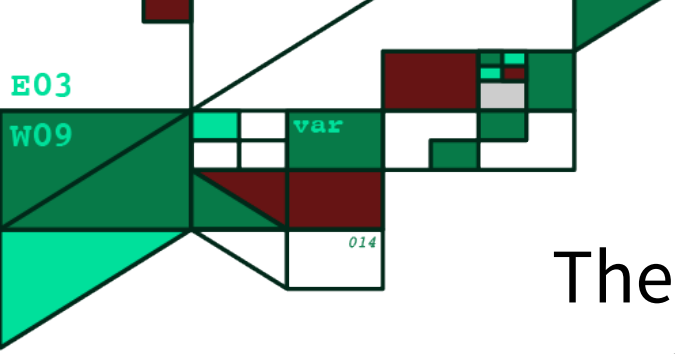


depositar
研究資料寄存所

The Development of *depositar* —An Open Repository for FAIR Data

貼近 FAIR 資料原則的開放儲存庫——研究資料寄存所 (depositar) 的發展

2024-02-01 @ National Yang Ming Chiao Tung University Library

Tyng-Ruey Chuang[1, 2, 3]
with Cheng-Jen Lee[1], Chia-Hsun Wang[1], and Ming-Syuan Ho[2]

莊庭瑞[1, 2, 3]、李承鑫[1]、王家薰[1]、何明諠[2]

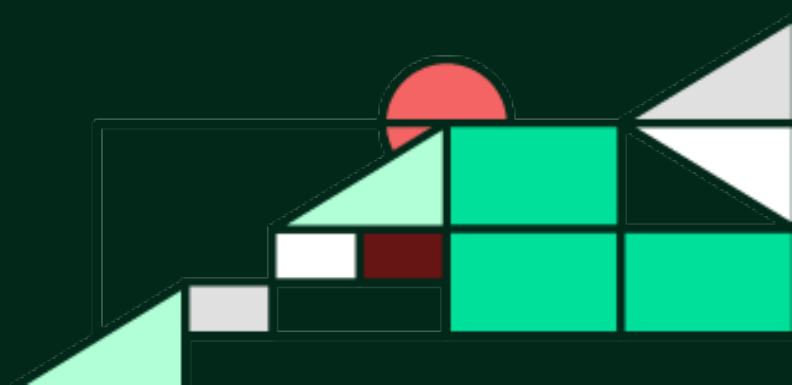
data.contact@depositar.io

[1] Institute of Information Science,
[2] Research Center for Information Technology Innovation &
[3] Research Center for Humanities and Social Sciences (GIS Center)
Academia Sinica, Taiwan

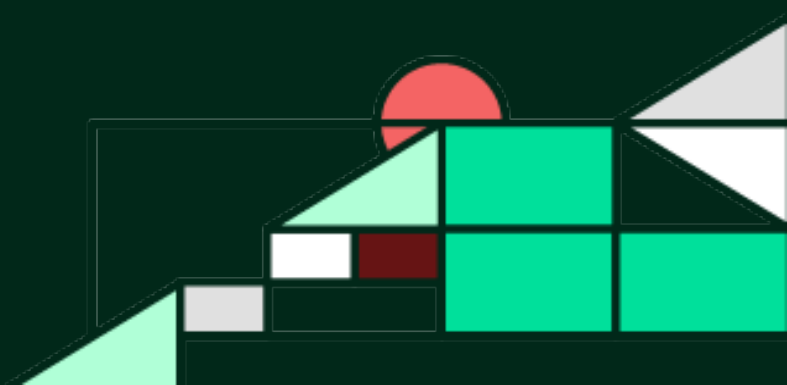
https://pid.depositar.io/ark:37281/k5c6h2g3n

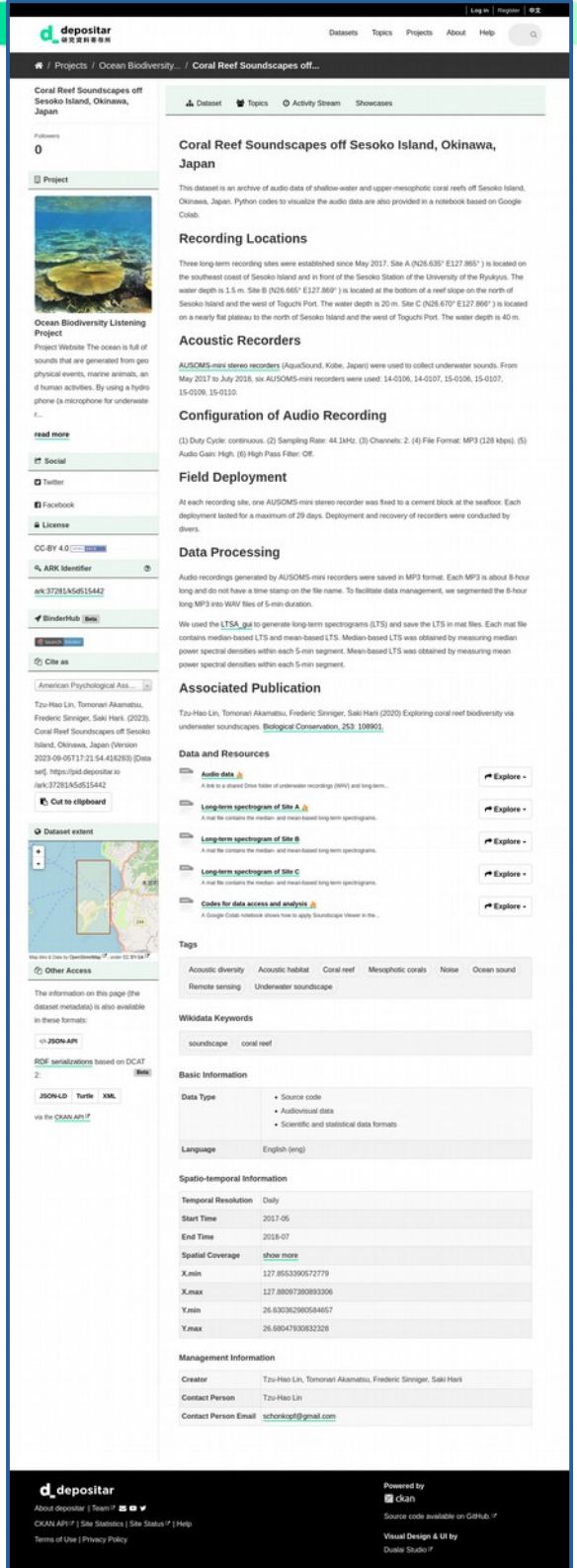
Outline

- A tour of *depositar* via a sample dataset
- *depositar*: A bottom-up data repository
- Repositories and “open science”
- Discussion



A tour of *depositar* via a sample dataset





A Sample Dataset

https://pid.depositar.io/ark:37281/k5d515442

A tour of the dataset

- Long description of the dataset and the project
- Data and (external) resources
- Tags and Wikidata keywords
- Basic information
- Spatio-temporal information
- Management information
- License
- Citation snippet
- Data endpoints
  - JSON-API
  - RDF serializations

Coral Reef Soundscapes off Sesoko Island, Okinawa, Japan

Followers

0

Project

Ocean Biodiversity Listening Project

Project Website The ocean is full of sounds that are generated from geophysical events, marine animals, and human activities. By using a hydrophone (a microphone for underwater... read more

Dataset | Topics | Activity Stream | History

Coral Reef Soundscapes off Sesoko Island, Okinawa, Japan

This dataset is an archive of audio data of shallow-water and upper-mesophotic coral reefs off Sesoko Island, Okinawa, Japan. Python codes to visualize the audio data were also provided in a notebook based on Google Colab.

Recording Locations

Three long-term recording sites were established since May 2017. Site A (N26.635° E127.865°) is located on the southeast coast of Sesoko Island and in front of the Sesoko Station of the University of the Ryukyus. The water depth is 1.5 m. Site B (N26.665° E127.869°) is located at the bottom of a reef slope on the north of Sesoko Island and the west of Toguchi Port. The water depth is 20 m. Site C (N26.670° E127.866°) is located on a nearly flat plateau to the north of Sesoko Island and the west of Toguchi Port. The water depth is 40 m.

Acoustic Recorders

AUSOMS-mini stereo recorders (AquaSound, Kobe, Japan) were used to collect underwater sounds. From May 2017 to July 2018, six AUSOMS-mini recorders were used: 14-0106, 14-0107, 15-0106, 15-0107, 15-0109, 15-0110.

Configuration of Audio Recording

(1) Duty Cycle: continuous. (2) Sampling Rate: 44.1kHz. (3) Channels: 2. (4) File Format: MP3 (128 kbps). (5) Audio Gain: High. (6) High Pass Filter: Off.

Field Deployment

At each recording site, one AUSOMS-mini stereo recorder was fixed to a cement

Social

Twitter

Facebook

License

CC-BY 4.0 OPEN DATA

ARK Identifier Beta

ark:37281/k5d515442

Cite as

American Psychological Asso…

Tzu-Hao Lin, Tomonari Akamatsu, Frederic Sinniger, Saki Harii. (2022). Coral Reef Soundscapes off Sesoko Island, Okinawa, Japan (Version 2022-07-11T10:14:54.103158) [Data set]. https://pid.depositar.io/ark:37281/k5d515442

Cut to clipboard

Dataset extent

(1) Duty Cycle: continuous. (2) Sampling Rate: 44.1kHz. (3) Channels: 2. (4) File Format: MP3 (128 kbps). (5) Audio Gain: High. (6) High Pass Filter: Off.

Field Deployment

At each recording site, one AUSOMS-mini stereo recorder was fixed to a cement block at the seafloor. Each deployment lasted for a maximum of 29 days. Deployment and recovery of recorders were conducted by divers.

Data Processing

Audio recordings generated by AUSOMS-mini recorders were saved in MP3 format. Each MP3 is about 8-hour long and do not have a time stamp on the file name. To facilitate data management, we segmented the 8-hour long MP3 into WAV files of 5-min duration.

We used the LTSA_gui to generate long-term spectrograms (LTS) and save the LTS in mat files. Each mat file contains median-based LTS and mean-based LTS. Median-based LTS was obtained by measuring median power spectral densities within each 5-min segment. Mean-based LTS was obtained by measuring mean power spectral densities within each 5-min segment.

Associated Publication

Tzu-Hao Lin, Tomonari Akamatsu, Frederic Sinniger, Saki Harii (2020) Exploring coral reef biodiversity via underwater soundscapes. Biological Conservation, 253: 108901.

Data and Resources

Audio data
A link to a shared Drive folder of underwater recordings (WAV) and long-term... — Explore

Long-term spectrogram of Site A
A mat file contains the median- and mean-based long-term spectrograms. — Explore

Long-term spectrogram of Site B
A mat file contains the median- and mean-based long-term spectrograms. — Explore

Long-term spectrogram of Site C
A mat file contains the median- and mean-based long-term spectrograms. — Explore

Codes for data access and analysis
Explore

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

Other Access

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

</>JSON-API

RDF serializations based on DCAT 2: Beta

</>JSON-LD | </>Turtle | </>XML

via the CKAN API

Tags

Acoustic diversity | Acoustic habitat | Coral reef | Mesophotic corals | Noise | Ocean sound | Remote sensing | Underwater soundscape

Wikidata Keywords

soundscape | coral reef

Basic Information

Data Type	• Source code • Audiovisual data • Scientific and statistical data formats
Language	English (eng)

Spatio-temporal Information

Temporal Resolution	Daily
Start Time	2017-05
End Time	2018-07
Spatial Coverage	show more
X.min	127.8553390572779
X.max	127.88097380893306
Y.min	26.630362980584657
Y.max	26.68047930832328

Management Information

Author	Tzu-Hao Lin, Tomonari Akamatsu, Frederic Sinniger, Saki Harii
Contact Person	Tzu-Hao Lin

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

其他存取方式

此頁面上的資訊 (資料集之後設資料) 也提供以下格式：

</>JSON-API

RDF 串列化輸出 (修改自 DCAT 2)： Beta

</>JSON-LD | </>Turtle | </>XML

經由 CKAN API

標籤

Acoustic diversity | Acoustic habitat | Coral reef | Mesophotic corals | Noise | Ocean sound | Remote sensing | Underwater soundscape

Wikidata 關鍵字

聲景 | 珊瑚礁

基本資訊

資料類型	• 原始碼 • 影音資料 • 科學與統計資料
語言	英文 (eng)

時空資訊

時間解析度	日
起始時間	2017-05
結束時間	2018-07
空間範圍	顯示更多
空間範圍.X.min	127.8553390572779
空間範圍.X.max	127.88097380893306
空間範圍.Y.min	26.630362980584657
空間範圍.Y.max	26.68047930832328

管理資訊

產製者	Tzu-Hao Lin, Tomonari Akamatsu, Frederic Sinniger, Saki Harii
聯絡人	Tzu-Hao Lin

Social

Twitter

Facebook

License

CC-BY 4.0 OPEN DATA

ARK Identifier Beta

ark:37281/k5d515442

Cite as

American Psychological Asso…

Tzu-Hao Lin, Tomonari Akamatsu, Frederic Sinniger, Saki Harii. (2022). Coral Reef Soundscapes off Sesoko Island, Okinawa, Japan (Version 2022-07-11T10:14:54.103158) [Data set]. https://pid.depositar.io/ark:37281/k5d515442

Cut to clipboard

Dataset extent

(1) Duty Cycle: continuous. (2) Sampling Rate: 44.1kHz. (3) Channels: 2. (4) File Format: MP3 (128 kbps). (5) Audio Gain: High. (6) High Pass Filter: Off.

