

Retooling An Open Data Repository for A Research Data Repository

ECAI Workshop

In conjunction with PNC 2018 Annual Conference and Joint Meetings
Co-located with the Digital Heritage 2018 Conference

2018-10-27

San Francisco, CA, USA

Tyng-Ruey Chuang, Cheng-Jen Lee, and Chia-Hsun Wang
莊庭瑞、李承鑫、王家薰

Institute of Information Science
Academia Sinica, Taiwan
台灣 中央研究院 資訊科學研究所



Outline

- CKAN
- *depositar* – on the shoulder of CKAN
- Features and show cases
- Infrastructure from below

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
- 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 – 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

Search and data discovery

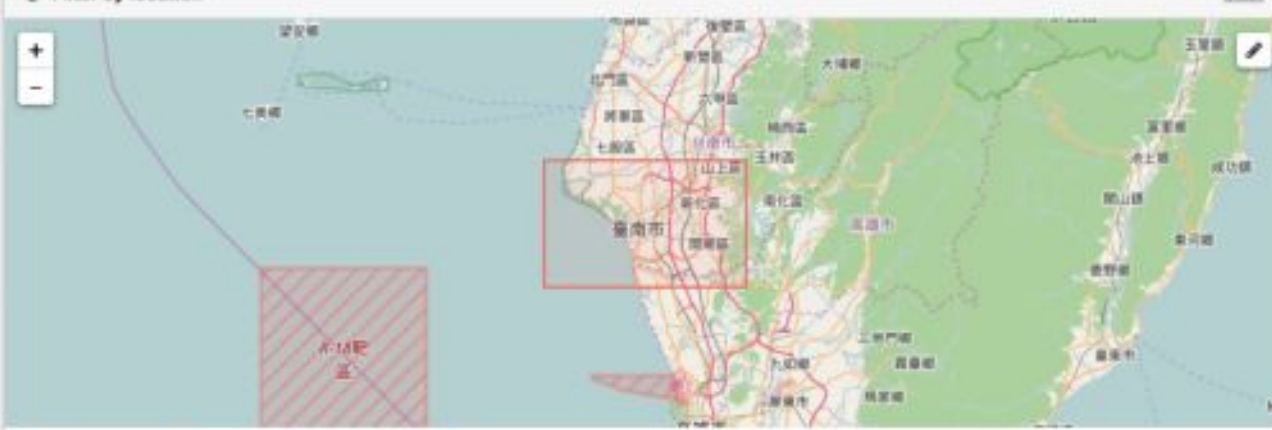
- full text search
- multiple filters
- spatial and temporal constraints

跨領域區域研究資料集

Datasets Organizations Groups Help Search

/ Datasets

Filter by location



Map tiles & Data by OpenStreetMap, under CC BY-SA

Cancel

Search datasets...

