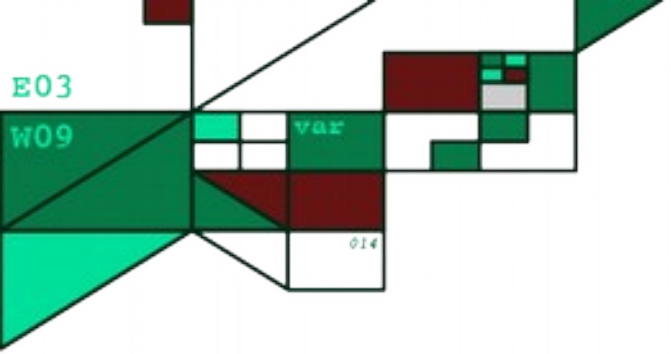


Collaboration in Soundscape Monitoring:
A Use Case for Research Data Management

Soundscape Conference, ISGC 2019
2019-04-01, Taipei, Taiwan

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

depositar is a collaboration at IIS, CITI, and GIS Center (RCHSS) in Academia Sinica, supported in part by MOST grants, Taiwan
「研究資料寄存所」是中央研究院資訊所、資創中心、 GIS 中心（人社中心）的協作專案，部份經費來自台灣科技部計畫

CC BY

Outline

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

Research Data Management (RDM)

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

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 clients

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
  - simplified metadata authoring

Publisher *vs.* Depository

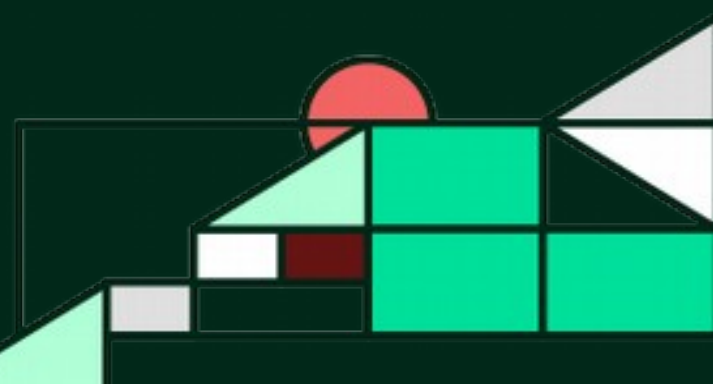
- CKAN – software for publishing open data
- *depositar* – a depository for research data

