

線上使用研究資料寄存所

A Hands-on Practice with *depositor*

2022-12-21

李承鑫

Cheng-Jen Lee

台灣 中央研究院

Academia Sinica – Taiwan



準備工作

右上選單

登入 | 註冊 | English



資料集 主題 專案 關於 支援



- 前往 **depositor 測試機** <https://demo.depositor.io>
- 右上選單「註冊」
- 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
- 前往信箱收信，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
- 下載本簡報檔案與範例資料 (下圖)
 - https://data.depositor.io/dataset/2022-12-21_course
 - 線上使用研究資料寄存所 (Hands-on.pdf)
 - 臺北郵局轄內營業據點 (Post_Taipei.csv)

已依照行前通知
完成註冊者，請直
接登入帳號

* 必填欄位

註冊帳號

密碼設定信件

範例資料：臺北郵局轄內營業據點

提供機關 / 中華郵政股份有限公司 [2021] [中華郵政全國營業據點] 此開放資料依政府資料開放授權條款 (Open Government Data License) 提供

資料 API 使用範例 Jupyter Notebook
可下載後於 Google Colab 或 ASGC DICOS Apps 執行

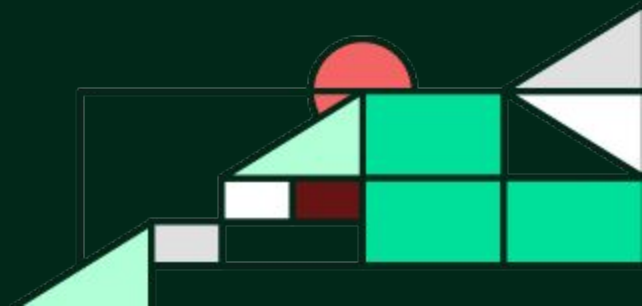
資料 API 使用範例 Jupyter Notebook (JupyterLite 專用)
可下載後於 JupyterLite 執行



大綱

- 新增資料集與資源
- 新增資源檢視
- 資料 API 使用 (以 Python 程式語言為例)

一、新增資料集與資源



新增資料集

(登入狀態下, 三擇一)



新增資料集 – 基本資訊

- 標題
- 網址 * (可自訂)
 - 資料集唯一識別碼
- 摘要
- 資料類型 *
- 滑鼠於選項上停留將顯示 說明
- Wikidata 關鍵字
 - 使用 [Wikidata](#) 項目作為關鍵字
- 標籤
- 語言
- 備註

📄 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

* 需要的欄位 § 接受多值 🔍 ARK 所需欄位

新增資料集 – 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🔍:

接受 'YYYY'、'YYYY-MM'、'YYYY-MM-DD'

ⓘ 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🔍:

接受 'YYYY'、'YYYY-MM'、'YYYY-MM-DD'

ⓘ 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:

填寫圖資空間四至範圍之最西經度值

空間範圍.X.max:

填寫圖資空間四至範圍之最東經度值

空間範圍.Y.min:

填寫圖資空間四至範圍之最南緯度值

空間範圍.Y.max:

填寫圖資空間四至範圍之最北緯度值

使用上述四至範圍填寫空間範圍

📏 空間解析度:

資料的空間解析度數值，以公尺為單位

可使用四至範圍填寫空間範圍

新增資料集 – 管理資訊

- 授權 *
- 產製者 *
- 資料產製時間
- 資料處理歷程
- 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* Q:

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

可按此帶入帳號資訊
以填寫聯絡人與電子郵件

● 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 Q ARK 所需欄位

完成資料集欄位編輯

下一步：新增資料

新增資源

- 檔案 (上傳檔案或填寫連結)
- 名稱
- 摘要
- 字元編碼
 - 僅用於 shapefile 資源, 其他資源免填
- 座標參考系統
 - 僅用於 shapefile 資源, 其他資源免填
- 格式
 - 通常為檔案之「副檔名」
 - 將決定資料可否被預覽
 - 未填寫時, 系統將自動猜測, 但仍建議填寫

1. 檔案上傳限制: 單一檔案大小 1GB、檔名長度 3~100 字元
2. 無資源數量限制

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

完成資源新增

先前的

儲存並新增其他項目

完成

練習一：新增資料集與資源

- 新增資料集
 - 盡量填寫完整，可參考[資料來源](#) (中華郵政全國營業據點 | 政府資料開放平臺) 填寫
 - 標題勿過於簡單，以免重複而不被接受
 - 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
 - 填寫完成後，按一下最下方「下一步：新增資料」
- 新增資源
 - 選取並上傳測試資料 (Post_Taipei.csv)
 - 填寫名稱，與格式 (CSV)
 - 按一下最下方「完成」
- 自動返回資料集頁面 (如右上圖)
- 點選剛才上傳之資源，確認是否有表格預覽 (如右下圖)
 - 如未顯示，請稍後再重新整理網頁

