

科技部「回應國家重要挑戰之人工智慧主題研究專案」

永續智慧城市之指揮與監控中心設計

「用於檢測行善之智慧電表資料」

**(C01-D01-1)**

**v1.0.0**

資料集說明文件

(含去識別化及 FAIR 原則)

子計畫名稱： 永續智慧城市之資料治理自動化作業 **(SP1)**

子計畫主持人： 熊博安教授

說明文件版本：**v1.0.0**

日期：中華民國 **111** 年 **08** 月 **29** 日

# 1 資料集基本說明

|                            |        |  |
|----------------------------|--------|--|
| 資料集名稱                      |        | 用於檢測行善之智慧電表資料  |
| 資料集編號 (版本/日期)              |        | C01-D01-1 (v1.0.0 / 111 年 08 月 29 日)   |
| 資料集摘要(含目的/用途/<br>價值/標註重點等) |        | 該文件包括從美國中西部地區 2009 年 RECS 數據集中隨機選擇的 200 個家庭的電力需求概況。  |
| 資料集聯絡人                     |        | Matteo Muratori ; Matteo.Muratori@nrel.gov   |
| 影像資料                       | 資料類型   | <input type="checkbox"/> 影像(靜態) <input type="checkbox"/> 影像(動態) <input checked="" type="checkbox"/> 數據(結構化)<br><input type="checkbox"/> 文字(非結構化) <input type="checkbox"/> 語音 <input type="checkbox"/> 其他：_____ |
|                            | 資料儲存格式 | <input type="checkbox"/> MP4 <input checked="" type="checkbox"/> CSV <input type="checkbox"/> PG/JPEG <input type="checkbox"/> PNG <input type="checkbox"/> 其他_____  |
|                            | 資料來源   | 來源名稱：NREL Data Catalog<br>網址： <a href="https://data.nrel.gov/submissions/69">https://data.nrel.gov/submissions/69</a>  |
|                            | 收集期間   | 2010/01/01~2010/12/31  |
|                            | 資料筆數   | 總筆數： <u>500</u> 訓練筆數：_____ 測試筆數：_____<br>其他筆數：_____ (Validation, etc.)   |
|                            | 資料處理情形 | <input checked="" type="checkbox"/> 原始資料，未經處理  |

|                  |                     |   |
|------------------|---------------------|---|
|                  |                     | <input type="checkbox"/> 有處理，處理方式：<br>(請詳細說明處理目的、方法、使用之工具、處理後結果)<br>_____   |
|                  | 總共容量                | 1.28 MB   |
|                  | 適法性基礎<br>(本項暫不對外公開) | ● 取用資料方式：<br><input checked="" type="checkbox"/> 開放型 <input type="checkbox"/> 同意授權條款 <input type="checkbox"/> 申請審核 <input type="checkbox"/> 需付費<br><br>● 授權合約類型：<br><input checked="" type="checkbox"/> Public Domain <input type="checkbox"/> Attribution <input type="checkbox"/> Share-Alike<br><br><input type="checkbox"/> Non-Commercial <input type="checkbox"/> Database Only <input type="checkbox"/> No Derivatives<br><br>(RE: <a href="https://docs.data.world/en/59261-59714-2--Common-license-types-for-datasets.html">https://docs.data.world/en/59261-59714-2--Common-license-types-for-datasets.html</a> )<br><br>● 授權合約名稱：_____ |
|                  | 其它說明                |   |
| 標<br>註<br>資<br>料 | 標註工具                | excel   |
|                  | 標註方式                | <input checked="" type="checkbox"/> 自動標註<br><br><input type="checkbox"/> 人工標註<br><br><input type="checkbox"/> 自動標註+人工校正<br><br><input type="checkbox"/> 其他：_____  |

|           |        |   |
|-----------|--------|---|
|           | 標註類型   | <input type="checkbox"/> Regression Value <input type="checkbox"/> Class (Category) <input type="checkbox"/> Polylines<br><input type="checkbox"/> Segmentation <input type="checkbox"/> Keypoints <input type="checkbox"/> Heatmap<br><input type="checkbox"/> Temporal Detection <input type="checkbox"/> Geolocation<br><input checked="" type="checkbox"/> 其他：Boolean |
|           | 標註儲存格式 | CSV   |
|           | 標註進度   | <input checked="" type="checkbox"/> 全部標註完成： 73000 筆<br><input type="checkbox"/> 僅標註部份項目： ____ 筆<br><input type="checkbox"/> 未標註： ____ 筆 ( <input type="checkbox"/> 未來持續標註 <input type="checkbox"/> 不標註)   |
|           | 其它說明   |   |
| 與其他資料庫的串接 |        | <input type="checkbox"/> 保留與機構內其他資料庫串接可能性：未來可透過假 id 串接同機構內其他的資料庫<br><input type="checkbox"/> 保留與國網中心資料庫串接可能性：與上傳國網中心的資料採同樣的假 ID，以利後續資料更新/維護（但僅限由本團隊主動進行更新與維護）   |