Field Deployment

At each recording site, one AUSOMS-mini stereo recorder was fixed to a cement block at the seafloor. Each deployment lasted for a maximum of 29 days. Deployment and recovery of recorders were conducted by divers.

Data Processing

Audio recordings generated by AUSOMS-mini recorders were saved in MP3 format. Each MP3 is about 8-hour long and do not have a time stamp on the file name. To facilitate data management, we segmented the 8-hour long MP3 into WAV files of 5-min duration.

We used the LTSA_gui to generate long-term spectrograms (LTS) and save the LTS in mat files. Each mat file contains median-based LTS and mean-based LTS. Median-based LTS was obtained by measuring median power spectral densities within each 5-min segment. Mean-based LTS was obtained by measuring mean power spectral densities within each 5-min segment.

Associated Publication

Tzu-Hao Lin, Tomonari Akamatsu, Frederic Sinniger, Saki Harii (2020) Exploring coral reef biodiversity via underwater soundscapes. Biological Conservation, 253: 108901.

Data and Resources

Audio data
A link to a shared Drive folder of underwater recordings (WAV) and long-term...
Explore

Long-term spectrogram of Site A
A mat file contains the median- and mean-based long-term spectrograms.
Explore

Long-term spectrogram of Site B
A mat file contains the median- and mean-based long-term spectrograms.
Explore

Long-term spectrogram of Site C
A mat file contains the median- and mean-based long-term spectrograms.
Explore

Codes for data access and analysis
Explore

Biological Conservation 253 (2021) 108901

Contents lists available at ScienceDirect

Biological Conservation

journal homepage: www.elsevier.com/locate/biocon

Exploring coral reef biodiversity via underwater soundscapes

Tzu-Hao Lin [a,*], Tomonari Akamatsu [b,**], Frederic Sinniger [c], Saki Harii [c]

[a] Biodiversity Research Center, Academia Sinica, Taiwan
[b] The Ocean Policy Research Institute, The Sasakawa Peace Foundation, Japan
[c] Tropical Biosphere Research Center, University of Ryukyus, Japan

ARTICLE INFO

Keywords:
Ocean sound
Mesophotic corals
Remote sensing
Noise
Acoustic habitat
Acoustic diversity

ABSTRACT

Information on biodiversity is essential to evaluate the ecological status of coral reefs. Sounds produced by reef-associated organisms have been used as a biodiversity indicator. However, the interference from abiotic sounds and the lack of a comprehensive audio library have impeded effective evaluation. This study investigated the application of underwater soundscapes as a remote-sensing method to detect biological and anthropogenic activities. Using techniques including the visualization of long-duration recordings, source separation, and clustering, soundscapes were separated into sounds of anthropogenic and biological sources. Our results revealed the dynamics of biological sounds among coral reefs off Sesoko Island, Okinawa, Japan. Biological sounds were much more prominent in shallow-water reefs than in upper-mesophotic reefs, but their spectral features and compositions differed. The shallow-water reefs were dominated by broadband sounds of crustaceans and low-frequency transient fish calls, whereas the upper-mesophotic reefs were characterized by a diverse array of fish choruses and transient sounds. We also discovered that shipping noise heavily interfered with the soundscapes from the upper-mesophotic reefs and represented an invisible threat to life in the low-light habitat. The applied techniques of soundscape information retrieval revealed the distinct ecological status of coral reefs and the behavior change of sound-producing organisms in high temporal resolution. Implementation of soundscape monitoring can generate ecological information on habitat quality, reef biodiversity, human activities, and their interactions. Global collaboration on underwater soundscapes will establish a data-informed platform and help stakeholders assess the resilience of coral reefs to environmental and anthropogenic stressors.

1. Introduction

Marine ecosystems provide irreplaceable services and currently face significant pressures due to climate change, human disturbance, and excessive use of marine resources. The United Nations has recognized these threats and placed the conservation of marine ecosystems as one of its sustainable development goals (UN General Assembly, 2015). Coral reefs support various social and economic activities, such as fisheries, coastal protection, and tourism, of many maritime tropical and subtropical nations (Moberg and Folke, 1999; Barbier, 2017; Spalding et al., 2017; Woodhead et al., 2019). These benefits rely on the abundant biodiversity in coral reefs. However, coral reefs have undergone recurrent high-frequency bleaching episodes over the past 20 years due to increased sea surface temperatures (Hughes et al., 2017, 2018). Therefore, detailed information on the spatiotemporal changing patterns of marine biodiversity and interactions with human activities is crucial for the conservation management of coral reefs.

Biodiversity monitoring in coral reefs remains challenging, partially due to the distinct reef environments and their unique fish assemblages (Pearman et al., 2018; Dumalagan et al., 2019). A comprehensive and long-term assessment of reef biodiversity, environmental characteristics, and human activities may not be feasible because of limited resources for observation and survey opportunities, especially for developing regions or remote reefs. An underwater sensing system capable of monitoring the changing patterns of marine biodiversity, with the ability to diagnose potential risks due to environmental and anthropogenic stressors, is required for establishing management strategies of coral reefs and for providing alerts to the early-warning signs of ecosystem changes (Schmeller et al., 2017; Obura et al., 2019).

A potential solution for such an underwater sensing platform is through monitoring ocean sounds. One autonomous recorder can store long-duration audio recordings, with improved time resolution of

* Correspondence to: T.-H. Lin, Biodiversity Research Center, Academia Sinica, 128 Academia Road, Sec. 2, Nankang, Taipei 11529, Taiwan.
** Correspondence to: T. Akamatsu, The Ocean Policy Research Institute, The Sasakawa Peace Foundation, 1-15-16 Toranomon, Minato, Tokyo 105-8524, Japan.
E-mail addresses: lintzuhao@gate.sinica.edu.tw (T.-H. Lin), akamatsu.tom@gmail.com (T. Akamatsu).

https://doi.org/10.1016/j.biocon.2020.108901
Received 18 June 2020; Received in revised form 23 November 2020; Accepted 27 November 2020
Available online 10 December 2020
0006-3207/

With the recent development of underwater technology and audio information retrieval techniques, a soundscape monitoring network can generate numerous acoustic data that contain ecological information in multiple dimensions, including the quality of the acoustic habitat, community of sound-producing organisms, and potential effects due to human activities. The generated information will allow managers and stakeholders to conduct a more comprehensive assessment of ecosystem health at scale.

The feasibility of using 2-h daytime recordings to obtain rapid ecoacoustic indicators to characterize the major ecosystem functions of coral reefs is currently being discussed (Elise et al., 2019a). Our result suggests that beyond a spatial survey based on snapshot recordings, several long-term recording sites must be established within a spatial range of interest to allow a full understanding of the temporal variations of underwater soundscapes and ecosystem dynamics. The exceptional power of time-series recordings can help establish a library to document

Data availability

The audio dataset used in preparing this paper are available from the corresponding authors on reasonable request. A dataset of the LTS is available on depositar (https://data.depositar.io/en/dataset/coral-reef-sesoko).

CRediT authorship contribution statement

Tzu-Hao Lin: Conceptualization, Methodology, Software, Validation, Data curation, Formal analysis, Resources, Writing – original draft. **Tomonari Akamatsu**: Conceptualization, Methodology, Resources, Data curation, Writing – reviewing and editing, Funding acquisition. **Frederic Sinniger**: Conceptualization, Visualization, Investigation, Data curation, Writing – reviewing and editing. **Saki Harii**: Conceptualization, Investigation, Writing – reviewing and editing, Funding

Data availability

The audio dataset used in preparing this paper are available from the corresponding authors on reasonable request. A dataset of the LTS is available on depositar (https://data.depositar.io/en/dataset/coral-reef-sesoko).

Coral Reef Soundscapes off Sesoko Island, Okinawa, Japan

Followers

0

Project

Ocean Biodiversity Listening Project

Project Website The ocean is full of sounds that are generated from geophysical events, marine animals, and human activities. By using a hydrophone (a microphone for underwater... read more

Dataset | Topics | Activity Stream | History

Coral Reef Soundscapes off Sesoko Island, Okinawa, Japan

This dataset is an archive of audio data of shallow-water and upper-mesophotic coral reefs off Sesoko Island, Okinawa, Japan. Python codes to visualize the audio data were also provided in a notebook based on Google Colab.

Recording Locations

Three long-term recording sites were established since May 2017. Site A (N26.635° E127.865°) is located on the southeast coast of Sesoko Island and in front of the Sesoko Station of the University of the Ryukyus. The water depth is 1.5 m. Site B (N26.665° E127.869°) is located at the bottom of a reef slope on the north of Sesoko Island and the west of Toguchi Port. The water depth is 20 m. Site C (N26.670° E127.866°) is located on a nearly flat plateau to the north of Sesoko Island and the west of Toguchi Port. The water depth is 40 m.

Acoustic Recorders

AUSOMS-mini stereo recorders (AquaSound, Kobe, Japan) were used to collect underwater sounds. From May 2017 to July 2018, six AUSOMS-mini recorders were used: 14-0106, 14-0107, 15-0106, 15-0107, 15-0109, 15-0110.

Configuration of Audio Recording

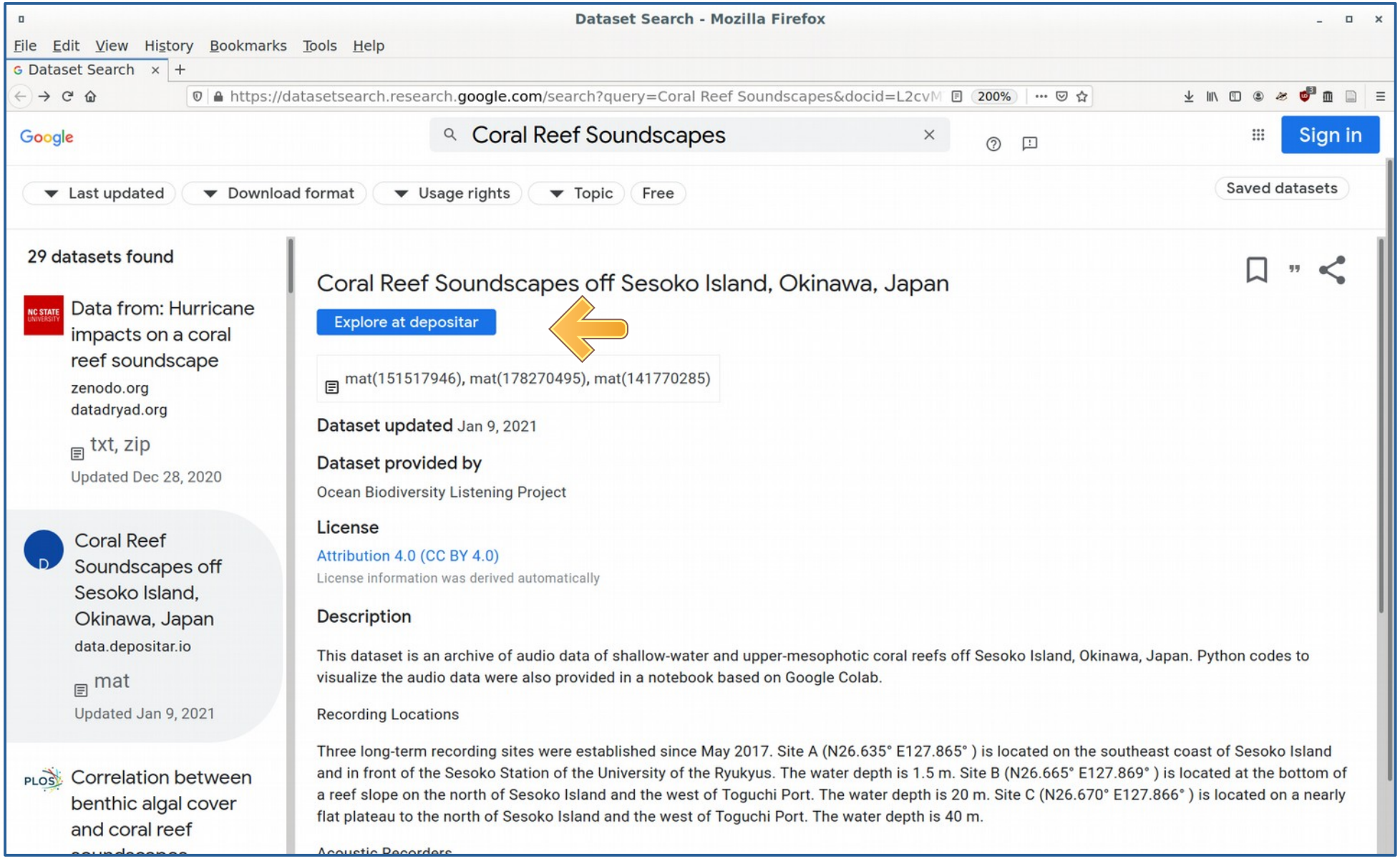
(1) Duty Cycle: continuous. (2) Sampling Rate: 44.1kHz. (3) Channels: 2. (4) File Format: MP3 (128 kbps). (5) Audio Gain: High. (6) High Pass Filter: Off.

