

---

## Kima: a text-based historical gazetteer in Hebrew Script

**Sinai Rusinek**

sinai.rusinek@mail.huji.ac.il

Tel Aviv University, Israel

**Glauco Mantegari**

mantegla@gmail.com

University of Salzburg, Germany

**Dimid Duchovny**

dimidd@gmail.com

The Hebrew University of Jerusalem, Israel

---

KIMA - Towards an Open Hebrew Gazetteer is a project funded by the Pelagios Commons Resource Grants, which sets the grounds for a comprehensive, dynamic and interoperable database of historical place names in languages written in the Hebrew Script. The gazetteer provides a stable and shared digital reference for toponyms and will be integrated with the Pelagios Commons annotation and mapping environment, [Recogito](#).

Each entry in this database consists of Preferred forms of a toponym (both in Hebrew script and in its English normalized form), Alternate Hebrew script names and their transcriptions, together with their extant historical attestations, a calculated historical span of use, and geographical coordinates where available. In order to enhance its interoperability and multilingual applicability we also matched our entities to existing open knowledge resources of contemporary and historical places: Geonames and Wikidata entities.

### Background

Hebrew place names are a complex, problematic and fascinating phenomenon which stretches far beyond the spaces in which Hebrew speakers nowadays dwell. Their history spans over three Millennia: from biblical place names that constitute a common world heritage of geographical, occasionally spiritual and of partly metaphorical nature, to the turbulent, expansive, multilingual and diverse diasporic experience of Jewry around the world, an experience which mapped the world using various

cultural, linguistic and religious spatial layers. The Hebrew Script was used not only for pre-modern and modern Hebrew, but also for Aramaic and Yiddish, as well as for several families of Jewish Languages, such as the Judeo-Arabic, Judeo-Spanish (Ladino) and others - a variety that was engendered by diglossia and multiple contacts with local population.

### Research potential

The KIMA Gazetteer is unique in being based on textual attestations, thus enabling beyond the study, visualization and use of changing historical placenames, also a thorough study of placenames as a linguistic, discursive phenomenon. Beyond searching for specific historical place names, finding all the variants of a place name or mapping all place names mentioned in a text, it would also enable answering and visualizing queries such as:

- When and in which literary genres was Istanbul referred to as Koshta, Constandina, Constantinople or Stambuli, and how did the name history relate to its political history?
- What are the relations between the Judeo Arabic or Hebrew naming chronology of Fustat and Kahir and the naming chronology of Cairo in other languages?
- What were the placenames associated with Zion, Jerusalem, Babylon and Gehenna, having been dissociated from their original geo-location?
- How did the spatial migration of area names like Ashkenaz, Zarfath and Sefarad play in the various Jewish languages?

### Data and work, present and future

As our primary core sources we processed and merged 6,351 place names from a digital version of the *Biblia Hebraica Stuttgartensia* (BHS) which was made available in the text database of the Hebrew Bible behind [SHEBANQ](#) (a system for the study of the Hebrew Bible, produced by the Eep Talstra Centre for Bible and Computer (ETCBC) at the Vrije Universiteit (WIVU) in cooperation with the German Bible Society and the Netherlands Bible Society); over 68,000 attestations of placenames in the body of texts compiled and analysed in the Maagarim corpus of the Academy of the Hebrew Language; 108,000 original place of publication fields from the Bibliography of the Hebrew Book (BHB) database and from around 300,000 Hebrew catalogue records from the National Library of Israel. Refining, matching, merging and

mapping the data meant a fascinating engagement with issues of Hebrew NLP, with human and machine error and doubt, with geopolitical and philosophical questions of the identity of places.

During Summer 2017 we will be working to integrate resources of Medieval Hebrew travel literature and early Geography compositions as well as Aramaic, Ladino and Judeo Arabic attestations, which will enhance the multilingual nature of gazetteer and its interoperability with other languages (respectively German, Spanish and Arabic). This endeavor presents new challenges of text annotation, of digitization of print Gazetteers and of dealing with OCR output. Also planned is an API and website for advanced query and mapping, a moderated input-export system, and collaboration with academic and other institutions for manual or semi-automatic extraction of place names from specific texts and corpora.

### In the poster:

- A description of the project's workflow, resources, and tools: the conception and construction of a text-based historical gazetteer
- Our annotated Schema (see figure), with explanations of our choices
- Examples for some of the challenges encountered in data resolution and merging, and their solutions (including, for example, different systems of notation for chronological information)
- Statistics of the Results in various stages of work
- Selected visualizations of the database and several use cases of its query, within and beyond Recogito.

### Bibliography

**Elliger, K., and Rudolph, W.** (1977). *Biblia Hebraica Stuttgartensia*. 5<sup>th</sup> Revised Edition, edited by Adrian Schenker. 1997. Deutsche Bibelgesellschaft, Stuttgart

