

Linking Knowledge Through Place: ISHI, the WHG, and the Future of Gazetteer Collaboration

For Presentation at Linked Pasts 11

December 9, 2025

Agenda

- **Introducing the University of Pittsburgh Institute for Spatial History Innovation (ISHI) and the Pelagios Place Activity (formerly Gazetteer Alignment Activity) (Ruth Mostern)**
 - Institute for Spatial History Innovation (ISHI) at University of Pittsburgh as new administrative home of World Historical Gazetteer. ISHI website: <https://ishi.pitt.edu/>
 - ISHI is the co-institutional stewardship, with [the Pelagios Network](#), of the network's Place Working Group (formerly Gazetteer Alignment Activity).
 - Planned work focuses on defining minimal standards for gazetteer alignment, providing clear guidance and templates for contributors, and maintaining an overview of URI-based gazetteers and alignment efforts. It also includes hosting workshops and collaborative activities while supporting community members preparing or aligning historical place datasets.
- **What's new at the World Historical Gazetteer (Stephen Gadd)**
 - Authentication via ORCID. The use of ORCID will ensure accurate attribution, persistent researcher identity, interoperability, and scholarly credit, while also maintaining accessibility beyond institutional affiliations, within the linked open data ecosystem.
 - New Entity and Reconciliation Service APIs: <https://docs.whgazetteer.org/content/technical/apis.html>
 - Improved reconciliation matching with Wikidata.
- **Revised WHG LP-TSV submission template (Ali Straub)**
 - Discussion focused on improving the LP-TSV template to make it more user-friendly.
 - A copy of the template can be found here: https://docs.google.com/spreadsheets/d/1DM4MmagkYkeUEMtB00QZdYGkrCfth3_FsUp6JGGPwa0/copy
- **New WHG Submission Readiness Checklist (Palak Vashist)**
 - A checklist outlining the required steps and criteria contributors must meet to successfully publish a dataset on the WHG.

- WHG Submission Readiness Checklist can be found on page 4 of this document.

Summary of Open Conversation by Topic

- **Modeling historical places from uncertain data**

- The [La Sfera project](#) involves building spatial representations (regions, polygons, lines) from largely point-based historical data. A major challenge is how to represent places that were real to historical actors (e.g., “China,” “Scythia”) but whose understood locations do not align with modern geographic knowledge.
- Vagueness, fuzzy boundaries, and uncertainty are inherent and need to be explicitly handled in the World Historical Gazetteer data model.
- Existing linked places formats, like those used in the WHG, allow expressions of uncertainty, but these features are under-documented and not well implemented. There is interest in improving or extending these models, possibly using numeric confidence scores (e.g., 1–100).
- One approach is allowing contributors to supply approximate geometries based on historical sources, or to contribute records without geometry but linked to broader place clusters.
- The goal is not perfect geometry but capturing historical reasoning about place. Recognition that many historical regions do not map neatly onto modern or classical boundaries, creating fuzzy intersections.

- **Gazetteers, linked data, and place clusters**

- Individual attestations of a place can coexist within larger clusters of possible meanings and locations.
- Gazetteers can accept diverse representations, including uncertain or source-specific geometries.
- Linking across datasets allows different interpretations of the “same” place to be preserved and contextualized.

- **Temporal scope and disciplinary boundaries**

- Ongoing tension and overlap between ancient, medieval, and modern place data.

