querying, and analysing the immense amount of digital data now available. By fielding a comprehensive spatial index of finds of cuneiform inscriptions, we hope to motivate usage of a standard index for naming and referring to archaeological proveniences, further integrated approaches to the study of this corpus in its entirety, and to make its contents more immediately available for researchers from other fields. In the following sections, we briefly sketch the relevance and potential application of this resource, the structure and contents of the assembled dataset, directions for its integration with existing catalogues and its potential application in future research. Regular updates of the dataset are released in .csv, .geojson, and .kml formats under a Creative Commons Attribution 4.0 International License⁴ from Uppsala University Department of Linguistics and Philology and hosted by Lecturer in Assyriology Jakob Andersson of the same institution.⁵ A stable version of all releases is deposited with the OpenAIRE repository Zenodo (<http://doi.org/10.5281/zenodo.4960710>).

¹ Funding: Swedish Research Council grant no. 2016-02028 and Riksbankens Jubileumsfond grant no. MXM19-1160:1 (see <https://www.vr.se/english/swecris> for details)

² We would like to express our sincere gratitude to the editors of the CDLJ and the two anonymous reviewers for their comments and suggestions to this paper. We should furthermore acknowledge the expert insight and advice of several colleagues, namely Émilie Pagé-Perron, Nathan Morello, Jamie Novotny, and Olof Pedersén. Any remaining errors or inaccuracies are our own.

³ <https://www.lingfil.uu.se/research/assyriology/glow/> (accessed 17 June 2021)

⁴ <https://creativecommons.org/licenses/by/4.0/> (accessed 17 June 2021)

⁵ <https://www.lingfil.uu.se/research/assyriology/earth/#CIGS> (accessed 17 June 2021)

§2. Research Background

§2.1 The immediate relevance of a dedicated spatial index for a catalogue of written sources counting at least 500,000 individual records (Streck, 2010) should not be difficult to appreciate. A textual corpus extending across all of the Middle East and adjoining regions and encompassing more than 3,000 years of recorded human history holds immense potential for the understanding of long-term, large-scale developments in the use and application of writing in early human history. Whether formally defined or anecdotally hinted at, spatial distribution forms an integral part of such perspectives, and even more decisively so in the digital age. Where extensive, corpus-wide analyses of cuneiform sources were previously well beyond the humanly possible of any one scholar, the array of computing applications now available offers rare and extraordinary opportunities for overseeing, understanding, and communicating the richness and diversity of this immense body of historical information at a global level (Bigot Juloux et al., 2018; articles in Rossi and De Santis, 2019). Such are also certain to benefit from formal reconsideration of the spatial dimension of this body of textual sources, as has already been amply demonstrated by the promising application of localised spatial datasets to various transects of the corpus, e.g. the *Hittite Epigraphic Findings in the Ancient Near East*⁶ and the *Ancient Records of Middle Eastern Politics* (ARMEP) (Novotny and Radner, 2019).⁷ Through the use of spatial data and web mapping applications, such initiatives allow the user to easily query immense numbers of texts according to spatial distribution. Further beyond, the spatial dimension of such corpora will also allow for the analytical integration with a range of entirely different types of data, e.g. material culture, by using spatial location as the ordering index.

§2.2 Digital humanities ecosystems relating to a considerable variety of textual corpora from the pre-modern world have for some time now incorporated comprehensive spatial indices as part of their basic data pool. These include, to name but a few, the spatial indices that accompany

the more than 8,000 inscriptions of the *Digital Archive for the Study of Pre-Islamic Arabian Inscriptions* (Avanzini et al., 2019),⁸ the aggregate 10,000 Runic inscriptions from Scandinavia and Northern Europe found in e.g. the *Scandinavian Runic-Text Database* (Peterson, 1994),⁹ or the more than 80,000 Latin inscriptions of the *Epigraphic Database Heidelberg*.¹⁰ Emphasising the very tangible materiality characteristic of many early writing systems, such indices serve a central role firstly in documenting the origin and provenience of material culture, secondly as a basic reference for indexing and ordering historical documentation, and third, and last, but certainly not least, as a critical means to survey and safeguard material cultural heritage. In the digital age, the ability to query, map, and inspect textual records on the basis of their spatial location goes well beyond a mere obligatory recording of archaeological context. Spatial location is a powerful variable in the analysis of artefact use patterns in archaeology (Hodder and Orton, 1976; Conolly and Lake, 2006), and the use of spatial data applications in historical research focusing on a much more recent past has seen significant gains in latter years (Gregory and Geddes, 2014). Such perspectives have only rarely been applied in analytical approaches to early writing, however (for a good example of how the spatial distribution of writing may be used to address historical questions from the ancient world, see e.g. Ebert et al. 2012). The technological means now available to bring such variables to bear on the formulation and execution of broader research questions are bound to permanently change many practical constraints that have hitherto remained unchallenged.

