Uploading and Reconciling Place Data in World Historical Gazetteer
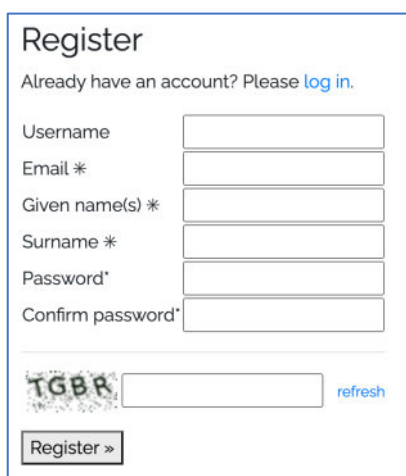*Getting started with the WHG platform*   *25 May 2023*

World Historical Gazetteer (WHG) provides services and tools allowing registered users to *upload*, *augment (reconcile)*, *manage*, and *publish* their place datasets. Other WHG features enable searching and browsing data, creation of annotated *Place Collections* and *Dataset Collections*, and using the WHG's API—but those are outside the scope of this guide.

The following steps walk through 1) registering as a WHG user, 2) uploading a delimited-format dataset file into a user's private workspace, 3) initiating a system task for reconciling (matching) the dataset records to WHG's internal index of 3.5 million Wikidata place records, and 4) reviewing the results—that is, making the matching decisions for those records that found prospective matches.
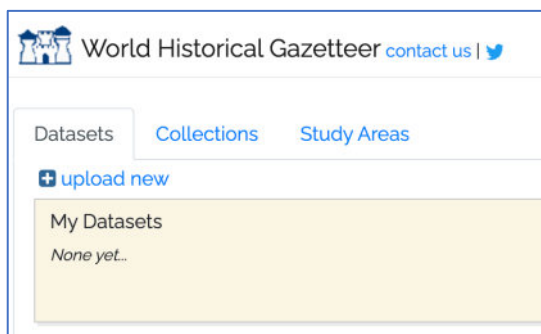
1. Registering as a WHG user
To work with data on the WHG site you must register (https://whgazetteer.org). Upon submitting the form you will receive an email asking you to confirm the registration.

Figure 1 - Registering as a WHG user

After logging in, you land on your "My Datasets" page, which is empty at this point.

Figure 2 - New empty private workspace

2. Uploading a delimited-format dataset file
WHG can accept data uploads in one of two formats: Linked Places format (LPF)—an expressive GeoJSON extension—and LP-TSV, a much simpler delimited tabular format (.csv, .tsv, or spreadsheet). In this guide we are working with the simpler LP-TSV.

Using the Quick Start
Click the "upload new" link, and you are brought to an "Upload dataset" form. On the right side is a brief explanation of requirements, including a "QUICK START" section providing a downloadable Google Sheets template for experimenting with the process; instructions for its use are in the spreadsheet itself. You will learn that you can either remove the data rows in it and copy/paste your own, or simply remove the explanatory material and load the 7 sample records.

Figure 3 - Upload form

Fill in minimum metadata

Only three fields are required for the initial dataset file upload: **title**, **label**, and **description**. A help popup explains the **label** field. All metadata fields except **label** can be added and/or edited at any future time.

Using sample data files

You will see a link to download a .zip file with several samples and template files in .xlsx and .ods format. Working with those is an alternative option.

Creating your own data files

The easiest way to prepare your own LP-TSV upload file is to use either the Quick Start Google Sheet template or one of the spreadsheets in the sample file download just mentioned. Copy columns of your data in to the appropriate columns within the template.

There are only four required fields (columns) in a valid LP-TSV file: **id**, **title**, **title_source**, and **start**. However, the more context you supply in a record, the higher the quality of candidate matches in the reconciliation step to follow. The fields **types** and **aat_types** are highly desirable, and of course geometry for those records when available: either **lon** and **lat** or **geowkt**.

The WHG site also provides a detailed tutorial for preparing an LP-TSV format spreadsheet.

