

運用大數據協助家暴風險預測—— 以臺北市家庭暴力案件資料為例

陳淑娟 · 游蕙瑜

壹、緣起

「面對家暴只能被動挨打嗎？公部門挑戰大數據，這樣防家暴」（註1），106年3月臺北市政府推出「家暴風險地圖」，透過圖資數據進行多重資料的比對與整合，讓數據來說話，並試圖建立家暴風險預警系統，做為輔助社工進行安全處遇決策與資源分配之依據，而非只等家暴案件發生、通報後才能作為，防治工作也試圖改變過去雨露均霑的表面宣導，於家庭暴力防治工作推動二十年來，社會已有改變，家庭暴力問題也隨著社會及家庭的改變而產生更大的挑戰，如何面對未來二十年家庭及社會安全的挑戰？累積二十年的家暴案件資料能告訴我們甚麼？並轉為具有風險預警的積極資料，「智慧治理」、「主動預防」，成為未來政府處理家暴事件的關鍵策略，但如何做？

目前全國各縣市皆使用衛生福利部保護資訊系統紀錄個案通報、評估及處遇與介接其他資料，該系統得以產出部分簡易

統計，卻無法進行交叉分析，保護資訊系統的功能主要係紀錄個案服務歷程，卻無法苛求該資訊系統能具備輔助決策功能，但如何透過資料分析輔助決策、回饋現有及未來工作，甚至能做為資源分配的依據？這些可能性，來自於筆者在網路上看到一個標題「Data For Social Good」（註2）。

貳、目的

本案以家庭暴力案件預警與風險管理為主要訴求，採熱點預警的概念，針對家庭暴力案件建置以「地點/場域」為主體的通報熱點為警示指標，透過有效評估工具篩選高危機案件，並對於具有風險之案件，藉由個案資訊系統各項表單萃取重要項目來建置警示指標，支援社工處遇決策預警及督導管理社工處遇服務品質，以維護被害人安全及降低其再度受暴之可能與提高社工服務效能與品質。

另一個較巨視的目的，是改變政府機關慣用的委託服務而以平等、創造及促

進社會人參與公私協力防治家暴，因此，臺北市家防中心結合智庫驅動公司辦理 D4SG(Data for Social Good 的縮寫) 公益計畫，透過資料分析，建立跨域對話機制，進而解決社會問題，也透過 Data Mixer, Workshops, Data Challenges, Fellowship 等方式，集體創造資料人與社工人的跨界合作機會，培力社工人對資料的敏感度，同時提升不同資料領域的社會人因為參與而更了解家暴防治工作的過程。

參、歷程

一、回到對資料詮釋的好奇心與決策痛點 (註 3)

長久以來我們可以做到統計，計算共處理多少通報案件，並進行基本資料描述統計例如性別等，但若要進一步計算到年齡或交叉變項統計，就必須要再次的人工計算，這對所有以處理案件為優先的直接服務單位來說，是不會想要去處理的統計問題。然而，社工服務所得資料如未妥適運用分析，難以回饋到家暴與性侵害防治業務，也無法用客觀依據去做人力資源分配運用，「發揮效能」常僅能是計畫目的的形容詞；再者，實務現場收集回來的資料與社工人員登錄的紀錄、評估及處遇計畫等，雖有專業經驗及智慧，但當資料輸入者沒有一致的輸入項目及關鍵字定義，即會造成有統計數字前後不一致、計算標準不同的情形出現，所得到之統計數據做為決策輔助的可信程度亦相對不高，防治策略就無法切中核心。

對於機關來說，機關（構）本身組織

資源不足是向外求助的一大痛點，包括資訊人力吃緊、僅作簡易統計、對資料進行多重條件交叉分析耗時費力等。

D4SG 資料英雄計畫 (註 4) 是透過專案招募有一定熱誠的志願參與者（資料英雄），考量其具公益性質，因此，家防中心提案參與，通過檢核之後由智庫驅動媒合專案團隊的資料英雄進行下一步社會工作與智慧科技的奇幻旅程。