## 2 欄位說明

| 欄位<br>編號     | 欄位名稱<br>(呈現之欄位標題) | 欄位格式<br>(數值/字串/文字等)<br>[限制：大小、小數點等] | 必<br>/<br>選 | 範例/欄位說明/<br>資料庫資訊或檔名    |
|--------------|-------------------|-------------------------------------|-------------|-------------------------|
| A2-<br>A366  | 日期<br>(datetime)  | 日期                                  |             | 2010/01/01-2010/12/31   |
| B1-<br>GS1   | 住戶<br>(Household) | 文字                                  |             | Household1-Household200 |
| B2-<br>GS366 | 電表異常值             | Boolean                             |             | 0 表示正常，1 表示異常           |
|              |                   |                                     |             |                         |
|              |                   |                                     |             |                         |
|              |                   |                                     |             |                         |
|              |                   |                                     |             |                         |
|              |                   |                                     |             |                         |
|              |                   |                                     |             |                         |

### 3 標註說明範例

本章節使用到下述檔案

| 標註項目    | 資料檔                      | 標註資料檔                          |
|---------|--------------------------|--------------------------------|
| 每個電表異常值 | ResidentialProfiles_days | ResidentialProfiles_days_label |

#### 3.1 資料與標註資料的對應關係

Label: 每個電表異常值  $\geq$  平均值 + 3 \* 標準差

Label 為異常標籤，0 表示正常，1 表示異常

#### 3.2 標註內容

標註欄位說明

| 欄位名稱 | 說明      | 檔案截圖 |
|------|---------|------|
| 橫軸   | 200 個電表 |      |
| 縱軸   | 365 天   |      |

### 4 工具

可以使用下列工具從儲存的檔案中取得資料

| 資料類別 | 公開/免費軟體 | 商用軟體  |
|------|---------|-------|
| 資料   | NA      | excel |
| 標註資料 | NA      | excel |

## 5 去識別化

|        |   |
|--------|---|
| 去識別化方式 | <input checked="" type="checkbox"/> 遵循本專案共同制訂方式                     |
|        | <input type="checkbox"/> 本資料集自行定義方式 (勾選本項請逐欄位說明各欄位之處理作法)<br><br>說明： |

### 本專案共同制訂方式

本專案進行去識別化之處理程序如下：

1. 定義需求:若資料有特定接收對象，需與資料需求者溝通，確定需要納入哪些資料，在此過程需要對資料審慎考量。因為有些資料對於計畫目的而言是有用，但不是

關鍵性資料，此資料是否需要保留就必須有所取捨。其考量因素包含對資料使用者信任程度、組織對隱私保護政策要求等。確定資料範圍後，必須進行分析出直接識別符，並進行相關處理。

- 步驟 1.瞭解需求，確定資料集範圍、規模
- 步驟 2.判定最小可接受使用之資料
- 步驟 3.判定資料屬性
- 步驟 4.將直接識別資料遮罩

2. 重新識別風險規劃:設計並記錄去識別化風險分析的方法及過程。內容包含採用去識別化之方法、威脅模型建立、處理步驟及相關風險參數之確立等。

- 步驟 5.設定組織重新識別風險門檻
- 步驟 6.建立重新識別風險評估模型
- 步驟 7.確定去識別化門檻
- 步驟 8.訂定去識別化相關參數

3. 去識別化實作及驗證：使用選擇的去識別化方法重複對資料進行去識別化處理，計算實施結果直到符合相關風險參數。若是資料量很大，就效率的考量，實做上可採用抽樣技術輔助事先預演操作，在確定符合風險參數後再套入原本的資料集進行處理。

- 步驟 9.抽樣資料
- 步驟 10.評估處理後資料之風險



- 步驟 11.比較風險與門檻值，並進行相應處理

- 步驟 12.特殊離群值處理

4. 解決方案驗證:在完成去識別化處理之後，需對特殊的資料處理，防止具特殊疏離群值的PII被重新識別。此外，針對防止攻擊者導入外部資料比對獲得PII，一般建議需抽樣相當數量的資訊至常被使用的網站搜尋，分析是否可查詢到個人資訊，以驗證去識別化的完整性。最後需記錄整個作業程序，作為善個資保護盡義務之佐證，並可供後續類似工作參考。

- 步驟 13.對解決方案進行診斷

- 步驟 14.記錄執行過程及相關參數

5. 週期性持續審視：在資料的整個生命週期過程中持續、週期性重新評估現行的去識別化機制是否仍符合風險需求。需因應威脅發展及技術改變，適度調整去識別化處理的程序。相關的審視需包含適當矯正措施及預防動作程序。

- 步驟 15.定期或是針對特事件檢核資料安全性

- 步驟 16.週期性風險衝擊評鑑

屬於本計畫資料集內部需要去識別化的資料整理如下：

| 資料集    | 識別資料與特性<br>(直接識別符、間接識別符) | 去識別動作 (壓抑、模糊、概化、縱向資料一致性、開放式本文處理、保留原狀),  |
|--------|--------------------------|---|
| UAV 影像 | 人臉 (直接識別符)               | 模糊化 (Automatic Face Redaction)          |
| 交通影像   | 車牌 (直接識別符)               | 模糊化 (Automatic License Plate Redaction) |

| 資料集 | 識別資料與特性<br>(直接識別符、間接識別符) | 去識別動作 (壓抑、模糊、概化、縱向資料一致性、開放式本文處理、保留原狀), |
|-----|--------------------------|--|
|     |                          |  |
|     |                          |  |
|     |                          |  |
|     |                          |  |
|     |                          |  |