51 datasets found

Order by: Relevance

Temporal Search Clear

Or use time period shortcut

Organizations

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

Data preview

臺南市里界圖

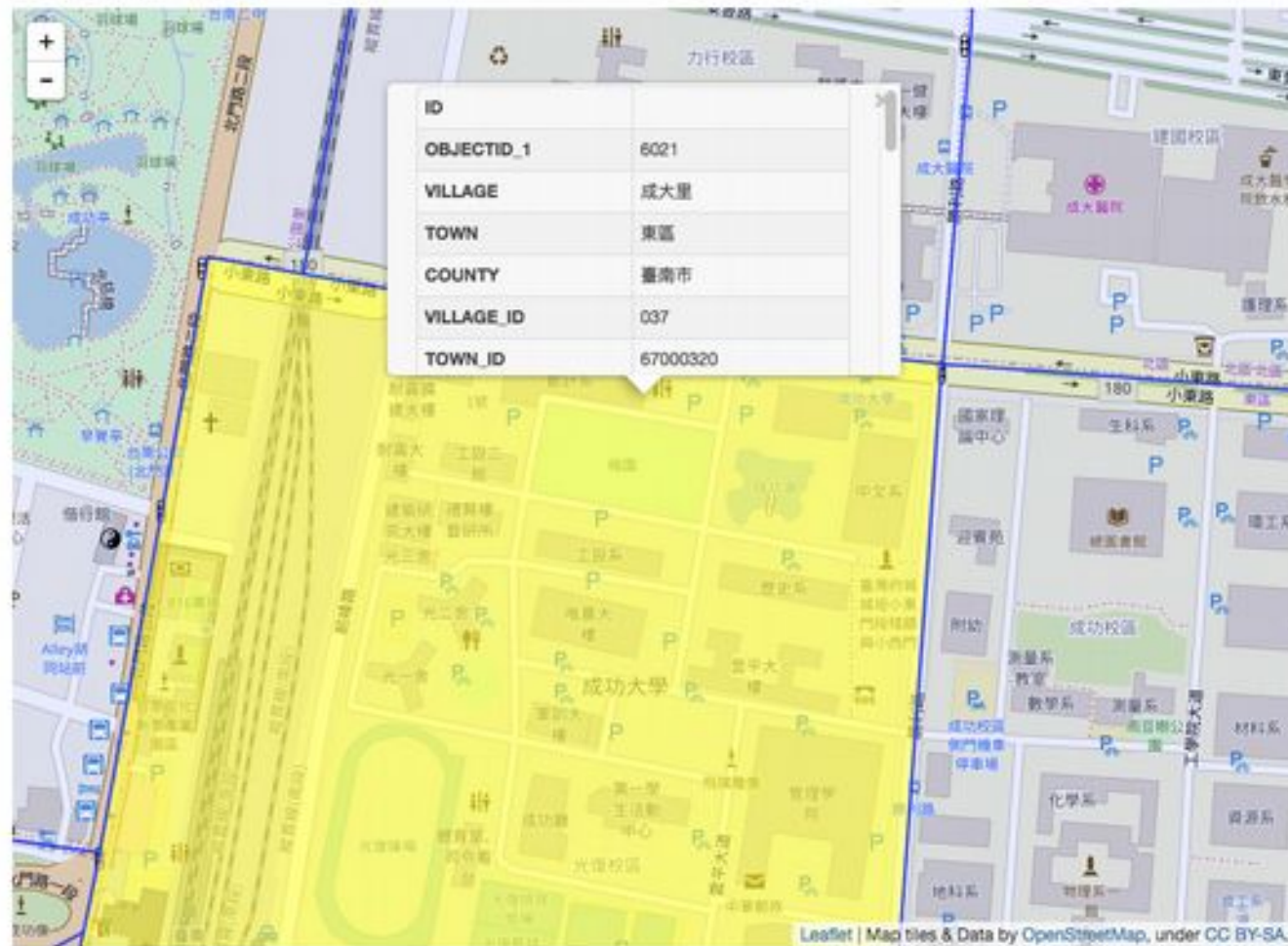
[Manage](#)[Go to resource](#)

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)

Simplified metadata authoring

- 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, ...

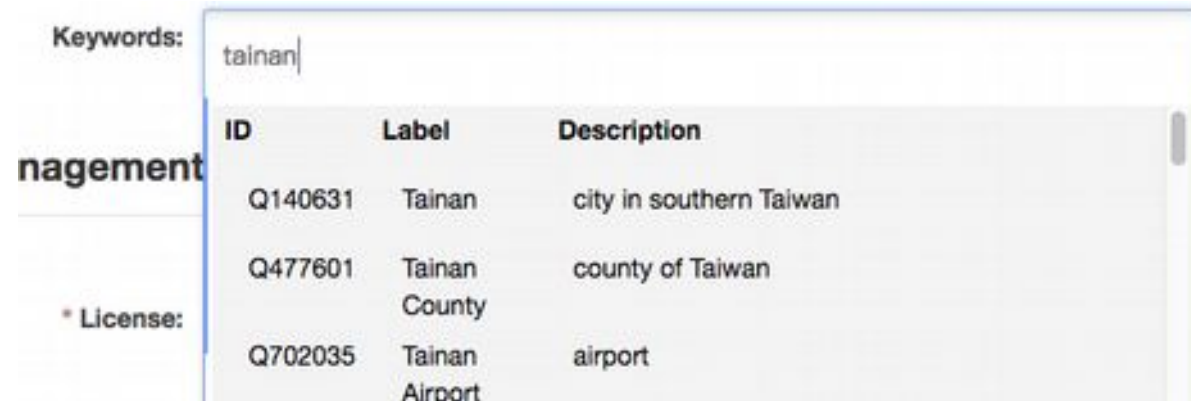
Wikidata-powered keywords

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



Wikidata entries as keywords

Search and select keywords:

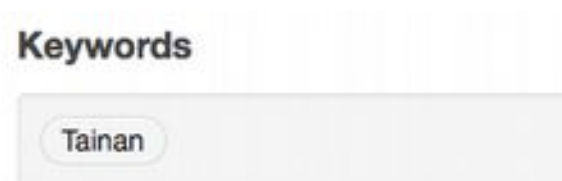


The screenshot shows a search interface with a text input field containing 'tainan'. Below the input field is a table of search results. The table has three columns: ID, Label, and Description. The results are as follows:

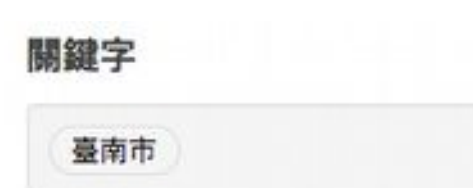
ID	Label	Description
Q140631	Tainan	city in southern Taiwan
Q477601	Tainan County	county of Taiwan
Q702035	Tainan Airport	airport

Keywords (as Wikidata IDs) are stored.

Viewed in English:



Viewed in Chinese:



Place Names & Maps – A Showcase

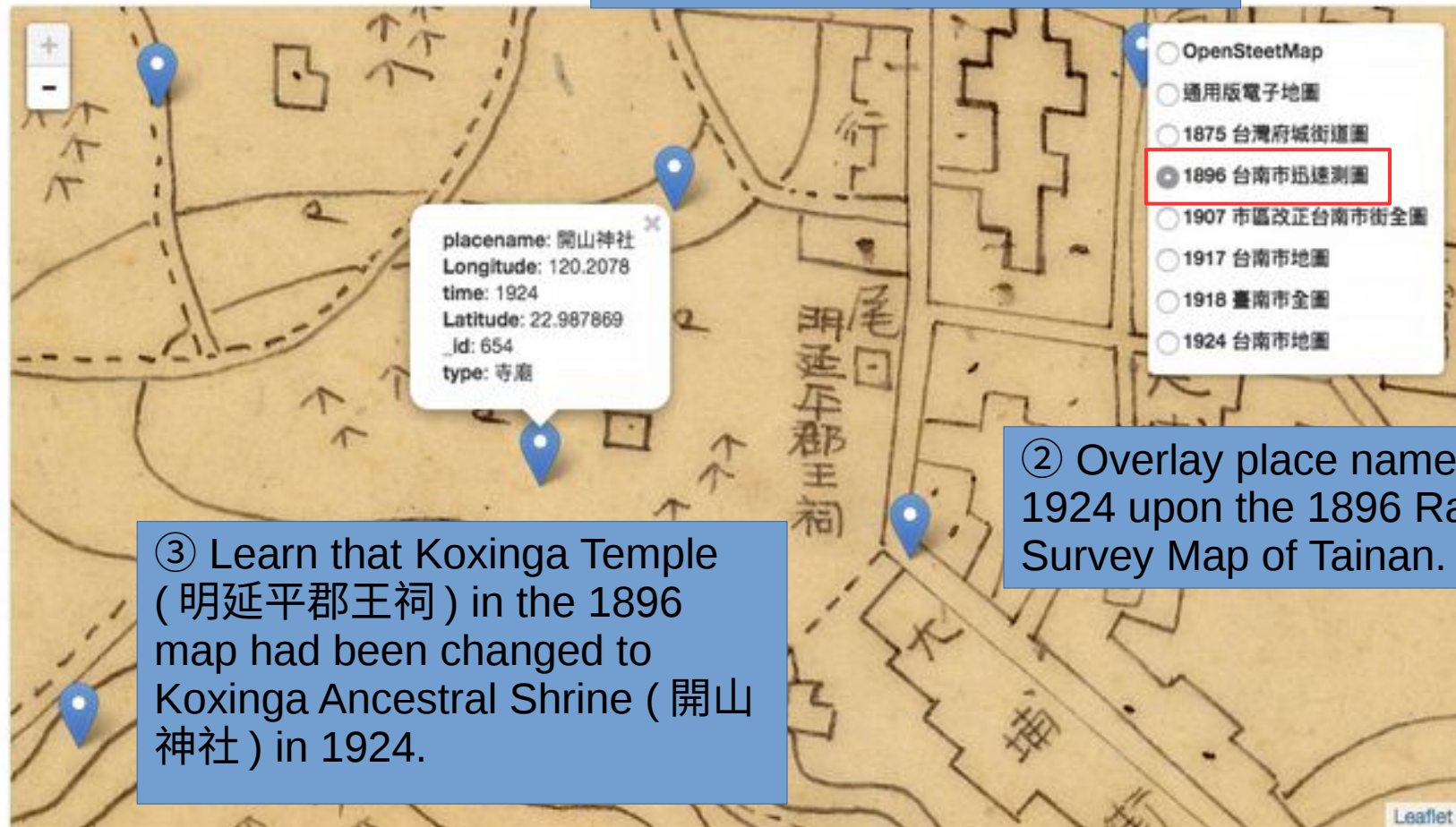
<https://data.depositar.io/dataset/place-names-in-west-central-district-of-tainan>

