ISGC 2026

How Open Science, Data Science, and Scientific Computing Revitalize and Make Hard-to-reach Population Visible: A Two-decade Research on Taiwan Indigenous Peoples

Ji-Ping Ln(林季平)
Academia Sinica
Nankang, Taipei
TAIWAN 115
Email: jplin@sinica.edu.tw
ORCID: https://orcid.org/0000-0003-2348-9243

Acknowledgements: The research acknowledges financial & administrative supports from Council of Indigenous Peoples, National Science & Technology Council, and Academia Sinica.

1. Introduction

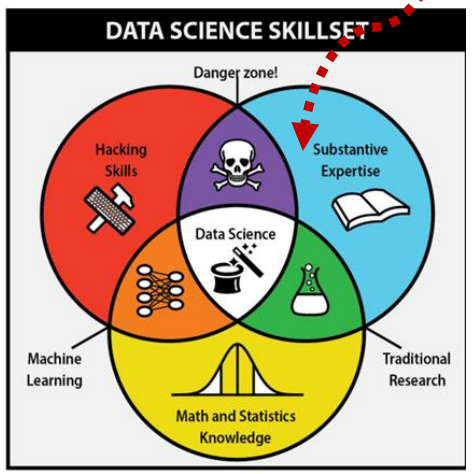
Main theme of my presentation

(1)Background stories of building open data on TIPs from 2000 to 2012

(2)Processes of building open data: 2013-now

(3)Focus on substantive expertise

1. Data Science



Source: O'Neil and Schutt 2013

2. Interview of Center for Open Science

Researcher Q&A: A Conversation About Research on Taiwan Indigenous Peoples and Making a Hard-to-Reach Population Visible

September 4th, 2024, Center for Open Science

Posted in: Research, OSF, Open Science

OSF

RESEARCHER Q&A:

A Conversation About Research on Taiwan Indigenous Peoples and Making a Hard-to-Reach Population Visible

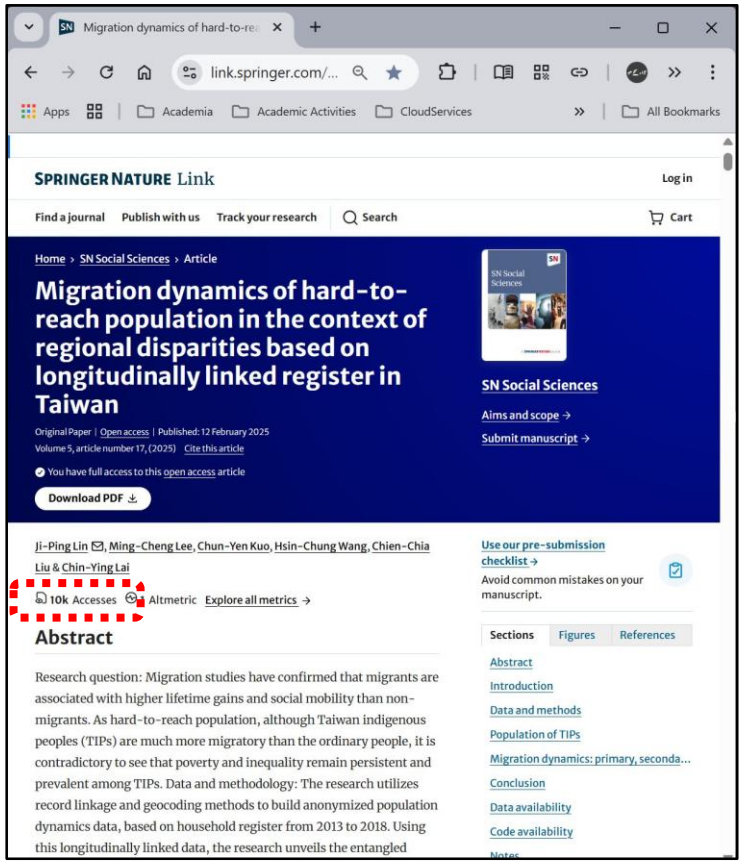
*With **Ji-Ping Lin, PhD**, Researcher at Academia Sinica in Taiwan*

Ji-Ping Lin, PhD, is a researcher at Academia Sinica in Taiwan who created an open access dataset to facilitate research and understanding of Taiwan Indigenous Peoples. In this Q&A, Lin answers questions about the project, including his use of the Open Science Framework (OSF) and the wide-ranging impact of his research.

Q. What inspired you to pursue this research project?

A. Taiwan Indigenous Peoples (TIPs) are a branch of Polynesian-Malaysian (or Austronesian) ethnic groups in a genetic and linguistic context. There was a rich body of ethnographic, official, and academic records on TIPs before 1940. However, the period of 1940–2000 was a data "Dark Ages" for TIPs due to political reasons. This lack of data marginalized them, making it seem lik

3. A publication by Springer Nature Social Science



SPRINGER NATURE Link

Log in

Find a journal | Publish with us | Track your research | Search | Cart

Home › SN Social Sciences › Article

Migration dynamics of hard-to-reach population in the context of regional disparities based on longitudinally linked register in Taiwan

Original Paper | Open access | Published: 12 February 2025

Volume 5, article number 17, (2025) Cite this article

You have full access to this open access article

Download PDF

SN Social Sciences

Aims and scope →

Submit manuscript →

Ji-Ping Lin, Ming-Cheng Lee, Chun-Yen Kuo, Hsin-Chung Wang, Chien-Chia Liu & Chin-Ying Lai

10k Accesses | Altmetric | Explore all metrics →

Use our pre-submission checklist →

Avoid common mistakes on your manuscript.

Abstract

Research question: Migration studies have confirmed that migrants are associated with higher lifetime gains and social mobility than non-migrants. As hard-to-reach population, although Taiwan indigenous peoples (TIPs) are much more migratory than the ordinary people, it is contradictory to see that poverty and inequality remain persistent and prevalent among TIPs. Data and methodology: The research utilizes record linkage and geocoding methods to build anonymized population dynamics data, based on household register from 2013 to 2018. Using this longitudinally linked data, the research unveils the entangled

Sections | Figures | References

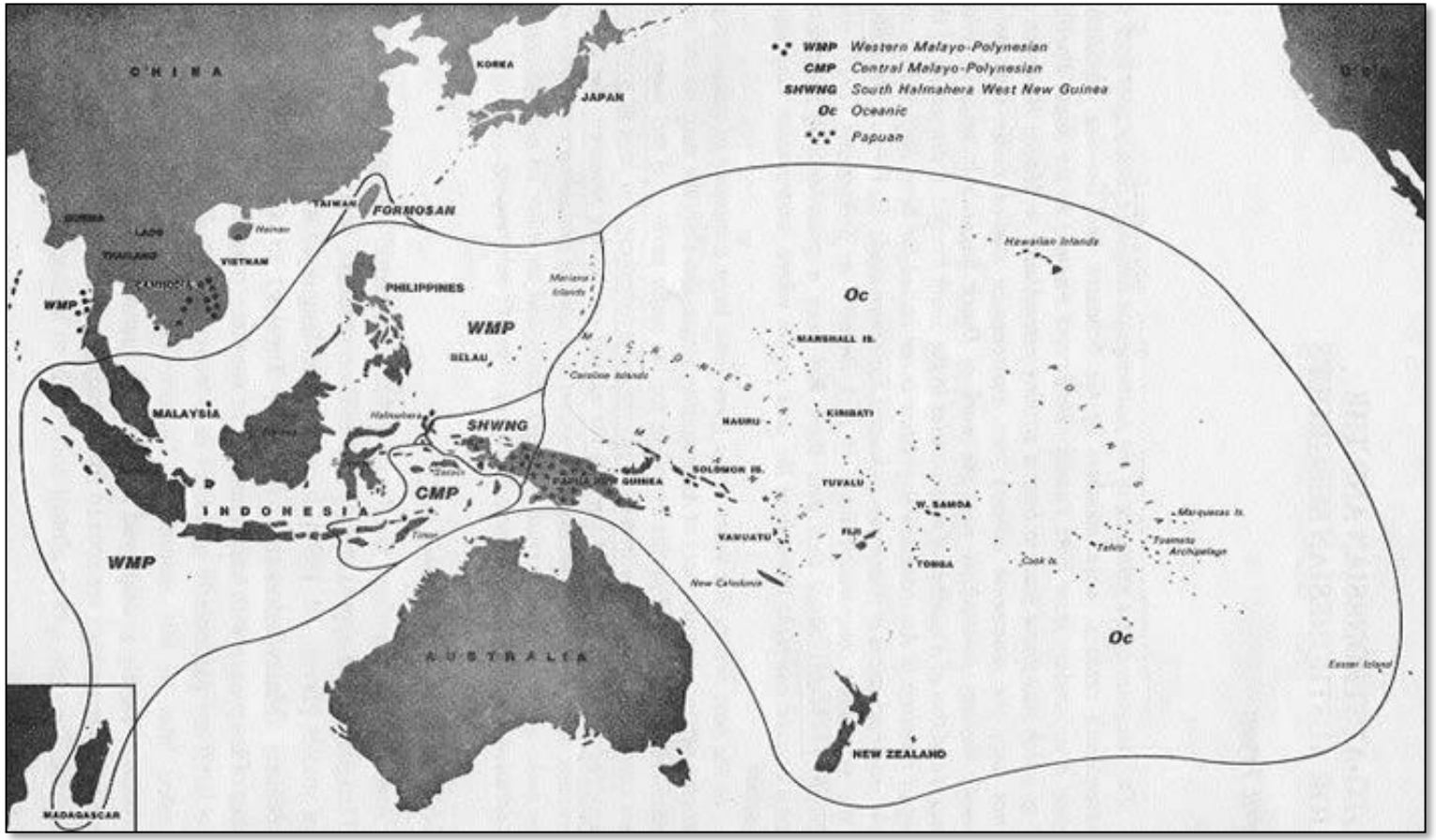
Abstract
Introduction
Data and methods
Population of TIPs
Migration dynamics: primary, seconda...
Conclusion
Data availability
Code availability
Notes

1. Introduction

- **Taiwan Indigenous peoples (TIPs):** A branch of Polynesian-Malaysian (or Austronesian) ethnic groups in genetic and linguistic context.

Contemporary TIPs

Austronesian distribution

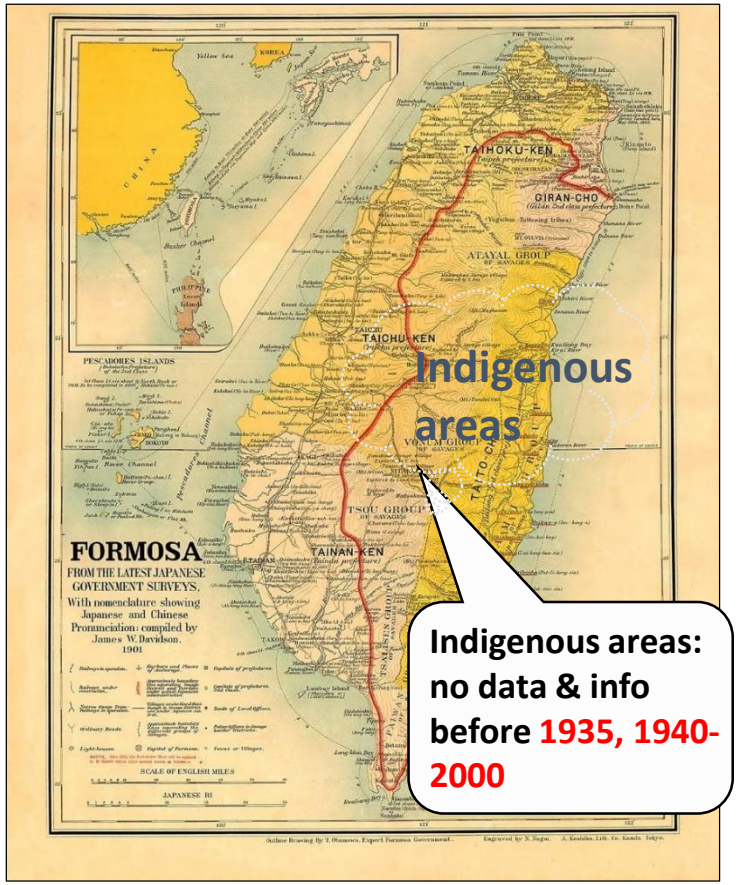


Source: http://www.taiwandna.com/AborigineAustronesia.jpg

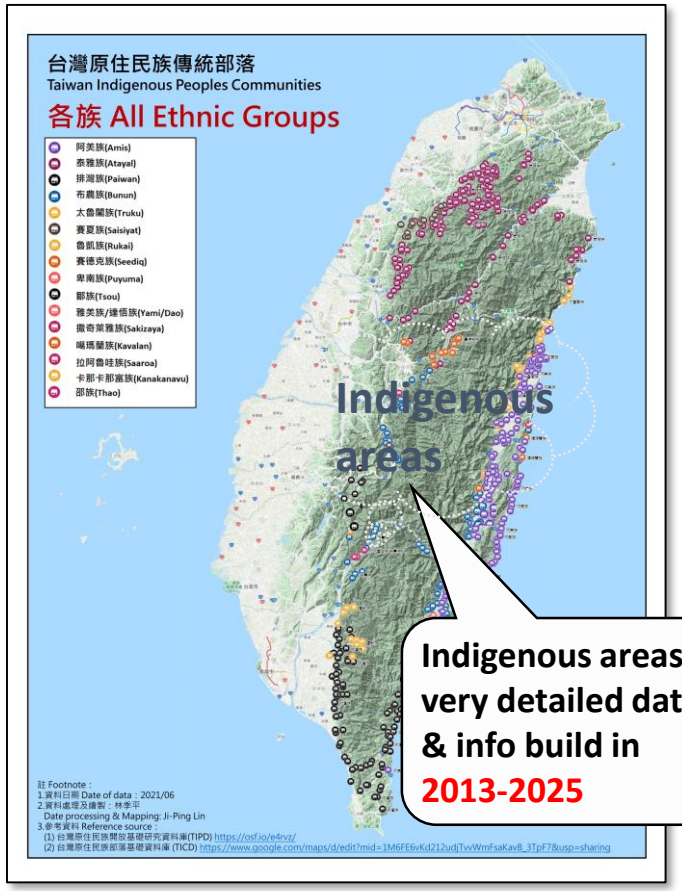
1. Introduction

- Impacts of past systemic barriers and ethnic isolation policy
- Why Taiwan indigenous peoples (TIPs) became a missing link of population in Taiwan
- How TIPs become from the least to the most known pop'n in 2000-2025

1901 Formosa "Han-Fan" boundary



2022 TIPs traditional communities



Taiwan landform

1. Introduction

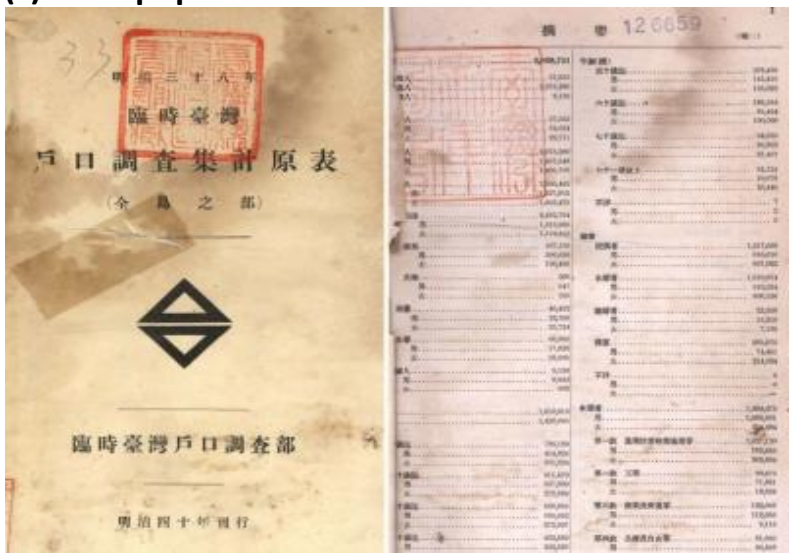
- **1895-1940 as the onset of collecting TIPs data**

1. **Pioneering studies** of Japanese scholars in indigenous areas
2. **Official data collected** by Japanese government of Formosa
   1) Pop'n censuses: 1905, 1915, 1920, 1925, 1930, 1935, 1940
   2) Household registers
   3) Official surveys

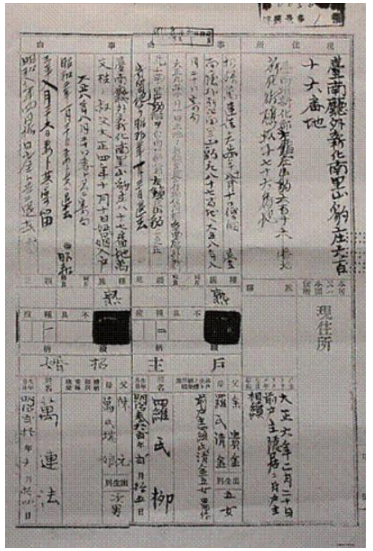
(a) Known Japanese scholars

(b) TIPs in early 20th century

(c) 1905 pop'n census



(d) Pop'n register

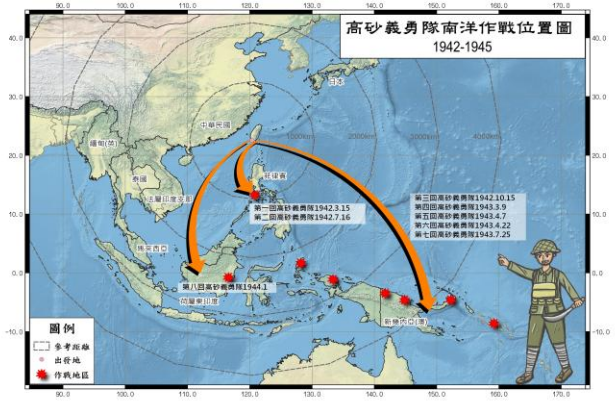


1. Introduction

- **1940-2000 as data "Dark Ages" for TIPs**
  - ✓ 1941-45: Pacific War & TIPs data collection stopped
  - ✓ 1946-1990: ethnic isolation policy of new regime from China
- Persistent lack of TIPs data led TIPs to become **isolated** and **marginalized** and thus **underdeveloped**

e.g.,

- ✓ prevalent ethnic discrimination
- ✓ much higher unemployment
- ✓ income level far below average
- ✓ life expectancy less than the average for about 11 years
- ✓ less educational & medical availability & accessibility
- ✓ very low social mobility & social status



