

# 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)<sup>1</sup> is a novel web software platform providing services to historical researchers in 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 was released in August 2021. To enable the integration of place datasets from multiple sources, the WHG project has led the collaborative development of Linked Places format (LPF)<sup>2</sup>, which adds temporal extensions to the GeoJSON<sup>3</sup> standard. It also adheres to the JSON-LD 1.1 specification<sup>4</sup>, making it compatible with current Linked Data practices<sup>5</sup>. This paper introduces the WHG platform and Linked Places format: their conceptual basis, their features, ongoing work-in-progress, and challenges being actively addressed.

## CCS CONCEPTS

• **Information systems** → **Information systems applications**;  
Spatial-temporal systems • **Information systems** → **Information retrieval**; Users and interactive retrieval

## KEYWORDS

Gazetteer, place, toponymy, data aggregation, historical research

### ACM Reference format:

Karl Grossner, Ruth Mostern. 2021. Linked Places in World Historical Gazetteer. In *5th ACM SIGSPATIAL International Workshop on Geospatial Humanities (GeoHumanities '21)*, Beijing, China. ACM, New York, USA, 4 pages.  
<https://doi.org/10.1145/3486187.3490203>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).

GeoHumanities '21, November 2–5, 2021, Beijing, China

© 2021 Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-9102-3/21/11...\$15.00

<https://doi.org/10.1145/3486187.3490203>

## 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 a single map marker is 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 can select a marker and in a *place portal* screen learn 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. hamlet to town to city; and the administrative territories and regions that have “contained” it in some respect. Pressing a ‘contained by’ button would display area geometry for those territories and regions. Beyond such 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, and an implementation in the WHG project [1] is described here. Challenges remain however, in three main areas: acquiring 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 Sciences 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

<sup>1</sup> World Historical Gazetteer v2; <https://whgazetteer.org>

<sup>2</sup> Linked Places format; <https://github.com/LinkedPasts/linked-places-format>

<sup>3</sup> GeoJSON; <https://geojson.org>

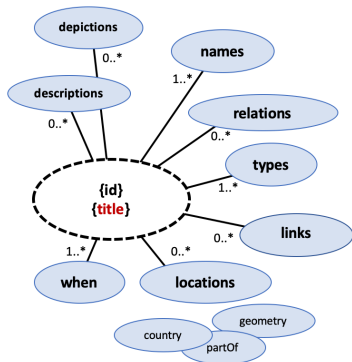
<sup>4</sup> JSON-LD 1.1; <https://w3c.github.io/json-ld-syntax/>

<sup>5</sup> Linked Data (W3C); <https://www.w3.org/standards/semanticweb/data>

disciplinary perspectives want to discover, visualize, 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, which in turn requires digital historical gazetteers and/or published historical GIS systems. There are several of both in existence, each focusing on a particular region and period, but they require individual manual lookups or the building of a scripted solution run against 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 also 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 interested to collaboratively assemble a comprehensive specialized gazetteer resource by linking their place datasets. Note that this model of collaboration has produced the successful and widely used Pleiades project<sup>6</sup>, self-described as “a community-built gazetteer and graph of ancient places,” focused on the Classical Mediterranean. WHG is a venue for assembling others, and at this writing, the creation of such focused domains using the WHG Collections feature are under way for colonial and pre-colonial Latin America, Dutch History, the Middle East, and Central Eurasia. Such collections can be exported for use in annotation platforms like Recogito<sup>7</sup>.



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

### 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<sup>8</sup>. But research concerned with geographical locations

and distributions is performed in almost every field. Although there is no standard way of representing place in historical research data [1]; there is some interest in GIScience in developing one [2][3].

The WHG project has taken a pragmatic first step, driven by our need for a common contribution format. Data integration systems like WHG cannot parse data from all models and formats in use by contributors, originating from spreadsheets, relational databases, GIS systems, and RDF files, with no common vocabularies.

In collaboration with the Pelagios 7 project<sup>9</sup>, a progenitor of WHG, we have developed the Linked Places format (conceptual model, Figure 1), which has the following characteristics: (i) it is valid JSON-LD 1.1, making it valid RDF; (ii) it is valid GeoJSON; (iii) it extends GeoJSON with a “when” element that can be used to temporally scope an entire Feature or several of its attributes in any combination: names, types, geometry, and relations with other places. The structure of a “when” element (Figure 2) enables some expression of uncertainty, in that it can include one or more timespans, each potentially bounded by earliest/latest pairs of ISO-8601 dates. It can also include dereferenceable URIs for named time periods from a temporal gazetteer such as PeriodO<sup>10</sup> or ChronOntology<sup>11</sup>. 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. While some data contributions are very detailed, the majority are not. For that reason, we have also developed a delimited file alternative, LP-TSV,<sup>12</sup> adequate for simpler datasets. For both, a conversion to an internal LP representation in the WHG relational database is made automatically upon upload, and exports as “full” LP format are available.