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).

Add Filter



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

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

Datasets / 路殺社範例資料集-2016

Dataset Groups Activity Stream History

路殺社範例資料集-2016

0 Followers

Social

Google+

Twitter

Facebook

Licenses

Licenses Not Specified

Dataset extent

Map

Other Access

The information on this page (the dataset metadata) is also available in these formats: [CSV](#) [JSON](#)

via the CKAN API

路殺社範例資料集-2016

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 122.68358369547873

Y.min 21.45913969982141

Y.max 26.44352482769591

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

Datasets / 路殺社範例資料集-2016 / 路殺社範例資料集

Dataset Groups Activity Stream History

路殺社範例資料集

121.35888882608885, 122.68358369547873 resource:1138a00f-199b-487c-b76c-76b2c33d57c0e0

From the dataset abstract

路殺社範例資料集-2016

Source: 路殺社範例資料集 (2016)

Map

Embed

Area Filter

Grid Graph Map 3275 records 1 - 50 Search data...

#	ID	科名	科名(英)	路殺社	路殺社	紀錄時間	地點	數量	備註
1	24932	Suburba	路殺社	Pana lin	路殺社	2016-01	路殺社	121.35888882608885	24.794442
2	24934	Suburba	路殺社	Turkey co	路殺社	2016-01	路殺社	122.38166	25.62861
3	24936	Suburba	路殺社	Randall	路殺社	2016-01	路殺社	120.196	22.894976
4	24938	Tulip	路殺社	Mogana	路殺社	2016-01	路殺社	120.788	25.773827
5	24971	Culiville	路殺社	Igreth	路殺社	2016-01	路殺社	120.94217	24.994201
6	24972	Viper	路殺社	Trois	路殺社	2016-01	路殺社	120.974	24.702631
7	25004	Suburba	路殺社	Suburba	路殺社	2016-01	路殺社	121.35888882608885	25.127258
8	25006	Suburba	路殺社	Suburba	路殺社	2016-01	路殺社	121.35888882608885	22.729732
9	25008	Suburba	路殺社	Suburba	路殺社	2016-01	路殺社	120.348	22.281891
10	25009	Suburba	路殺社	Suburba	路殺社	2016-01	路殺社	120.724	24.231546
11	25023	Suburba	路殺社	Suburba	路殺社	2016-01	路殺社	118.412	24.588998
12	25025	Columbia	路殺社	Suburba	路殺社	2016-01	路殺社	118.219	24.471515
13	25026	Suburba	路殺社	Suburba	路殺社	2016-01	路殺社	121.35888882608885	25.368142
14	25027	Columbia	路殺社	Columbia	路殺社	2016-01	路殺社	121.35888882608885	24.388222
15	25028	Rhacoph	路殺社	Rhacoph	路殺社	2016-01	路殺社	121.734	25.391321
16	25029	Rhacoph	路殺社	Rhacoph	路殺社	2016-01	路殺社	121.734	25.391321
17	25030	Rhacoph	路殺社	Rhacoph	路殺社	2016-01	路殺社	121.734	25.391321
18	25032	Columbia	路殺社	Physa	路殺社	2016-01	路殺社	120.7871	24.78225
19	25033	Suburba	路殺社	Suburba	路殺社	2016-01	路殺社	121.35888882608885	24.388222
20	25034	Suburba	路殺社	Suburba	路殺社	2016-01	路殺社	118.412	24.587998

