

# 大數據匯流分析平台之建置 ——就業追蹤資料庫案例分享

梁德馨

關鍵字：大數據分析、就業追蹤資料庫、職業訓練

## 摘 要

大數據（Big Data）價值展現在於大洞察（Insight），在現今豐富的數據量下，已被洞察的訊息仍只是冰山一角。在大數據洞察的過程中，基礎的資料匯流工作雖為大數據分析中之黑手苦工，但卻是大數據分析上最重要的基石。是以，本文將針對建置大數據匯流分析平台之打造手法提出建議，由資料之取得、整理與匯流，到以議題導向為出發，建置橫縱切面分析所需的四類層級資料庫，逐一提出概念方法。文中亦將分享勞動部勞動及職業安全衛生研究所之「就業追蹤資料庫輔助決策支援平台」的成功建置經驗及其架構內容。並以「職業訓練後之就業追蹤分析研究」為例，展現大數據匯流分析資料庫建置後之研究應用。

最後，本文建議國軍在建置大數據匯流分析資料庫時，可與人才招募息息相關之軟性議題切入，針對國軍退除役就業輔導進行研究，剖析參與各類職業訓練之退除役官兵其在就業率、就業穩定度及薪資水平。此大數據分析的成果除可對職業訓練成效做長期的追蹤評估，亦可能透過大數據分析明瞭退除役官兵所從事的產業，提供未來職訓課程規劃所參用。

## 壹、前 言

在資訊技術的快速發展下，資料取得及儲存技術效率快速躍進，造就了大數據（Big Data）年代的興起。「大數據」之「大」不只是單一資料量的大，也代表著資料來源的廣及雜，更追求資料處理效率上的快。而大數據價值是立基在資料的易使用性及真確性，大數據價值展現則在於大洞察（Insight），在豐富的數據量下，被洞察的訊息仍只是冰山一角。

在現今的資訊科技下，數據之取得及儲存已不再是科學研究最大的困難。因儲存成本大幅下降，在過去公務體系因為費用過高而無法長期留存的資料，也可以不再受限於儲存年限的規範能長期大量保留。在面對已取得及儲存大量數據下，能彙整串接多個資料來源，並以領域專業研究所需來重整建置出最適於議題分析之資料架構，則是完成大數據分析成功的最關鍵環節。

在各政府機關內，因資訊相關單位擁有專業資訊技能，所以在在大數據分析軟硬體環境建置及資料庫整合的工作常扮演著主導規劃的角

色。在大數據分析的過程中，最耗時耗工的往往是在資料清理準備，此階段的工作往往占了整體研究期程的50%以上。各組織中之資訊單位通常僅以資料儲存的觀念來管理大數據資料，此乃因資訊單位並非政策議題之領域專業研究人員，在清理建置資料庫時，難以統計或議題分析之角度來轉換產生研究所需之指標。因此，在各項議題研究上，資訊單位通常皆僅以提供原始資料，再由個別研究者自行加工轉換處理。在各項研究的每次清理都耗工費時，在無系統化的分散作業下，更無法長期累積及分享研究經驗。在研究人力及時間資源重複的循環投入下，無形間造成了大量預算的浪費。

在政府各單位皆大力推廣大數據分析之下，以策略議題導向來建置大數據分析資料庫，為各領域議題研究者先行建置好實用的資料平台，應是組織內大數據分析能普及與成功的核心關鍵。是以，本文首先將針對大數據匯流分析平台之建置理念及設計原則做一闡述，再針對依本理念所成功建置之「就業追蹤資料庫輔助決策支援平台」及其應用分析做介紹。最後，在此平台成功建置之經驗下，探討國軍大數據分析之可能分析議題，初步建議可參考此就業追蹤資料輔助決策支援平台之建置經驗，並結合運用其已有之資源，由軟性且無國家資訊安全疑慮考量之關懷照顧議題出發，建議可在「國軍退除役後之轉業輔導追蹤」等議題上進行大數據分析。

## 貳、大數據匯流分析平台建置

大數據分析執行的起手式並非最熱門的人工智慧（Artificial Intelligence）模型應用，也非美觀的視覺商業智慧（Business Intelligence）分

析展現，而是無趣的資料擷取（Extract）、轉置（Transform）、載入儲存（Load）等工作。這是建造大數據分析之資料流高速公路的基礎建設，用以收集不同來源的資料，依據商務、政策或議題分析所需規則轉換資料，再將其載入至目的地資料存放區。這項資料匯流工作實為大數據分析中之黑手苦工，但卻是大數據最重要的基石。

### 一、跨域人才齊步行

資訊技術、統計分析及領域知識等三大類專業人才是在進行大數據洞察時皆不可或缺的。大數據分析進行時的四大步驟依序為取得、儲存、運算及視覺化，在大多數的建置過程中，總會以序列步驟來進行階段性的分工；在取得與儲存的階段由資訊技術人員來進行，而在運算及視覺化的階段才由統計分析及領域知識者接手來參與。而這樣的分工法，往往會使得資料之建置無法由議題研究及數據分析的角度來準備妥當，常需要重新清理、整合及串接資料，才能為研究者所用。為能改善此現象，建置的過程應該是採取各類專業人力在四階段皆共同參與，依階段分工比重不同的方式來完成；所以即使在大數據資料基礎建置的起步階段就需要有三類跨領域人才齊步合作。

另外，整合跨專業領域的專業人才來共同完成大數據分析平台的建置，是件知易行難的事。因此，擔負著方向主導及跨領域整合協調與決策角色的專案計畫主持人或專案經理，扮演著極重要的協調仲裁溝通的角色（Coordinator），是帶領專案成功建置的關鍵人。

### 二、垂直水平十字手法

大數據分析平台建置成功的核心要件為需

要明確的策略及應用方向，但短期的建置成效則是大数据專案是否繼續獲得組織上層決策者支持的關鍵點。傳統資訊系統與平台建置之手法，總是先由使用者提出需求，而資訊部門進行溝通評估後來規劃軟硬體環境並建置之，而此建置的過程往往需時甚久。若一開始只著重於基礎分析軟硬體環境的建置而未能即時有分析成果產生，組織上層決策者無法想像出成效，就很難在預算及資源上給予合理的配置及全力的支持，而造成專案半途而廢或形成半套平台的狀況。

「垂直水平十字手法」是建置大数据匯流分析平台時極佳之規劃及專案推進手法。在進行垂直向往下紮根的資料取得及儲存的基礎建置時，也要能同步平行的進行演算與視覺化的工作。在第一年的建置專案中，同時有部分基礎的匯流資料庫建置及應用資料庫所進行的議題分析成果，則較容易協助組織上層決策者評估專案成效，也較容易爭取到更多的預算及支援。

### 三、議題導向指標之必要

那麼數據是新的黃金嗎？數據是像黃金礦，它需要被精煉才能展現其價值，它的價值需要被洞察後，才能發光展現。產生具策略參考價值之分析與研究成果是大数据匯流分析平台建置的最終目的。資料匯流處理者在面對眾多大数据分析使用者的需求期待下，或因資料可取得之完整性不足，或因使用者需求無法明確定義，或因負責整併之資訊專業者的領域知識不足無法理解……等眾多原因，最後總是難

