

Datasets from Maps: Extraction and Remix

Workshop on Spatiotemporal Knowledge
PNC 2019 Annual Conference and Joint Meetings

Oct 15-18, NTU, Singapore

Tyng-Ruey Chuang
Academia Sinica, Taipei, Taiwan
trc@iis.sinica.edu.tw

Joint work with Chih-Chuan Hsu and Huang-Sin Syu

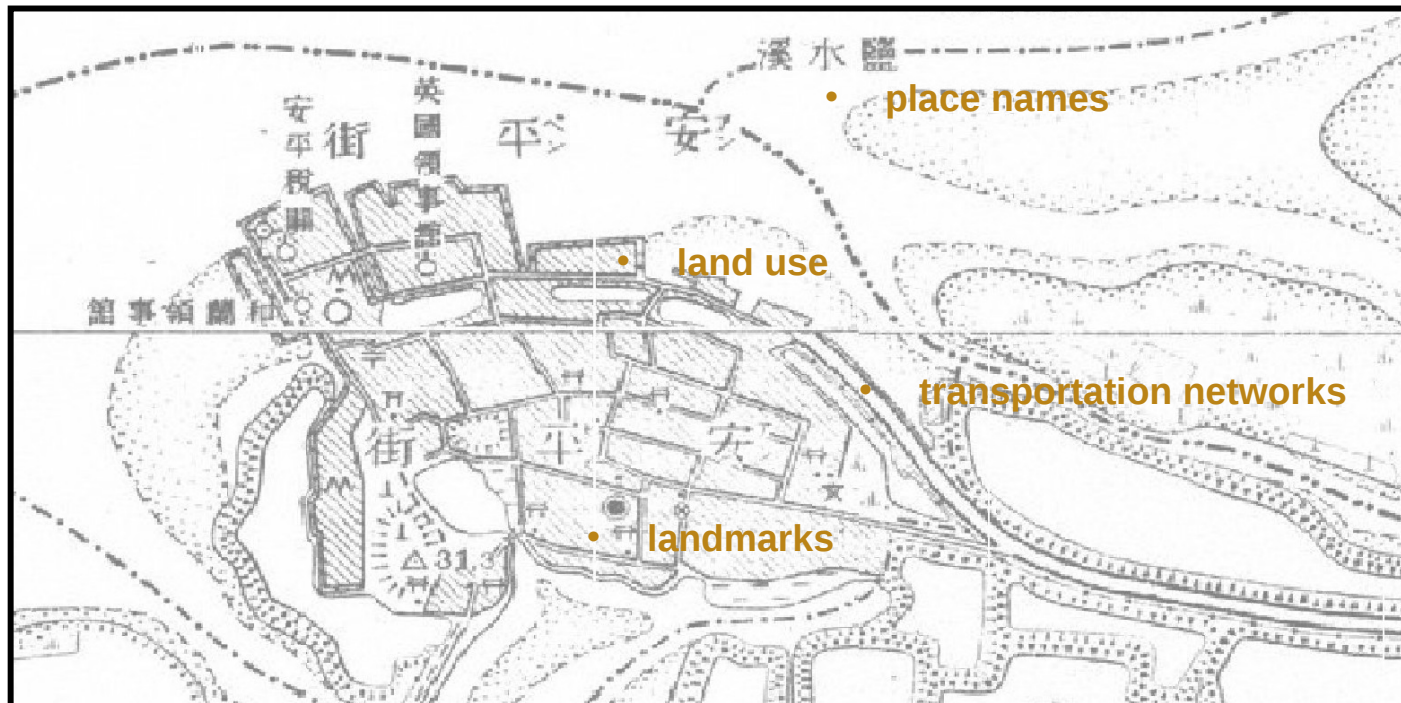
Distributed under the Creative Commons BY-ND 4.0 License

Outline

- Taiwan Baotu (台灣堡圖)
- OpenStreetMap
- Taiwan Baotu Re-imagined in OpenStreetMap
- Data Extraction Workflow
- Reuse and Remix of Datasets

Taiwan Baotu

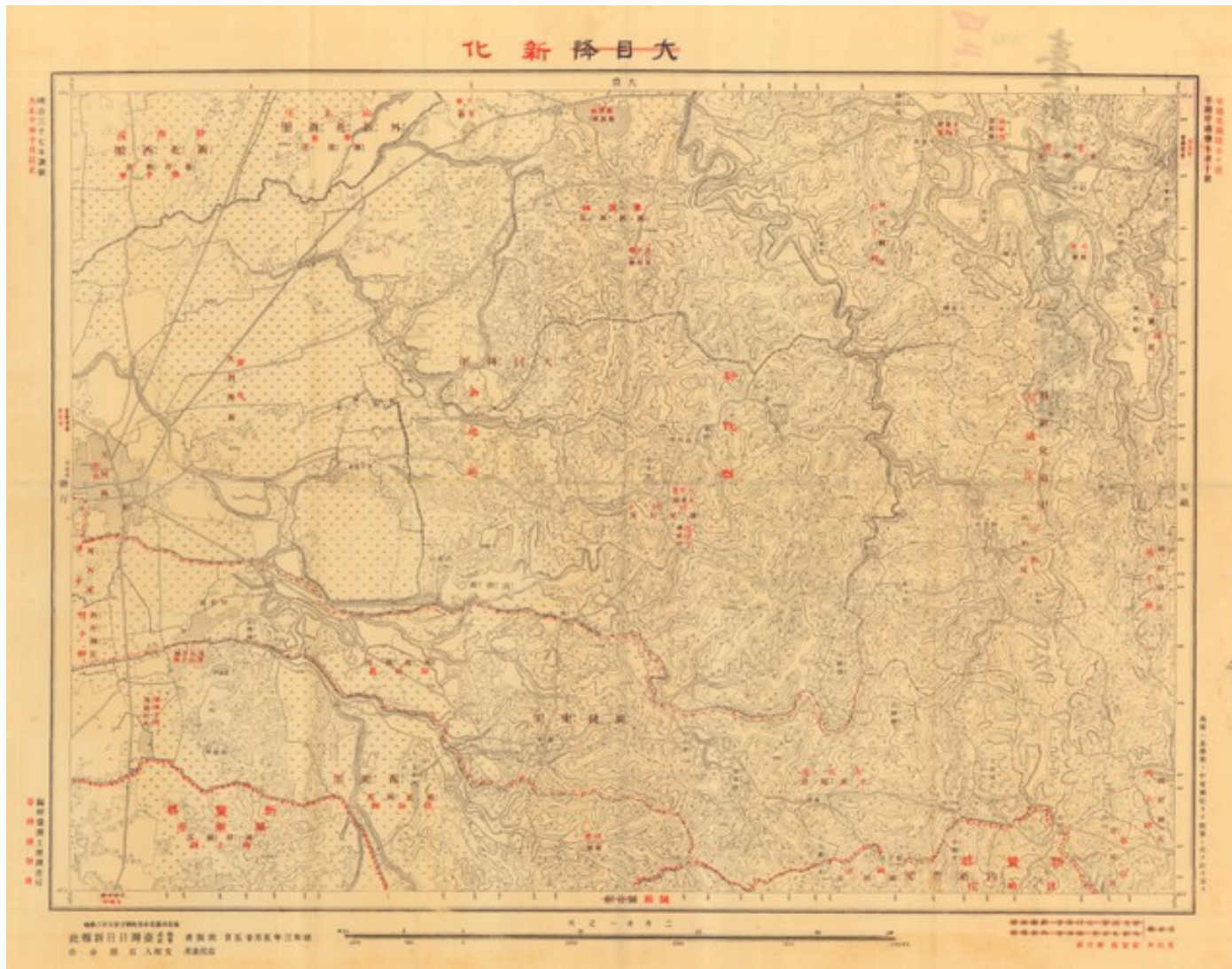
- Taiwan Baotu (台灣堡圖) , 1904
 - a set of 457 topographic maps of Taiwan, produced from systematic land surveys when the island was ruled by Japan
 - including information about **place names, land use, transportation networks, landmarks**, and others
 - scale of the map: 1/20,000



