

# 永續科學團隊研究資料管理的行動研究

盧鏡臣、周素卿、王瓊芯、鄭詠晨

12/21 2023

# 開放研究資料的研究

隨著網路發展，網路可做研究成果、資料資訊化、傳播的載台

- 2003年Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities
- 不僅是研究論文開放，原始資料、數位化素材及學術多媒體資料應該被搜尋到及使用

影響研究員人員分享/使用開放研究資料的因素，2004-19 36篇相關文章的綜述研究

- 研究人員背景、要求和義務、個人動力和內在動機、促進條件(如設備、網路、平台、資金等)、信任、預期效益、社會影響力和從屬關係、付出之成本(時間、精力等)、研究人員的經驗和技能、法規或政策要求(如投稿期刊要求提供資料)、資料特徵
- <https://doi.org/10.1371/journal.pone.0239283>

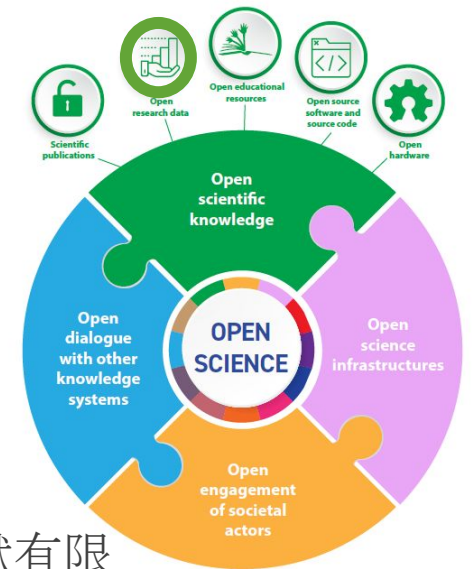
不同學科分享/使用開放資料的程度有異

- 和地球科學等學科比，醫學、社會科學分享的程度較低

開放研究資料存在巨大障礙

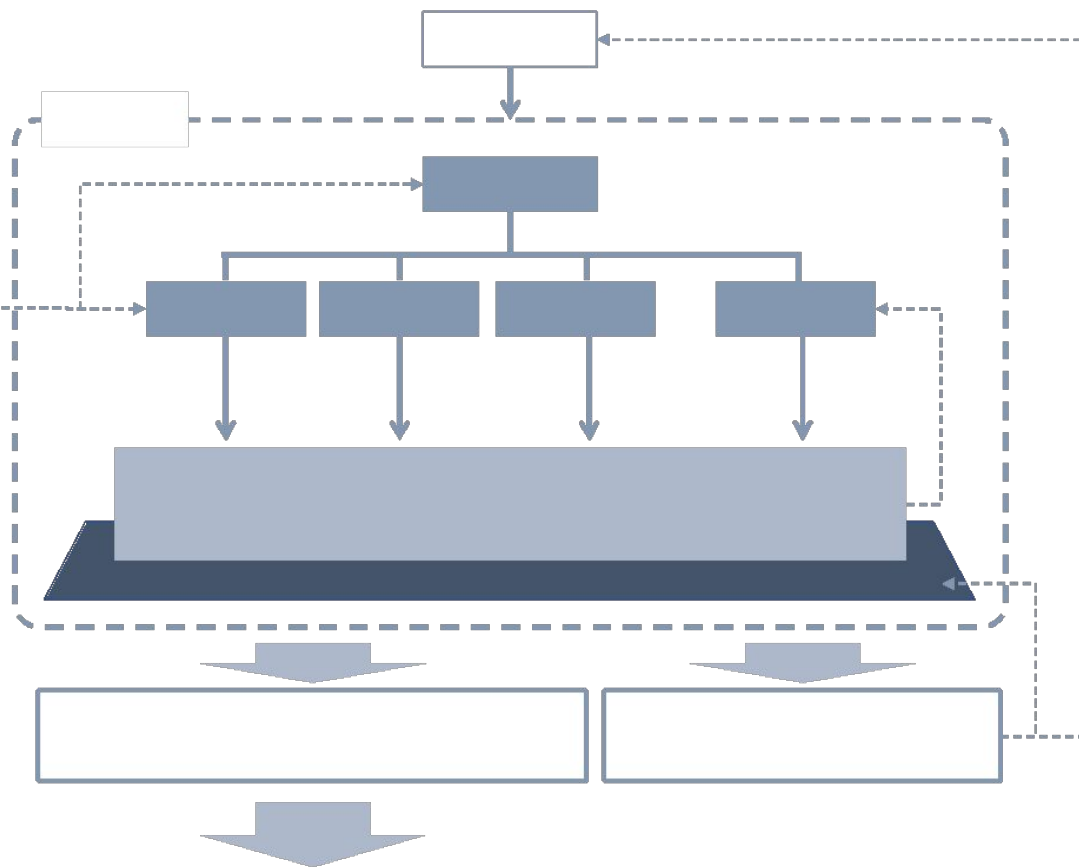
- 研究者不願意分享研究的資料
- 缺乏時間、資源、技術來進行資料分享
- 缺乏開放資料平台/基礎設施、智慧財產權議題、資料誤用、資料格式不符