(a) Pacific war and TIPs army of IJA

(b) Ethnic isolation period after WWII

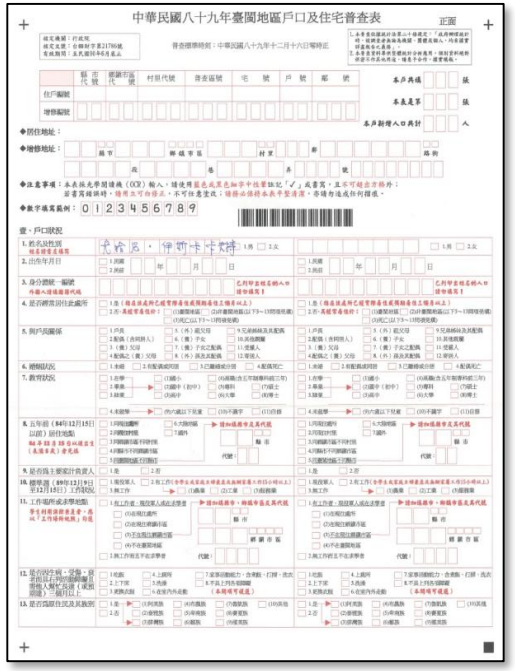
2. Data Renaissance and Revolution

■ 2000-2010 as "Data Renaissance" of TIPs

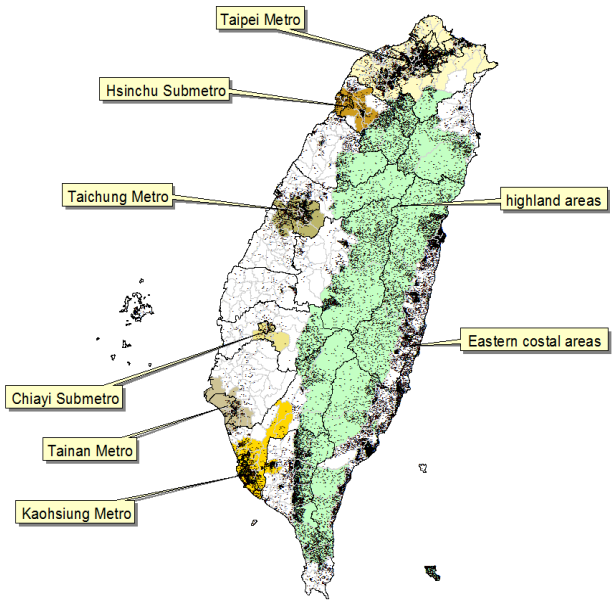
- ✓ Political democratization and formation toward pluralistic system in the 1990s, e.g., two-decade Indigenous Movement
- ✓ After six decades, Taiwan government resumes TIPs census in 2000 Population & Household Census

(a) Taiwan indigenous movement in 1980s-1990s

(b) 2000 Taiwan Pop'n Census



(c) First individual-level pop'n map of TIPs (Lin 2002)



2. Data Renaissance and Revolution

- **2011-2025 as "Data Revolution" of TIPs**

✓ Household register becomes the new source data of TIPs

1. 2010 Pop'n Census was cancelled after 2010
2. Administrative data: Taiwan Household Registration Data (THRD).
3. THRD data sets are archived for the study on a monthly base.
4. Information in THRD: Household ID, Time of data creation, PIN, name, spouse name, parents' names, education, age, marital status, address, birth place, mobility…

- **Overcoming issues of accessing source data**

1. My thought: *be brave if you want to do something big!*
2. As a PI, I proposed to sign a legal document indicating that "*The PI takes the **SOLE** responsibility in case of having legal issues & law suits*".

- Building new TIPs data based on data science, open science, and scientific computing

2. Data Renaissance and Revolution

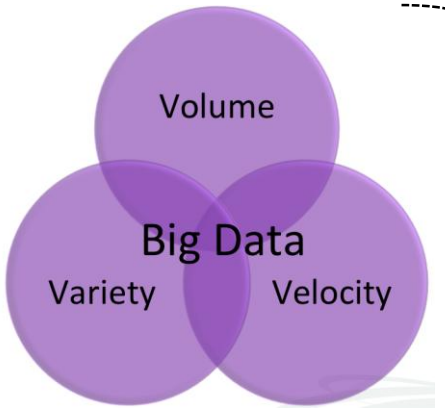
■ Foundation 1: Open Science and Open Data

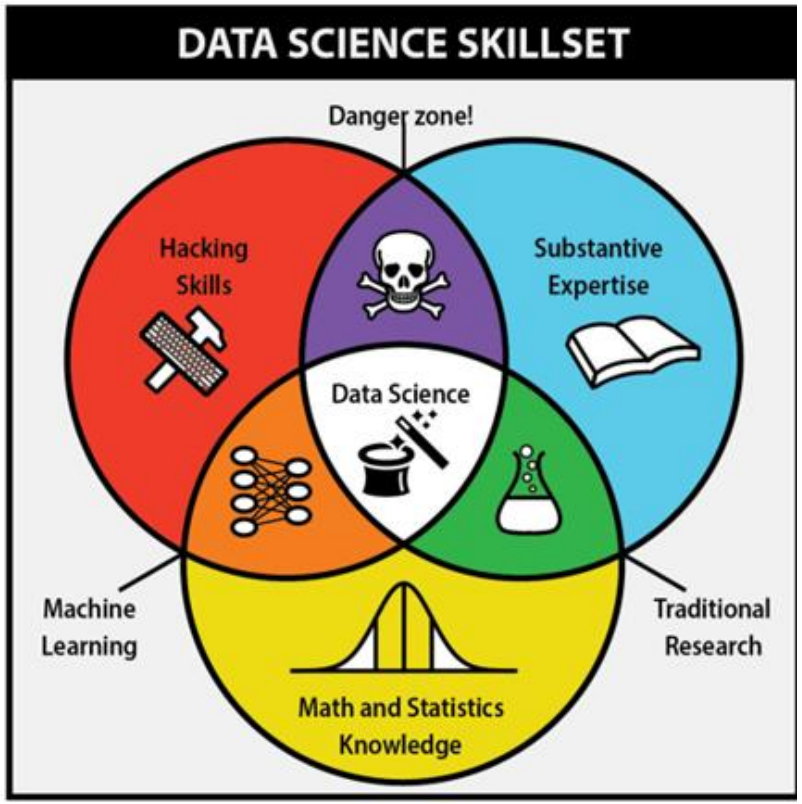
1. **Open science** must comply with six principles: open data, open source, open methods, open peer review, open access, and open educational research.
2. **Open data** refers to the data that can be available without restrictions of any forms.
3. **Open source** serves as an effective ways of building collective wisdom & unleashing creativity.
4. Open science and open data can serve as a foundation of overcoming replicability and reproducibility issues in research.

2. Data Renaissance and Revolution

■ Foundation 2: Data Science

- ✓ Data science refers to the extraction of knowledge from data, with the main goals being to extract meaning from data and to produce data products.
- ✓ It employs techniques and theories drawn from many fields within the broad areas of mathematics, statistics, and information technology.





Source: O'Neil and Schutt 2013

2. Data Renaissance and Revolution

■ Foundation 3: Scientific Computing

➢ Hacking skills: Big advances of digital Infrastructure in 2011~12: CPUs, DRAM, chipsets, & x64 OSs & applications…

Fig. Digital infrastructure before 2012

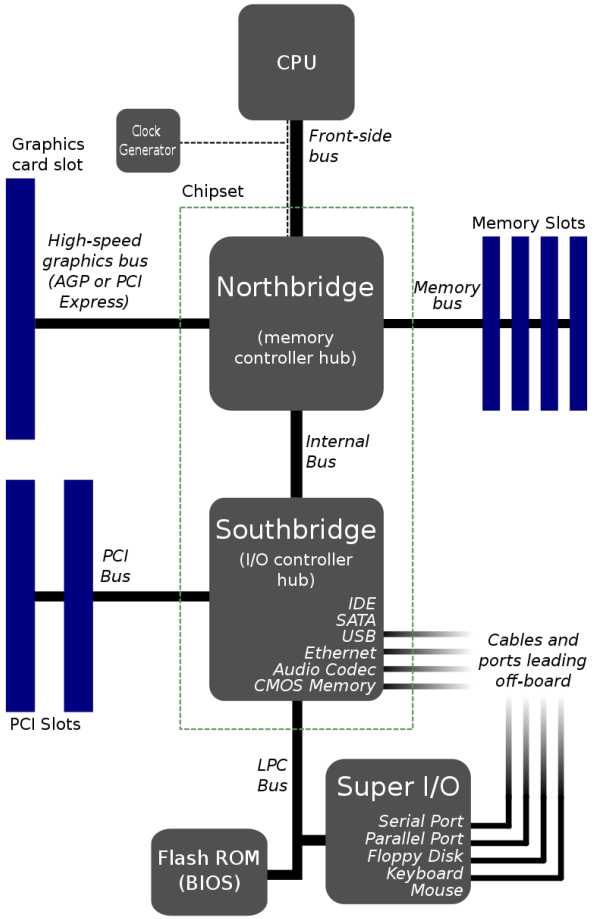
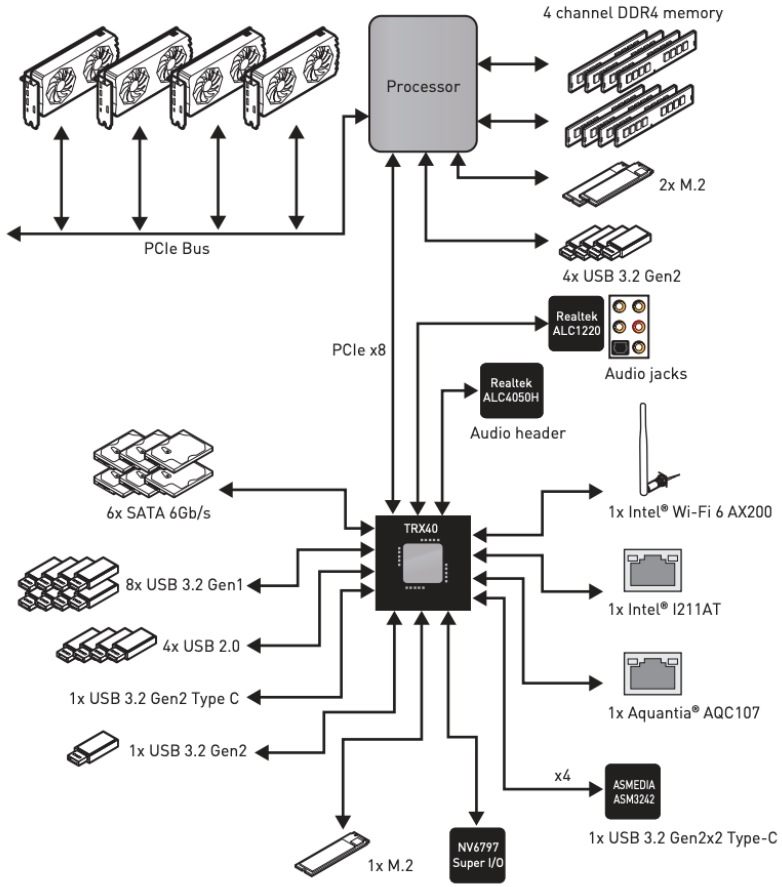


Fig. Digital infrastructure after 2012

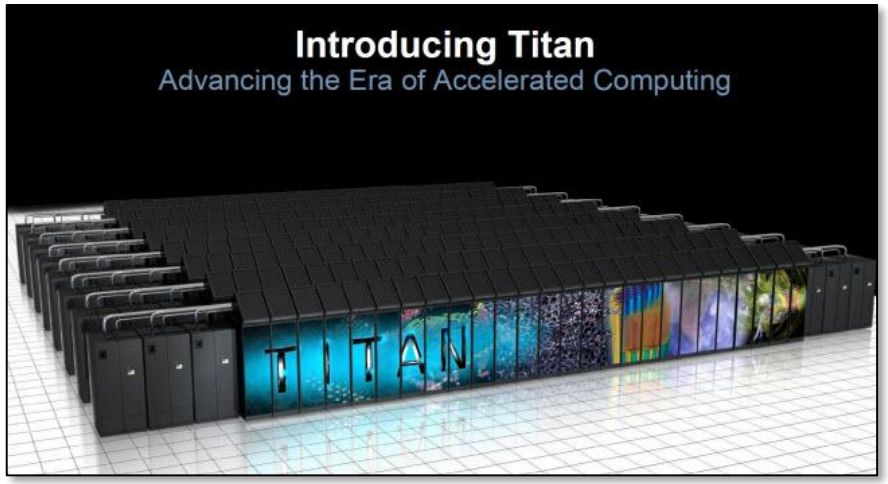


2. Data Renaissance and Revolution

■ Foundation 3: Scientific Computing

➢ Hacking skills: Emergence of advanced super computers after 2012, e.g., Oakridge Titan, Fujisu Kyo, Wuhan Tienho, Oakridge Frontier…

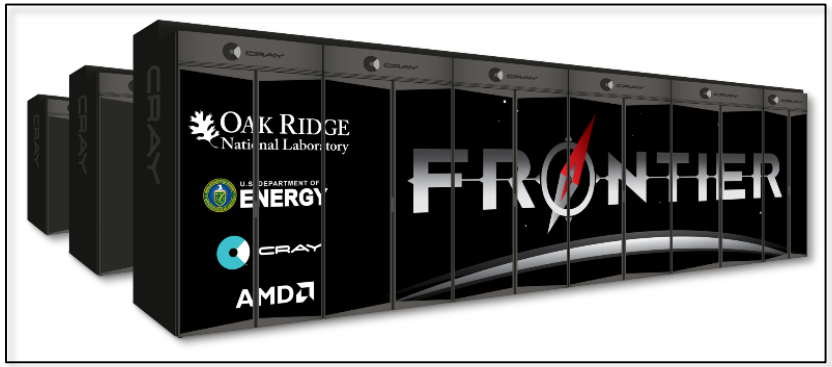
Oakridge Titan



Fujisu Kyo

Wuhan Tehho

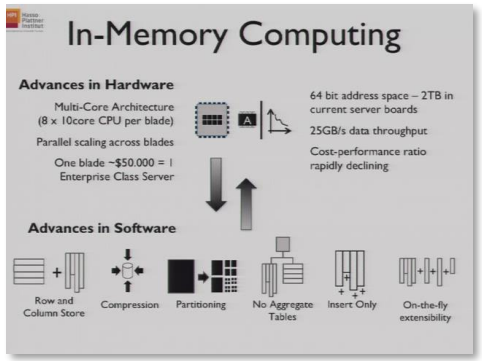
Oakridge Frontier

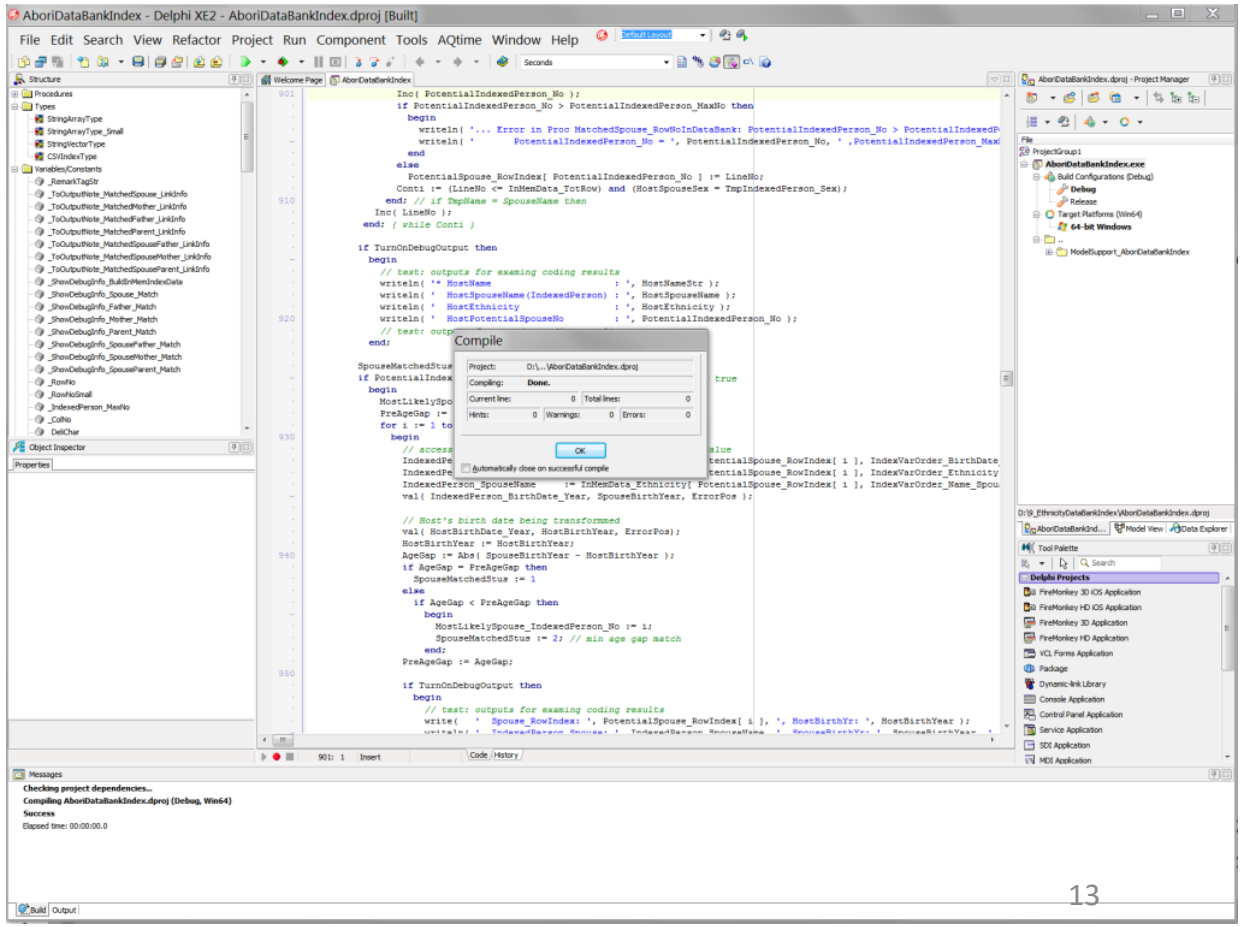


2. Data Renaissance and Revolution

■ Foundation 3: Scientific Computing

- ✓ Manipulation of High Performance Computing (HPC)
- ✓ Programming tools: Win11 x64 + x.64 programming language object Pascal & coding in RAD Studio Delphi
- ✓ In-memory computing as Fundamentals of HPC
  1) Record-linkage computing techniques: matching involves thousands of billion matching in accumulated data bank;
  2) To accelerate computing, we use: In-memory computing to achieve genealogy computing by overclocking digital hardware (1) CPUs & (2) IO bus & (3) DRAM.