§2.3 Another related argument for developing a formal spatial dimension to existing digital text catalogues is the changing nature of research infrastructures and resources of the 21st century CE more generally. Growing digitisation and the increasing automation and linking of formerly isolated databases and the wider operationalisation of data repositories will, in a not too distant future, prompt a rethinking of how we integrate textual sources within broader and more regionally or super-regionally focused archaeological

⁶ <http://www.hittiteepigraphs.com> (accessed 17 June 2021)

⁷ <https://www.armep.gwi.uni-muenchen.de> (accessed 17 June 2021)

⁸ <http://dasi.cnr.it> (accessed 17 June 2021)

⁹ <http://www.runforum.nordiska.uu.se/en/srd/> (accessed 17 June 2021)

¹⁰ <https://edh-www.adw.uni-heidelberg.de> (accessed 17 June 2021)

and historical perspectives (consider for example Kintigh et al., 2014; Smith et al., 2012; Wright and Richards, 2018). Research agendas are bound to become increasingly embedded in and reliant on fully digital research ecosystems in the future, ecosystems where spatial data plays an important, structuring role. Archaeology, to mention a discipline quite proximal to cuneiform studies, is currently seeing a rapid growth in the scale and quality of spatial data repositories also covering much of the Middle East, with records numbering in the many hundreds of thousands (see for example Harrison, 2018; Zerbini, 2018). Education, dissemination, public information, and management of archaeological and historical heritage will only become more reliant on the lasting integration of digital repositories and data collections in years ahead. Such repositories include locational data, which holds tremendous potential as a means of easily overseeing and conveying otherwise huge bodies of standardised information.

§3. Method

§3.1 The dataset introduced here represents the first comprehensive, digital index of findspots for inscriptions in cuneiform and derived scripts. The index currently (version 1.x) numbers more than 500 records at site level. A draft version of the index (0.x) was developed in 2018-2019 as part of the catalogue of the *Memories For Life* project of Uppsala University and the University of Cambridge, headed by Jakob Andersson and Christina Tsouparopoulou. A preliminary version of this index has been used with the catalogue of the *Cuneiform Digital Library Initiative*¹¹ (CDLI) for development purposes over the summer of 2019 and forms the basis for spatial indices implemented with the new, future platform of the CDLI. Over the coming years, we will continue to develop and expand upon the index as part of the *Geomapping Landscapes of Writing* project of Uppsala University, so as to stimulate broader standardisation and integration of spatial data in the field of cuneiform studies.

§3.2 The index is developed by combining archival research with the inspection of a broad range of cartographic, imagery, and web mapping resources using GIS software. Information on sites with known finds of cuneiform sources are acquired from archaeological and philological publications or online text catalogues. The archaeological site is then visually located on high-resolution satellite imagery and checked against web mapping resources and older printed maps to verify identification and to add identifier concordance and additional toponymic data. Imagery resources used are freely accessible, high-resolution satellite image repositories such as Google Earth or Bing Maps, complemented by web mapping gazetteers such as OpenStreetMap, Geonames, and Google Maps, and specialist indices of archaeological sites, such as e.g. the *ANE.kmz*¹² (Pedersén, 2012), *Pleiades*,¹³ and *Ancient Locations*.¹⁴ Over time, we aim to expand the index to include also bibliographical references that will facilitate independent verification of individual record locations. The canonical index as well as associated datasets are stored with a dedicated project server maintained by the Uppsala University IT Services Division.

§4. Dataset

§4.1 The following sections lay out some general definitions of the principal concepts guiding this index, namely the type of written source material to which it relates and our notion of an archaeological site. Secondly, we provide a brief summary of individual variables and data fields contained within the current version (1.x) of the index. Regular updates of the dataset are released in .csv, .geojson, and .kml formats, along with a complete version history, under a Creative Commons Attribution 4.0 International License¹⁵ from Uppsala University Department of Linguistics and Philology.¹⁶

§4.2 Cuneiform

§4.2.1 Discrete provenience location records included in this index are contingent on the discovery, open or clandestine, of cuneiform in-

¹¹ <http://cdli.ucla.edu> (accessed 17 June 2021)

¹² <https://www.lingfil.uu.se/research/assyriology/earth/> (accessed 17 June 2021)

¹³ <https://pleiades.stoa.org> (accessed 17 June 2021)

¹⁴ <http://www.ancientlocations.net> (accessed 17 June 2021)

¹⁵ <https://creativecommons.org/licenses/by/4.0/> (accessed 17 June 2021)