Sinhua 新化 大目降

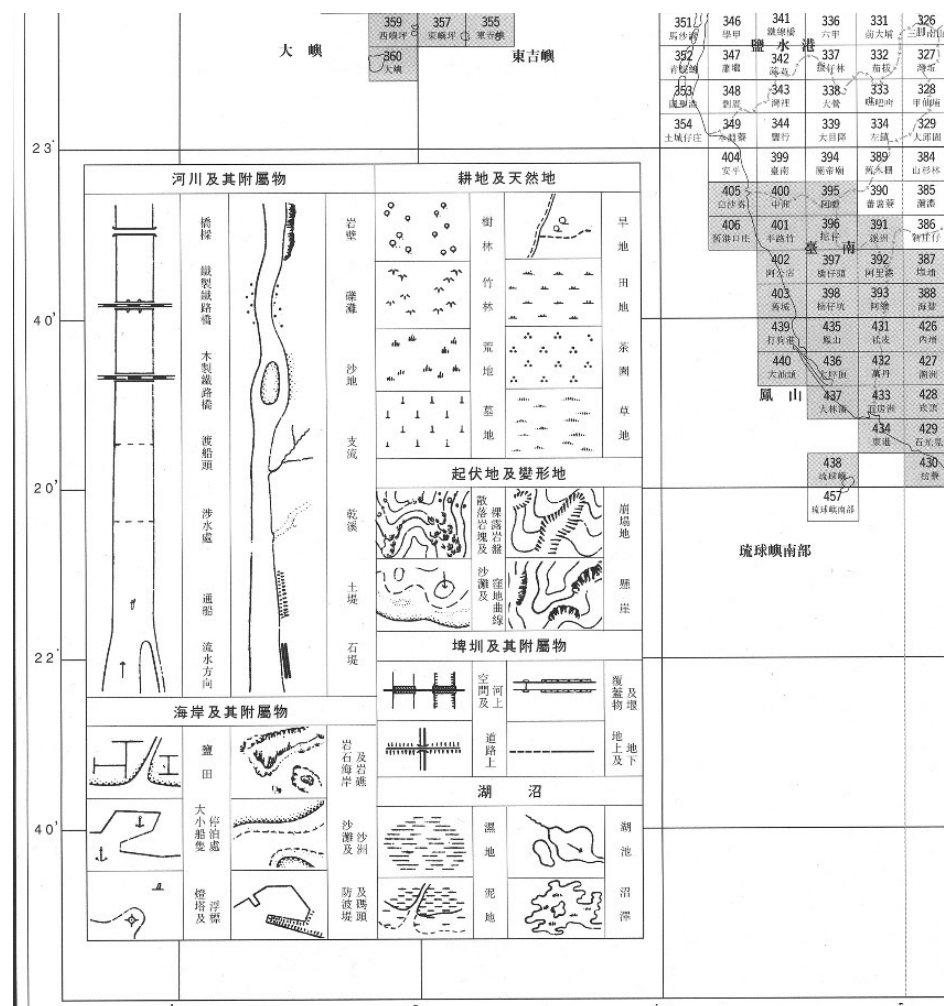
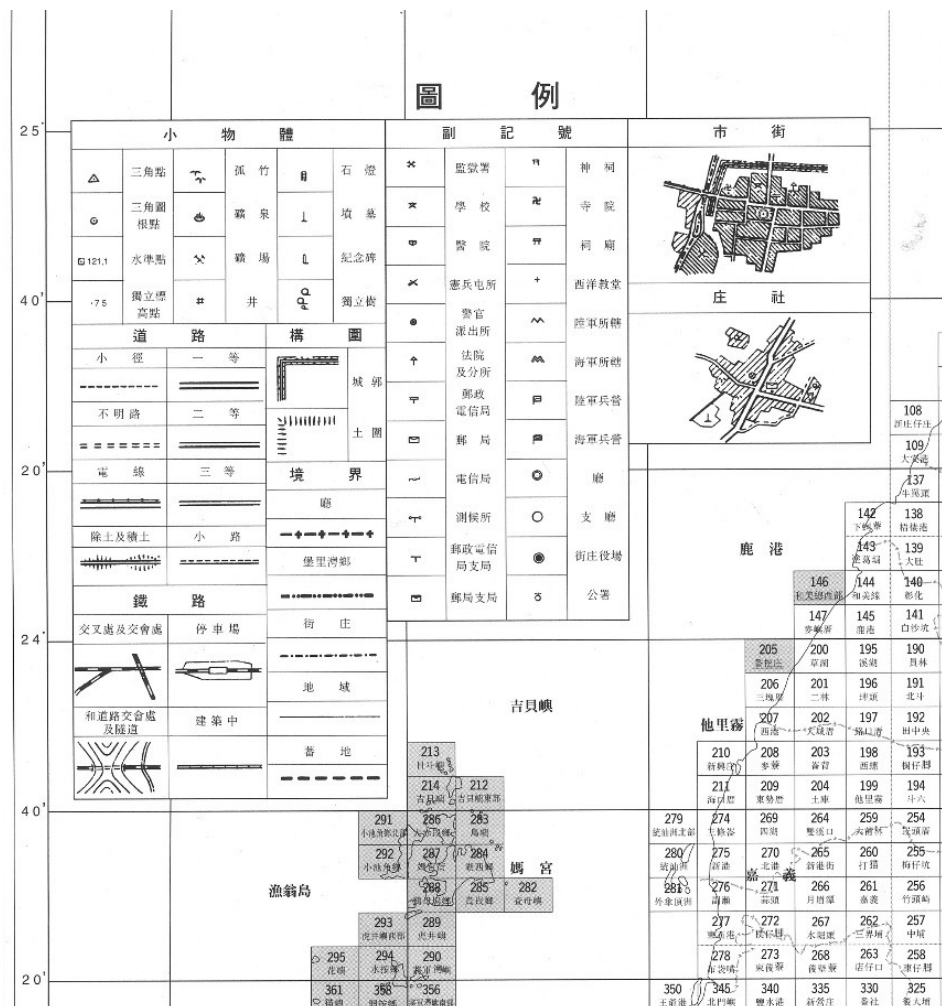
(Tovacon in the Siraya language)