- 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

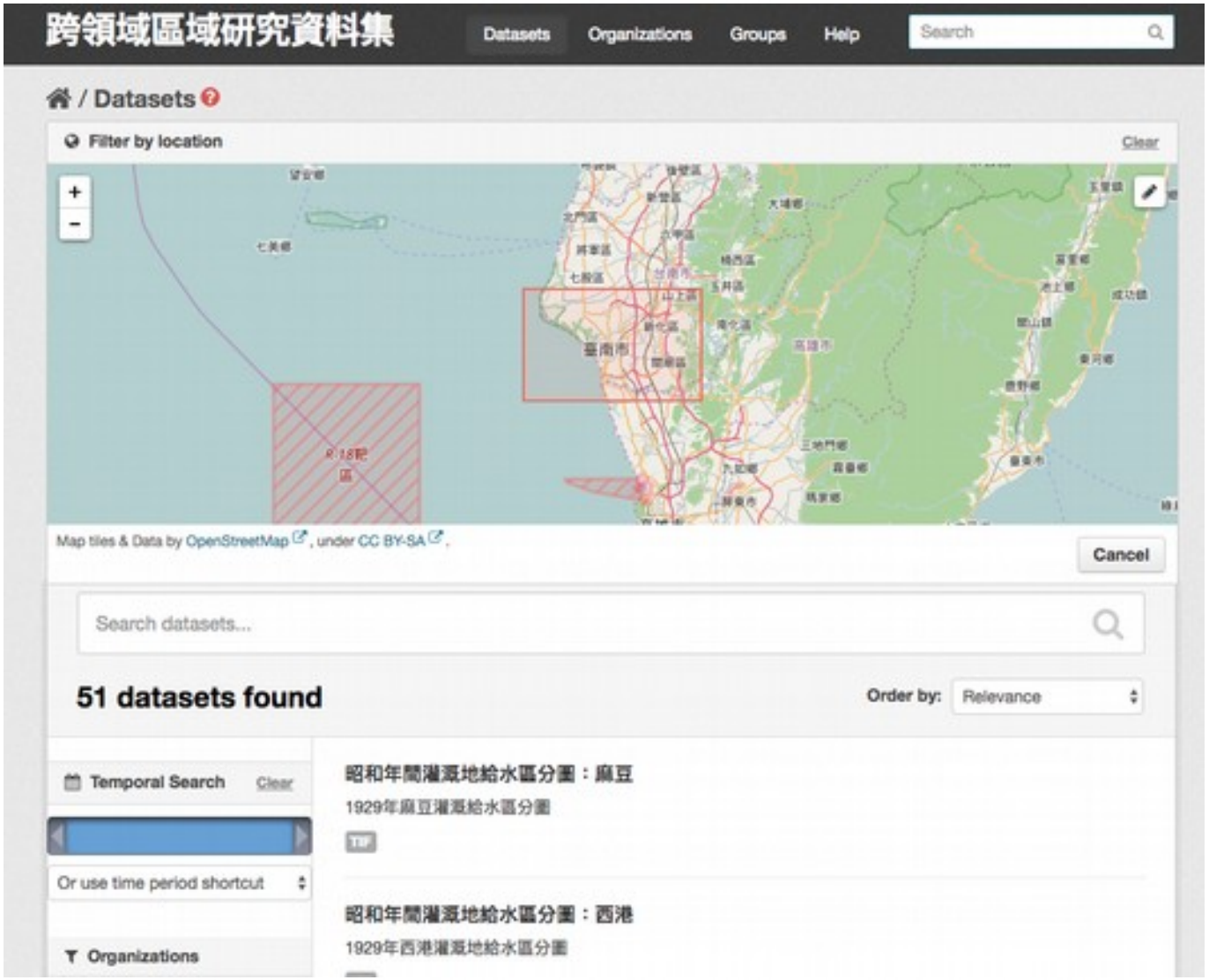
depositar

Features and Show Cases

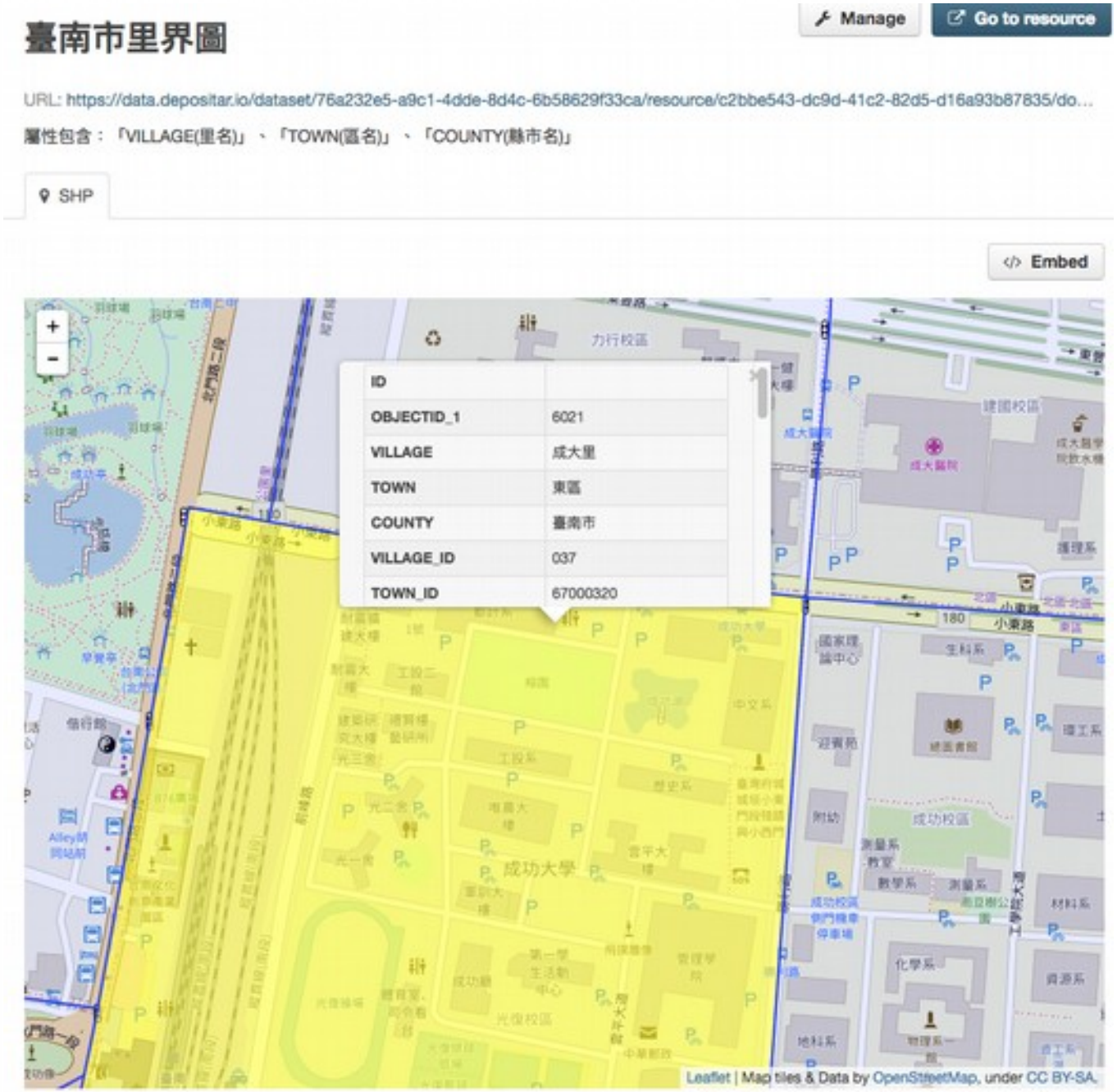


Search and Data Discovery

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

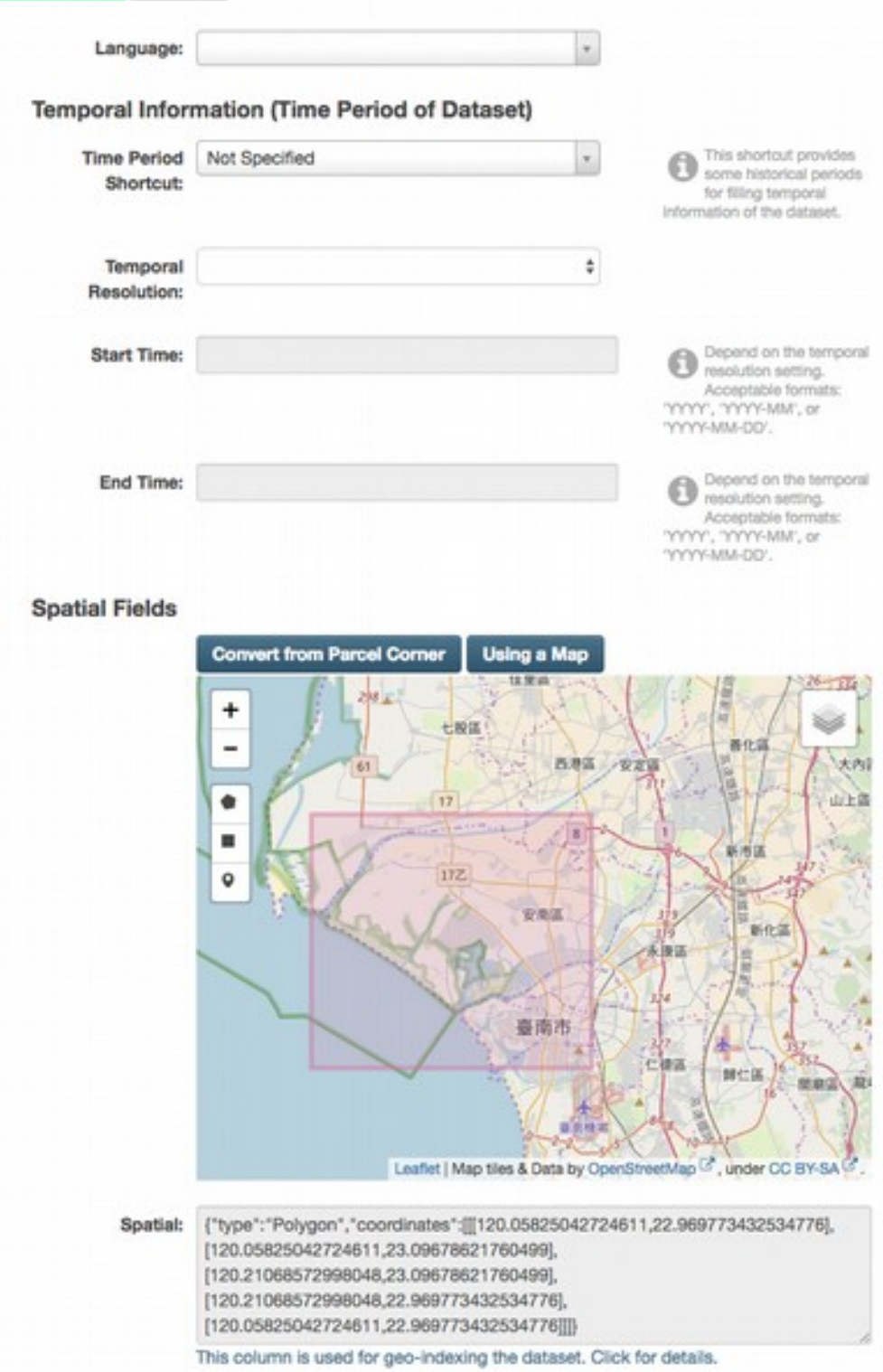


Data Preview



Metadata for Datasets

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



- Bilingual
  - English
  - Chinese
- Metadata
- User interface

Simplified Metadata Authoring

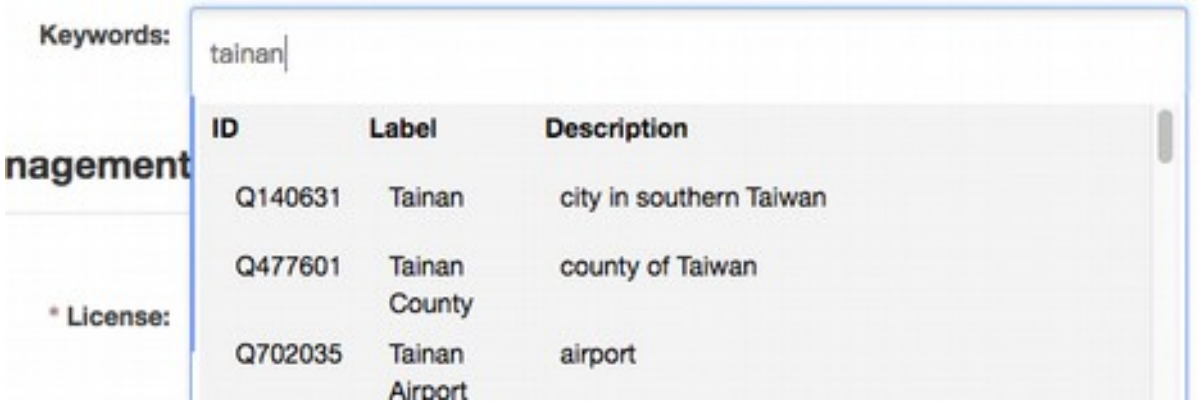
- One set of simple metadata for all datasets; three categories:
  - **Basic metadata:** title, description, data type, …
  - **Descriptive metadata:** language, temporal, and spatial information, keywords, …
  - **Management metadata:** license, author, created time, organization, maintainer, …

Wikidata-powered Keywords

- Keywords
  - controlled vocabularies for tagging datasets
  - DIY vocabulary design: a never-ending story …
- Use Wikidata as the source for keywords
  - 47M+ entries
  - multilingual
  - semantic relations enable data inference
    - e.g. Taipei is part of Taiwan

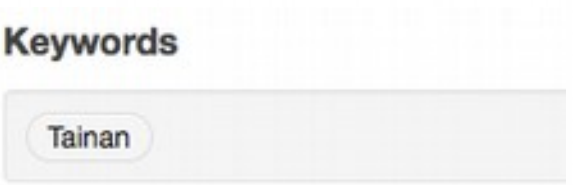
Wikidata Entries as Keywords

Search and select keywords:

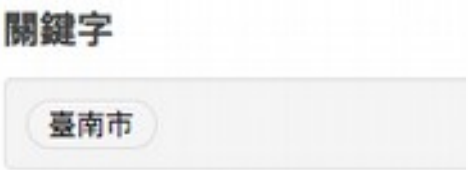


Keywords (as Wikidata IDs) are stored.