以整理出合乎大部分研究者所需的大數據匯流分析資料庫。要能提升大数据匯流分析資料庫之實用性，在起步建置之初，就先有目的性的選定幾個應用廣度較夠的議題，並同步採用「指標設計方法論」收斂審視應整併產生之資料庫及欄位（即指標）。

指標設計原則可由兩大面向進行，包含「指標價值」以及「指標策略」。（圖1）所謂「指標價值」即為以學理及社會或國家發展等各類研究上來評估指標的價值性，判斷該指標是否為研究者廣泛應用所需，且具有應被長期追蹤觀察的必要性，此類指標應有明確的定義且屬於長期不會隨意更動之指標；所謂「指標策略」則是因應現行市場或政府策略評估時是否需採用之指標。用以評估現行策略而提出的創新指標，其在「指標價值」上的重要性也許不高，但其「指標策略」就有其很強的重要性。

在進行資料集建置時，應依據大数据分析範疇（例如：就業追蹤研究）以「指標建立漏斗法」分別列出與議題研究有關可能需要的各類指標，然後分別依「指標價值」以及「指標策略」來評估指標的價值重要性或策略需用性，分別收斂出二種評估原則下的指標長清單，接著再將其比對後，彙整出指標短清單，最後再依議題主題加以歸類檢視是否足用，另外亦需針對該指標應分析的維度（例如：地區別、年份、性別……等）加以釐清。透過此一過程，可以較有條理的彙整出有意義且必須建置的大数据分析所需衍生欄位了。

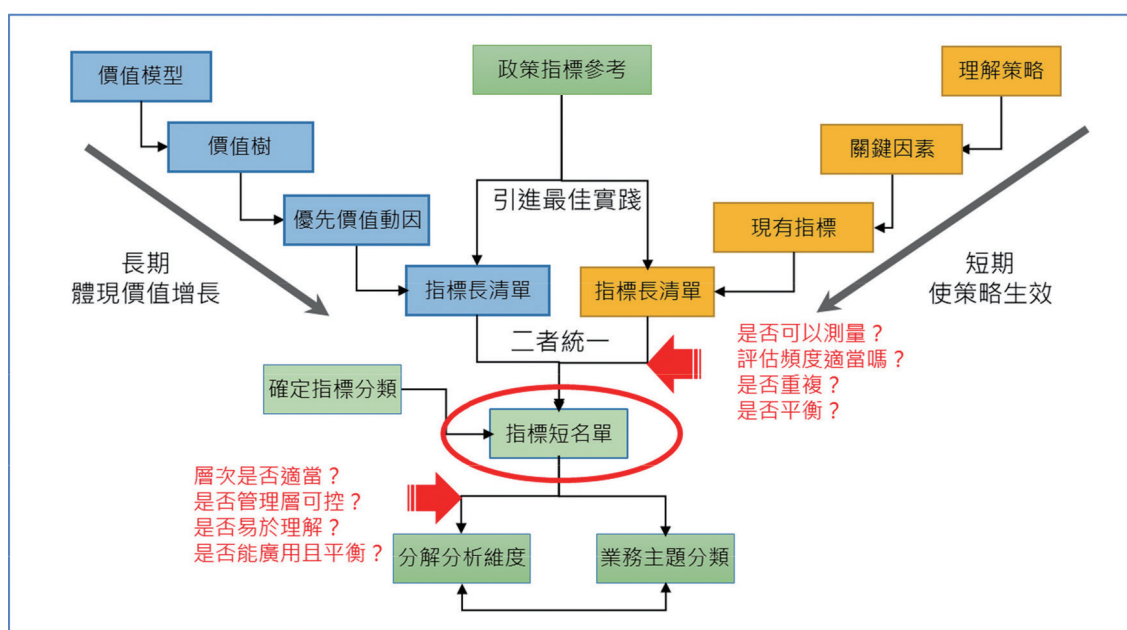


圖1 指標建立漏斗法

#### 四、大數據之資料庫層級

為了便於大數據分析資料庫能更有效率進行管理與維護，資料結構的設計應使用資料分層設計概念。資料庫分層讓資料的脈絡更具條理及系統化，每一層都有其作用與目的，當資料有異常時，能夠快速找到問題點，針對問題做區域階段性修復，毋需整體重新執行所有的處理流程，確保以最低的成本，修正資料。分層架構也能減少重複開發，重複計算的問題。

在此分層設計理念下，資料結構區分為「檔案來源（Data Source）」、「資料暫存區（Staging Area）」、「操作資料庫（Operational Data Store, ODS）」與「資料市集」（Data Mart）。「檔案來源」屬於資料提取層，「資料暫存區」則屬於資料轉換與清潔層，而「資料市集」則屬於資料存放層，至於「操作資料庫」則會因為不同情境，扮演資料轉換與清潔層或資料存放層的角色。

若以數據分析應用的角度來將資料庫分

層，則一般可分為「基礎資料集（Foundational Dataset）」、「加值資料集（Value Added Dataset）」、「追蹤資料集（Tracking Dataset）」及「統計資料集（Statistical Dataset）」等四層。（圖2）「基礎資料集」為資訊匯流的最基礎資料集。可能為單一資料源或由多個資料源整理合併而成。「加值資料集」是針對不同的分析主題新增合併而成的資料集。「追蹤資料集」是將研究主體對象（人、公司…等）依時間歷程串整歸戶具時間序列之資料集。「統計資料集」是依分析主題或維度進行統整之資料集，最常用於BI分析時之資料整合。因應不同分析需求建置不同層次的資料集，將有助於研究者能針對議題分別完成橫切面的剖析以及縱切面的追蹤觀察探究。

上述大數據資料平台的建置原則及理念，本文將以勞動及職業安全衛生研究所之「就業追蹤資料庫建置專案」來做成功實踐案例說明於後。



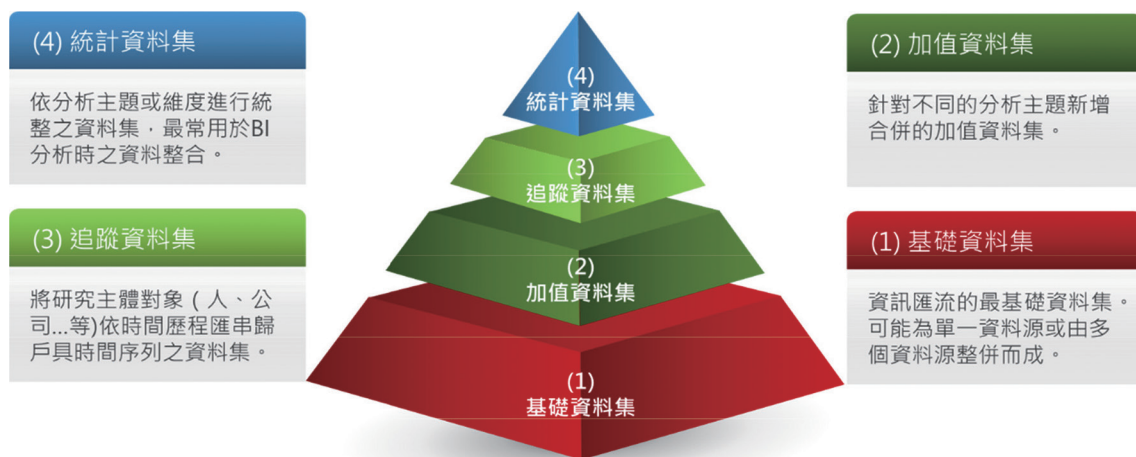


圖2 大數據資料集層級

## 五、大數據匯流分析平台之效益

在大數據資料即是黃金的理念下，各公務機關也開始著手推廣運用公務資料進行有價值的訊息挖掘。而由機關組織全面分析的角度來建立「大數據匯流分析平台」可以使數據決策的實踐更有效的達成。由機關組織全面考量來建立整合研究所需之資料庫，共同由專業的資訊、統計及領域專業人員來建置大數據匯流分析資料庫，其所解決的問題及達成的效益歸納闡述如下。