<http://publicdomain.tw/u/pd4pd/m/Xin-Hua/>



See also <https://stanford.maps.arcgis.com/apps/SimpleViewer/index.html?appid=37dd40c1402440d985fd783a60d60443>

Taiwan Baotu Keys and Symbols



OpenStreetMap

- OpenStreetMap (OSM), 2004 –
 - Open Source and Open Data: Free software (GNU GPLv2) with collaborative mapping data (ODC ODbL)
 - a flexible toolkit for mapping and other online services

















OpenStreetMap

- Why in OpenStreetMap?
 - Reusable common infrastructure and data format
 - Abundant tools for processing, styling, and layering maps
 - Customizable map styling

OSM Features in (Key, Value) Pair

Vegetation or surface related

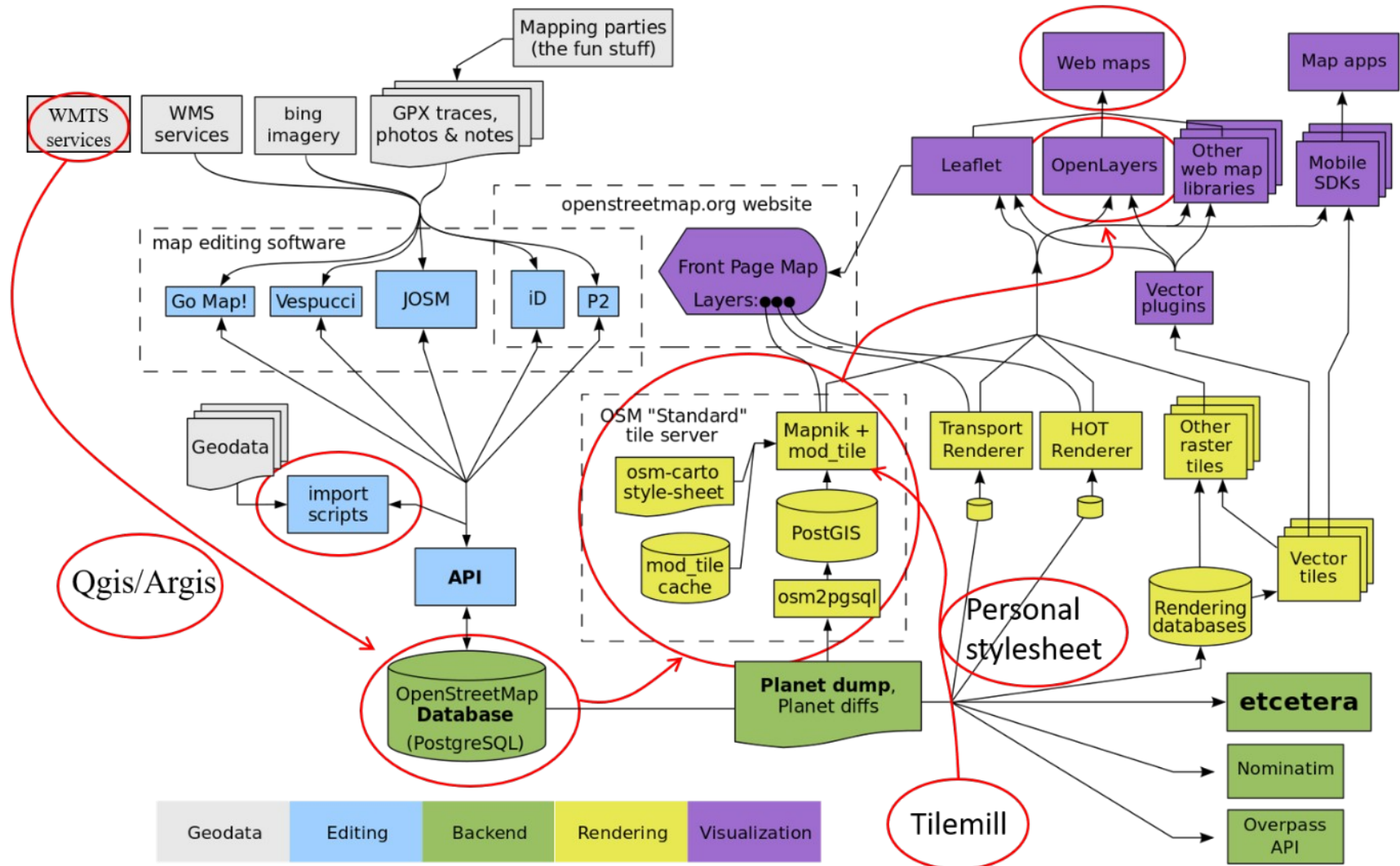
Key	Value	Element	Description	Image
natural	wood		Tree-covered area (a 'forest' or 'wood')	
natural	tree_row		A line of trees	
natural	tree		A single tree	
natural	scrub		Uncultivated land covered with bushes or stunted trees.	
natural	heath		A dwarf-shrub habitat, characterised by open, low growing woody vegetation, often dominated by plants of the Ericaceae.	
natural	moor		Don't use, see wikipedia	
natural	grassland		Natural areas where the vegetation is dominated by grasses (Poaceae) and other herbaceous (non-woody) plants.	