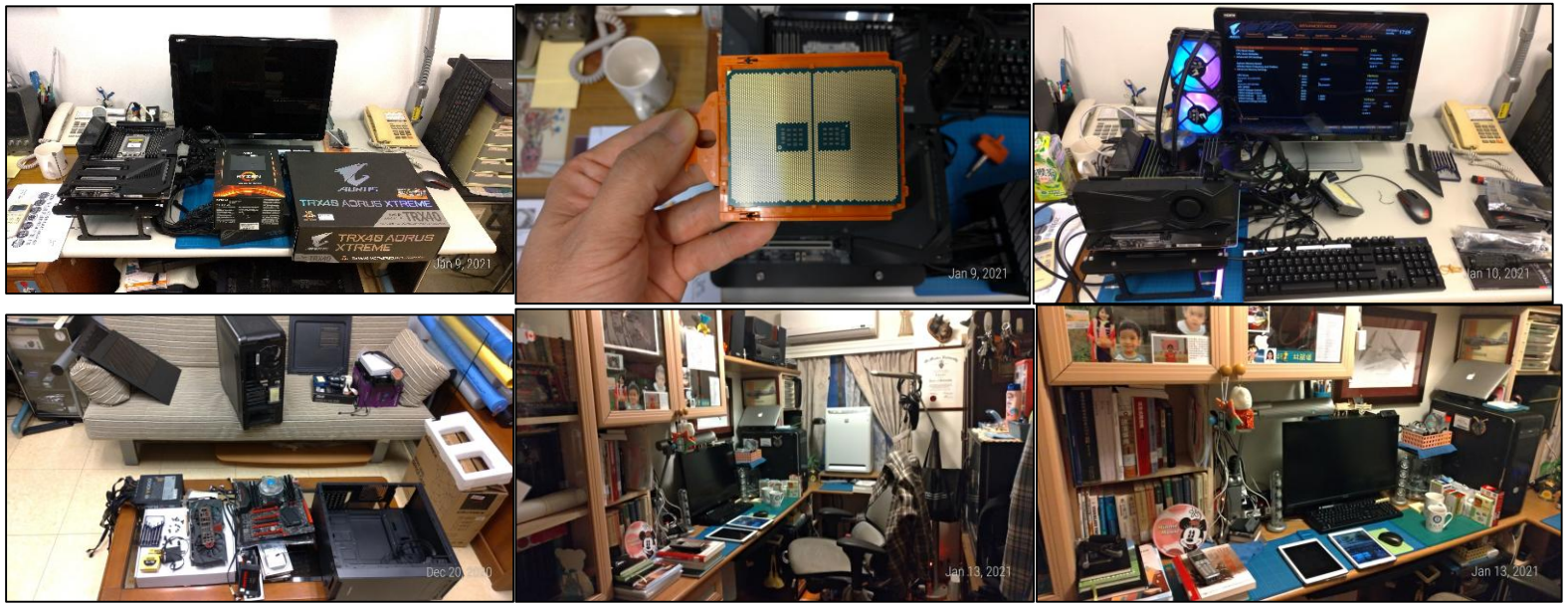


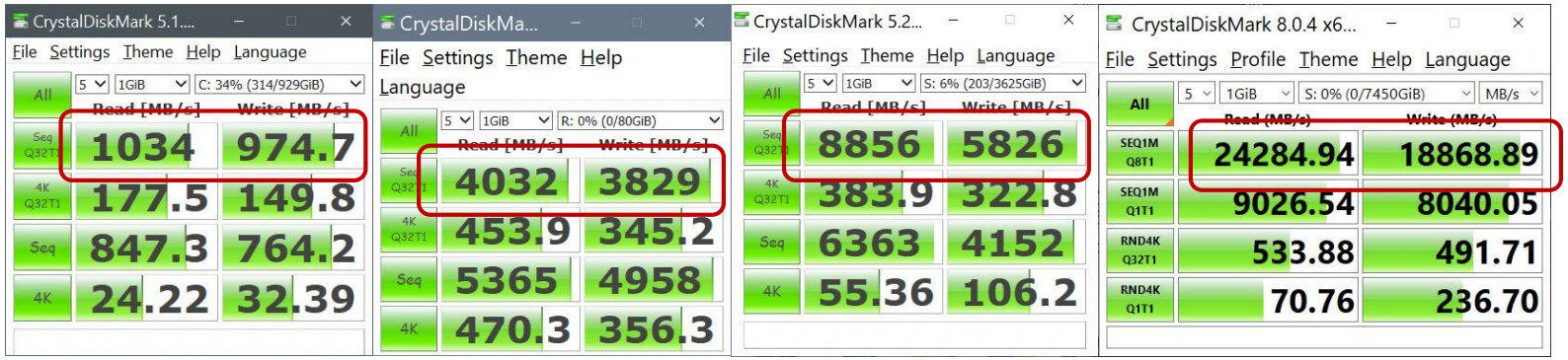


2. Data Renaissance and Revolution

■ Foundation 3: Scientific Computing

✓ Data Engineering: Assembling & fine-tuning computing systems



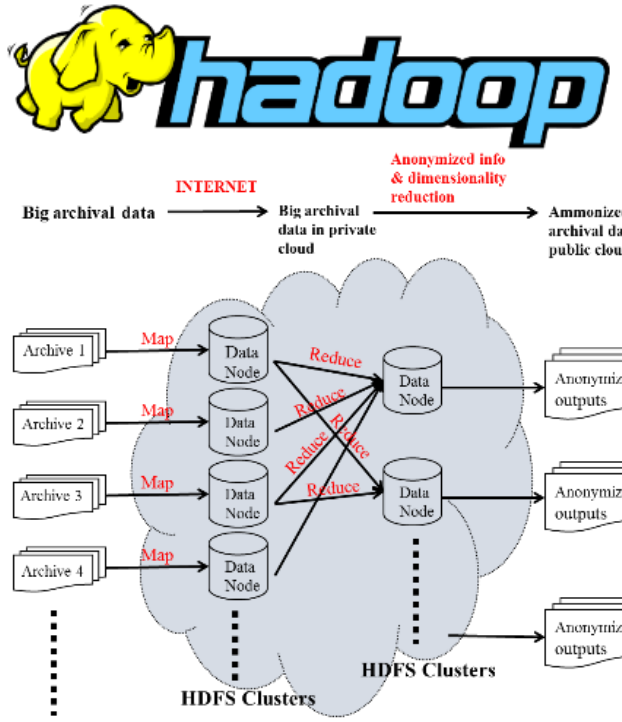


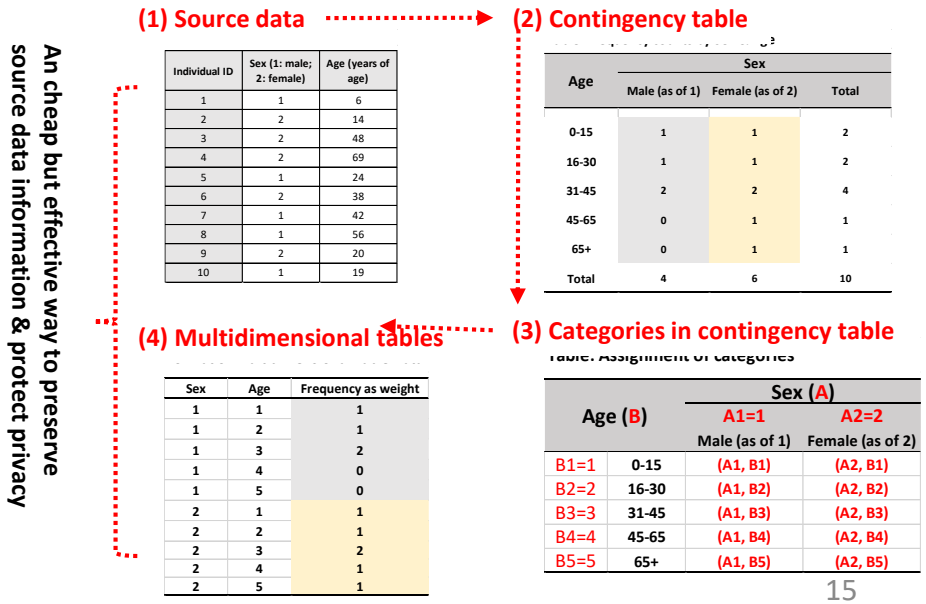
2013 → 2016 → 2019 → 2023

3. Building, Enriching, and Integrating Data

■ Foundation of Overcoming Legal & Ethical Issues

1. Giving up conventional in-house lab method, and adopting distributed computing infrastructure
2. The first tool that was considered to use at the beginning of research: Apache Hadoop (open sourced version of Google GDFD+MapReduce)
3. Building conventional “old-school” multi-dimensional tables is adopted as means for “distributed data storage” and “centralized data integration”





3. Building, Enriching, and Integrating Data

■ Why I Decided to Build & Share Open Data

1. To **overcome privacy, legal, and ethic issues**:

   ✓ the research won't succeed unless we can devise a computational method that enables us to transform raw individual data to a new form of data which still preserves most information of raw data (e.g. say 95%), but the transformed data won't allow identification of any personal information.

2. To enhance **research efficiency and flexibility** for team members:

   ✓ open data enables us to promote interpersonal collaboration and research efficiency.

3. Building, Enriching, and Integrating Data

■ Open Data & Importance

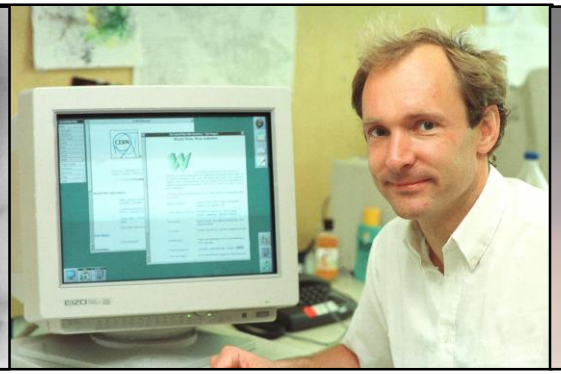
1. **Open data** refer to the data that can be available to everyone to use and republish without restrictions of any forms, including copyright, patents, and/or other mechanism controls.
2. Academic community has acknowledged that the practices of open data not only facilitate collaboration and drive data analysis, but also help promote transparency and reproducibility.
3. Open science and Open Data can serve as a foundation of overcoming replicability issues in research.
4. Open sources as an effective ways of collective wisdom & improvement;
5. The main goal of open sources is free that serves as the role of unleashing creativity;

e.g. **CopyLeft (L)** WWW, UNIX, Linux etc…

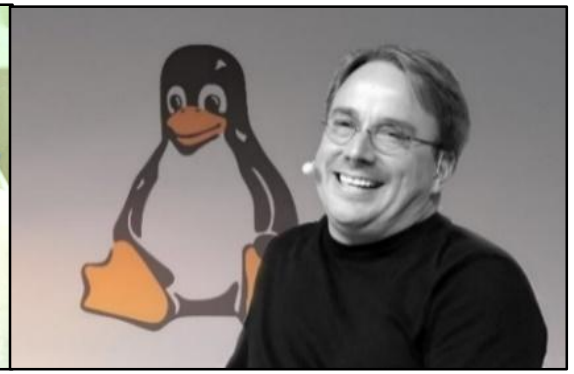
Bell Lab's Unix

IBM PC

CERN's Tim Bernes-Lee



Linus & Linux



3. Building, Enriching, and Integrating Data

■ Procedure of building open data

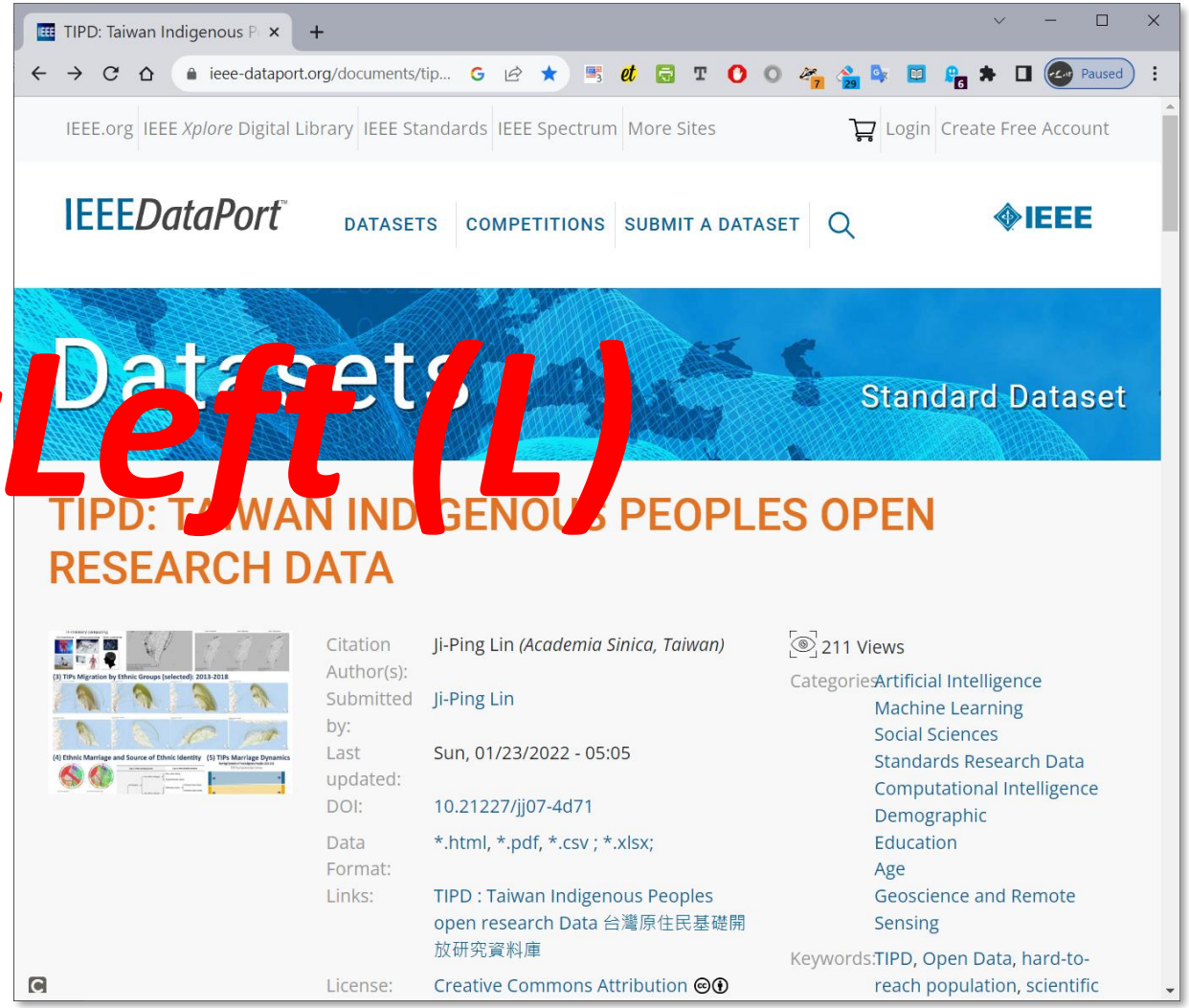
1. Collect raw data at the end of each month
2. Process raw data and build open data in a closed data lab
3. Synchronize newly built TIPD with primary data repository
4. Synchronize newly built TIPD with secondary data repository
5. Synchronize first and secondary open data repositories with open science framework's platform
6. Disseminate new open data to participants of our joint research program
7. Renew contents on open science platform twice a year
8. Announce new release of open data and research materials to the public for open download
9. Repeat the above procedures

3. Building, Enriching, and Integrating Data

■ Open data: TIPD (Taiwan Indigenous Peoples Open Research Data)

TIPD on OSF

TIPD on IEEE DataPort



3. Building, Enriching, and Integrating Data

- **Principles of building open data: being friendly, easy access, ease of use for "ordinary people".**

- **Available data formats in TIPD**

  PDF, HTML, RTF, XLS, CSV, Excel, dBase, Access, Matlab, Gauss, HTML, JMP, SAS, SPSS, Stata, Access & R data formats.

- **Types of data in TIPD**
  - ✓ Cross-sectional data
  - ✓ Longitudinally linked data
  - ✓ Household structure and characteristics data
  - ✓ Population dynamics data
  - ✓ Individual-level genealogy data
  - ✓ Traditional community data
  - ✓ Ethnic genealogy data
  - ✓ Individual-level kinship data
  - ✓ TIPS in urban communities/cluster data
  - ✓ Household & individual geocoding data

3. Building, Enriching, and Integrating Data

■ Building spatial information: locating TIPs

1. Geo-coding: Locate each individual TIPs based on household full address;
2. Methods & processes of obtaining & integrating geolocation data:
   1) develop a new command-line program in Object Pascal that can parse geolocation info in Google Map;
   2) Let cleansed address info be geolocated by Google Map & use the program to parse geolocation info from Google Map;
   3) Using redirection operators and pipeline to integrate geolocation obtained from Google Map;
   4) Integrate geolocation data with TIPD open data.