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 ↻ 探索

Wikidata 關鍵字

台北郵局 中華郵政

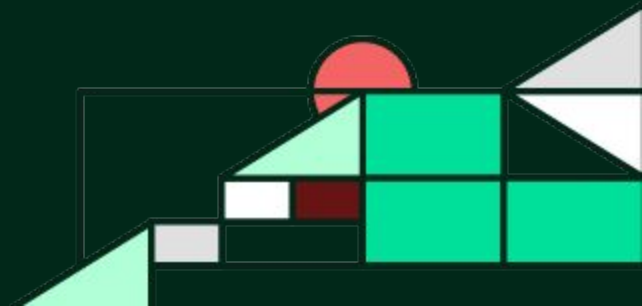
新增篩選條件

新增完成的資源

網絡 圖表 地圖 172 records « 1 - 100 »

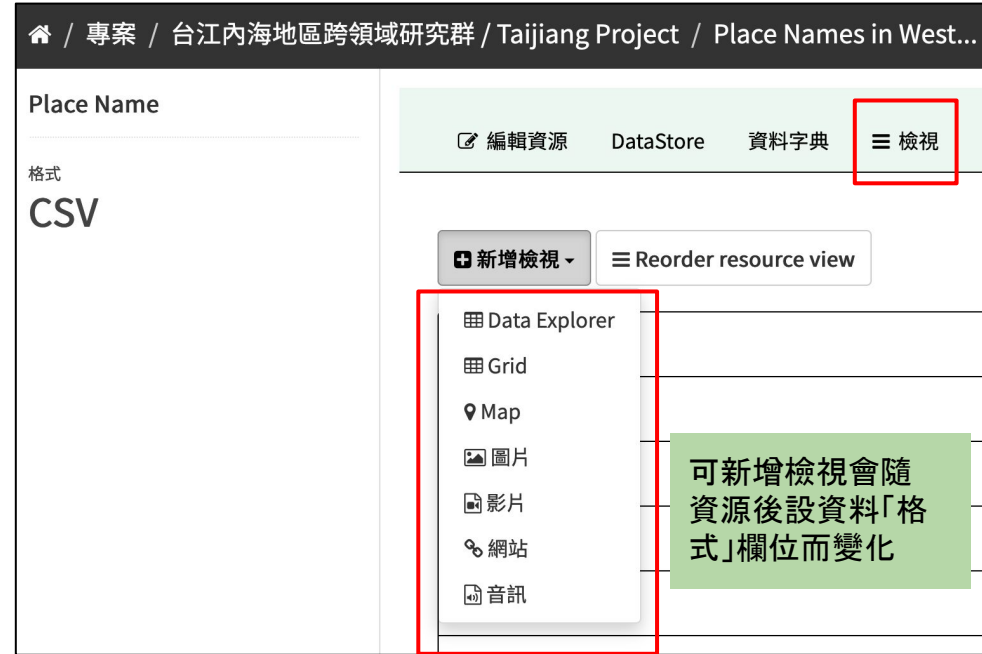
_id	Zip Code	Office N...	City / C...	Townshi...	Longitude	Latitude
1	100003	臺北北門...	臺北市	中正區	121.511...	25.047355
2	100016	臺北東門...	臺北市	中正區	121.528...	25.03408
3	108207	臺北漢中...	臺北市	萬華區	121.507...	25.04132
4	108010	臺北西園...	臺北市	萬華區	121.501...	25.040769
5	100056	臺北南海...	臺北市	中正區	121.514...	25.031833
6	100053	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

二、新增資源檢視



資源預覽－自訂檢視

- 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
- 資源管理頁面上方「檢視」
- 新增檢視
 - Data Explorer: 表格瀏覽 + 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

網格式 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

Data Explorer



資料集 / 資源管理頁面

- 位於資料集 / 資源頁面右上方「管理」
- 可在此編輯與刪除資料集 / 資源 需擁有編輯權限

🏠 / 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)

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

depositar 團隊於 COSCUP 2022 的報告簡報 (pdf)