Linked Places format is seeing growing interest and adoption. We consider it a draft standard that is being adjusted and improved as more exemplar data is encountered.

```

"when": {
  "timespans": [
    { "start": { "in": "yyyy-mm?" },
      "end": {
        "earliest": "-yyyy%",
        "latest": "yyyy-mm-dd" }
    }
  ],
  "periods": [
    { "name": "Anglo-Saxon Period, 449-1066",
      "uri": "http://n2t.net/ark:/99152/p06c6g3whtg" }
  ],
  "label": "for a century during the Anglo-Saxon period",
  "duration": "P100Y"
}

```

**Figure 2: Structure of a “when” object in Linked Places format**

## 4 World Historical Gazetteer

### 4.1 Data stores and dependencies

The WHG 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

<sup>6</sup> Pleiades; <https://pleiades.stoa.org>

<sup>7</sup> Recogito; <https://recogito.pelagios.org>

<sup>8</sup> For example, Linked Art is a CIDOC-CRM profile; <https://linked.art/>

<sup>9</sup> Now the Pelagios Network; <https://pelagios.org>

<sup>10</sup> PeriodO (Periods, Organized); <https://perio.do/en/>

<sup>11</sup> ChronOntology; <https://chronontology.dainst.org/>

<sup>12</sup> LP-TSV; [https://github.com/LinkedPasts/linked-places-format/blob/master/tsv\\_0.3.md](https://github.com/LinkedPasts/linked-places-format/blob/master/tsv_0.3.md)

index. Datasets are uploaded to the database and maintained there, with permanent identifiers minted for each record. A further accessioning step indexes records as JSON documents, linking them using the 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`<sup>13</sup> to it are tagged as a “child.” In this way, WHG is assembling over time a union record for each place, composed of attestations (records) from one or more dataset. Individual datasets are maintained as discrete data objects in the relational database, and can be made public for search and browsing 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 facets, including name, feature class and type, attested year, and spatial containment. Searches can be performed against all public datasets in the database or against the union index described above. Public browse screens for datasets and dataset collections present maps and individual record details.

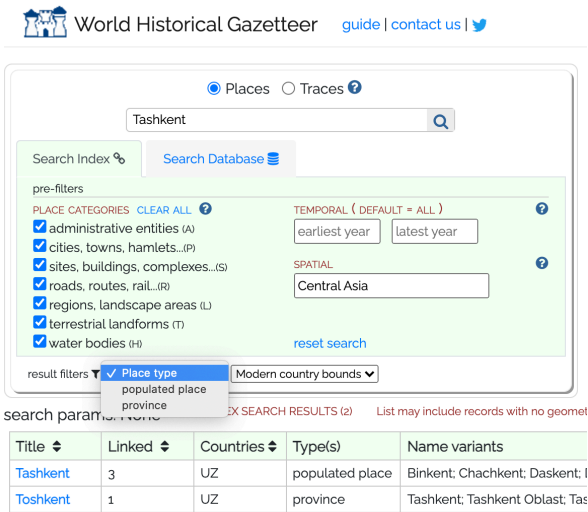


Figure 3: Pre- and post-search faceted filtering

### 4.3 Reconciliation and review

After contributors export their place records from a project dataset and convert them to either full Linked Places format or the LP-TSV alternative, the typical workflow 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<sup>14</sup> or Getty Thesaurus of Geographic Names (TGN)<sup>15</sup> or both.
- Review the results of the reconciliation task: for each record with one or more potential matches (hits), decide whether

any are a `skos:closeMatch` with it. This matching step requires expert human decisions and *cannot be automated*.

- Request the dataset be flagged as public (subject to review by WHG editorial staff)
- Request accessioning, i.e. reconciliation to the WHG index, with semi-automated review. When an incoming record shares a linked external identifier with an index record, it can be automatically linked as a “child” of the set’s “parent.”

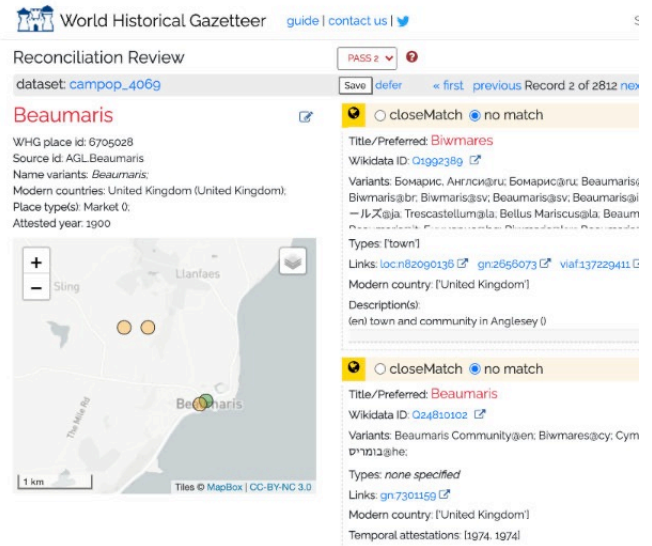


Figure 4: Reconciliation review. Uploaded record (left) is compared against possible matches (right)

