

depositor
研究資料寄存所

資料寄存實作

2023 研究資料管理工作坊

2023-12-21

李承鑫

Cheng-Jen Lee

台灣 中央研究院

Academia Sinica – Taiwan





大綱

- 新增資料集與資源
- 新增資源檢視
- 使用資料 API
- 操作練習

研究資料寄存所 操作手冊

- <https://docs.depositar.io/zh-tw/stable/user-guide.html>

- 操作手冊

- CKAN 簡介

- 資料集與資源

- 使用本平台

- 註冊帳號與登入

- 資料發布功能

- 新增資料集
- 本平台補充功能——將資料集加入選定的主題
- 本平台補充功能——填寫輔助功能介紹
- 編輯資料集
- 資料集協作者
- 新增、刪除及修改資源
- 刪除資料集
- 建立專案
- 管理專案
- 邀請專案成員（共同協作者）

- 查找資料

- 全站搜尋
- 本平台補充功能——時間搜尋功能介紹
- 本平台補充功能——空間搜尋功能介紹
- 搜尋專案內的資料集
- 探索資料集
- 本平台補充功能——引用資料集
- 本平台補充功能——資料預覽功能介紹

- 資料 API

- RDF Serializations (序列化輸出)

- 方法一：RDF Endpoints
- 方法二：內容協商 (Content Negotiation)
- 範例

- ARK 持續識別碼

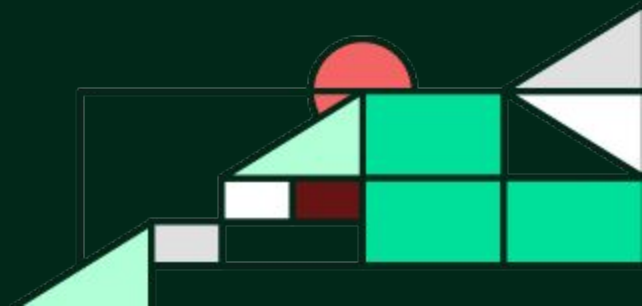
- ARK 識別碼發放條件
- 透過 ARK 識別碼連結至資料集
- ARK 識別碼後設資料 (ERC)
- 已失效 ARK 識別碼