Viewed in English:

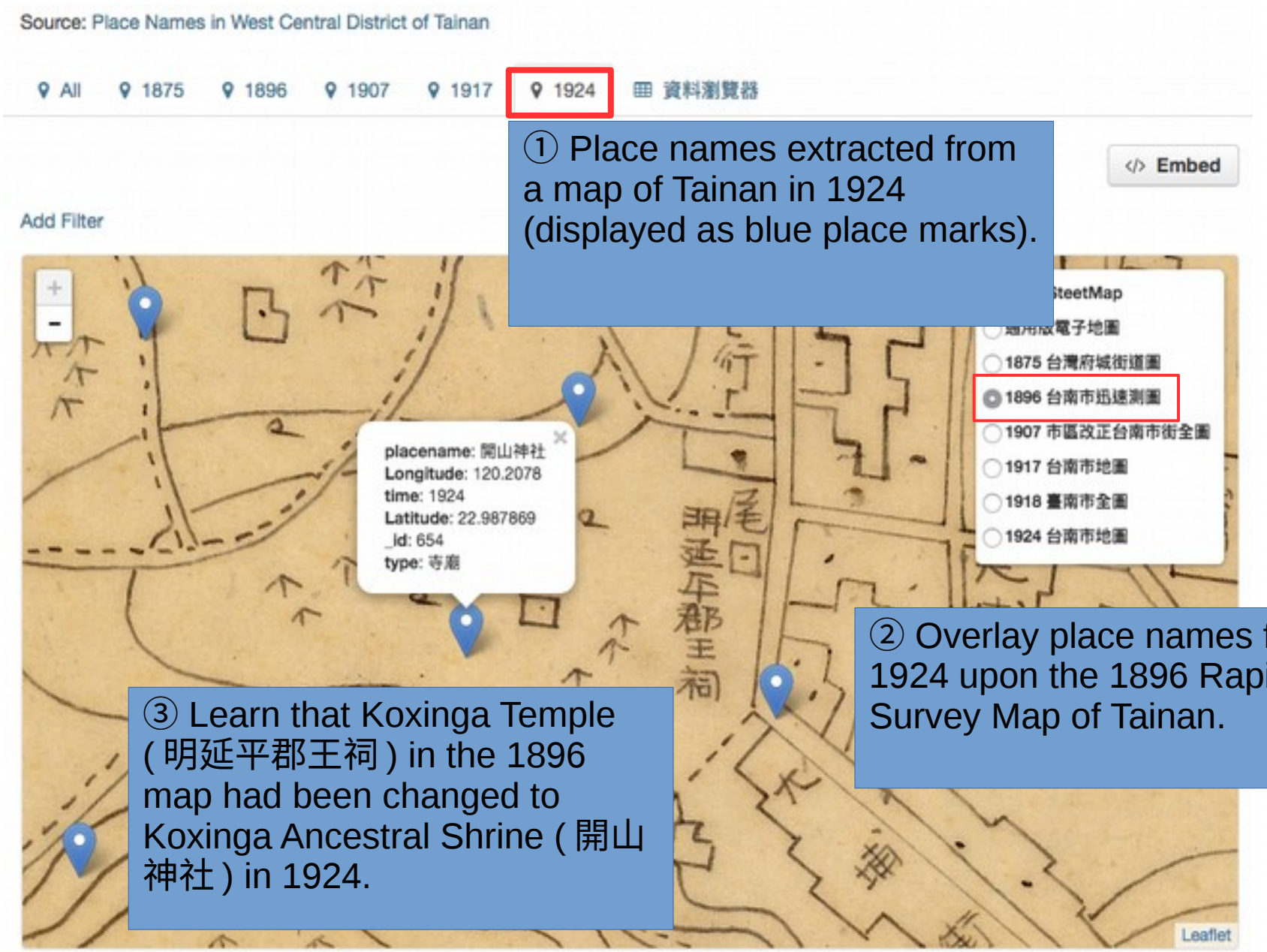


Viewed in Chinese:



Place Names & Maps – A Showcase

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



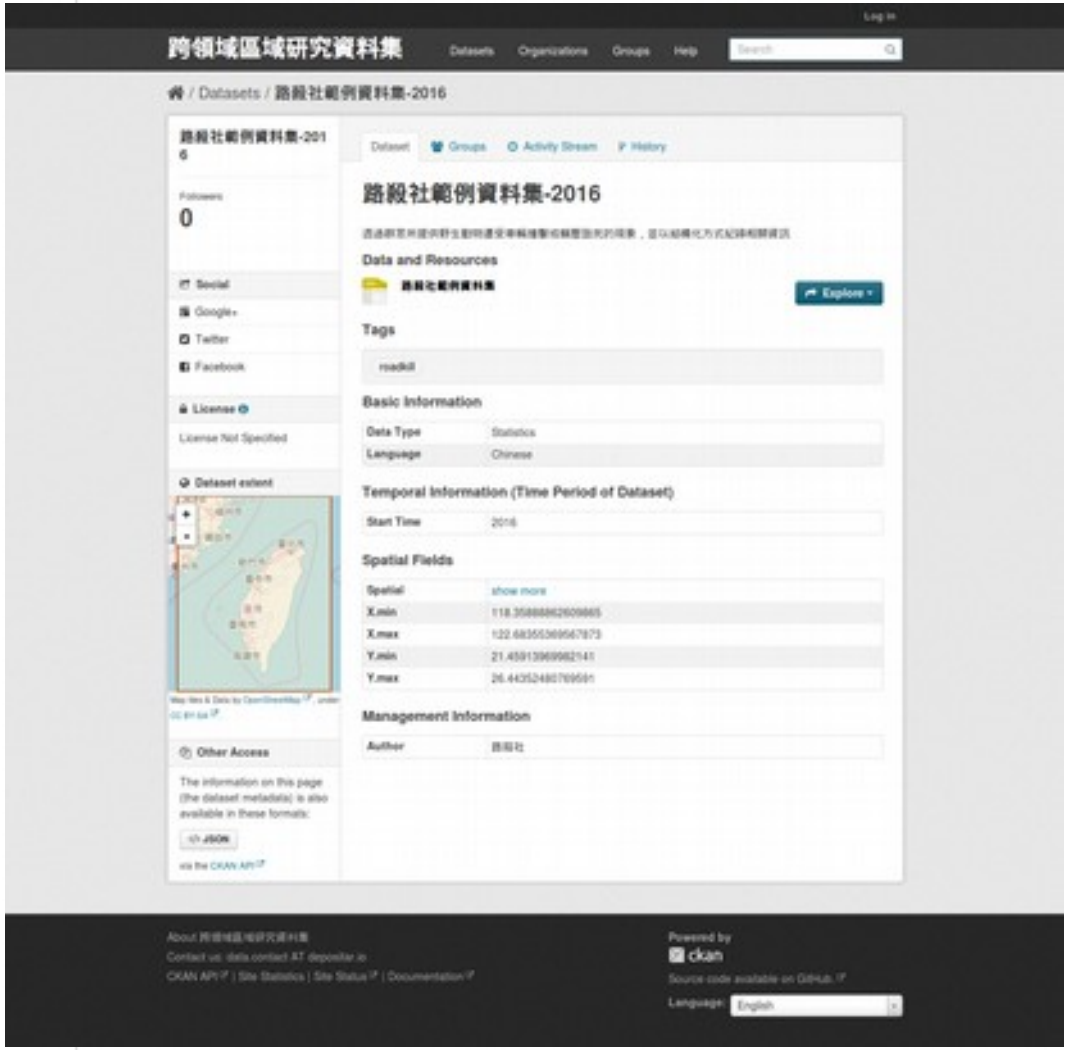
① Place names extracted from a map of Tainan in 1924 (displayed as blue place marks).