二、專案的合作過程—各司其職、跨領域對話 (註 5)

家防中心、智庫驅動與資料英雄如何互動協力，游蕙瑜（2018）以深度訪談加以觀察個案中的脈絡，先釐清各自三方定位，進而了解協力的互動關係如何建立，進而試圖歸納該專案中在互動過程中的幾個關鍵成功要素，以下進行專案過程中三方行動者之功能與職責介紹：

（一）家防中心—資料製造者、領域知識的提供者

家防中心是資料製造者，先以虛擬資料庫讓專案成員了解政府所擁有的資訊格式，當然基於資料隱私權保護，會先將所有資料以去識別化 (註 6) 方式處理，進而將資料彙整給智庫驅動複核後，提供給資料分析成員。

再者，家防中心亦是領域知識的提供者，協助專案成員釐清相關重要名詞與領域概念，簡單來說也是解惑者；另也使資料英雄了解家暴處理流程，畢竟外部人士並不會想像到家暴案件的通報流程涉及多少單位，包括中心本身通報專線、113 專

線、警政、移民、司法、教育等單位，但通報並不代表案件成立，亦需要專業評估、後續追蹤是否繼續提供協助。因此，家防中心主要角色為協助具有資訊處理能力的資料英雄與智庫顧問釐清其家暴專業知識與流程說明。

社工提供資訊系統運用及說明，並非僅是資料全部提供就能解讀，讓不了解家暴防治實務的人及非具有專業及實務經驗者可以了解該資料領域的知識，因此跨領域交流是重要且有益的。讓資料英雄可從現有資料中找到可以分析及具意義的資訊，經由使用者與其他領域的專家們進行多次討論，發掘問題、疑問，並找到解決方案。

(二) 智庫驅動－媒合者、顧問、轉(通)譯者

智庫驅動的角色為提供媒合平臺去湊合專案機關與資料科學家，機關經提案審核後確認合作與否，並在提案單位確認前會進行訪談了解單位需求與所擁有資料，智庫驅動進而挑選自願報名的資料科學家(資料英雄)參與此專案。另外，智庫驅動會在該專案中進行進度控管、提供先前經驗，意即「顧問」的角色。

跨領域溝通是困難的，因此智庫也扮演「中介溝通者」、「轉譯者」的角色，必須先釐清提案單位需求與目的，進而引導資料英雄以資料分析的角度去思考如何運用現有資料想出解決問題的方法。另外亦擔任家防中心與資料科學家的窗口進行溝通，由於機關對資料釋出持保留態度，

因此有先與智庫驅動簽訂保密協議，針對家防中心所提供的資料再做一次去識別化的檢核，以確保資料無從辨識特定個人。

(三) 資料英雄－協作者、執行者

由於此為志願參與的計畫，除智庫檢核提案單位之外，亦同時徵集資料英雄，資料英雄則為專案的「協作者」、「執行者」，透過資料試圖找解決方式處理問題。資料英雄不拘於其科系與背景、在職或在學，但專案團隊成員工作背景有學生、金融業員工、程式工程師等，各自依據其專長提供知識與技術，像負責地圖的成員是資訊工程師，由他進行網頁地圖建置、另一位因為本身工作負責使用者介面與使用者經驗設計；還有一位則是處理資料分析、檢閱大量國外文獻釐清專案問題及關鍵特點，並協助建立模型解釋分析；該專案分為兩組－地圖組與建模組，前者建立家暴案件風險地圖的網頁資訊界面呈現，後者主要為建構統計模型找出關鍵特徵進行統計預測。