¹⁶ <https://www.lingfil.uu.se/research/assyriology/earth/#CIGS> (accessed 17 June 2021)

scriptions at that location. Our definition of ‘cuneiform’ is inclusive in its outlook and reflects the current state of digital catalogues in the field, most notably those of the CDLI and the *Open Richly Annotated Cuneiform Corpus*¹⁷ (ORACC). We have accepted as valid records to include in this index all known archaeological localities with finds of inscriptions utilising the cuneiform script or derived writing systems (Finkel and Taylor, 2015; Walker, 1990). The current index then includes proveniences for inscriptions utilising the generally known cuneiform script used to render the language isolate Sumerian (Michalowski, 2008), the various dialects of the Akkadian language branch, including Assyrian, Babylonian, and Eblaite (Huehnergard and Woods, 2008) and their western relative, Ugaritic (Pardee, 2008), early Indo-European tongues such as Hittite (Watkins, 2008), Luvian (Melchert, 2008a), and Palaic (Melchert, 2008b) and isolate languages of the Taurus and Zagros foothills and the Eastern Anatolian and Armenian highlands, namely Hurrian (Wilhelm, 2008a) and Urartian (Wilhelm, 2008b). Next to these are scripts associated with the Elamite of southwestern Iran, including Proto-Elamite, Linear Elamite, and Elamite cuneiform (Dahl, 2018; Stolper, 2008), and of course Old Persian cuneiform containing the earliest renditions of Indo-Iranian languages (Schmitt, 2008). While no formal, easily definable, and broadly accepted scholarly definition of cuneiform as a discrete writing system exists, we feel that the extent and duration sketched above reflect both specialist and general convention as well as the principal material and technological characteristics observable in cuneiform writing.

§4.3 Sites

§4.3.1 A provenience record refers to an archaeological site. We have maintained a notion of an archaeological ‘site’ as a geographically delineable and largely contiguous area of archaeological remains, even if such a definition is not without problems from a formal analytical perspective. Perspectives emphasising artefact distribution as the primary object of archaeological inquiry have, for example, criticised the notion of a ‘site’ as conceptually flawed and largely artificial (Dunnell and Dancey 1983). But even if the notion of an archaeological site is almost universally accepted, individual provenience records contained

within the present index will also, to a certain extent, reflect scholarly tradition in the field. In our efforts to balance conceptual rigour and scholarly convention, some cases will inevitably stand out. For example, we consider the complex of large mounds that make up ancient Kiš (KSH) south of modern Baghdād a single record, even if some of the site’s component parts are occasionally considered discrete properties by others. At the other end of this spectrum, Dīqdīqqah (DQD), two kilometres northeast of Ur, is considered a discrete record on account of the relatively speaking substantial distance from its much larger neighbour, even if material remains of the former are virtually always considered part of the assemblage deriving from the latter.

§4.3.2 The index also introduces a simple locational accuracy estimate reflecting the certainty with which locational data provided for individual entries can be said to reflect an accurate geographical location of a site. This measure seeks to qualify the otherwise undifferentiated authority assigned to records that may, on closer inspection, illustrate very different levels of certainty as far as our knowledge about where a text comes from is concerned. Accuracy estimates employ a formal, four-tier scale explained in more detail below. It is important to stipulate that our definition of a site entity refers strictly to an archaeological, not a historical, locale, and that this is a locale of archaeological provenience, not origin of production or place of use. No attempt has been made to include approximate locations of historical toponyms where these cannot be associated with a known or reasonably firmly associated geographical feature. Such a functionality abides by a very different and more complex historical geographical ontology already well integrated in other applications, e.g. the digital gazetteer of *Pleiades*.¹⁸

§5. Data Fields

§5.1 The following sections describe individual data fields contained in the index and the procedures guiding their population. The current version 1.x index contains a total nineteen fields, namely one primary ID, one spatial accuracy field, six integer and string fields for external data links, nine string fields with toponyms, and two integer fields making up the point coordinate of the

¹⁷ <http://oracc.museum.upenn.edu> (accessed 17 June 2021)

¹⁸ <https://pleiades.stoa.org> (accessed 17 June 2021)

record. These are described in turn below and in the associated table.

§5.2 Record identifiers

§5.2.1 The table supplies a range of primary ID keys suitable for linking different datasets, which will enable users to link and match data from different repositories. Next to our native record key (`site_id`), the index also contains future (`cdli_id`) and legacy (`cdli_legacy`) CDLI provenience record identifiers, along with identifiers of concordant records in *Pleiades: A Gazetteer of Ancient Places* (`pleiades_id`),¹⁹ OpenStreetMap (`osm_id` and `osm_type`),²⁰ and Geonames (`geonames_id`).²¹ Future versions of the index will also seek to include links to corresponding records at Wikipedia and, by extension, the considerable wealth of spatial data links stored with Wikidata. Geographical locations given in the current index and in the provenience index of the CDLI should be identical in all respects. For the latter three sets of concordances, the record entity is identical, whereas the geographical location may differ to varying degrees, depending on the geolocation procedures utilised by the respective data consortiums. While widely used, crowdsourced geographical data repositories such as Geonames and OpenStreetMap offer no transparent information on how individual data points within their databases were originally identified or captured (Goodchild, 2007).