- 個人化設定

- 新聞消息來源
- 帳號資訊

- 系統限制

一、新增資料集與資源

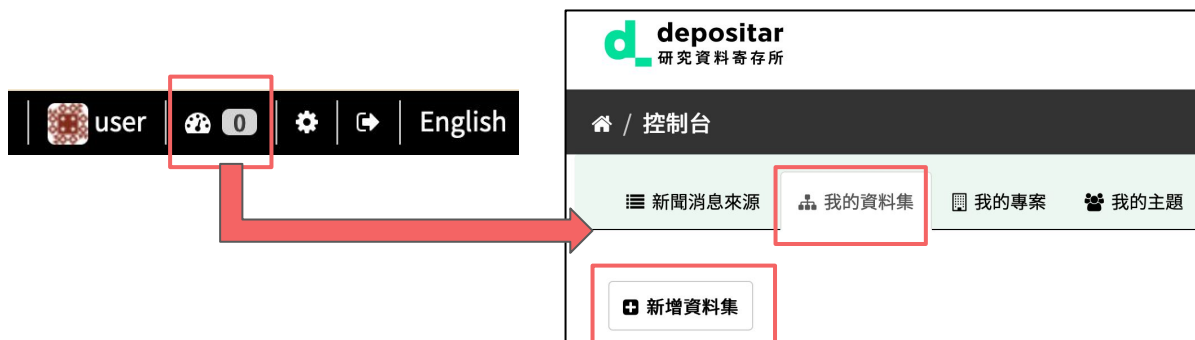




新增資料集

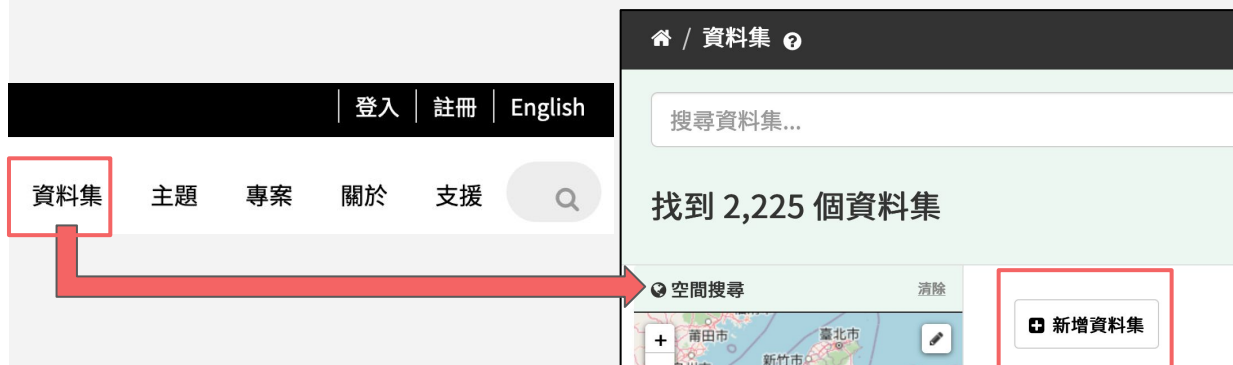
(登入狀態下, 三擇一)



首頁按鈕



右上  個人控制台 >
 我的資料集



右上選單「資料集」

填寫資料集基本資訊

- 標題
- 網址 *
- 英數字，可自訂
- 勿過於簡單，避免重複
- 摘要 (支援 [Markdown](#))
- 資料類型 *
- 滑鼠於選項上停留將顯示說明
- [Wikidata](#) 關鍵字
- 標籤
- 語言
- 備註 (支援 [Markdown](#))

* 需要的欄位 § 接受多值 🔍 ARK 所需欄位

📄 標題 🔍:

例如：一個描述性的標題

* 網址: <demo.depositar.io/dataset/<dataset>> [編輯](#)

✍ 摘要:

關於資料集簡單扼要的說明

您可以在此使用 [Markdown](#) 格式

📄 資料類型 §:

- 📄 壓縮檔資料
- </> 原始碼
- ⚙ 組態資料
- 🗄 資料庫
- 📁 辦公軟體文件
- 📊 結構化圖形
- 🖼 影像
- 🎧 影音資料
- 📶 網路通訊資料
- 📄 原始輸出資料
- 📊 科學與統計資料
- 🛠 應用程式
- 📄 結構化文字資料
- 📄 純文字資料
- * 其他

📄 Wikidata 關鍵字 §:

加入維基資料項目以描述資料集...

🔍 搜尋 Wikidata (維基資料) 項目選取關鍵字以描述資料集。只用於您專案或資料集的標記 (如: 計畫編號) , 請填寫於「標籤」欄位。

📄 標籤 §:

例如：第一期計畫、與 Joe 的訪談

🔍 只用於您專案或資料集的標記。

📄 語言 §:

搜尋 ISO 639-3 語言編碼...

🔍 資料集內容所使用之語言。ISO 639-3 語言編碼請參見列表。

📄 備註:

您可以在此使用 [Markdown](#) 格式

填寫資料集時空資訊

- 時間解析度
- 起始時間
- 結束時間
- 空間範圍 (圖台、四至範圍)
- 空間解析度

時空資訊

時間資訊 (資料內容時間)

⌚ 時間解析度:

ⓘ 時間解析度為表示資料內容對應時間所能精確描述之程度。可以指「相鄰兩次觀測的最小時間間隔」，或「在不確定開始與結束的明確時間時，最大的誤差範圍」。

📅 起始時間🔍:

ⓘ 請填寫與資料集內涵所關聯之時間，而非資料集內資源的產出或完成之時間。

📅 結束時間🔍:

ⓘ 請填寫與資料集內涵所關聯之時間，而資料集內資源的產出或完成之時間。

空間資訊

📍 空間範圍:

以 GeoJSON 描述資料集之空間範圍。您亦可使用圖台，或自四至範圍轉換，以填寫此欄位。

使用圖台填寫空間範圍



ⓘ 填寫空間範圍資訊後，方可以空間搜尋獲得此資料集。

空間範圍.X.min:

空間範圍.X.max:

空間範圍.Y.min:

空間範圍.Y.max:

使用上述四至範圍填寫空間範圍

📏 空間解析度:

可使用四至範圍填寫空間範圍

圖

填寫資料集管理資訊

- 授權 *
- 產製者 *
- 資料產製時間
- 資料處理歷程 (支援 Markdown)
- 專案
- 聯絡人
- 聯絡人的電子郵件

● 授權*

授權條款未指定

ⓘ 宣告本資料集所使用的授權。若不在候選清單內，請選擇「其他授權」，並於上方「備註」欄位註明授權條款。

● 產製者*🔍:

Joe Bloggs

ⓘ 資料生產者或單位的名稱

📅 資料產製時間:

接受 'YYYY'、'YYYY-MM'、'YYYY-MM-DD'

ⓘ 資料集檔案產出時間

🔄 資料處理歷程:

以文字描述資料形成所經過之處理過程，建議以各階段或步驟為導向進行填寫

您可以在此使用 Markdown 格式

🏠 專案:

不屬於任何專案

ⓘ 設定資料集所屬專案。若選擇「不屬於任何專案」，則此資料集會強制設定為「公開」狀態。

使用您的帳號資訊填入聯絡人與電子郵件

● 聯絡人:

Joe Bloggs

ⓘ 本資料集主要維護人員

✉ 聯絡人的電子郵件:

joe@example.com

ⓘ 本資料集維護人員之電子郵件信箱

* 需要的欄位 S 接受多值 🔍 ARK 所需欄位

可按此帶入帳號資訊
以填寫聯絡人與電子郵件

完成資料集填寫

下一步：新增資料

新增資源 (檔案)

- 檔案 (上傳檔案或填寫連結)
- 名稱
- 摘要 (支援 Markdown)
- 字元編碼
- 座標參考系統
- 格式
 - 通常為檔案之「副檔名」
 - 將決定資料可否被預覽
 - 例: CSV (前面無需加“.”)
 - 未填寫時, 系統將自動猜測

僅用於 shapefile
其他檔案類型免填

- 檔案上傳限制: 單一檔案大小 1GB、檔名長度 3~100 字元, 無數量限制
- 詳細欄位填寫說明請見 [後設資料項目](#)

請參考使用手冊、後設資料說明, 與平台限制填寫

上傳檔案大小限制: 1 GB

檔案:

上傳

連結

名稱:

例如: 2011 年 1 月金價

摘要:

關於資源簡單扼要的說明

您可以在此使用 Markdown 格式

字元編碼:

此資源所使用之編碼系統。目前僅用於 shapefile 資源。

座標參考系統:

EPSG 例如: 4326

當您所新增之資源為 shapefile 檔案, 且未提供投影格式 (.prj) 檔案時, 則需另外填寫此欄位, 否則將無法開啟預覽功能。常用座標系統: EPSG:4326→WGS84 (預設), EPSG:3826→TWD97/TM2 台灣, EPSG:3821→TWD67, EPSG:3825→TWD97/TM2 澎湖, EPSG:3828→TWD67/TM2 台灣。

格式:

例如: CSV, XML 或 JSON

Shapefile 檔案請填寫「shp」, 並將 .shp、.shx、.dbf、.prj 檔案置於一 zip 壓縮檔內,

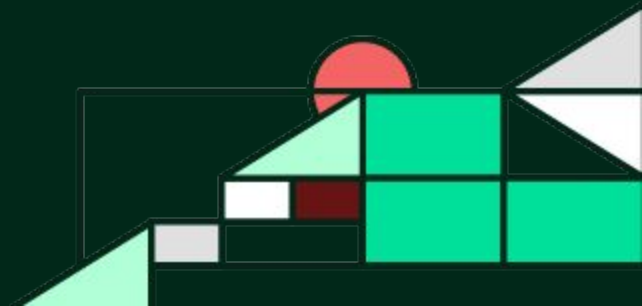
完成資源新增

先前的

儲存並新增其他項目

完成

二、新增資源檢視



資源管理 (編輯) 頁面

- 位於資源頁面右上方「管理」按鈕
- 可在此編輯與刪除資源

- 需有編輯權限
- 資料集管理 (編輯) 頁面同理

🏠 / 專案 / depositar / COSCUP 2022 — 資料長期保存的好所在：用. 資源頁面
/ 資料長期保存的好所在：用 CKAN...