### (一) 資料品質面向

在公務體系中，資訊單位與研究業務單位在大數據分析上的角色分工上也有所差別。資訊單位通常負責原始資料的倉儲庫，以管理來源資料的「完整性」、「正確性」、「系統穩定性」與「定期更新」為任務。研究業務單位則是在特定的研究議題下，對資料進行再分解與組合，進而創造衍生指標，並做出有策略評估或建議參考的研究報告。（吳慧娜、梁德馨，2021）

機關組織內各個研究業務單位為達成議題分析的任務，皆重複在處理資料，其

過程中也都常會遭遇「欄位缺失值」、「資料不正確」、「多資料源資料不一致」、「交易型資料（Transaction Data）轉置或歸戶」……等問題；當需要進行追蹤資料時，則更會面對需要較專業資訊能力才能處理的「觀察值左右設限問題（Left/Right Censor）」及「資料起/迄時間對齊處理」等問題。不同的研究者其資訊處理能力不盡相同，對於指標定義及計算邏輯也不完全一致，因而可能創造出同一指標在不同的報告中，以相同的名詞但不同的定義或不同的數值來呈現的問題。是以，需由跨領域的專業共同來進行規劃建置，以一致性的資料定義與處理原則進行處理，才能提供出高品質的大數據分析資料。

### (二) 成本節省面向

在大數據分析的過程中，資料清理準備此階段的工作往往占了大半以上研究期程。在不同研究者皆分別進行的狀況下，也造成了人力工時的浪費。因此建置大數據匯流分析資料庫，將重工耗時與繁瑣性問題資料處理工作畢其功於一役，則必可減少大量過去各專案各別處理資料的重複工時。在節省工時成本外，也讓研究單位

能省卻苦工，直接專注於數據分析、模型建立及議題解析，必能提升研究之質與量及速度，更有效率的提供組織所需之各類策略建議。

另外，若各專案皆自行處理進行資料之再處理，相關的資訊軟硬體資源亦會產生重複建置的問題。若能整合建置，可全面規劃並精實資料倉儲空間，避免硬體無效耗用。在依據勞動及職業安全衛生研究專案建置經驗的估計，若所建置之資料庫可供其他研究應用，可減少1個月資料清整時間，估計每項計畫可節省人力成本34萬元。研究人員觀察使用統計軟體資料儲存空間，與導入資料庫應用資料硬體使用空間比較，後者減少65%。（吳慧娜、梁德馨，2021）

### （三）智識管理面向

IBM首席分析長兼首席經濟學家弗萊明（Martin Fleming）坦言，領導與治理，就是企業數據長的第一個、也是最重要的挑戰。（天下雜誌，2016）在大數據分析時代，各類大數據分析的資料採礦（Data Mining）軟體及開源分析軟件（Open Source Software）已廣為數據分析師所採用。各類網路公開的開源教學及論壇討論分享使得大數據分析的執行，似乎成為是只要用就可人人自學而為之。在組織大力推展數據決策下，若無建置整合性的數據匯流資料庫及分析平台，則各業務研究人員所累積的各類智識將四處分散，在研究過程中所產生有關資料匯流、指標建置、分析維度、追蹤分析…等各類智識皆無法完整留存於組織。而在大量資料處理到指標建立的各項經驗等大多存在於個人，而每個人在留存智識的完整度亦不盡

相同，造成智識無法有系統的整合及管理，也無法達成智識之傳承或再被運用。亦即，各研究專案的補助與執行，的確創造了研究者個人的智識成長，但對組織而言，卻未達到應有的智識累積。

「全民健康保險研究資料庫」是我國能將公務資料匯整而提供給產官學各界為用的最佳成功典範。透過分享運用此一資料庫，創造了大量的智識擴散與分享。25年來，共收載671億筆的龐大數據，乃全世界少數擁有大數據基礎的健保資料庫，更是醫藥衛生研究極珍貴且具代表性之實證資料；根據中央健康保險署統計，從民國94年至110年3月底，國內學者利用健保資料庫所進行的研究，發表於國際期刊的論文累計達6,550篇論文。

因此，可借鏡於健保資料庫之分享及管理，於組織內部先建構大數據匯流資料庫，以形成組織內部的數據共享及智識分享，降低研究者在進行分析之資料取得及儲存的難度，以提升組織內部研究的能量及品質，使數據決策成功的運行。接著，可針對資訊及個人資料安全保護的考量下，研議大數據資料庫對外分享的管理機制，以創造更大跨域智識結合與擴散。

## 參、就業追蹤資料庫

### 一、資料庫之建置介紹

優質及充分的勞動力是各國經濟體系成長之主要核心動能。隨著新興科技快速發展，使得國內外產業板塊與結構轉移改變，造成勞動力在產業及職能上出現供需失衡現象。（張玉山、陳仁惠、黃春長、吳慧娜，2019）歷年

來，政府相關單位及勞動力研究相關學者無不針對這些重要議題持續的進行著各項研究，並提出報告與政策建議。

勞動及職業安全衛生研究所為政府勞動政策的研究智庫，每年針對勞動力相關的研究不在少數。為提供官學各屆進行勞動議題之相關研究，勞動及職業安全衛生研究所提供實體隔離分析實驗室來進行大數據研究。但是每項專案研究的資料清理，不但耗工耗時，且分別佔據龐大軟硬體資源，且無法跨期間或跨專案累積資料及經驗成果。另外，勞動力的研究也常需要能進行多年持續性的研究，一直以來許多官方研究報告提出建立就業追蹤績效指標的重要性，也有學者建議建立長期性資料庫系統作為政策導引。因此勞動及職業安全衛生研究在上述的政策研究需求下，於2018年啟動「就業追蹤資料庫建置專案」，結合各類資料來源建置就業追蹤資料庫。透過資料庫能穩定且長期進行各項就業研究，提供更具延續性及長期觀察上之政策評估參考。建立就業追蹤資料庫將可成為政策評估研究的加速器。（吳慧娜、梁