- Ancient-world resources (e.g., Pleiades) are foundational, but there is a push to expand later temporal coverage.
- **Graph data models and infrastructure**
 - Ongoing development by WHG Technical Director of a graph-based data model, with clearer documentation and a public development roadmap. This documentation can be found here: <https://docs.whgazetteer.org/content/v4.html>
 - Emphasis on transparency about future versions and technical direction.
- **Curation strategy and content gaps**
 - The WHG index has grown opportunistically and now needs systematic analysis Identifying spatial and temporal gaps is a priority.
 - Future work should involve targeted data solicitation and clearer standards, drawing on historical texts, atlases, and maps.
- **Courses, training, and knowledge sharing**
 - Strong interest in creating a centralized portal or clearinghouse for syllabi, courses, and training materials related to spatial history and gazetteers. But concerns about sustainability and keeping resources up to date as well as not duplicating existing efforts.
 - Interest in a focused hub highlighting gazetteer-based and linked open data projects rather than duplicating broader GIS listings.
- **Partnerships and capacity building**
 - Future partnerships should focus on content areas beyond the core team's expertise. A potential model could be where the project provides standards and infrastructure while partners contribute domain expertise.
 - Specific interest in collaborations with projects like Enslaved.org and others working on underrepresented historical spaces.

WHG Submission Readiness Checklist

This checklist helps contributors confirm that their dataset is ready to upload into the WHG Workbench.

1. Choose the Correct Upload Format

(from “*Choosing an upload data format: LPF or LP-TSV?*”)

- I have selected **LPF** or **LP-TSV** based on my dataset needs.
- I understand that **both formats require one record per place**.
- If my dataset has multiple names, geometries, or temporal attributes per place → I chose **LPF**.
- If my dataset is simple (name, type, coordinate, single timespan) → I chose **LP-TSV**.

2. Extract One Record per Place

(from “*Preparing data for upload – extracting places*”)

- If my source file had one row per event, artifact, or observation → I extracted **unique places** into their own table.
- I confirmed that places are not repeated across multiple rows.
- I resolved ambiguities (e.g., different names that refer to the same place).
- I decided whether to normalize place scales (e.g., settlement vs county).

3. Prepare LP-TSV Data Correctly

(from “*Preparing data for upload – the simple case*”)

If I chose **LP-TSV**:

- I copied my data into the WHG **LP-TSV spreadsheet template**.
- My file contains **one row per place**.
- My place types and name variants fit within single columns (e.g., “name1;name2”).
- I did **not** edit TSV files using Notepad (because tabs will break).
I used Excel, Google Sheets, or LibreOffice.

4. Prepare LPF Data Correctly

(based on *LPF note in documentation*)

If I chose **LPF**:

- I understand LPF is JSON-based and requires a transformation script.
- My LPF includes temporally scoped attributes where applicable.
- My LPF includes citations for names, geometries, types, or time spans if my dataset uses them.

5. Uploading the Dataset

(from “*Upload dataset workflow*”)

- I accessed “Upload dataset” via **My Data** → + **button**.
- I filled in all required fields in the upload form (as required by the UI).
- I verified upload progress through the **Status** tab.
- I used the **Browse** tab to visually inspect the uploaded places on the map/table.

6. Reconciliation Preparation

(from “*Reconciliation & Accessioning*”)

Before initiating reconciliation:

- My place names are clear enough to be matched to Wikidata/GeoNames.
- My coordinates (if present) are positioned correctly.
- I understand that reconciliation returns prospective matches (“hits”).

Understanding reconciliation workflow:

- I know I must review each potential match.
- I know my options: **close Match**, **no match**, **defer**.
- I know deferred records go into a special queue.
- I understand PASS 0, PASS 1, PASS 2 (automatic vs scored matches).
- I know I can add notes to assist future review.

7. After Reconciliation

(from “*Reviewing accessioning results*”)

- I understand that if I accept a match, it becomes a **closeMatch** link.
- I understand the meaning of **closeMatch** based on SKOS definition.
- I understand “no match” means the record may become a **seed**.

- I know I can download augmented data after reconciliation.

8. Publication and Accessioning

(from “*Publishing Data*” and “*Accessioning to the WHG Union Index*”)

- After reconciliation, I completed required metadata in the **Metadata** tab.
- I submitted a publication request when the dataset was ready.
- I understand that accessioning (final integration) is initiated by WHG staff.
- I understand that accessioning may link my dataset with existing clusters.
- I know that unmatched records become “seeds” in the Union Index.