Field Deployment

At each recording site, one AUSOMS-mini stereo recorder was fixed to a cement

Data Discovery via Google Dataset Search

https://datasetsearch.research.google.com/search?query=Coral Reef Soundscapes



Coral Reef Soundscapes off Sesoko Island, Okinawa, Japan

Followers

0

Project

Ocean Biodiversity Listening Project

Project Website The ocean is full of sounds that are generated from geophysical events, marine animals, and human activities. By using a hydrophone (a microphone for underwater... read more

Dataset | Topics | Activity Stream | History

Coral Reef Soundscapes off Sesoko Island, Okinawa, Japan

This dataset is an archive of audio data of shallow-water and upper-mesophotic coral reefs off Sesoko Island, Okinawa, Japan. Python codes to visualize the audio data were also provided in a notebook based on Google Colab.

Recording Locations

Three long-term recording sites were established since May 2017. Site A (N26.635° E127.865°) is located on the southeast coast of Sesoko Island and in front of the Sesoko Station of the University of the Ryukyus. The water depth is 1.5 m. Site B (N26.665° E127.869°) is located at the bottom of a reef slope on the north of Sesoko Island and the west of Toguchi Port. The water depth is 20 m. Site C (N26.670° E127.866°) is located on a nearly flat plateau to the north of Sesoko Island and the west of Toguchi Port. The water depth is 40 m.

Acoustic Recorders

AUSOMS-mini stereo recorders (AquaSound, Kobe, Japan) were used to collect underwater sounds. From May 2017 to July 2018, six AUSOMS-mini recorders were used: 14-0106, 14-0107, 15-0106, 15-0107, 15-0109, 15-0110.

Configuration of Audio Recording

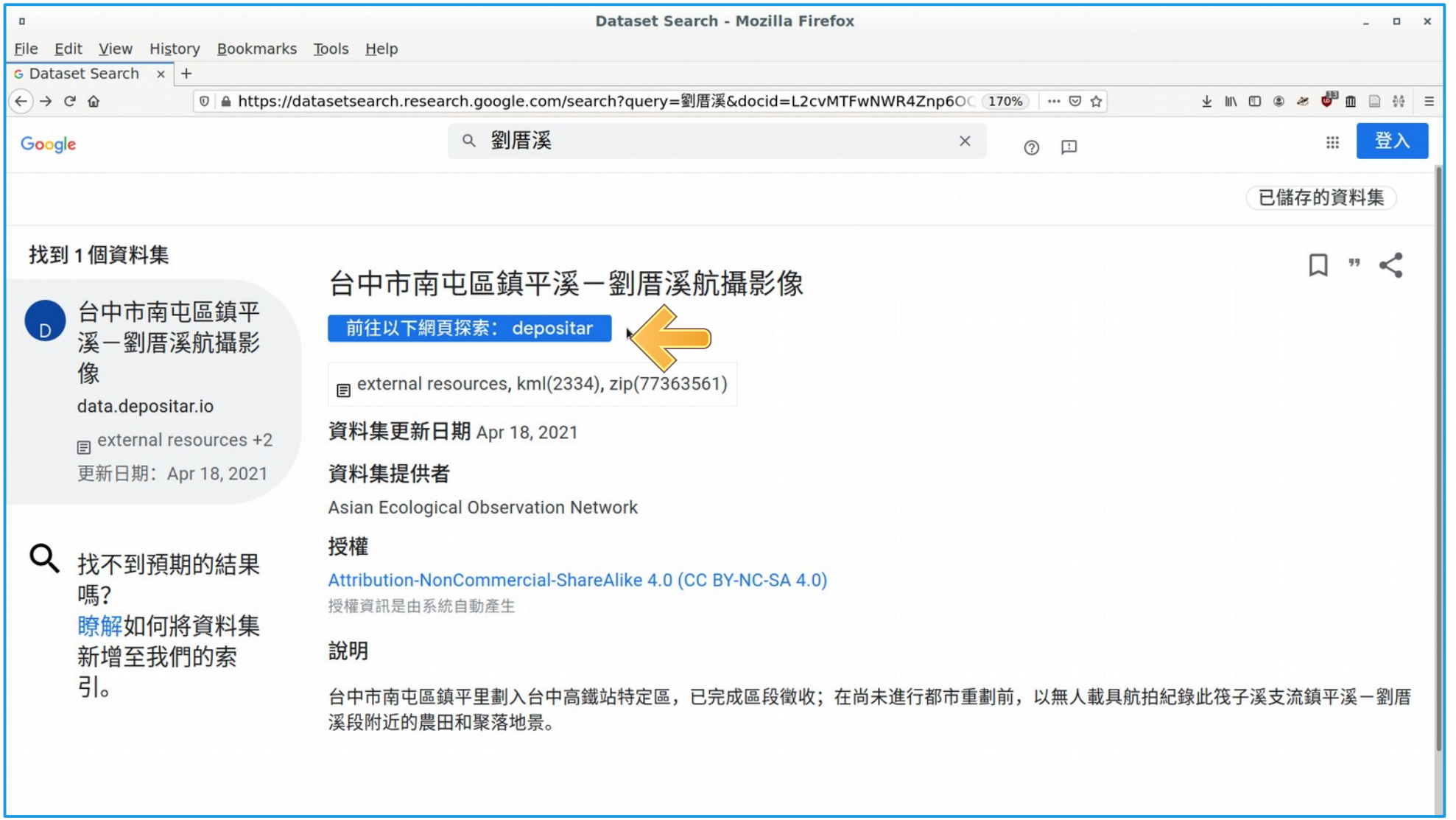
(1) Duty Cycle: continuous. (2) Sampling Rate: 44.1kHz. (3) Channels: 2. (4) File Format: MP3 (128 kbps). (5) Audio Gain: High. (6) High Pass Filter: Off.

Field Deployment

At each recording site, one AUSOMS-mini stereo recorder was fixed to a cement

More Google Dataset Search

https://datasetsearch.research.google.com/search?query=劉厝溪



https://pid.depositar.io/ark:37281/k5f912n4j

資料與資源

航拍規劃中心線

規劃航線中心線KML檔

探索

OAM正射影像連結 (Link to OpenAerialMap)

發布於OAM的正射影像連結 (Link to accessing the ortho-mosaics published on the...

探索

2021-04-13 正射影像Google圖磚 (Google Earth tiles)

Google圖磚壓縮檔。解壓縮後，點選開啟資料夾中的kml檔，即可使用Google Earth...