德馨，2021）

此專案至今執行二年（圖3），以19億多筆的勞保、就保單位被保險人檔為基礎資料，彙整了18個其他資料檔，相關檔案如表1所示。為了兼顧長期策略及短期效益展現，此專案以大數據分析之「取得、儲存、運算、視覺化」四大步驟來規劃建置「就業追蹤資料庫輔助決策支援平台」。為因應就業市場及勞動力之研究，讓所建立之平台在分析上更具實用性，除了已有明確定義及被研究者廣泛使用的就業率、失業率……等指標外，在此專案彙整完成橫縱向資料後，亦設計並計算出「轉職頻率」、「就業穩定度」、「平均轉職年資」、「轉職變化度」、「返回職場時間」、「薪資成長率」……等多項指標，以便於各政策議題上研究追蹤觀察所用。為了提升本平台的易用及親和性，亦打造視覺化的自主分析環境，以利非資訊專業技術的政策研究人員能更容易的進行相關議題分析。「就業追蹤資料庫輔助決策支援平台」建置匯流流程如圖4所示。

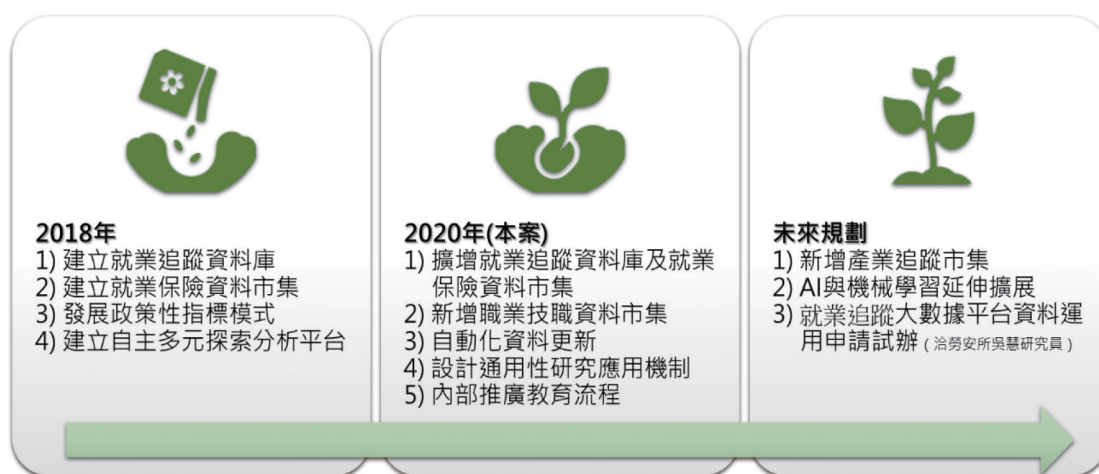


圖3 就業追蹤創新指標應用平台之發展歷程



單位：千筆

**資料平台整合多元資料**

**資料來源**

- 匯入資料前：事業單位、勞務關係、勞動關係、數位化、資料庫、數據庫、SQL Server、Oracle、MySQL、Microsoft Access、Excel、PDF、Word、PowerPoint、PPT、DOCX、XLSX、XML、JSON、CSV、TXT、RTMP、HLS、MPEG-DASH、WebRTC、SIP、RTP、UDP、TCP、HTTP、HTTPS、FTP、SSH、Telnet、SMTP、POP3、IMAP、LDAP、NFS、CIFS、SMB、AFP、Fibre Channel、iSCSI、FC、SAS、ATA、IDE、USB、FireWire、Thunderbolt、PCI Express、PCI、VMEbus、ISA、EISA、AGP、PCI-X、Serial ATA、Parallel ATA、IEEE 1394、IEEE 802.11、IEEE 802.15.4、IEEE 802.16、IEEE 802.22、IEEE 802.3、IEEE 802.4、IEEE 802.5、IEEE 802.6、IEEE 802.7、IEEE 802.8、IEEE 802.9、IEEE 802.10、IEEE 802.11、IEEE 802.12、IEEE 802.13、IEEE 802.14、IEEE 802.15、IEEE 802.16、IEEE 802.17、IEEE 802.18、IEEE 802.19、IEEE 802.20、IEEE 802.21、IEEE 802.22、IEEE 802.23、IEEE 802.24、IEEE 802.25、IEEE 802.26、IEEE 802.27、IEEE 802.28、IEEE 802.29、IEEE 802.30、IEEE 802.31、IEEE 802.32、IEEE 802.33、IEEE 802.34、IEEE 802.35、IEEE 802.36、IEEE 802.37、IEEE 802.38、IEEE 802.39、IEEE 802.40、IEEE 802.41、IEEE 802.42、IEEE 802.43、IEEE 802.44、IEEE 802.45、IEEE 802.46、IEEE 802.47、IEEE 802.48、IEEE 802.49、IEEE 802.50、IEEE 802.51、IEEE 802.52、IEEE 802.53、IEEE 802.54、IEEE 802.55、IEEE 802.56、IEEE 802.57、IEEE 802.58、IEEE 802.59、IEEE 802.60、IEEE 802.61、IEEE 802.62、IEEE 802.63、IEEE 802.64、IEEE 802.65、IEEE 802.66、IEEE 802.67、IEEE 802.68、IEEE 802.69、IEEE 802.70、IEEE 802.71、IEEE 802.72、IEEE 802.73、IEEE 802.74、IEEE 802.75、IEEE 802.76、IEEE 802.77、IEEE 802.78、IEEE 802.79、IEEE 802.80、IEEE 802.81、IEEE 802.82、IEEE 802.83、IEEE 802.84、IEEE 802.85、IEEE 802.86、IEEE 802.87、IEEE 802.88、IEEE 802.89、IEEE 802.90、IEEE 802.91、IEEE 802.92、IEEE 802.93、IEEE 802.94、IEEE 802.95、IEEE 802.96、IEEE 802.97、IEEE 802.98、IEEE 802.99、IEEE 802.100、IEEE 802.101、IEEE 802.102、IEEE 802.103、IEEE 802.104、IEEE 802.105、IEEE 802.106、IEEE 802.107、IEEE 802.108、IEEE 802.109、IEEE 802.110、IEEE 802.111、IEEE 802.112、IEEE 802.113、IEEE 802.114、IEEE 802.115、IEEE 802.116、IEEE 802.117、IEEE 802.118、IEEE 802.119、IEEE 802.120、IEEE 802.121、IEEE 802.122、IEEE 802.123、IEEE 802.124、IEEE 802.125、IEEE 802.126、IEEE 802.127、IEEE 802.128、IEEE 802.129、IEEE 802.130、IEEE 802.131、IEEE 802.132、IEEE 802.133、IEEE 802.134、IEEE 802.135、IEEE 802.136、IEEE 802.137、IEEE 802.138、IEEE 802.139、IEEE 802.140、IEEE 802.141、IEEE 802.142、IEEE 802.143、IEEE 802.144、IEEE 802.145、IEEE 802.146、IEEE 802.147、IEEE 802.148、IEEE 802.149、IEEE 802.150、IEEE 802.151、IEEE 802.152、IEEE 802.153、IEEE 802.154、IEEE 802.155、IEEE 802.156、IEEE 802.157、IEEE 802.158、IEEE 802.159、IEEE 802.160、IEEE 802.161、IEEE 802.162、IEEE 802.163、IEEE 802.164、IEEE 802.165、IEEE 802.166、IEEE 802.167、IEEE 802.168、IEEE 802.169、IEEE 802.170、IEEE 802.171、IEEE 802.172、IEEE 802.173、IEEE 802.174、IEEE 802.175、IEEE 802.176、IEEE 802.177、IEEE 