

# Retooling An Open Data Repository for A Research Data Repository

International Data Week – SciDataCon 2018  
Gaborone, Botswana

Tyng-Ruey Chuang, Cheng-Jen Lee, and Chia-Hsun Wang

Institute of Information Science &  
Research Center for Information Technology Innovation  
Academia Sinica  
Taipei, Taiwan



# CKAN

- **C**omprehensive **K**nowledge **A**rchive **N**etwork
- Software for powering hubs of open datasets
- Developed by the Open Knowledge International; supported by an active contributor community
- Widely used to manage open government data portals, e.g. [catalog.data.gov](https://catalog.data.gov)
- Written in Python, released under AGPL 3.0
- The CKAN API exposes all core features to API clients
- DKAN – Drupal-based CMS based on CKAN

# *depositar* – on the shoulder of CKAN

- *depositar* – v. to deposit  
*depositario* – n. trustee, depository
- [data.depositar.io](http://data.depositar.io) – built on top of CKAN
- New in *depositar*
  - **open to academic user registration**
  - support spatiotemporal description and query
  - more data preview and base map
  - keywords sourced from [Wikidata](https://www.wikidata.org/)
  - simplified metadata authoring

# publisher vs. depository

- CKAN – software for publishing open data
- *depositar* – a depository for research data
- publisher – engage in “acquisition, copy editing, production, (e-)printing, marketing and distribution”
- depository – “a place where something is deposited, as for storage, safekeeping or preservation”
- one-to-many vs. many-to-one; push vs. pull
- for polished products vs. for things still in progress

*depositar*

features and show cases

# Data preview

## 臺南市里界圖

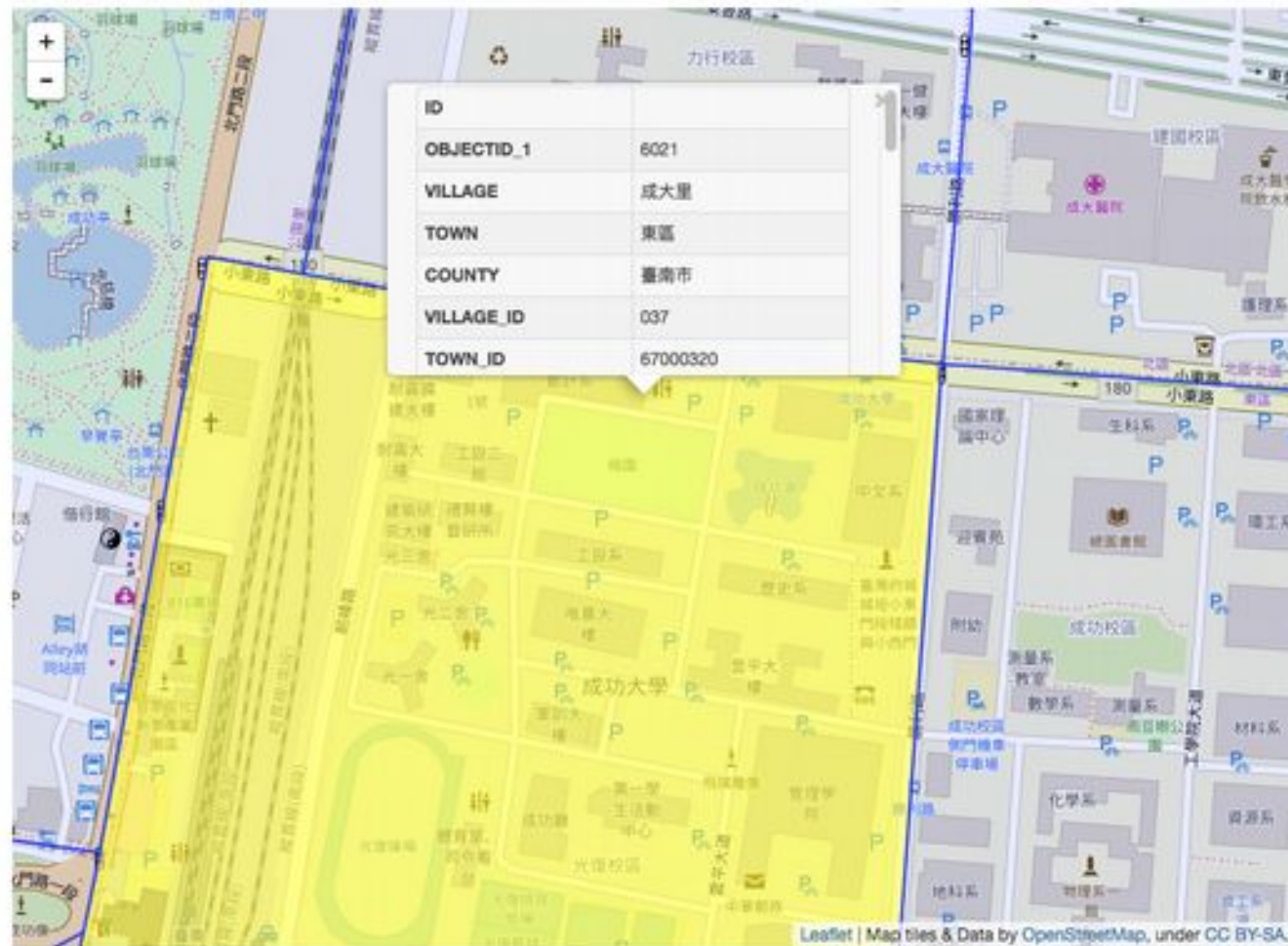
[Manage](#)[Go to resource](#)