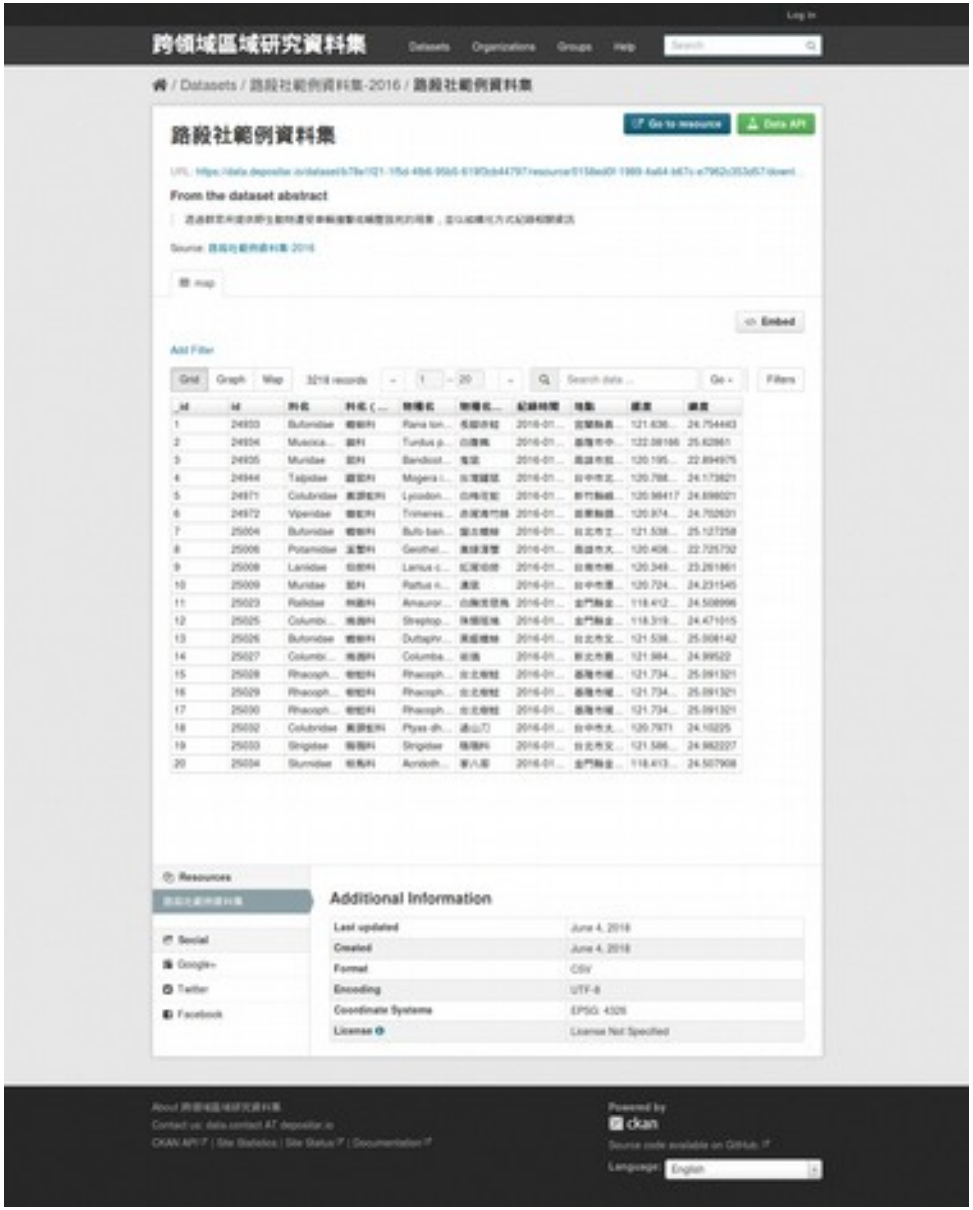
② Overlay place names from 1924 upon the 1896 Rapid Survey Map of Tainan.

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.

Roadkill Observations – A Showcase

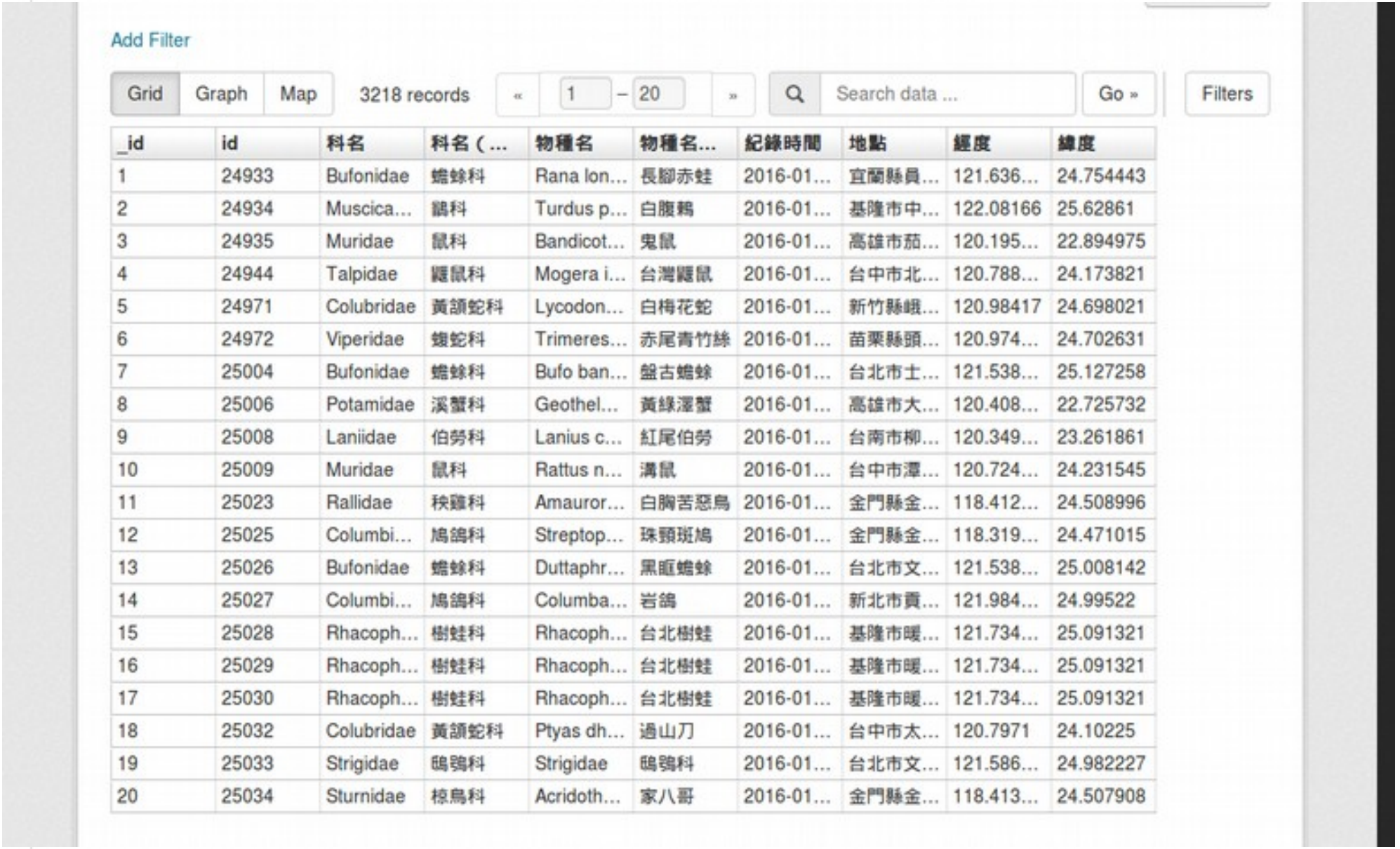
https://data.depositar.io/dataset/2016





Roadkill Observations – A Showcase

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



Add Filter

Grid | Graph | Map | 3218 records | « 1 – 20 » | 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	Bufonidae	蟾蜍科	Rana lon...	長腳赤蛙	2016-01...	宜蘭縣員...	121.636...	24.754443
2	24934	Muscica...	鶲科	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	鼴鼠科	Mogera i...	台灣鼴鼠	2016-01...	台中市北...	120.788...	24.173821
5	24971	Colubridae	黃頷蛇科	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	Bufonidae	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	Bufonidae	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	黃頷蛇科	Ptyas dh...	過山刀	2016-01...	台中市太...	120.7971	24.10225
19	25033	Strigidae	鴟鴞科	Strigidae	鴟鴞科	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