802.178、IEEE 802.179、IEEE 802.180、IEEE 802.181、IEEE 802.182、IEEE 802.183、IEEE 802.184、IEEE 802.185、IEEE 802.186、IEEE 802.187、IEEE 802.188、IEEE 802.189、IEEE 802.190、IEEE 802.191、IEEE 802.192、IEEE 802.193、IEEE 802.194、IEEE 802.195、IEEE 802.196、IEEE 802.197、IEEE 802.198、IEEE 802.199、IEEE 802.200、IEEE 802.201、IEEE 802.202、IEEE 802.203、IEEE 802.204、IEEE 802.205、IEEE 802.206、IEEE 802.207、IEEE 802.208、IEEE 802.209、IEEE 802.210、IEEE 802.211、IEEE 802.212、IEEE 802.213、IEEE 802.214、IEEE 802.215、IEEE 802.216、IEEE 802.217、IEEE 802.218、IEEE 802.219、IEEE 802.220、IEEE 802.221、IEEE 802.222、IEEE 802.223、IEEE 802.224、IEEE 802.225、IEEE 802.226、IEEE 802.227、IEEE 802.228、IEEE 802.229、IEEE 802.230、IEEE 802.231、IEEE 802.232、IEEE 802.233、IEEE 802.234、IEEE 802.235、IEEE 802.236、IEEE 802.237、IEEE 802.238、IEEE 802.239、IEEE 802.240、IEEE 802.241、IEEE 802.242、IEEE 802.243、IEEE 802.244、IEEE 802.245、IEEE 802.246、IEEE 802.247、IEEE 802.248、IEEE 802.249、IEEE 802.250、IEEE 802.251、IEEE 802.252、IEEE 802.253、IEEE 802.254、IEEE 802.255、IEEE 802.256、IEEE 802.257、IEEE 802.258、IEEE 802.259、IEEE 802.260、IEEE 802.261、IEEE 802.262、IEEE 802.263、IEEE 802.264、IEEE 802.265、IEEE 802.266、IEEE 802.267、IEEE 802.268、IEEE 802.269、IEEE 802.270、IEEE 802.271、IEEE 802.272、IEEE 802.273、IEEE 802.274、IEEE 802.275、IEEE 802.276、IEEE 802.277、IEEE 802.278、IEEE 802.279、IEEE 802.280、IEEE 802.281、IEEE 802.282、IEEE 802.283、IEEE 802.284、IEEE 802.285、IEEE 802.286、IEEE 802.287、IEEE 802.288、IEEE 802.289、IEEE 802.290、IEEE 802.291、IEEE 802.292、IEEE 802.293、IEEE 802.294、IEEE 802.295、IEEE 802.296、IEEE 802.297、IEEE 802.298、IEEE 802.299、IEEE 802.300、IEEE 802.301、IEEE 802.302、IEEE 802.303、IEEE 802.304、IEEE 802.305、IEEE 802.306、IEEE 802.307、IEEE 802.308、IEEE 802.309、IEEE 802.310、IEEE 802.311、IEEE 802.312、IEEE 802.313、IEEE 802.314、IEEE 802.315、IEEE 802.316、IEEE 802.317、IEEE 802.318、IEEE 802.319、IEEE 802.320、IEEE 802.321、IEEE 802.322、IEEE 802.323、IEEE 802.324、IEEE 802.325、IEEE 802.326、IEEE 802.327、IEEE 802.328、IEEE 802.329、IEEE 802.330、IEEE 802.331、IEEE 802.332、IEEE 802.333、IEEE 802.334、IEEE 802.335、IEEE 802.336、IEEE 802.337、IEEE 802.338、IEEE 802.339、IEEE 802.340、IEEE 802.341、IEEE 802.342、IEEE 802.343、IEEE 802.344、IEEE 802.345、IEEE 802.346、IEEE 802.347、IEEE 802.348、IEEE 802.349、IEEE 802.350、IEEE 802.351、IEEE 802.352、IEEE 802.353、IEEE 802.354、IEEE 802.355、IEEE 802.356、IEEE 802.357、IEEE 802.358、IEEE 802.359、IEEE 802.360、IEEE 802.361、IEEE 802.362、IEEE 802.363、IEEE 802.364、IEEE 802.365、IEEE 802.366、IEEE 802.367、IEEE 802.368、IEEE 802.369、IEEE 802.370、IEEE 802.371、IEEE 802.372、IEEE 802.373、IEEE 802.374、IEEE 802.375、IEEE 802.376、IEEE 802.377、IEEE 802.378、IEEE 802.379、IEEE 802.380、IEEE 802.381、IEEE 802.382、IEEE 802.383、IEEE 802.384、IEEE 802.385、IEEE 802.386、IEEE 802.387、IEEE 802.388、IEEE 802.38

(資料來源：勞動部勞動及職業安全衛生研究所，2020)

及育嬰留停榮工資料市集」及「失業給付勞工市集」四大市集（詳見表2）。另外再依縱切面資料所需，建構了「職訓技職十年追蹤」（包



含就服、職訓與技檢十年追蹤)、「失業給付勞工十年追蹤」(包含出生世代觀測)及「婦女生育留停十年追蹤」等三大跨時間串聯市集(詳見表3),以及產業勞動力存流量觀測橫切

面分析所需之八大市集(詳見表4)。未來亦會持續更新擴充追蹤資料庫包納的內容,並創造更多不同議題之視覺化自主分析平台。

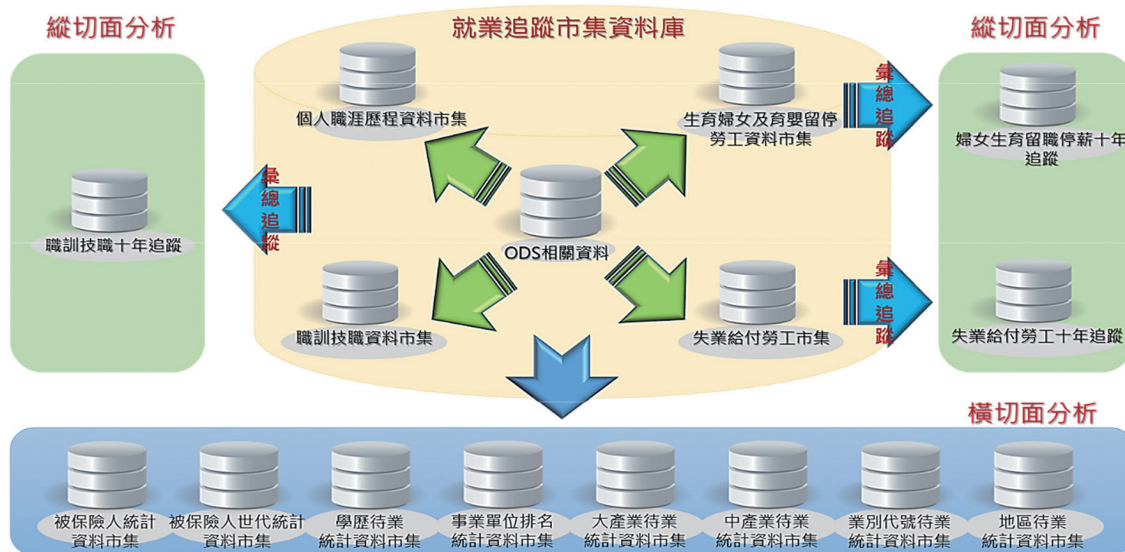


圖5 就業追蹤資料庫關係概念架構圖

(資料來源：勞動部勞動及職業安全衛生研究所，2020)

