

檔 號：

保存年限：

科技部 函

地址：台北市和平東路2段106號

聯絡人：黃士育

電話：02-2737-7374

傳真：02-2737-7673

電子信箱：syuhuang@most.gov.tw

受文者：國立嘉義大學

發文日期：中華民國105年2月24日

發文字號：科部工字第1050012975號

速別：普通件

密等及解密條件或保密期限：

附件：「政府巨量資料技術工具研發計畫」徵求公告 1件，附件一：計畫徵求主題說明 1件 (105D2004642.PDF, 105D2004643.PDF) GSSATTCH1 GSSATTCH2

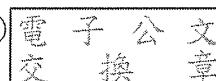
主旨：本部工程司推動之「政府巨量資料技術工具研發計畫」，自即日起接受申請，請查照轉知。

說明：

- 一、請申請人依本部補助專題研究計畫作業要點，研提計畫申請書(採線上申請)，申請人之任職機構須於105年4月20日(星期三)下午5時前備函「送達」本部(請彙整造冊後專案函送)，逾期恕不受理。
- 二、計畫之研究主題必須符合下列要項：1. 巨量資料擷取與管理 (Big Data Acquisition and Management) 技術工具。2. 巨量資料分析 (Big Data Analytics) 技術工具。3. 巨量資料視覺化呈現 (Big Data Visualization) 技術工具。
- 三、計畫主持人於全程計畫執行期限截止後三個月內至本部網站線上繳交研究成果報告，由本司邀請學者專家進行書面審查及召開成果評鑑會議。
- 四、本專案相關徵求計畫書說明及詳細內容業已公佈於網站(本部工程司網站<https://www.most.gov.tw/eng/ch>)-最新消息。

正本：專題研究計畫受補助單位(共295單位)

副本：本部綜合規劃司、工程司、前瞻應用司(均含附件)



附件一：「政府巨量資料技術工具研發計畫」徵求主題細部說明

I. 巨量資料擷取與管理技術工具

巨量資料之前處理對於後端之資料分析應用極為重要，前處理主要包含資料之擷取 (Extraction)、轉換 (Transformation) 與載入 (Loading) 等三大步驟，從擷取面來看巨量資料中強調的資料流處理、大量非結構化文字處理以及各種格式間之快速轉換技術工具為目前仍較缺乏部分。同時，一個能整合各種前處理模組以發揮串聯各模組功能之友善界面平台於政府巨量資料之應用上亦極為需要。以下為相關徵求研究主題：

1. 資料流前處理技術 (Streaming Data Preprocessing)

目前資料分析系統對於即時資料流的處理仍侷限在靜態資料的處理方式，使用者必須自行在時間軸上進行資料流處理，增加了分析程式撰寫上之複雜度。如何發展能針對即時資料流的資料介接與前處理界面，可以讓後端分析程式不需考慮資料源特性與時間特性，在前處理時即能對即時資料流併行分析為重要之議題。

2. 非結構化文字資料擷取與前處理技術 (Unstructured Text Data Preprocessing)

現有資料處理系統中對於非結構化文字資料，多僅提供字串處理功能，在文字上需要開發泛用之擷取與前處理功能，例如自動網頁資料擷取以及斷字與詞頻統計等，或是加入語法結構與詞性分析等，並結合現有相關文字處理工具，提供使用者透過此文字處理模組可直接將文字資料源直接處理成後續分析所需要的結構資料。

3. 快速與自動資料格式轉換技術 (Fast and Automatic Transformation of Data Formats)

同時面對不同資料格式並一起進行分析時的快速與自動格式轉換相當重要，如資料來源是網頁內容中特定且固定的資訊欄位，自動截取資訊並轉換成 JSON 格式進行資料分類、轉換與儲存，在面對未處理的網頁形式開放資料為一重要議題。

4. 快速及具可適性之前處理模組 (Fast and Adaptive Data Preprocessing Module)

現存資料前處理演算法在處理巨量資料時常有記憶體消耗與運行時間過長的問題，多數演算法在巨量資料平台上需另予最佳化改寫，如何達到快速及具備資源可適性為主要挑戰。這部分雖多屬學術研究領域，目前亦逐漸出現具體之系統開發，為一重要課題。

5. 巨量資料特徵標註技術 (Big Data Feature Labeling)

資料前處理中特徵的選取相當重要，除了可利用資訊分析的方式選取有效資料特徵外，能夠有效地整理並提供資料特徵標註的模組對於資料分析上都相當有幫助，提供使用者在資料分析前能利用這些工具選擇並標註有用或需要之特徵資料，為一重要議題。