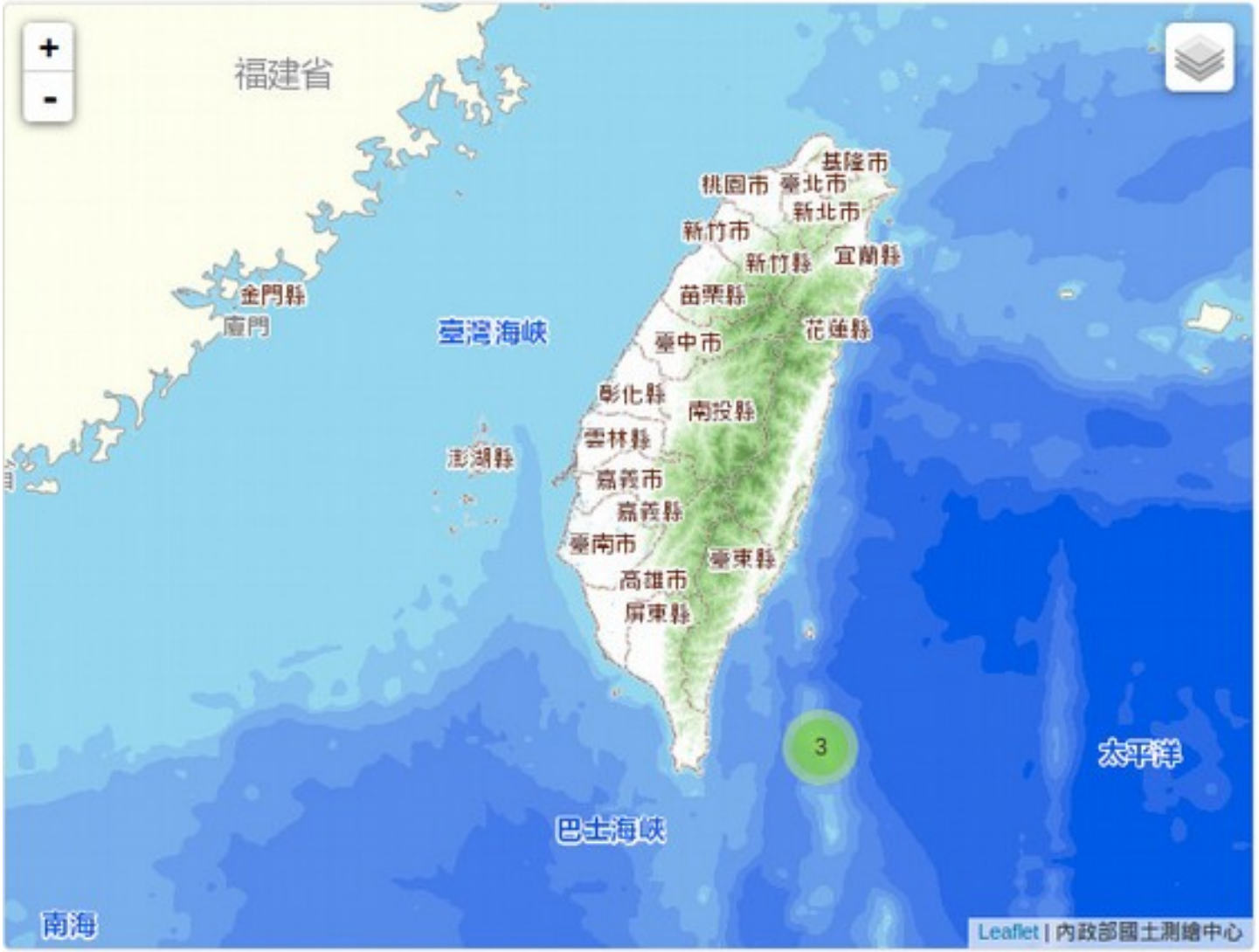
Add filter

科名（中文） ×

蟾蜍科

Update

Grid | Graph | Map | 3 records | « | 1 – 3 | » | Search data ... | Go » | Filters



Latitude / Longitude fields

GeoJSON field

Latitude field

緯度

Longitude field

經度

Update

Auto zoom to features

Cluster markers

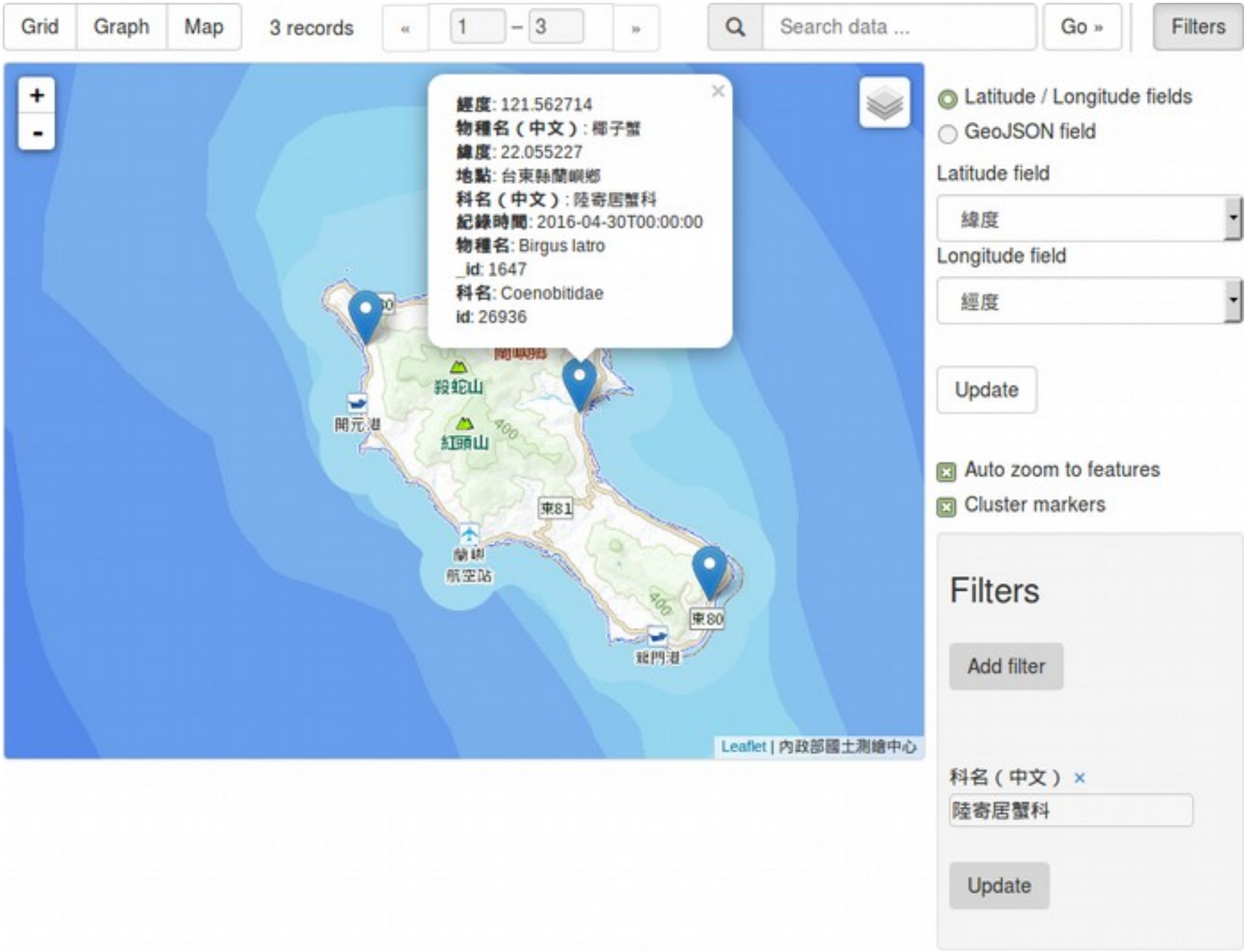
Filters

Add filter

科名（中文） ×

陸寄居蟹科

Update


Grid
Graph
Map
3 records
«
1
– 3
»
Search data ...
Go »
Filters
+
-
經度: 121.562714
物種名（中文）: 椰子蟹
緯度: 22.055227
地點: 台東縣蘭嶼鄉
科名（中文）: 陸寄居蟹科
紀錄時間: 2016-04-30T00:00:00
物種名: Birgus latro
_id: 1647
科名: Coenobitidae
id: 26936
殺蛇山
紅頭山
東81
東80
Leaflet | 內政部國土測繪中心
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