## 6 FAIR 評估

本計畫依據科技部之要求，將檢驗計畫中所產出之每一資料集是否符合 FAIR 原則：

- 可搜尋(Findable)：資料必須有唯一的識別符並於檢索資源中有效的標示出來；
- 可近用(Accessible)：必須易於透過開放系統和有效且安全的身份驗證與授權程序進行檢索；
- 可互操作(Interoperable)：必須使用標準的語彙及格式；
- 可再利用(Reusable)：必須向使用者充分描述資料及使用授權，並可追溯資料來源。

### (1) FAIR 15項細部原則

FAIR 原則最初僅是 4 項大原則，2016 年 Mark D. Wilkinson 發表的 “The FAIR Guiding Principles for Scientific Data Management and Stewardship” 一文中，定義了 15 項細部原則（詳表 1），讓 4 項大原則可更具體被實施。

### (2) FAIR 原則符合度之評估方法

本指引採用「研究資料聯盟」(The Research Data Alliance，簡稱 RDA)所發表的“FAIR Data Maturity Model Specification and Guidelines”所定義的 41 項評估指標，做為資料集之 FAIR 原則符合度(即成熟度)評估基準。RDA 提出「衡量進展」與「衡量“通過”或“不通過”」兩種評估方式，第一種(衡量進展)評估方式，可看出各指標的完成程度，以及待改善之處，較適合剛要導入 FAIR 原則之實作者。因此，本指

引採用「衡量進展」評估方式，其評估表詳指引之附件一。「衡量“通過”或“不通過”」的評估方式可參考指引之附件二。

以下為本資料集之 **FAIR** 自評表。

## 附件 5-7.「符合 FAIR 原則」達成程度最高的 1 項資料集之 FAIR 自評表

註：請總計畫自行評估後提供。

資料集名稱：用於檢測行善之智慧電表資料

所屬子計畫編號及名稱：子計畫一：永續智慧城市之資料治理自動化作業

欄位填寫說明：

1. 各計畫依該評估指標的完成狀況就-不適用、尚未考慮、規劃階段、實施階段、全面實施等 5 個欄位中以打勾「V」符號勾選出 1 個符合的狀態。
2. 「佐證資料說明」欄位是填寫該項評估指標的實施內容或佐證資料

| 指標代號           | 評估指標           | 評估指標說明                | 不適用 | 尚未考慮 | 規劃階段 | 實施階段 | 全面實施 | 佐證資料說明   |
|----------------|----------------|-----------------------|-----|------|------|------|------|--|
| 可搜尋 (Findable) |                |                       |     |      |      |      |      |  |
| RDA-F1-01M     | 詮釋資料由持久性識別符所標識 | 查證詮釋資料的識別符是列在持久性註冊服務中 |     |      |      |      | V    | 已在 <a href="http://purl.org/">http://purl.org/</a> 網站上為資料集說明文件登記一個PURL識別符。<br>以下為此計畫ICCC平台實際登錄PURL之網站<br><a href="https://purl.archive.org/purl/iccc-platform/smart-meter-outage-prediction">https://purl.archive.org/purl/iccc-platform/smart-meter-outage-prediction</a> |
| RDA-F1-01D     | 資料由持久性識別符所標識   | 查證資料的識別符是列在持久性註冊服務中   |     |      |      |      | V    | 已在 <a href="http://purl.org/">http://purl.org/</a> 網站上為資料集登記一個PURL識別符。<br>以下為此計畫ICCC平台實際登錄PURL之網站<br><a href="https://purl.archive.org/purl/iccc-platform/smart-meter-outage-prediction">https://purl.archive.org/purl/iccc-platform/smart-meter-outage-prediction</a>     |
| RDA-F1-02M     | 詮釋資料由全球唯一      | 查證詮釋資料的識別符            |     |      |      |      | V    | 已在 <a href="http://purl.org/">http://purl.org/</a> 網站上為資料集說明文件登  |

