開放 **IORG**



- 作者 = IORG

note

本文件以下列貢獻者為作者，採創用 CC 姓名標示－相同方式分享 4.0 國際授權釋出。

貢獻者

- IORG
- Chihhao Yu
- Andrea Wang
- Pomin Wu
- …

本文件透過 Google 文件開放協作，權限設定為：任何人皆可註解。上列貢獻者包括本文件正文內容，及透過 Google 文件註解、建議修正功能貢獻內容之使用者，列表由 IORG 主動整理，列出使用者帳號之顯示名稱。

- 本文永久網址 = https://iorg.tw/open
- 本文協作文件 = Google Doc

原則：預設開放、公眾信任

開放、當責，是民主的核心價值。秉持預設開放（open by default）的精神，IORG 的研究、方法、資料、原始碼，將以最大程度開放、開源釋出。

IORG 作為資訊操弄的研究者，發布資訊必須格外慎重。IORG 致力於可公開驗證的資料科學研究，我們釋出開放資料、充分揭露研究方法，實踐開放、透明，對公眾負責，以期獲得社會大眾的信任，並協助重建基於事實、科學的線上公共討論空間。

quote
知識的「開放」指的是任何人皆可自由存取、使用、修改、分享知識，最多受到標示來源及存續開放的限制。
——〈開放定義〉版本 2.1

---

photo


研究報告、報導

IORG 各項研究報告及報導內容，除另有標示外，皆以 IORG 為作者，以創用 CC 姓名標示－相同方式分享 4.0 國際條款，於 IORG 網站 https://iorg.tw 授權釋出。

研究架構

IORG 致力紀錄、開放各項資料蒐集及科學研究方法，以實踐可公開驗證、可重製之科學研究，推進研究方法透明化，提升整體領域研究品質。

IORG 研究架構永久網址 = https://iorg.tw/open/rm

辭典

預計開源項目：

- 資訊操弄、資訊判讀相關詞彙定義
- 詞彙華英對照
- 同義詞
- 停用詞

書單

- 開源計畫待定

典藏

經 IORG 研究員人工備份的各類文字、影音多媒體內容。典藏項目皆有編號，且多附有螢幕截圖，可於 IORG 網站 https://iorg.tw/archive/ 瀏覽，原始碼、資料可由下列連結取得。

- https://github.com/iorg-tw/archive/
- https://github.com/iorg-tw/archive/blob/master/data/archive.json

---

IORG 資訊環境資料庫

以科學研究為目的，IORG 以自動、人工各種方法取得社交媒體、即時通訊及其他網路平台各類內容及詮釋資料，並建置「IORG 資訊環境資料庫」(IORG Information Environment Archive)，保存上述資料，作為科學研究使用。

為實踐本文所宣示的開放、信任 2 項原則，IORG 提出以下計畫，逐步開放「IORG 資訊環境資料庫」之資料，釋出各項資料集。

IORG 以科學研究為目的，蒐集各類網路內容。IORG 並非內容的原始作者，僅對蒐集而來之資料及詮釋資料（metadata）予以整理、儲存，為科學研究之用。IORG 已盡最大努力，完整呈現資料原始樣貌，IORG 不保證所提供各項資料集之完整性。

使用 IORG 所提供之各項資料集，表示你同意遵守下列約定事項。

- 標示資料來源為 IORG，並提供連結至 IORG 網站 https://iorg.tw
- 將衍生成果以符合〈開放定義〉之方法、形式向公眾開放
- 不傷害、騷擾、歧視他人
- 不對 IORG 造成傷害
- 遵守當地、國際法規
- 自行承擔風險

references
- Open Definition. https://opendefinition.org/   - 什麼是開放資料？Open Data Handbook. https://opendatahandbook.org/guide/zh_TW/what-is-open-data/

note
IORG 所使用，由他人產製的各項資料，遵守其原始授權條款。 部分網路內容、社交媒體、即時訊息、田野調查、調查員、專家、學者所提供之資料，IORG 為保護資料提供者，或為遵守資料原始授權條款，不予公開。

按照公開程度，IORG 所釋出的資料集可分為「公開資料集」、「全文資料集」，二者在資料欄位、精準度有所差異。

- 公開資料集：任何人皆可自由下載、使用，資料欄位受限程度較高，使用時請遵守上述約定事項。
- 全文資料集：向 IORG 提出申請後，由 IORG 提供資料集，資料欄位受限程度較低。