§5.3 Record geodata and accuracy

§5.3.1 The index provides a coordinate location and a locational accuracy assessment of each provenience record (Fig. 2). Record geographical location is provided as a single point location, consisting of two fields giving the *x* and *y* value, namely longitude (`lon_wgs1984`) and latitude (`lat_wgs1984`). Both are given in decimal degrees, using the extremely common and universally used coordinate reference system of the World Geodetic System (WGS) 1984 datum (EPSG 4326) maintained by the National Geospatial-Intelligence Agency of the United States Department of Defense.

§5.3.2 The locational accuracy of the point location (accuracy) is given according to a formalised, four-tier scale, 3 being certain, 2 being represen-

tative, 1 being tentative, and 0 being unknown. Accuracy levels reflect site visibility and ease of delineation, but some measure of authority as to whether cuneiform inscriptions have actually been found at the site in question will also be reflected from the locational accuracy value. Where a discrete site outline can be identified and traced, the site has been drawn as a polygon and the location derived from the resulting centroid, giving a value of 3. Where the site can be positively located, but not drawn, the value is given as 2, e.g. the mound of Ḥabūbah (HBK) currently below the surface of Lake Assad. Where a site location can be placed with reasonable certainty, but not positively located or delineated, the value is given as 1, e.g. the modern city of Kirmānšāh (KRM), from which a number of texts that might equally well have been illicitly excavated elsewhere are said to derive. Where the location cannot be defined with any reasonable degree of certainty, the value is 0.

§5.3.3 Regardless of the level of relative accuracy obtainable from imagery, the geographical accuracy of the locations that the index contains will only be as accurate as the underlying imagery source on which drawn polygon and point geometry is based. As we rely on publicly available satellite imagery resources, generally recognised horizontal deviation in the range of up to 50 metres and more in developing countries should be kept in mind (Mohammed et al., 2013; Pedersén, 2012; Potere, 2008).

§5.4 Record place-names

§5.4.1 In order to allow for cross-referencing between different languages and scripts, and to introduce some measure of standardisation as far as nomenclature is concerned, the index also provides a range of names for each record. The primary historical name of the location (`anc_name`) as it relates to cuneiform culture is given if known with a reasonable degree of certainty. Places can, of course, have many names, and the current index is not intended to provide an exhaustive collection of all variant ancient writings or toponyms attested for individual records. The table also offers a suite of fields for the rendition of modern toponyms in pertinent languages that do not make use of the Roman script, e.g. Arabic (`ara_name`),

¹⁹ <https://pleiades.stoa.org/> (accessed 17 June 2021)

²⁰ <https://www.openstreetmap.org/> (accessed 17 June 2021)

²¹ <http://www.geonames.org> (accessed 17 June 2021)

Armenian (arm_name), Farsi (fas_name), Georgian (geo_name), Greek (gre_name), Hebrew (heb_name), and Russian (rus_name). A Romanised version of the modern place-name (transc_name) is maintained from the original when dealing with Azerbaijani, Maltese, Romanian, or Turkish toponyms, or transcribed according to the guidelines of the American Library Association – Library of Congress (ALA-LC) (see <https://www.loc.gov/catdir/cpsd/roman.html>) when dealing with Arabic, Armenian, Azerbaijani, Farsi, Georgian, Greek, Hebrew, or Russian

toponyms. Where names in multiple languages are available, the transcribed name is drawn from the official language of the national entity currently associated with the record in question. One should note that spelling of toponyms may vary, so discrepancies between the values given in this index and those of other repositories will occasionally occur. Values are derived from archaeological reports or site gazetteers or, alternatively from online resources, e.g. Wikipedia, OpenStreetMap, and Geonames.