探索

2021-04-13 台中市南屯區鎮平溪－劉厝溪航攝影像

中央研究院網格計算中心WebODM計算成果下載連結；建議使用Firefox瀏覽器開啟連結，瀏覽影像2D、3D影像資料。

探索

空中360影像

空中360影像Google Street View連結。

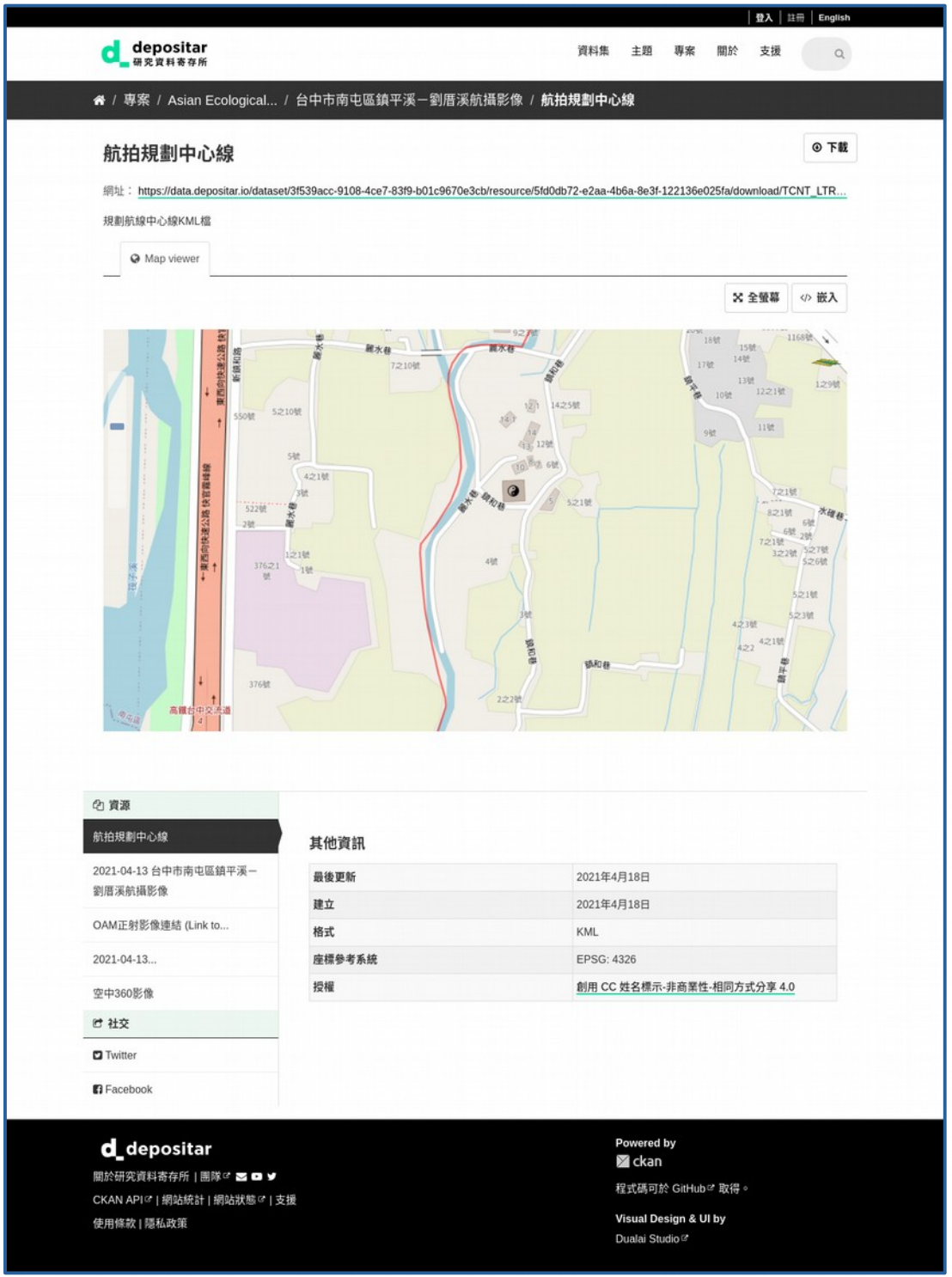
探索

標籤

南屯區　台中市　地景變遷　筏子溪　都市重劃

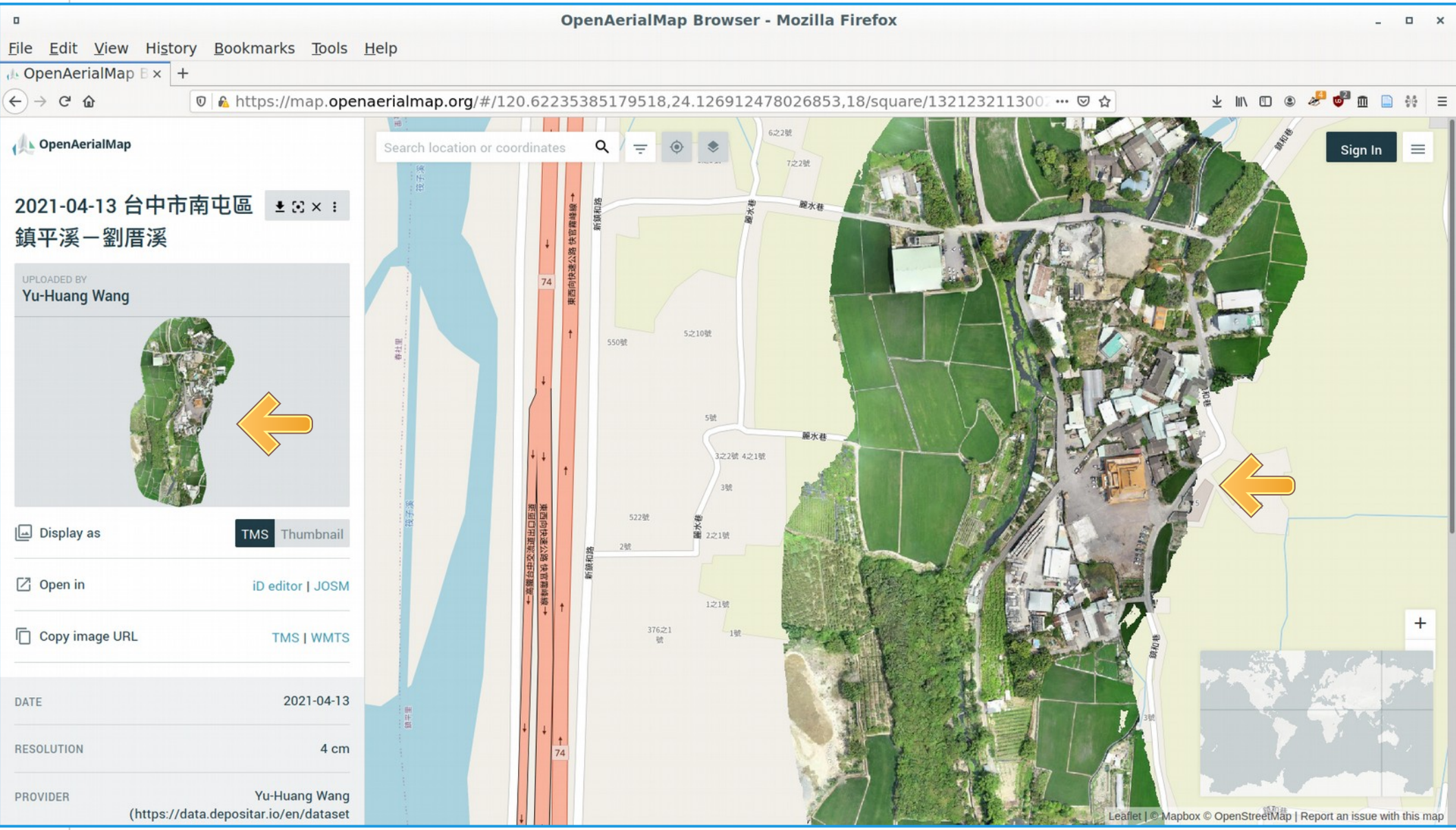
Wikidata 關鍵字

正射影像　riverscape　筏子溪　光球　南屯區　無人航空載具　臺中市



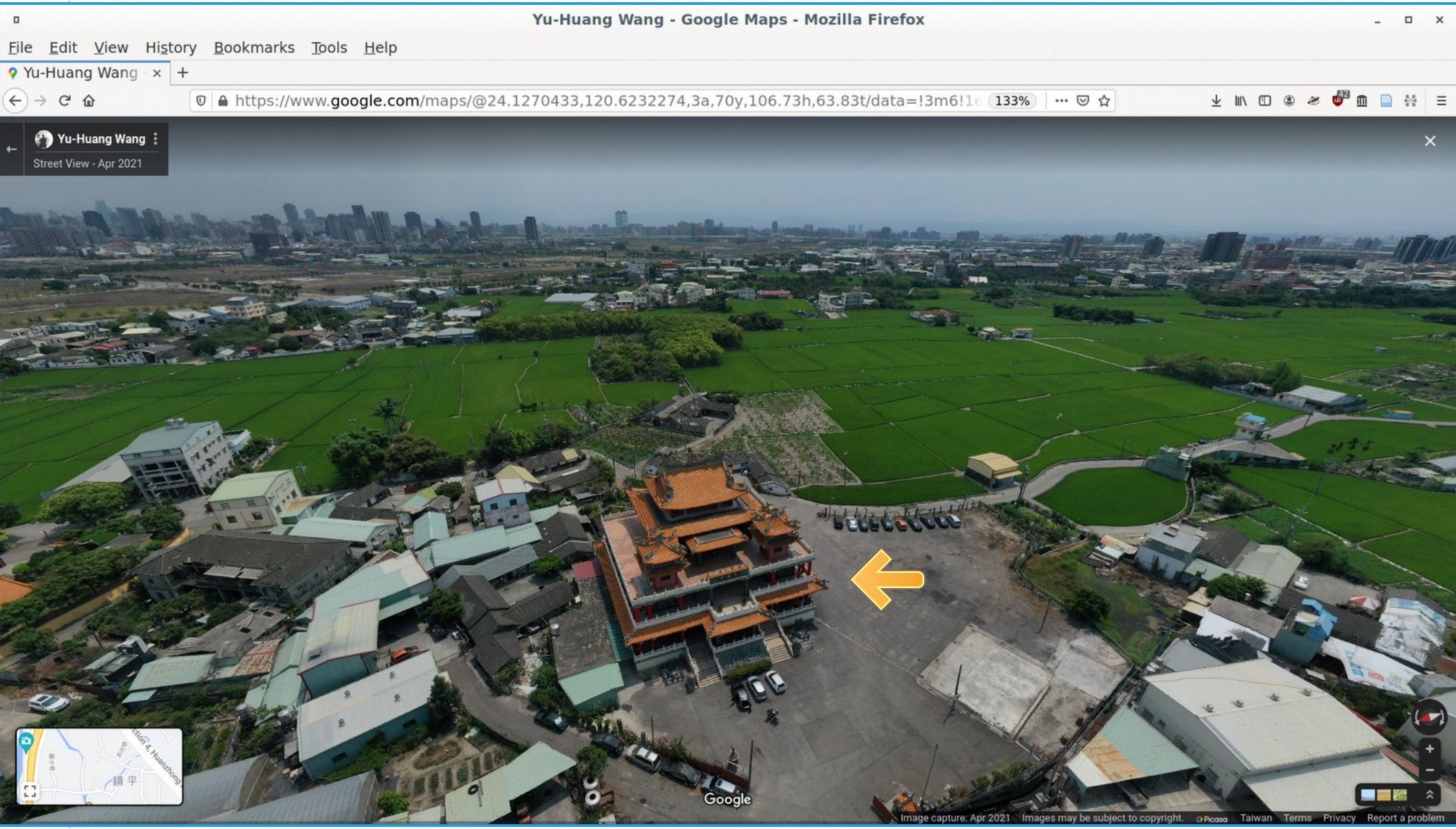
KML
(in browser preview)

Orthophoto (link to Open Aerial Map)



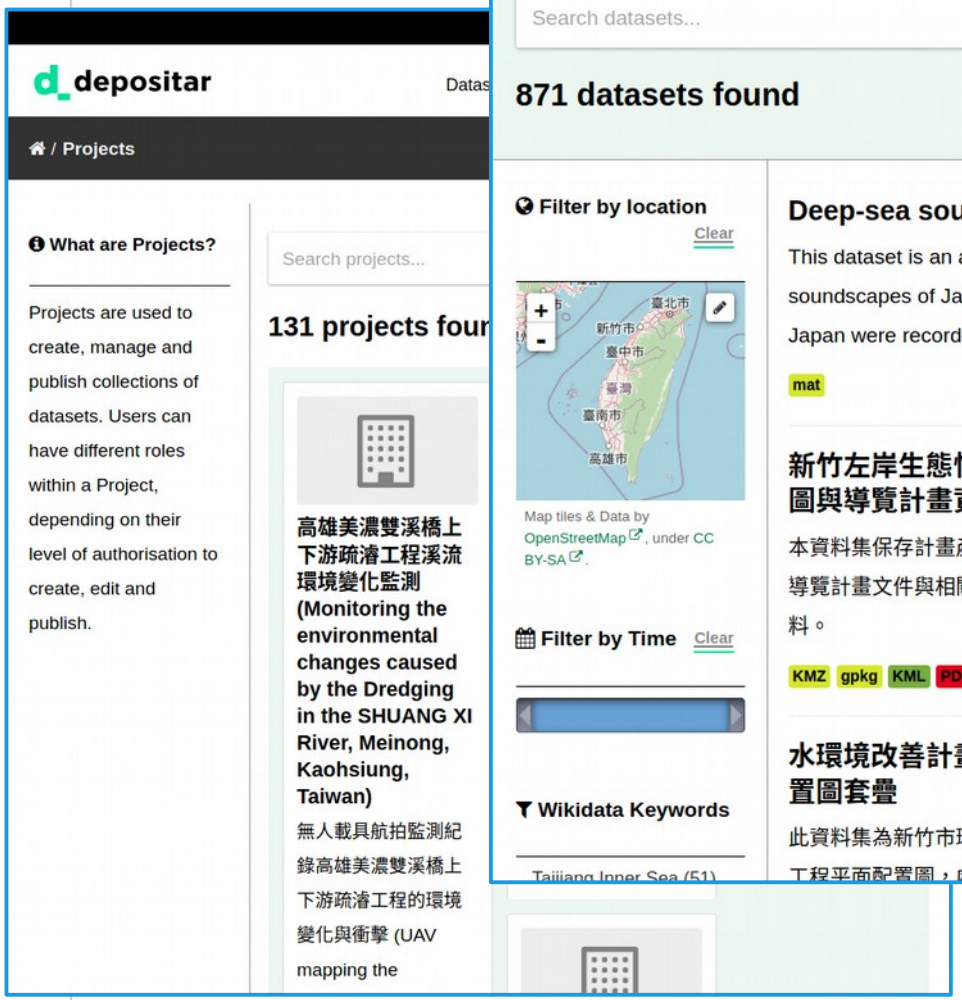
https://map.openaerialmap.org/#/120.62299489974974,24.127955278046894,15/square/13212321130023323/607c4f45bfb5350008e665a2

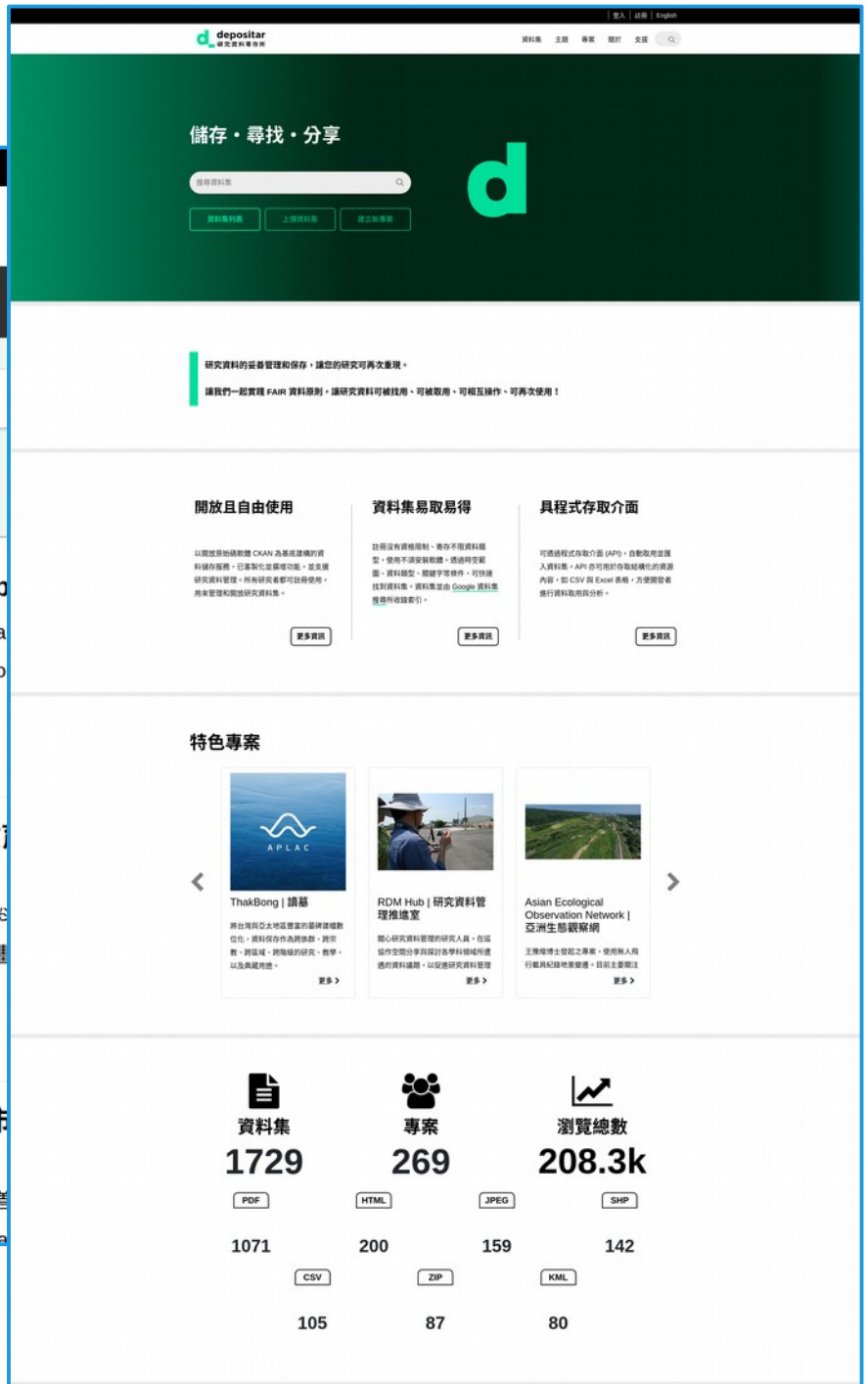
360° Panorama (link to Google Street View)



https://goo.gl/maps/zZZwQ3PkstQzrXYN7

https://data.depositar.io/





Collaborative Badlands | 惡地協作

作者

Collaborative Badlands

前往網站

展示案例包含的資料集

西南惡地點位資料庫

【英國攝影家湯姆生1871台灣線性文化遺產】

社交

Twitter

Facebook

Collaborative Badlands | 惡地協作

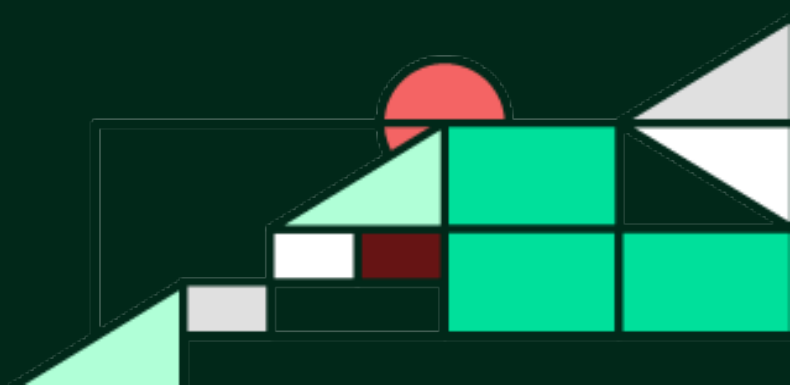
以台南左鎮為實驗場域，對外延伸至其他惡地形地區，包括台南龍崎、以及高雄田寮、內門等地區，建立地方研究、教學、服務等各項工作。由國立成功大學推動的協力計畫。

A collaborative project at the National Cheng Kung University on the rural regeneration and transdisciplinary research in the Southwestern Taiwan Badlands Region.