按照內容原始來源，IORG 釋出資料可分為下列 3 項，細節於下詳列。

- Facebook 粉專貼文資料
- 微博貼文資料
- 可疑訊息資料

Facebook 粉專貼文資料

- 涵蓋範圍：在貼文中大量使用華語的 Facebook 粉專。
- 釋出方式：分階段釋出。

資料欄位	公開資料集	全文資料集
貼文內容	✅ 前 280 字元	✅ 全文
發布時間	✅ 精度：秒	✅ 精度：秒
備份時間	□	✅ 精度：秒
粉專名稱	□	✅
貼文 URL	□	✅

微博貼文資料

- 涵蓋範圍：已認證，或追蹤數超過 IORG 所設門檻的微博帳號。
- 釋出方式：分階段釋出。

資料欄位	公開資料集	全文資料集
貼文內容	✅ 前 280 字元	✅ 全文
發布時間	✅ 精度：秒	✅ 精度：秒
備份時間	□	✅ 精度：秒
微博帳號名稱	□	✅
貼文 URL	□	✅

可疑訊息資料

- 涵蓋範圍：經不特定使用者回報，或 IORG 研究員回報，原始來源包括 LINE 的可疑訊息。
- 釋出方式：分階段釋出。

資料欄位	公開資料集	全文資料集
訊息內容	✅ 前 140 字元	✅ 全文
回報時間	✅ 精度：秒	✅ 精度：秒
訊息原始來源	無資料	無資料

關於 **IORG** 資訊環境資料庫的 **FAQ**

備份時間、粉專或帳號名稱、貼文 **URL** 在「公開資料集」裡沒有，原因？

- 自 2020 年年初，IORG 持續調整、修正備份系統。
- 在 IORG 備份 Facebook 粉專貼文、微博帳號貼文的同時，Facebook、微博也同時有反制備份的機制。
- 期間，IORG 曾發現平台業者疑似針對特定帳號，在特定情況下，延遲或不顯示其貼文內容的狀況。
- 若為真，則表示平台業者有「針對特定帳號反制備份」的相關機制。
- 為維持備份系統持續、穩定運作，持續發展資料驅動研究方法，IORG 選擇自「公開資料集」排除部分欄位。
- 未列入「公開資料集」之資料欄位，皆可經向 IORG 提出申請後取得。

你們的「公開資料集」只有部分內文，是怎樣？

- 承上，完整內文可能對備份系統穩定運作、研究方法持續發展造成障礙，故於「公開資料集」僅提供部分內文。

為什麼這些資料集沒有明定授權條款？

- IORG 並非資料集中原始內容之作者，因此無法使用常見的各式開放授權條款。

所以「完整資料集」要怎麼申請？

- 請來信 hi [at] iorg.tw

資料標準呢？

- 在樓下說明

IORG 資料標準

為建置「IORG 資訊環境資料庫」，IORG 參考 0archive 資料標準，修改、制定 IORG 資料標準。

「IORG 資料標準」預計公告於 IORG 網站及 GitHub。

資訊科技工具

IORG 的資料科學研究借重各項開源軟體。IORG 亦以研究為目的，自行開發，並計畫開源各項資訊科技工具。

IORG 自行產製的各項原始碼，除非另有標示，皆以 IORG 為作者，於 IORG GitHub 帳號 https://github.com/iorg-tw 以 MIT 條款授權釋出。

計畫開源項目：

- 內容管理系統
- Facebook 自動備份系統
- 微博自動備份系統
- 可疑訊息人工回報系統
- 資料備份工具

內容管理系統

- 狀態：已開源
- 授權條款：MIT
- https://github.com/iorg-tw/iorg-tw-nuxt

Facebook 自動備份系統

- 狀態：開源計畫待定

微博自動備份系統

- 狀態：開源計畫待定

可疑訊息人工回報系統

- 狀態：開源計畫待定

資料備份工具

- 狀態：已公開
- 授權條款：待定
- https://github.com/iorg-tw/archive/

訪談內容

note
開放計畫待定

內部文件

為共享知識，促進研究，IORG 計畫將各項內部文件開源釋出。

計畫開放項目：

- 組織架構、分工文件
- 工作、會議記錄
- 內部檢核報告

組織架構、分工文件

- 狀態：開放計畫待定

工作、會議記錄

- 狀態：開放計畫待定

內部檢核報告

- 狀態：開放計畫待定