

Update on the Research Data Repository data.depositar.io

ISGC 2018
ECAI / ApSTi Workshop

March 18, 2018

Tyng-Ruey Chuang
Cheng-Jen Lee

Institute of Information Science
Academia Sinica



Outline

- Collaborative Research
- Research Data Management
- data.depositar.io
 - Retooling an open data repository for a research data repository
 - Update and current focus

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.

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.

Free Software for Serving Open Source/Data/Media



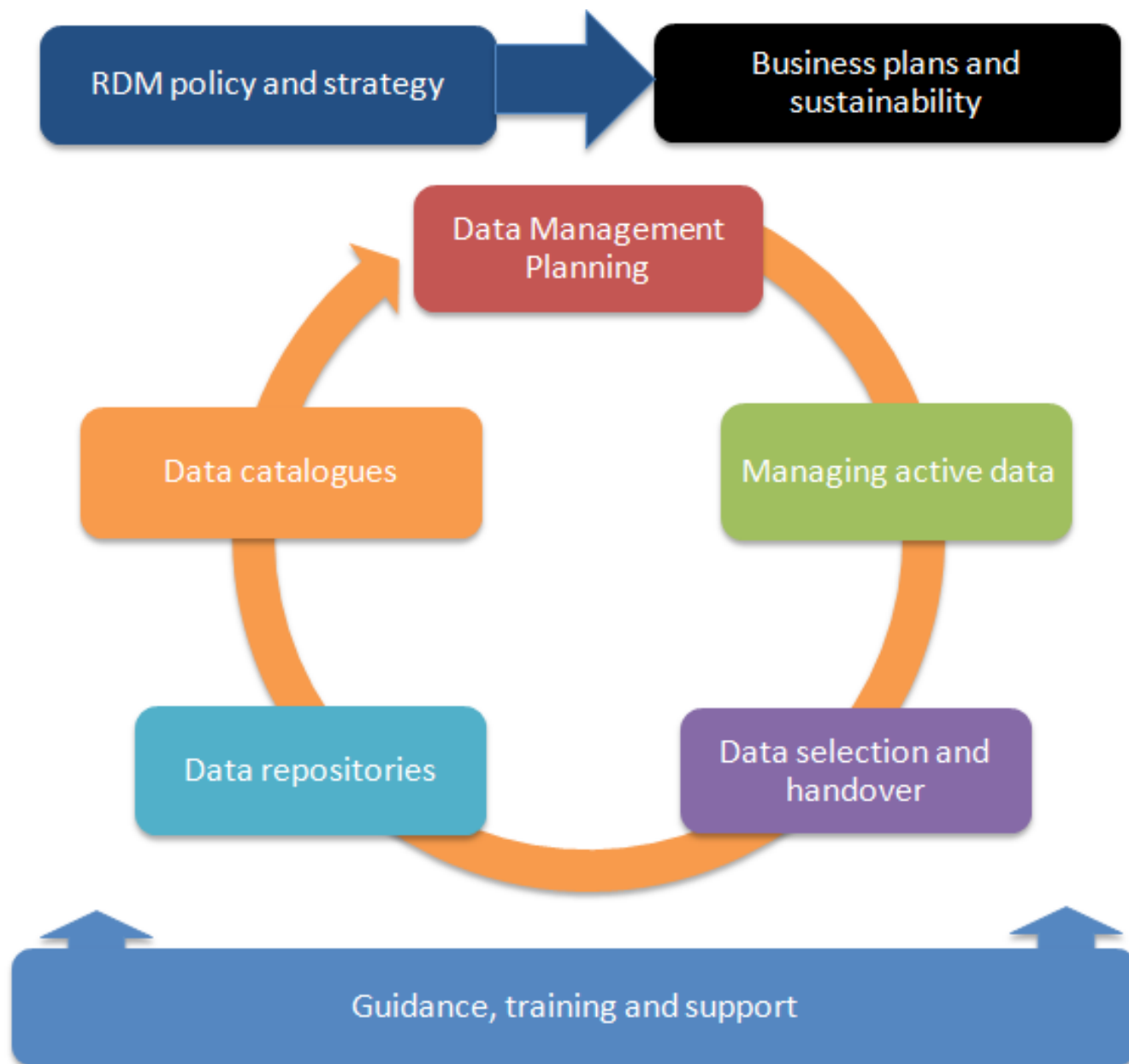
ckan



GitLab

Research Data Management (RDM)