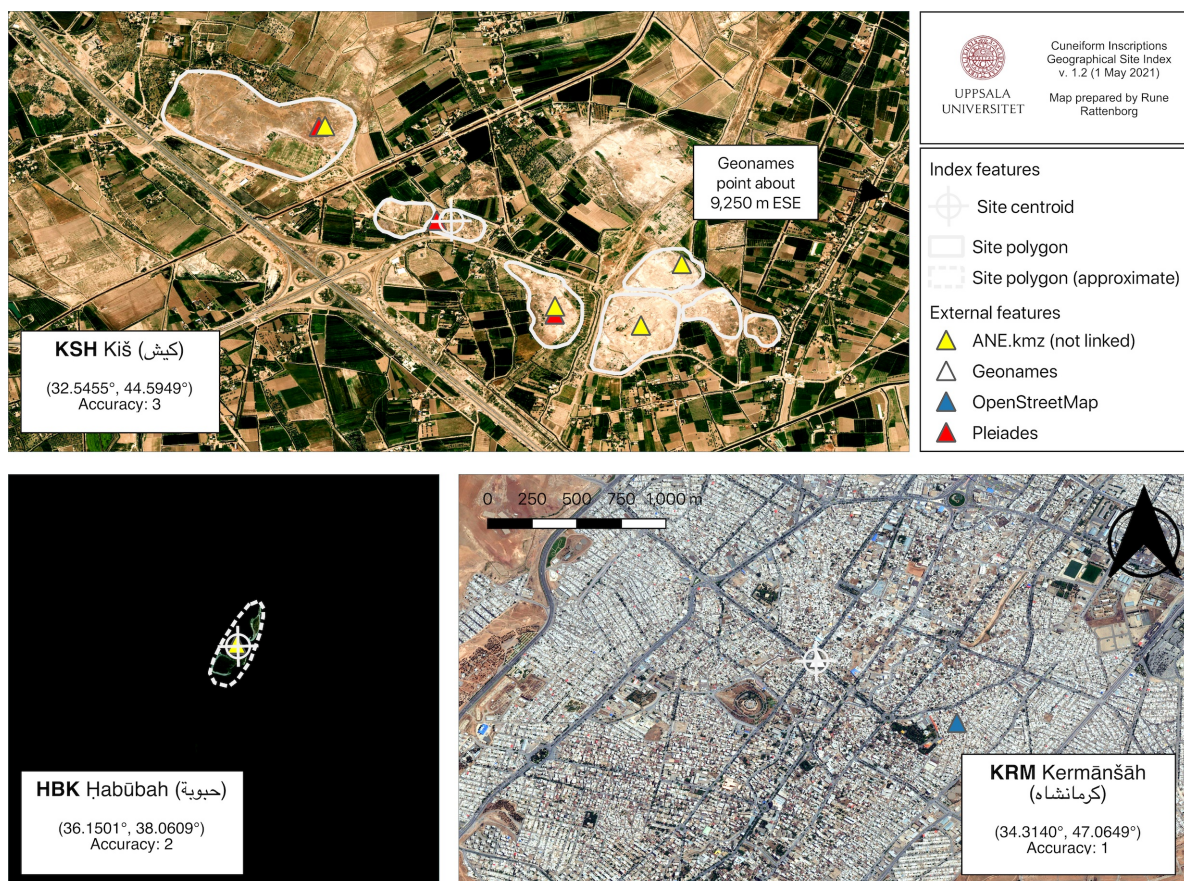


Figure 1: A visualisation of spatial vector data associated with individual records in CIGS and the different levels of spatial accuracy for each record. Kiš (KSH, top left) is a typical mounded site with a well-defined topographical outline. The location and general outline of Ḥabūbah (HBK, bottom left), a mound now inundated by an artificial lake, is well known, but not accurately traceable on modern imagery. The modern city of Kirmānshāh (KRM, bottom right), while given as the presumed provenience of multiple inscriptions, cannot be meaningfully represented as an archaeological locale and so is given a representative point location only. All maps to scale. Background imagery courtesy of Bing Maps (www.bing.com/maps).

§6. Usage

§6.1 A provisional visualisation of the current version of this index (Fig. 1) neatly illustrates the immense geographical spread of the cuneiform script, from Rome to Kabul and from central Russia to southern Egypt and the Strait of Hormuz. Considering the comparatively high level of spatial and archaeological resolution, as well as the sheer scale, of this corpus, cuneiform holds significant potential to decisively inform data-driven comparative studies on the emergence, spread, and use of writing in early human history (see for quantitative studies on long-term trends in the production and consumption of writing, albeit from much later periods, e.g. Buringh and van Zanden, 2009; Xu, 2013). Linking up existing digital text catalogues, such as those of the CDLI, ORACC, and others, with versatile meta-

data frameworks for querying associated temporal and spatial information will allow for extended, formal analyses of a vastly increased number of variables than previously seen.

§6.2 Our initial aim with the index introduced here is, however, more prosaic. First and foremost, we wish to introduce a measure of standardisation to text provenience information, thereby allowing for the lasting and dynamic integration of spatial information with digital research tools and existing online text catalogues. The index combines a formal set of record identifiers and basic attribute data with accurate locational information generated according to a uniform and easily reproducible set of procedures and allows for the linking of these entities with corresponding records in leading open access geodata repositories.

