

Development of Research Data Repository 研究資料典藏庫的發展

December 14, 2017

TESRI

莊庭瑞 Tyng-Ruey Chuang
和 with

黃韋菁 Andrea Wei-Ching Huang

李承鑫 Cheng-Jen Lee

許煌鑫 Huang-Sin Syu

Institute of Information Science
Academia Sinica



Outline

- Collaborative Research
- Research Data Management
- data.depositar.io
 - Retooling a Research Data Repository

Collaborative Research

- Collaboration is the process of two or more people or organizations working together to realize or achieve something successfully.
 - *Wikipedia*
- For collaborative research, we should make
 - the research project and
 - the research dataopen to project members and even to others.

Openness

- Libre
 - can be used by people
- Digital
 - can be used by machines and put online
- Raw
 - can be modified and re-purposed
- Common (format & vocabulary)
 - can be exchanged and interlinked
- Transparent
 - (the process) can be fixed; meta-level

Openness Benefits Research

- Help disseminate research findings.
- Help reproduce and re-purpose research results.
- Help encourage research collaborations.

Free Software and Openness



ckan



GitLab

Research Data Management (RDM)

<http://www.dcc.ac.uk/resources/how-guides/how-develop-rdm-services>



FAIR Data

- **Findable:** Have sufficiently rich metadata and a unique and persistent identifier.
- **Accessible:** Retrievable by humans and machines through a standard protocol; open and free by default; authentication and authorization where necessary.
- **Interoperable:** Metadata use a “formal, accessible, shared, and broadly applicable language for knowledge representation”.
- **Reusable:** Metadata provide rich and accurate information; clear usage license; detailed provenance.

A Research Data Repository

- Designed for interdisciplinary regional study
 - A collaboration between Academia Sinica (IIS and GIS Center) and NCKU in the last few years
- Built with CKAN
 - CKAN: software from OKI for “open data”
 - Adapt and extend CKAN for [research data](#) deposit, publishing, preview, and reuse
 - Hosted at Academia Sinica
 - Soon open to the research community

Search and Data Discovery

- free-text search
- multiple filters
- spatial and temporal constraints

The screenshot shows a web application interface for "跨領域區域研究資料集" (Cross-domain Regional Research Data Collection). The interface includes a navigation bar with "Datasets", "Organizations", "Groups", and "Help" links, along with a search bar. The main content area is titled "Datasets" and features a "Filter by location" section with a map of Taiwan. The map shows various regions, with a red hatched area labeled "R-18 區" (R-18 Area) highlighted. Below the map, there is a search bar for datasets, a "Cancel" button, and a list of 51 datasets found, ordered by relevance. The first two datasets listed are:

- 昭和年間灌溉地給水區分圖：麻豆
1929年麻豆灌溉給水區分圖
TIF
- 昭和年間灌溉地給水區分圖：西港
1929年西港灌溉給水區分圖

Dataset Visualization

臺南市里界圖

Manage

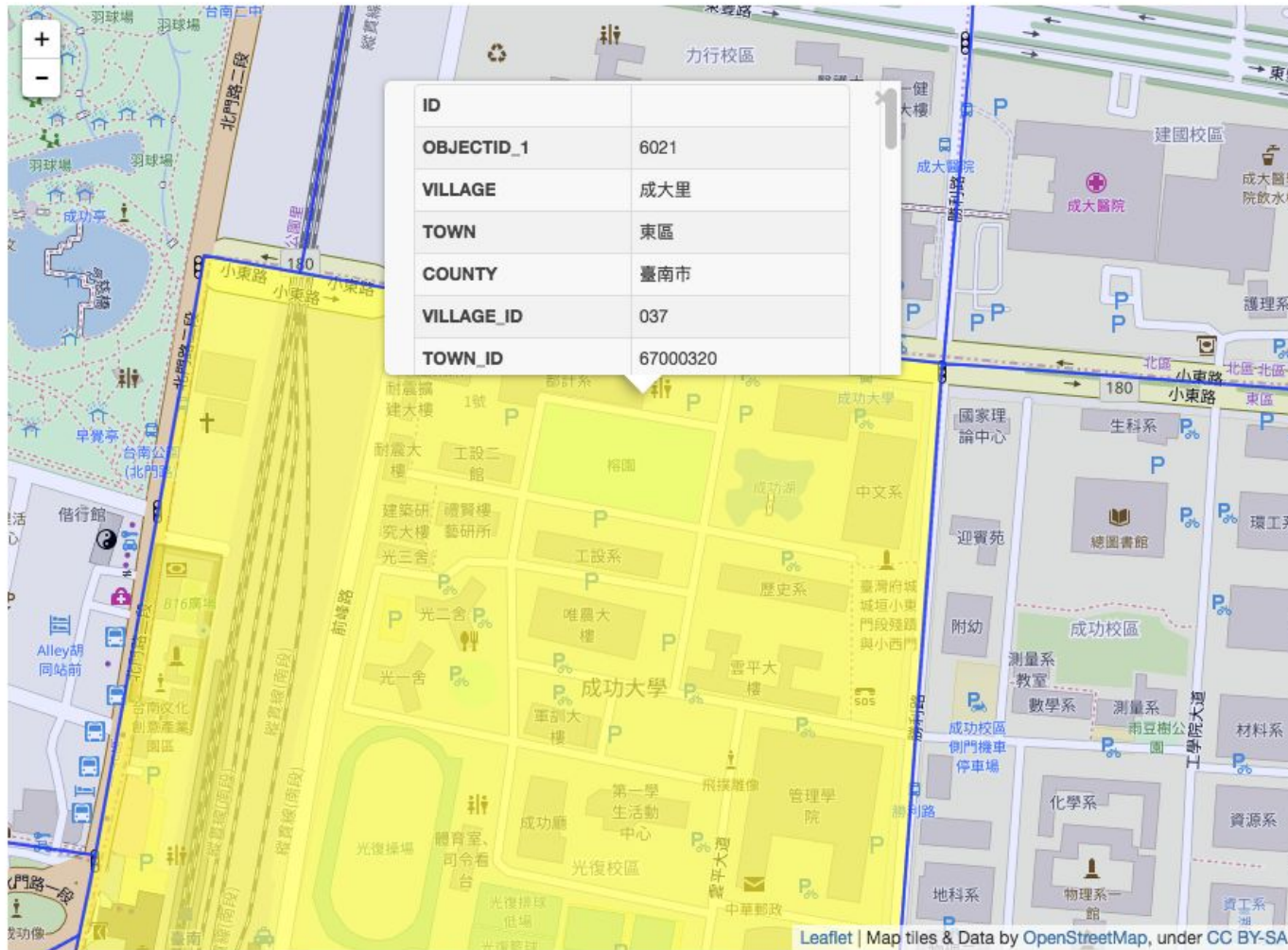
Go to resource

URL: <https://data.depositar.io/dataset/76a232e5-a9c1-4dde-8d4c-6b58629f33ca/resource/c2bbe543-dc9d-41c2-82d5-d16a93b87835/do...>

屬性包含：「VILLAGE(里名)」、「TOWN(區名)」、「COUNTY(縣市名)」

SHP

Embed



Metadata

- designed for cross-disciplinary research
- emphasizing spatial and temporal metadata authoring

Language:

Temporal Information (Time Period of Dataset)

Time Period Shortcut:

i This shortcut provides some historical periods for filling temporal information of the dataset.

Temporal Resolution:

Start Time:

i Depend on the temporal resolution setting.
Acceptable formats: 'YYYY', 'YYYY-MM', or 'YYYY-MM-DD'.

End Time:

i Depend on the temporal resolution setting.
Acceptable formats: 'YYYY', 'YYYY-MM', or 'YYYY-MM-DD'.

Spatial Fields



Spatial:

```
{ "type": "Polygon", "coordinates": [[[120.05825042724611, 22.969773432534776], [120.05825042724611, 23.09678621760499], [120.21068572998048, 23.09678621760499], [120.21068572998048, 22.969773432534776], [120.05825042724611, 22.969773432534776]]]] }
```

This column is used for geo-indexing the dataset. Click for details.