| 指標代號       | 評估指標           | 評估指標說明                                  | 不適用 | 尚未考慮 | 規劃階段 | 實施階段 | 全面實施 | 佐證資料說明   |
|------------|----------------|---|-----|------|------|------|------|--|
|            | 識別符所標識         | 是列在註冊服務中                                |     |      |      |      |      | 記一個PURL識別符。<br>以下為此計畫ICCC平台實際登錄PURL之網站<br><a href="https://purl.archive.org/purl/iccc-platform/smart-meter-outage-prediction">https://purl.archive.org/purl/iccc-platform/smart-meter-outage-prediction</a><br>以下為DOI(數位物件識別碼)<br><a href="https://doi.org/10.5281/ZENODO.7134896">https://doi.org/10.5281/ZENODO.7134896</a>  |
| RDA-F1-02D | 資料由全球唯一識別符所標識  | 查證資料的識別符是列在註冊服務中                        |     |      |      |      | V    | 已在 <a href="http://purl.org/">http://purl.org/</a> 網站上為資料集登記一個PURL識別符。<br>以下為此計畫ICCC平台實際登錄PURL之網站<br><a href="https://purl.archive.org/purl/iccc-platform/smart-meter-outage-prediction">https://purl.archive.org/purl/iccc-platform/smart-meter-outage-prediction</a><br>以下為DOI(數位物件識別碼)<br><a href="https://doi.org/10.5281/ZENODO.7134896">https://doi.org/10.5281/ZENODO.7134896</a> |
| RDA-F2-01M | 提供豐富的詮釋資料以利被發現 | 查證有提供豐富的詮釋資料                            |     |      |      |      | V    | 資料集說明文件可以清楚說明資料集內容及資料集存放位置。<br>以下為此計畫ICCC平台實際登錄PURL之網站可供參照<br><a href="https://purl.archive.org/purl/iccc-platform/smart-meter-outage-prediction">https://purl.archive.org/purl/iccc-platform/smart-meter-outage-prediction</a>  |
| RDA-F3-01M | 詮釋資料內含資料的識別符   | 查證在詮釋資料元素(Metadata element)中有包含"資料的識別符" |     |      |      |      | V    | 資料集說明文件中有記載資料集的識別符。<br>以下為此計畫ICCC平台實際登錄PURL之網站<br><a href="https://purl.archive.org/purl/iccc-platform/smart-meter-outage-prediction">https://purl.archive.org/purl/iccc-platform/smart-meter-outage-prediction</a><br>以下為DOI(數位物件識別碼)<br><a href="https://doi.org/10.5281/ZENODO.7134896">https://doi.org/10.5281/ZENODO.7134896</a>  |
| RDA-F4-01M | 詮釋資料可以被收集      | 查證詮釋資料是可以索                              |     |      |      |      | V    | 資料集的說明頁面可被瀏覽器找到。   |

| 指標代號                  | 評估指標                | 評估指標說明                             | 不適用 | 尚未考慮 | 規劃階段 | 實施階段 | 全面實施 | 佐證資料說明  |
|-----------------------|---------------------|------------------------------------|-----|------|------|------|------|---|
|                       | 和索引                 | 引                                  |     |      |      |      |      | 在計畫所自建的共享平台上有提供查詢介面以找到資料集的說明頁面或資料集說明文件。<br>可透過在 <a href="https://github.com/">https://github.com/</a> 上搜尋關鍵字找到。   |
| 可近用(可取用) (Accessible) |                     |                                    |     |      |      |      |      |   |
| RDA-A1-01M            | 詮釋資料包含讓使用者能夠取用資料的資訊 | 查證詮釋資料中提供的資訊讓使用者能夠知道如何取用資料         |     |      |      |      | V    | 資料集的說明文件中有記載可以讓使用者能夠取用資料的資訊(如：取用資料的限制，或申請方法，或身份驗證/授權之說明)。<br>可透過以下網站前往且說明中含有資料取得方式<br><a href="https://purl.archive.org/purl/iccc-platform/smart-meter-outage-prediction">https://purl.archive.org/purl/iccc-platform/smart-meter-outage-prediction</a>   |
| RDA-A1-02M            | 可以人工取用詮釋資料 (即人工干預)  | 查證需在人為干預下取用詮釋資料                    | V   |      |      |      |      | 下載資料集說明文件，需額外人工聯絡申請。  |
| RDA-A1-02D            | 可以人工取用資料 (即人工干預)    | 查證詮釋資料中有描述“如何透過人為干預獲得對數位物件的取用權”的資訊 | V   |      |      |      |      | 下載資料集，需額外人工聯絡申請。<br>資料集說明文件中有記載“如何人工申請資料集的使用權”的資訊。  |
| RDA-A1-03M            | 詮釋資料識別符可以解析為詮釋資料記錄  | 查證可以使用識別符來取用其詮釋資料                  |     |      |      |      | V    | 可透過 <a href="https://purl.org/">https://purl.org/</a> 詮釋資料PURL識別符，連至該PURL識別符所代表的資料集的說明頁面或資料集說明文件。<br>以下為此計畫ICCC平台實際登錄PURL之網站<br><a href="https://purl.archive.org/purl/iccc-platform/smart-meter-outage-prediction">https://purl.archive.org/purl/iccc-platform/smart-meter-outage-prediction</a> |