6. 整合性架構資料前處理工具 (Integrated Big Data Preprocessing Framework)

上述前處理模組相當多且複雜，實務上在進行巨量資料處理與分析時，往往不僅只需要單一前處理方法，目前在整合性的 framework 上仍鮮有工具，需要一個能夠提供友善界面之平台以整合多項前處理工具與加速處理流程。

II. 巨量資料分析技術工具

針對巨量資料之大量性 (Volume)、多樣性 (Variety)、快速成長 (Velocity) 等繁複特性，已有為數不少的資料分析技術與軟體以支援巨量資料的特性。然針對政府開放資料或所建構之相關雲端服務應用，例如自然環境之感測資料、交通雲之感測資料、醫療雲之健康照護資料與社群網路資料等，這些資料反映了更多樣性的特質及分析複雜度。針對政府巨量資料之多樣性、複雜性及相關應用情境需求等，進行統計分析技術研發，開發巨量資料共用軟體。相關徵求主題如下：

1. 時空資料分析技術 (Spatial-Temporal data analytics)

在政府目前的開放資料中，如環境感測資料、氣象資料、即時路況資料等，這些資料均同時具有時間與空間維度的重要資訊。因此，未來在政府巨量資料分析的工具中需要具備更多可有效分析巨量空間與時間資料之資料分析工具。

2. 圖形資料分析技術 (Graph data analytics)

針對政府巨量資料的多樣性，如健保資料、交通雲之即時路況與路網資料、社群網路資料等，提供可支援高多樣性及複雜性之巨量圖形分析演算法為重要之研究議題。

3. 預測最佳化技術 (Optimization of prediction tasks)

巨量資料環境中預測模型的精準度要達到最佳的預測準度，往往需要透過不斷的嘗試不同的建模演算法與參數調校，此過程極為耗時。針對給定之預測標的物，如何選取適當之建模演算法與最佳化參數達到最佳之預測精準度為一重要議題。

4. 前端裝置資料分析技術 (Front-End data analytics)

在某些資料收集的應用情境中，如交通雲之即時路況視訊監控等，這些多媒體或感測資料等需要更多前端的基本資料分析，以降低後台分析平台及資料傳輸的成本等。如何開發適用於前端裝置的資料分析工具亦為一重要研究議題。

5. 隱私保護的資料分析技術 (Privacy preserving data analytics)

政府資料中常具有敏感性資料，如健保就醫資料、戶籍資料、個人金融資料等，如何於資料分析過程中保有適度之資料隱私，使探勘出來的模式或是參數僅能允許資料提供者掌握與擁有為一重要研究議題。

6. 以 GPU 為基礎之資料分析技術 (GPU-based data analytics)

隨著資訊硬體之快速發展，以 GPU 為主之伺服器已漸趨普遍，例如國網中心已建置台灣最大之 GPU 叢集運算環境 Formosa 5，GPU 目前已經廣泛用於各種巨量資料之分析處理。因此，以 GPU 為基礎之資料分析技術為值得探索之議題。

7. 非結構化資料分析 (Unstructured data analytics)

政府巨量資料中存在許多非結構化資料，如文字資料、影像視訊、即時交通視訊等。針對非結構化資料之分析，目前仍缺少有效的分析平台與工具。針對政府巨量資料中的非結構化資料提出有效之分析技術工具為一重要課題。

III. 巨量資料視覺化呈現技術工具

巨量資料工具應用中，資料呈現技術為前端介面最重要的挑戰之一，政府巨量資料視覺化議題，不像以往只限於專家與專家之間的圖形溝通介面，如今更是政府，專家學者與民眾多方溝通的橋樑。而過往資料呈現多依賴單一作業系統上之統計模擬商業軟體所搭配的報表呈現功能來達成複雜的資料視覺化，因作業系統介接等問題，這類的資料呈現多不利於大眾化分享。如何考量到資訊傳達的互動性、有效性、即時性、直覺性、便利性，是這部分議題徵求的主要方向。因此，本計畫即徵求基於開源形式能支援開放性跨平台架構，符合 HTML5 架構之資料視覺化技術工具（基於 R、Python、Javascript 等程式語言之函式庫）或整合性資料呈現服務，以便利各樣式的巨量資料呈現。相關徵求主題如下：