Place Names & Maps - A Showcase

<https://data.depositar.io/dataset/place-names-in-west-central-district-of-tainan>

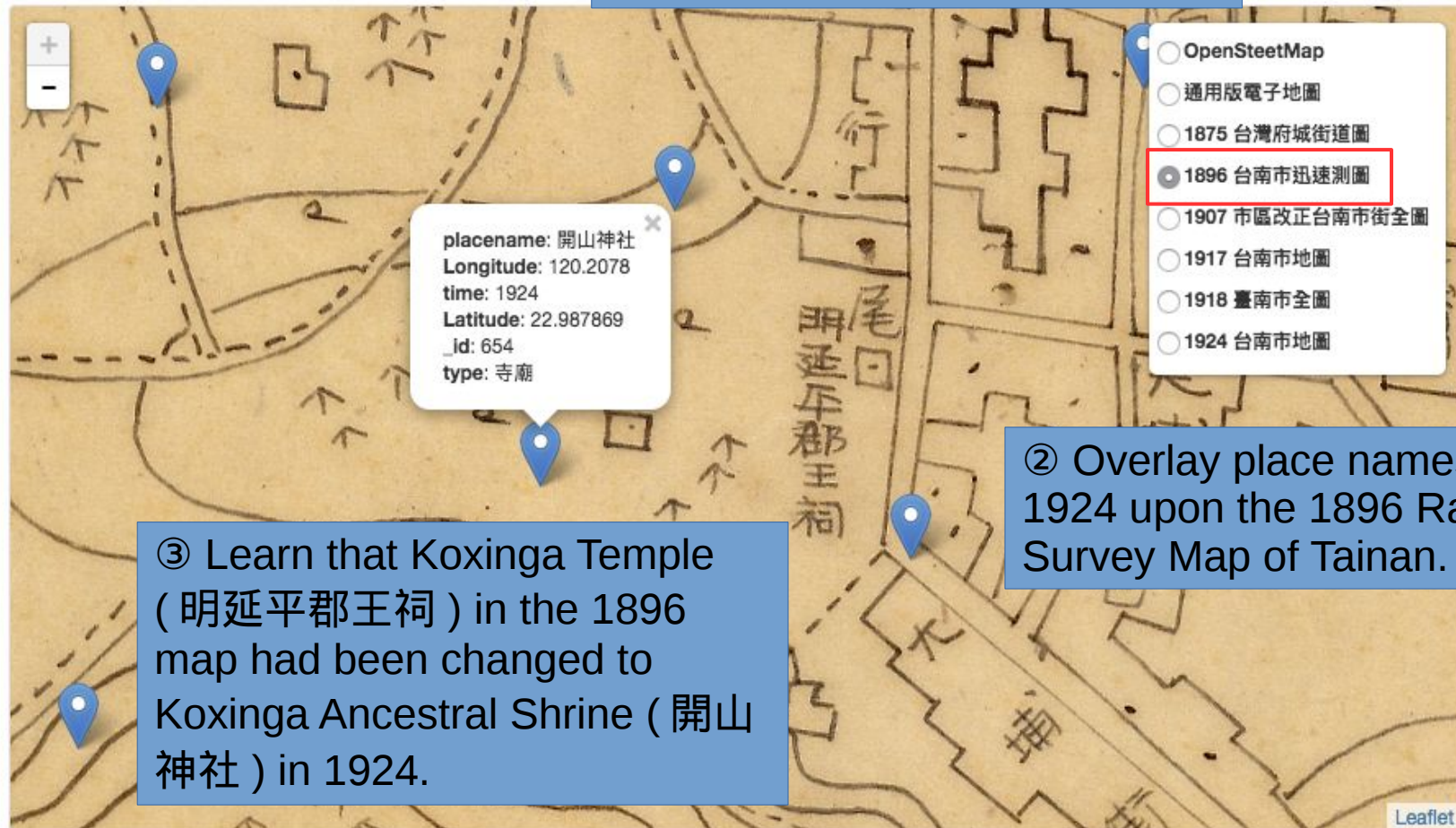
Source: Place Names in West Central District of Tainan

All 1875 1896 1907 1917 1924 資料瀏覽器

① Place names extracted from a map of Tainan in 1924 (displayed as blue place marks).