| 指標代號         | 評估指標                 | 評估指標說明                         | 不適用 | 尚未考慮 | 規劃階段 | 實施階段 | 全面實施 | 佐證資料說明   |
|--------------|----------------------|--------------------------------|-----|------|------|------|------|--|
| RDA-A1-03D   | 資料識別符可以解析為數位物件       | 查證可以使用識別符來得到數位物件               |     |      |      |      | V    | 可透過 <a href="https://purl.org/">https://purl.org/</a> 資料集PURL識別符，連至該PURL識別符所代表的資料集的下載頁面。<br>以下為此計畫ICCC平台實際登錄PURL之網站<br><a href="https://purl.archive.org/purl/iccc-platform/smart-meter-outage-prediction">https://purl.archive.org/purl/iccc-platform/smart-meter-outage-prediction</a> |
| RDA-A1-04M   | 透過標準化協定取用詮釋資料        | 查證詮釋資料的取用方式是透過標準化協定            |     |      |      |      | V    | 這是Github建立的共享Source code網址：<br><a href="https://github.com/ICCC-Platform/Smart-Meter-Outage-Prediction">https://github.com/ICCC-Platform/Smart-Meter-Outage-Prediction</a>   |
| RDA-A1-04D   | 透過標準化協定取用資料          | 查證資料的取用方式是透過標準化協定              |     |      |      |      | V    | 這是Github建立的共享Source code網址：<br><a href="https://github.com/ICCC-Platform/Smart-Meter-Outage-Prediction">https://github.com/ICCC-Platform/Smart-Meter-Outage-Prediction</a>   |
| RDA-A1-05D   | 可以自動化取用資料(即透過計算機程序)  | 查證透過解析持久識別符可自動化取用資料            |     | V    |      |      |      | 目前此資料集因為很簡單，沒有提供API之必要性。   |
| RDA-A1.1-01M | 詮釋資料可經由免費取用協定取用      | 查證此協定之使用為免費                    |     |      |      |      | V    | 這是Github建立的共享Source code網址：<br><a href="https://github.com/ICCC-Platform/Smart-Meter-Outage-Prediction">https://github.com/ICCC-Platform/Smart-Meter-Outage-Prediction</a>   |
| RDA-A1.1-01D | 資料可經由免費取用協定取用        | 查證此協定之使用為免費                    |     |      |      |      | V    | 這是Github建立的共享Source code網址：<br><a href="https://github.com/ICCC-Platform/Smart-Meter-Outage-Prediction">https://github.com/ICCC-Platform/Smart-Meter-Outage-Prediction</a>   |
| RDA-A1.2-01D | 資料可經由支持身份驗證和授權取用協定取用 | 評斷協定中存在身份驗證和授權處理機制             | V   |      |      |      |      | 取用資料集前需要先登入帳號。   |
| RDA-A2-01M   | 詮釋資料保證在資料不再可用後仍然可用   | 根據有關詮釋資料和資料的生命週期之資訊來評估此指標，此資訊應 |     |      |      |      | V    | 資料集的說明頁面有說明如果資料集不再可用，此說明網頁將仍然存在。<br>可透過PURL方式保留詮釋資料，以下為此計畫   |



| 指標代號                 | 評估指標                 | 評估指標說明                                      | 不適用 | 尚未考慮 | 規劃階段 | 實施階段 | 全面實施 | 佐證資料說明  |
|----------------------|----------------------|---|-----|------|------|------|------|---|
|                      |                      | 表明如果資料不再可用，詮釋資料仍然將保持可用                      |     |      |      |      |      | ICCC平台實際登錄PURL之網站<br><a href="https://purl.archive.org/purl/iccc-platform/smart-meter-outage-prediction">https://purl.archive.org/purl/iccc-platform/smart-meter-outage-prediction</a>  |
| 可互操作 (Interoperable) |                      |   |     |      |      |      |      |   |
| RDA-I1-01M           | 詮釋資料使用以標準化格式表示的知識表達  | 查證詮釋資料有使用管制語彙表來表示知識                         |     | √    |      |      |      | 目前尚無此需求，未來有需求時提供。   |
| RDA-I1-01D           | 資料使用以標準化格式表示的知識表達    | 查證資料所使用的標準是適用於數位物件的領域及類型                    |     | √    |      |      |      | 目前尚無此需求，未來有需求時提供。   |
| RDA-I1-02M           | 詮釋資料使用機器可理解的知識表達     | 查證用於表示詮釋資料的知識表達模型為機器可理解                     |     | √    |      |      |      | 目前尚無此需求，未來有需求時提供。   |
| RDA-I1-02D           | 資料使用機器可理解的知識表達       | 查證用於表示資料的知識表達模型為機器可理解                       |     | √    |      |      |      | 目前尚無此需求，未來有需求時提供。   |
| RDA-I2-01M           | 詮釋資料使用符合 FAIR 標準的語彙表 | 查證詮釋資料中使用的每個語彙表都記錄在案並可使用全球唯一和持久的識別符對語彙表進行解析 |     |      |      |      | √    | 資料集的詮釋資料所使用到的每個語彙表都記錄在案並可使用全球唯一和持久的識別符對語彙表進行解析。<br>以下為此計畫ICCC平台實際登錄PURL之網站<br><a href="https://purl.archive.org/purl/iccc-platform/smart-meter-outage-prediction">https://purl.archive.org/purl/iccc-platform/smart-meter-outage-prediction</a> |
| RDA-I2-01D           | 資料使用符合 FAIR          | 查證資料中使用的每個                                  |     | √    |      |      |      | 目前尚無此需求，未來有需求時提供。   |