Resource

Additional information

Lead updated June 4, 2016

Created June 4, 2016

Format CSV

Encoding UTF-8

Coordinate Systems EPSG: 4326

Licenses 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

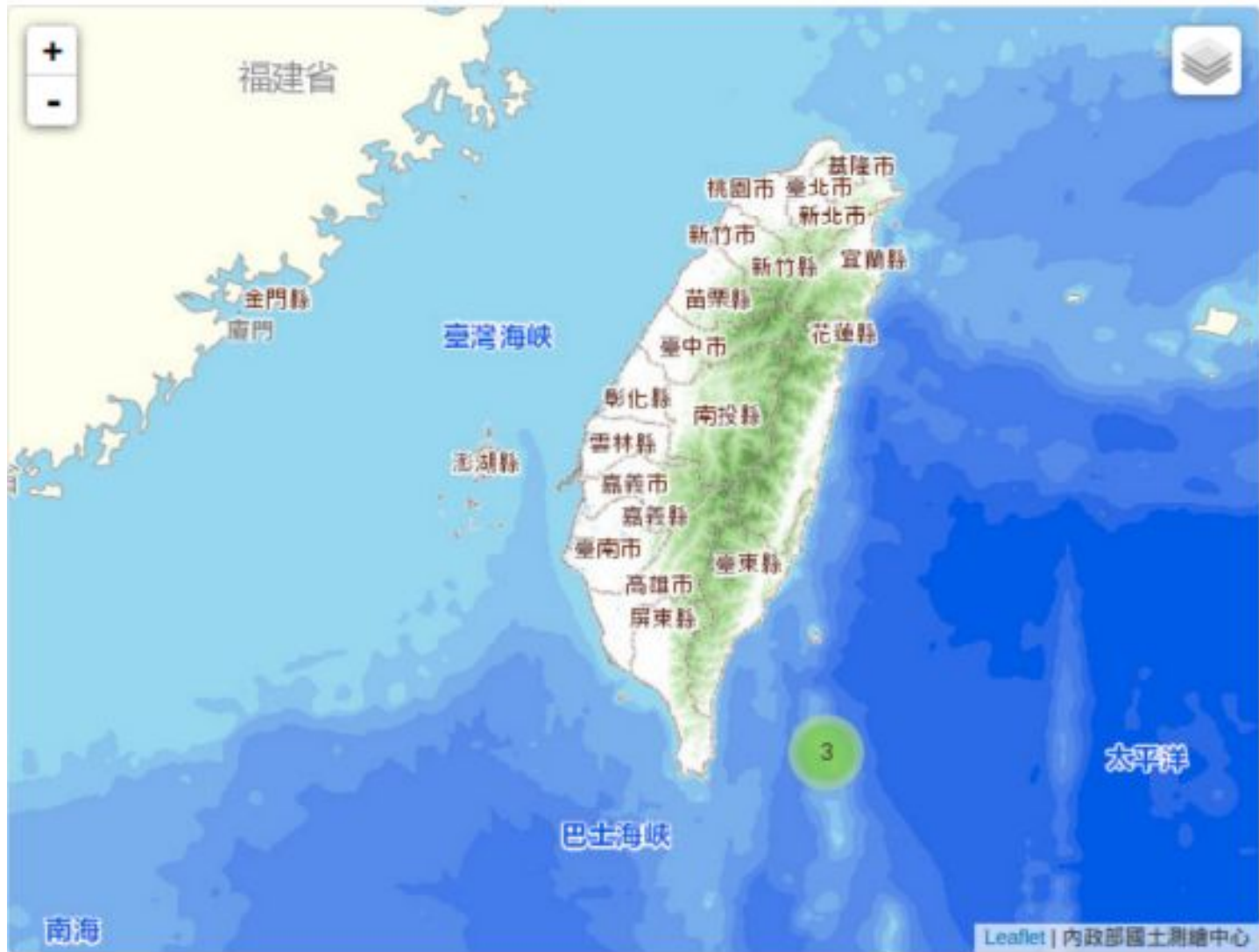
Filters

Add filter

科名 (中文) ×

蟾蜍科

Update



- ☒ Latitude / Longitude fields
☐ GeoJSON field

Latitude field

緯度

Longitude field

經度

Update

- ☒ Auto zoom to features
☒ Cluster markers

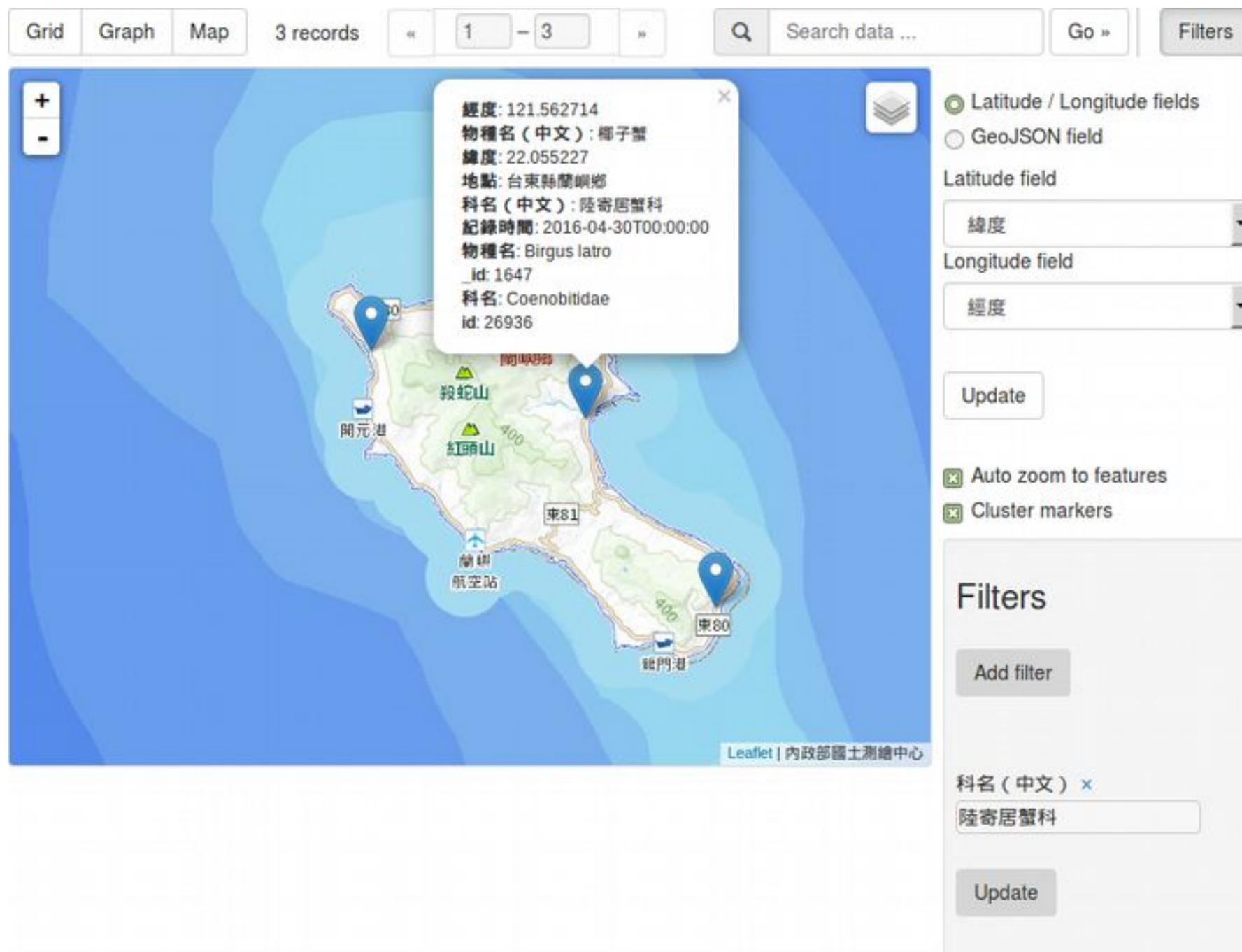
Filters

Add filter

科名 (中文) ✕

陸寄居蟹科

Update



Grid

Graph

Map

3 records

«

1

- 3

»



Search data ...

Go »

Filters

+

-

經度: 121.510102

物種名 (中文): 短掌陸寄居蟹

緯度: 22.070254

地點: 台東縣蘭嶼鄉

科名 (中文): 陸寄居蟹科

紀錄時間: 2016-05-02T00:00:00

物種名: Coenobita brevipanus

_id: 1795

科名: Coenobitidae

id: 27120

☒ 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) x

陸寄居蟹科

Update

Leaflet | 內政部國土測繪中心

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.

depositar – an experiment in progress

- Challenge: Transition from a tool to a service
 - maintaining a stable service by and for resource-poor academic institutes, projects, and individuals
- Understand and respond to user needs
 - serving small projects and emerging topics
 - multilingual support, batch dataset upload
 - data preview and analysis; links to/from external resources
- What about a *Terms of Use*?
- Toward a collaborative infrastructure from below