三、資料分析方法

(一) 繪製地圖：以「區域」觀點探索家暴案件

1. 運用 Google 地圖 API：依據案件里鄰轉成資訊，無法判斷者佐以人工辨識修正。

2. 資料清理 ETL：去除重複案件、不屬臺北市案件及無住址案件，將資料清理成可用的資料(見圖 1)。

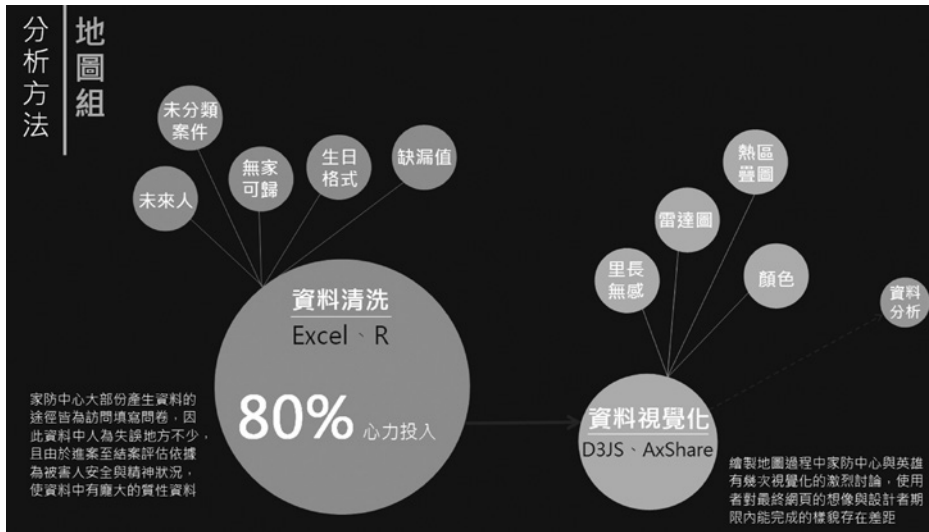


圖 1 專案中地圖組之資料分析方法示意圖

資料來源：D4SG 家暴案件風險與預警管理系統專案簡報

3. **設計地圖**：透過每週腦力激盪，與社工人員進行討論，就使用者需求繪製友善的地圖畫面。

4. **完成地圖**：依據 wireframe 形成網頁框架、參考網站顏色，最後將清理完成的資料檔連結資料呈現於介面 (參見圖 2)。



圖 2 家暴熱點地圖網站介面示意圖

資料來源：D4SG 家暴案件風險與預警管理系統專案簡報

(二) 建立再發預測模組：以「個案」觀點探索多次受暴者特徵，並建立預測模型