有關研究團隊如何導入開放研究資料？遭遇到哪些障礙？如何因應？的相關文獻有限



Source: UNESCO, 2021

# 研究問題與方法



## 國科會計畫：都市永續轉型與韌性建構 執行DMP之目標

- A.** 促進團隊研究經驗的傳承
- B.** 應用分析指認研究知識系譜
- C.** 強化團隊資料取得、分享與整合
- D.** 透過開放研究資料促進都市永續轉型研究與實務應用

如何透過DMP促成上述目標？開放研究資料有哪些過程、特性與障礙？

採行動研究取徑：以本整合型計畫為案例，透過設計推動開放資料的過程，歸納分析團隊的操作經驗、困難障礙，以探索討論導入開放研究資料/DMP的策略、議題及效益討論

# 研究問題與方法

---

## 研究問題

- 研究有哪些資料初始需求、中間處理及成果呈現？
  - 次級資料 vs. 原始資料
  - 中間資料
  - 成果呈現
- 研究團隊如何執行資料蒐集與處理？
  - 人力配置、處理方法、檢核機制
  - 內部資料儲存、分享架構
  - 研究團隊執行開放研究資料的階段成果、效益與遭遇困難為何？

## 研究方法—行動研究

- 資料管理與分享平台評估
  - 篩選開放研究資料的存放平台: Depositar
  - 由單一子計畫先試行
- 撰寫資料寄存指引及內部操作訓練
- 請各子計畫上傳研究資料
- Depositar開放研究資料之階段評估--盤點開放研究資料之特性、品質及分析原因
  - 盤點資料內容
  - 訪談--操作DMP的經驗、權衡考量與原因

# 哪些資料被蒐集、處理、分享？

## ● 實際使用的資料—次級

### ● 空間圖資類

- 基本圖
- 1995/2013/2019國土利用調查圖資、土地使用分區、地形圖、建物圖
- 開放資料—衛星影像

### ● 特定主題圖

- 開放資料—北市水利處主集水區圖、水綠地圖
- 行文申請—台北市公有林範圍圖

### ● 政府調查統計資料

- 人口、社會、經濟
- 人口指標、所得、原住民人口數、教育程度、售電資料、冷氣普及率

### ● Co-design—新北市循環經濟推動之數值

### ● 土地統計

- 統計要覽—作物種植面積與、營建材料、已登錄土地使用用途

## ● 研究文獻資料(含政府研究報告)

- 政策法規計畫文件
- 都市計畫、空間規劃—北士科
- 特定主題之政策計畫、政府/議會公報-氣候、循環經濟、防洪、公有林、造林

## ● 新聞媒體報導剪報

- 特定主題—北士科、都市森林、韌性都市林
- 搜尋管道以官方新聞公告、聯合新聞知識庫為主

## ● 實際使用的資料—原始

### ● 訪談/工作坊之書信、影音及逐字稿

### ● 調查清冊、問卷及調查資料

- 現地調查—地圖、照片紀錄
- 問卷

### ● 質性田野調查記錄—參與觀察記錄、照片、錄音檔與逐字稿

# 哪些資料被蒐集、處理、分享？

- 資料處理過程之樣態
  - 空間圖資—精度/分類校正
    - 多時期國土利用調查圖資；河道資料分類標準不一，參考前後期圖資進行修補組合
  - 空間圖資—三維化、擬真化、VR化
    - 將土地使用分區圖、道路圖、建物圖導入 City engine 軟體中，以產製三維模型
    - 針對重點建物進行現況擬真，使用台灣的紅綠燈、窗框、停車格軟體素材，以及透過現地調查，建置現況便利商店的招牌、鐵欄杆、鐵窗、斑馬線、摩托車等元素。並以VR呈現模型
  - 衛星影像分析
    - 藉Landsat 7&8衛星影像數據，獲取植生指數 (NDVI)與地表溫度(LST)資料
    - 藉Sentinel-2數據，獲取NDVI
- 統計資料空間化
  - 人口、社會、經濟統計資料為例
    - 部分數據在原始資料的基礎上，加以合併/重組/平均等處理。
    - 藉由最小統計區、行政區界空間化，進行空間分析/統計分析
- 統計資料指標化，以進行能值分析
  - 參考土地統計數據，估算投入、產出指標。並將指標數據轉化能值參數，以操作能值分析。如作物種植面積X產量，以估算農產品的產出；如營建材料、石材，以估算政府對地方建設投入...等指標。
- 次級資料的內容分析/文本分析
  - 依照「關鍵字：Circular economy, indicators, factors, index」整理文獻，進而歸納出CE指標—「nano、micro、meso、macro」之定義及其界定方式
  - 依照「關鍵字：都市、中和四號公園」整理林業試驗所相關之新聞剪報資料，以歸納出都市林一詞的討論重點、相關利害關係人等脈絡