Embed

Add Filter



② Overlay place names from 1924 upon the 1896 Rapid Survey Map of Tainan.

③ Learn that Koxinga Temple (明延平郡王祠) in the 1896 map had been changed to Koxinga Ancestral Shrine (開山神社) in 1924.

Retooling a Research Data Repository

- From the “Taijiang Inland Sea” Project's Research Data Repository
 - taijiang.tw
- To a general-purpose research data repository
 - data.depositar.io
- With many adjustments and enhancements
 - generalized and multilingual metadata
 - wikidata-powered keywords
 - default values
 - latest CKAN features

Generalized and Multilingual Metadata

- One set of simple metadata for all datasets; three categories:
 - **Basic metadata:** title, description, data type, ...
 - **Descriptive metadata:** language, temporal, and spatial information, keywords, ...
 - **Management metadata:** license, author, created time, organization, maintainer, ...

Multilingual metadata

▼ Data Type
Books (455)
Vector (144)
Statistics (103)
Pictures (Spatial) (34)
Pictures (Non spatial) (25)
▼ Language
Chinese (730)
Dutch (22)
English (8)
Spanish (Castilian) (1)
▼ Historical Material
Local chronicles in Qing dynasty (36)
Maps (20)
Japan officials (19)
Newspapers (16)
Inscriptions (9)
Taiwanese Governor Office Files (9)
Dutch Formosa (5)
Officials in postwar period (1)

▼ 資料類型
文獻書籍 (455)
向量資料 (144)
統計資料 (103)
圖像 (空間類) (34)
圖像 (非空間類) (25)
▼ 語言
華語/中文 (730)
荷蘭語 (22)
英語 (8)
西班牙語 (1)
▼ 使用史料
清代方志 (36)
地圖 (20)
日治出版品與官方資料 (19)
報紙 (16)
碑文 (9)
台灣總督府檔案 (公文類纂、府報、調查報告、職員錄、衛生法規) (9)
荷治時期檔案 (5)
戰後官方資料與鄉鎮志 (1)

Wikidata-powered Keywords

- Keywords
 - controlled vocabularies for tagging datasets
- Define the list(s) of keywords
 - a never-ending process ...
- Use [Wikidata](#) as the source for keywords
 - 37M+ entries
 - multilingual
 - semantic relations enable data inference
 - e.g. Tainan is part of Taiwan



Wikidata-powered Keywords

Search and select keywords from Wikidata:

Keywords: tainan|

ID	Label	Description
Q140631	Tainan	city in southern Taiwan
Q477601	Tainan County	county of Taiwan
Q702035	Tainan Airport	airport

* License:

Keywords (as Wikidata IDs) are stored.

Viewed in English: **Keywords**

Tainan

Viewed in Chinese: **關鍵字**

臺南市

Default Value and Auto-completion

- ✓ A checkbox to open the dataset to organization members only (default is to open to all).
- ✓ Auto-completion of management metadata (e.g. maintainer name and email from user account).
- ✓ Generate better URLs from dataset titles (e.g. when titles include Chinese characters).

Organization: 中研院人社中心GIS專題中心 Open for organization members only

* Title: 台南成功大學

* URL: 140.109.23.195:9999/en/dataset/

Use your account information to fill in maintainer's name and email

Maintainer: Cheng-Jen Lee

Maintainer Email: cjlee@iis.sinica.edu.tw

Latest CKAN Features

- Private datasets (available only to members of an organization) are now included in the search results (for those who have access).
- Speed improvement for dataset preview.

Thank you!

<http://data.depositar.io/>

data.contact@depositar.io