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） ×

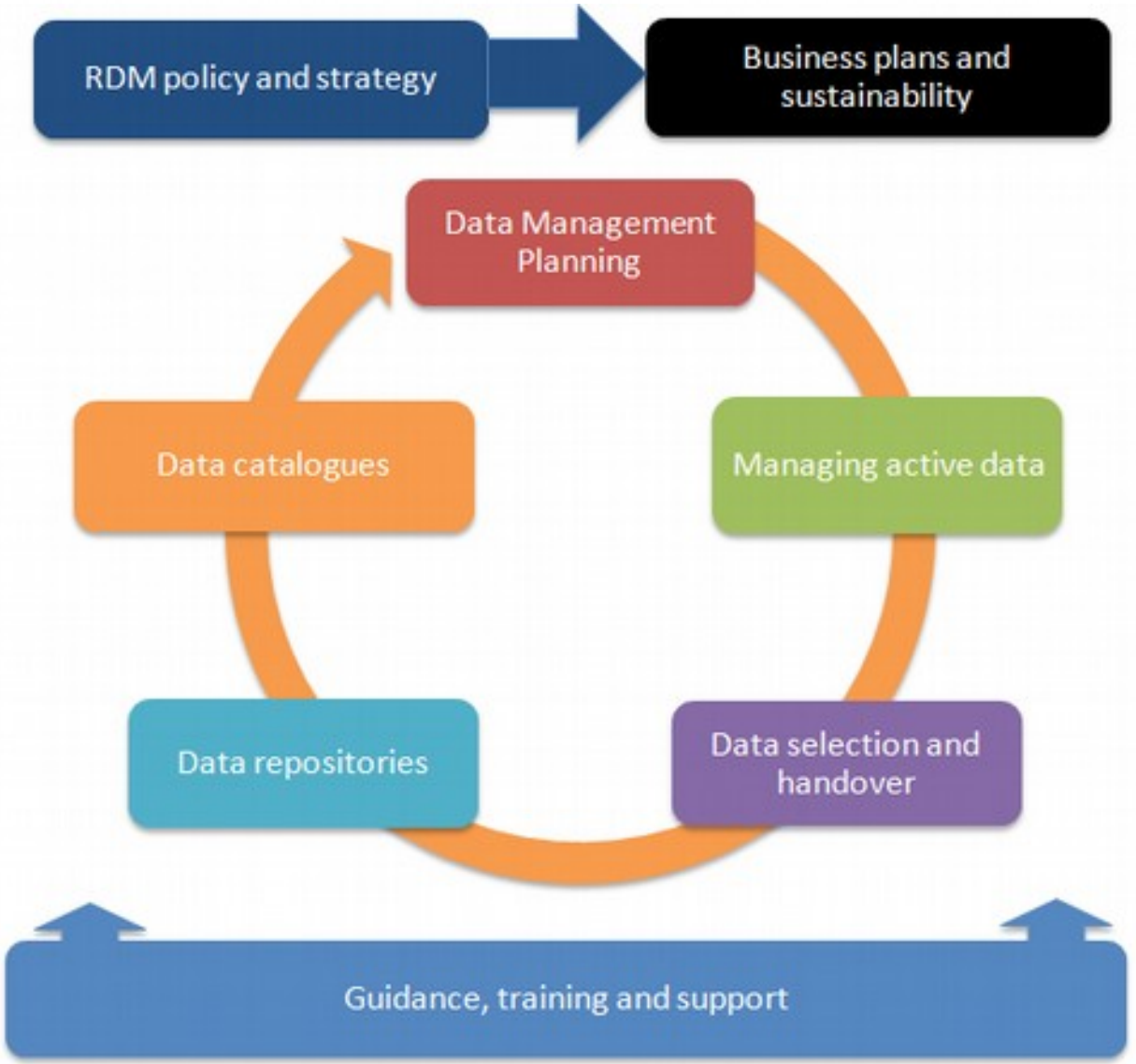
陸寄居蟹科

Update

Collaboration in Soundscape Monitoring

Research Data Management (RDM)

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



Data Management Plan

- **Data Collection**
  - What data will you collect or create?
  - How will the data be collected or created?
- **Documentation and Metadata**
  - What documentation and metadata will accompany the data?
- **Ethics and Legal Compliance**
  - How will you manage any ethical issues?
  - How will you manage copyright and Intellectual Property Rights (IP/IPR) issues?
- **Storage and Backup**
  - How will the data be stored and backed up during the research?
  - How will you manage access and security?
- **Selection and Preservation**
  - Which data are of long-term value and should be retained, shared, and/or preserved?
  - What is the long-term preservation plan for the dataset?
- **Data Sharing**
  - How will you share the data?
  - Are any restrictions on data sharing required?
- **Responsibilities and Resources**
  - Who will be responsible for data management?
  - What resources will you require to deliver your plan?

This plan is based on the "DCC Template" template provided by Digital Curation Centre.

data.depositar.io

Topics

- Asian Soundscape Monitoring Network
- Natural and physical sciences

Project
Asian Aquatic Soundscape

Dataset **site1**

Dataset **site2**

. . .

Project
Asian Forest Soundscape

Dataset **site1**

Dataset **site2**

. . .

Project
Ultrasoundscape in Taiwan

Dataset **site1**

Dataset **site2**

. . .

Keywords

- biodiversity
- geophony
- biophony
- anthrophony
- Vietnam 、 Thailand
- Soundscape …

Tags

- biological sounds
- long-term monitoring
- Chiroptera …

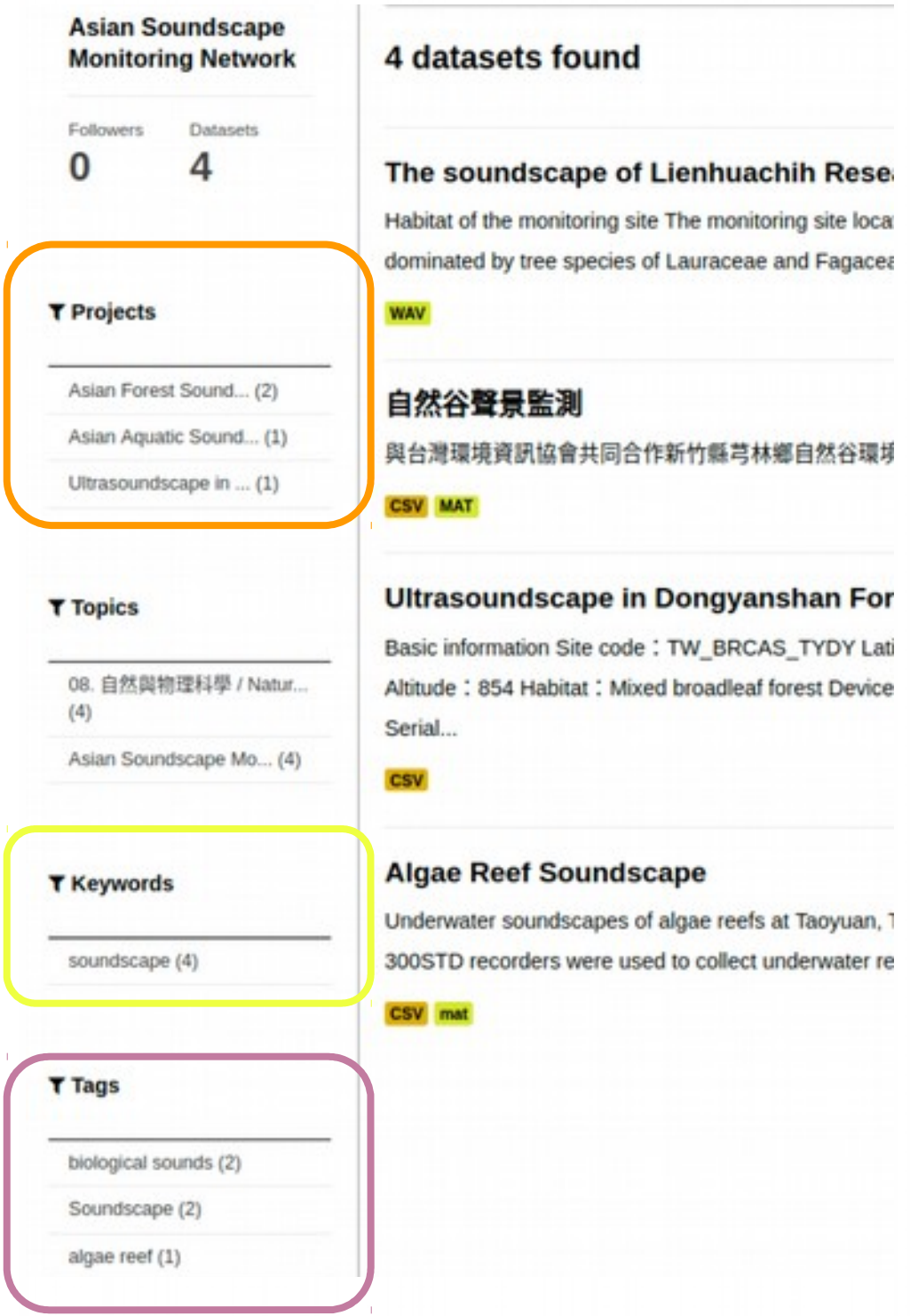
Formats

- csv 、 mat ...

