

# Asian Directories and Chronicles: Foreign Residents Benchmark Dataset

---

Peter Cornwell, Institute for Modern and Contemporary Culture, University of Westminster, U.K.  
Madeleine Herren, Europa Institute, Basel, Switzerland.

---

## Dataset Description

This deposit comprises a benchmark dataset based on the listings of foreign residents in the 1896, 1899 and 1934, 1937 volumes of the Asian Directories & Chronicles serial, which was published annually by *The Hong Kong Daily Press* between 1863 and 1941. With the current exceptions of 1866, 1867, 1872, 1875 and 1884 all of the volumes of the Directories & Chronicles have been assembled by the Europa Institute at the University of Basel<sup>1</sup>. In a collaboration with Data Futures<sup>2</sup>, high-resolution digitization of the pages of the volumes and analysis of OCR data has enabled automated detection of each person record in the foreign resident listings and generation of 60,712 annotations. The latter are represented here as OADM, a precursor of WADM<sup>3</sup> which is currently in widespread use, although this dataset will be upgraded to WADM as soon as more applications emerge which support it. The OCR text has subsequently been corrected and tokenized with the aid of surname and location dictionaries created from the corpus, to produce searchable person 'instance' data using the schema at <https://schemata.hasdai.org/historic-persons/historic-person-entry-v0.0.2.json> and attached hereto.

The years selected for this benchmark serve the dual purposes of developing dynamic dictionaries for automating correction and tokenization of the remaining volumes of the serial, and they are also pivotal in relation to historic events in East Asia. The First Sino-Japanese War, waged from July 1894 to April 1895, was followed by establishment of large numbers of small communities of foreign residents throughout East Asia. In the early 20<sup>th</sup> century consolidation of foreign residents in larger communities in coastal cities was followed by a marked exodus during escalating conflict in the Second Sino-Japanese War between July 1937 and September 1945, which some sources date back to the Japanese invasion of Manchuria in 1931. These population shifts are visible when the benchmark dataset is rendered on to maps as shown in the following illustrations. The annotations and instance data are presented at <https://fr-benchmark.unibas.hasdai.org/> as an Invenio<sup>4</sup> repository enabling searching by person and location, as well as viewing person instances in context of the page of the serial where the person was listed—via Mirador<sup>5</sup> and a IIIF<sup>6</sup> service.

---

<sup>1</sup> <https://europa.unibas.ch/>

<sup>2</sup> <https://data-futures.org/>

<sup>3</sup> <https://www.w3.org/TR/annotation-model/>

<sup>4</sup> <https://invenio-software.org/>

<sup>5</sup> <https://projectmirador.org/>

<sup>6</sup> <https://iiif.io/>