Taiwan Baotu Re-Imagined in OSM

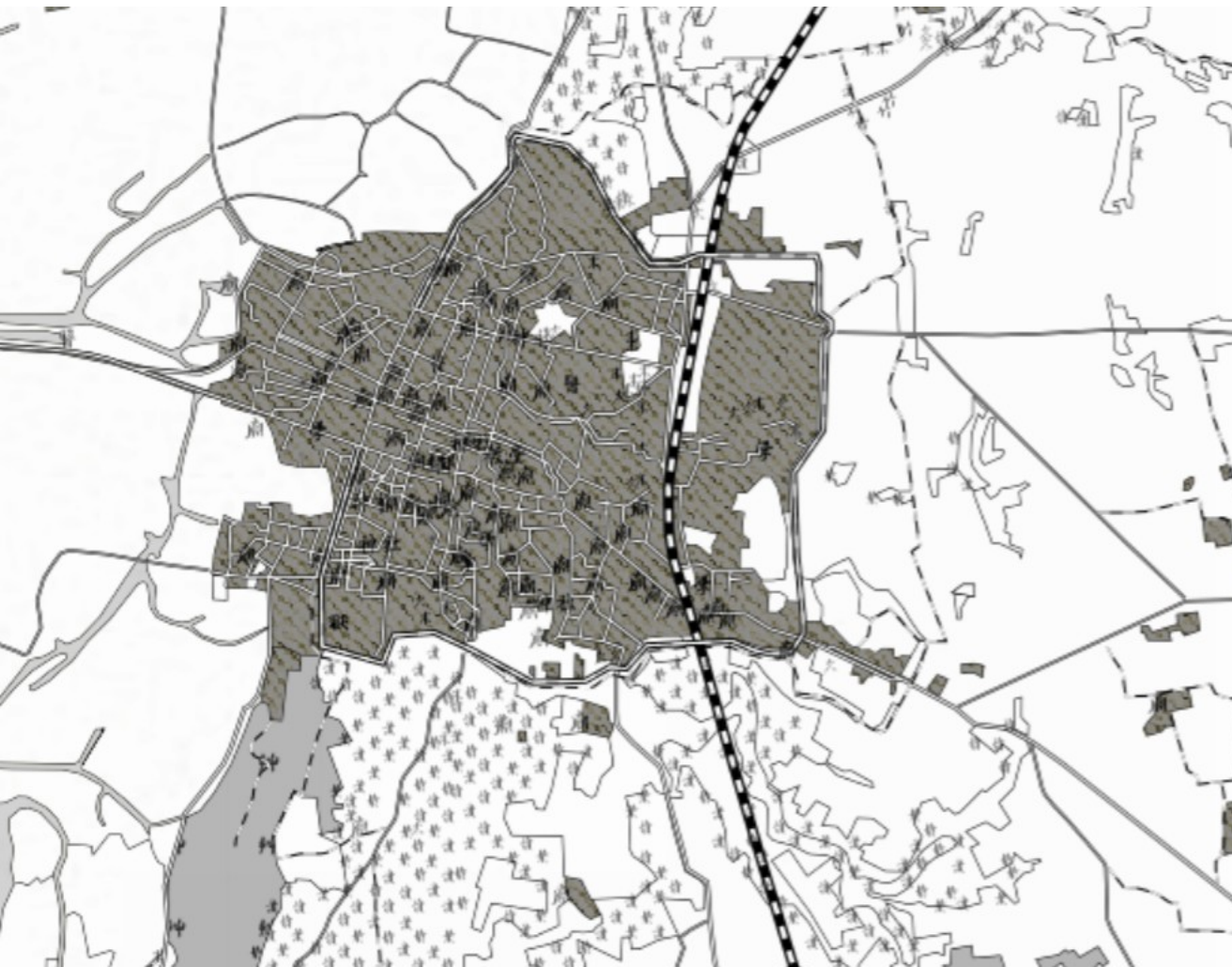
- Digitize map sheets and merge images (done by the Center for GIS, Academia Sinica)
- Extract and collect features from images (QGIS)
- Convert feature collections into OSM datasets
- Render datasets by self-hosted OSM tile servers
- Experiment with various map styles (Tilemill)

OSM Components (modifications in red)

