Linked Places in World Historical Gazetteer

Karl Grossner
World History Center
University of Pittsburgh
Pittsburgh, PA USA
grossner@pitt.edu

Ruth Mostern
Department of History
University of Pittsburgh
Pittsburgh, PA USA
rmostern@pitt.edu

ABSTRACT

World Historical Gazetteer (WHG)[1] is a novel web software platform providing several services to historical researchers from social science and humanities disciplines who are concerned with places and the geographic dimensions of their research subjects. WHG was launched in July 2020 and is in continual development. Version 2, released in August 2021, is fully functional but not yet feature-complete.

To enable the integration of place datasets modeled in a wide variety of ways, the WHG project has led the collaborative development of Linked Places format (LPF)[2]. LPF adds temporal extensions to the GeoJSON standard widely used in web mapping applications. It also adheres to the JSON-LD 1.1 specification[3], making it compatible with current Linked Data practices[4].

This paper introduces the WHG platform and Linked Places format: their conceptual basis, their features, some ongoing work-in-progress, and some as-yet unmet challenges.

CCS CONCEPTS

• Spatial-temporal systems • Users and interactive retrieval • Semantics and reasoning

KEYWORDS

Gazetteers, place, toponomy, place data aggregation, historical research

1 Motivation

It is not hard to imagine the following: a digital map of the world displayed on a web page with an accompanying time control, the earliest date for which is 9600 B.C., the latest 2021. The control's pointer is all the way to the left and there is a single marker displayed, indicating the city of Jericho. A map user can manually move the pointer to any position (year), or press "play" and watch as dots accumulate as the earth is settled by humans over time. At any given position of this time slider, one could select a marker and in a *place portal* screen learn many basic facts about the place: its known or estimated span of existence; the names it has been known by and where in historical sources they appeared; the kind of place it has been over time, e.g. transitioning from hamlet, to town, to city; and the administrative territories and asserted regions that have "contained" it in some respect. Pressing a 'contained by' button would display area geometry for those territories and regions. Beyond those basic facts, the portal screen would list any number of scholarly research projects in which the place had been referenced, and provide links to project pages where one could learn why.

We can conceive such a system—a comprehensive world historical gazetteer supporting an atlas web application—but realizing it is an enormous undertaking. The technology to assemble the gazetteer has become available in recent years [1], and an implementation in the WHG project is described here. Significant challenges remain however, in three main areas: acquiring the content, representing the uncertainty and vagueness inherent in historical data in data models and on maps, and sustaining the resource as a permanent piece of Digital Humanities and Social Science infrastructure.

The intellectual and disciplinary perspective of the WHG project is that of World History. As such, its development aims to facilitate representing connections between places. Some ways this is being accomplished are discussed below.

2 Use cases

The atlas described above is only one type of application historical gazetteers can support. One of the most important use cases WHG is designed to support is the geocoding of references to places found in historical sources. Researchers studying the past from numerous disciplinary perspectives want to discover and understand the geography of their sources. This requires resolving the places references found in textual and tabular material to actual or estimated geographic coordinates; and this requires digital historical gazetteers and/or published historical GIS systems. There are several of both in existence, each focusing on a particular region

ACM SIGSPATIAL Workshop, November, 2021, Beijing, China (virtual)

[1] World Historical Gazetteer v2; https://whgazetteer.org
[2] Linked Places format; https://github.com/LinkedPasts/linked-places-format
[3] JSON-LD 1.1; https://w3c.github.io/json-ld-syntax/
[4] Linked Data (W3C); https://www.w3.org/standards/semanticweb/data

and period, but they require individual manual lookups or the building of a scripted solution using a large dump file.

The tasks of digitizing texts, identifying place references in them, then discovering usable coordinate locations for them and creating maps and/or performing spatial or network analyses can be an excellent teaching exercise, and the reconciliation services built in to WHG support key pieces of that workflow.

Another of the primary use cases addressed so far in WHG development involves researchers from multiple projects who are studying a particular region and period from any combination of disciplines and thematic interests. We have identified several such communities of interest who are interested to collaboratively assemble a comprehensive specialized gazetteer resource by linking their place datasets. We note that this model of collaboration has produced the successful and widely used Pleiades project[5], self-described as "a community-built gazetteer and graph of ancient places," focused on the Classical Mediterranean. WHG aims to be a venue for assembling others, and at this writing, efforts to build "focus domains" within WHG for colonial and pre-colonial Latin America, Dutch History, the Middle East, and Central Eurasia are under way.