3. Building, Enriching, and Integrating Data

- **Building spatial information**: develop highly precise automated geocoding to locate TIPs

Figure – 全部原住民(All Indigenous Peoples)

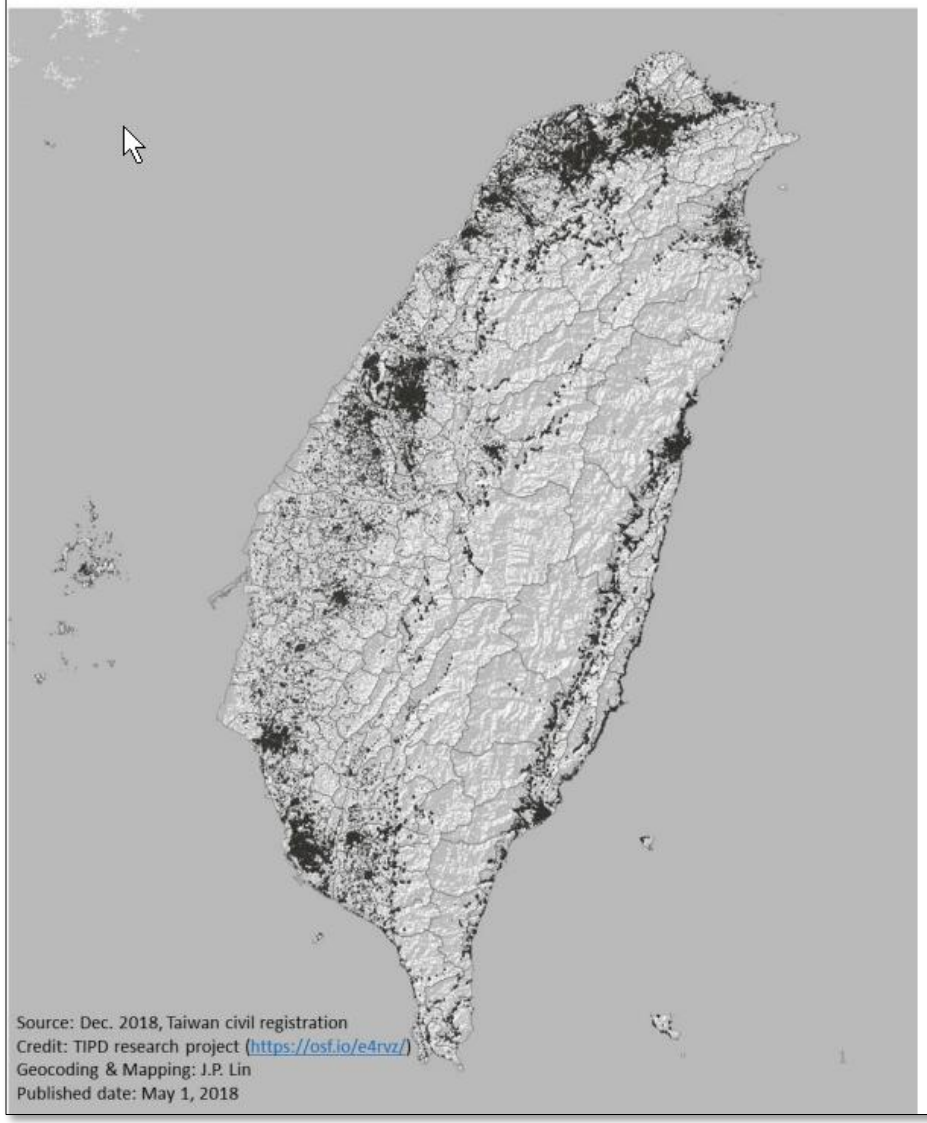


Figure - 阿美(Amis)

Figure - 泰雅(Atayal)

Figure - 排灣(Paiwan)

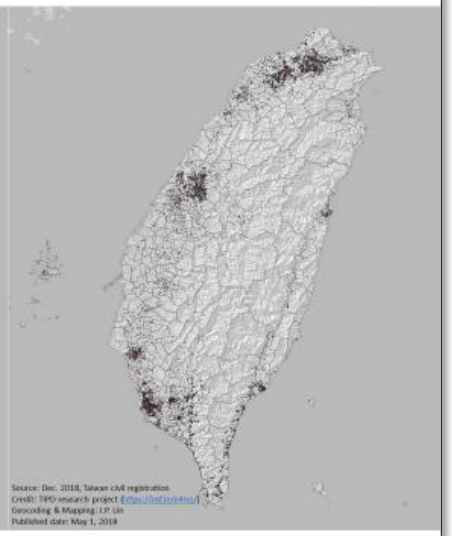


Figure - 布農(Bunun)

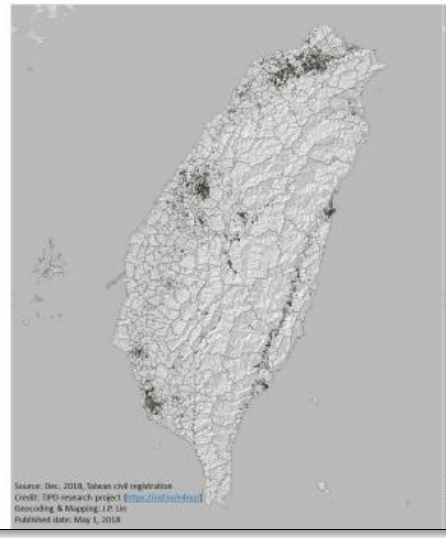


Figure - 魯凱(Rukai)

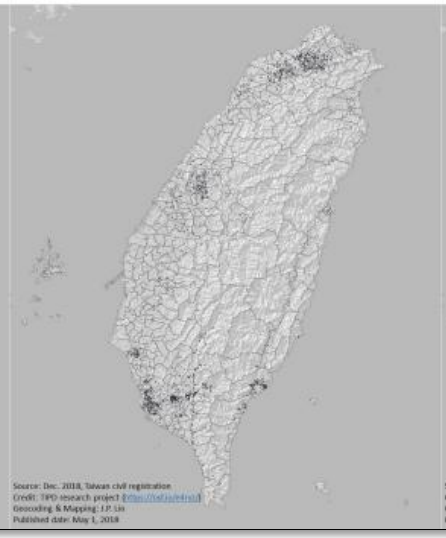


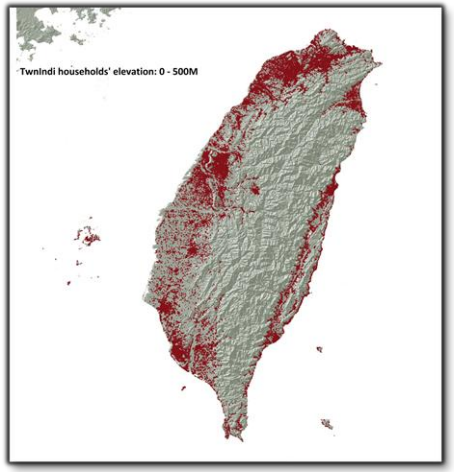
Figure - 卑南(Puyuma)

3. Building, Enriching, and Integrating Data

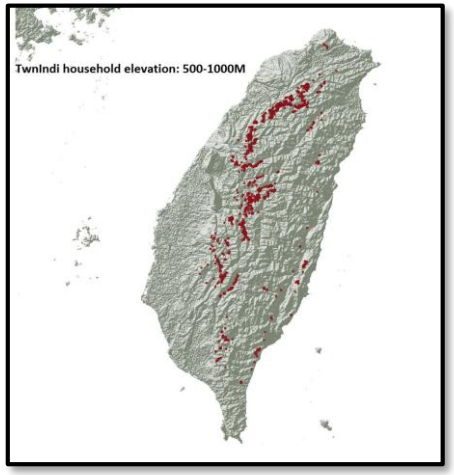
- **Building spatial information**: develop highly precise automated geocoding to locate TIPs

Distribution by elevation:

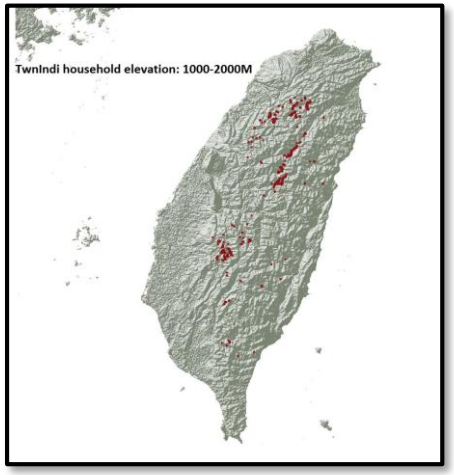
Elevation: 0-500m



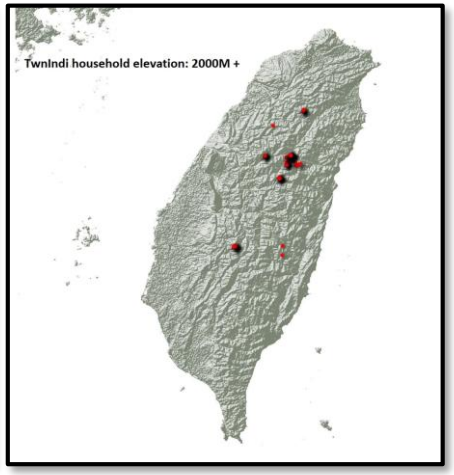
Elevation: 500-1000m



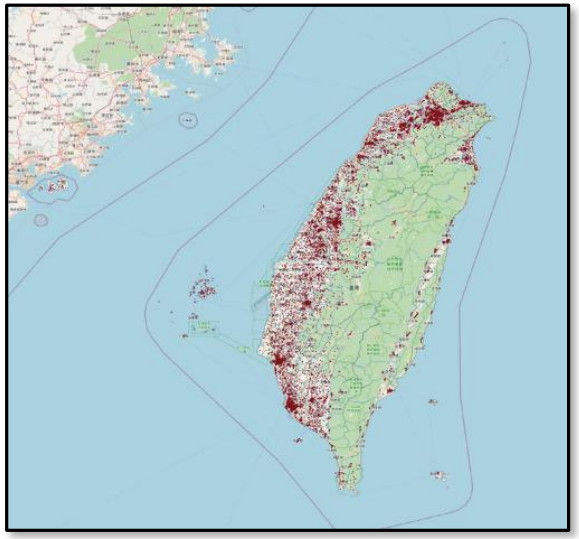
Elevation: 1000-2000m

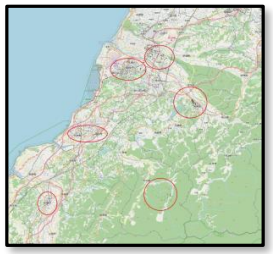


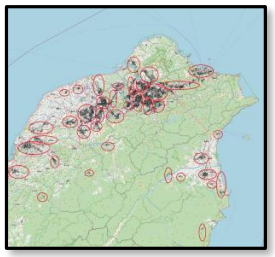
Elevation: 2000m +

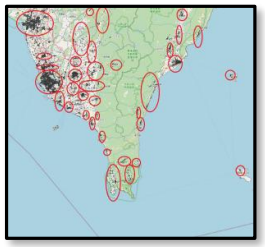


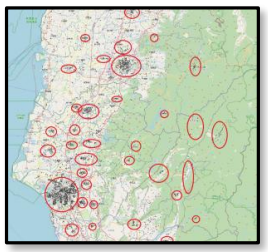
Localities: distribution by clusters:









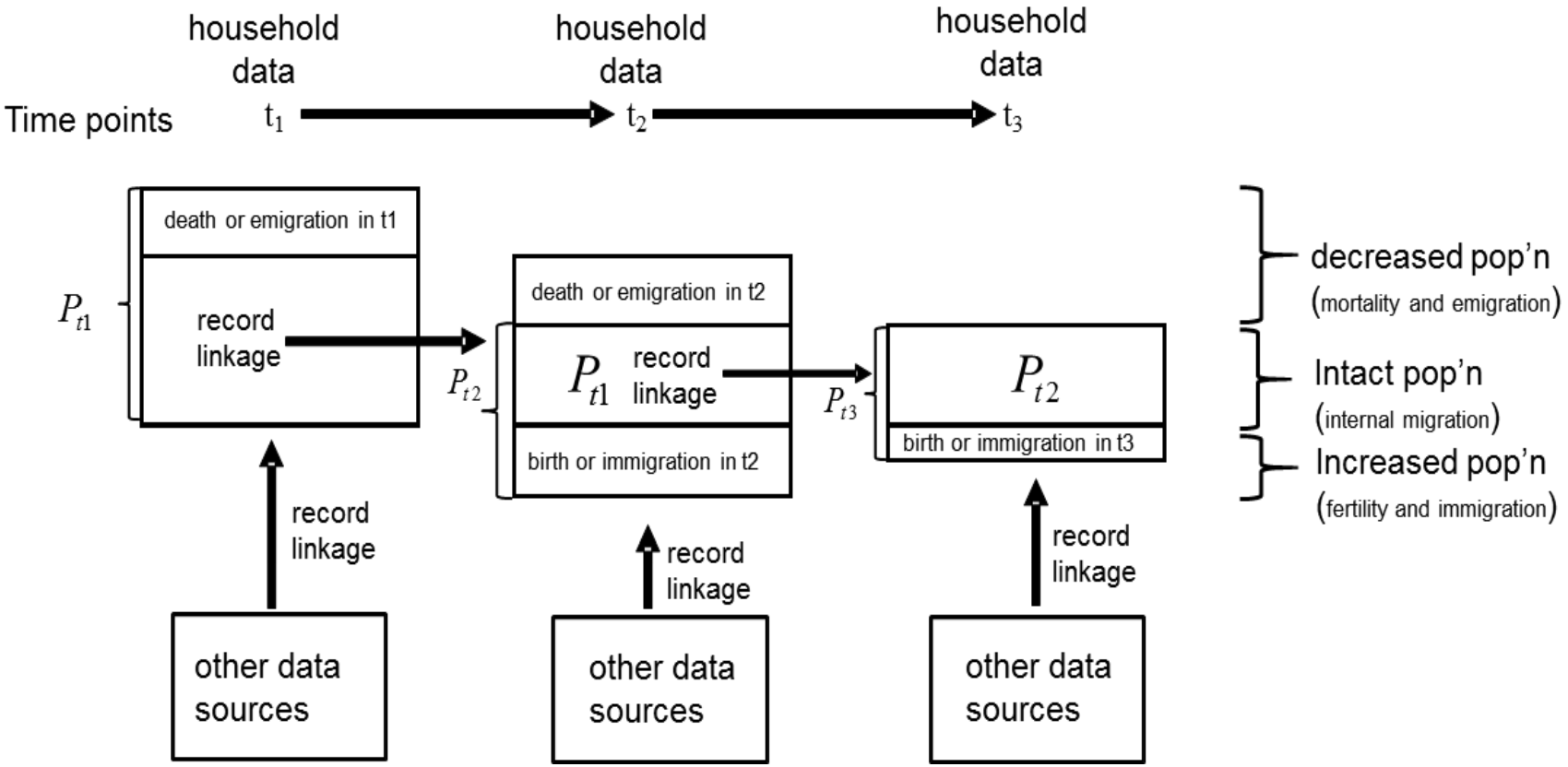


3. Building, Enriching, and Integrating Data

- **Building longitudinally linked register**

✓ **population dynamics data**

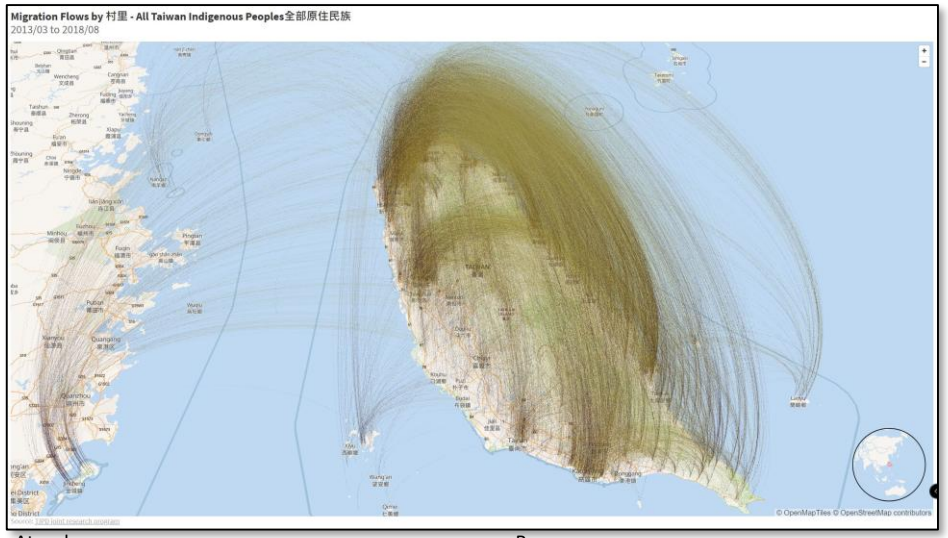
1. Pop'n of increase: comprising of "birth" & "immigration"
2. Pop'n of decrease: comprising of “death" & "emigration"
3. Pop'n of intact: comprising of "staying-put" & "internal migrants“



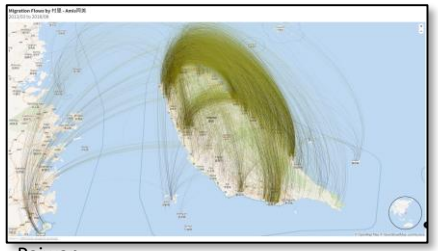
3. Building, Enriching, and Integrating Data

- **Population dynamics enable us to explore a number of dynamics by ethnicity (**e.g., birth, death, migration, marriage, labor force, social mobility, income, health, infectious diseases, life expectancy, family & household, etc**.)**

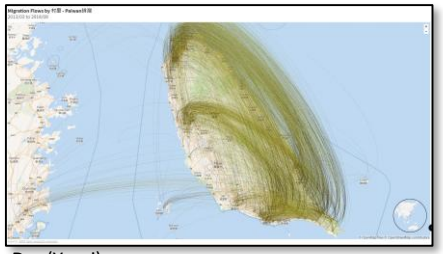
e.g., individual-level migration flows reflect kinship network & spreading paths of infectious diseases



Amis

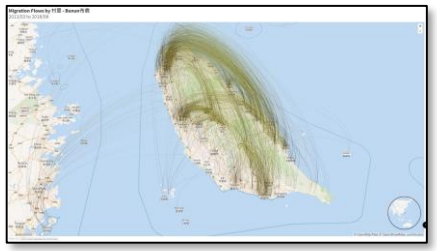


Paiwan

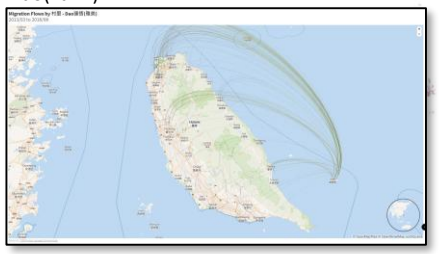


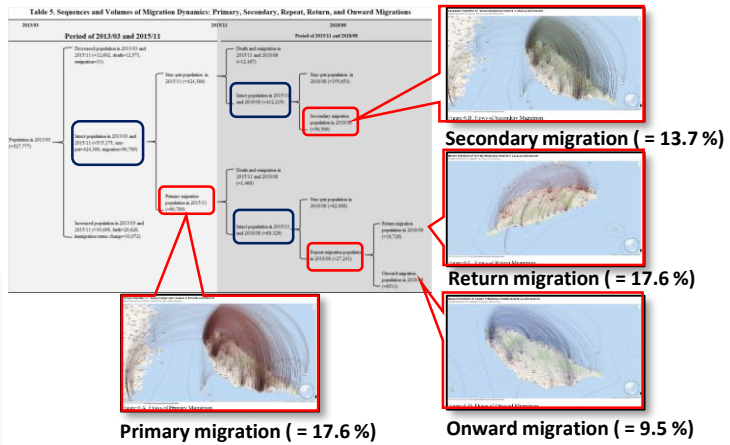
Atayal

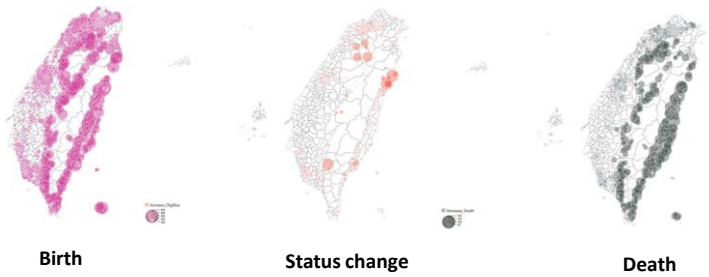
Bunun



Dao(Yami)