Data on *depositar*

Asian Soundscape Monitoring Network

https://data.depositar.io/group/asian-soundscape-monitoring-network



Algae Reef Soundscape

Followers
0

Follow

Project

Asian Aquatic Soundscape

There is no description for this project

Social

Google+
Twitter
Facebook

License

CC-BY 4.0

Dataset extent

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

Dataset | Topics | Activity Stream | History | Manage

Algae Reef Soundscape

Underwater soundscapes of algae reefs at Taoyuan, Taiwan.

Acoustic Recorders. SoundTrap 300STD recorders were used to collect underwater recordings. Two SoundTrap recorders were used in three surveys: SoundTrap 300STD #1 SN1208004643; SoundTrap 300STD #2 SN805347354.

Configuration of Audio Recording. (1) Duration: 5 min. (2) Sampling Rate: 96kHz. (3) Channels: 1. (4) File Format: WAV (converted by SUD). (5) Audio Gain: High. (6) High Pass Filter: Off.

Field Deployment. Each SoundTrap recorder was fixed vertically on an angle steel at the bottom. The recorder was exposed in the air during low tide, so the collection of underwater recordings were only possible 3 hours before and after the high water.

Data Processing. We used the LTSA_gui (https://github.com/schonkopf/long-term-spectrogram) to visualize the spectral-temporal change of long-duration recordings. By the median-based long-term spectrogram, we employed the periodicity-coded non-negative matrix factorization (https://github.com/schonkopf/soundscape-viewer) to separate the audio data into two components: (1) high-frequency biological chorus, and (2) low-frequency ambient noise. The relative intensities of the two components were analyzed.

Data and Resources

- **Spatiotemporal change of soundscape** — This csv describes the Location, Latitude, Longitude, Survey Number, Tide... — Explore
- **Field recordings at BY** — Audio archive on Asian Soundscape — Explore
- **Field recordings at G1** — Audio archive on Asian Soundscape — Explore
- **Field recordings at G2** — Audio archive on Asian Soundscape — Explore
- **Field recordings at YX** — Audio archive on Asian Soundscape — Explore
- **Field recordings at YA** — Audio archive on Asian Soundscape — Explore
- **Long-term spectrogram of the Survey 1 at BY** — Please use Soundscape_Viewer (https://github.com/schonkopf/soundscape-viewer)... — Explore
- **Long-term spectrogram of the Survey 1 at G1** — Please use Soundscape_Viewer (https://github.com/schonkopf/soundscape-viewer)... — Explore

Basic site information

Spatiotemporal change of soundscape (.cvs)

Audio archive (Pumilio)

Visualization of acoustic data (.mat)

Lessons Learned

- Communication is always important
- Plan early for integration and cooperation
- Adequate training and support is essential
- Patience and understanding

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
- Toward a collaborative infrastructure from below

depositar — Deposit, Discover, Reuse - Mozilla Firefox

File Edit View History Bookmarks Tools Help

depositar — Deposit, [×

https://data.depositar.io/en/ 110%

Search

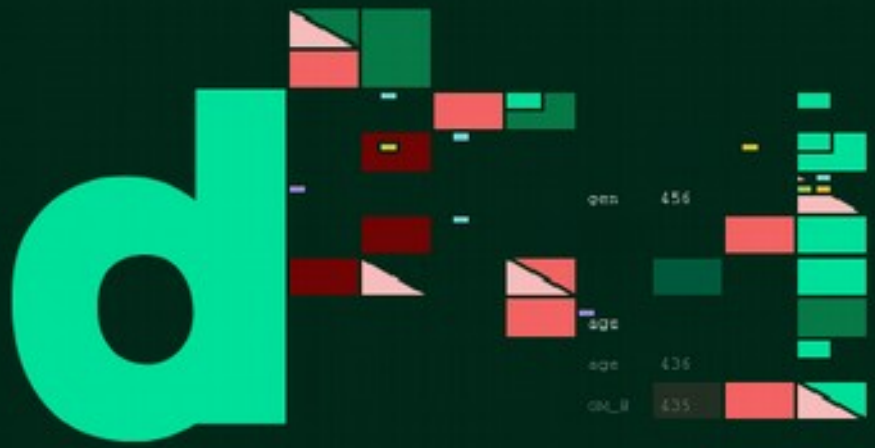
Log in | Register | 中文

depositar

Datasets Topics Projects About Help

deposit • discover • reuse

Learn More



depositar — 儲存、尋找、分享 - Mozilla Firefox

File Edit View History Bookmarks Tools Help

depositar — 儲存、尋 ×

https://data.depositar.io/zh_TW/ 110% Search

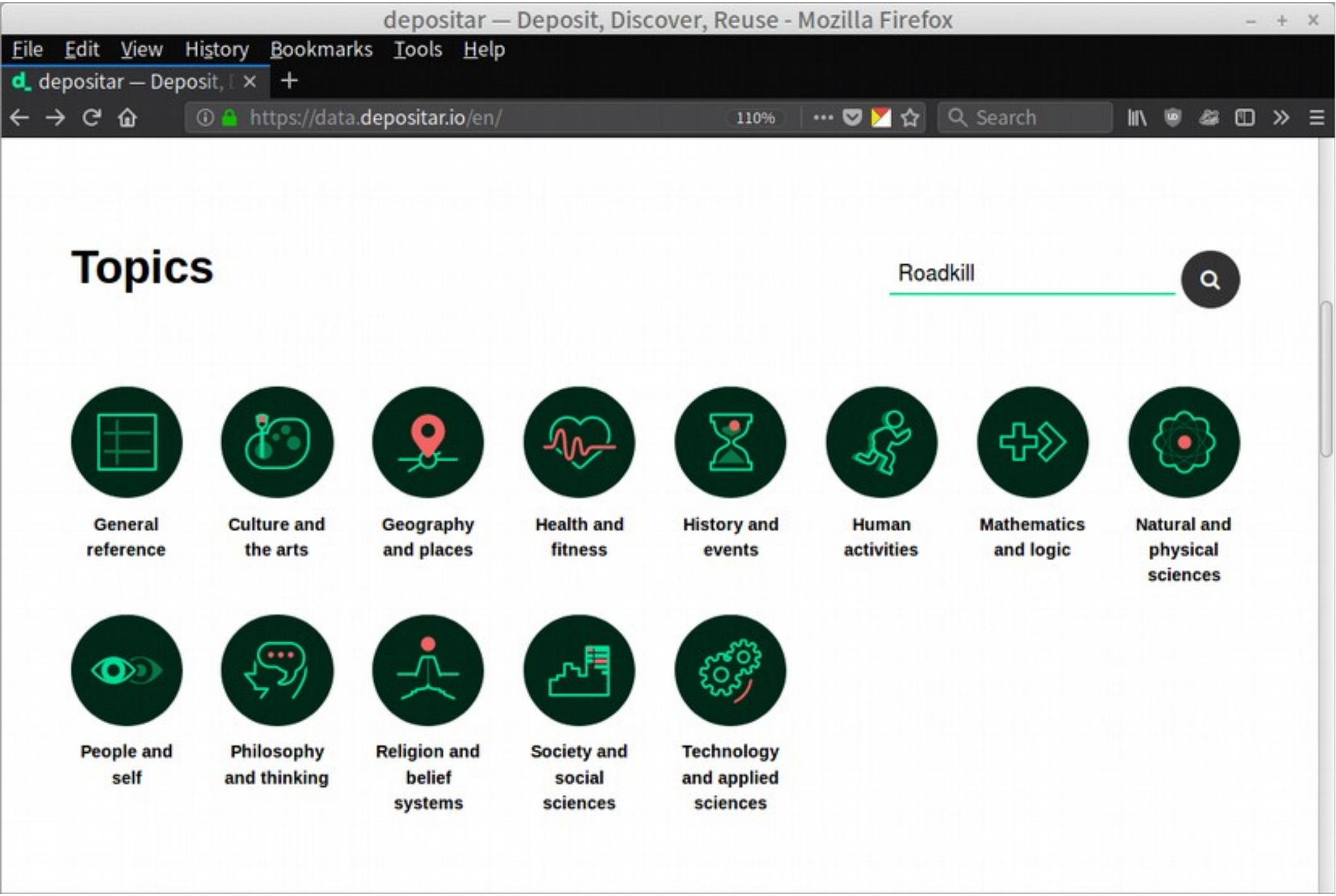
登入 | 註冊 | English

depositar

資料集 主題 專案 關於 支援

儲存・尋找・分享

了解更多


depositar — Deposit, Discover, Reuse - Mozilla Firefox
File Edit View History Bookmarks Tools Help
depositar — Deposit,
https://data.depositar.io/en/
110%
Search
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