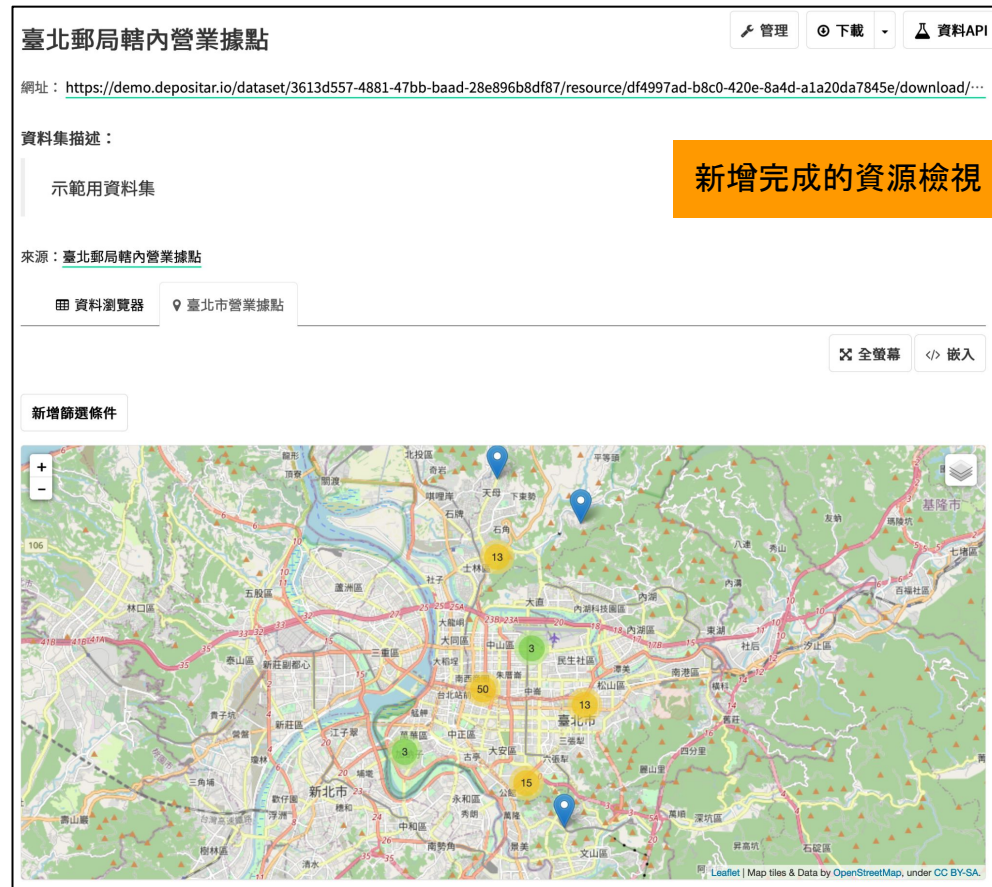
 管理  下載

須先完成練習一
請參照本簡報 12 至 13 頁內容操作

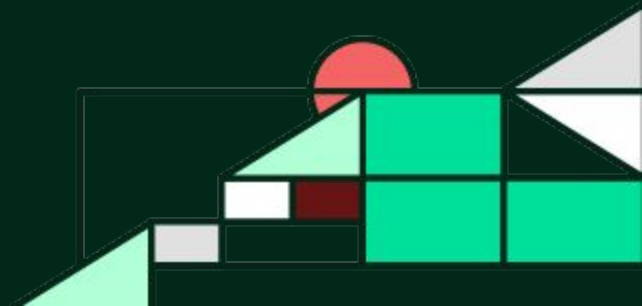
練習二：新增資源檢視



1. 【調整欄位類型】**管理頁面**上方**資料字典**分頁
 - a. 將「欄位 5. Longitude (text)」的「覆蓋類型」調整為 numeric
 - b. 將「欄位 6. Latitude (text)」的「覆蓋類型」調整為 numeric
 - c. 按一下最下方「儲存」按鈕
2. 【更新資料庫】**管理頁面**上方 **DataStore** 分頁
 - a. 按一下「上傳至 DataStore」按鈕
 - b. 重新整理畫面，直到最下方顯示 Express Load completed, 繼續下一步驟
3. 【新增檢視】**管理頁面**上方**檢視**分頁
 - a. 按一下「新增檢視」按鈕，選擇 Map
 - b. 在下一畫面，「標題」填寫「臺北市營業據點」
 - c. 按一下篩選條件，選取「臺北市」
 - d. 欄位類型確認為 Latitude / Longitude fields
 - e. 緯度欄位選取 Latitude; 經度欄位選取 Longitude
 - f. 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)