Datasets

Topics

Projects

About

Help



deposit • discover •
reuse

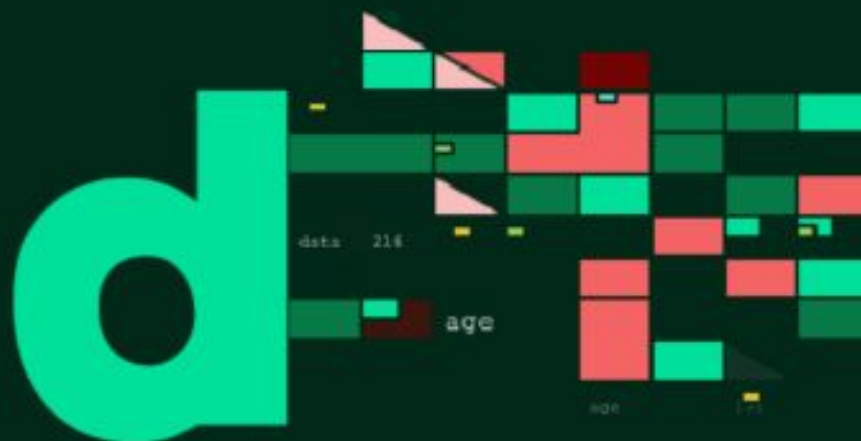
Learn More





儲存 · 尋找 · 分享

了解更多



Topics

Roadkill



General
reference



Culture and
the arts



Geography
and places



Health and
fitness



History and
events



Human
activities



Mathematics
and logic



Natural and
physical
sciences



People and
self



Philosophy
and thinking



Religion and
belief
systems

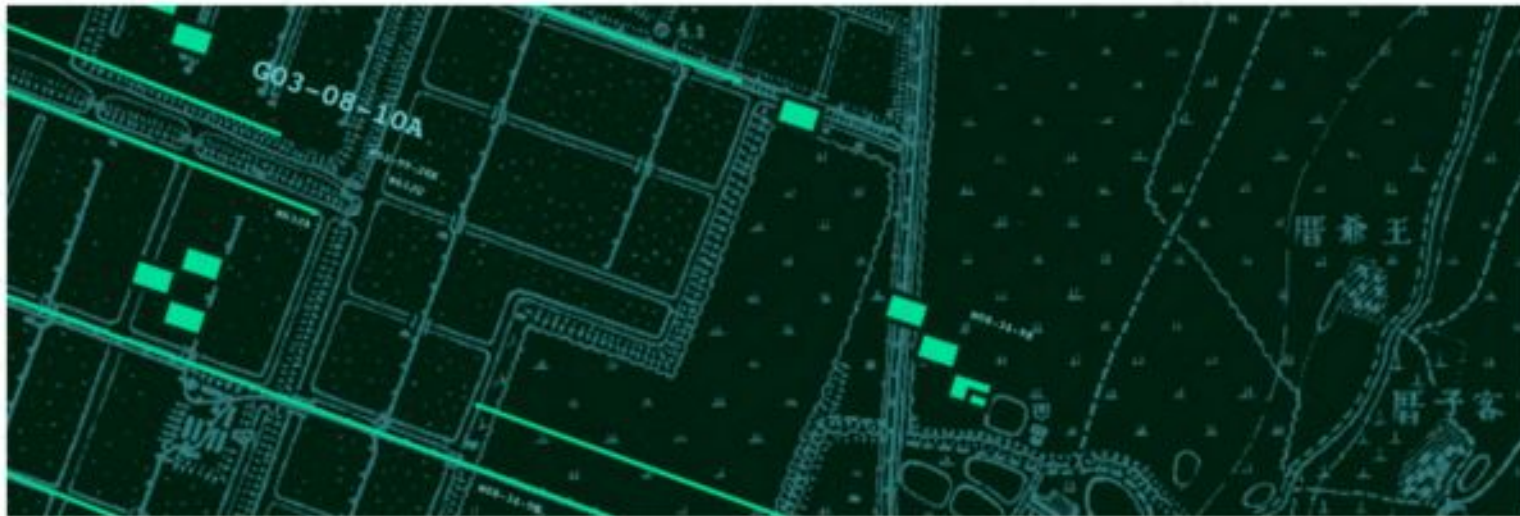


Society and
social
sciences



Technology
and applied
sciences

Selected Datasets



Taiwan History and Culture in Time and Space

Do you know that there are ancient waterway routes around Taiwan?