https://wiki.openstreetmap.org/wiki/Component_overview



Re-styled Taiwan Baotu in OSM



Marker map

☒ word

☐ ink

☐ modern

☐ default

Base map

☒ ukiyo

☐ green

☐ texture

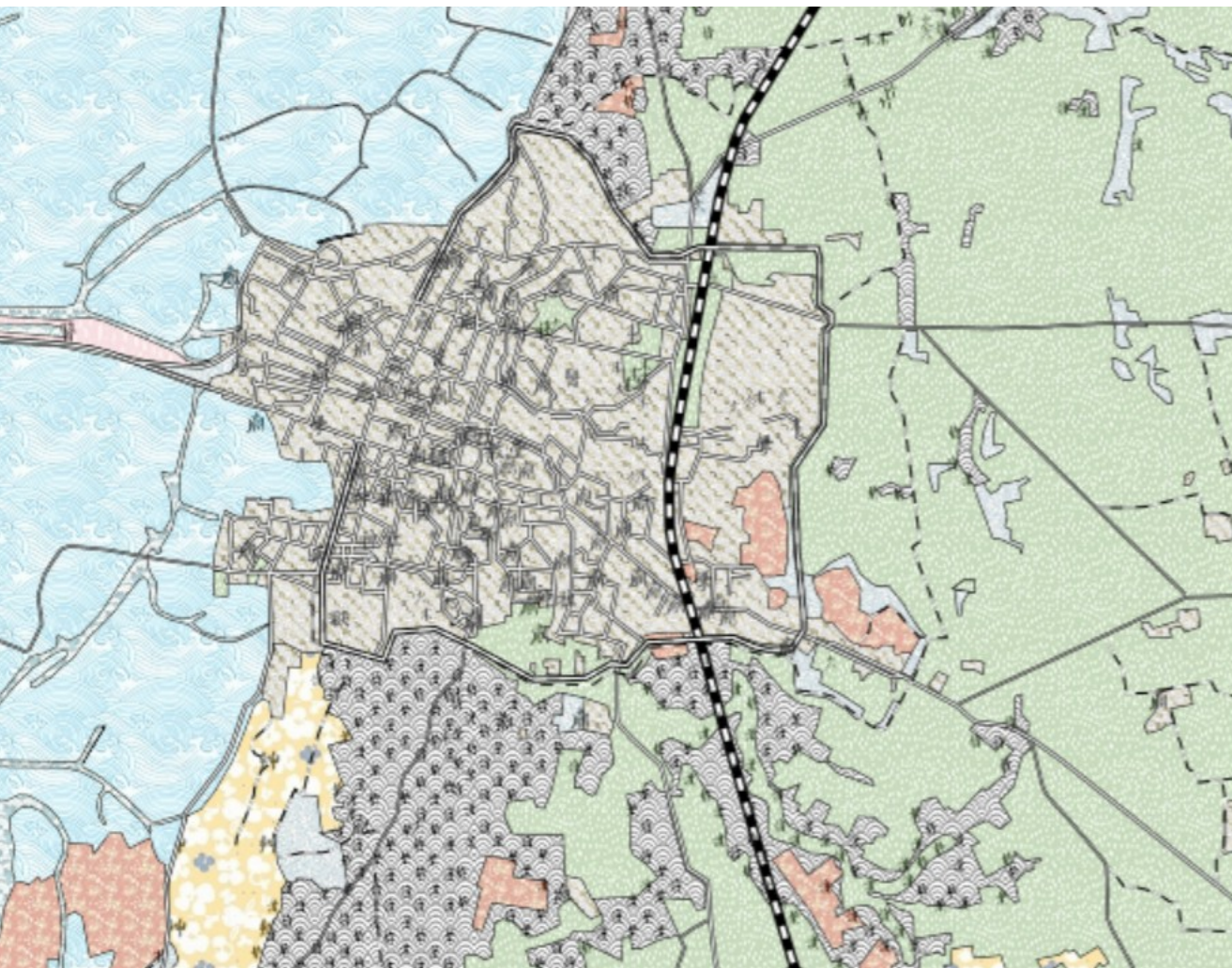
☐ Open Street Map

Transparency



<https://bcfuture.github.io/tileserver/Switch.html>

Re-styled Taiwan Baotu in OSM



Marker map

- ☒ word
- ☐ ink
- ☐ modern
- ☐ default

Base map

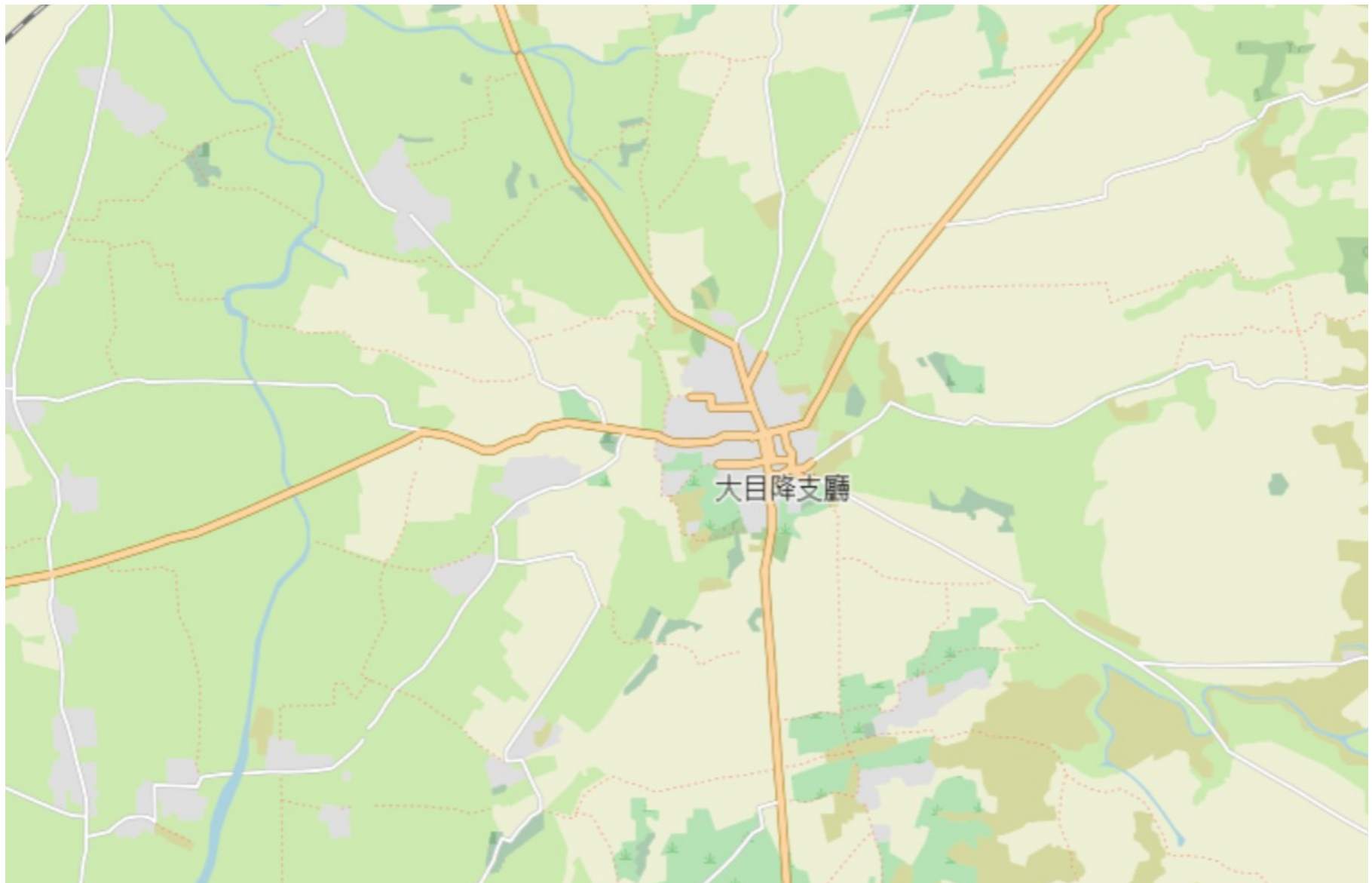
- ☒ ukiyo
- ☐ green
- ☐ texture
- ☐ Open Street Map