| 指標代號            | 評估指標               | 評估指標說明                          | 不適用 | 尚未考慮 | 規劃階段 | 實施階段 | 全面實施 | 佐證資料說明  |
|-----------------|--------------------|---------------------------------|-----|------|------|------|------|---|
|                 | 標準的語彙表             | 語彙表都記錄在案並可使用全球唯一和持久的識別符對語彙表進行解析 |     |      |      |      |      |   |
| RDA-I3-01M      | 詮釋資料內含對其他詮釋資料的引用   | 查證詮釋資料是存在對其他詮釋資料的引用             |     |      |      |      | V    | 資料集的說明文件中提到對其他詮釋資料的引用。<br>引用以下網站的資料來訓練<br><a href="https://www.ausgrid.com.au/Industry/Our-Research/Data-to-share/Past-outage-data">https://www.ausgrid.com.au/Industry/Our-Research/Data-to-share/Past-outage-data</a> |
| RDA-I3-01D      | 資料內含對其他資料的引用       | 查證資料中是存在對其他資料的引用                |     | V    |      |      |      | 目前尚無此需求，未來有需求時提供。   |
| RDA-I3-02M      | 詮釋資料內含對其他資料的引用     | 查證詮釋資料中是存在對其他資料的引用              |     | V    |      |      |      | 目前尚無此需求，未來有需求時提供。。  |
| RDA-I3-02D      | 資料內含對其他資料的限定引用     | 查證相關資源的引用和數位物件間是否有特定關係角色        |     | V    |      |      |      | 目前尚無此需求，未來有需求時提供。   |
| RDA-I3-03M      | 詮釋資料內含對其他詮釋資料的限定引用 | 查證相關資源的引用與描述資源是否有特定關係           |     | V    |      |      |      | 目前尚無此需求，未來有需求時提供。   |
| RDA-I3-04M      | 詮釋資料內含對其他資料的限定引用   | 查證相關資源的引用與描述資源是否有特定關係角色         |     | V    |      |      |      | 目前尚無此需求，未來有需求時提供。   |
| 可再利用 (Reusable) |                    |                                 |     |      |      |      |      |   |

| 指標代號         | 評估指標                 | 評估指標說明                         | 不適用 | 尚未考慮 | 規劃階段 | 實施階段 | 全面實施 | 佐證資料說明  |
|--------------|----------------------|--------------------------------|-----|------|------|------|------|---|
| RDA-R1-01M   | 提供了多個準確和相關的屬性以便再利用   | 查看詮釋資料的內容和品質                   |     |      |      |      | V    | 資料集的說明文件中的對資料集的說明是足夠充分與準確。<br>可參照以下網站，此網站為計畫ICCC平台PURL<br><a href="https://purl.archive.org/purl/iccc-platform/smart-meter-outage-prediction">https://purl.archive.org/purl/iccc-platform/smart-meter-outage-prediction</a>   |
| RDA-R1.1-01M | 詮釋資料內含資料可再利用的授權合約之資訊 | 在詮釋資料中查看已提供資料可再利用的授權合約之資訊      |     |      |      |      | V    | 資料集的說明文件中有記載”資料可再利用的授權合約”的資訊。<br>可參照以下網站，該詮釋資料是CC0且免費使用<br><a href="https://www.ausgrid.com.au/Industry/Our-Research/Data-to-share/Past-outage-data">https://www.ausgrid.com.au/Industry/Our-Research/Data-to-share/Past-outage-data</a>  |
| RDA-R1.1-02M | 詮釋資料參照標準可再利用授權合約     | 查證授權合約確實是標準授權合約                |     |      |      |      | V    | 資料集的說明文件中有記載”資料可再利用的授權合約”的資訊，且授權合約確實是標準授權合約(如：Creative Commons licenses , Open Data Commons)。<br>可參照以下網站，該詮釋資料是CC0且免費使用<br><a href="https://www.ausgrid.com.au/Industry/Our-Research/Data-to-share/Past-outage-data">https://www.ausgrid.com.au/Industry/Our-Research/Data-to-share/Past-outage-data</a>                    |
| RDA-R1.1-03M | 詮釋資料參照機器可理解的可再利用授權合約 | 查證連到授權合約的連結可以解析出機器可理解的授權條件之表達式 |     |      |      |      | V    | 連到授權合約的連結可以解析出機器可理解的授權條件之表達式(如：RDF expression of Creative Commons licences 或是 Open Data Rights Language (ODRL))。<br>可參照以下網站，該詮釋資料是CC0且免費使用<br><a href="https://www.ausgrid.com.au/Industry/Our-Research/Data-to-share/Past-outage-data">https://www.ausgrid.com.au/Industry/Our-Research/Data-to-share/Past-outage-data</a> |
| RDA-R1.2-01M | 詮釋資料內含採用社            | 查證資料出處的資訊有                     |     |      |      |      | V    | 資料集的說明文件中有記載資料集出處的資訊，   |