# 哪些資料被蒐集、處理、分享？

類別	初始擬蒐集之資料類型與內容	資料實際使用情形	資料開放情形 (資料集數量)
次級資料	研究文獻資料(含政府研究報告)	√	4
	政策法規計畫文件	√	8
	新聞媒體報導剪報	√	3
	政府調查統計資料	√	1
	空間圖資	√	1+1
	其他(網站)	√	1
原始資料	訪談、工作坊、書信、影音及逐字稿	√	0
	調查清冊、問卷及調查資料	√	1
	內容分析(如計畫內容評估)表單及資料	—	0
	實驗研究資料	—	0
	質性田野調查紀錄(含影音)	√	1
	參數、分析流程、程式原始碼及程式	√	0
加值資料	GIS及空間影像加值資料	√	10
	加值統計資料(如新增變項、降尺度資料)	—	0
研究成果	如報告書、論文、圖、表...等	√	1

## ● 資料特性及障礙歸納

- 空間圖資，多有版權問題，且購置成本高。
- 遠期官方統計資料蒐集具難度
  - 遠期缺乏數化資料
  - 須從紙本資料，回推數化、校正關鍵字，再搜尋
  - 每年調查項目不同，如農業生產成本
- 資料獲取管道，除開放資料、行文申請外，則為co-design的途徑

本資料集為全台範圍內的相關圖資，用以與社子島相關數據進行橫向比對。資料集包括土地使用與社會經濟統計資料、人口統計資料、以及本研究所統整數據。


街區開放空間模擬情境將針對標的建物所在之處建立四個不同時間點的模型，包含該案重後(Now)、假設該案旁建物重建(Future)、假設臨街建物皆重建(Far Future)。此資料集為「三段5巷5弄7號」在「Past&Now」時間點的3D與VR模型。


利用台北市工務局建置之水綠地圖平台，檢視北士科範圍內的都市綠基盤、水資源再利用設施分布，

誤公開, 已修正

## 資料與資源



**六都社會經濟整合資料**    
資料年份：2013 空間範圍：六都 空間層級：鄉里 資料說明：整合型數據



**人口統計指標**    
資料年份：2013、2019、2020 資料頻率：每年 空間範圍：六都 空間層級：鄉里 資料來源：社會經濟資料服務平台

**所得資料**    
資料年份：2013、2019、2020 資料頻率：每年 空間範圍：六都 空間層級：鄉里 資料來源：社會經濟資料服務平台

**原住民人口**    
資料年份：2013、2019、2020 資料頻率：每年 空間範圍：六都 空間層級：鄉里 資料來源：社會經濟資料服務平台



**教育程度**    
資料年份：2013、2019、2020 資料頻率：每年 空間範圍：六都 空間層級：鄉里 資料來源：社會經濟資料服務平台

**冷氣普及率**    
資料年份：2013、2019 資料頻率：每年 空間範圍：六都 空間層級：城市 資料來源：經濟部統計

**台北盆地國土利用調查**    
資料年份：2013、2019 資料頻率：每年 空間範圍：台北盆地 資料來源：國土測繪中心

**台北千分之一地形圖**    
資料年份：2017 空間範圍：台北盆地 資料來源：都發局(都市測量科)

## 資料與資源

**情境模擬二 社區道路 Past&Now**  

### 標籤

CityEngine Now Past

### 基本資訊

資料類型	結構化圖形
------	-------

### 時空資訊

起始時間	2015
結束時間	2022

### 管理資訊

產製者	
資料產製時間	2022
聯絡人	
聯絡人的電子郵件	


缺少資料處理歷程

管理資訊	
產製者	石曉瑜, 王常暢
資料產製時間	2013
資料處理歷程	根據對社子島的研究需求, 部分資料在基礎圖資加工後, 再進行二次處理。
聯絡人	常暢
聯絡人的電子郵件	wcc910921@163.com

次級資料的產製者

應為原資料供應者。

## 資料與資源

**佚名的資源** 

### 標籤

保水防災 台北市 水綠地圖 田園城市 都市綠地

### 基本資訊

資料類型	科學與統計資料
語言	中文 (zho)