Transparency

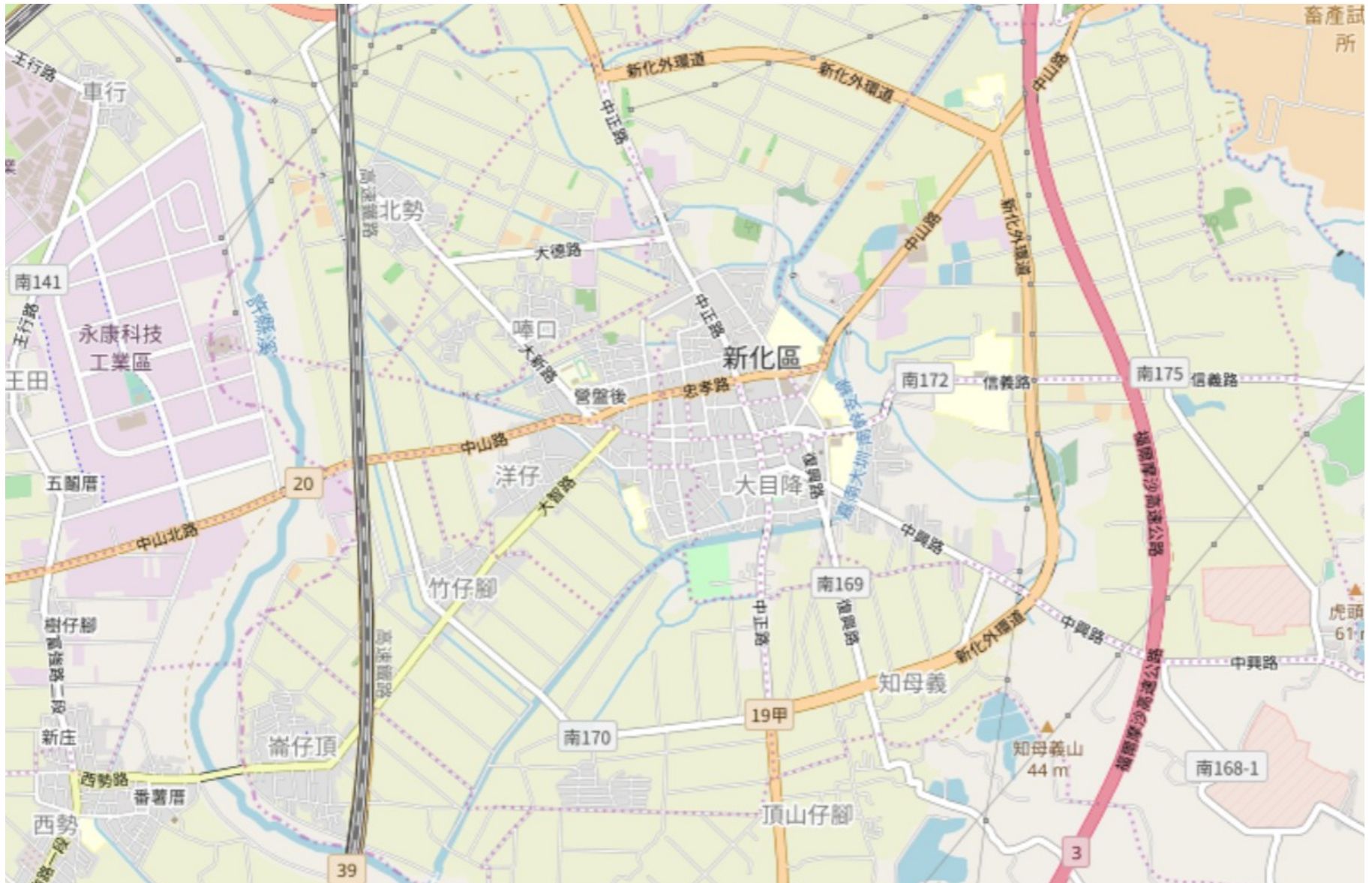


<https://bcfuture.github.io/tileserver/Switch.html>

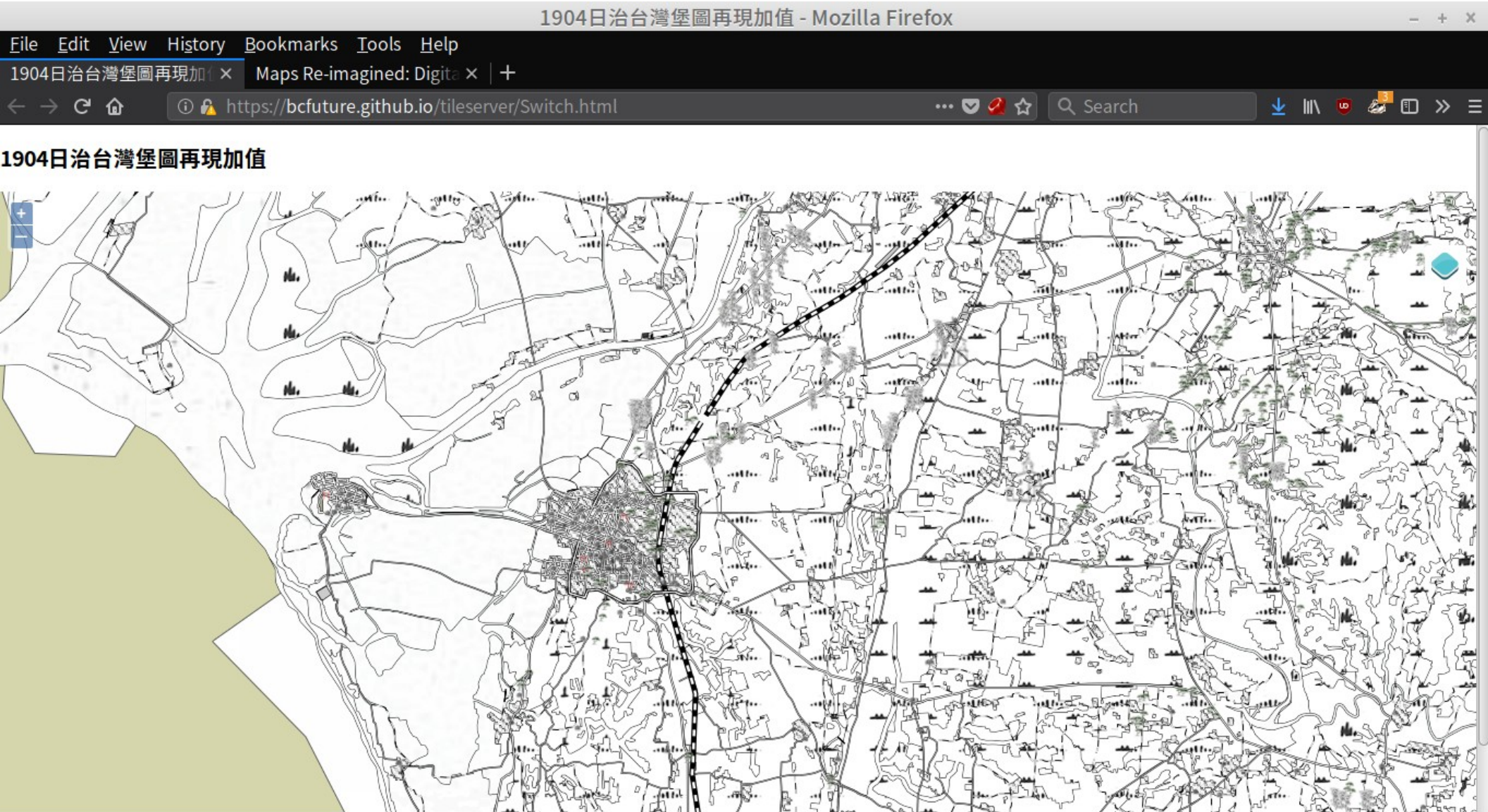