1. 適應各式前端裝置的資料呈現技術

隨著手持式及穿戴式裝置的普及，透過各式設備不同類型螢幕來存取資訊已成為十分普遍的做法。因此能適應各式前端裝置的螢幕，解決螢幕碎片化所帶來的挑戰，為一重要議題。

2. 大量資料下的互動式視覺化技術

對於大量資料之呈現，如合有效解決前後端大量資料傳輸延遲情況，以及克服大量資料圖表互動操作時 lag 等問題，為新型視覺化工具之重要議題。

3. 異質性資料的互動式視覺化技術

過往的 Web-based 資料呈現套件如 D3.js，多著重於單一資料的互動式呈現。然現今各式資料均能快速的累積，能簡易 mashup 異質資料集，呈現不同角度的 dashboard 式視覺化介面，為極具挑戰之重要議題。

4. 即時性資料的互動式視覺化技術

現今的 Web-based 視覺化工具主要透過 Javascript 將資料完整傳輸到使用者端後開始繪製圖表，對於處理資料流(streaming data)的呈現，將十分不便利。如何有效地只針對新的資料做運算，並將之與舊的資料及其視覺化圖表結合，也是新型態工具發展的方向之一。

5. 智慧性資料視覺化技術

巨量資料中樣式相當多元，常用格式如文字 json，伺服器存取 log，關聯式表格，影像等，如何減少非資訊背景使用者在處理整合資料呈現上會面臨的技術門檻，能智慧化判斷常見的資料應用情境，主動的產生相對應的基礎圖表以及相關程式碼，為一重要議題。

6. 整合性資料視覺式服務

現今已有許多提供資料分析與視覺化網站建置的開源整合服務平台，如 CartoDB 以及 Rstudio 的 Shiny package 等可自建視覺化服務網站。然此類服務提供的功能多需再經過大幅修改調校，方能滿足特定領域的使用者需求。如何提供更方便的視覺呈現整合服務以加速政府各領域資料之有效呈現，是一重要之研究議題。

科技部

「政府巨量資料技術工具研發計畫」徵求說明書

壹、計畫背景

隨著資通訊技術及感測科技之快速進步，網際網路、行動網路、物聯網以及社群網路等應用蓬勃發展，各種來源的資料以驚人的速度快速累積成為巨量資料，而巨量資料運算也已成爲世界各國競相投入發展之重要領域。政府各部會單位中累積有各種對國家發展極有價值之各領域巨量資料，如何開發及結合巨量資料技術於政府各部會資料之整合運用，以提高施政效率及改善民生狀況已成為世界各國政府最重視課題之一。針對巨量資料於「大量、複雜、快速累積」三大特性的處理上，目前雖存在多種開源及商業性之技術工具可資運用，但其在資料擷取管理、資料分析、視覺呈現等面向上，於政府巨量資料之應用上皆有不盡完善之處。本專案目標為徵求開發巨量資料分析技術，以建立巨量資料相關技術平台，俾利政府巨量資料分析效益及提升施政效率，為政府巨量資料之應用發展與推動奠定良好基石。

貳、研究議題範疇

一、徵求主題：依下列三方向對於政府巨量資料（含政府各部會已開放資料或未開放之內部資料）之技術工具開發，細部說明詳見附件一。

1. 巨量資料擷取與管理 (Big Data Acquisition and Management) 技術工具
2. 巨量資料分析 (Big Data Analytics) 技術工具
3. 巨量資料視覺化呈現 (Big Data Visualization) 技術工具

二、上述徵求開發之巨量資料技術工具需能運用於政府巨量資料之相關應用，相關應用範例列舉部分如下(但不限)：

1. 健康照護—運用健保就醫等資料建立分析及預測模型，提供疾病成因與趨勢分析及預測、疫情警示等應用，促進全民健康福祉。
2. 自然環境保護—運用台灣各區之河川、土壤、雨量、空氣品質以及物種清單、物種時空分佈等資料建立各種分析預測模型，掌握生活環境變化及保護。
3. 災害預警—運用政府之各種環境監測資料、氣象資料、災害情資彙整資訊網絡、住民部落位置等資料，整合社群網路等資訊建立分析及預測模型，強化現有之災害防治預警監測系統，保障人民生命財產安危。
4. 財政收支控管—運用政府債務資料、稅收資料、政府預算及決算等資料建立國內財政發展情形之分析及預測模型，以利財政規劃及監督控管。
5. 毒藥品管制—運用藥物濫用與檢驗系統之資料、刑案紀錄，以及醫療機構之就醫紀錄、門診紀錄等資料對毒藥品濫用狀況建立分析及趨勢預測模型，以利政策擬定。