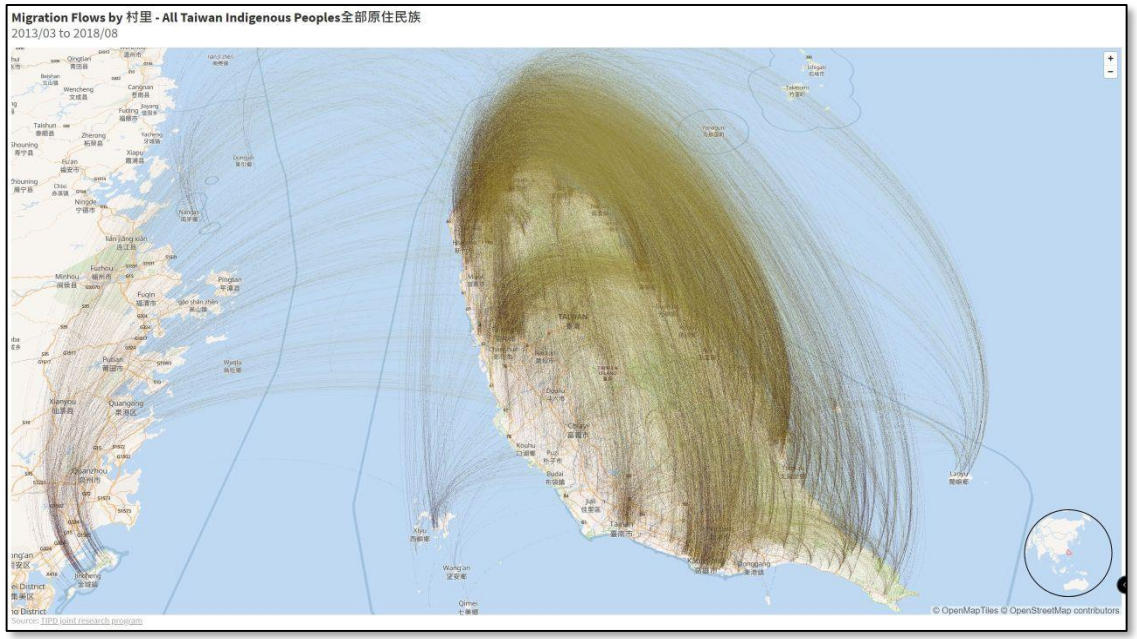
3. Building, Enriching, and Integrating Data

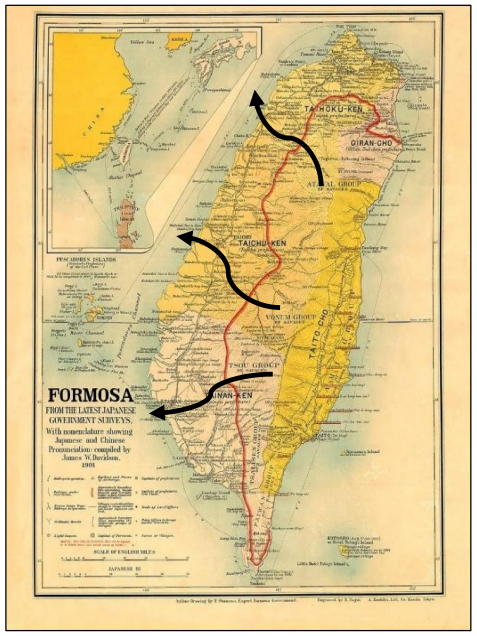
■ Population dynamics & social inclusion/integration: TIPs vs non-TIPs migrations

e.g., Social inclusion & integration

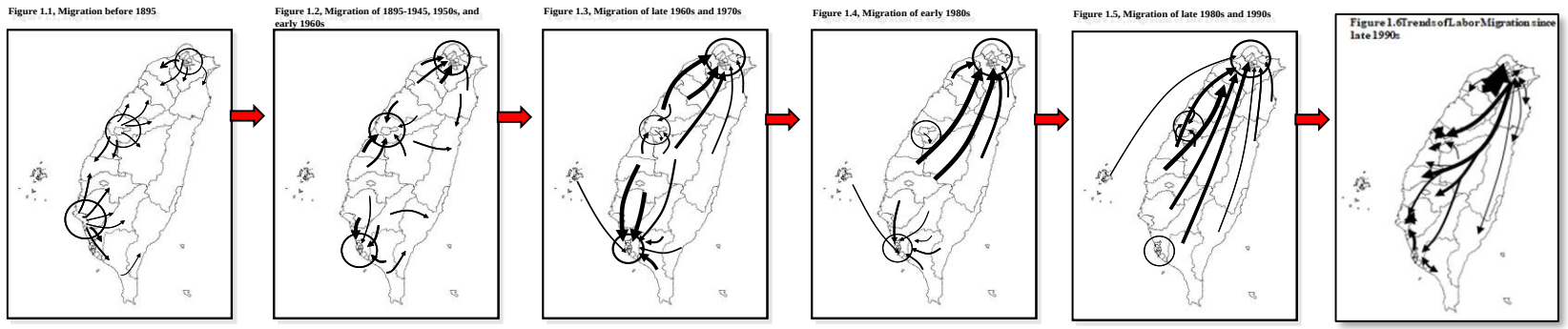
TIPs vs Non-TIPs

Figure, TIPs Population Dynamics in 2013-2018: Internal Migration





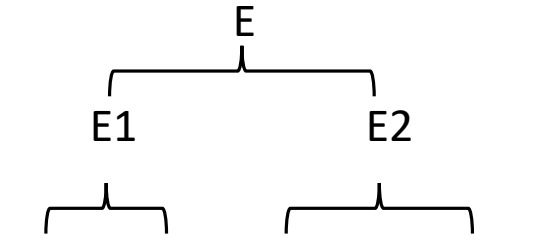
Figure, Migration Dynamics in Taiwan



3. Building, Enriching, and Integrating Data

■ Building Individual-level Genealogy Data

- ✓ A genealogical classification of a given ethnic group refers to a “standardized” hierarchal grouping classifications in the context of genealogy.
- ✓ Genealogy data enables us to systematically study how migration dynamics shape contemporary ethic population distribution and the evolutions of the ethnic group, i.e., evolution of family tree.
- ✓ Genealogy data enables us to systematically study: e.g.,
  1. Effect of migration dynamics on shaping contemporary ethic population
  2. ethnic relationships
  3. ethnic identity
  4. ethnic marriage practice
  5. socio-ethnical network



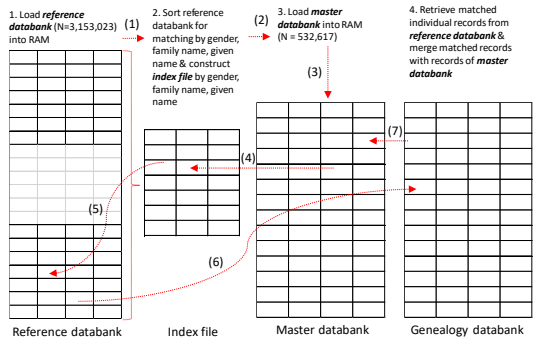


Figure 1 procedures of record matching using in-memory computing

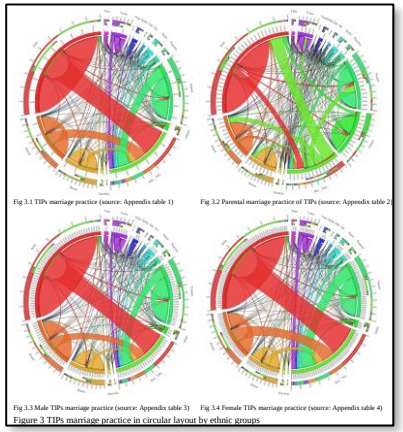


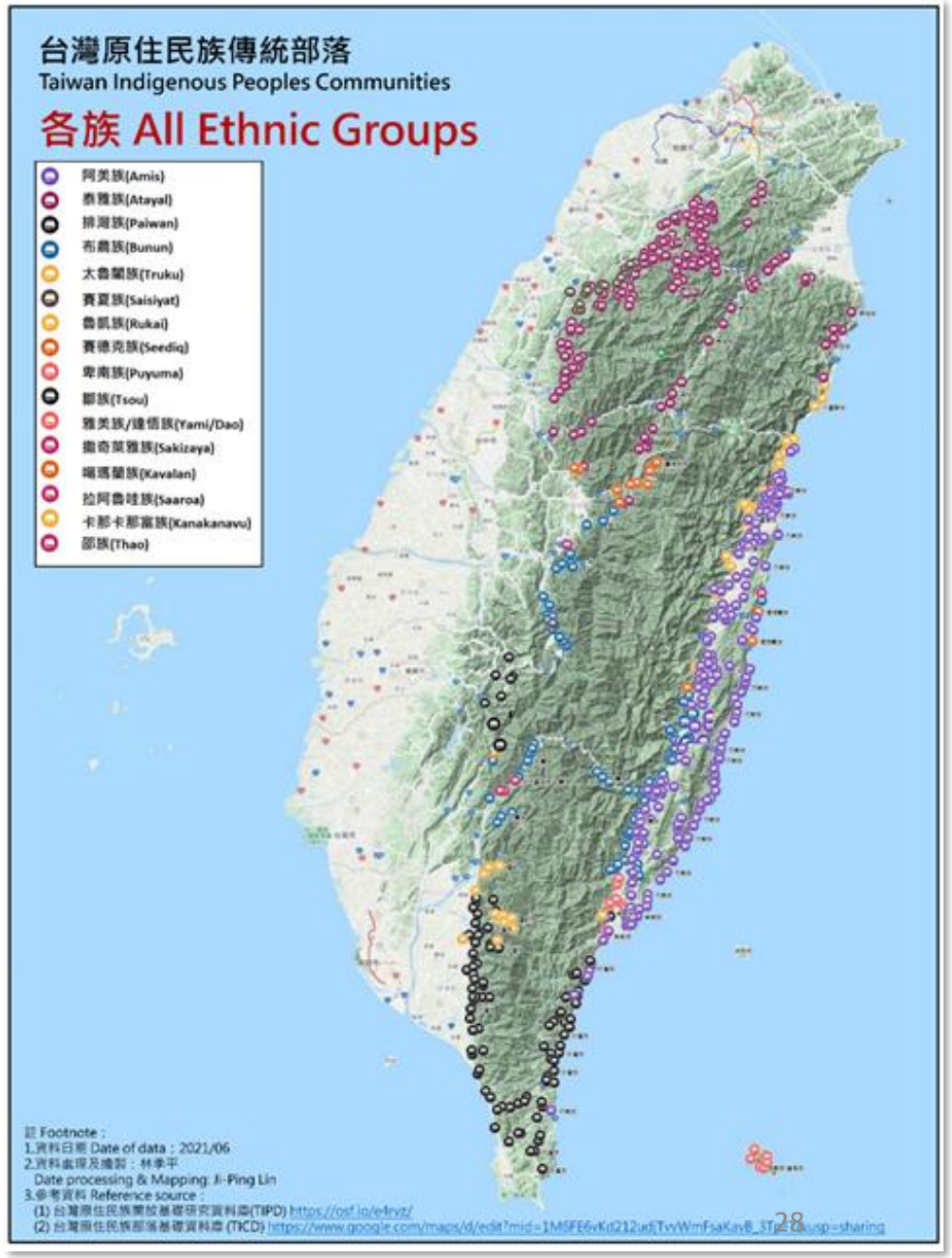
Figure 3 TIPs marriage practice in circular layout by ethnic groups

Table 1 Marriage practice and category of ethnic identity formation

Type of ethnic marriage practice	Type of ethnic identify formation
Ethnic marriage Practice	
Endogamy	
Intra-ethnic endogamy	Mono-ethnic identity
	Unspecified ethnic identity
Inter-ethnic endogamy	Multi-ethnic identity
	Patrilineal ethnic identity
	Matrilineal ethnic identity
	Unspecified ethnic identity
Exogamy	
	Multi-ethnic identity
	Patrilineal ethnic identity
	Matrilineal ethnic identity
	Unspecified ethnic identity

3. Building, Enriching, and Integrating Data

- **Building traditional TIPs community open data**
- **Impact of more than 8-million accesses**
  - ✓ Promote tourism
  - ✓ Promote ethnic identity & mutual understandings
  - ✓ Resource allocation policy (e.g., medical, educational, transportation
  - ✓ Hazard mitigation design
- **Methods of building community data**
  - ✓ Each indigenous community is a set of households.
  - ✓ There are 716 indigenous communities in Taiwan.
  - ✓ Use address information in source data to define the scope of each community.
  - ✓ Household spatial information is used to calculate the population center for each indigenous community.



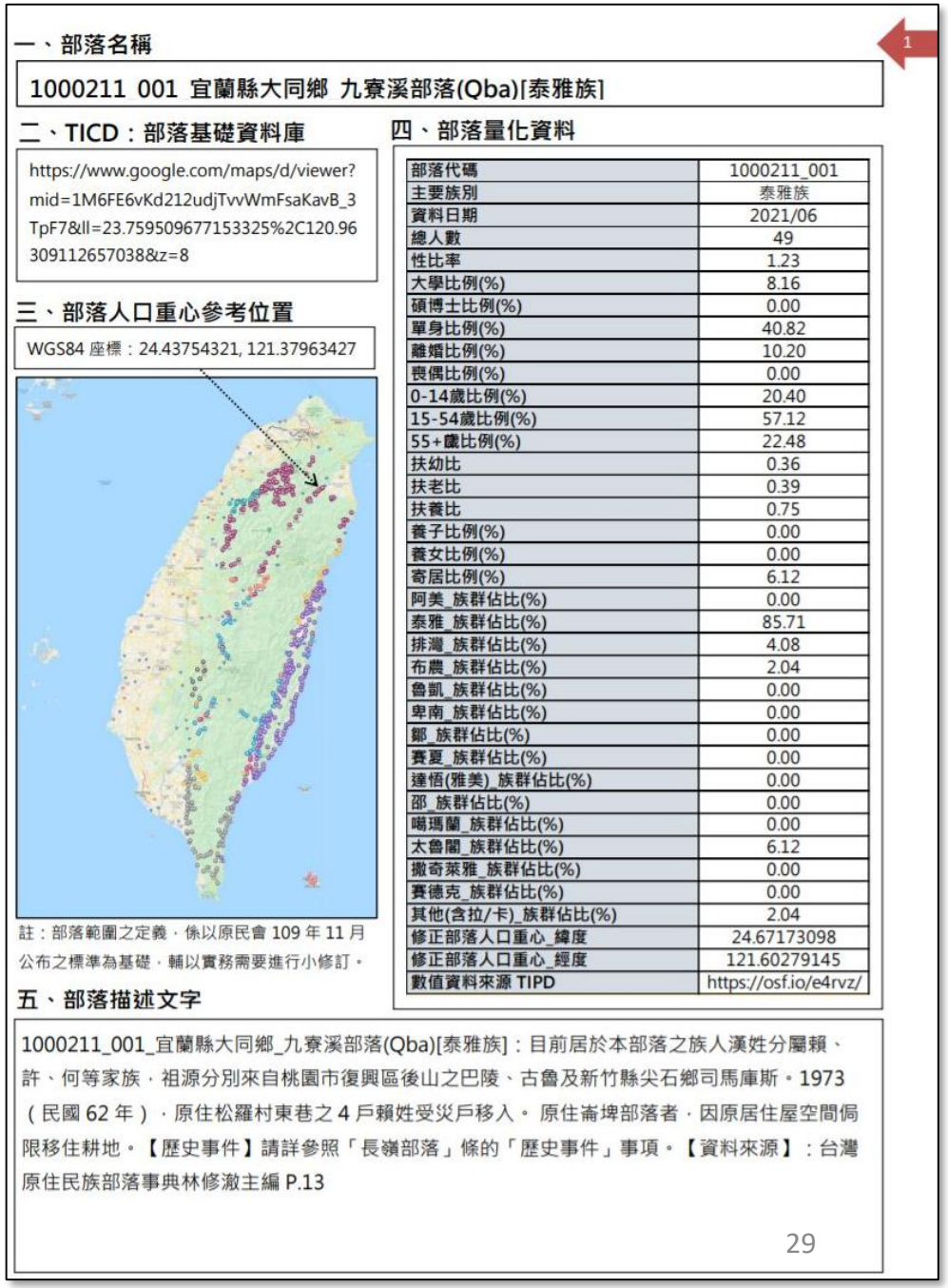
3. Building, Enriching, and Integrating Data

■ Building TICD distribution by ethnicity

✓ Creating TICD: Compute standardized statistical figures:

ethnical lineage,

1) population size,
2) sex ratio,
3) educational composition,
4) age composition,
5) marital status composition,
6) child dependency ratio,
7) old-age dependency ratio,
8) dependency ratio,
9) population composition by ethnic group,
10) population center.