Have you heard about the routes for gold prospecting in Eastern Taiwan during the Dutch Formosa period?

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](#)

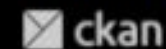
d_depositar

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

About depositar

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

Powered by



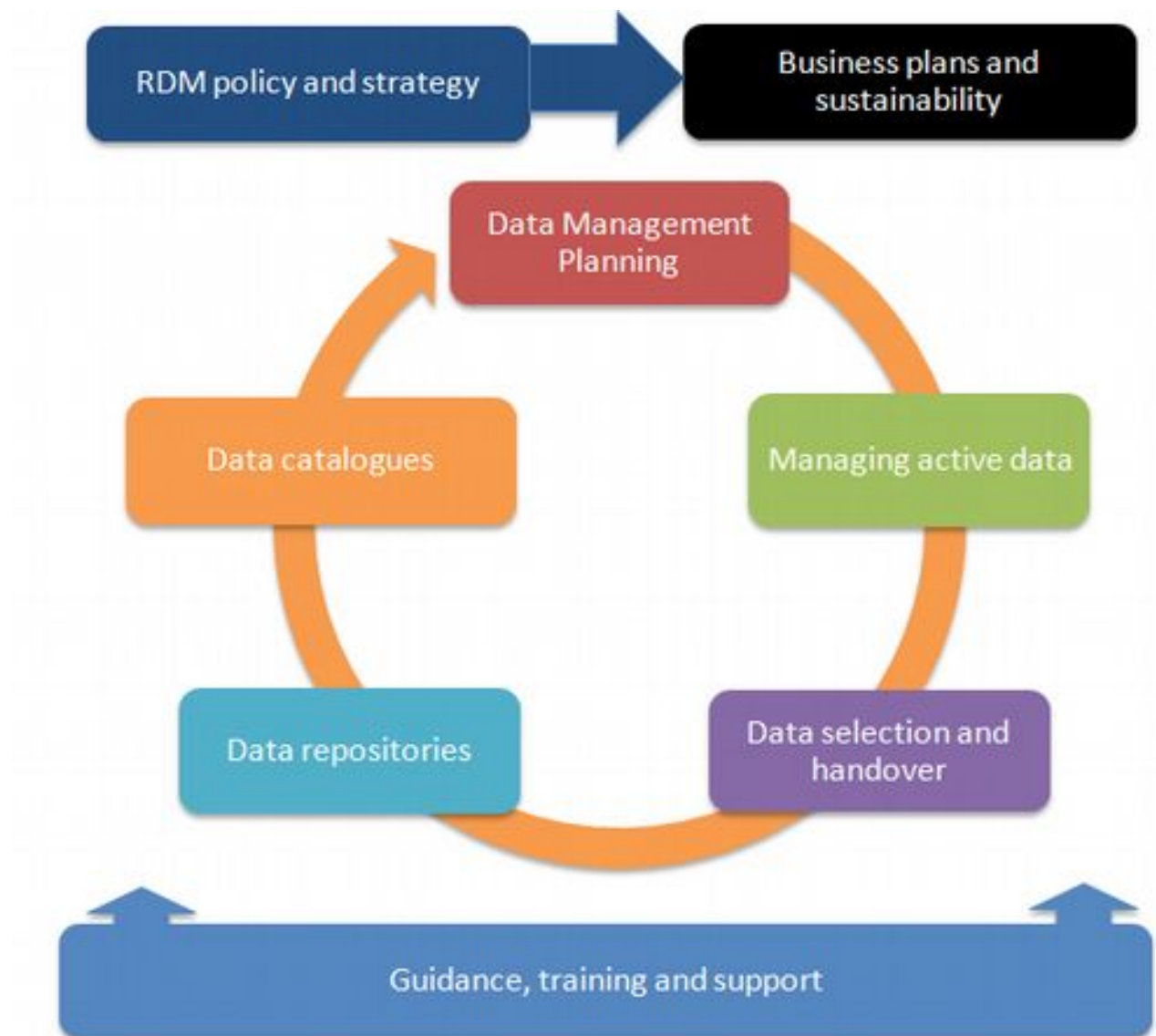
Source code available on [GitHub](#).

Visual Design & UI by

[Dualai Studio](#)

Research Data Management (RDM)

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



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



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

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

data.contact@depositar.io