表2 匯流併整之相關市集列表

(資料年月紀錄至2020年6月)

單位：千筆

|                 | 筆數      | 資料說明                              |
|-----------------|---------|-----------------------------------|
| 個人職涯歷程檔         | 132,040 | 紀錄個人職涯工時狀態、加退保日期與加退保薪資            |
| 失業給付勞工市集        | 1,611   | 紀錄失業勞工離職時間、離職薪資、失業津貼起訖日與金額        |
| 生育婦女及育嬰留停勞工資料市集 | 2,775   | 紀錄生育婦女請領生育津貼金額起訖日與金額，以及回到職場時的薪資狀況 |
| 職訓技職市集          | 916     | 紀錄受訓人員所上課的相關資訊                    |

表3 縱切面分析相關市集列表

(資料年月紀錄至2020年6月)

單位：千筆

| 中文名稱          | 筆數    | 資料說明                                                          |
|---------------|-------|---------------------------------------------------------------|
| 失業給付勞工十年追蹤檔   | 1,615 | 追蹤失業勞工之失業十年追蹤，紀錄勞工因為失業事件後，每年是否有工作、工作是正職或兼職，以及該工作的最高勞退投保薪資。    |
| 婦女生育留職停薪十年追蹤檔 | 712   | 追蹤婦女生育留職停薪之十年追蹤，紀錄婦女因為生留職停薪後，每年是否有工作、工作是正職或兼職，以及該工作的最高勞退投保薪資。 |
| 職訓技職十年追蹤檔     | 916   | 追蹤參與職訓練勞工之十年追蹤，紀錄勞工因為參與職訓後，每年是否有工作、工作是正職或兼職，以及該工作的最高勞退投保薪資。   |

表4 橫切面分析相關市集列表

(資料年月紀錄至2020年6月)

單位：千筆

| 中文名稱         | 筆數      | 資料說明                                  |
|--------------|---------|---------------------------------------|
| 被保險人統計資料市集   | 20,737  | 統計分析各事業單位的勞退保加保、退保及在職人員數              |
| 被保險人世代統計資料市集 | 159,240 | 以勞工出生年代為分析基礎，統計分析各事業單位的勞退保加保、退保及在職人員數 |
| 事業單位排名統計資料市集 | 7,474   | 以各事業單位的勞退保加保、退保及在職人員數做為排名基礎           |
| 地區待業統計資料市集   | 408     | 以地區別為統計基礎，計算勞工平均的待業時間                 |
| 大產業待業統計資料市集  | 26      | 以大產業類別為統計基礎，計算勞工平均的待業時間               |
| 業別代號待業統計資料市集 | 1,610   | 以業別代號為統計基礎，計算勞工平均的待業時間                |
| 中產業待業統計資料市集  | 53      | 以中產業類別為統計基礎，計算勞工平均的待業時間               |
| 學歷待業統計資料市集   | 120     | 以勞工學歷類別為統計基礎，計算勞工平均的待業時間              |

## 二、平台應用與議題分析

在本節將以「職業訓練後之就業追蹤分析研究」為議題，運用「就業追蹤資料庫輔助決策支援平台」進行的應用範例分析。以2016年參加職訓者共178,098人觀察，分析追蹤3年之職場狀況如下。

### (一)職訓後就業率追蹤指標<sup>1</sup>

受訓者就業率指標第1年為82.7%、第2年81.3%、第3年80.5%，表現最好以在職訓練如充電起飛計畫、主題產業職業訓練、在職進修訓練就業率均在90%以上，但

其第3年的就業率一樣會低於第1年的就業率。就業率較低者多屬於邊際弱勢就業族群，包括如準更生受保護人訓練計畫、推動原住民團體辦理原住民地區失業者職業訓練、原住民專班訓練等就業率低於40%。在圖6中可得知，圖中紅色直線左方的各項職訓，其第三年的就業率高於第1年，越往左第3年就業率高出越多，其中接受企業委託訓練的第3年就業率較第一年多出12.44個百分點；反之，越往右第3年就業率則相對低，其中產學攜手合作計畫第3年的就業率降至28.44%，值得再去深究探討其原因。

<sup>1</sup> 職訓就業率之定義為「第n年職訓者有就業總人數／職訓課程當年O月完訓人數」。追蹤指標之計算方式為，職業訓練課程於O年O月完訓者，往後計算第n年當月有多少比率就業，自動化程式最長可追蹤到10年。

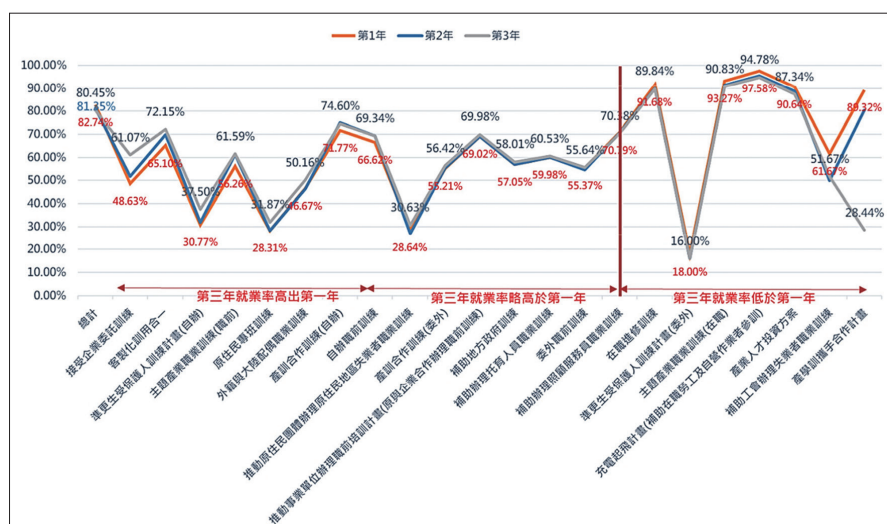


圖6 各職訓後就業率3年追蹤

(資料來源：就業追蹤資料庫輔助決策支援平台，2020)