1. 資料清理：透過基本資料與親密關係暴力危險量表(TIPVDA)間的比對(如圖3所示)。

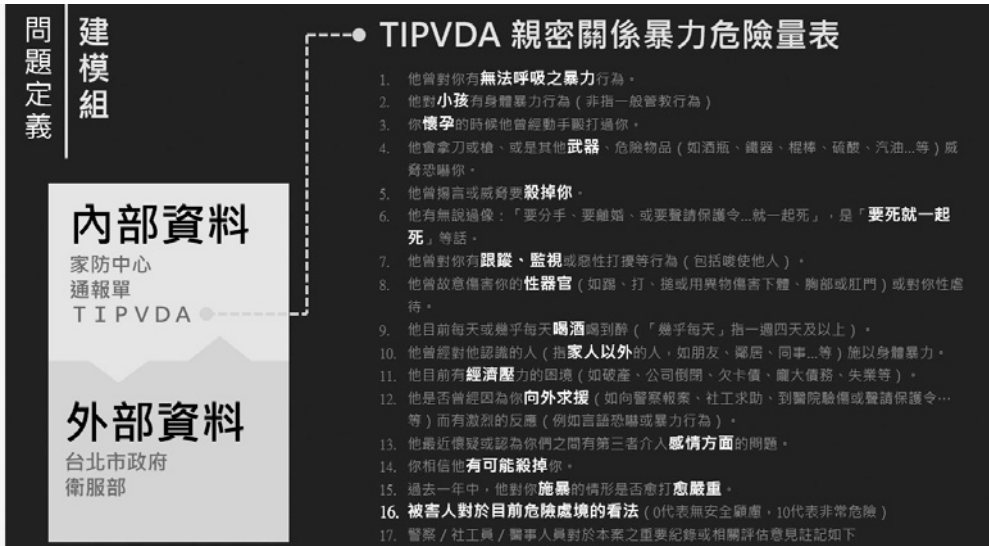


圖 3 專案中建模組之問題定義過程示意圖

資料來源：D4SG 家暴案件風險與預警管理系統專案簡報

2. 資料探索：依不同類型、年齡及再犯原因檢視危險因子(見圖4)。

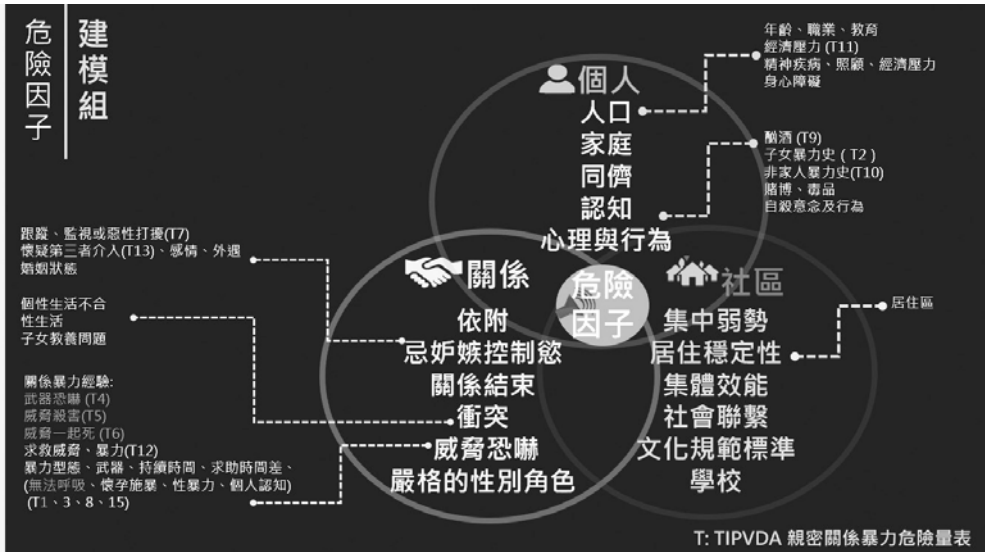


圖 4 專案中建模組之資料探索過程示意圖

資料來源：D4SG 家暴案件風險與預警管理系統專案簡報

3. 模型建構：經測試廣義線性模型 (Generalized Linear Model)、支持向量機器 (Support Vector Machine) 與隨機森林模型 (Random Forest)，最終採用隨機森林模型 (Random Forest)，將 46 個危險因子生成 500 棵決策樹 (參見圖 5)。