參、申請注意事項

一、申請資格：符合本部補助專題研究計畫作業要點之申請機構及計畫主持人與共同主持人資格者。

二、計畫類型及申請、執行時間：

1. 計畫研究型別：請擇一申請

a. 單一整合型計畫：由總計畫主持人將所有子計畫彙整成一本計畫書提出申請，每案申請總經費以 500 萬元為原則。

b. 個別型計畫：每案申請經費以 150 萬元為原則。

2. 申請方式與申請期限：自即日起接受申請，請至本部線上申辦系統製作申請書，申請機構應備文檢具申請名冊於 105 年 4 月 20 日(星期三)下午五時前函送本部（以郵戳為憑，電子公文亦可），逾期恕不受理。

3. 計畫執行期間：自 105 年 7 月 1 日至 106 年 6 月 30 日止，本部得視計畫作業時程作必要之調整。

三、計畫撰寫注意：

1. 依計畫徵求主題細部說明(詳見附件一)提出計畫書，可針對單一主題或整合多項主題進行開發，計畫內容應詳述下列各項，作為審查重點：

a. 計畫團隊於巨量資料技術及政府巨量資料應用領域之相關經驗與能力。

b. 擬開發技術工具之技術定位(技術方法之類型、前瞻性與原創性)與國內外相關現況，並請具體說明其於政府巨量資料之相關應用，及其可彌補現存開源或商用軟體技術工具不足之處。

c. 擬開發技術工具之研究開發方法，及其運用於政府巨量資料之驗證規劃。

d. 軟體程式源碼開放及提供佈署於國網中心資料分析平台之規劃（請參見四、執行規範第 2 條相關說明）。

e. 技術發展路程圖、評量指標及查核點，以作為期中與期末查核之依據。

2. 計畫書內容以 25 頁為限，請以中文撰寫。

3. 計畫類別為「一般型研究計畫」；計畫歸屬為「工程司」；學門代碼為「E9827-政府巨量資料技術工具研發計畫」。

四、執行規範

1. 計畫考評：本部對執行計畫進行期中與期末等成果追蹤及查核，必要時將進行實地訪視，各執行團隊須能實體展示該計畫所開發之技術與系統成果。

2. 本專案所開發之技術工具需開放其軟體程式源碼，並於計畫執行完成時需能將開發之技術工具提供佈署於國網中心之資料分析平台(請參見 <https://www.nchc.org.tw/tw/>)，供政府相關單位將來依其需要予以運用。計畫執

行過程所需之開發與實驗環境，執行單位可自行規劃準備或洽國網中心依其使用辦法申請使用(請注意相關收費規定)。

3. 執行團隊須配合本部進行計畫執行成果之散佈、推廣應用及交流等工作推動。

五、成果歸屬及運用:

1. 本計畫執行成果之智慧財產歸屬執行單位，但本計畫執行成果之產出（包含但不侷限於原始程式碼、編譯後二進位碼、影音圖像多媒體、相關文件等）需提供予政府各部會單位使用。
2. 若有使用任何開源及商業軟體，需說明其授權模式及將來政府單位使用時之相關規範及影響。

肆、審查

本部將邀集巨量資料領域產官學研專家就計畫核心構想之前瞻性與原創性、研究方法之可行性與應用價值、計畫執行規劃(技術發展路程圖、評量指標及查核點等)、軟體程式源碼開放及提供佈署於國網中心資料分析平台之規劃、計畫主持人與研究團隊之相關經驗與計畫執行能力、申請經費之合理性等要項進行審查。

伍、其他注意事項

- 一、本計畫每人限提一件，並列入計畫件數計算，共同主持人不列入執行件數。
- 二、本計畫屬專案計畫，恕無申覆機制。
- 三、計畫成果發表除須註明本部補助外，亦請註明本計畫名稱或計畫編號。
- 四、本計畫之簽約、撥款、延期與變更、經費報銷及報告繳交等應依本部補助專題研究計畫作業要點、專題研究計畫經費處理原則、專題研究計畫補助合約書與執行同意書及其他有關規定辦理。
- 五、其餘未盡事宜，請依本部頒定之補助專題研究計畫作業要點及其他相關規定辦理。

陸、計畫聯絡人

專案召集人：國立交通大學資訊工程學系（所）曾新穆教授

Email: vtseng@cs.nctu.edu.tw

電話：(03)571-2121#54722

科技部工程司 黃士育助理研究員

Email: syuhuang@most.gov.tw

電話：(02) 2737-7374

科技部自然司 陳美慧副研究員

Email: mhchen@most.gov.tw

電話：(02) 2737-7811