### 4.4 Collections

Communities of interest collaboratively assembling a “focus domain” within the WHG platform are supported by the *Collections* feature. A collection can be defined by any registered user as a set of two or more public datasets. Datasets need not 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 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 18c Spanish colonial Americas. In the second case, six datasets created by multiple researchers whose work concerns Dutch history have been linked so far, with many more anticipated.

### 4.5 Connecting places with traces

Many of the datasets contributed to WHG derive from research concerned with particular themes. The places in such datasets are by definition connected, and constitute a spatial footprint of the phenomena the project studied. For example, the recent

<sup>13</sup> `skos:closeMatch`; <https://www.w3.org/TR/skos-reference/#L4858>

<sup>14</sup> Wikidata; <https://wikidata.org>

<sup>15</sup> Getty TGN; <http://www.getty.edu/research/tools/vocabularies/tgn/>

contribution of 16c birthplaces of priests in the Netherlands asserts a meaningful connection between those places.

Another way of representing connection is with “traces,” an experimental feature in WHG, and an alternative 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 trace annotation data<sup>16</sup>, WHG is able to display on place portal pages links to research that concerns 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 Meeting challenges

The vision described in Section §1 is an ambitious “moonshot,” so at this development stage the WHG platform remains a work-in-progress and is in a sense experimental. The challenges mentioned in §1 have been and are being addressed, in an ongoing process.

**5.1.1 Data acquisition.** Creating a copy of a project’s place data in LP format places a burden on researchers. The simplified LP-TSV format helps; more and better conversion tools and documentation are essential, as is continued community-building.

**5.1.2 Uncertainty and vagueness.** The 4-part structure of the LP format “when” element timespan (earliest/latest ranges for both start and end, and URIs to named periods) help with respect to dating. Geometries can be tagged as ‘certain’, ‘less certain’ or ‘uncertain’. Cartographic representations for these remain an area of current collaborative experimentation.

**5.1.3 Sustainability.** Two challenges shared by virtually all DH infrastructure projects are ongoing institutional support, and technological upkeep. We have established a collaborative partnership with the Humanities Cluster of the Royal Dutch Academies, and several organizational units within the University of Pittsburgh—steps towards building a supporting consortium. Permanent support can ensure technology updates as required.

### 5.2 Ongoing and future work

Several enhancements to existing functionality are under way at this writing, and many more are planned.

**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) 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 LP-compatible 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 Selected planned enhancements**

**5.2.2.1 Expanded Collection capabilities.** Currently, collections are defined 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, and significantly expand the collection-level search and download capabilities.

**5.2.2.2 Trace annotation datasets.** The current exemplar trace data was manually added to a traces index. We will enable the routine uploading of trace datasets and significantly enhance mapping, search and browse functionality for them.

**5.2.2.3 Historical mapathons.** We are collaborating with the Allmaps project<sup>17</sup> to facilitate integration with crowd-sourced feature extraction data from IIIF-formatted geo-referenced maps.

## 6 Summary

Numerous web-based digital resources catalogue place information, however the WHG and its implementation of LP format are distinctive in combining (i) global geographic coverage; (ii) structured, granular temporal information for both place records as a whole and important individual attributes; (iii) records that are attestations of place drawn from historical sources by researchers; (iv) multiple attestations for a given place linked using tools embedded in the system, thereby linking research products across projects and disciplines; (v) faceted record-level search across datasets; (vi) serving as a Linked Data publishing platform for smaller, specialist historical gazetteers; and (vii) providing machine access via an application programming interface (API).

## ACKNOWLEDGMENTS

World Historical Gazetteer is a project of the World History Center at the University of Pittsburgh (WHC)<sup>18</sup>, which has since 2017 provided in-kind and financial support. Ruth Mostern, Director of the WHC, serves as the WHG project director. The initial 3-year phase of WHG development was funded by a U.S. National Endowment for the Humanities 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 performed invaluable work in support of WHG throughout.

## 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.

<sup>16</sup> Patterns of W3C Web Annotation Format; <https://github.com/LinkedPasts/linked-traces-format>

<sup>17</sup> Allmaps; <https://allmaps.org/>

<sup>18</sup> World History Center; <https://www.worldhistory.pitt.edu/>