Upload

After preparing a file in one of the ways described above, and filling out the minimum three form fields, click the Upload button. The file will be validated, and if successful, your data will be written to the WHG database—each record receiving a permanent **pid** identifier that is kept along with your id (now called **src_id** in the WHG database). *It is not unusual for something to go wrong the first time around!* Spellings of column headings, and leaving out a required column are common errors. But it's easy once you get the hang of it :^) Note that not all errors are captured and reported back perfectly—some result in an email response (copied to WHG staff), and we will troubleshoot any issues quickly.
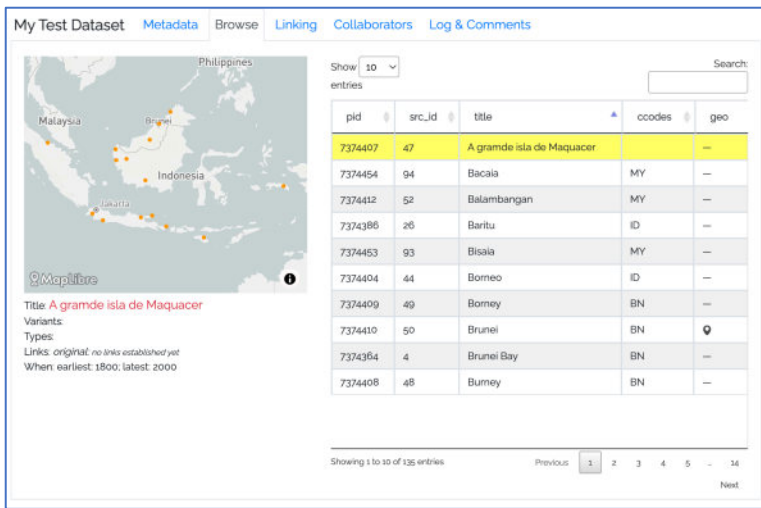

Figure 4 - Dataset Summary page

Figure 5 - Dataset Browse tab

3. Initiating a reconciliation task

Once your dataset is uploaded, you land on its "Summary" page where you can update its metadata, assign collaborators, and reconcile your records against WHG's Wikidata index—a way to augment your data with geometry and/or authority identifiers obtained by accepting a "hit" as a "close match" with your record. What constitutes a close match is explained in a help screen available on the "Reconciliation Review" screen.

From your dataset's Summary page, click the Linking tab and then the "Add new task" link.

A new "Add task" tab is generated, with a form and a map. Wikidata is preselected as the only available source, and you have an option to constrain the search spatially, with either a pre-defined region or a "Study Area" of your own (see section below). If you know the approximate coverage of your data, it is always best to create a Study Area, so that records lying outside that are not considered.
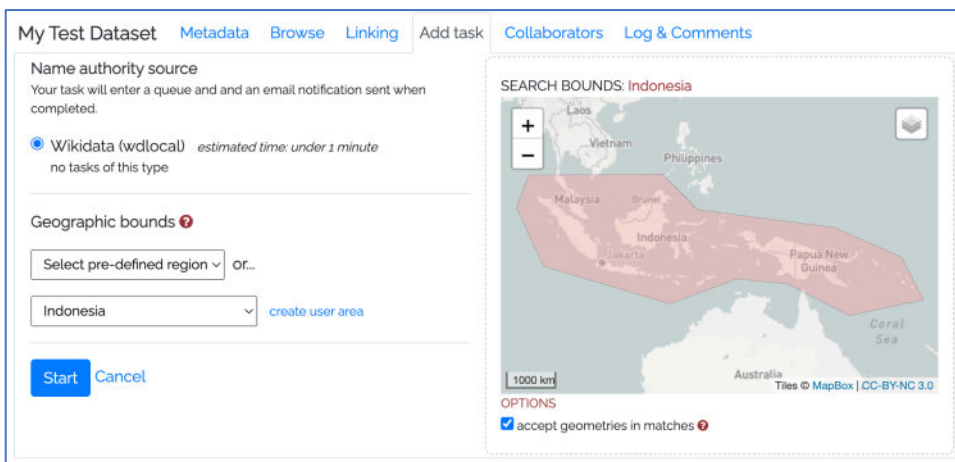

Figure 6 - Add Task tab with a custom Study Area chosen

After selecting or creating a study area (if you choose) click the "Start" button, and your reconciliation task is launched. The search processes about three records per second. When it is completed, you will receive an email notification.