一、部落名稱

1000211 001 宜蘭縣大同鄉 九寮溪部落(Qba)[泰雅族]

二、TICD：部落基礎資料庫

https://www.google.com/maps/d/viewer?mid=1M6FE6vKd212udjTvvWmFsaKavB_3TpF7&ll=23.759509677153325%2C120.963091126570 38&z=8

三、部落人口重心參考位置

WGS84 座標：24.43754321, 121.37963427

註：部落範圍之定義，係以原民會 109 年 11 月公布之標準為基礎，輔以實務需要進行小修訂。

四、部落量化資料

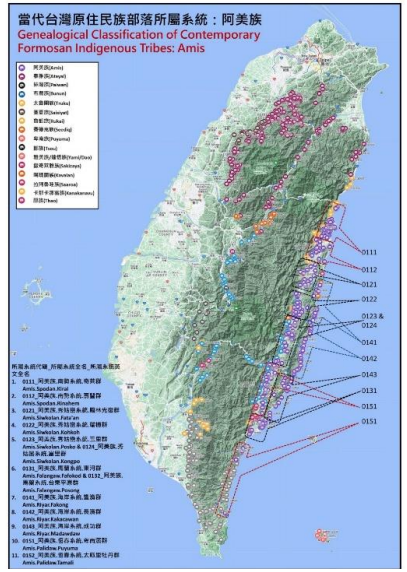
部落代碼	1000211_001
主要族別	泰雅族
資料日期	2021/06
總人數	49
性比率	1.23
大學比例(%)	8.16
碩博士比例(%)	0.00
單身比例(%)	40.82
離婚比例(%)	10.20
喪偶比例(%)	0.00
0-14歲比例(%)	20.40
15-54歲比例(%)	57.12
55+歲比例(%)	22.48
扶幼比	0.36
扶老比	0.39
扶養比	0.75
養子比例(%)	0.00
養女比例(%)	0.00
寄居比例(%)	6.12
阿美_族群佔比(%)	0.00
泰雅_族群佔比(%)	85.71
排灣_族群佔比(%)	4.08
布農_族群佔比(%)	2.04
魯凱_族群佔比(%)	0.00
卑南_族群佔比(%)	0.00
鄒_族群佔比(%)	0.00
賽夏_族群佔比(%)	0.00
達悟(雅美)_族群佔比(%)	0.00
邵_族群佔比(%)	0.00
噶瑪蘭_族群佔比(%)	0.00
太魯閣_族群佔比(%)	6.12
撒奇萊雅_族群佔比(%)	0.00
賽德克_族群佔比(%)	0.00
其他(含拉/卡)_族群佔比(%)	2.04
修正部落人口重心_緯度	24.67173098
修正部落人口重心_經度	121.60279145
數值資料來源 TIPD	https://osf.io/e4rvz/

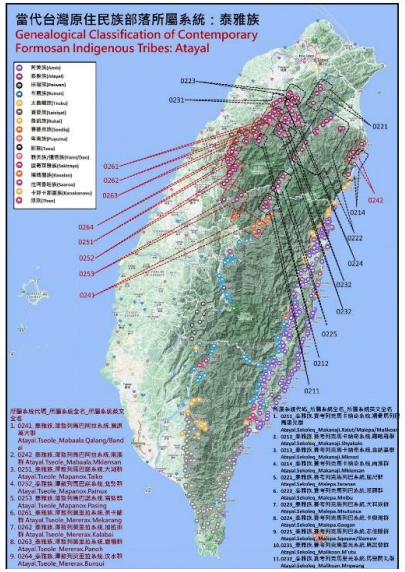
五、部落描述文字

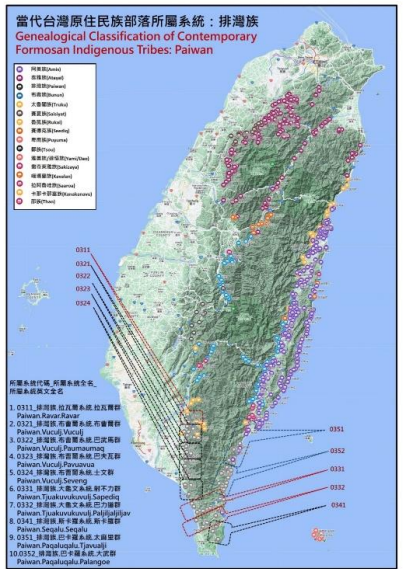
1000211_001_宜蘭縣大同鄉_九寮溪部落(Qba)[泰雅族]：目前居於本部落之族人漢姓分屬賴、許、何等家族，祖源分別來自桃園市復興區後山之巴陵、古魯及新竹縣尖石鄉司馬庫斯。1973（民國 62 年），原住松羅村東巷之 4 戶賴姓受災戶移入。原住崙埤部落者，因原居住屋空間侷限移住耕地。【歷史事件】請詳參照「長嶺部落」條的「歷史事件」事項。【資料來源】：台灣原住民族部落事典林修澈主編 P.13

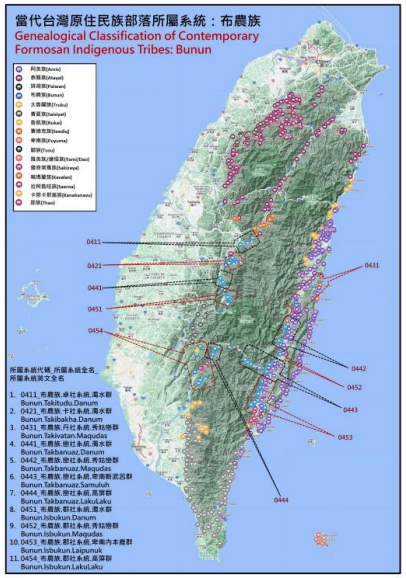
3. Building, Enriching, and Integrating Data

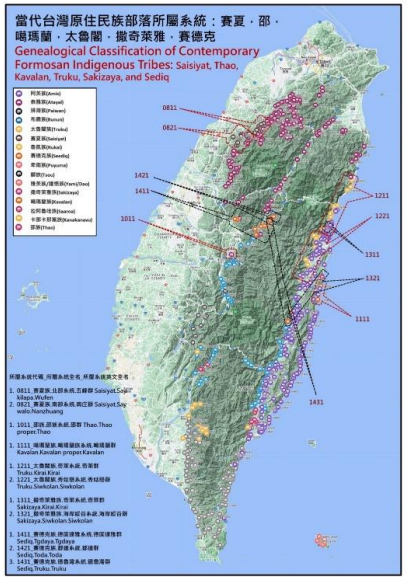
■ Building ethnic genealogical lineage

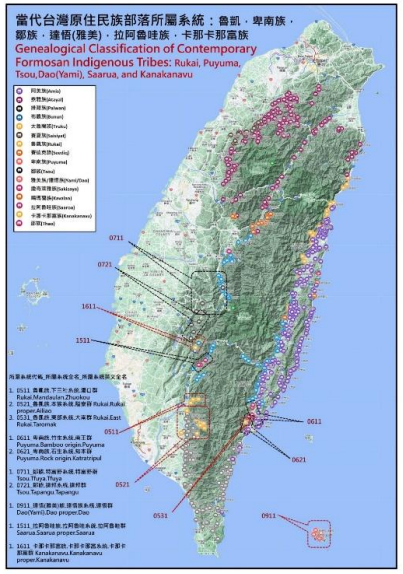


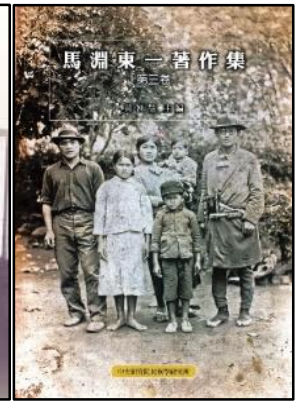


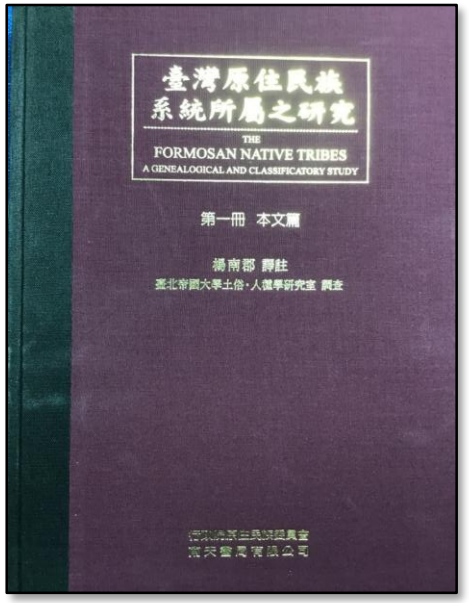


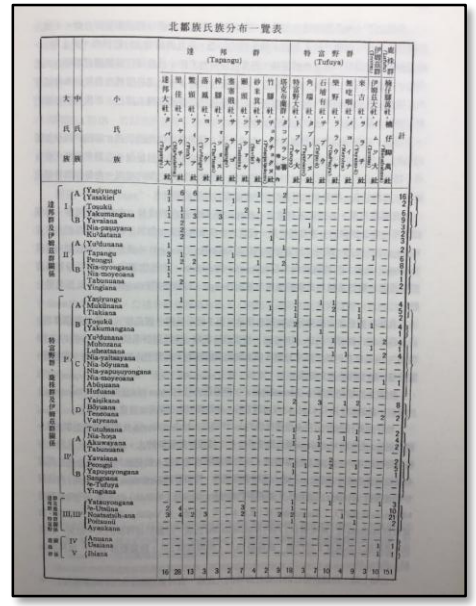












Utsurikawa N, Miyamoto N, Mabuchi T (1935) The Formosan native tribes: a genealogical and classificatory system. Institute of Ethnology, Taipei Imperial University.
https://tm.ncl.edu.tw/article?u=008_106_000013

3. Building, Enriching, and Integrating Data

■ Classification of Ethnic genealogical lineage

✓ Three-category classification:

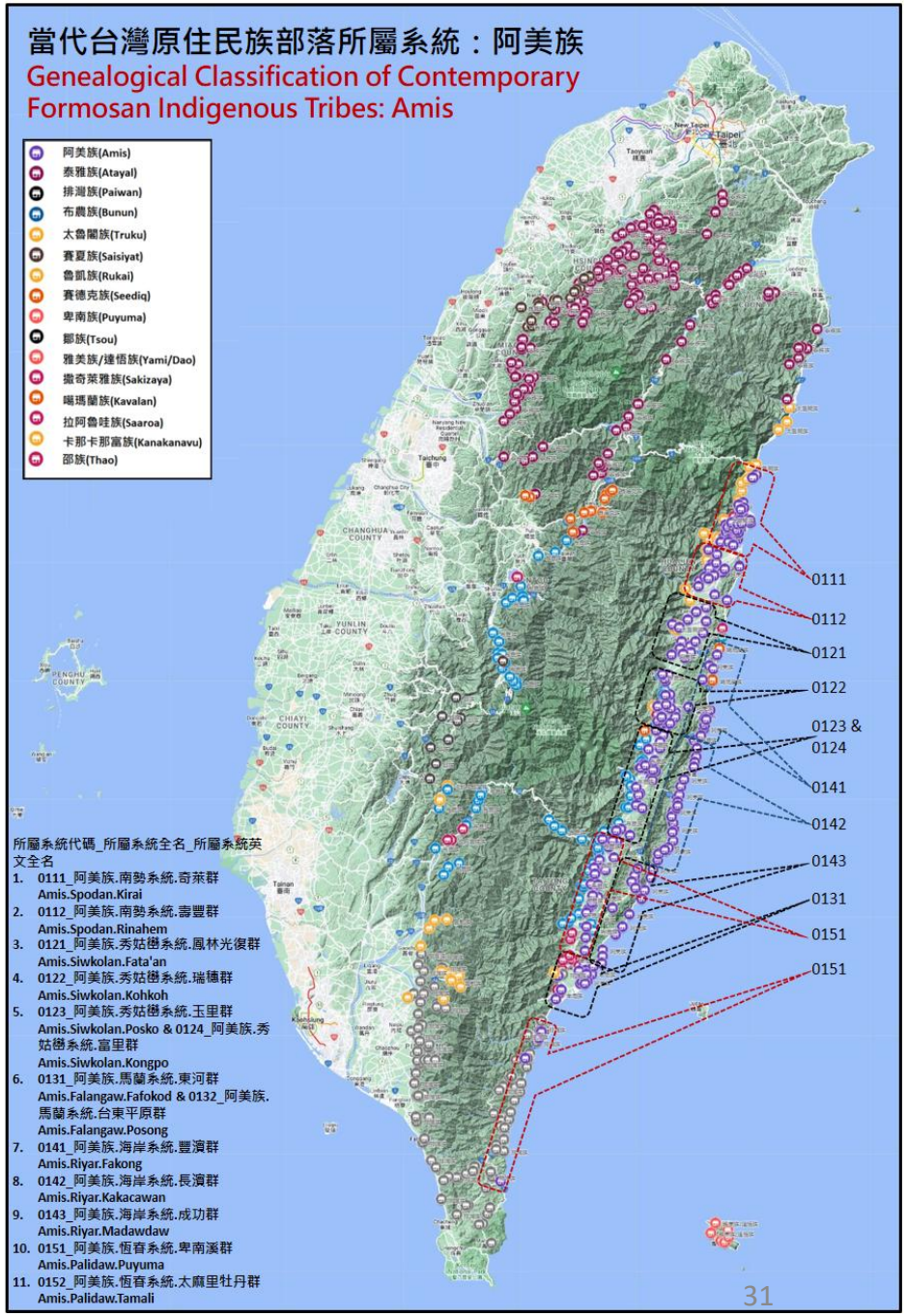
Ethnicity.PrimaryLineage.SecondaryLineage

1) Lineage code: 0132 (01: Amis code; 3: Falangaw; 2: Posong)
2) 所屬系統：阿美族.馬蘭系統.台東平原群
3) Lineage classification: Amis.Falangaw.Posong

■ Building ethnic genealogical lineage

✓ Genealogical Classification: Amis

所屬系統代碼	所屬系統全名	Genealogical Classifcation Name
0111	阿美族.南勢系統.奇萊群	Amis.Spodan.Kirai
0112	阿美族.南勢系統.壽豐群	Amis.Spodan.Rinahem
0121	阿美族.秀姑巒系統.鳳林光復群	Amis.Siwkolan.Fata'an
0122	阿美族.秀姑巒系統.瑞穗群	Amis.Siwkolan.Kohkoh
0123	阿美族.秀姑巒系統.玉里群	Amis.Siwkolan.Posko
0124	阿美族.秀姑巒系統.富里群	Amis.Siwkolan.Kongpo
0131	阿美族.馬蘭系統.東河群	Amis.Falangaw.Fafokod
0132	阿美族.馬蘭系統.台東平原群	Amis.Falangaw.Posong
0141	阿美族.海岸系統.豐濱群	Amis.Riyar.Fakong
0142	阿美族.海岸系統.長濱群	Amis.Riyar.Kakacawan
0143	阿美族.海岸系統.成功群	Amis.Riyar.Madawdaw
0151	阿美族.恆春系統.卑南溪群	Amis.Palidaw.Puyuma
0152	阿美族.恆春系統.太麻里牡丹群	Amis.Palidaw.Tamali

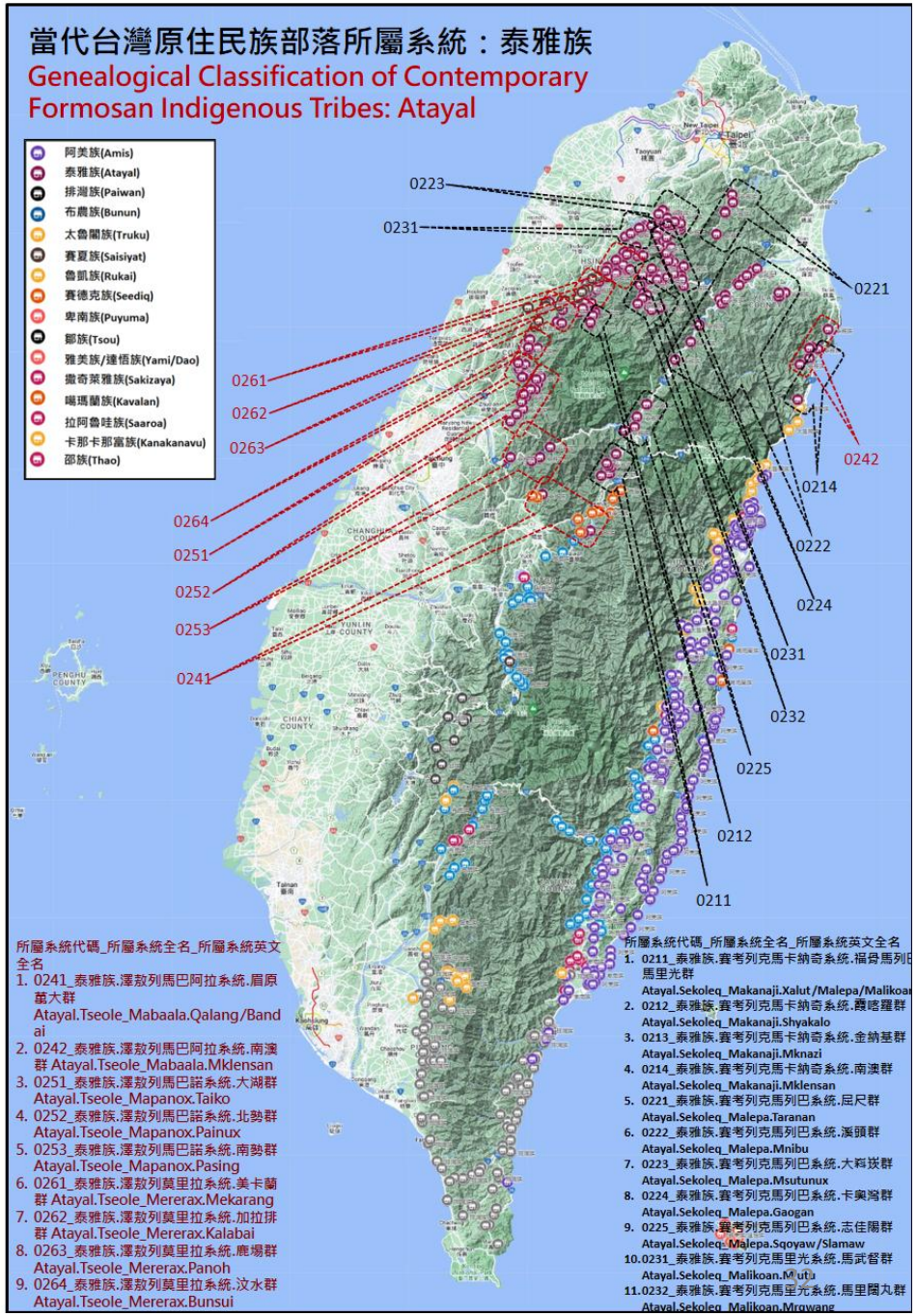


3. Building, Enriching, and Integrating Data

■ Building ethnic genealogical lineage

✓ Genealogical Classification: Atayal

所屬系統代碼	所屬系統全名	Genealogical Classifcation Name
0211	泰雅族.賽考列克馬卡納奇系統.福骨馬列巴馬里光群	Atayal.Sekoleq_Makanaji.Xalut/Malepa/Malikoan
0212	泰雅族.賽考列克馬卡納奇系統.霞喀羅群	Atayal.Sekoleq_Makanaji.Shyakalo
0213	泰雅族.賽考列克馬卡納奇系統.金納基群	Atayal.Sekoleq_Makanaji.Mknazi
0214	泰雅族.賽考列克馬卡納奇系統.南澳群	Atayal.Sekoleq_Makanaji.Mklensan
0221	泰雅族.賽考列克馬列巴系統.屈尺群	Atayal.Sekoleq_Malepa.Taranan
0222	泰雅族.賽考列克馬列巴系統.溪頭群	Atayal.Sekoleq_Malepa.Mnibu
0223	泰雅族.賽考列克馬列巴系統.大嵙崁群	Atayal.Sekoleq_Malepa.Msutunux
0224	泰雅族.賽考列克馬列巴系統.卡奧灣群	Atayal.Sekoleq_Malepa.Gaogan
0225	泰雅族.賽考列克馬列巴系統.志佳陽群	Atayal.Sekoleq_Malepa.Sqoyaw/Slamaw
0231	泰雅族.賽考列克馬里光系統.馬武督群	Atayal.Sekoleq_Malikoan.M'utu
0232	泰雅族.賽考列克馬里光系統.馬里闊丸群	Atayal.Sekoleq_Malikoan.Mrqwang
0241	泰雅族.澤敖列馬巴阿拉系統.眉原萬大群	Atayal.Tseole_Mabaala.Qalang/Bandai
0242	泰雅族.澤敖列馬巴阿拉系統.南澳群	Atayal.Tseole_Mabaala.Mklensan
0251	泰雅族.澤敖列馬巴諾系統.大湖群	Atayal.Tseole_Mapanox.Taiko
0252	泰雅族.澤敖列馬巴諾系統.北勢群	Atayal.Tseole_Mapanox.Painux
0253	泰雅族.澤敖列馬巴諾系統.南勢群	Atayal.Tseole_Mapanox.Pasing
0261	泰雅族.澤敖列莫里拉系統.美卡蘭群	Atayal.Tseole_Mererax.Mekarang
0262	泰雅族.澤敖列莫里拉系統.加拉排群	Atayal.Tseole_Mererax.Kalabai
0263	泰雅族.澤敖列莫里拉系統.鹿場群	Atayal.Tseole_Mererax.Panoh
0264	泰雅族.澤敖列莫里拉系統.汶水群	Atayal.Tseole_Mererax.Bunsui