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 for datasets

- designed for cross-disciplinary research
- emphasizing spatiotemporal description
- nice user interface for metadata authoring

Language:

**Temporal Information (Time Period of Dataset)**


Time Period Shortcut:

Temporal Resolution:

Start Time:

End Time:

**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.

# Bilingual metadata & user interface

▼ 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)



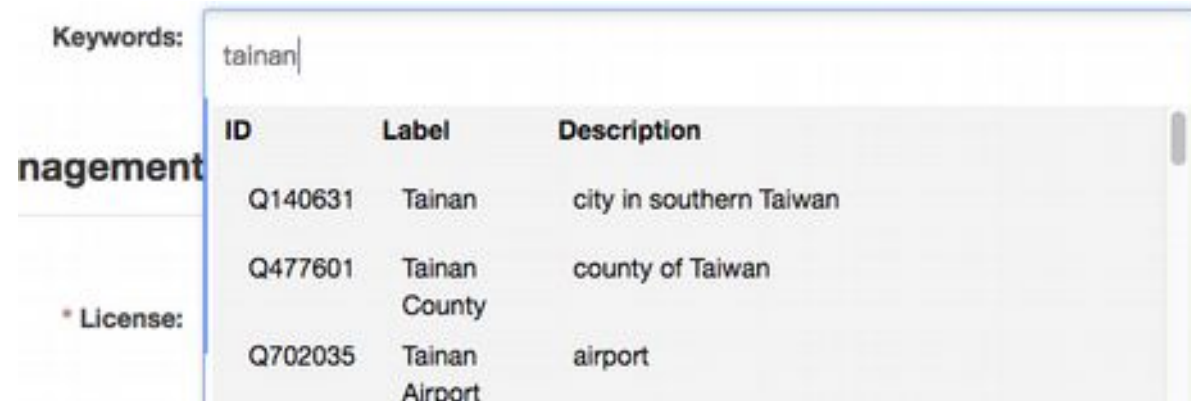
# Wikidata-powered keywords

- Keywords
  - controlled vocabularies for tagging datasets
  - DIY vocabulary design: a never-ending story ...
- Use [Wikidata](#) as the source for keywords
  - 47M+ entries
  - multilingual
  - semantic relations enable data inference
    - e.g. Gaborone is part of Botswana



# Wikidata entries as keywords

Search and select keywords:

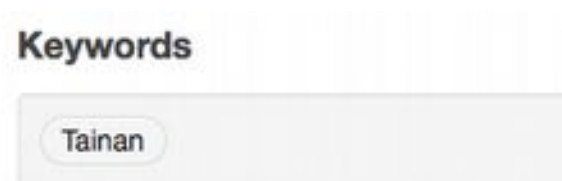


The screenshot shows a search interface with a text input field containing 'tainan'. Below the input field is a table of search results. The table has three columns: ID, Label, and Description. The results are as follows:

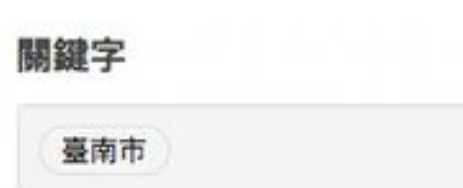
ID	Label	Description
Q140631	Tainan	city in southern Taiwan
Q477601	Tainan County	county of Taiwan
Q702035	Tainan Airport	airport

Keywords (as Wikidata IDs) are stored.

Viewed in English:



Viewed in Chinese:



# Roadkill Observations – A Showcase

<https://data.depositar.io/dataset/2016>

跨領域區域研究資料集

Datasets Organizations Groups Help

Log in

Dataset / 路殺社範例資料集-2016

路殺社範例資料集-2016

0 Followers

0 Social

Google+ Twitter Facebook

License

License Not Specified

Dataset extent

Other Access

The information on this page (the dataset metadata) is also available in these formats:

JSON

via the CKAN API

路殺社範例資料集-2016

Dataset Groups Activity Stream History

路殺社範例資料集-2016

透過觀察并建立路殺事件數據集來瞭解路殺事件發生的現象，並以結構化方式記錄相關資訊

Data and Resources

路殺社範例資料集

Tags

roadkill

Basic Information

Data Type Statistics

Language Chinese

Temporal Information (Time Period of Dataset)

Start Time 2016

Spatial Fields

Spatial show more

X.min 118.35888882608885

X.max 122.68350369547873

Y.min 21.45913969982141

Y.max 26.44324802769591

Management Information

Author 路殺社

About 跨領域區域研究資料集

Contact us: data.contact AT depositar.io

CKAN API | Site Statistics | Site Status | Documentation

Powered by ckan

Source code available on GitHub

Language: English

跨領域區域研究資料集

Datasets Organizations Groups Help

Log in

Dataset / 路殺社範例資料集-2016 / 路殺社範例資料集

路殺社範例資料集

https://data.depositar.io/datasets/16152-154-456-865-4195d44757/resource/1138d0f109-494-6d7c-476d033d57c0om

From the dataset abstract

透過觀察并建立路殺事件數據集來瞭解路殺事件發生的現象，並以結構化方式記錄相關資訊

Source: 路殺社範例資料集 (2016)

Embed

Map

Grid Graph Map 3275 records

Grid

ID	lat	lon	科名	科名(中)	物種名	物種名(中)	紀錄時間	地點	數量	備註
1	24.932	121.838	Suburbae	雜草科	Pennisetum	五節草	2016-01	宜蘭縣	121.838	24.794423
2	24.934	121.838	Suburbae	雜草科	Turkicop	五節草	2016-01	宜蘭縣	121.838	24.794423
3	24.938	121.838	Suburbae	雜草科	Randallia	五節草	2016-01	宜蘭縣	121.838	24.794423
4	24.944	121.838	Suburbae	雜草科	Magnolia	五節草	2016-01	宜蘭縣	121.838	24.794423
5	24.971	121.838	Suburbae	雜草科	Lycopersicon	五節草	2016-01	宜蘭縣	121.838	24.794423
6	24.972	121.838	Suburbae	雜草科	Turkicop	五節草	2016-01	宜蘭縣	121.838	24.794423
7	24.974	121.838	Suburbae	雜草科	Suburbae	五節草	2016-01	宜蘭縣	121.838	24.794423
8	24.976	121.838	Suburbae	雜草科	Suburbae	五節草	2016-01	宜蘭縣	121.838	24.794423
9	24.978	121.838	Suburbae	雜草科	Suburbae	五節草	2016-01	宜蘭縣	121.838	24.794423
10	24.980	121.838	Suburbae	雜草科	Suburbae	五節草	2016-01	宜蘭縣	121.838	24.794423
11	24.982	121.838	Suburbae	雜草科	Suburbae	五節草	2016-01	宜蘭縣	121.838	24.794423
12	24.984	121.838	Suburbae	雜草科	Suburbae	五節草	2016-01	宜蘭縣	121.838	24.794423
13	24.986	121.838	Suburbae	雜草科	Suburbae	五節草	2016-01	宜蘭縣	121.838	24.794423
14	24.988	121.838	Suburbae	雜草科	Suburbae	五節草	2016-01	宜蘭縣	121.838	24.794423
15	24.990	121.838	Suburbae	雜草科	Suburbae	五節草	2016-01	宜蘭縣	121.838	24.794423
16	24.992	121.838	Suburbae	雜草科	Suburbae	五節草	2016-01	宜蘭縣	121.838	24.794423
17	24.994	121.838	Suburbae	雜草科	Suburbae	五節草	2016-01	宜蘭縣	121.838	24.794423
18	24.996	121.838	Suburbae	雜草科	Suburbae	五節草	2016-01	宜蘭縣	121.838	24.794423
19	24.998	121.838	Suburbae	雜草科	Suburbae	五節草	2016-01	宜蘭縣	121.838	24.794423
20	25.000	121.838	Suburbae	雜草科	Suburbae	五節草	2016-01	宜蘭縣	121.838	24.794423