### 時空資訊

空間範圍	<a href="#">顯示更多</a>
空間範圍.X.min	121.45053864922376
空間範圍.X.max	121.6341019095853
空間範圍.Y.min	24.964577458486353
空間範圍.Y.max	25.212079081767925



# 哪些資料被蒐集、處理、分享？

---

- 開放資料的詮釋資料歧異度大
  - 加值型的開放資料，缺少原始/基本資料的描述、資料處理歷程記述不完全
    - 空間圖資的取得方式、針對空間圖資的加值方法步驟...
  - 詮釋資料欄位註記不完全或填寫錯誤
    - 未填寫標籤
    - 產製者：次級資料的產製者應為原始產製單位。
- 誤公開未取得授權之資料 (著作權議題)
- 現象原因探討
  - 工作導向，較非開放研究資料導向
    - 未能從資料使用者角度來說明資料
  - 團隊分工問題—上傳者非資料產製者
    - 資料缺乏適當的分類跟標註
  - 對開放研究資料/詮釋資料的認知有落差
    - 執行人員因時間、經驗限制，未能較細緻說明資料處理歷程

# 資料蒐集與處理機制

## 團隊資源/組成樣態

### 團隊組成

- 有專兼任助理
- 兼任助理
- 助理空缺期較長

### 分工方式

- 以個案地域分工
  - 北士科
  - 都市森林
- 依事務分工

## 資料蒐集與處理方法

- 有SOP機制, 資料可重複操作性較高
  - City engine/Arc Pro軟體操作手冊
- 藉由PI經驗指導蒐集方向

## 資料品質檢核機制

### 制度化機制

- 定期工作會議+會議記錄、逐字稿
- 定期工作會議
- 定期研究生meeting)

### 資料品質檢核方式

- 依多位PI、Co-PIs知識經驗, 與助理共同討論
- 根據PI知識經驗, 與助理共同討論
- PI依據知識經驗檢核大方向問題, 小細節由助理把關

## 內部資料儲存/分享架構

各有不同的資料儲存平台, 但依共享程度區分

- Dropbox, 研究資料同步全整合及共享, 且成員可編輯彼此的資料
- Depositar/Google Drive, 研究資料經成員統整篩選後上傳共享
- 學生畢業後, 提供一份研究資料於辦公室桌機; 成員間以USB共享

# 資料蒐集與處理機制

- 資料蒐集與處理運作方式，和研究者的資源、研究室的特性有關
  - 標準作業程序助於資料蒐集/處理的可重複操作性。但前提須有研究生/助理。
  - 研究資料品質和團隊穩定性有關；而研究資料品質檢核，與研究者投入的心力有關。
- 團隊在資料的儲存/分享架構上各有不同；但多是延續過去的經驗
- 研究資料的標準化程度較多元
- 資料開放共享程度有異，但基於信任全部開放共享存在風險
  - **基於信任全部開放共享，對資料安全存在風險**
    - 優點：進度整合性高、開放性高、雲端備份
    - 缺點：
      1. 若研究成員操作雲端資料習慣不佳，則研究資料容易被修改/刪除
      2. 具版權/個資敏感之研究資料有風險性
  - **經成員統整篩選後上傳共享**
    - 優點：資料另有備份
    - 缺點：
      1. 時間成本
      2. 資料經篩選，資料的開放性易受到研究成員觀點影響

# 研究發現與討論

---

- 使用Depositator平台開放資料的助益
  - 團隊成員可隨時取用資料
  - 對研究經驗傳承/交接或有幫助
- 操作Depositator平台之困難與障礙
  - 資料對接性較差, 無法直覺上傳研究資料與填寫
  - 單層資料結構, 資料需花時間重新適當的分類跟註記
- 資料寄存平台架構與使用者行為
  - 開放資料基礎建設 對 開放資料 的促進
  - 平台設計架構 vs 使用者的平台行為
    - 如何促進團隊的資料處理經驗傳承?
      - 資料來源、處理過程等註記
    - 如何促進團隊內部分享資料?
      - 資料開放與授權
      - 資料集 vs. 檔案儲存架構

# 研究發現與討論

---

- 研究團隊之研究資料開放
  - 永續研究資料的開放意願，和研究者對資料被使用的公共利益評估、發表進度有關
  - 開放研究資料為額外的服務工作，且處理資料以符合開放科學的過程，伴有額外的行政成本
    - 研究團隊的量能/人力資源/經費穩定性
  - 部分資料類型存在法律與倫理問題：智財權議題、敏感資料
- 環境資料與開放科學
  - 相當程度仰賴公部門建置之環境資料
  - 資料因跨時段/長期的調查及編碼方式改變，以及調查精度差異，研究者須進行中間處理
  - 加值資料受到資料取得成本、及使用授權的限制
  - 研究成果資料的開放與被使用仍存在限制

謝謝，請指教

---