3 Linked Places format

Researchers gather and manage their data using a wide variety of models and formats. With the advent of semantic computing and the Linked Data paradigm, many projects endeavor to implement standard vocabularies and ontological models to facilitate interoperability. Arguably, these efforts are most successful within disciplines[6]. But research concerned with geographical locations and distributions is performed in almost every field. As a result there is no standard way of representing place in historical research data [1]; however there is some interest in GIScience in developing one [2][3].

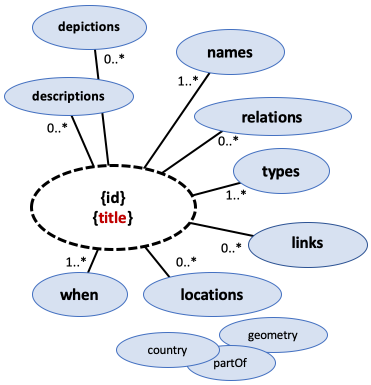

Figure 1: Conceptual model of a Place attestation in Linked Places format

The WHG project has taken a very pragmatic first step, driven by our need for a common contribution format. A data integration system like WHG cannot possibly parse all of the data models and formats in use by contributors. Data comes from spreadsheets, relational databases, GIS systems, and RDF files, with no commonality in terms used.

In collaboration with the Pelagios 7 project[7], a progenitor of WHG, we have developed the Linked Places format (cf. conceptual model, Figure 1), which has the following characteristics: (i) it is valid JSON-LD 1.1; (ii) it is valid GeoJSON; (iii) it extends GeoJSON with a temporal element, "when," which can be used to temporally scope a Feature as well as many of its attributes: names, types, geometry, and relations with other places. The structure of a "when" object (Figure 2) enables some expression of uncertainty, in that it can include one or more timespans, each potentially bounded by an earliest/latest pair of ISO-8601 dates. It can also include dereferenceable URIs for named time periods from a temporal gazetteer such as PeriodO[8] or ChronOntology[9]. The temporal extension to GeoJSON derives from the "Setting" pattern presented by Grossner, et al. in [4].

For WHG, the majority of Linked Places elements are optional; some data contributions are elaborate and detailed, but the majority are not. For that reason, we have also developed a delimited file alternative for contribution, LP-TSV,[10] that is adequate for datasets that do not temporally scope individual attributes. For both, a conversion to an internal representation in a relational database is made automatically upon upload.

Linked Places format is seeing growing interest and adoption. We consider it a "living standard," which may be adjusted and improved as more exemplar data is encountered.

```
"when": {
  "timespans": [{"start": {"in":"0676"},"end": {"in":"1066"}}],
  "periods": [
    { "name": "Anglo-Saxon Period, 449-1066",
      "@id": "periodo:p06c6g3whtg"},
    { "name": "Anglo-Saxon (culture or style)",
      "@id": "http://chronontology.dainst.org/period/O5r960WKERYr"}
  ],
  "label": "sample 'when' w/timespans, periods, duration",
  "duration": "P100Y"
},
```

Figure 2: Structure of a "when" object in Linked Places format

4 World Historical Gazetteer

4.1 Data stores

The World Historical Gazetteer platform is a Django (Python) project that makes extensive use of several JavaScript libraries, including MapLibre.js, Leaflet, and Turf.js. It has two data stores—a PostgreSQL/PostGIS relational database, and an Elasticsearch index. Datasets are uploaded to the database initially and

[5] Pleiades; https://pleiades.stoa.org
[6] For example, Linked Art is a CIDOC-CRM profile; https://linked.art/
[7] Now the Pelagios Network; https://pelagios.org
[8] PeriodO (Periods, Organized); https://perio.do/en/
[9] ChronOntology; https://chronontology.dainst.org/
[10] LP-TSV; https://github.com/LinkedPasts/linked-places-format/blob/master/tsv_0.3.md

maintained there, with permanent identifiers minted for each record. A further accessioning step indexes records as JSON documents, linking them using a parent/child structure offered by Elasticsearch. The first record indexed for a place is tagged as a "parent." Subsequently contributed record(s) identified by their creators as being a skos:closeMatch[11] to are tagged as a "child." In this way, WHG assembles over time a union record for each attested place, composed of records from one or more dataset. Individual datasets remain as discrete data objects in the relational database, and can be made public (and queried) prior to the accessioning/linking step.

4.2 Search and browse

The WHG web site includes search functionality with several pre- and post-result filtering options. Searches can be performed against all public datasets in the database or against the union index described above. Search facets include name, feature class and type, attested year, and spatial containment. Public browse screens for datasets and dataset collections present maps and individual record details.