前往網站

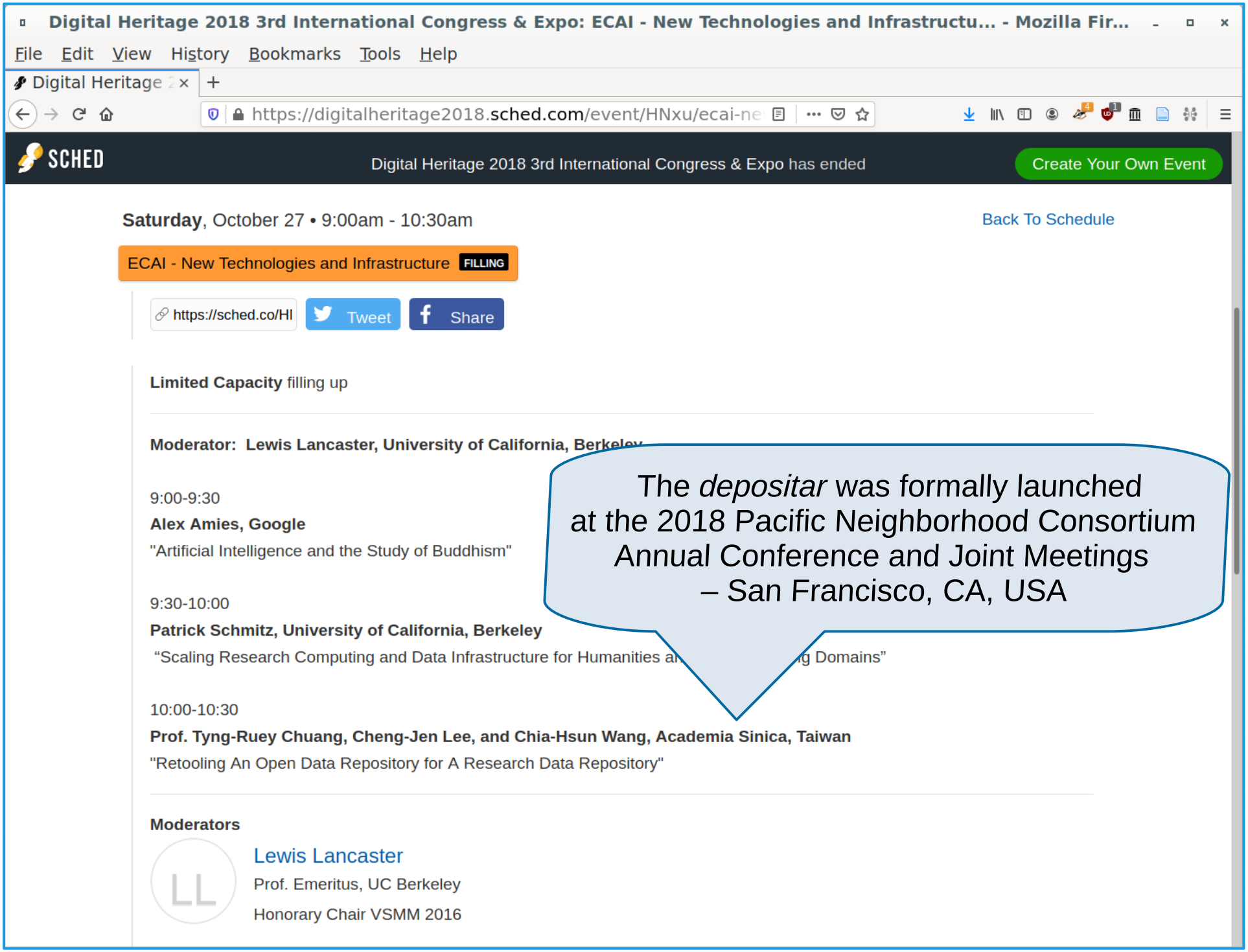
https://data.depositar.io/zh_Hant_TW/organization/about/collaborative-badlands-project

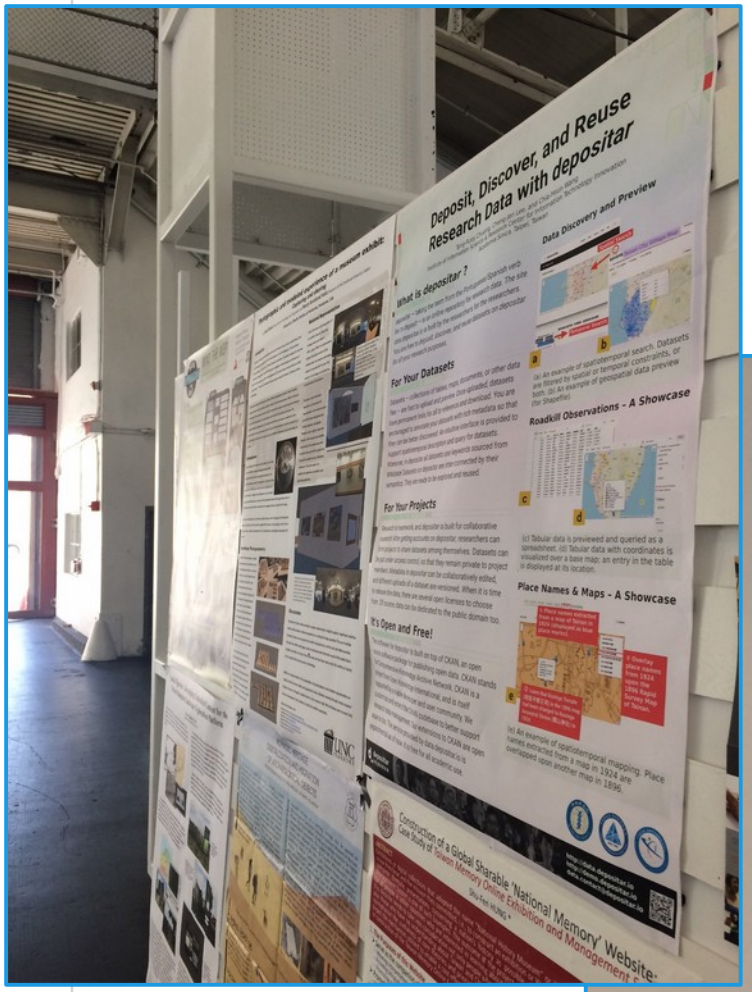
depositar: A bottom-up data repository

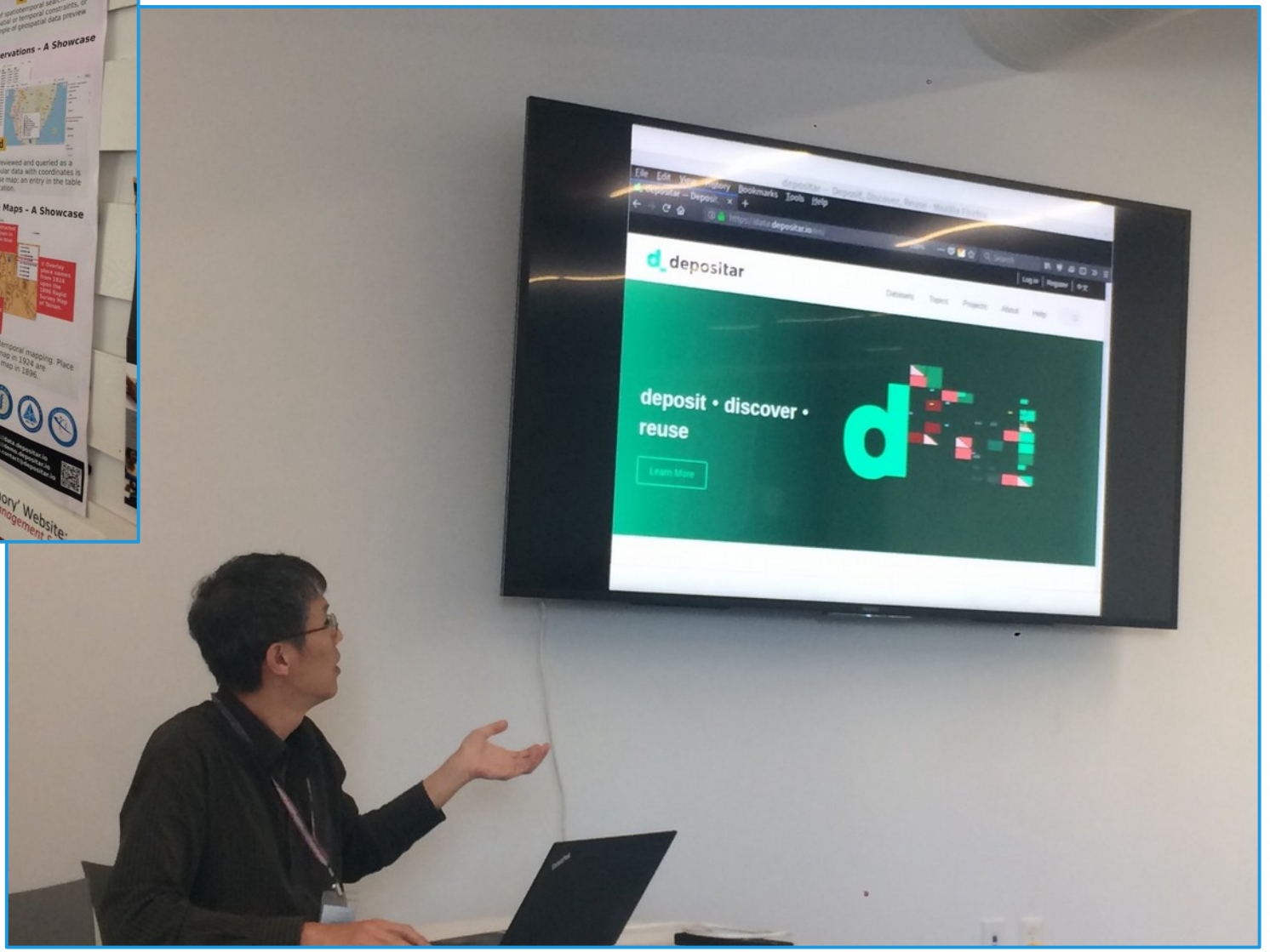


The road to *depositar*

- Developed to facilitate data sharing for two “integral” research projects funded by the Ministry of Science and Technology (MOST), Taiwan
  - 2013 – 2015 and 2016 – 2019
  - First presented at Open Repositories 2015
- Open to the public for general use in Oct 2018
  - Launched at the Pacific Neighborhood Consortium (PNC) 2018 Annual Conference, San Francisco, CA, USA
  - From “data sharing portal for some” to “data depository for all”
- Celebrating the five-year anniversary at PNC 2023, Okinawa, Japan
  - Progress report at Open Repositories 2021, 2022, 2023 too


Digital Heritage 2018 3rd International Congress & Expo: ECAI - New Technologies and Infrastructu... - Mozilla Fir...
File Edit View History Bookmarks Tools Help
Digital Heritage
https://digitalheritage2018.sched.com/event/HNxu/ecai-ne
SCHED
Digital Heritage 2018 3rd International Congress & Expo has ended
Create Your Own Event
Saturday, October 27 • 9:00am - 10:30am
Back To Schedule
ECAI - New Technologies and Infrastructure FILLING
https://sched.co/HI
Tweet
Share
Limited Capacity filling up
Moderator: Lewis Lancaster, University of California, Berkeley
9:00-9:30
Alex Amies, Google
"Artificial Intelligence and the Study of Buddhism"
9:30-10:00
Patrick Schmitz, University of California, Berkeley
"Scaling Research Computing and Data Infrastructure for Humanities a g Domains"
10:00-10:30
Prof. Tyng-Ruey Chuang, Cheng-Jen Lee, and Chia-Hsun Wang, Academia Sinica, Taiwan
"Retooling An Open Data Repository for A Research Data Repository"
Moderators
LL
Lewis Lancaster
Prof. Emeritus, UC Berkeley
Honorary Chair VSMM 2016
The depositar was formally launched
at the 2018 Pacific Neighborhood Consortium
Annual Conference and Joint Meetings
– San Francisco, CA, USA


Deposit, Discover, and Reuse Research Data with depositar
What is depositar ?
For Your Datasets
For Your Projects
It's Open and Free!
Data Discovery and Preview
Roadkill Observations - A Showcase
Place Names & Maps - A Showcase
Construction of a Global Sharable 'National Memory' Website:


depositar
deposit • discover • reuse
Learn More

p. 22, “Retooling An Open Data Repository For A Research Data Repository” slideset for the ECAI Workshop in PNC 2018.

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.