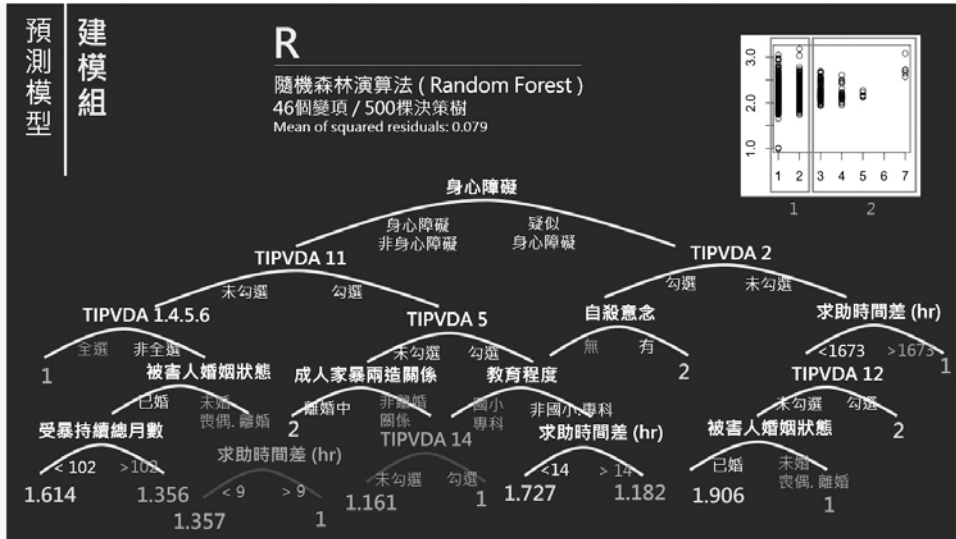


圖 5 專案中建模組之模式建構示意圖

資料來源：D4SG 家暴案件風險與預警管理系統專案簡報

肆、資料運用

一、資料來源

臺北市每年 1 萬餘件之家庭暴力案件通報，透過「家庭暴力案件預警與風險管理」系統的建置，本市家庭暴力案件資料透過去識別化，得以轉化成能統計分析且具預測風險之效能的資料，使得資料的運用與開放是持續且能作不同時間區間之比較 (例如：每年家暴案件分佈趨勢比較)，進而找出有效區域防治策略，讓社區防暴心裡有數，提升社區人民參與之意願與說服力。

二、資料探勘發現

本次運用科學方法整理 104 年 1 萬 3 千餘件家庭暴力通報案件，建立臺北市「家暴案件地圖」，透過區域及數據資料，期以有效分配資源；另以受暴案件特徵及紀錄資料建立「親密關係暴力再發生風險預測模型」，發展科學工具提前辨識再發生風險因素。過程中花最多時間是在作資料的清洗與歸納，進而探勘資料過程，亦發現由於每個社工用字遣詞的習慣不同、觀念不同，產生了同一情況可能有不同的描述之現象，也因著上述發現，臺北市家防中心社工人員透過參與本項計畫，實際且深刻經驗到建立有用的資料影響力，因

此主動重新檢視及建立社工對工作過程中問題或工作方法描述的關鍵字操作型定義，並自行檢核對於個案資料中哪些欄位是必填、服務過程中資料的核對與更正，讓資料正確性提高，使分析統計及預測模型更趨近於真實、正確與準確。

三、發展家暴案件地圖及再度進案預測模組

(一) 家暴案件地圖

主要透過每年家庭暴力案件的分佈地圖(斑點圖)，供社區相關人員線上查詢，了解發生案件類型統計、分布趨勢等，進而於 105、106 年提供社區營造、治安防治策略等方案規劃之科學參考；同時亦可作為規劃家庭暴力防治政策宣導重點之依據。

(二) 再度進案預測模組

此部分透過年度內重複通報的案件分析，找出關鍵指標，運用統計分析，進而設計出系統程式，提供本中心第一線接案社工運用系統預測該案再度通報之機率，進而提升社工面對高案量時評估之機制，以提升社工敏感度及減少再次受暴情形發生。

伍、成果與效益

一、政府統計資料開放，減少社區資源盤點及自行調查統計的成本

為使社區內規劃家庭暴力防治宣導相關參據，業已評估規劃透過資料已有去識別化處理，僅以數字作呈現方式，先行開放於相關專業網絡相關單位及研究使用，

提供專業人員、里長及民間團體查詢案件統計資料，輔助社區規劃家暴防治切合社區需求重點。

二、協力過程的亮點與挑戰

(一) 機關開放態度釋放資料與民間協作

資料必須要有所活用，而非在資訊鍵入電腦之後變成死的資料無法回饋工作業務，無論是用以主管或基層提供服務者決策的輔助，都應有所助益而非僅是數字作業、堆積資料；另外，對於資料的釋出應該抱持開放態度提供數據，與資料科學家討論過後確認何種數據有價值得以運用，並運用專業領域知識互相激盪現有問題的解決方案，實踐用資料來解決社會問題、對社會有助益的精神。