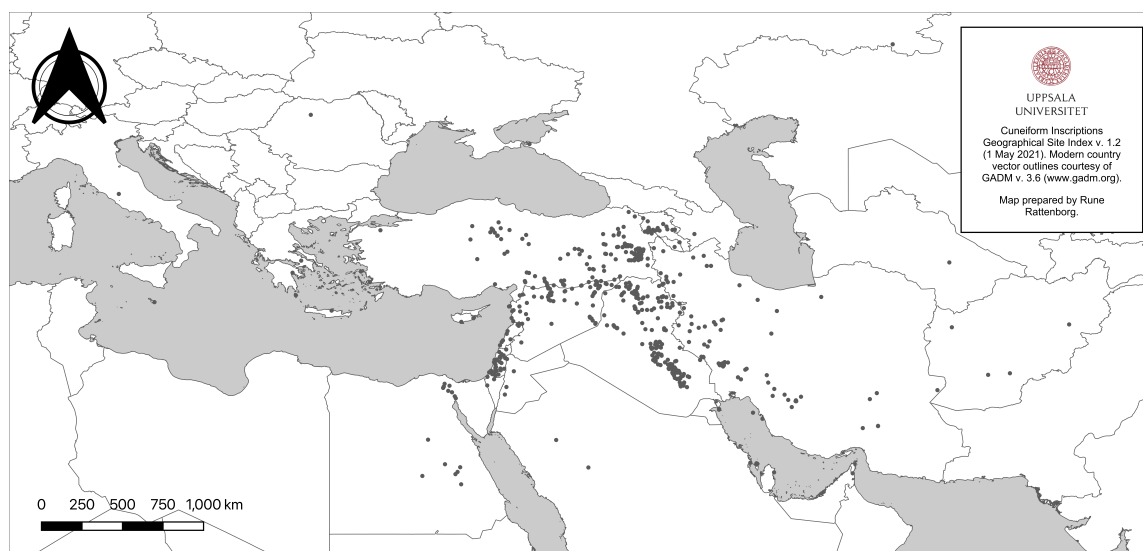


Figure 2: Point distribution of the 553 location records contained in CIGS v. 1.2. Modern country vector outlines courtesy of GADM v. 3.6 (www.gadm.org).

§6.3 Most pressing here is, of course, the need to adopt a formally sound and generally recognised means of referencing locational data in association with the study and safeguarding of cuneiform culture, a requirement that this index seeks to address by introducing a basic and universally applicable reference resource. As an example, comparable geographical indices have been in widespread use in Mayan studies since the 1980s (Fash, 2016; Prager et al., 2014), offering an easy and transparent way to identifying and referencing provenience locations across dif-

ferent research projects, initiatives, and publications. The rise of crowdsourced or volunteer geographic information platforms over the last two decades has brought with it an unrivalled wealth of locational information accessible to professionals and laypersons alike. These resources hold immense potential for the future integration and free dissemination of spatial data by researchers, if recognising and helping to amend the emergent need to generate and curate geographical information with the same rigour as other types of metadata (Goodchild 2008).

§6.4 Similar sentiments underlie our inclusion of toponymical information in a wide variety of languages and scripts, and the adoption of a formal means of transliteration into Roman letters. The exact rendering of many toponyms in cuneiform studies are a product firstly of the rich linguistic landscape of the Middle East and adjoining regions, secondly the myriad ways of naming, spelling, and rendering place names in native languages and scripts, and thirdly of the different orthographical conventions of various European languages in which these toponyms have often ultimately settled in the literature. By providing a formal and digitally durable set of names for text proveniences, we hope that this index will bring some degree of order to the very confusing toponymical landscape that we ourselves have encountered.

§6.5 Finally, and as noted in the introduction,

building infrastructures for the digital rendering of provenience metadata is a critical prerequisite for the future integration of research databases in cuneiform studies and archaeology. While the digital divide between these fields is understandable from an epistemological, and especially disciplinary, point of view, it is lamentable nevertheless, even more so when considering the urgent need for comprehensive digital infrastructures able to aid the safeguarding of cultural heritage in the Middle East and beyond. The increased technological and computer-driven abilities of archaeological research in the Middle East, and here especially the emergence of extensive, spatial databases, certainly represents one of the most obvious interfaces between textual and material research going forward (Gates, 2020). We hope that researchers and cultural heritage professionals alike will be able to make use of the resource presented here in comparable ways.

Value	Field	Type	Description
Site ID key	site_id	str	Primary ID. Each record is identified with a unique three-letter code.
Locational accuracy	accuracy	int	A formal assessment of the level of geographical accuracy with which position given can be said to relate to the historical location in question on a four-tier scale, 3 being certain, 2 being representative, 1 being tentative, and 0 being unknown.
CDLI ID key	cdli_provenience_id	int	The numerical provenience ID for the corresponding site in the Cuneiform Digital Library Initiative (www.cdli.ucla.edu) catalogue, if available.
Ancient name	anc_name	str	Common rendering of the ancient name of the site in question, if known, based on readings from cuneiform texts.
Transcribed name	transc_name	str	The transcribed name of the site as drawn from the principal language of the national entity currently associated with the record in question.
Arabic name	ara_name	str	Arabic name of the site, if applicable and available.
Armenian name	arm_name	str	Armenian name of the site, if applicable and available.
Farsi name	fas_name	str	Farsi name of the site, if applicable and available.
Georgian name	geo_name	str	Georgian name of the site, if applicable and available.
Greek name	gre_name	str	Greek name of the site, if applicable and available.
Hebrew name	heb_name	str	Hebrew name of the site, if applicable and available.
Russian name	rus_name	str	Russian name of the site, if applicable and available.
Pleiades ID	pleiades_id	int	The primary ID of the corresponding place record in Pleiades: A Gazetteer of Ancient Places, if available. The stable link will be https://pleiades.stoa.org/places/[pleiades_id] .
OpenStreetMap ID	osm_id	int	The primary ID of the corresponding place record in OpenStreetMap, if available. The stable link will be https://www.openstreetmap.org/[osm_type]/[osm_id] .
OpenStreet Map geometry type	osm_type	str	The geometry type of the corresponding place record in OpenStreetMap, if available.