Organic bottom-up growth

- Uptake by NPOs (non-profit organizations) since 2019
  - *e.g.* from the groups monitoring ecological impacts resulted from governmental construction works
- Two grants from MOST (NSTC) with some focus on Research Data Management (2019 – 2022, 2022 – 2025)
  - interacting with other project teams working on Sustainability Research and on Long-Term Social and Ecological Research (LTSER research stations)
- Use by researchers at Academia Sinica, Taiwan
  - as well as by others in Taiwan and from abroad
- Formalized *Terms of Use* and *Privacy Policy*, July 2021

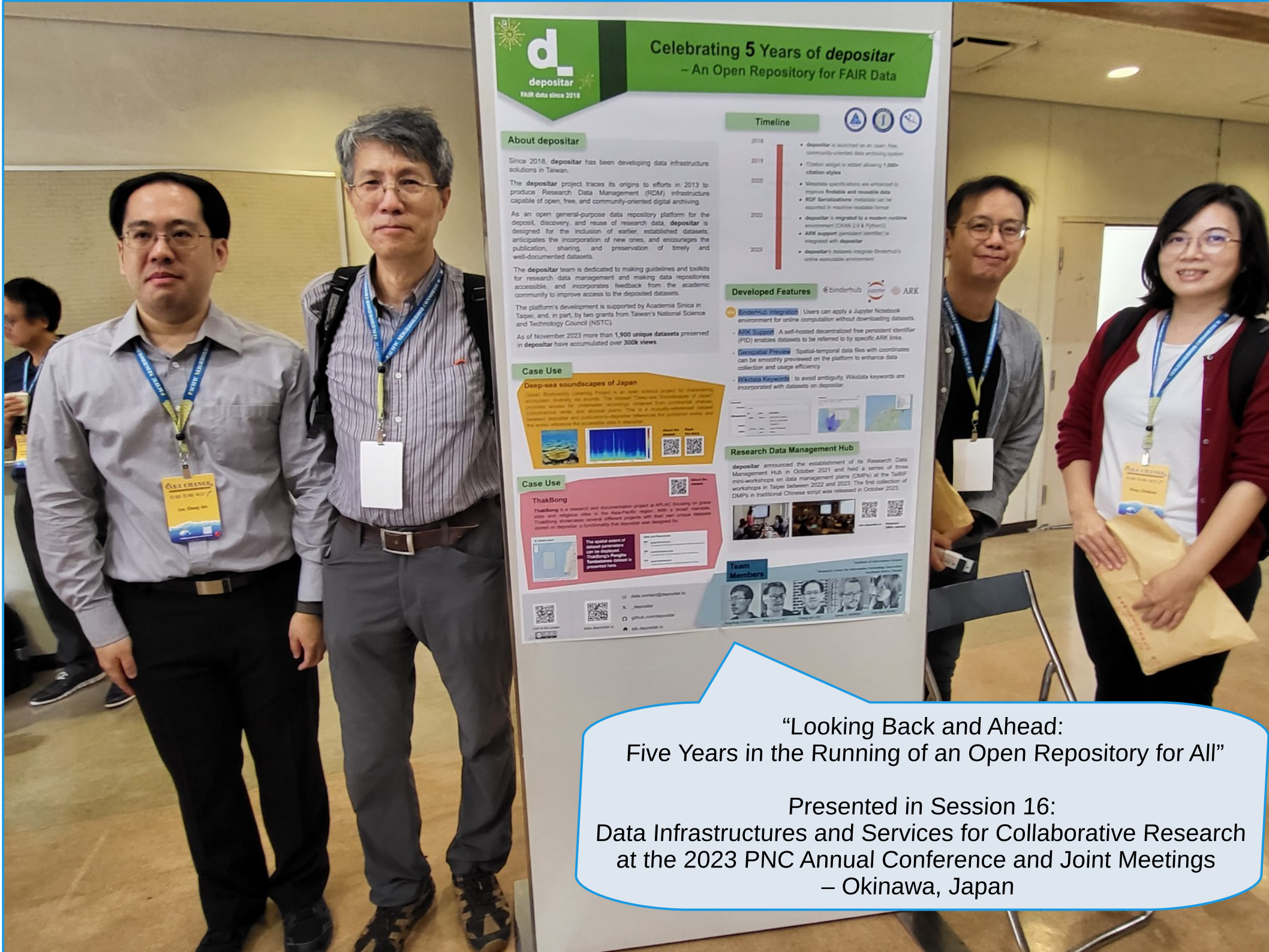
Workshops co-organized by NPOs on Research Data Managment (RDM)

Translation of the RDM Guide from Science Europe

Research Data Management Hub (RDM Hub)

https://rdm.depositar.io/

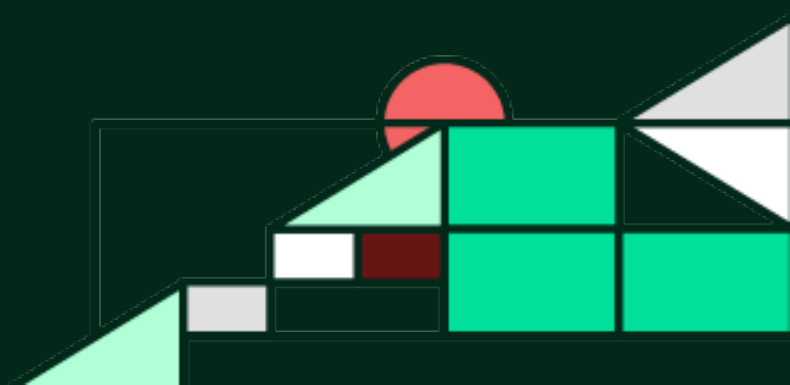
- Resources
- News and Reporting
- Communities of Practices



"Looking Back and Ahead:
Five Years in the Running of an Open Repository for All"

Presented in Session 16:
Data Infrastructures and Services for Collaborative Research
at the 2023 PNC Annual Conference and Joint Meetings
– Okinawa, Japan

depositar: A bottom-up data repository



Building on top of CKAN

- CKAN from OKFN: open source and in Python
  - repurposed and customized for research datasets
  - enriched metadata & connected to more resources
- Added features
  - Wikidata (2017)
  - Citation Widget (2018)
  - RDF Serialization (2020)
  - Archival Resource Key (2022 Q2)
  - BinderHub Integration (2023 Q4)

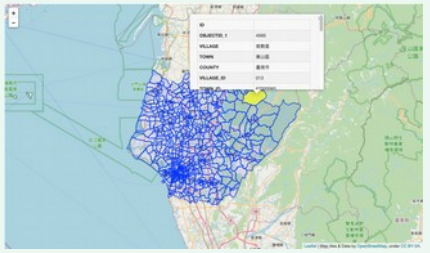
Feature Highlights

Spatio-temporal Search

Google site:data.depositar.io

Dataset Search

ARK Persistent Identifier



Data Previewers

W3C DCAT-based Metatada

WIKIDATA

Wikidata Keywords

depositar 研究資料寄存所

FAIR data since 2018

Project Management

/api/3/action/

JSON Data API

JSON-LD | XML | Turtle

RDF Serializations

jupyter

BinderHub Integration

Dataset Citation CSL

ckan

Open Data License Widget

Why Use Wikidata?

Are we flying from Taiwan to Japan or flying from Japan to Taiwan?

「松山飛松山」首航 華航創先例

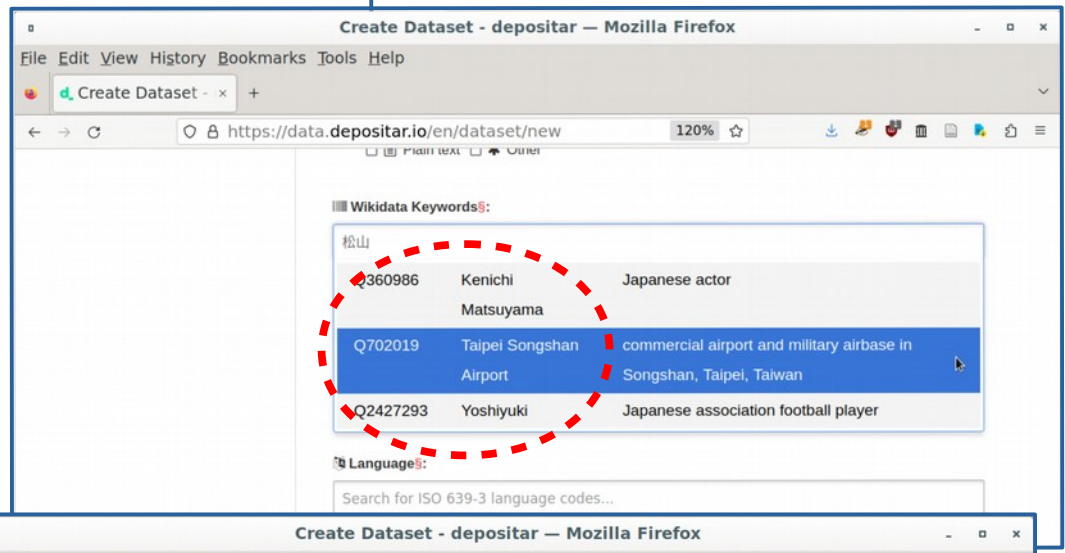
台灣松山機場往日本松山機場的同名航線今首航！（圖擷取自年代新聞台）

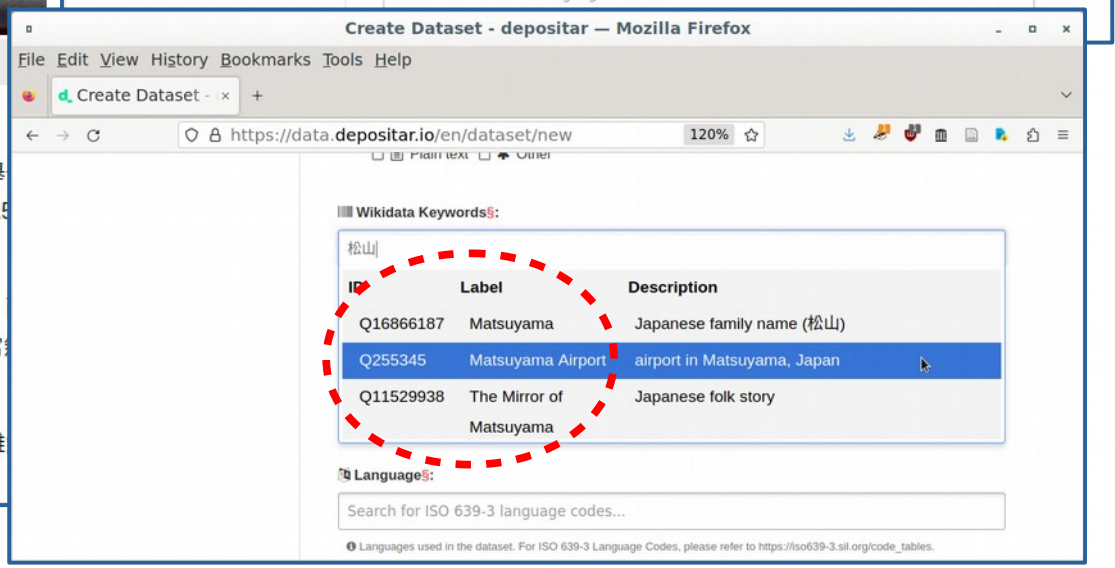
2013/10/11 11:33

〔即時新聞／綜合報導〕台灣松山機場往日本松山機場的同名航線今首航！此舉
機場後，破例開放二線城市日本愛媛松山機場航線，首航班機採包機制，共有15
民眾紛紛表示，全球首例當然要捧場，不過最重要的還是只需2小時就到日本。

中華航空公司創下全球首例，加開台灣松山機場對飛日本松山機場的對飛航線，
多將有158位幸運乘客創下歷史先例，搶到首航機票的民眾直呼，既然是首航當
了，只要兩小時。

