

# Data Science as a Foundation Toward Open Data and Open Science: The Case of Taiwan Indigenous Peoples open research Data (TIPD)

*Wednesday, 8 March 2017 14:30 (30 minutes)*

The research aims are threefold: to (1) demonstrate the methodology of data science in constructing Taiwan Indigenous Peoples open research Data (TIPD, see enter link description here <http://TIPD.sinica.edu.tw>, and enter link description here <https://osf.io/e4rvz/>) based on Taiwan Household Registration (THR) administrative data; (2) to illustrate automated or semi-automated data processing as a determinant of constructing effective open data; and (3) to demonstrate appropriate utilization of “old-school” data format such as multi-dimensional tables as an effect means to overcome legal and ethic issues.

The research extracts valuable information embedded in micro data of THR and enriches the extracted information through the processes of cleaning, cleansing, crunching, reorganizing, reshaping the source data. Major outputs of TIPD amount to 7,300 files in number and around 32 GB in size. TIPD now consist of three categories of open research data: (1) categorical data, (2) household structure and characteristics data, and (3) population dynamics data. Categorical data include two broad dimensions. The data enrichment processes produce a number of data sets that contain no individual information but retain most of source data information. The enriched data sets thus can be open to the public for open administrative data study. The open data are systematically constructed in an automated or semi-automated way through the integration of compiler programming language, software, and script languages.

Keywords: data science, administrative data, open data, open science, TIPD

**Primary author:** Dr LIN, Ji-Ping (Academia Sinica)

**Presenter:** Dr LIN, Ji-Ping (Academia Sinica)

**Session Classification:** Humanities, Arts & Social Sciences I

**Track Classification:** Humanities, Arts, and Social Sciences (HASS) Applications