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



A Research Data Repository

- Originally built for interdisciplinary regional study
 - A collaboration between Academia Sinica (IIS and GIS Center) and NCKU in the last few years
- Built upon CKAN
 - CKAN: software from OKI for “open data”
 - Adapt and extend CKAN for **research data** deposit, publishing, 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 displays a web interface for a cross-domain regional research data collection. The header includes the title '跨領域區域研究資料集' and navigation links for 'Datasets', 'Organizations', 'Groups', and 'Help'. A search bar is located in the top right corner.

The main content area features a map of Taiwan with a red hatched area labeled 'R-18 區' in the southern part of the island. The map is titled 'Filter by location' and includes a 'Clear' button. Below the map, there is a search bar for datasets and a list of 51 datasets found, ordered by relevance. The first two datasets 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年西港灌溉地給水區分圖

The interface also includes a 'Temporal Search' section with a 'Clear' button and a 'Use time period shortcut' dropdown, and an 'Organizations' section.

Dataset Visualization

臺南市里界圖

Manage

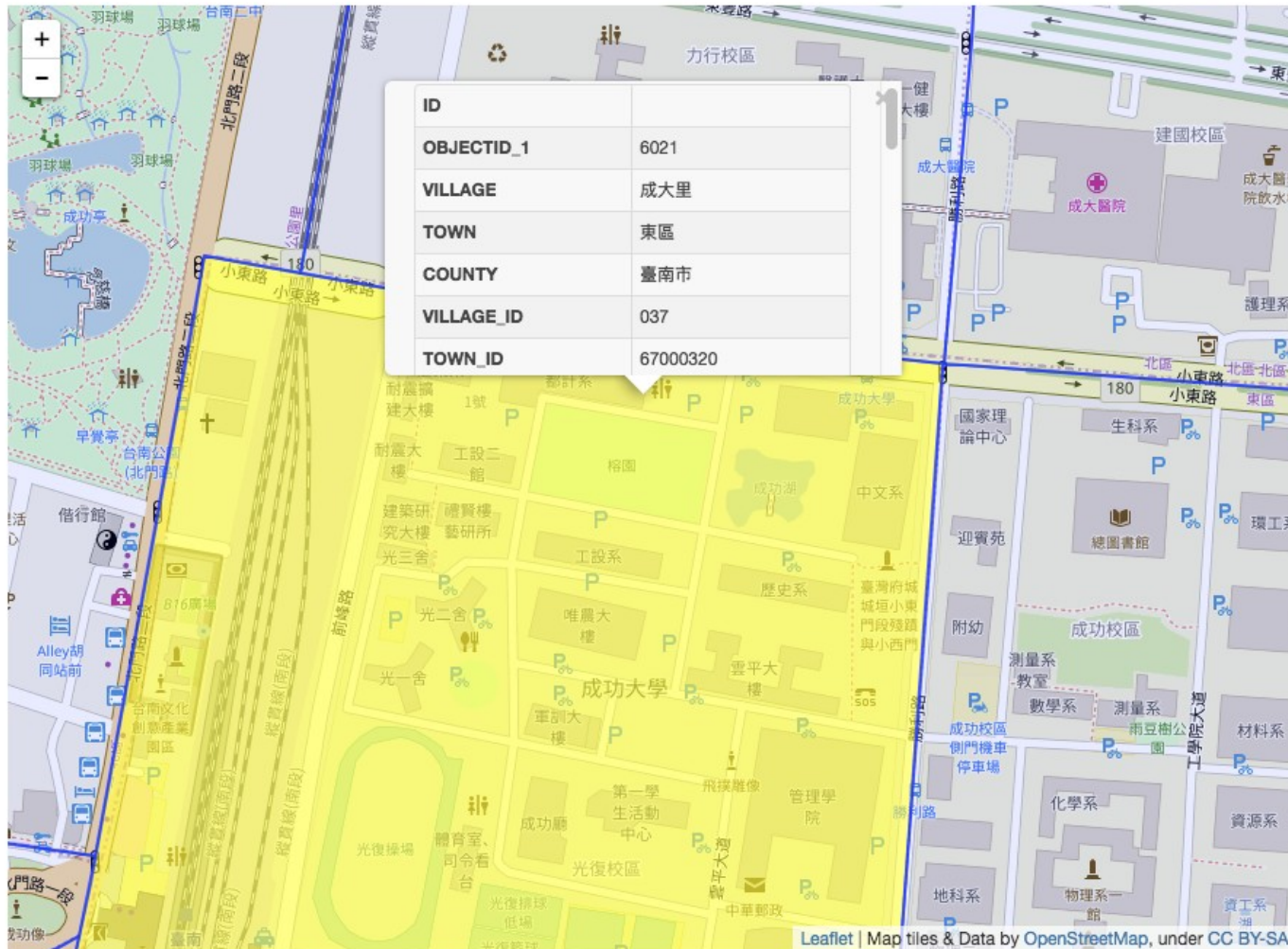
Go to resource

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>

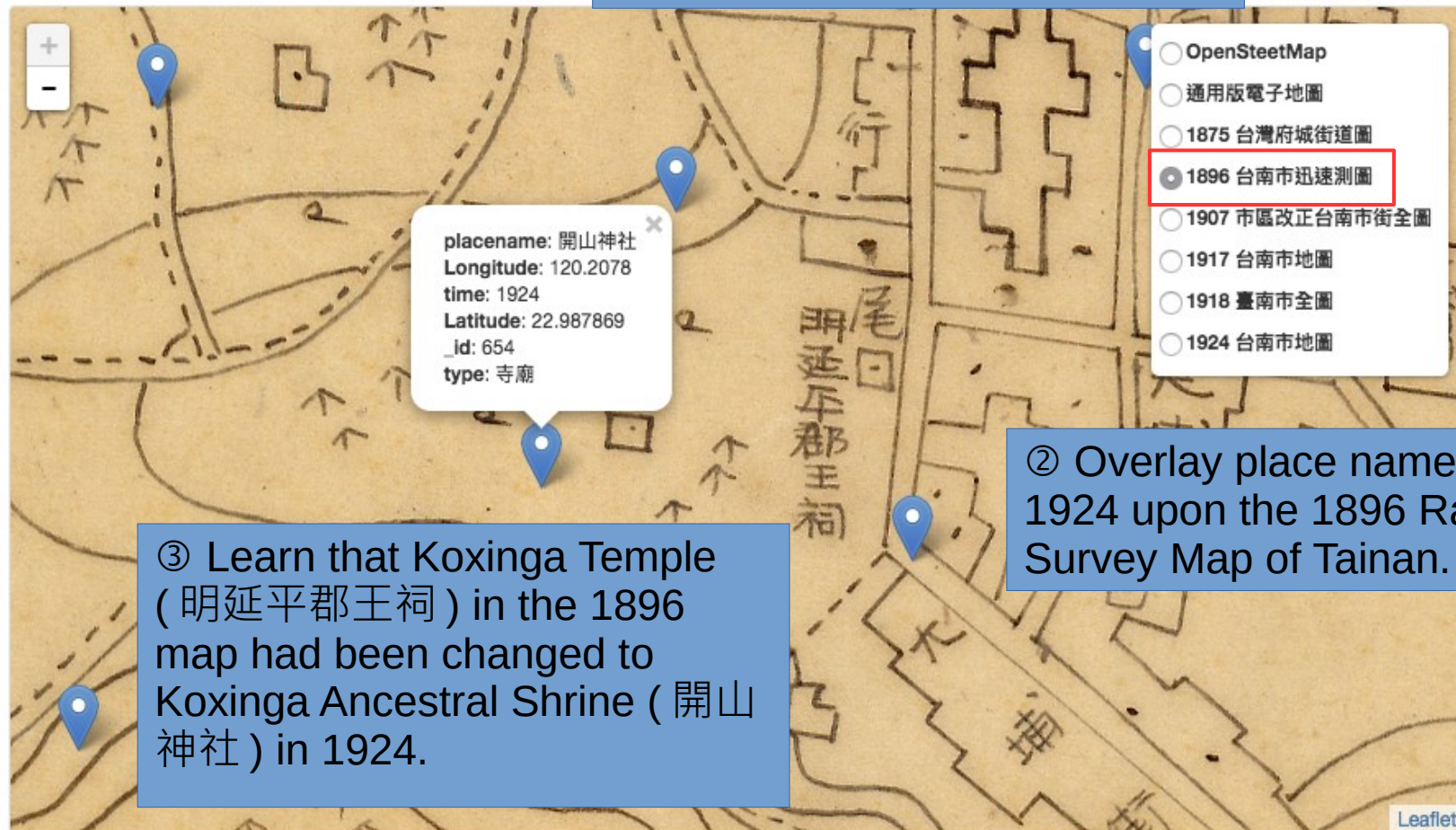
Source: Place Names in West Central District of Tainan