Geonames ID	geonames_id	int	The primary ID of the corresponding place record in Geonames, if available. The stable link will be https://www.geonames.org/[geonames_id] .
CDLI legacy value	cdli_legacy	str	All associated legacy provenience values found in the current catalogue of the Cuneiform Digital Library Initiative (www.cdli.ucla.edu).
Longitude (x)	lon_wgs1984	int	Longitude of the record location in decimal degrees in the WGS 1984 geographic coordinate reference system (EPSG 4326).
Latitude (y)	lat_wgs1984	int	Latitude of the record location in decimal degrees in the WGS 1984 geographic coordinate reference system (EPSG 4326).

Table 1: Overview and explanation of data fields (CIGS v. 1.2).

BIBLIOGRAPHY

- Avanzini, A., A. De Santis, and I. Rossi (2019). "Encoding, Interoperability, Lexicography: Digital Epigraphy Through the Lens of DASI Experience". In: *Crossing Experiences in Digital Epigraphy, From Practice to Discipline*. Ed. by I. Rossi and A. De Santis. Berlin & Boston: De Gruyter, 1–18.
- Bigot Juloux, V., A.R. Gansell, and A. Di Ludovico (2018). *CyberResearch on the Ancient Near East and Neighboring Regions*. Leiden & Boston: Brill.
- Buringh, E. and J.L. van Zanden (2009). "Charting the 'Rise of the West': Manuscripts and Printed Books in Europe, A Long-Term Perspective from the Sixth through Eighteenth Centuries." *The Journal of Economic History* 69, 409–445.
- Connolly, J. and M. Lake (2006). *Geographical Information Systems in Archaeology*. Cambridge: Cambridge University Press.
- Dahl, J. L. (2008). "The Proto-Elamite Writing System". In: *The Elamite World*. Ed. by J. Álvarez-Mon, G.P. Basello, and Y. Wicks. London and New York: Routledge, 383–396.
- Dunnell, Robert C. and William S. Dancey (1983). "The Siteless Survey: A Regional Scale Data Collection Strategy". In: *Advances in Archaeological Method and Theory* 6. Ed. by Michael Schiffer. New York: Academic Press, 267–287.
- Ebert, C.E., K.M. Prufer, and D.J. Kennett (1999-2000). "Maya Monuments and Spatial Statistics: A GIS-Based Examination of the Terminal Classic Period Maya Collapse". *Research Reports in Belizean Archaeology* 9, 91–105.
- Fash, B.W. (2016). *Site Codes. Corpus of Maya Hieroglyphic Inscriptions*. URL: <https://www.peabody.harvard.edu/cmhi/writing.php>.
- Finkel, I. and J. Taylor (2015). *Cuneiform*. London: The British Museum.
- Gates, M.-H. (2020). "Archaeology and the Ancient Near East: Renewing a Culture-Historical Partnership". In: *A Companion to the Ancient Near East, Blackwell Companions to the Ancient World*. Ed. by D.C. Snell. ABAW NF 79. John Wiley & Sons, 74–93.
- Goodchild, M.F. (2007). "Citizens as sensors: the world of volunteered geography". *GeoJournal* 69, 211–221. URL: <https://doi.org/10.1007/s10708-007-9111-y>.
- (2008). "Spatial Accuracy 2.0". In: *Proceedings of the 8th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*.
- Gregory, I.N. and A. Geddes (2014). *Toward Spatial Humanities: Historical GIS and Spatial History*. Bloomington, IN: Indiana University Press.
- Harrison, T.P. (2018). "Computational research on the Ancient Near East (CRANE): large-scale data integration and analysis in Near Eastern archaeology". *Levant* 52:1-2, 1–4. URL: <https://doi.org/10.1080/00758914.2018.1492784>.
- Huehnergard, J. and C. Woods (2008). "Akkadian and Eblaite". In: *The Ancient Languages of Mesopotamia, Egypt, and Aksum*. Ed. by R.D. Woodard. Cambridge: Cambridge University Press, 83–152.
- Kintigh, K. et al. (2014). "Grand Challenges for Archaeology". *American Antiquity* 79, 5–24.
- Melchert, H.C. (2008a). "Luvian". In: *The Ancient Languages of Asia Minor*. Ed. by R.D. Woodard. Cambridge: Cambridge University Press, 31–39.
- (2008b). "Palaic". In: *The Ancient Languages of Asia Minor*. Ed. by R.D. Woodard. Cambridge: Cambridge University Press, 40–45.
- Michalowski, P. (2008). "Sumerian". In: *The Ancient Languages of Mesopotamia, Egypt, and Aksum*. Ed. by R.D. Woodard. Cambridge: Cambridge University Press, 6–46.