另外，台灣松山機場也破格開放二線城市機場對飛，對此，搭機乘客則認為，雖
本都逛大城市，藉機來趟祕境之旅也會有不一樣的感受。





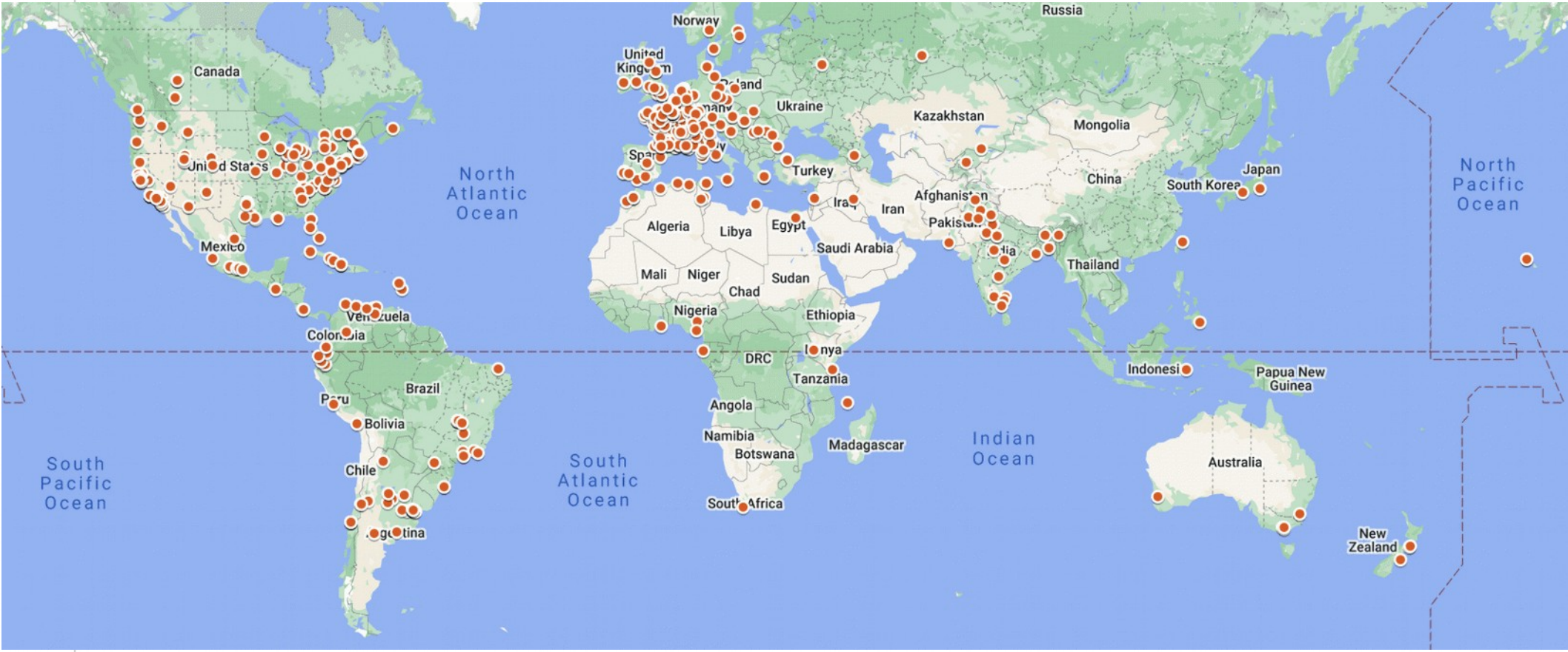
Language	Label	Description	Also known as
English	Taiwan	country in East Asia	Republic of China Chunghwa Minkwo Chunghwa Minkuo Formosa Nationalist China Free China Free area of the Republic of ... Republic of China (Taiwan) Chinese Taipei 🇹🇼 TW ROC Taiwan ROC China, Republic TWN TPE Nanasian Taywan Taivang Ilaod Taivangu The ROC
Traditional Chinese	中華民國	東亞國家	自由中國 臺灣 臺澎金馬 中華台北 ROC 台灣 民國 中華民國臺灣 ROC 台灣 中國（1970年前）

https://www.wikidata.org/wiki/Q865

Why Use ARK (Archival Resource Key)?

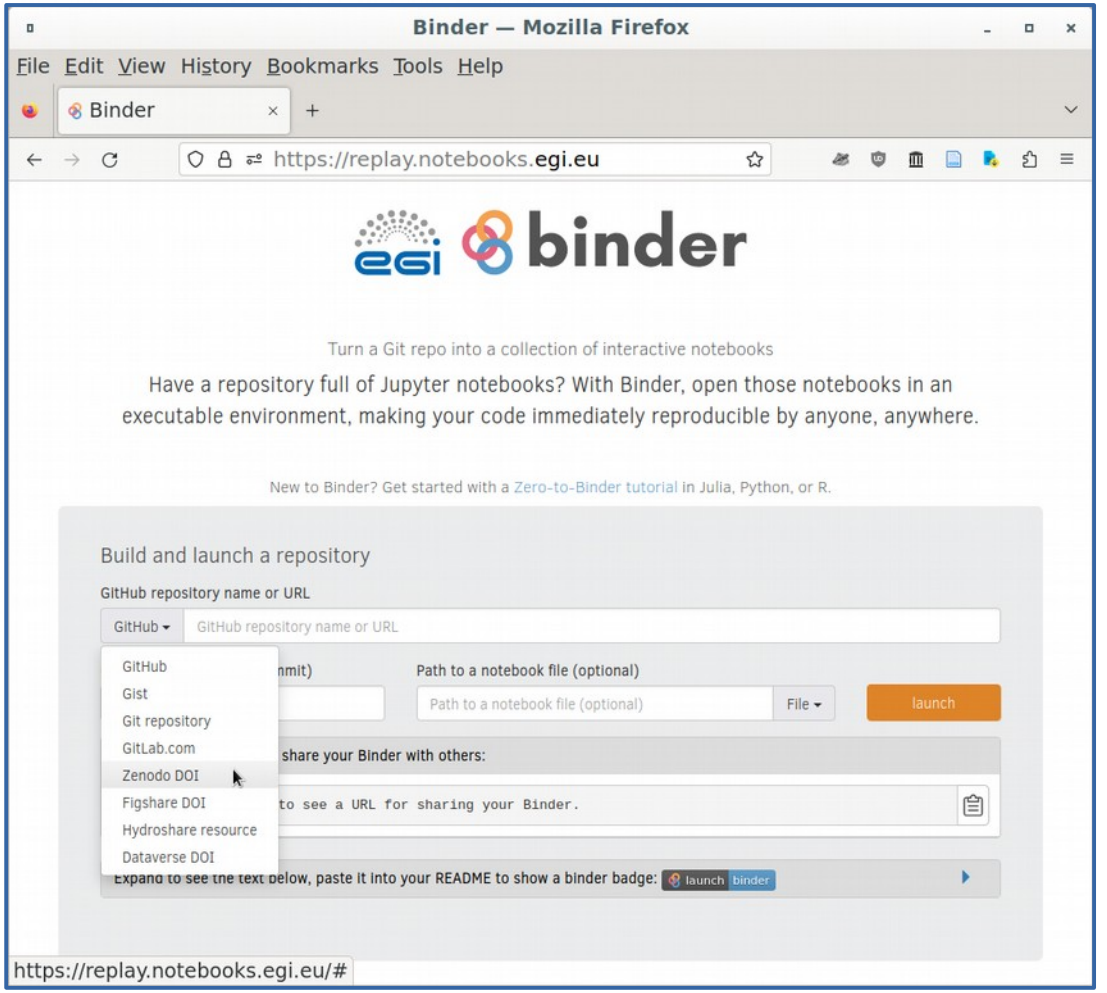
Issue Persistent Identifiers (PIDs) for Free and Support Each Others in Resolving Them!

- Institute of Information Science, Academia Sinica, mints and resolves its own ARKs … so are BnF, Caltech, Crossref, Internet Archive, Smithsonian Institution, …
- ARKs not issued by ourselves can be passed to their minters to resolve them…
- https://n2t.net/ark:37281/k5d951q1h →
  https://pid.depositar.io/ark:37281/k5d951q1h →
  https://data.depositar.io/dataset/binder-example-sea-turtle-sightings-in-taiwan



Bringing Computation to Data Repositories

- Our inspirations
  - EGI Notebook and Replay
  - GESIS Notebooks
  - Code Package Function from NII (Japan)
- Repositories for reproducible research
  - Article, code, and data are all open access and at the same place in a repository
  - Analytical results in the article can be automatically generated with the code and the data (by clicking a single button)
  - The repository coordinates computational and storage resources that are needed for the reproducibility of the results



Online Interactive Computing: Support and Challenge

- Jupyter (Jupyter Notebook, JupyterLab)
  - A web-based interactive computing platform
  - Support many kernels (programming languages)

- JupyterHub
  - A multi-tenancy JupyterLab service for group of users

- The current JupyterHub service for specific data partners
  - Limited and inelastic processing power as a single server
  - The need for downloading the data from the repository and then uploading it to Jupyter

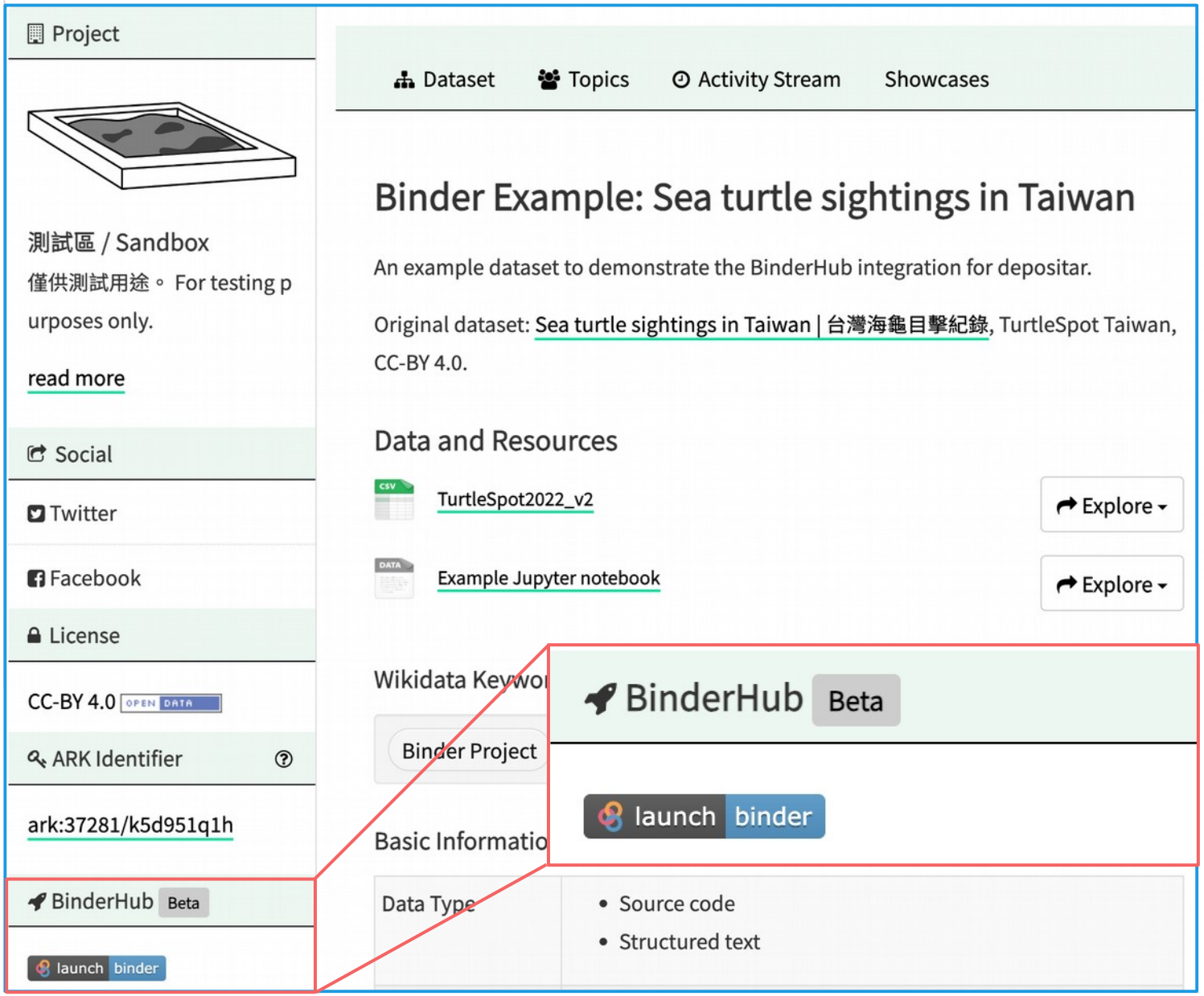
BinderHub Integration for *depositar*

- BinderHub: establish a JupyterHub in the Kubernetes (k8s) environment to create Jupyter notebooks from data repositories such as GitHub, Zenodo, or Dataverse
  - MyBinder.org: a public Binder service

- BinderHub integration for *depositar* (available as of October 2023)
  - binder.depositar.io: a customized Binder service with a “CKAN provider” to support datasets on the *depositar*
  - Launch a Jupyter environment containing resources from a dataset by **clicking a button** on the dataset page without downloading the resources
  - Highly scalable thanks to the k8s technology