資料長期保存的好所在：用 CKAN 打
造的「研究資料寄存所」(depositar)

🔧 管理

📄 下載

網址：<https://data.depositar.io/dataset/e281bf47-9556-4844-9a20-c32a16dc46d9/resource...>

depositar 團隊於 COSCUP 2022 的報告簡報 (pdf)

資源預覽－新增檢視

- 資源檢視 Resource View
 - 資源預覽的不同方式
 - 資源可以擁有多個檢視
- 位於資源管理頁面上方「檢視」
- 資源檢視類型
 - Data Explorer: 表格 + 簡易圖表
 - Grid: 單獨表格瀏覽
 - Map: 地圖
 - 圖片、影片、網站、音訊...等



新增篩選條件

Data Explorer

網絡 圖表 地圖 172 records « 1 - 100 »

_id	Zip Code	Office N...	City / C...	Townshi...	Longitude	Latitude
1	100003	臺北北門...	臺北市	中正區	121.511...	25.047355
2	100016	臺北東門...	臺北市	中正區	121.528...	25.03408
3	108207	臺北漢中...	臺北市	萬華區	121.507...	25.04132
4	108010	臺北西園...	臺北市	萬華區	121.501...	25.040769
5	100056	臺北南海...	臺北市	中正區	121.514...	25.031833
6	100053	臺北螢橋...	臺北市	中正區	121.517...	25.02463
7	106203	臺北青田...	臺北市	大安區	121.531...	25.026851
8	100009	臺北復興...	臺北市	中正區	121.519...	25.046041

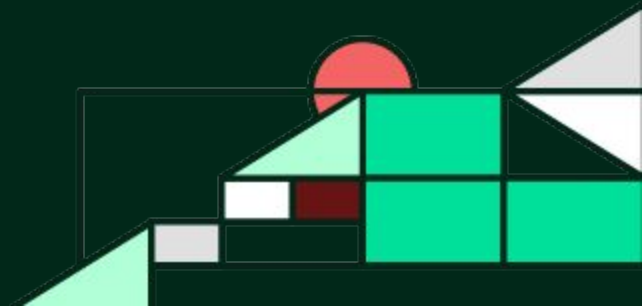




上傳檔案後會自動產生檢視的格式

- 文字檔案 (TXT、JSON 與 XML 等)
- 圖片 (PNG、JPEG 與 GIF)
- PDF
- 試算表 (CSV、XLS 與 XLSX)
- 網頁 (HTML)
- 空間資料 (KML、WMS/WMTS、GeoJSON、Shapefile 等)
- Audio (MP3、WAV、OGG)
- Video (MP4、WebM、OGG)

三、使用資料 API



示範資料集

- Sea turtle sightings in Taiwan | 台灣海龜目擊紀錄

<https://n2t.net/ark:37281/k5f374z25>

資料瀏覽器

全螢幕

嵌入

新增篩選條件

網格

圖表

地圖

about 4000 records

«

1

- 100

»

Q

Search data ...

Go »

篩選條件

_id	occurre...	basisOf...	catalog...	recorde...	year	month	day	eventDate	eventRe...	eventTime	continent	country	country...
1	819757d...	HumanO...	201805_02	Willy Liang	2010	11	18	2010/11/...		11:45+0...	Asia	Taiwan	TW
2	a66994a...	HumanO...	201906_15	DrSea Chu	2010	3	23	2010/3/23			Asia	Taiwan	TW
3	8af5e7b...	HumanO...	202004_...	林珈薇	2012	4	11	2012/4/11			Asia	Taiwan	TW
4	285b96b...	HumanO...	201804_34	Chialing ...	2013	4	11	2013/4/11		12:14+0...	Asia	Taiwan	TW
5	b933729...	HumanO...	su_sighti...	Su Huai	2014	9	11	2014/9/11			Asia	Taiwan	TW
6	4a7822e...	HumanO...	201803_22	彭貴莉	2014	4	26	2014/4/26			Asia	Taiwan	TW
7	d9e3b90...	HumanO...	su_sighti...	Su Huai	2015	1	1	2015/1/1			Asia	Taiwan	TW
8	908e9e1...	HumanO...	su_sighti...	Su Huai	2015	1	2	2015/1/2		14:51+0...	Asia	Taiwan	TW
9	f1d3bc0...	HumanO...	su_sighti...	Su Huai	2015	12	5	2015/12/5			Asia	Taiwan	TW
10	44cf098...	HumanO...	su_sighti...	Su Huai	2015	12	5	2015/12/5			Asia	Taiwan	TW

取得資源 ID

- 於資料集頁面，點選資源

資料與資源



- 於資源頁面，點選「資料 API」按鈕

🏠 / 資料集 / Sea turtle sightings in... / TurtleSpot2022_v2

TurtleSpot2022_v2

📄 下載

🔗 資料API

網址：<https://data.depositar.io/dataset/ed9c9422-0e34-448b-8b3f-4...>

註：僅符合規定的 CSV、XLS 與 XLSX 檔案可使用資料 API 功能，此時會顯示「資料 API」按鈕


- 於範例中即可取得資源 ID (resource_id)