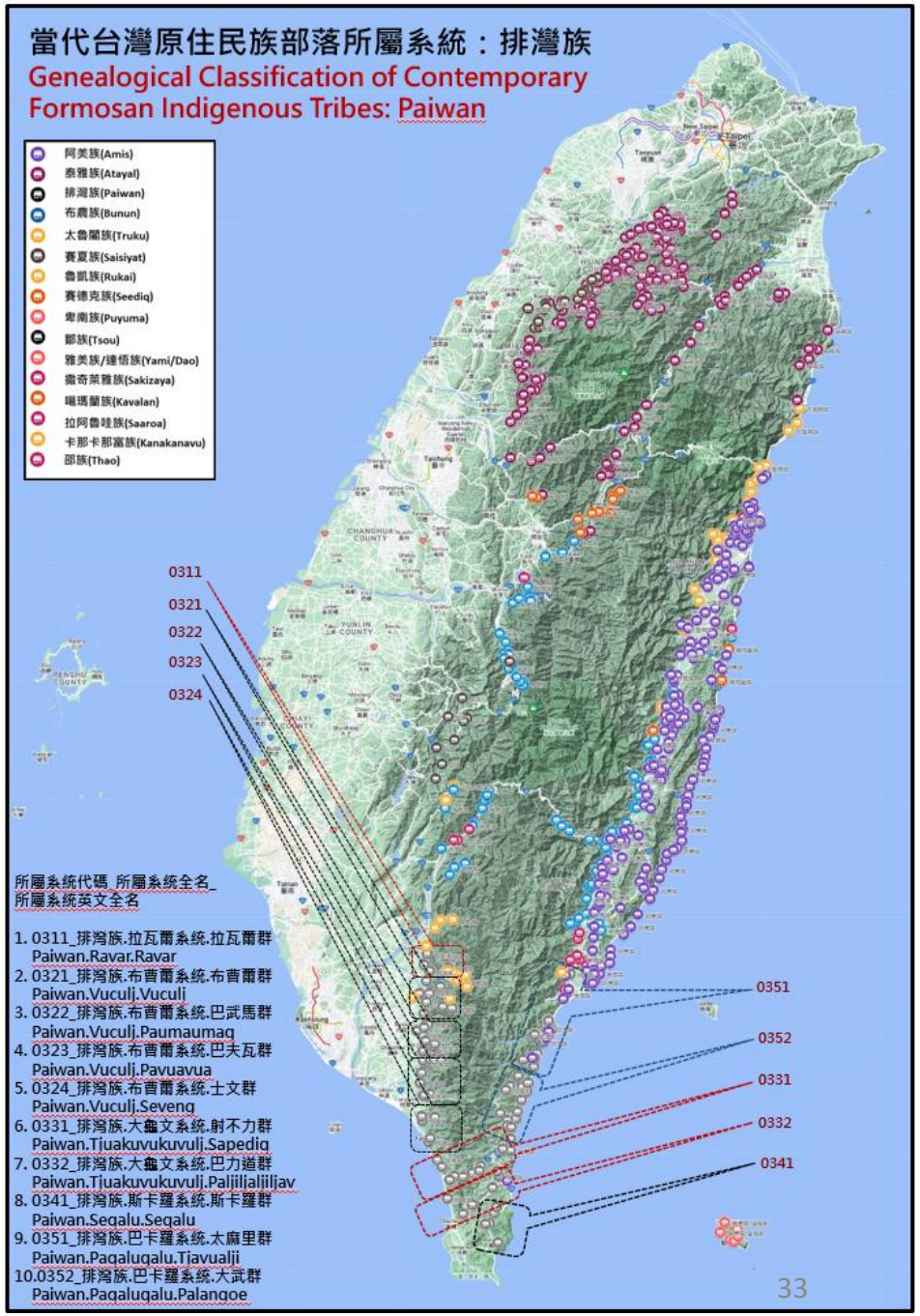


3. Building, Enriching, and Integrating Data

■ Building ethnic genealogical lineage

✓ Genealogical Classification: Paiwan

所屬系統代碼	族群	所屬系統_主系統	所屬系統_次系統	所屬系統英文全名
0311	排灣族	拉瓦爾系統	拉瓦爾群	Paiwan.Ravar.Ravar
0321	排灣族	布曹爾系統	布曹爾群	Paiwan.Vuculj.Vuculj
0322	排灣族	布曹爾系統	巴武馬群	Paiwan.Vuculj.Paumaumaq
0323	排灣族	布曹爾系統	巴夫瓦群	Paiwan.Vuculj.Pavuavua
0324	排灣族	布曹爾系統	士文群	Paiwan.Vuculj.Seveng
0331	排灣族	大龜文系統	射不力群	Paiwan.Tjuakuvukuvulj.Sapediq
0332	排灣族	大龜文系統	巴力道群	Paiwan.Tjuakuvukuvulj.Paljiljaljiljav
0341	排灣族	斯卡羅系統	斯卡羅群	Paiwan.Seqalu.Seqalu
0351	排灣族	巴卡羅系統	太麻里群	Paiwan.Paqaluqalu.Tjavualji
0352	排灣族	巴卡羅系統	大武群	Paiwan.Paqaluqalu.Palangoe

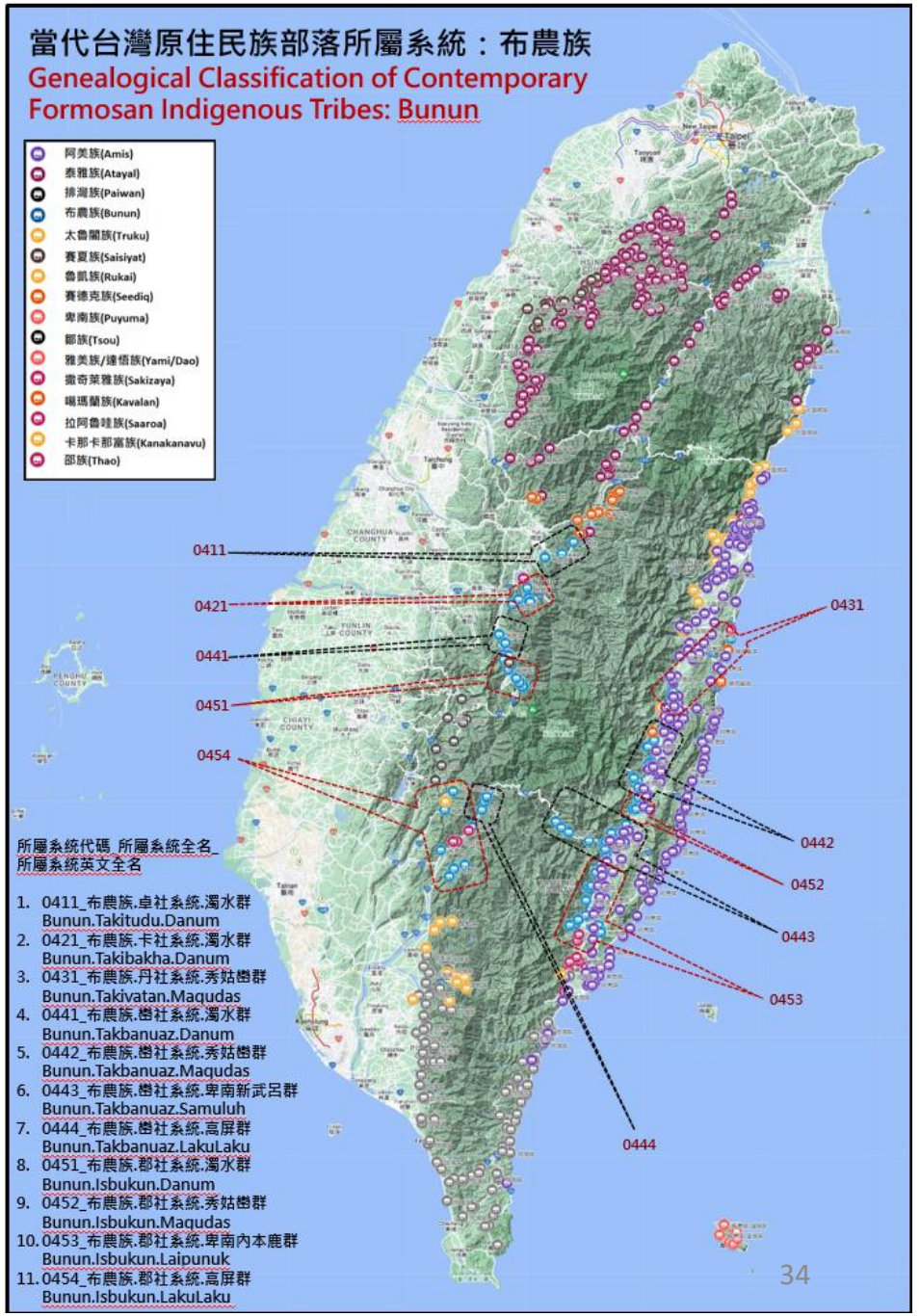


3. Building, Enriching, and Integrating Data

■ Building ethnic genealogical lineage

✓ Genealogical Classification: Bunun

所屬系統代碼	族群	所屬系統_主系統	所屬系統_次系統	所屬系統英文全名
0411	布農族	卓社系統	濁水群	Bunun.Takitudu.Danum
0421	布農族	卡社系統	濁水群	Bunun.Takibakha.Danum
0431	布農族	丹社系統	秀姑巒群	Bunun.Takivatan.Maqudas
0441	布農族	巒社系統	濁水群	Bunun.Takbanuaz.Danum
0442	布農族	巒社系統	秀姑巒群	Bunun.Takbanuaz.Maqudas
0443	布農族	巒社系統	卑南新武呂群	Bunun.Takbanuaz.Samuluh
0444	布農族	巒社系統	高屏群	Bunun.Takbanuaz.LakuLaku
0451	布農族	郡社系統	濁水群	Bunun.Isbukun.Danum
0452	布農族	郡社系統	秀姑巒群	Bunun.Isbukun.Maqudas
0453	布農族	郡社系統	卑南內本鹿群	Bunun.Isbukun.Laipunuk
0454	布農族	郡社系統	高屏群	Bunun.Isbukun.LakuLaku

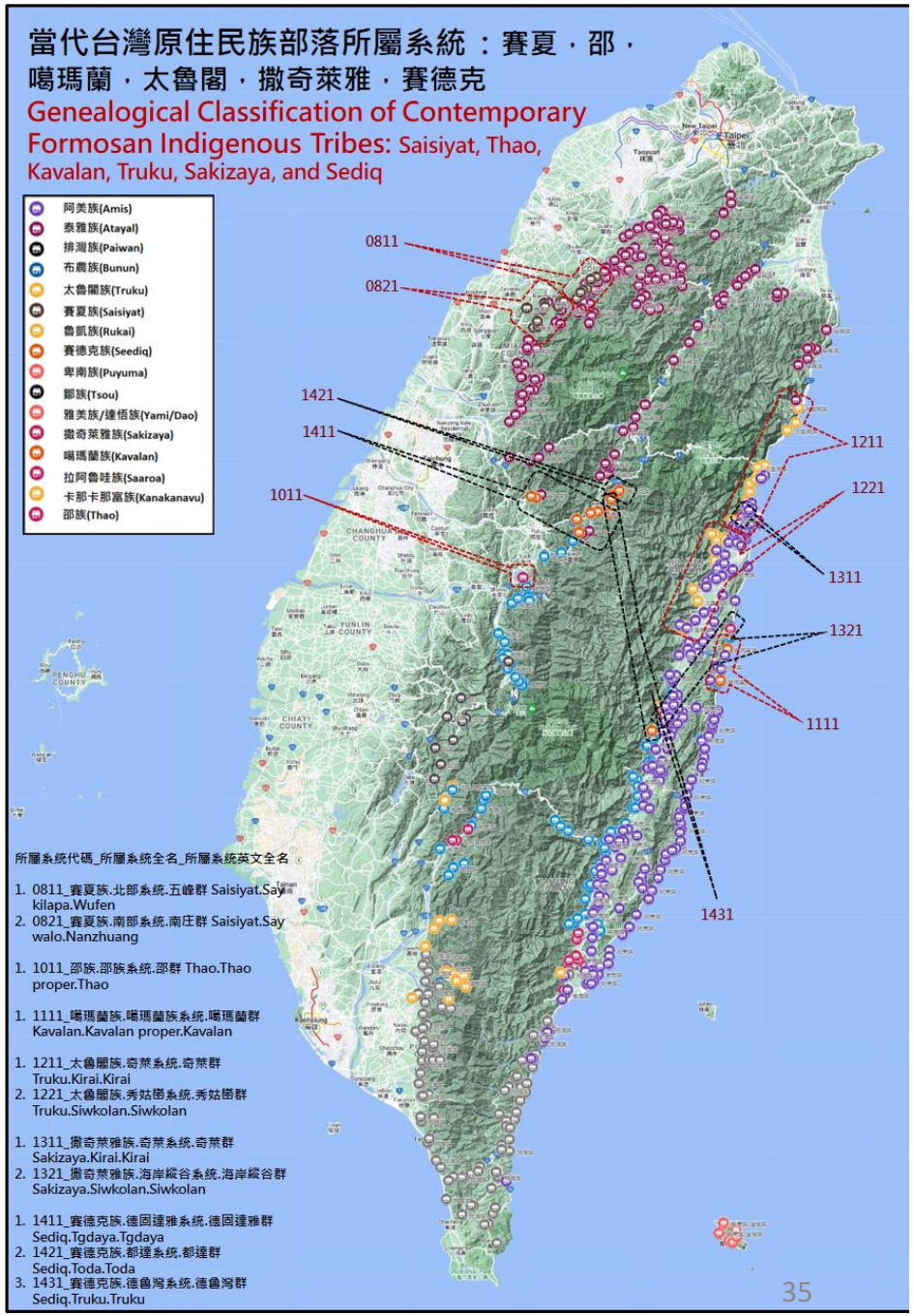


3. Building, Enriching, and Integrating Data

■ Building ethnic genealogical lineage

✓ Genealogical Classification: Saisiyat, Thao, Kavalan, Truku, Sakizaya, and Sediq

所屬系統代碼	族群	所屬系統_主系統	所屬系統_次系統	所屬系統英文全名
0811	賽夏族	北部系統	五峰群	Saisiyat.Say kilapa.Wufen
0821	賽夏族	南部系統	南庄群	Saisiyat.Say walo.Nanzhuang
1011	邵族	邵族系統	邵群	Thao.Thao proper.Thao
1111	噶瑪蘭族	噶瑪蘭族系統	噶瑪蘭群	Kavalan.Kavalan proper.Kavalan
1211	太魯閣族	奇萊系統	奇萊群	Truku.Kirai.Kirai
1221	太魯閣族	秀姑巒系統	秀姑巒群	Truku.Siwkolan.Siwkolan
1311	撒奇萊雅族	奇萊系統	奇萊群	Sakizaya.Kirai.Kirai
1321	撒奇萊雅族	海岸縱谷系統	海岸縱谷群	Sakizaya.Siwkolan.Siwkolan
1411	賽德克族	德固達雅系統	德固達雅群	Sediq.Tgdaya.Tgdaya
1421	賽德克族	都達系統	都達群	Sediq.Toda.Toda
1431	賽德克族	德魯灣系統	德魯灣群	Sediq.Truku.Truku

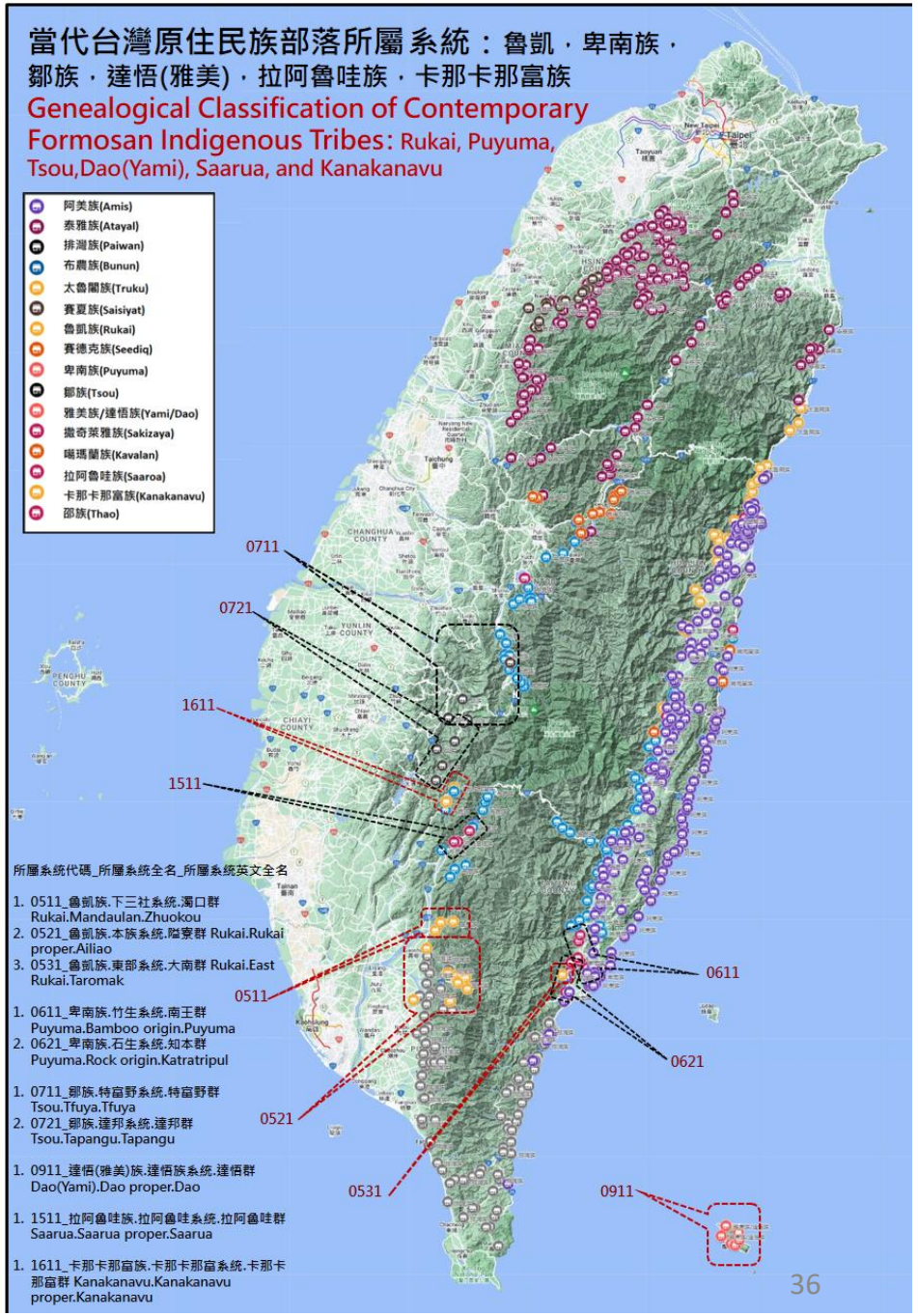


3. Building, Enriching, and Integrating Data

■ Building ethnic genealogical lineage

✓ Genealogical Classification: Rukai, Puyuma, Tsou,Dao(Yami), Saarua, and Kanakanavu

所屬系統代碼	族群	所屬系統_主系統	所屬系統_次系統	所屬系統英文全名
0511	魯凱族	下三社系統	濁口群	Rukai.Mandaulan.Zhuokou
0521	魯凱族	本族系統	隘寮群	Rukai.Rukai proper.Ailiao
0531	魯凱族	東部系統	大南群	Rukai.East Rukai.Taromak
0611	卑南族	竹生系統	南王群	Puyuma.Bamboo origin.Puyuma
0621	卑南族	石生系統	知本群	Puyuma.Rock origin.Katratripul
0711	鄒族	特富野系統	特富野群	Tsou.Tfuya.Tfuya
0721	鄒族	達邦系統	達邦群	Tsou.Tapangu.Tapangu
0911	達悟(雅美)族	達悟族系統	達悟群	Dao(Yami).Dao proper.Dao
1511	拉阿魯哇族	拉阿魯哇系統	拉阿魯哇群	Saarua.Saarua proper.Saarua
1611	卡那卡那富族	卡那卡那富系統	卡那卡那富群	Kanakanavu.Kanakanavu proper.Kanakanavu



3. Building, Enriching, and Integrating Data

■ Building and harmonizing individual kinship network

✓ Base record information: “relationship with household head”.