(二) 協作是問題 導向－欲解決什麼問題

回歸專案的目的，也就是家防中心、智庫驅動與資料英雄們在討論對話過程中的核心是為了解決何種問題，亦呼應資料治理是問題導向、目標導向，奠基在提供行政組織內部決策的輔助，多鎖定在剖析資料樣態及應該以何種角度切入詮釋，因此以區域作為切入複雜資料的呈現是較為省力且容易閱讀的，無須過多統計背景或技術門檻即可輕鬆了解顏色深淺、區域分布，進而提升社會大眾對於家暴議題的關注，因此該專案亦嘗試以「區域/地圖」觀點切入，先了解家暴案件發生分布，且易助於家防中心管理者與社工進行決策輔助。另外，該專案更為關鍵在於建立預警風險系統以供社會工作者預先了解個案風

險程度，釐清高再發的家暴案件較為顯著的特徵為何。

預警模型則在於用數據輔助相關社工處遇決策，例如提醒個案風險、定期追蹤等，希冀提早避免受害者再次受暴暴露在更危險的施暴環境中，畢竟就家暴案件的處理來說，若得以透過客觀數據的統計提早了解再發風險高低程度，社會上家暴再犯的憾事得以減少，較為有所助益的是對於前線社工、高階主管來說客觀數據做為預警分析可以適時提早做出關鍵處遇決策，不再僅是事後彌補遺憾而是事前預防的積極作為。

(三) 互惠性—協力過程中的知識學習與分享

若要了解資料可以如何運用，先去釐清所擁有的資料有哪些，再者透過討論確認何為關鍵數據，進而發揮資料的真正價值，在此專案中相當值得學習在於討論前段會先以文獻回顧熟悉該領域的專業知識，另外家防中心亦於前期討論讓智庫與資料英雄了解家暴案件的處理流程，包括通報程序、處理程序等，這樣的做法有助於對於家暴議題完全門外漢的資料英雄的理解與熟悉，達成跨專業領域間的知識學習與分享。

三、效益

(一) 內部效益

1. **支援社工決策，增進案件處理品質：**以家庭暴力案件預警與風險管理為主要訴求，採熱點預警的概念，針對家庭暴

力案件建置以「地點/場域」為主體的通報熱點為警示指標，透過有效評估工具篩選高危機案件，並對於具有風險之案件，藉由個案資訊系統各項表單萃取重要項目來建置警示指標，支援社工處遇決策預警及督導管理社工處遇服務品質。

2. **發展科學工具提前辨識再發生風險，降低受暴風險：**運用科學方法整理104年1萬3千餘件家庭暴力通報案件，建立本市「家暴案件地圖」，透過區域及數據資料，有效分配資源；另以受暴案件特徵及紀錄資料建立「親密關係暴力再發生風險預測模型」，發展科學工具提前辨識再發生風險因素。

3. **建立被害人再度受暴評估參考工作決策的機會：**投入透過區域資料建立家暴案件特徵地圖，以個案特徵（親密關係暴力），嘗試建立預測模型，此模型能作為預警系統，新進個案若呈現高再受暴特徵，藉由識別受暴者的特質來預測期再次通報受害的可能性，在進案的第一時間便可採取防治降低再度受暴的機會；社工亦能協助被害人提早覺察，防止多次家暴發生，避免受害者身心狀況惡化以及過多的人力資源投入，以維護被害人安全及降低其再度受暴之可能與提高社工服務效能。

(二) 外部效益

1. **繪製家暴案件地圖，提高運用資源效能，落實社區預防，減少盲目宣導：**建立本市「家暴案件地圖」，並用顏色區隔熱區，透過區域及數據資料整合，將需要投注人力、物力及財力的重點區域標註，經由通報案件數的分布，讓社區了解本身