查詢 »

查詢之範例 (前5個結果)

https://data.depositar.io/api/3/action/datastore_search?resource_id=661e0cc0-3567-4fa3-9d6e-1fe56f53d05d&limit=5

資料 API 範例

 如何於瀏覽器內瀏覽 JSON 檔案:

- ① 點選並複製 JSON 連結
- ② 於 [JSON Viewer](#) 頁面上方點選「URL」按鈕，並貼上前一步取得的連結

● 簡易搜尋

https://demo.depositar.io/api/action/datastore_search?resource_id=RES_ID

- 取前 100 筆資料: [&limit=100 連結](#)
- 取第 101 筆~200 筆資料: [&offset=100&limit=100 連結](#)
- 取 county 與 individualCount 欄位: [&fields=county,individualCount 連結](#)
- 搜尋包含 Taitung 的資料: [&q=Taitung 連結](#)

RES_ID: 資源 ID

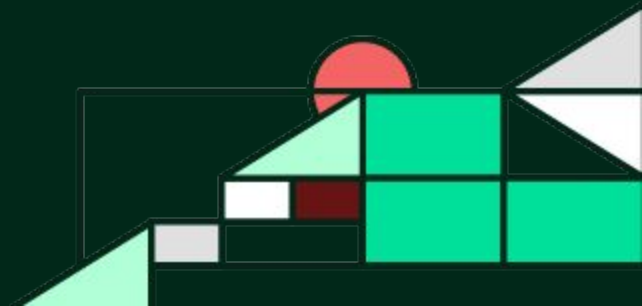
● 進階 SQL 搜尋

https://demo.depositar.io/api/action/datastore_search_sql?

- 取得 county 為 Taitung 的資料: [sql=SELECT * from "RES_ID" WHERE county = 'Taitung' 連結](#)

● [更多使用方式](#)

四、操作練習



- 前往 depositar 測試站 <https://demo.depositor.io>
- 點選右上選單「註冊」，於下一頁填寫註冊資訊
 - 使用者 *：文數字、- 與 _ 符號
 - 全名：顯示於個人頁面上的名稱
 - 電子郵件 *
 - 個人資料圖片
- 收信 (來自 no-reply@data.depositor.io)，並完成密碼設定
 - 若未收到密碼設定信件，請檢查垃圾郵件匣
- 右上選單「登入」，並登入帳號

* 必填欄位

以下操作均於 **測試站** 進行，請確認目前登入的網站
另因 **測試站系統資源有限**，若無法連線或連線緩慢，請稍後再嘗試練習

練習一：新增資料集與資源

- [下載](#)範例資料：臺北郵局轄內營業據點 (Post_Taipei.csv)
- 新增資料集
 - 盡量填寫完整，可參考[下載來源](#)，與原始[資料來源](#) (中華郵政全國營業據點 | 政府資料開放平臺) 與填寫
 - **資料類型、產製者 (位於表單尾端) 為必填欄位**，未填寫將無法建立資料集
 - 按一下最下方「下一步：新增資料」按鈕
- 新增資源
 - 選取並上傳範例資料
 - 填寫資源名稱，與格式 (CSV)
 - 按一下最下方「完成」按鈕

練習一：新增資料集與資源 (續)

- 自動返回資料集頁面 (如右上圖)
- 點選剛才上傳之資源，確認有表格預覽 (如右下圖)
 - 如未顯示，請稍後重新整理網頁
- 也可以嘗試上傳其他檔案、連結

臺北郵局轄內營業據點

新增完成的資料集

示範用資料集

資料與資源

CSV 臺北郵局轄內營業據點

點選

Wikidata 關鍵字

台北郵局 中華郵政

探索

新增篩選條件

新增完成的資源

表格 圖表 地圖 172 records << 1 - 100 >>

_id	Zip Code	Office N...	City / C...	Townshi...	Longitude	Latitude
1	100003	臺北北門...	臺北市	中正區	121.511...	25.047355
2	100016	臺北東門...	臺北市	中正區	121.528...	25.03408
3	108207	臺北漢中...	臺北市	萬華區	121.507...	25.04132
4	108010	臺北西園...	臺北市	萬華區	121.501...	25.040769
5	100056	臺北南海...	臺北市	中正區	121.514...	25.031833
6	100053	臺北螢橋...	臺北市	中正區	121.517...	25.02463
7	106203	臺北青田...	臺北市	大安區	121.531...	25.026851
8	100009	臺北復興...	臺北市	中正區	121.519...	25.046041