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 a special purpose research data repository
 - taijiang.tw for the Taijiang Inland Sea Project
- To a general-purpose research data repository
 - data.depositar.io
- With many adjustments and enhancements
 - generalized and multilingual metadata
 - wikidata-powered keywords
 - metadata 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 your own 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 Entries for Keywords

Search and select keywords:

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 limit 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

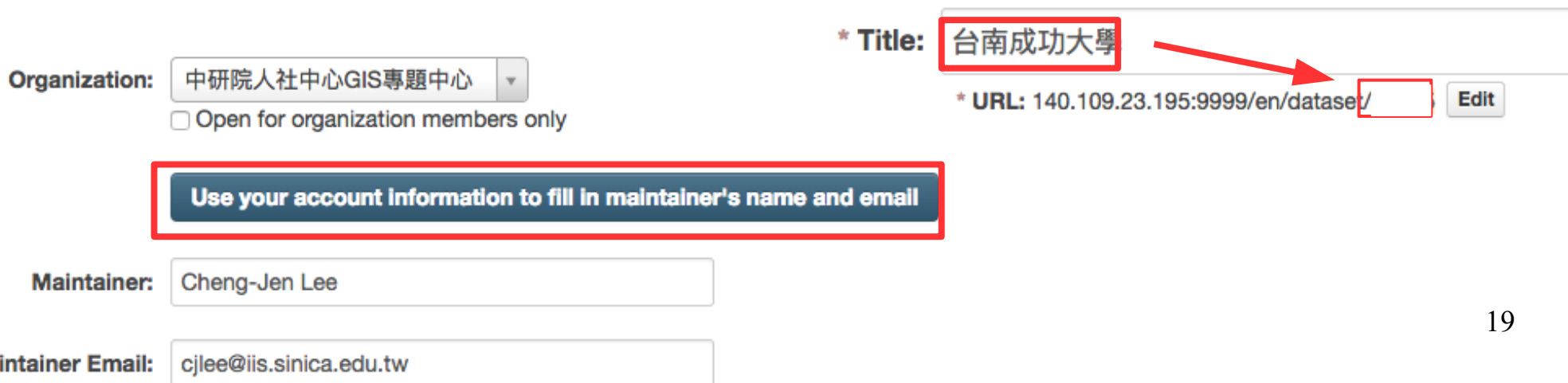
* Title: 台南成功大學

* URL: 140.109.23.195:9999/en/dataset/ Edit

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

Maintainer: Cheng-Jen Lee

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

A screenshot of a web form for dataset management. The form includes an 'Organization' dropdown menu with '中研院人社中心GIS專題中心' selected and an unchecked checkbox for 'Open for organization members only'. The 'Title' field contains '台南成功大學' (National Sun Yat-sen University), which is highlighted with a red box. A red arrow points from this box to the 'URL' field, which contains '140.109.23.195:9999/en/dataset/' followed by an empty input box, also highlighted with a red box. Below the form is a blue button with white text that reads 'Use your account information to fill in maintainer's name and email'. The 'Maintainer' field is filled with 'Cheng-Jen Lee' and the 'Maintainer Email' field is filled with 'cjlee@iis.sinica.edu.tw'.