3a. Creating a Study Area

For example, if your data is for Indonesia, a roughly drawn study area for it will exclude potential matches from other historical Dutch territories and the Netherlands, where places with the same names may be found. Instructions for defining a Study area by either drawing it on a map or listing modern countries are found on that page.
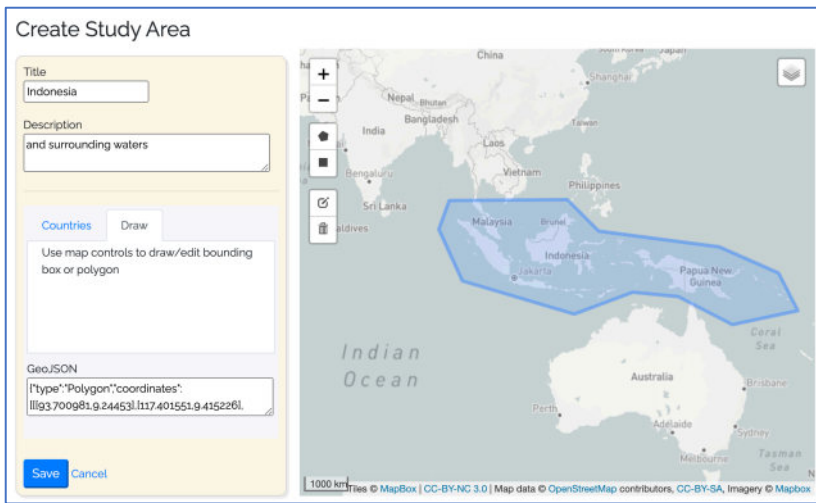
Figure 7 - Create Study Area

4. Reviewing reconciliation results

When your reconciliation task has completed, you will receive an email and your Linking tab will display a summary of the task results. The task makes up to 3 separate "passes" (queries) for each of your records (0, 1, 2). And these appear in separate queues. Candidate matches in Pass 1 will be higher quality than Pass 2, because Pass 2 "relaxed" the query of Pass 1 if nothing was found. Viable matches are normally found in all passes.
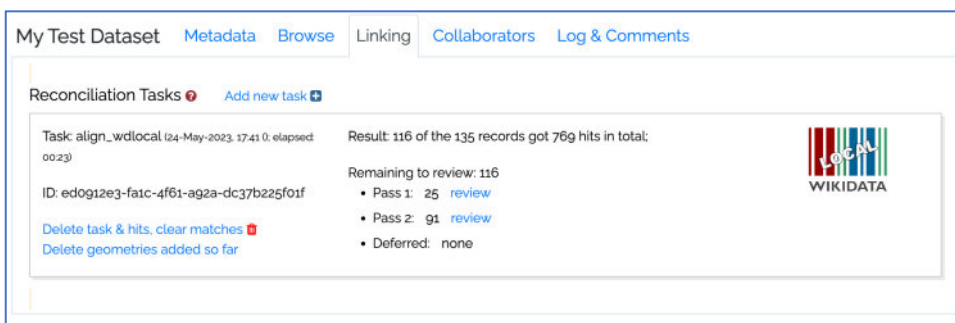