須先完成練習一
請參照本簡報 11 至 12 頁內容操作

練習二：新增資源檢視

- 資源管理頁面上方「檢視」
- 按一下「新增檢視」按鈕，選擇「Map」
- 下一畫面標題填寫「臺北市營業據點」
- 按一下篩選條件，City / County 選取「臺北市」
- 勾選「自動縮放至圖徵」與「將附近的圖徵標記為叢集」
- 按一下最下方「新增」按鈕，新增的資源檢視將顯示於下一畫面 (如右圖)



練習三：使用資料 API (以 Python 為例)

- [連結](#)至範例資料集
- 點選左方工具列
BinderHub 區塊下方
 按鈕
- 在新分頁的檔案總管開啟 Example.ipynb
- 於上方選單選取 Run, 再選取 Run All Cells
- 如正確執行, 應會產生如右三張圖表

可嘗試自由變更參數 (如 query 範圍等)

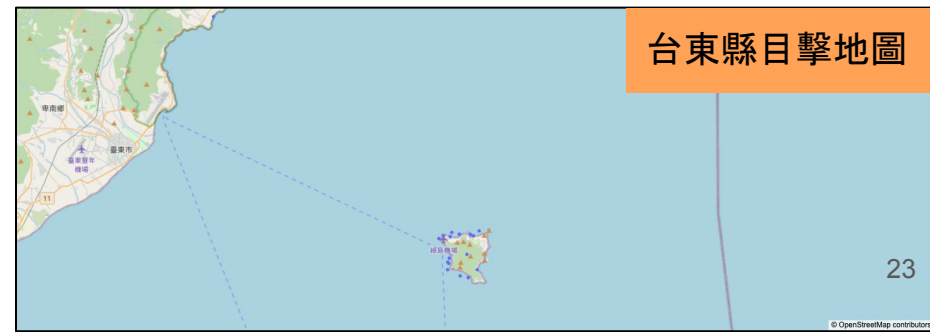
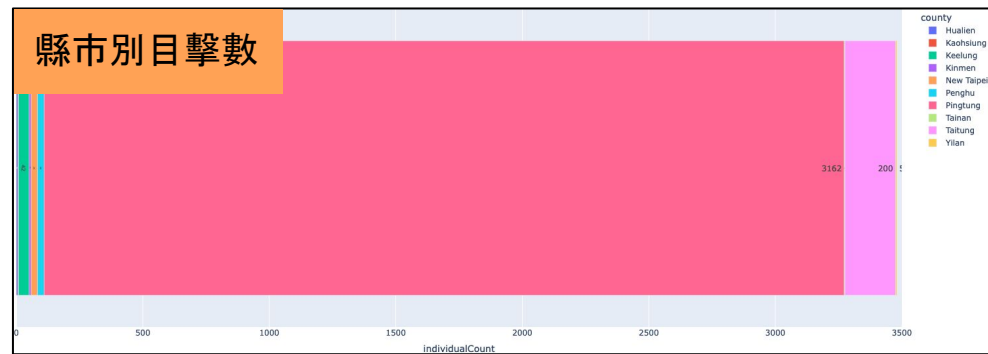
Get the data 所有資料

```
[2]: df = pd.read_csv("../TurtleSpot2022_v2_gbifV1.7.csv")
df2 = df.copy().query('county == "Taitung"')

df
```

```
[2]:
```

	occurrenceID	basisOfRecord	catalogNumber	recordedBy
0	819757d5-5a06-4332-b2e6-41d307f98526	HumanObservation	201805_02	Willy Liang
1	a66994ab-582a-445e-b811-66f5a938a1e7	HumanObservation	201906_15	DrSea Chu



@_depositar



謝謝！ Thank You!

<https://data.depositar.io/> 研究資料寄存所
<https://rdm.depositar.io/> 研究資料管理推進室

data.contact@depositar.io
<https://lab.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.

The *depositar* project team: T-R Chuang, M-S Ho, C-J Lee & C-H Ally Wang.

「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。

研究資料寄存所計畫成員: 莊庭瑞、何明誼、李承 龕、王家薰。