(二)職訓後就業穩定度指標<sup>2</sup>

整體就業穩定度指標，追蹤第1年0.9、第2年之後下滑0.7、第3年維持0.7。各項計畫均以第1年表現最好，之後呈現下滑趨勢。第2、3年以後表現落差最大（超過

0.3），包括有為外籍與大陸配偶職業訓練、原住民專班訓練、準更生受保護人訓練計畫、接受企業委託訓練、補助辦理托育人員職業訓練等。（圖7）

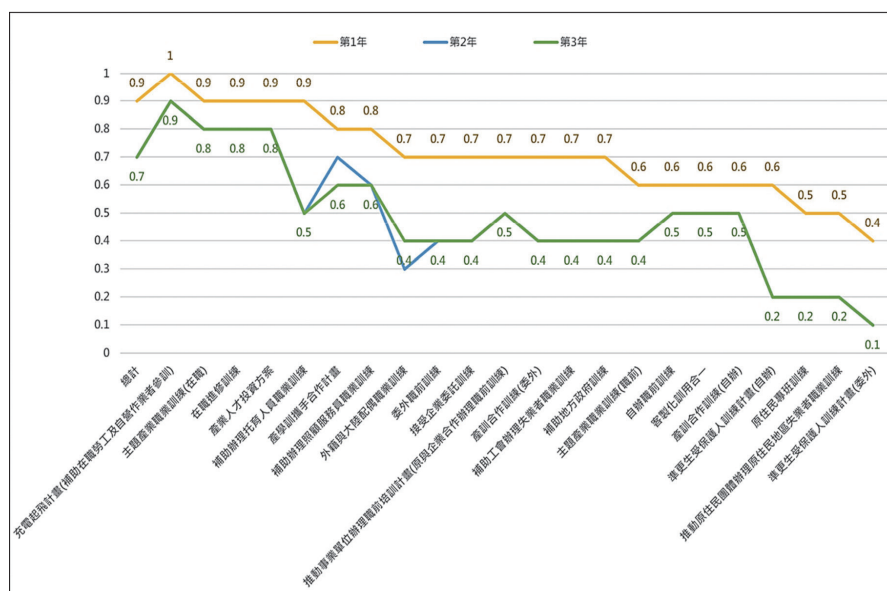


圖7 各職訓後勞工薪資及成長率3年追蹤

(資料來源：就業追蹤資料庫輔助決策支援平台，2020)

<sup>2</sup> 職訓後就業穩定度指標之定義為「職訓課程完訓後實際工作期間／完訓後一定追蹤期間（單位：月）」。職訓後實際的工作期間佔追蹤期間比例（越接近1表示越穩定就業），已設計自動化程式可每年累計最長追蹤10年。以2012年5月完訓者為例，若追蹤到2020年5月（共5年計60個月），其中有48個月有投勞保，則就業穩定度為48/60=0.8。



### (三)職訓後薪資成長率指標<sup>3</sup>

本指標之計算需串接勞退雇主提撥投保薪資，資料串接後勞退人數約佔整體55-56%，未有勞退者多屬於職業工會、自營作業者。就業追蹤第1年平均勞退投保薪資為37,390元、第2年38,843元、第3年40,281元均有成長趨勢，成長幅度分別為123%、

127%、132%。薪資表現最佳為接受企業委託訓練計畫、主題產業職業訓練（在職）、補助辦理照顧服務員職業訓練；而產業人才投資方案薪資成長幅度第三年表現也達126%，而薪資表現比原先低者為補助工會辦理失業者職業訓練，追蹤第3年薪資成長率僅為原先95%。（圖8）

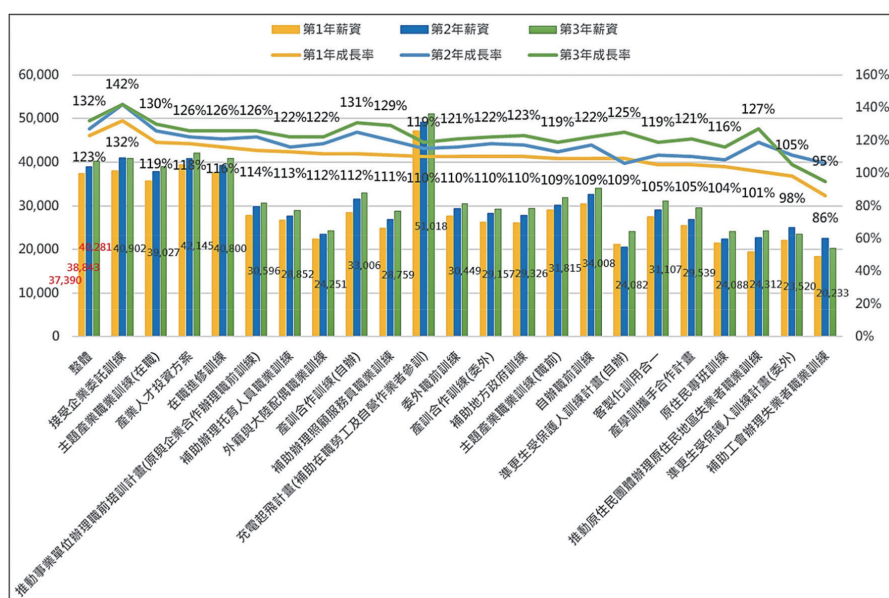


圖8 各職訓後勞工薪資及成長率3年追蹤

（資料來源：就業追蹤資料庫輔助決策支援平台，2020）

透過上述指標在追蹤資料庫上的分析，可為各項職業訓練活動建立長期成效評估之追蹤指標。針對特殊身份或對象之職業訓練，經追蹤分析後若發現對輔導就業的成效不彰時，除了在技能或培訓方式上強化外，亦應考量是否應再增列預算或再舉辦相關的業界媒合活動，來提升其就業率及就業穩定度。

## 肆、國軍退撫職業訓練及追蹤輔導

依據國家未來發展情勢，我國的經濟結構與人力資源的運用已從勞力密集轉型為高科技高附加價值的方向發展。隨著時代的演變，國防建軍不應再是以人多取勝的人海戰術，而

<sup>3</sup> 職訓薪資成長率指標之定義為「（第n年職訓者有就業者勞退投保薪資）—（職訓課程當月的勞退投保薪資）／（職訓課程結束後當月的勞退投保薪資）」。職訓課程結束後，每一年薪資成長幅度，已設計自動化程式最長可以追蹤到第10年。

應循現代化的軍隊發展，走職業化、專業化的「精兵」路線。（詹中一，2012）我國亦於2018年1月1日起，開始實施完全募兵制，將國軍人力完全朝職業化的人才招募來邁進，並期能以「精兵」培育的方式來確保國軍在各類人才上的專業素質。

國軍展開募兵制變革，常以「在軍中培養工作技能，習得一技之長」吸引適齡青年投身軍旅，但民間勞動市場與部隊環境之差異，常使具有技術專長的軍人退伍後無法立即銜接社會專業需求。最近10年軍職人員退伍之平均年齡為30.3歲，軍職人員請領退伍金為26.1歲，請領退休俸為43.0歲。（公務人員退休撫卹基金統計年報，2020）在正值青壯年時，職涯就面臨職場轉換情境，若無法順利在新的工作上有所發展，勢必會成為優質人才選擇投入軍旅考量時的極大疑慮。因此，退除役官兵適當的職業訓練及就業輔導推介，不只是國軍退除役官兵輔導委員會（以下簡稱退撫會）的職責，也會溯源影響到國軍優質人才之招募，與國軍人力素質之維持及提升息息相關。

根據退撫會委外的調查，近九成的屈退官兵在退役後預計會繼續找工作，其未來退役後希望找尋的工作職類中發現，屈退官兵以希望找尋「餐飲／旅遊／美容美髮類」職類為主（占16.1%），其次為「操作／技術／維修類」職類（占15.3%）。在針對產業界目前需求的員工職類的調查，則以「客服／門市／業務／貿易類」為主，占23.1%，其次為「操作／技術／維修類」，占15.9%。（黃榮護、張緯良，2017）在供需雙面向比對來看，屈退官兵之期望與企業之需求間似乎並未有很大的落差。