區域內的家庭暴力案件的現況，亦即將錢花在刀口上，把力量集中於特定的區域，方能有效分配資源。

2. 促進公私協力創新、改變而進步：

(1) 機關開放資料與民間協作，發揮資料詮釋力，用資料解決社會問題，發揮社會公益精神。為專案計畫 (project) 的協力合作，於特定時間完成一系列的工作流程，透過彼此間溝通互動的動態過程，完成特定目標，此方案主要結合民間資源所辦理 D4SG(Data for Social Good) 資料英雄計畫，經由民間資訊、統計等各項專業人士、以志願服務方式與本中心社工人員利用 3 個月時間，定期每周三晚上密集討論 2 小時，參與人員為民間專業人員 7 人、智庫驅動公司顧問 2 人及本中心社工 5 人、資訊 1 人跨域對話，透過跨界資料科學團隊的協作，創造各種新的公共服務及工具。

(2)D4SG(Data for Social Good) 資料英雄計畫的各項專業人士 (包含資料統計、腦神經研究、企業管理等背景人士) 以不同的角度與視野看家暴資料；在社會科學的領域，對於家暴危險因子的定義通常是質性的描述，例如「過往受虐經驗」、「關係衝突」或是「居住穩定性」，較無法以量呈現，然為了建立統計模型，資料英雄們和社工師進行困難的對話，共同嘗試將這些質性指標轉換為可量化的指標，充分展現公部門與民間在不同領域與立論點下，多元參與以導人民間的創意動能，多

方反映問題複雜性，並建立夥伴關係，共同改善社會問題，為預防問題發生而努力。

陸、結語

資料需要被處理分析，才能產生生命力與資料力，本次透過區域資料建立家暴案件地圖，主要用於輔助社區進行家暴防治策略規畫，減少重複盤點社區案件特徵需求的成本，促進更多民間社區組織藉由家暴案件地圖更有效及正確了解各社區的防治需求，投入進行社區扎根的方案，達到促進社區參與、預防宣導。另經由建立預測模型作為個案風險預警系統，輔助社工進行問題診斷及與被害人擬定安全計畫等處遇決策參考。此外，在運用已去識別化資料的同時，仍需要不斷的來回檢視確保資料安全及預警模型的校準，更要時時省思社會工作是對人的工作，運用資料產生的相關工具，主要是輔助及預警功能，其無法替代社工專業技術或其中任何一部份，這也是本案在協作過程中最常被反思討論的。

(本文作者：陳淑娟為臺北市家庭暴力暨性侵害防治中心主任；游蕙瑜現為國立政治大學公共行政研究所研究生，106 年以本案進行學期研究)。

關鍵詞：家庭暴力 (domestic violence)、資料運用 (data use)、風險預警 (risk prediction)、輔助決策 (decision-making support)

註 釋

- 註 1 王彥喬（2017）。「面對家暴 只能被動挨打嗎？」公部門大挑戰 大數據這樣防家暴，風傳媒，取自 <http://www.storm.mg/article/230480>（瀏覽日期：2017 年 10 月 17 日）
- 註 2 智庫驅動。D4SG 資料英雄計畫網站，取自 <https://d4sg.org/>（瀏覽日期：2018 年 12 月 9 日）
- 註 3 游蕙瑜（2018）。被攤在陽光下的家暴問題—以協力探討家暴預警與風險管理的專案，政治大學公行研究所學期報告，頁 10。
- 註 4 智庫驅動。D4SG 資料英雄計畫網站，取自：<https://d4sg.org/fellowship/>（瀏覽日期：107 年 12 月 22 日）
- 註 5 同註 3，頁 12-15。
- 註 6 個人資料保護法施行細則第 17 條規定，無從識別特定當事人，指個人資料以代碼、匿名、隱藏部分資料或其他方式，無從辨識該特定個人者。

參考文獻

- 游蕙瑜（2018）。被攤在陽光下的家暴問題—以協力探討家暴預警與風險管理的專案，政治大學公共行政研究所學期報告，未出版。
- 智庫驅動（2017）。D4SG 家暴案件風險與預警管理系統專案簡報。
- 王彥喬（2018）。「面對家暴 只能被動挨打嗎？」公部門大挑戰 大數據這樣防家暴，風傳媒，取自 <http://www.storm.mg/article/230480>。
- 智庫驅動。D4SG 資料英雄計畫網站，取自 <https://d4sg.org/>。