- Mohammed, N.Z., A. Ghazi, and H.E. Mustafa (2008). "Horizontal Positional Accuracy of Google Earth's High-Resolution Imagery Archive". *Sensors* 8, 7973–7981. URL: <https://doi.org/10.3390/s8127973>.
- (2013). "Positional Accuracy Testing of Google Earth". *International Journal of Multidisciplinary Sciences and Engineering* 4, 6–9.
- Novotny, J. and K. Radner (2019). "Official Inscriptions of the Middle East in Antiquity: Online Text Corpora and Map Interface". In: *Crossing Experiences in Digital Epigraphy*. Ed. by I. Rossi and A. De Santis. Warsaw: De Gruyter, 141–154.
- Pardee, D. (2008). "Ugaritic". In: *The Ancient Languages of Syria-Palestine and Arabia*. Ed. by R.D. Woodard. Cambridge: Cambridge University Press, 5–35.
- Pedersén, O. (2012). "Ancient Near East on Google Earth: Problems, Preliminary Results, and Prospects". In: *Proceedings of the 7th International Congress on the Archaeology of the Ancient Near East 12 April - 16 April 2010, the British Museum and UCL, London. Vol. 3: Fieldwork & Recent Research*. Ed. by R. Matthews and J. Curtis. Wiesbaden: Harrassowitz, 385–393.
- Peterson, L. (1994). "Scandinavian Runic-text Database: a Presentation". In: *The Twelfth Viking Congress: Developments Around the Baltic and North Sea in the Viking Age, Birka Studies*. Ed. by B. Ambrosiani and H. Clarke. Stockholm, 305–309.
- Prager, C. et al. (2014). *A Checklist of Archaeological Sites with Hieroglyphic Inscriptions*. URL: <https://mayawoerterbuch.de/archaeologicalsites/>.
- Rossi, I. and A. De Santis (2019). *Crossing Experiences in Digital Epigraphy, From Practice to Discipline*. Berlin & Boston: De Gruyter. URL: <https://doi.org/10.1515/9783110607208>.
- Schmitt, R. (2008). "Old Persian". In: *The Ancient Languages of Asia and the Americas*. Ed. by R.D. Woodard. Cambridge: Cambridge University Press, 76–100.
- Smith, M.E. et al. (2012). "Archaeology as a Social Science". *Proceedings of the National Academy of Sciences* 109, 7617–7621.
- Stolper, M.W. (2008). "Elamite". In: *The Ancient Languages of Mesopotamia, Egypt, and Aksum*. Ed. by R.D. Woodard. Cambridge: Cambridge University Press, 47–82.
- Streck, M.P. (2010). "Großes Fach Altorientalistik: Der Umfang des keilschriftlichen Textkorpus". *Mitteilungen der Deutschen Orient-Gesellschaft* 142, 35–58.
- Walker, C.B.F. (1990). "Cuneiform". In: *Reading the Past: Ancient Writing From Cuneiform to the Alphabet*. Ed. by J.T. Hooker. University of California Press and British Museum, 15–74.
- Watkins, C. (2008). "Hittite". In: *The Ancient Languages of Asia Minor*. Ed. by R.D. Woodard. Cambridge: Cambridge University Press, 6–30.
- Wilhelm, G. (2008a). "Hurrian". In: *The Ancient Languages of Asia Minor*. Ed. by R.D. Woodard. Cambridge: Cambridge University Press, 81–104.
- (2008b). "Urartean". In: *The Ancient Languages of Asia Minor*. Ed. by R.D. Woodard. Cambridge: Cambridge University Press, 105–123.
- Wright, H. and J.D. Richards (2018). "Reflections on Collaborative Archaeology and Large-Scale Online Research Infrastructures". *Journal of Field Archaeology* 43, 60–67. URL: <https://doi.org/10.1080/00934690.2018.1511960>.
- Xu, T. (2013). "Knowledge Formation and the Great Divergence between China and Europe: Manuscripts and Printed Books, ca. 581–1840". *Journal of Comparative Asian Development* 12, 245–284. URL: <https://doi.org/10.1080/15339114.2013.792455>.
- Zerbini, A. (2018). "Developing a Heritage Database for the Middle East and North Africa". *Journal of Field Archaeology* 43, 9–18. URL: <https://doi.org/10.1080/00934690.2018.1514722>.