在接受職業訓練的需求上，61.3%的已核認的第二類退除役官兵<sup>4</sup>未來想要參加職業訓練，45.4%的未核認者也表達有意願參加職業訓練。第二類官兵未來選擇參加職業訓練的主要種類為「輔導會委外訓練」（已核認占29.5%、未核認占26.0%），其次為「輔導會自辦養成訓練」（已核認占24.9%、未核認占22.5%）。在針對未來有意願參加職業訓練者中，不論是已核認者及未核認者，都以「餐飲服務職群」為其最想要參與的職訓類型，二者選擇讓職訓的比例分別為12.4%及12.6%；在已核認者中，有11.2%對「商業與行銷職群」的職訓類型有興趣，未核認者則較多對於「機械修復職群」（10.2%）職訓有興趣。（黃榮護、張緯良，2017）

透過調查，雖然已可掌握國軍官兵對職業訓練及未來轉業的需求，但是不同職業訓練的成效，則需由透過收集退除役官兵轉業後的職涯現況或歷程才能評估。若能透過大數據匯流分析之整合，將退除役官兵及職業訓練參與相關資料集建置大數據匯流分析資料庫，再與前述勞動及職業安全衛生研究所之「就業追蹤資料庫」踴檔串整，則可以定期自動的不斷追蹤退除役官兵轉業後的職涯發展，透過大數據分析了解官兵退役後10年內的就業率、就業穩定度，以及其工作產業類別以及薪資成長狀況，除了可做為退撫會各項職訓成效的評估依據，也可依數據決策的理念，來修正或提出更好的就業輔導方案。

## 伍、結語

「大數據」成為近十年最熱門的話題，在

<sup>4</sup> 第二類退除役官係指志願服軍官、士官、士兵役，其服役四年以上未滿十年，或服現役期滿，志願留營或再入營服現役，先後合計四年以上未滿十年退伍、除役或解除召集者。

未來的十年依然仍會是主流。而政府組織運用大數據決策的效益，則應該由建置「大數據匯流資料庫」著手。而大數據匯流分析資料庫之建置不是一件短期的任務，所需耗用的人力及預測資源亦不少，因此在政府相關組織在評估建置的過程中，常會有方案難選與決策難定的困境。本文所述之理念及應用成功案例，應可讓國防相關單位在邁向以大數據分析為決策的治理之路上能有所參循。如何在有限的預算資源下，能逐年逐步來實踐大數據匯流資料整合與應用分析實踐，本文提列重點建議彙整如后。

### 一、跨部門或部會的合作

大數據分析的取得、儲存、運算及視覺化等四大步驟，需有資訊技術、統計分析及領域知識等三大類專業人才來共同完成。因此在大數據匯流分析平台建置的起步，就應能有跨部門的合作，邀集資訊單位及業管研究單位來共同合作規劃。而其分工方式不宜以大隊接力法為之，應以兩人三腿齊心合作的方式為之。

另外，如本文所提及之勞動及職業安全衛生研究所之「就業追蹤資料平台」，已將勞動力就業研究之相關資料進行整併，且預計於今年開始試行提供外界申請使用。未來在針對國防相關大數據分析時，若需使用相關資料，則應以跨部會合作方式為之，無需重複匯流及整併。

### 二、議題及指標導向為引導

大數據分析平台建置成功的核心要件為需要明確的策略及應用方向，但短期的建置成效則是大數據專案是否繼續獲得組織上層決策者支持的關鍵點。由議題來引導大數據匯流分析平台之建置，在進行資料集建置時，應依據「指標建立漏斗法」分別由「指標價值」以及

「指標策略」二面向來評估大數據指標的價值重要性或策略需用性，有條理的彙整出有意義且必須建置的大數據分析所需衍生欄位，可使分析成果能在較早期有所展現，必使建置專案能更順利之推展。

### 三、依分析需求規劃資料庫層級

產生具策略參考價值之分析與研究成果是大數據匯流分析平台建置的最終目的。各類議題研究分析所需用的資料結構型態亦不盡相同。因此在建置大數據分析匯流分析資料庫時，應考量未來政策議題研究所需的各類指標，考量縱切面及橫切面分析的各項資料結構需求，分層建置「基礎資料集」、「加值資料集」、「追蹤資料集」及「統計資料集」等四類資料集。不同的資料庫，以便研究人員可快取用。亦即，在建置匯流平台時，要能以使用者導向去思考規劃，讓平台能提供研究人員需用指標的成品或半成品，以加速研究之進行，並節省不必要的重複資料取得及處理工作。

### 四、國軍可以軟性議題為出發

國防數據資料多具有機敏性，因此在大數據匯流分析應用與執行上，國防部各級相關單位通常會以更審慎評估的態度為之。因此，在建置大數據匯流分析資料庫時，建議可由與人才招募息息相關之國軍退除役就業輔導研究為議題來進行。針對退除役官兵在退役後之生涯發展歷程做追蹤分析，並結合國軍職業訓練之相關留存資料，針對參與各類職業訓練之退除役官兵其在就業率、就業穩定度及薪資水平上做分析及評估。此大數據分析的成果除可對職業訓練成效做長期的追蹤評估，亦可能透過大數據分析明瞭退除役官兵所從事的產業，提共未來職訓課程規劃所參用。

大數據資料即是黃金，大數據匯流分析資料庫則是將大數據礦石提煉為黃金的加速器，若能成功建置此加速器，則能在資源共享下，使預算擲節，亦可使組織之智識留存、分享及擴散，最終使數據治理得以達成。

## 參考文獻

1. 吳慧娜、梁德馨(2021),就業追蹤資料庫系統優化與推廣應用規劃工作計畫.
2. 勞動部勞動及職業安全衛生研究所.
3. 彭子珊(2016),資料變黃金數據長該怎麼做,天下雜誌,603期,2016-08-02.
4. 詹中一(2012),國軍全面實施募兵制對廣儲後備影響之研究,海軍學術雙月刊,46(4),頁124-139.
5. 張玉山；陳仁惠；黃春長；吳慧娜(2019)從勞動力供需變化趨勢,找尋經濟新未來,產業管理評論10卷2期,頁9-36.
6. 黃榮護、張緯良(2017),第二類退除役官兵就

學就業職訓需求調查研究,國軍退除役官兵輔導委員會.

7. Matt Shepherd(2018),"Is Data The New Gold?",CEO Today Magazine Today.(from <https://www.ceotodaymagazine.com/2018/04/is-data-the-new-gold/>)



### 梁德馨

#### 現職：

輔仁大學統計資訊學系教授  
輔仁大學管理學院副院長  
台灣析數資訊整合策略研發執行顧問長

#### 學歷：

淡江大學管理科學博士碩士

#### 經歷：

實務經歷包括製造、零售、保全、金融、保險、公營行庫、政府機關等產業的各類大數據分析專案，目前領導台灣析數公司，負責發展策略、研發及解決方案等之規劃及推展