Figure 8 - Reconciliation task results summary

The next step is to review the candidate matches and decide if any are a close match with your record(s). That work is performed on the "Reconciliation Review" screen (below). On the left, those records from your dataset that got possible matches are displayed, one-at-a-time. On the right are one or more potential matches. At the top of the screen is a red help icon that loads a screen explaining every element of this page, *including what we mean by closeMatch!*
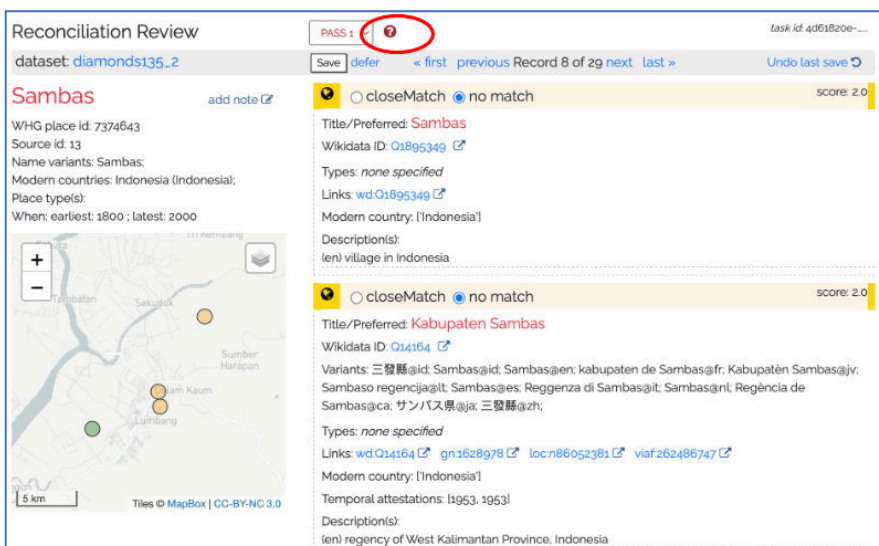

Figure 9 - Reconciliation Review screen

If your record included a geometry, its marker will be green. Map markers for any potential matches are orange. Hovering on the globe icon highlights the corresponding marker, and vice-versa: clicking a map marker will highlight the corresponding record in the list on the right.

If one or more of the "hits" on the right are a close match with yours, select "closeMatch" in place of "no match" and click "Save." If none of the hits are a match, just click "Save" and the page advances to your next record. Note that there is also a "defer" choice. Clicking "defer" will put that record and its hits in a separate queue you can return to at any time. There is also an "Undo last save" that can come in handy.

Note that you can add any number of registered WHG users as collaborators in a "member" or "co-owner" role; in this way multiple people can tackle the review work together— simultaneously or not.

Note also that you can monitor the progress of this Review step with features on the Browse tab of your dataset.

Show 50 ∨ entries		Review filters	✓ All		Search:	
			-----wikidata-----			
pid	src_id	title	Needs review (wd)	odes	geo	wikidata ?
			Reviewed (wd)			
7374723	93	Bisaia	Deferred (wd)	—	no hits
7374674	44	Borneo		ID	◉	☑
7374679	49	Borney		BN	—	no hits
7374680	50	Brunei		BN	◉	☐
7374634	4	Brunei Bay		BN	◉	☑
7374678	48	Burney		BN	—	no hits

Figure 10 - Monitoring review progress

FAQ: *"What happens when I accept a match?"*

When you assert a close match, one or more records are written to the database, associating your place record with the accepted match's geometry and links (authority identifiers from the Wikidata record). The system maintains these new *place_geom* and *place_link* records separately, so that the task, hits, and matches can be removed easily if desired. The new geometry and authority identifiers do however become part of your now augmented dataset. You can always download your original upload file or the newly augmented dataset at any stage.

5. What's next?
When all of the potential matches have been reviewed and match/no match decisions made, your dataset is publishable. WHG staff will be notified and will review the metadata to see if it is adequate for publication. When that is complete, we flag it as public, and it becomes searchable and appears on lists of public data.

A further step of indexing your data (*accessioning*) is highly desirable, but beyond the scope of this guide. Accessioning to the union index links your individual records with records for closely matched places from other datasets. As a result, attestations for the same place are returned in index searches, linking projects (and disciplines) in novel ways.