Resource

Additional information

Last updated: June 4, 2016

Created: June 4, 2016

Format: CSV

Encoding: UTF-8

Coordinate Systems: EPSG:4326

License: License Not Specified

About 跨領域區域研究資料集

Contact us: data.contact AT depositar.io

CKAN API | Site Statistics | Site Status | Documentation

Powered by ckan

Source code available on GitHub

Language: English

# Roadkill Observations – A Showcase

<https://data.depositar.io/dataset/2016>

Add Filter

Grid	Graph	Map	3218 records		«	1	–	20	»	Q	Search data ...	Go »	Filters
_id	id	科名	科名 ( ...	物種名	物種名...	紀錄時間	地點	經度	緯度				
1	24933	Bufo	蟾蜍科	Rana lon...	長腳赤蛙	2016-01...	宜蘭縣員...	121.636...	24.754443				
2	24934	Muscica...	鶇科	Turdus p...	白腹鶇	2016-01...	基隆市中...	122.08166	25.62861				
3	24935	Muridae	鼠科	Bandicot...	鬼鼠	2016-01...	高雄市茄...	120.195...	22.894975				
4	24944	Talpidae	鼯鼠科	Mogera l...	台灣鼯鼠	2016-01...	台中市北...	120.788...	24.173821				
5	24971	Colubridae	黃頭蛇科	Lycodon...	白梅花蛇	2016-01...	新竹縣峨...	120.98417	24.698021				
6	24972	Viperidae	蝮蛇科	Trimeres...	赤尾青竹絲	2016-01...	苗栗縣頭...	120.974...	24.702631				
7	25004	Bufo	蟾蜍科	Bufo ban...	盤古蟾蜍	2016-01...	台北市士...	121.538...	25.127258				
8	25006	Potamidae	溪蟹科	Geothel...	黃綠澤蟹	2016-01...	高雄市大...	120.408...	22.725732				
9	25008	Laniidae	伯勞科	Lanius c...	紅尾伯勞	2016-01...	台南市柳...	120.349...	23.261861				
10	25009	Muridae	鼠科	Rattus n...	溝鼠	2016-01...	台中市潭...	120.724...	24.231545				
11	25023	Rallidae	秧雞科	Amauror...	白胸苦惡鳥	2016-01...	金門縣金...	118.412...	24.508996				
12	25025	Columbi...	鳩鴿科	Streptop...	珠頸斑鳩	2016-01...	金門縣金...	118.319...	24.471015				
13	25026	Bufo	蟾蜍科	Duttaphr...	黑眶蟾蜍	2016-01...	台北市文...	121.538...	25.008142				
14	25027	Columbi...	鳩鴿科	Columba...	岩鴿	2016-01...	新北市貢...	121.984...	24.99522				
15	25028	Rhacoph...	樹蛙科	Rhacoph...	台北樹蛙	2016-01...	基隆市暖...	121.734...	25.091321				
16	25029	Rhacoph...	樹蛙科	Rhacoph...	台北樹蛙	2016-01...	基隆市暖...	121.734...	25.091321				
17	25030	Rhacoph...	樹蛙科	Rhacoph...	台北樹蛙	2016-01...	基隆市暖...	121.734...	25.091321				
18	25032	Colubridae	黃頭蛇科	Ptyas dh...	過山刀	2016-01...	台中市太...	120.7971	24.10225				
19	25033	Strigidae	鴞科	Strigidae	鴞科	2016-01...	台北市文...	121.586...	24.982227				
20	25034	Sturnidae	椋鳥科	Acridoth...	家八哥	2016-01...	金門縣金...	118.413...	24.507908				