Sinhua in 1904, rendered in OSM



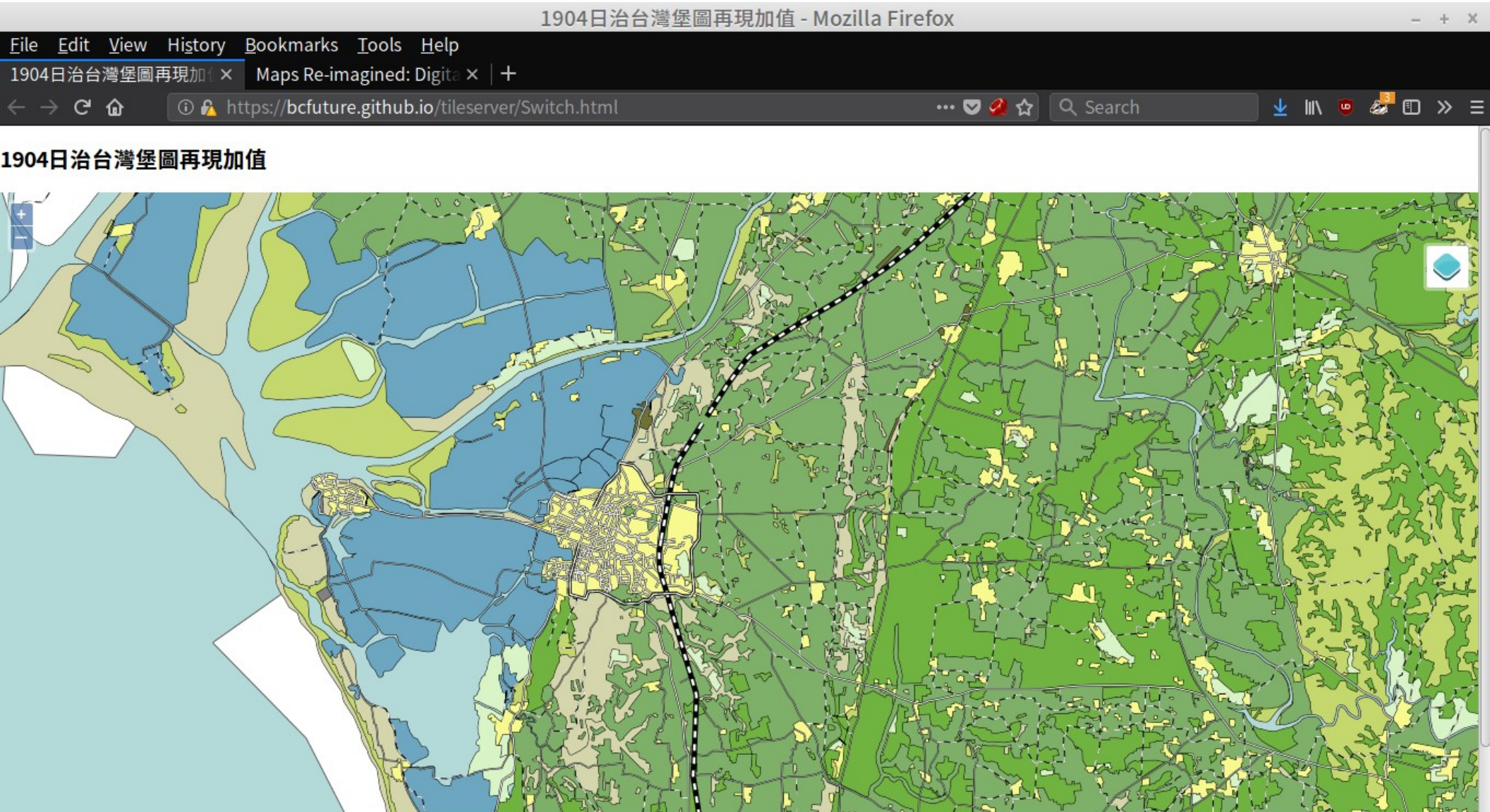
Current Sinhua in OSM



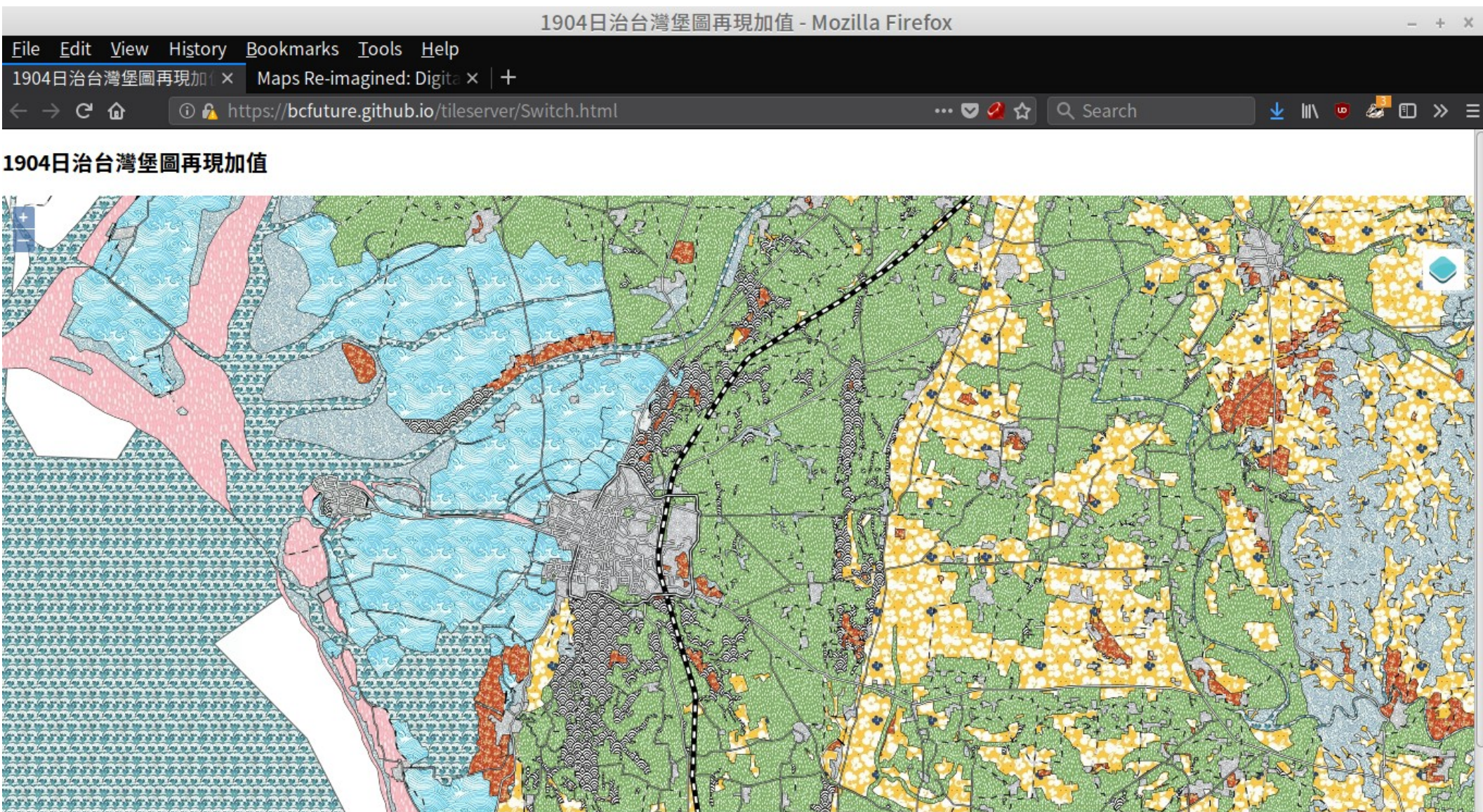
Tainan & Anping (Ink Marker)