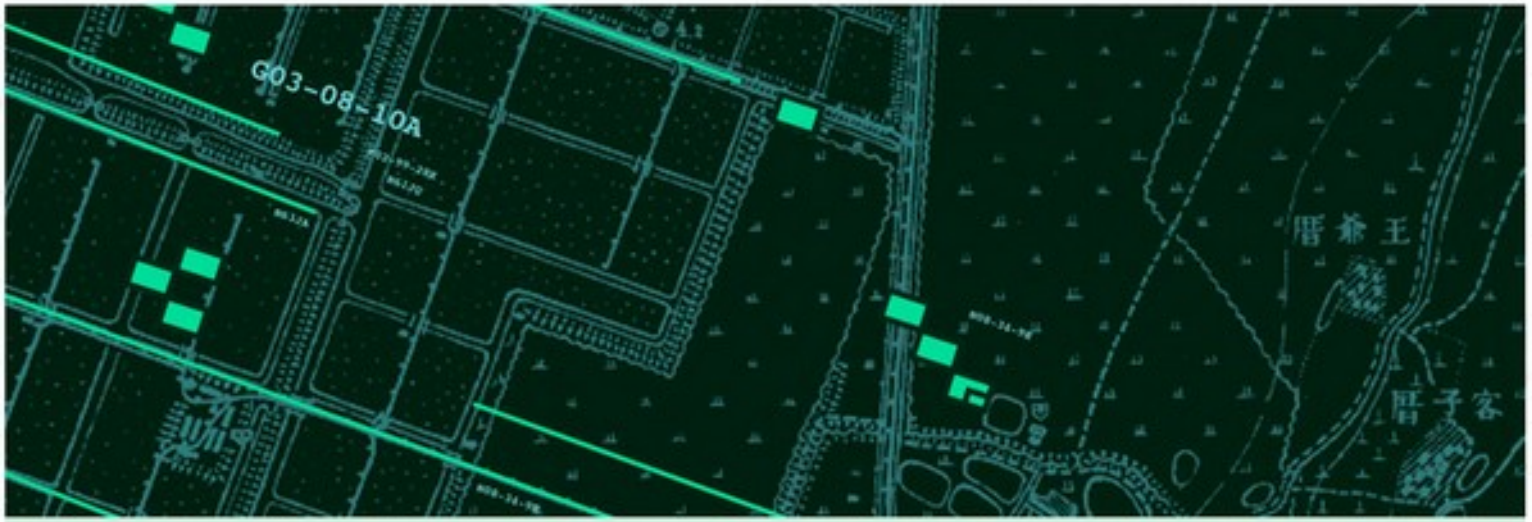
depositar — Deposit, Discover, Reuse - Mozilla Firefox

File Edit View History Bookmarks Tools Help

depositar — Deposit,

https://data.depositar.io/en/ 90% Search

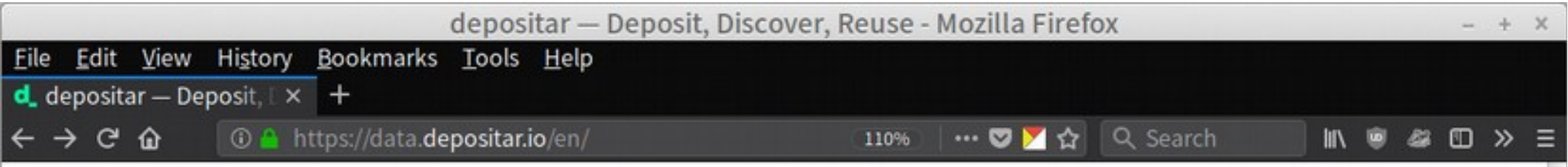
Selected Datasets



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

depositar

Contact Us: data.contact at depositar.io

About depositar

CKAN API | Site Statistics | Site Status | Help

Powered by

ckan

Source code available on GitHub.

Visual Design & UI by

Dualai Studio

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.

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

Deutsch
Die freie Enzyklopädie
1 430 000+ Artikel

Русский
Свободная энциклопедия
875 000+ статей

Français
L'encyclopédie libre
1 270 000+ articles

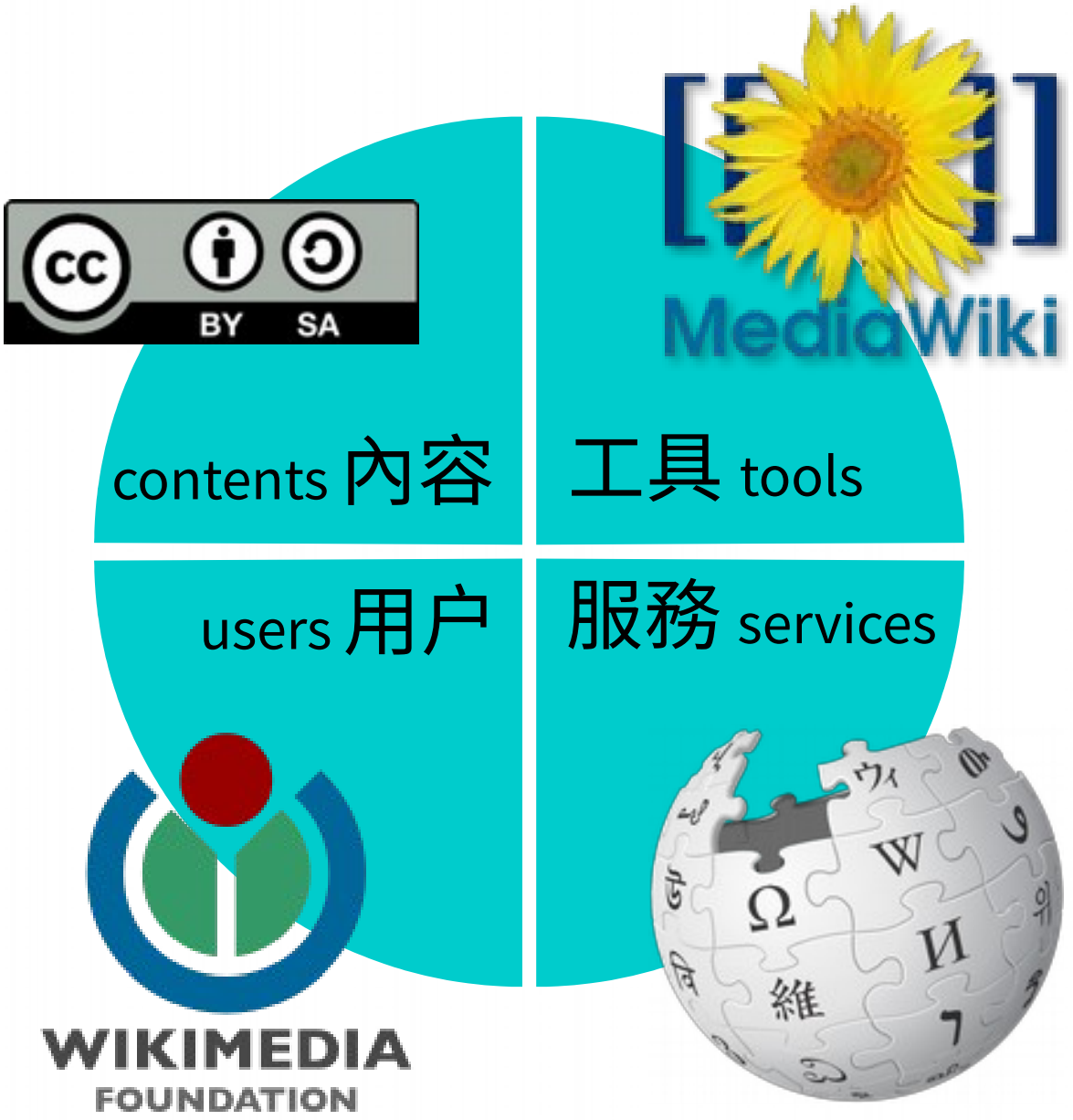
Italiano
L'enciclopedia libera
939 000+ voci

Polski
Wolna encyklopedia
908 000+ haseł

Português
A enciclopédia livre
742 000+ artigos

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 · Hľadať · Søg · Serĉu · Претрага · Paleška · Pošči · Cari · חיפוש · Търсене · Іздеу · Bilatu · Suk · Bilinga · Traži · खोजें


BY SA
MediaWiki
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