Demo

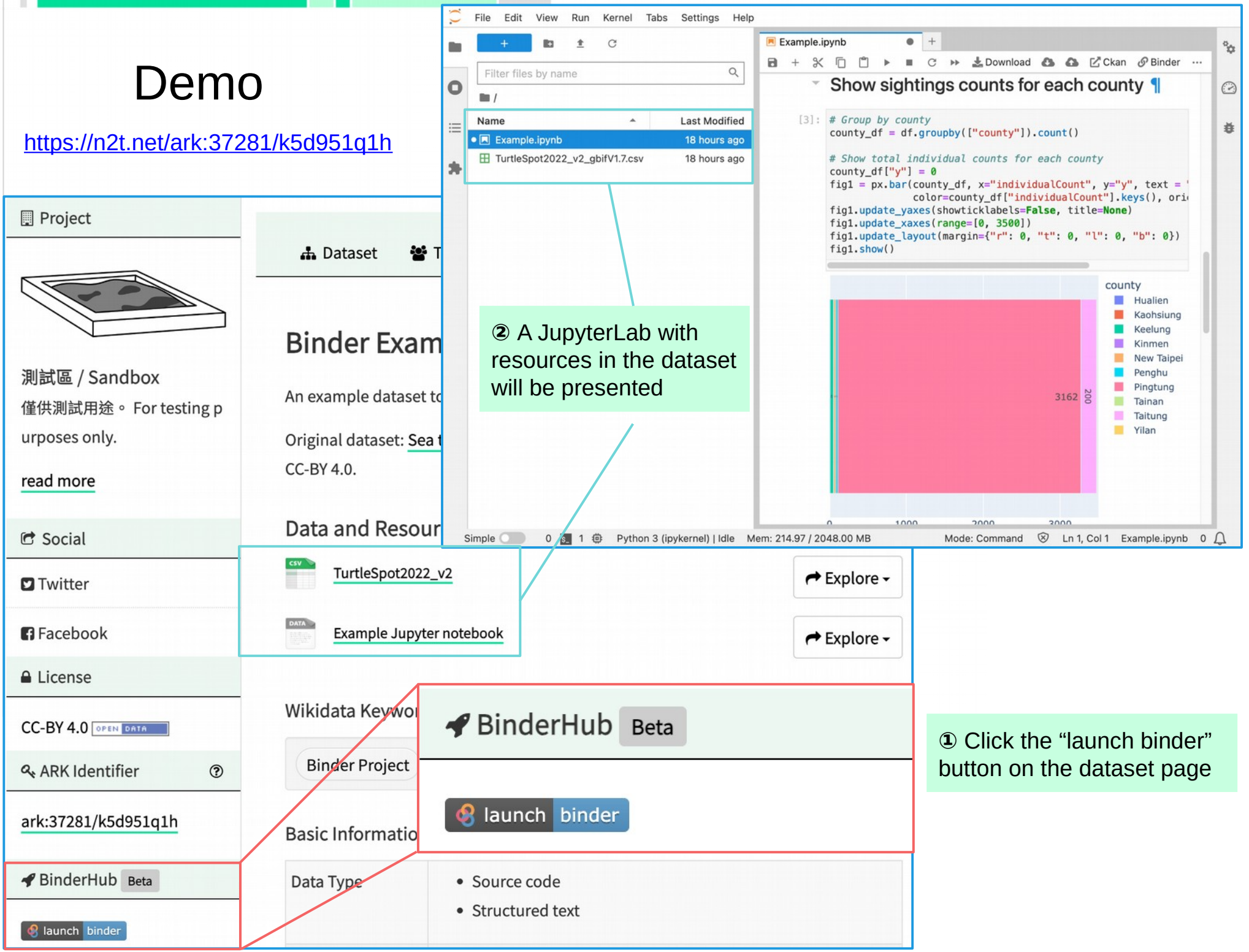
https://n2t.net/ark:37281/k5d951q1h



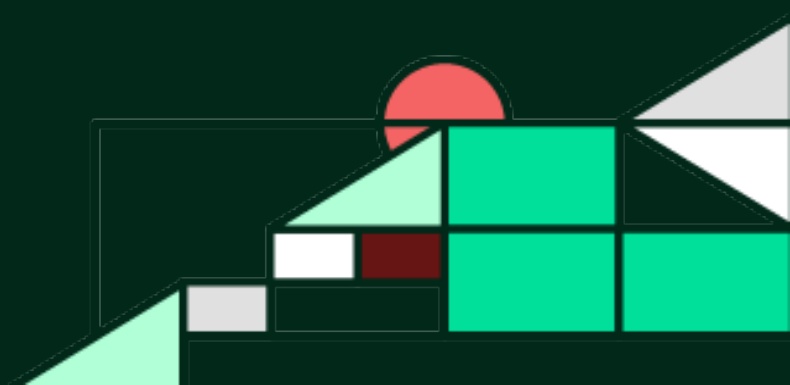
① Click the “launch binder” button on the dataset page

Demo

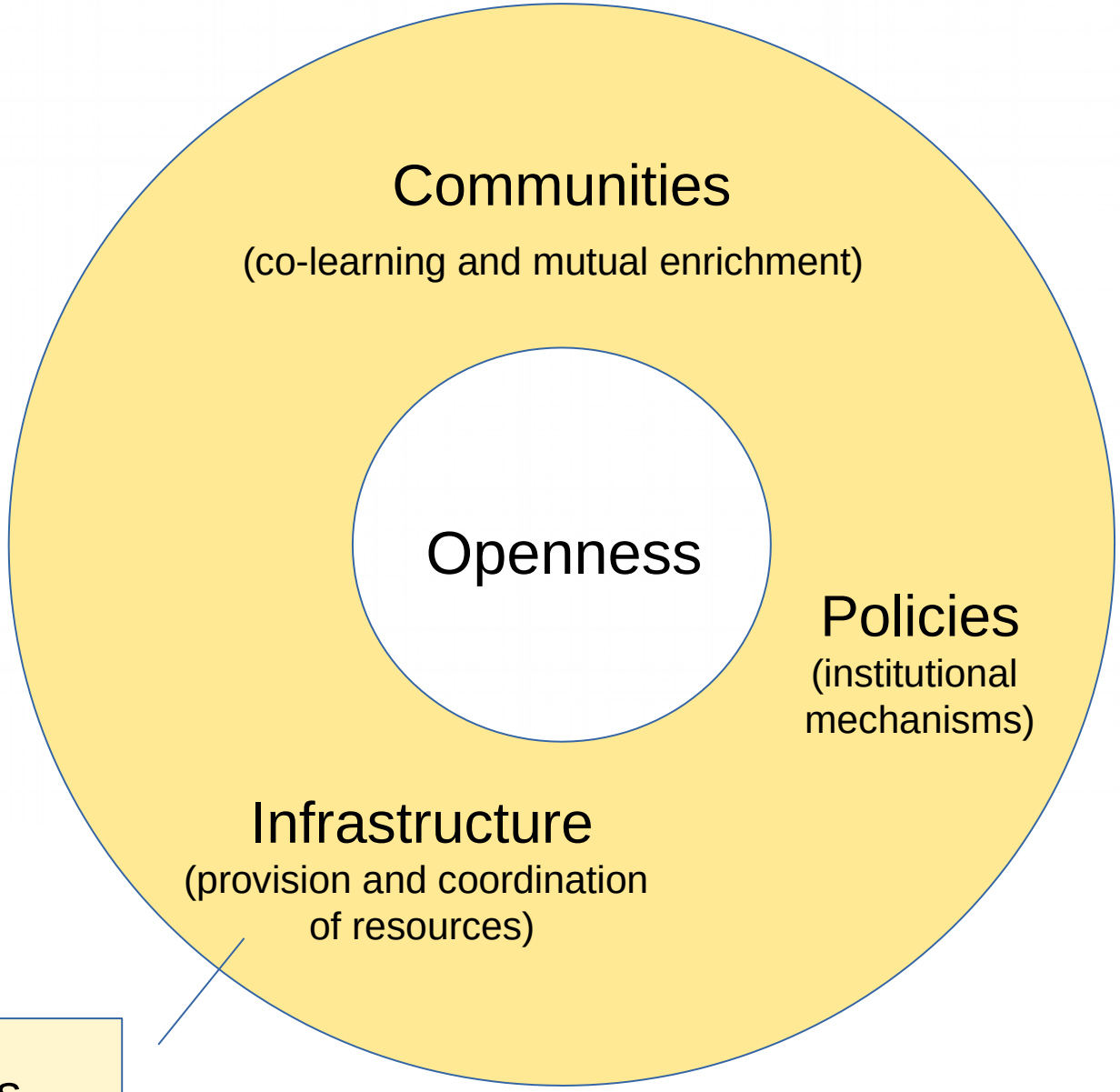
https://n2t.net/ark:37281/k5d951q1h



Repositories and “open science”



Repositories in Open Source/Data/Science…



Repositories

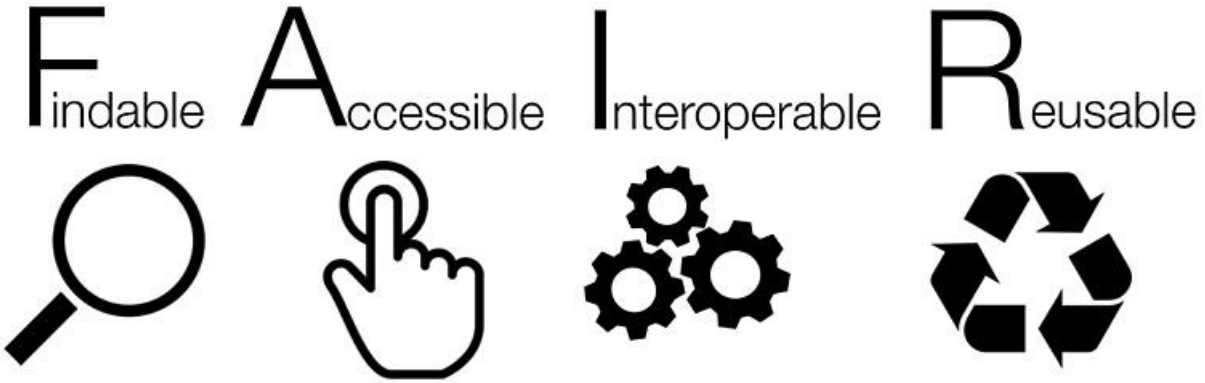
The FAIR Data Principle

Findable 可被找到	人或機器能輕鬆找到資料或後設資料。	F1. 配發持續識別碼 (PIDs) F2. 使用豐富的後設資料 (metadata) F3. 資料在可搜尋的資源中註冊或索引 F4. 後設資料指明資料的識別碼
Accessible 可被取用	人或機器可存取或下載資料或後設資料。	A1. 資料的識別碼使用標準化的通訊協定 A2. 即使資料不可取得，後設資料仍可得
Interoperable 可相互操作	不同來源的資料可被交換或結合。	I1. 使用正規、共享及廣泛使用的語言以達成知識再現 (e.g. RDF, JSON-LD) I2. 使用符合 FAIR 原則的語彙 (vocabulary) I3. 資料包括對其它資料的合適參照
Reusable 可再次使用	資料能容易被未來的研究複製或結合。	R1. 以精確且相關的多種屬性，豐富地描述（後設）資料 R1.1 以清晰、可取用的資料授權條款釋出（後設）資料 R1.2 （後設）資料連結到詳細的溯源資訊 R1.3. （後設）資料符合該領域相關的社群標準

https://force11.org/info/the-fair-data-principles/

The TRUST Principles for Digital Repositories

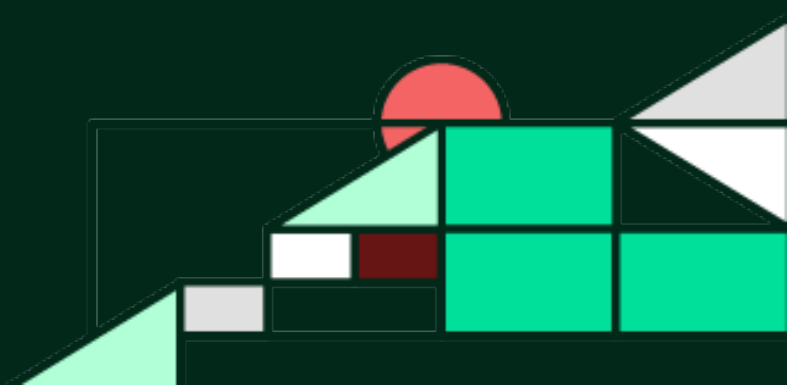
- Transparency 服務透明
  - 儲存庫的服務與其所保有的資料等資訊是透明的，且可被公眾取用並驗證。
  - To be transparent about specific repository services and data holdings that are verifiable by publicly accessible evidence.
- Responsibility 確保責任
  - 確保資料的真實性和完整性，並對於所提供服務的可靠性與持續性負起責任。
  - To be responsible for ensuring the authenticity and integrity of data holdings and for the reliability and persistence of its service.
- User Focus 以使用者為重
  - 確保使用者社群的資料管理需求獲得滿足，並關照社群的期待。
  - To ensure that the data management norms and expectations of target user communities are met.
- Sustainability 永續發展
  - 服務永續與資料保存是長期性的工作。
  - To sustain services and preserve data holdings for the long-term.
- Technology 基礎設施與技能
  - 提供基礎設施與技能以支撐安全、持續和可靠的服務。
  - To provide infrastructure and capabilities to support secure, persistent, and reliable services.



Data and Repositories

https://doi.org/10.1038/sdata.2016.18
https://doi.org/10.1038/s41597-020-0486-7
https://commons.wikimedia.org/wiki/File:FAIR_data_principles.jpg
https://www.rd-alliance.org/trust-principles-rda-community-effort

Discussion



Discussion

- Paths to sustainability
  - Tools (keep no data) →
    Systems (keep our own data) →
    Services (keep other people's data)
  - Persistent services and/or persistent datasets?
- Collaborations among diverse data communities
- Pooling resources together

depositar
研究資料寄存所

Thank You!

https://data.depositar.io/
https://rdm.depositar.io/
https://lab.depositar.io/

data.contact@depositar.io

The *depositar* is a collaboration at the Institute of Information Science, the Research Center for Information Technology Innovation, and the Research Center for Humanities and Social Sciences (GIS Center) in Academia Sinica, Taiwan. The project has been supported, in part, by grants from Taiwan's National Science and Technology Council.

「研究資料寄存所」是中央研究院資訊科學研究所、資訊科技創新研究中心、人文社會科學研究中心（地理資訊科學研究專題中心）的協作專案，部份經費來自台灣國科會的專題研究計畫。

depositar
研究資料寄存所

Please Write to Us 請與我們聯繫

data.contact@depositar.io

We Are Hiring! 人員招募中!

企劃推廣、系統發展、技術文件撰寫、……

Training and Outreach, System Development, Tech Writing,...