

法務部政府資料開放諮詢小組 114 年第 2 次會議紀錄

一、時間：114年10月3日(星期五)上午9時30分

二、地點：本部4樓401會議室

三、主席：黃政務次長世杰

四、出(列)席單位及人員：(如簽到表)

五、上次會議決議事項辦理情形：(詳如簡報)

主席裁示：同意解除列管。

六、報告事項：

(一)報告「本部及所屬機關資料盤點表盤點結果」及「民間需求回應說明」。
說明：

1. 依政府資料開放諮詢小組設置要點第 10 點，略以「應按開會頻率檢視該機關與所屬機關提報之盤點資料及民間需求之回應說明，涉機關主管法規疑義者，應會同機關法制單位檢討修正，並確認回應說明之適切性；涉機關業務流程或資訊系統問題者，應排定改善期程，定期追蹤列管；涉跨部會法規疑義者，應提報行政院諮詢小組」規定辦理。
2. 提會報告「本部及所屬機關資料盤點表最新結果」及 114 年 5 月至 9 月「民間需求回應說明」辦理情形，俾提交委員檢視及確認回應說明之適切性。

主席裁示：依本部所屬機關(單位)建議修正「本部及所屬機關資料盤點表」之「系統、資料庫描述」欄位內容，以臻完善，並同意整併第 55 項及第 87 項重複盤點之廉政業務管理系統，另為利於委員審查盤點情形，請資訊處後續於會議現場準備相關紙本資料供委員參閱。

(二)報告「本部及所屬機關於政府資料開放平臺之資料開放推動成果」。

說明：提會報告本部於政府資料開放平臺(data.gov.tw) 114 年 5 月至 9 月開放推動成果。

主席裁示：洽悉。

(三)報告「本部及所屬機關資料品質標章精進措施執行成果」。

說明：為依行政院「政府資料開放優質標章及深化應用獎勵措施」提升政府資料品質與加值應用效益，爰提會報告本部及所屬機關執行資料集品質標章精進措施之辦理情形。

主席裁示：請本部所屬機關(單位)就尚未完成改善之資料集，確實於 114 年底前完成。

(四) 報告「本部配合數位發展部之建構臺灣主權 AI 訓練語料庫行動計畫推動情形」。

說明：數位發展部於 114 年 7 月 7 日訂定發布「建構臺灣主權 AI 訓練語料庫行動計畫」，爰提會報告本部及所屬機關配合推動情形。

主席裁示：請本部所屬機關(單位)依資料盤點結果，依限將資料上架至政府資料開放平台，餘適用「臺灣主權 AI 訓練語料授權條款」者，請俟數位發展部完成建置語料庫平台後，辦理資料上傳事宜。

(五) 報告「本部及所屬機關資料及申請下架情形」

說明：提會報告 114 年 5 月至 9 月本部所屬各機關(單位)申請下架情形。

主席裁示：有關本部秘書處、矯正署臺南看守所及雲林監獄因資料集內容重複，爰申請下架之 5 筆資料集，同意下架。

七、臨時動議：

主席裁示：有關數位發展部刻正訂定之「促進資料創新利用發展條例(草案)」，本部針對其中未釐清政府資訊公開及政府資料開放兩個不同議題之條文，已向該部提出修正建議，請各位委員後續參加相關會議或活動時，再協助適時對外傳達本部意見，以使該草案更為完善。

散會：上午 10 時 10 分

紀錄：潘威廷

主席：黃世杰

會議議程	委員發言紀要
報告事項(一)	黃召集人世杰：請問各位委員對於報告事項(一)，本部及所屬機關資料盤點表跟民間需求回應的說明，是否有意見提出？ 黃委員心怡：主席好，想請問一下，在最近收到的電子郵件及紙本資料中，沒有看到資料盤點表，若日後方便，在沒有機密性疑慮並節省紙張的情況下，是否可以提供電子檔給我們參考，以利我們可以先行檢視內容？ 資訊處：向主席及委員說明，考量這份資料為非公開資料，而電子郵件為明文傳輸，有資安風險，所以沒有先利用電郵提供，後續將於會場提供紙本資料供委員參考。 黃召集人世杰：後續請資訊處於會議中將紙本資料提供各委員參考，俾利委員審查及提供意見。
報告事項(三)	黃召集人世杰：有關報告事項(三)資料集改善情形，兩個待改善事項皆將於今年底完成，尤其為改善事項 2 壓縮檔的部分，事實上已算完成改善。 謝謝各單位跟資訊處同仁的配合與協助，請大家於年底前確實完成改善作業。
報告事項(四)	黃召集人世杰：關於報告事項(四)，我做一些補充，讓各位委員更為了解。因我們目前還未充分了解數發部 AI 訓練語料庫的實際運作與機制，先前於行政院開會時，主辦機關有說明，其文本資料分類方式主要是依照是否涉及著作權問題進行區分，由於我們業務性質較特殊，因此比較少有著作權的問題，比較關注的部分反而是風險管控的問題。我曾於會議中多次詢問有關風險管控的議題，但目前尚未得到具體回應。 有關 AI 訓練語料庫所蒐集之非公開資料，當提供給有訓練需求的後端使用者，如政府機關或民間單位，於資料取得後會如何去應用，如何避免資料外洩等，目前也尚未得到具體的答案。因此，我們現行在配合推動該計畫時，考量其風險，故先採取相對保守的方式。 此外，有關 AI 語料的授權條款，主要內容亦多著重於著作權的部分，未說明於系統運作後，相關資料外洩的風險。針對此部分，我們將會再持續向相關單位請教，在此之前，將維持現行的運作模