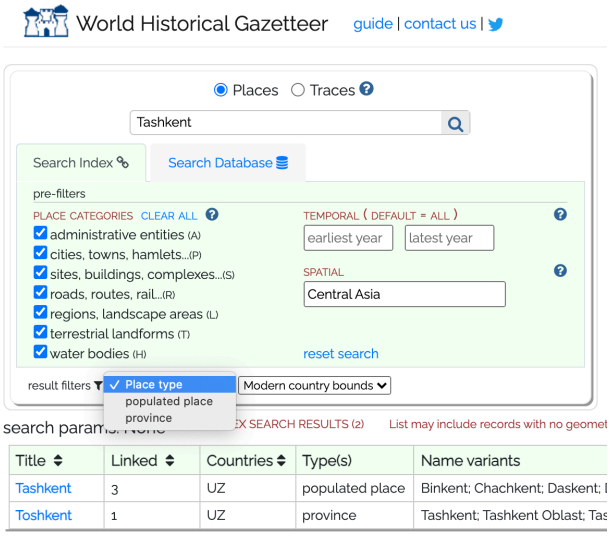

Figure 3: Pre- and post-search faceted filtering

4.3 Reconciliation and review

After exporting place records from a project dataset and converting them to either full Linked Places format or the LP-TSV alternative, the typical workflow for a contribution includes these steps, supported by WHG screens:

- Upload the LP file, and confirm its contents in dataset management screens
- In order to enhance the uploaded dataset and prepare it for future accessioning, initiate reconciliation task(s), seeking matches with either Wikidata[12] or Getty Thesaurus of Geographic Names (TGN)[13] or both.
- Review the results of the reconciliation task: for each record that got one or more potential matches (hits), decide whether any are a skos:closeMatch with it
- Request dataset be flagged as public (subject to review by WHG editorial staff)
- Request accessioning to the union index. This amounts to reconciling to the WHG index, and is followed by a review step.

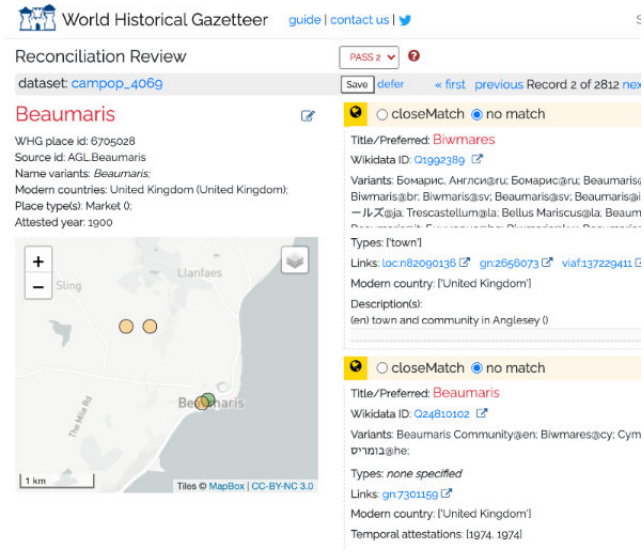

Figure 4: Reconciliation review. Uploaded record (left) compared against possible matches (right)

4.4 Collections

Support for the use case of a community of interest collaboratively assembling a "focus domain" within the WHG platform takes the form of a *Collections* feature. A collection can be defined by any registered user, as a set of two or more datasets which have been flagged as public. Datasets need not be be fully accessioned into the WHG index to be added, and can come from multiple sources/creators.

To date, collections are presented in a public-facing page with a description, an integrating map and somewhat limited browse and search functionality. Two examples are "HGIS de las Indias" and "Dutch History." In the first case, a single creator has joined separate datasets of settlements and administrative territories for the Spanish colonial Americas of the 18th century. In the second case, five datasets created by multiple researchers whose work concerns Dutch history have been linked, with many more anticipated.

4.5 Traces

One way connections between places are revealed in WHG is that many contributed datasets derive from research concerned with particular themes. The places in such datasets are by definition

[11] skos:closeMatch; https://www.w3.org/TR/skos-reference/#L4858
[12] Wikidata; https://wikidata.org
[13] Getty TGN; http://www.getty.edu/research/tools/vocabularies/tgn/

connected, and constitute a spatial footprint of the phenomena the project studied. For example, a recent contribution of the birthplaces of priests in 16c. Netherlands asserts that meaningful connection between those places.