三、資料 API 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 »

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

資料 API (續)

- 簡易搜尋

https://data.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](#)

- 進階 SQL 搜尋

https://data.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](#)

- 回傳格式: JSON

- 更多使用方式

- <http://docs.ckan.org/en/2.9/maintaining/datastore.html>



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
```

查詢之範例 (包含字串 'jones' 的結果)

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

查詢之範例 (透過SQL指令敘述)

```
https://data.depositar.io/api/3/action/datastore_search_sql?sql=SELECT * from "661e0cc0-3567-4fa3-9d6e-1fe56f53d05d" WHERE title LIKE 'jones'
```

RES_ID 可由資料 API 範例獲得

練習三：資料 API 使用 (以 Python 語言為例)

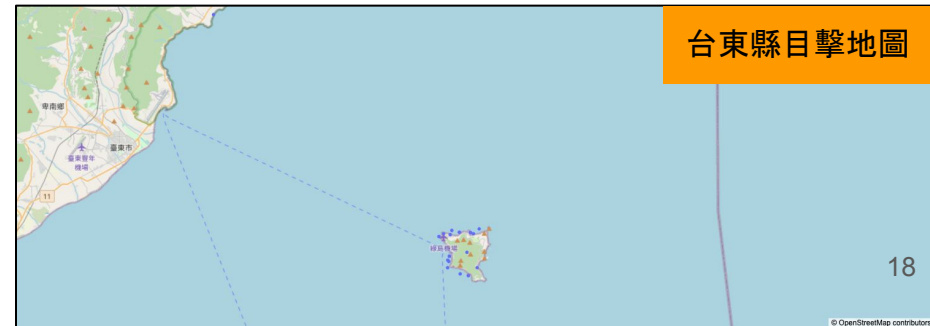
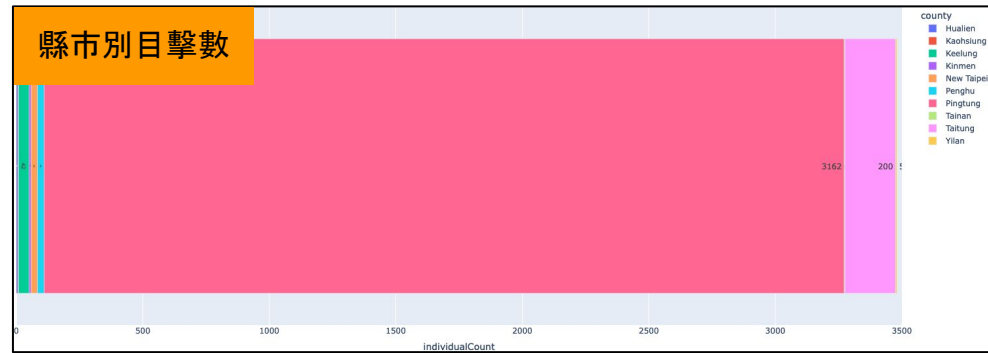
- 下載範例 Jupyter Notebook
 - https://data.depositar.io/dataset/2022-12-21_course
 - 資料 API 使用範例 Jupyter Notebook
- 方法一：於 Google Colab 執行
 - 連結至 [Google Colab](#), 選擇「上傳」頁籤, 並上傳 API_Example.ipynb
 - 執行階段 > 全部執行
- 方法二：於 ASGC DiCOS Apps 執行
 - 連線至 [DiCOS Apps](#), 新增並開啟 Jupyter Lab gpu 3090
 - 於開啟的畫面拖入 API_Example.ipynb
 - 點兩下開啟 ipynb 檔案
 - Run > Run All Cells
- 方法三：於 JupyterLite 執行
 - 連結至 [JupyterLite](#)
 - 於開啟的畫面拖入 API_Example_Lite.ipynb
 - 點兩下開啟 ipynb 檔案
 - Run > Run All Cells
- 如正確執行, 應會產生如右三張圖表
- 可自由變更資料來源、資料 API 參數 (如 SQL 搜尋範圍) 等

瀏覽所有資料

所有資料

```
[ ] df = json_normalize(dic["result"]["records"])
df
```

	_id	occurrenceID	basisOfRecord	catalogNumber
0	1	819757d5-5a06-4332-b2e6-41d307f98526	HumanObservation	201805_02
1	2	a66994ab-582a-445e-b811-66f5a938a1e7	HumanObservation	201906_15
2	3	8af5e7b5-0e53-47fa-ab53-03b6dcb46183	HumanObservation	202004_100



@_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

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。