Tainan & Anping (Green Basemap)



Tainan & Anping (Ukiyo Basemap)

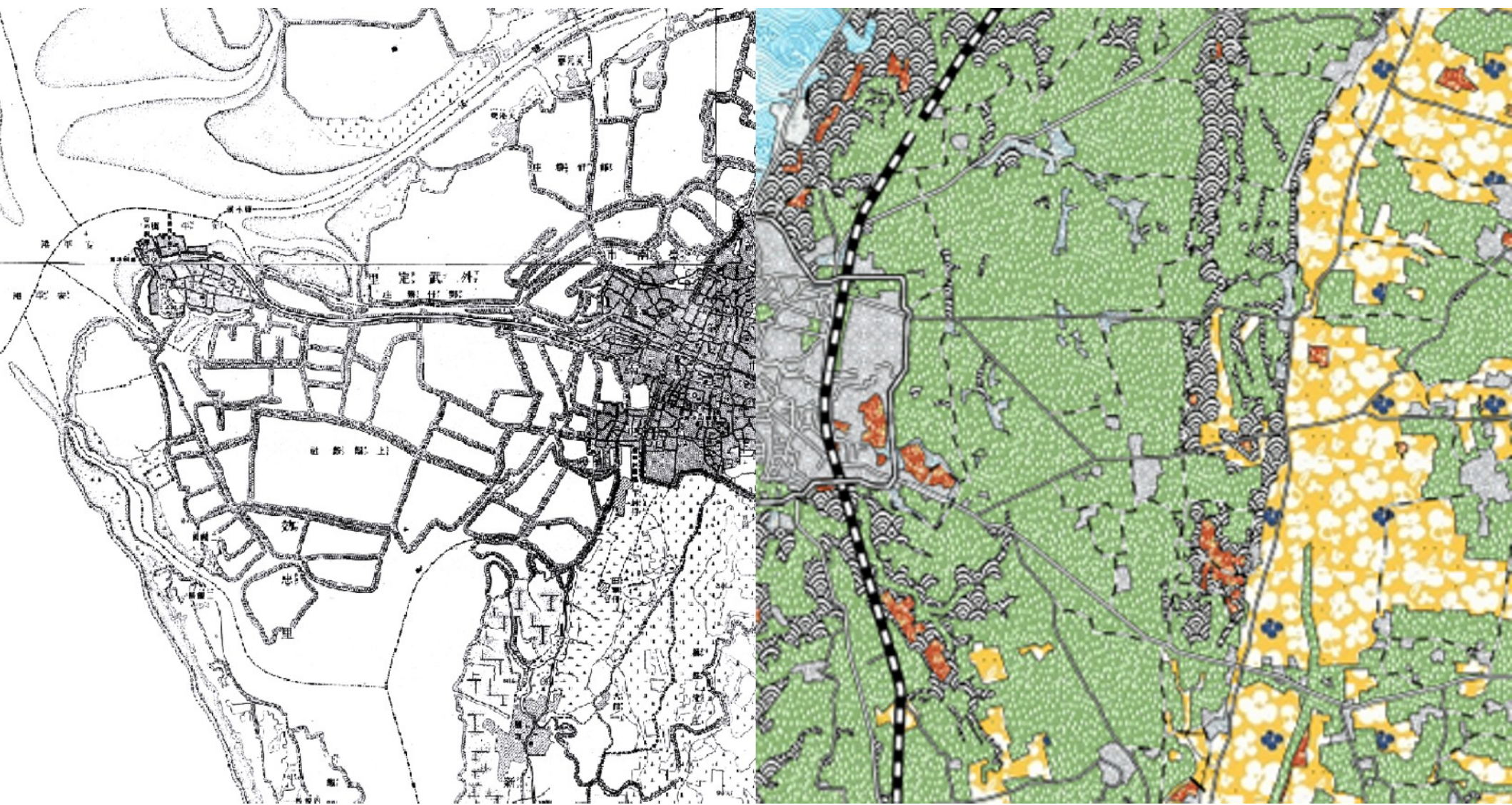


Data Workflow Reconsidered

- Task components per image
 - georectification
 - geometric feature tracing (points/lines/polygon)
 - feature type assignment (c.f. map legend)
 - text transcription, feature association, annotation
- Task components per collection
 - image alignment and merging
 - feature stitching and merging
- Task components related to collection management
 - metadata development; format conversion; making Web services
- Tasks performed in sequence or parallel; manually, automatically, and/or crowdsourced; issues of dataset revision and version control

Datasets Reuse and Remix

- Historical maps distilled as datasets for reuse
- Comparing and linking multiple datasets by
 - spatial relation
 - temporal variation
 - feature composition
 - text label matching
- Remix and reuse with external and communal resources
 - gazetteers, newspaper archives, OpenStreetMap, Wikidata
- Retooling the entire digitalization workflow for reuse and remix



Thank You!