✓ Processes of data cleansing, reorganizing, and harmonization: 99.5% automated & 0.5% manual data processing from 96 million records from 2007~2022 THRD.

✓ Harmonized kinship name and code:

和戶長關係_StandardizedCategories:

RelaWithHusHead_StdStr	RelaWithHusHead_StdCode
祖父母	1
外祖父母	2
父母	3
父母親屬	4
養父母	5
戶長	6
手足/手足配偶	7
手足/手足配偶親屬	8
配偶	9
配偶親屬	10
兒子	11
女兒	12
養子	13
養女	14
媳婦	15
女婿	16
孫子女	17
外孫子女	18
孫子女配偶及親屬	19
贅女婿	20
贅夫	21
家屬	22
寄居	23
領養親屬	24
其他	25
無資料	99

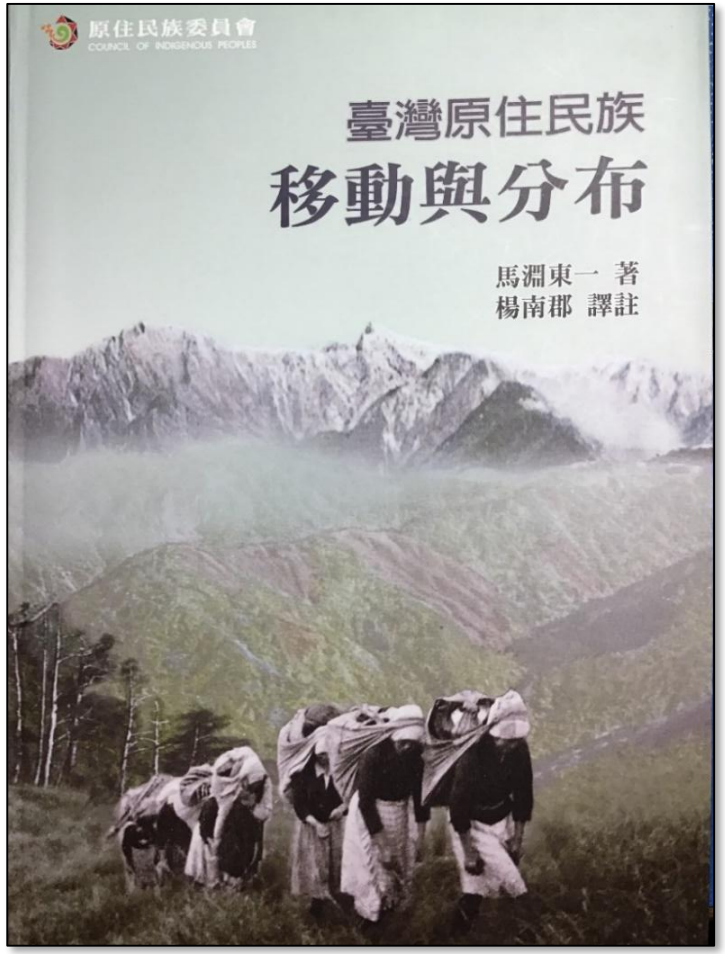
e.g., volume and share kinship compostion

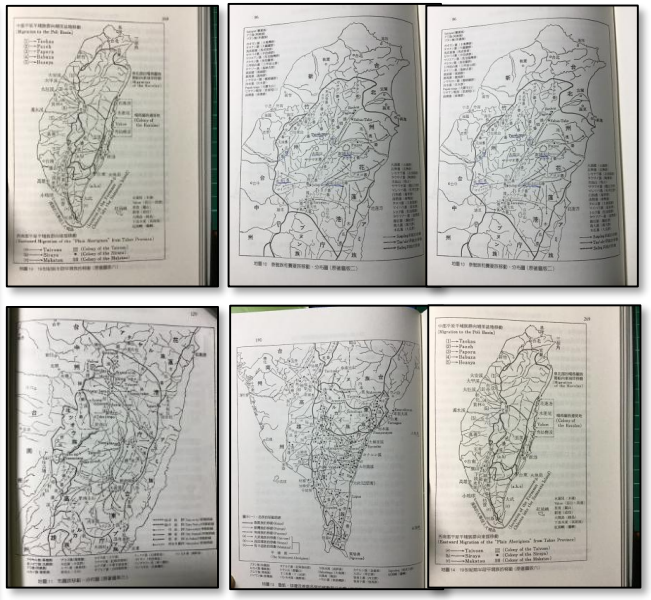
和戶長關係_標準化代碼	和戶長關係相對人數(人, 年平均)					
	2013	2014	2015	2016	2017	2018
全部人數	530,386	536,319	543,821	549,996	556,642	562,858
祖父母	259	262	273	278	275	271
外祖父母	7	8	9	10	9	12
父母	8,209	8,338	8,703	8,915	8,988	8,955
父母親屬	1,502	1,611	1,720	1,823	1,914	2,071
養父母	41	39	38	38	35	36
戶長	141,106	138,888	137,055	136,092	135,383	134,028
手足/手足配偶	18,612	18,958	19,225	19,284	19,504	20,041
手足/手足配偶親屬	11,577	12,116	12,577	12,876	13,288	13,976
配偶	47,882	47,829	48,270	48,705	49,094	49,209
配偶親屬	3,317	3,387	3,400	3,431	3,495	3,620
兒子	109,008	109,350	109,753	110,161	110,477	110,821
女兒	83,544	83,734	83,942	84,238	84,599	85,060
養子	1,447	1,431	1,420	1,408	1,390	1,382
養女	1,284	1,244	1,222	1,217	1,199	1,167
媳婦	10,790	11,053	11,882	12,478	12,987	13,346
女婿	1,224	1,326	1,507	1,624	1,761	1,934
孫子女	40,360	42,154	44,259	46,138	48,254	50,757
外孫子女	562	533	490	454	429	415
孫子女配偶及親屬	220	239	264	285	300	310
贅女婿	241	224	209	197	187	180
贅夫	1,102	1,054	986	925	866	813
家屬	8,018	8,241	8,378	8,449	8,675	8,928
寄居	26,398	27,816	28,625	29,134	29,646	30,324
領養親屬	70	66	58	51	48	47
其他	212	288	169	143	138	145
無資料	13,393	16,129	19,388	21,640	23,699	25,011

和戶長關係_標準化代碼	和戶長關係相對人數(%, 年平均)					
	2013	2014	2015	2016	2017	2018
全部人數	100.0	100.0	100.0	100.0	100.0	100.0
祖父母	0.0	0.0	0.1	0.1	0.0	0.0
外祖父母	0.0	0.0	0.0	0.0	0.0	0.0
父母	1.5	1.6	1.6	1.6	1.6	1.6
父母親屬	0.3	0.3	0.3	0.3	0.3	0.4
養父母	0.0	0.0	0.0	0.0	0.0	0.0
戶長	26.6	25.9	25.2	24.7	24.3	23.8
手足/手足配偶	3.5	3.5	3.5	3.5	3.5	3.6
手足/手足配偶親屬	2.2	2.3	2.3	2.3	2.4	2.5
配偶	9.0	8.9	8.9	8.9	8.8	8.7
配偶親屬	0.6	0.6	0.6	0.6	0.6	0.6
兒子	20.6	20.4	20.2	20.0	19.8	19.7
女兒	15.8	15.6	15.4	15.3	15.2	15.1
養子	0.3	0.3	0.3	0.3	0.2	0.2
養女	0.2	0.2	0.2	0.2	0.2	0.2
媳婦	2.0	2.1	2.2	2.3	2.3	2.4
女婿	0.2	0.2	0.3	0.3	0.3	0.3
孫子女	7.6	7.9	8.1	8.4	8.7	9.0
外孫子女	0.1	0.1	0.1	0.1	0.1	0.1
孫子女配偶及親屬	0.0	0.0	0.0	0.1	0.1	0.1
贅女婿	0.0	0.0	0.0	0.0	0.0	0.0
贅夫	0.2	0.2	0.2	0.2	0.2	0.1
家屬	1.5	1.5	1.5	1.5	1.6	1.6
寄居	5.0	5.2	5.3	5.3	5.3	5.4
領養親屬	0.0	0.0	0.0	0.0	0.0	0.0
其他	0.0	0.1	0.0	0.0	0.0	0.0
無資料	2.5	3.0	3.6	3.9	4.3	4.4

3. Building, Enriching, and Integrating Data

- **Building indigenous knowledge & linking to the real world, e.g., interview, survey, focus group, participation observation, etc.**





4. What's Next

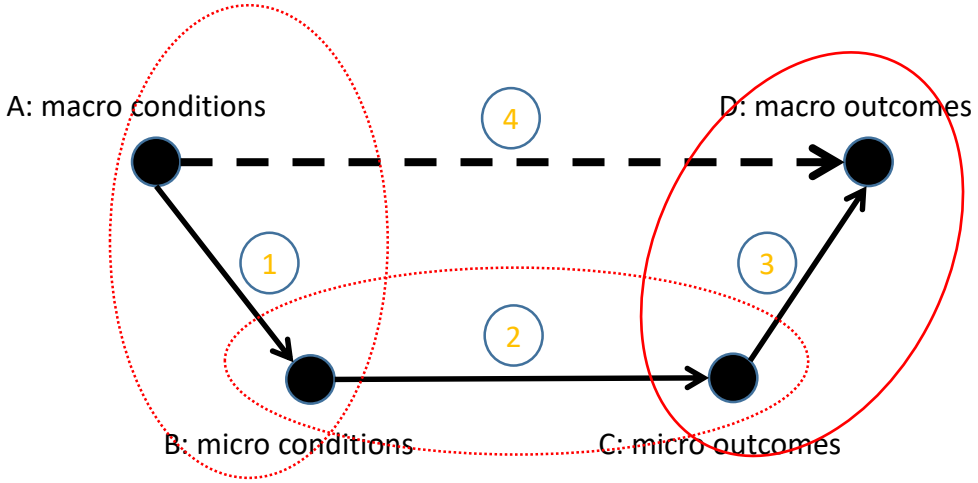
■ Building Micro-to-macro links & Model Training

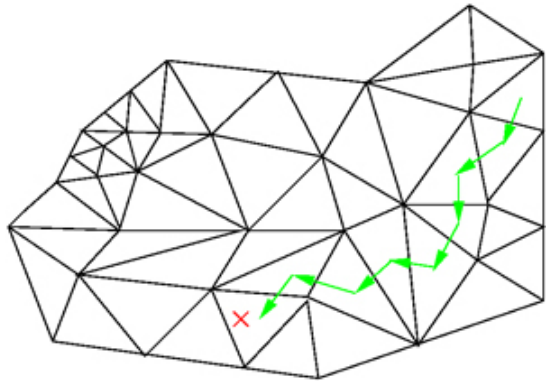
✓ Utilizing a more **generalized micro-to-macro** theoretical framework of transformation to explore

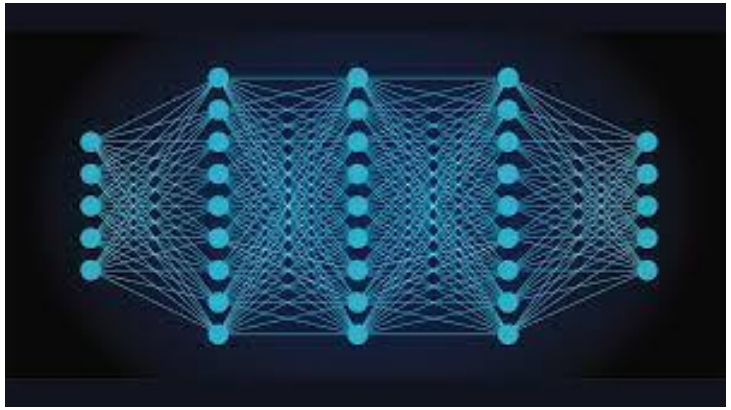
1) mechanism of system formation
2) evolution a complex system

✓ Solving **computing issues of optimization**, when

(1) volume and complexity of data grow very fast
(2) temporal length goes shorter and shorter
(3) spatial unit becomes smaller and smaller







4. What's Next

■ Building strong international research collaboration network

e.g., Academic conferences, congress, & keynote in 2024/09~2025/10

IGC 2024, Dublin

PANZ 2025, Wellington

IPC 2025, Brisbane

IDW 2025, Brisbane

4. What's Next

■ Building strong international research collaboration network

e.g., Academic visits in 2024/09~2025/10

4. What's Next

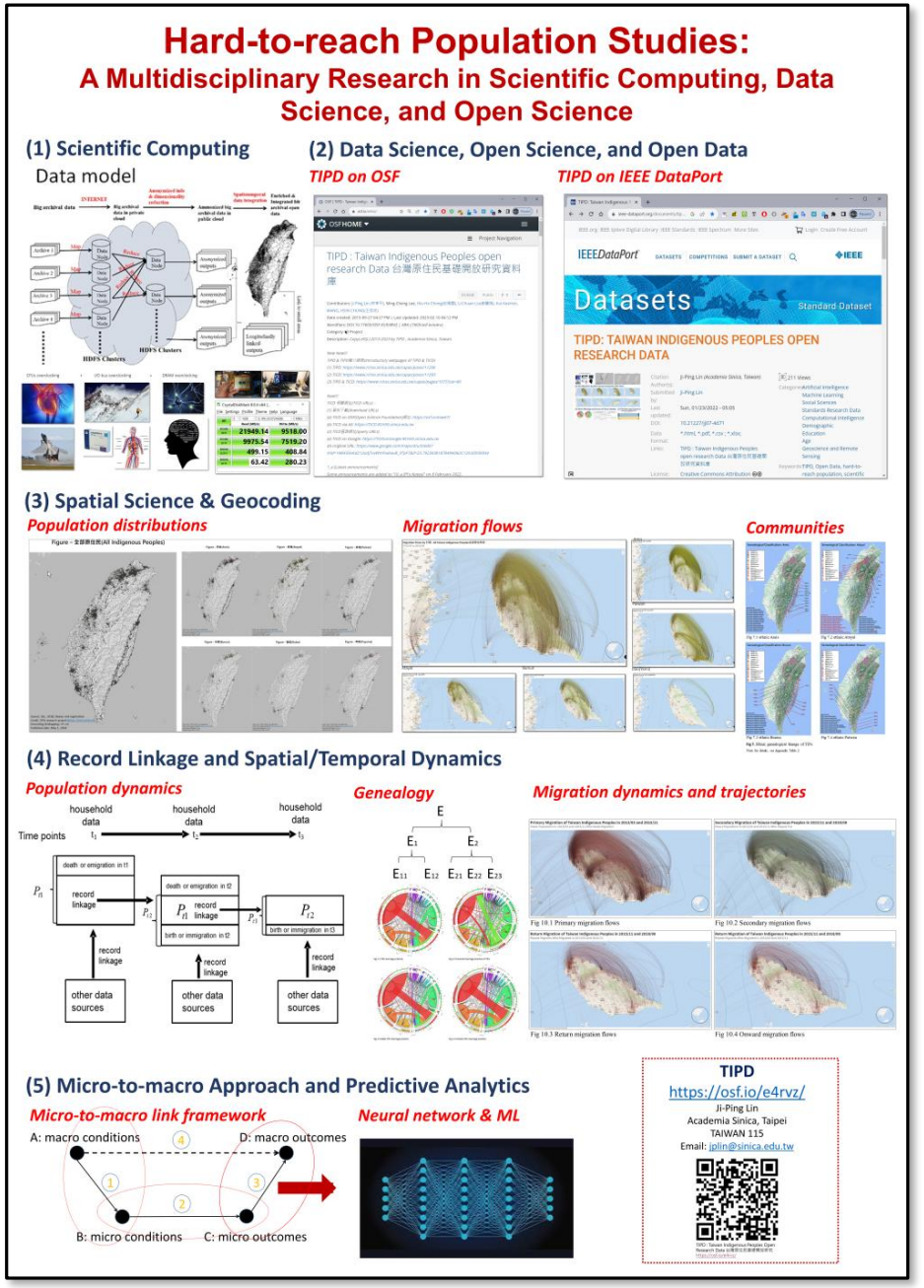
■ 技術移轉: in addition to open data, we need to build source code repositories of TIPD on,

e.g. GitHub

+ GitHub

5. Concluding Remarks

- Reorganizing raw data as open data to overcome legal & ethic issues boosts academic & crowd sourcing (civil) research, e.g., Taiwan indigenous peoples study and international cooperation.
- To allow us to enrich data through the process of data integration methodology, making longitudinally linked administrative data less expensive and more efficient, e.g., population dynamics data, birth & data & migration processes…
- To allow us to do what was not able to do before, e.g., individual genealogy, identity, ethnic marriage pattern



Thanks for your attentions!

… “Salute to the Angels”