| 指標代號         | 評估指標                 | 評估指標說明                    | 不適用 | 尚未考慮 | 規劃階段 | 實施階段 | 全面實施 | 佐證資料說明  |
|--------------|----------------------|---------------------------|-----|------|------|------|------|---|
|              | 群標準的出處資訊             | 遵循社群標準                    |     |      |      |      |      | 且出處資訊有遵循社群標準。<br>可參照以下網站，該資料為CC0且遵循社群標準<br><a href="https://www.ausgrid.com.au/Industry/Our-Research/Data-to-share/Past-outage-data">https://www.ausgrid.com.au/Industry/Our-Research/Data-to-share/Past-outage-data</a>  |
| RDA-R1.2-02M | 詮釋資料內含依據跨領域語言的出處資訊   | 查證出處資訊有使用跨領域語言            | V   |      |      |      |      | 出處資訊是是使用跨領域語言(例如 PROV-O)。   |
| RDA-R1.3-01M | 詮釋資料符合社群標準           | 查證詮釋資料是遵循社群標準             |     |      |      | V    |      | 資料集的說明文件的格式符合社群標準，例如該詮釋資料標準已收錄在註冊表中(如 FAIRsharing 所開發的)。<br>我們將資源放在Github上，可參照以下網站<br><a href="https://github.com/ICCC-Platform/Smart-Meter-Outage-Prediction">https://github.com/ICCC-Platform/Smart-Meter-Outage-Prediction</a>                                       |
| RDA-R1.3-01D | 資料符合社群標準             | 查證資料是遵循社群標準               |     |      |      | V    |      | 資料集的資料模型與格式符合社群標準，該資料模型與格式之標準已收錄在註冊表中(如 FAIRsharing 所開發的)。<br>可參照以下網站，該詮釋資料是CC0且符合社群標準<br><a href="https://www.ausgrid.com.au/Industry/Our-Research/Data-to-share/Past-outage-data">https://www.ausgrid.com.au/Industry/Our-Research/Data-to-share/Past-outage-data</a> |
| RDA-R1.3-02M | 詮釋資料以符合機器可理解的社群標準來表示 | 查證詮釋資料採用的社群標準是具有機器可理解的表達式 |     |      |      | V    |      | 資料集的詮釋資料的格式符合社群標準，具有機器可理解的表達式。<br>我們使用python，故是具機器可理解之表達式<br>可參照以下網站 <a href="https://github.com/ICCC-Platform/Smart-Meter-Outage-Prediction">https://github.com/ICCC-Platform/Smart-Meter-Outage-Prediction</a>  |

| 指標代號         | 評估指標               | 評估指標說明                  | 不適用 | 尚未考慮 | 規劃階段 | 實施階段 | 全面實施 | 佐證資料說明  |
|--------------|--------------------|-------------------------|-----|------|------|------|------|---|
| RDA-R1.3-02D | 資料以符合機器可理解的社群標準來表示 | 查證資料採用的社群標準是具有機器可理解的表達式 |     |      |      |      | V    | 資料集的資料模型與格式符合社群標準，具有機器可理解的表達式。<br>可參照以下網站，該資料符合社群標準是具機器可理解<br><a href="https://www.ausgrid.com.au/Industry/Our-Research/Data-to-share/Past-outage-data">https://www.ausgrid.com.au/Industry/Our-Research/Data-to-share/Past-outage-data</a> |

## 附件二 衡量「通過」或「不通過」的評估方式

(註：本專案未採用此方式，此附件僅供參考)

此種評估方式的重點在於確定被評估內容是否滿足“通過”或“不通過”二分法判定的要求，完全滿足該評估指標要求就給 1 分，反之不能完全滿足的就給 0 分，此方式中 RDA 將 41 個評估指標依其重要性分成 “必要(Essential)”、“重要(Important)”、“有用(Useful)”三類，如下表(評估指標的重要性分類表)所示。

表 2 評估指標的重要性分類表

| FAIR原則<br>代號 |    | 評估指標代號     | 評估指標(indicator)     | 必要<br>(Essential) | 重要<br>(Important) | 有用( Useful<br>) |
|--------------|----|------------|---------------------|-------------------|-------------------|-----------------|
| <b>F</b>     | F1 | RDA-F1-01M | 詮釋資料由持久性識別符所標識      | ●                 |                   |                 |
|              |    | RDA-F1-01D | 資料由持久性識別符所標識        | ●                 |                   |                 |
|              |    | RDA-F1-02M | 詮釋資料由全球唯一識別符所標識     | ●                 |                   |                 |
|              |    | RDA-F1-02D | 資料由全球唯一識別符所標識       | ●                 |                   |                 |
|              | F2 | RDA-F2-01M | 提供豐富的詮釋資料以利被發現      | ●                 |                   |                 |
|              | F3 | RDA-F3-01M | 詮釋資料內含資料的識別符        | ●                 |                   |                 |
|              | F4 | RDA-F4-01M | 詮釋資料的可以被收集和索引       | ●                 |                   |                 |
| <b>A</b>     | A1 | RDA-A1-01M | 詮釋資料包含讓使用者能夠取用資料的資訊 |                   | ●                 |                 |
|              |    | RDA-A1-02M | 可以人工取用詮釋資料 (即人工干預)  | ●                 |                   |                 |
|              |    | RDA-A1-02D | 可以人工取用資料 (即人工干預)    | ●                 |                   |                 |
|              |    | RDA-A1-03M | 詮釋資料識別符可以解析為詮釋資料記錄  | ●                 |                   |                 |