會議議程	委員發言紀要
	式。 張委員心玲：謝謝主席及委員。我剛剛想詢問的內容，其實也是主席所提到的問題，看起來目前主辦機關尚未針對此部分有進一步的結論。另外請教，這次部裡提出的資料與原先已開放的資料有什麼不同？ 資訊處：請委員可以先參考簡報第 20 頁的運作流程說明，數發部會定期將已經開放在政府資料開放平臺上的開放資料，同步至 AI 訓練語料庫內，因此，既有的開放資料將可納入新建的 AI 訓練語料庫內，一併提供給 AI 做為訓練。 至於目前提供的資料與這次盤點的文本資料是否有差異，兩者差異並不大，皆是依循數發部提出的十大文本類型為原則進行盤點。例如，我們開放較多的文本資料類型，無論是已在政府資料開放平臺上，或是這次配合盤點的資料，大多為研究報告，如出國報告、委託研究報告等，或者為施政文件，如每年度的施政績效報告，以及活動宣傳文本，如新聞稿等文本資料。 如同前面報告，本部在中央部會開放資料集數量已為第一名，所以其實不論是在數發部提出 AI 訓練語料庫計畫前後，本部各機關(單位)均已持續努力推動資料開放。 黃召集人世杰：另外詢問，依對報告事項(四)第 21 頁表格內容之理解，除報告事項(一)、(二)所列的資料外，我們又多盤點這些資料，並依適用的授權條款將其區分為兩類。又依報告事項(一)資料盤點表，甲類可開放資料應已全數上架於政府資料開放平臺，而甲類計有 27 項，報告事項(四)所列預計新增至政府資料開放平台之項目為 29 項，故想再確認是否甲類資料已全數上架？亦或是兩者統計方式有所不同？ 資訊處：向主席報告，前述資料盤點表中屬於甲類可開放資料的 27 項，係指本部共有 27 個系統或資料庫，其所含的資料是可以對外開放的，因此，本部於政府資料開放平臺上架之資料多源自於這 27 個系統。而本次新增盤點的 29 項資料，是在前述 27 個系統中，再進一步檢視後發現過去尚未開放的資料，故我們併同本次盤點作業一併上架至政府資料開放平臺。

會議議程	委員發言紀要
臨時動議	黃召集人世杰：了解，簡報第 21 頁統計表格下方第二欄所列的資料目前尚未上架，請本部相關機關(單位)於數發部的平臺建置完成後，再行辦理上架作業。 黃召集人世杰：有關前面所提數發部 AI 訓練語料庫風險評估的問題，若各位委員日後有機會，再請協助向主辦機關了解相關情形，讓我們後續作業更能順利進行。 此外，跟各位委員分享及討論另一項議題，數發部目前正在擬訂「促進資料創新利用發展條例」(草案)，目前我們與該部對部分條文內容尚在溝通討論，係有關該草案將政府資訊公開與政府資料開放兩個不同議題混淆之狀況。 政府資訊公開，其原則在於政府運作的透明化，會將現有資料原始提供，並不會特別進行處理，然而基於各種考量，有許多資訊屬於不公開的；而政府資料開放，其目的是為了應用而存在，資料必須經過處理後再釋出，這樣資料才能夠發揮它的功能。因此，政府資訊公開與政府資料開放是不同的概念，這兩者公開的方式與後端應用的機制涉及許多細節，應由相關法律加以規範，不能僅以是否依政府資訊公開法主動公開或已依申請提供者等方式，來認定是否應將資料作成開放資料(Open Data)。 該草案的目的是希望各機關盡可能釋出資料，立意很好，但不應是以政府資訊是否公開作為是否應開放的標準。我們希望相關主管機關能制定標準，明確規範哪些資料可以轉化為開放資料，像我們有些現行不得公開的資料，經適當處理後，其實可以做成開放資料使用。 因此，利用這個機會向委員們傳達我們的想法，目前我們和草案主管機關的意見尚未達成共識，希望各位委員給我們一些支持，假設有機會在相關會議或活動中，可向外界說明。其實各機關在推動 Open Data 的工作都相當努力投入，過去幾年的經驗，不論在行政院或各機關的會議上，我想各位都可以看到這樣的努力。但是，我們需要的是明確的資料開放標準讓各機關可以依循辦理，而不是在另一個法律內去找資料開放標準。如各位委員有需要，也可以向法律事務司取得相關資料或意見，法律事務司為「政府資訊公開法」的主管單位，可協助提供相關資訊，以便各位委員更清楚了解。

會議議程	委員發言紀要
	主席結論： 黃召集人世杰：今天感謝各位委員的指教，讓本部政府資料開放業務能更加精進，謝謝。