Adding New 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.

Update and Current Focus

- User manual (Chinese and English languages)
- Interoperable base map services when previewing spatial datasets
- Outreach
 - to researchers who have been curating datasets
 - to students by being part of teaching in class
- Research Data Management Workshop
 - currently planned in 2018 September



搜尋資料

例如：環境



熱門標籤

台江國家公園

海圖

黑面琵鷺

最新消息

系統更新 / System Upgrade v6.0 (2017/11)

十一月 3, 2017

[English Below]

[跨領域區域研究資料集](#) 版本升級為 v6.0。

同時更新 CKAN 核心版本至 2.6.4。 (depositar-io/ckan@85bb3fe)

更新內容如下：

- ...

系統更新 / System Upgrade (2017/9)

九月 5, 2017

[English Below]

[跨領域區域研究資料集](#) 版本升級作業已於日前完成。

組織

測試區 / Sandbox

僅供測試用途。 For testing purposes only.

嘉南大圳研究群 / Chia-nan Project

本組織包含「嘉南平原水文、農業空間及聚落生活空間之百年變遷與運動關係之研究」之各子計畫單位內成員，各子計畫於計畫執行期間，可將尚未對外開放之資料發布於本組織內。

中研院人社中心GIS專題中心

隸屬中央研究院人文社會科學研究中心，從事地理資訊科學相關基礎研究與資源整合，建立大型學術資料庫，致力於推動地理資訊系統技術應用在人文社會科學研究上。

登入

跨領域區域研究資料集

資料集

組織

主題

支援

搜尋



🏠 / 組織

📘 組織是什麼？

您可以使用 CKAN 組織來建立、管理與發佈多個資料集。在組織中，每位使用者可擁有不同角色，而得以建立、編輯與發佈資料集。

搜尋組織



找到4個組織

排序依照：

根據名稱遞增排序



as-rchss-gis

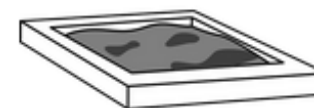
中研院人社中心GIS專題中心

隸屬中央研究院人文社會科學研究中心，從事地理資訊科學相關基礎研究與資源整合，建立大型學術資料庫，致力於推動地理資訊系統技術應用在人文社會科學研究上。

Chia-Nan

嘉南大圳研究群 / Chia-nan Project

本組織包含「嘉南平原水文、農業空間及聚落生活空間之百年變遷與運動關係之研究」之各子計畫單位內成員，各子計畫於計畫執行期間，可將尚未對外開放之資料發布於本組織內。



測試區 / Sandbox

僅供測試用途。 For testing purposes only.



台江內海地區跨領域研究群 / Taijiang Project

本組織包含「台江內海地區跨領域研究」計畫之各子計畫單

Thank you!

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

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

Please contact us at:

data.contact@depositar.io