| FAIR原則<br>代號 |      | 評估指標代號       | 評估指標(indicator)        | 必要<br>(Essential) | 重要<br>(Important) | 有用( Useful<br>) |
|--------------|------|--------------|------------------------|-------------------|-------------------|-----------------|
|              |      | RDA-A1-03D   | 資料識別符可以解析為數位物件         | ●                 |                   |                 |
|              |      | RDA-A1-04M   | 透過標準化協定取用詮釋資料          | ●                 |                   |                 |
|              |      | RDA-A1-04D   | 透過標準化協定取用資料            | ●                 |                   |                 |
|              |      | RDA-A1-05D   | 可以自動化取用資料 ( 即透過計算機程序 ) |                   | ●                 |                 |
|              | A1.1 | RDA-A1.1-01M | 詮釋資料可經由免費取用協定取用        | ●                 |                   |                 |
|              |      | RDA-A1.1-01D | 資料可經由免費取用協定取用          |                   | ●                 |                 |
|              | A1.2 | RDA-A1.2-01D | 資料可經由支持身份驗證和授權取用協定取用   |                   |                   | ●               |
|              | A2   | RDA-A2-01M   | 詮釋資料保證在資料不再可用後仍然可用     | ●                 |                   |                 |
| I            | I1   | RDA-I1-01M   | 詮釋資料使用以標準化格式表示的知識表達    |                   | ●                 |                 |
|              |      | RDA-I1-01D   | 資料使用以標準化格式表示的知識表達      |                   | ●                 |                 |
|              |      | RDA-I1-02M   | 詮釋資料使用機器可理解的知識表達       |                   | ●                 |                 |

| FAIR原則<br>代號 |      | 評估指標代號       | 評估指標(indicator)      | 必要<br>(Essential) | 重要<br>(Important) | 有用( Useful<br>) |
|--------------|------|--------------|----------------------|-------------------|-------------------|-----------------|
|              |      | RDA-I1-02D   | 資料使用機器可理解的知識表達       |                   | ●                 |                 |
|              | I2   | RDA-I2-01M   | 詮釋資料使用符合 FAIR 標準的語彙表 |                   | ●                 |                 |
|              |      | RDA-I2-01D   | 資料使用符合 FAIR 標準的語彙表   |                   |                   | ●               |
|              | I3   | RDA-I3-01M   | 詮釋資料內含對其他詮釋資料的引用     |                   | ●                 |                 |
|              |      | RDA-I3-01D   | 資料內含對其他資料的引用         |                   |                   | ●               |
|              |      | RDA-I3-02M   | 詮釋資料內含對其他資料的引用       |                   |                   | ●               |
|              |      | RDA-I3-02D   | 資料內含對其他資料的限定引用       |                   |                   | ●               |
|              |      | RDA-I3-03M   | 詮釋資料內含對其他詮釋資料的限定引用   |                   | ●                 |                 |
|              |      | RDA-I3-04M   | 詮釋資料內含對其他資料的引用       |                   |                   | ●               |
| R            | R1   | RDA-R1-01M   | 提供了多個準確和相關的屬性以便再利用   | ●                 |                   |                 |
|              | R1.1 | RDA-R1.1-01M | 詮釋資料內含資料可再利用的授權合約之資訊 | ●                 |                   |                 |



| FAIR原則<br>代號 |      | 評估指標代號       | 評估指標(indicator)      | 必要<br>(Essential) | 重要<br>(Important) | 有用( Useful<br>) |
|--------------|------|--------------|----------------------|-------------------|-------------------|-----------------|
|              |      | RDA-R1.1-02M | 詮釋資料參照標準可再利用授權合約     |                   | ●                 |                 |
|              |      | RDA-R1.1-03M | 詮釋資料參照機器可理解的可再利用授權合約 |                   | ●                 |                 |
|              | R1.2 | RDA-R1.2-01M | 詮釋資料內含採用社群標準的出處資訊    |                   | ●                 |                 |
|              |      | RDA-R1.2-02M | 詮釋資料內含依據跨領域語言的出處資訊   |                   |                   | ●               |
|              | R1.3 | RDA-R1.3-01M | 詮釋資料符合社群標準           | ●                 |                   |                 |
|              |      | RDA-R1.3-01D | 資料符合社群標準             | ●                 |                   |                 |
|              |      | RDA-R1.3-02M | 詮釋資料以符合機器可理解的社群標準來表示 | ●                 |                   |                 |
|              |      | RDA-R1.3-02D | 資料以符合機器可理解的社群標準來表示   |                   | ●                 |                 |

此「通過」或「不通過」的評估方式，FAIR 成熟度共分成了 0~5 共六個級，只要未做到”必要”類的所有評估指標(共 20 項)就只會停留在第 0 級成熟度，通過”必要”類的所有評估指標才會進到第 1 級成熟度，依序完成 3 類的所有評估指標後才會到最後的第 5 級成熟度，各級

別的成熟度的判斷標準如以下表(**FAIR** 成熟度等級表)所示。

表 3 FAIR 成熟度等級表

| 成熟度     | 必要 | 重要 | 有用 | 說明                               |
|---------|----|----|----|----------------------------------|
| Level 0 | ○  |    |    | 不是 FAIR                          |
| Level 1 | ●  |    |    | 須滿足所有「必要」項評估指標                   |
| Level 2 | ●  | ◎  |    | 須滿足所有「必要」項加上半數「重要」項評估指標          |
| Level 3 | ●  | ●  |    | 須滿足所有「必要」項加上所有「重要」項評估指標          |
| Level 4 | ●  | ●  | ◎  | 須滿足所有「必要」項加上所有「重要」項以及半數「有用」項評估指標 |
| Level 5 | ●  | ●  | ●  | 須滿足所有「必要」項加上所有「重要」項以及所有「有用」項評估指標 |

符號說明：

○ 評估指標沒有一個通過

◎ 評估指標有一半通過

● 評估指標全部通過

依據上述說明，本資料集目前自評之成熟度為：**Level 5**