Another way of representing connection is with "traces." Trace annotations are an experimental feature in WHG, and an alternate way to represent connection. We define *traces* as any historical phenomena of interest—events, people, objects, etc. *Trace data* are web resources with permanent URIs that in some way describe trace phenomena. *Trace annotations* are records that annotate those resources with URIs for places that are or were related to the trace in some way.

By accessioning such annotation data, WHG is able to display on "place portal" pages, links to research that concerned the place in some way: e.g. journeys for which it was a waypoint, objects for which it was a find spot, texts or images for which it is a subject.

5 Going forward

5.1 Outstanding challenges

The WHG vision in Section §1 is ambitious—a "moonshot" even—and so even at this development statge the WHG platform remains a work-in-progress. The goal of enabling the collaborative creation of multi-vocal descriptive temporally-scoped graphs for all the world's places has not to our knowledge been undertaken previously, so WHG is in a sense experimental. Some outstanding conceptual issues include:

- The conversion of place data from formats used in research projects to Linked Places format places a burden on researchers, so more and better tools and instruction aiding that task are essential
- There is no way to enforce criteria for making skos:closeMatch assertions. As a result, e.g. a record for a settlement may in some cases be linked with those for its containing municipality, and in other cases not. It remains to be seen whether such inconsistency is acceptable to the community of users.
- WHG needs long-term institutional support if it is to one day be the critical piece of historical research infrastructure it aspires to become.

5.2 Future work

Several enhancements to existing functionality and usability are under way at this writing, and more are outlined in a recent application for significant new funding.

5.2.1 Work in progress

5.2.1.1 *Teaching with WHG*. A new section of the WHG site will present (i) an overview of how place data, and WHG specifically may be used in teaching at secondary and post-secondary levels; and (ii) several lesson plans making use of existing WHG features

5.2.1.2 *Format conversion tools*. Users require tools to assist conversion of data from their formats to either LP format or LP-TSV. We are building tools to guide users through conversion of their data from RDF, shapefile, and delimited file formats to upload files WHG can parse.

5.2.1.3 *Ontology development*. LP format is valid JSON-LD, and therefore valid RDF, but it introduces a number of terms for entities and properties. Creation of a Linked Places Ontology (lpo:) to formally define those terms is under way.

5.2.2. Planned enhancements

5.2.2.1 *Expanded Collection capabilities*. Currently, collections can be defined only by listing as members a set of public datasets. We will allow collections to also be created by tagging individual records discovered during browse and search activity on the site, and significantly expand the collection-level search and download capabilities.

5.2.2.1 *Trace annotation datasets*. Example trace data now on the site was manually added to a traces index. We will enable the uploading of trace datasets, much as place datasets are now, and significantly enhance mapping, search and browse functionality for them.

5.2.2.1 *Historical mapathons*. An enormous amount of historical geographic data is locked away in print maps held as high resolution scans in web-based map collections. These maps represent an important potential source for historical place names and estimated geographic coordinates, which can be obtained by tracing features from them. We will continue early discussions with developers of the IIIF-related Allmaps project [14], who are developing ground-breaking tools for that digitization work.

ACKNOWLEDGMENTS

World Historical Gazetteer is a project of the World History Center at the University of Pittsburgh (WHC), which has since 2017 provided continuing in-kind and financial support for it. Ruth Mostern is Director of the WHC and also serves as the WHG project director. The initial 3-year phase of WHG development was funded by a US National Endowment for the Humanities (NEH) grant. Version 2 development in 2021 was made possible by financial support and collaborative input from the Humanities Cluster of the Royal Dutch Academy (KNAW). WHC staff and Research Assistants have been indispensable in the work performed to date.

REFERENCES

[1] Berman, M.L., Mostern, R., & Southall, H. 2016. *Placing Names: Enriching and integrating gazetteers*. Indiana University Press: Bloomington, IN.

[2] Tang, V., & Painho, M. 2021. Operationalizing places in GIScience: A review. *Transactions in GIS*, 25(3), 1127-1152.

[3] Ballatore, A. 2016. Prolegomena for an ontology of place. *Advancing geographic information science*, 91-103.

[4] Grossner, K., Janowicz, K., & Keßler, C. (2016). Place, period, and setting for linked data gazetteers. in Berman, M.L., Mostern, R., & Southall, H. (Eds.) *Placing names: Enriching and integrating gazetteers*, 80-96.

[14] Allmaps (Bert Spaan, principal developer); https://allmaps.org/