Grid

Graph

Map

406 records

«

1

–

406

»



Search data ...

Go »

Filters

☒ Latitude / Longitude fields☐ GeoJSON field

Latitude field

緯度

Longitude field

經度

Update

☒ Auto zoom to features☒ Cluster markers

## Filters

Add filter

科名 (中文) ×

蟾蜍科

Update



Grid

Graph

Map

406 records

«

1

-

406

»



Search data ...

Go »

Filters

☒ Latitude / Longitude fields☐ GeoJSON field

Latitude field

緯度

Longitude field

經度

Update

☒ Auto zoom to features☒ Cluster markers

## Filters

Add filter

科名 (中文) ×

蟾蜍科

Update

# Infrastructure from below

- Research data management is infrastructure work
  - basic but not sexy; mismatch in needs and resources
  - culture of collaboration; priority in coordination
  - need to be a community of practices
- Infrastructure with small pieces from below
  - common licenses, vocabularies, formats, protocols, etc.
  - reusable tools (e.g. CKAN) and references (Wikidata)
  - resource pooling: people, CPU, storage, bandwidth, etc.
  - engaging in communication: code, data, experience, etc.



Datasets

Topics

Projects

About

Help



deposit • discover •  
reuse

Learn More





# Partners

[Center for GIS, RCHSS, Academia Sinica](#)

[Electronic Cultural Atlas Initiative \(ECAI\)](#)

[Chia-nan Interdisciplinary Project](#)

[Collaborative ZuoJhen Project](#)

[Asian Soundscape Monitoring Network](#)

[Taiwan Roadkill Observation Network](#)

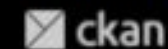
**d\_depositar**

Contact Us: [data.contact at depositar.io](mailto:data.contact@depositar.io)

About depositar

[CKAN API](#) | [Site Statistics](#) | [Site Status](#) | [Help](#)

Powered by



Source code available on [GitHub](#).

Visual Design & UI by

[Dualai Studio](#)

# WIKIPEDIA

## English

*The Free Encyclopedia*  
4 000 000+ articles

## 日本語

フリー百科事典  
815 000+ 記事

## Español

*La enciclopedia libre*  
900 000+ artículos

## Русский

Свободная энциклопедия  
875 000+ статей

## Italiano

*L'enciclopedia libera*  
939 000+ voci

## Português

*A enciclopédia livre*  
742 000+ artigos

## Deutsch

*Die freie Enzyklopädie*  
1 430 000+ Artikel

## Français

*L'encyclopédie libre*  
1 270 000+ articles

## Polski

*Wolna encyklopedia*  
908 000+ haseł



## 中文

自由的百科全書  
500 000+ 條目

Search · Suchen · Rechercher · Zoeken · Ricerca · Szukaj · Buscar · Поиск · 検索 ·  
Busca · Sök · 搜尋 · Tìm kiếm · Пошук · Cerca · Søk · Haku · Hledání · Keresés · 찾기 · Cari ·  
Ara · جستجو · Căutare · بحث · Hifadaf · Søg · Serçu · Претрага · Paleška · Polišči · Cari · חיפוש ·  
Търсене · Izdey · Bilatu · Suk · Bilnga · Traži · खोजे



contents 内容

工具 tools

users 用户

服務 services



**WIKIMEDIA**  
FOUNDATION



Thank You!

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

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

[data.contact@depositar.io](mailto:data.contact@depositar.io)

