

# 大數據的政策敘事分析： 基本框架與案例演示

莊文忠\*

資料向來是闡述公共政策中重點故事所不可或缺。自 1980 年代起，由於網路、電腦及行動裝置的興起，闡述的管道變得五花八門。如今，資料不僅變得更多，也產生出新型態的資料。…新的分析及視覺呈現工具，可以讓說故事時如虎添翼，也更能讓資料說話。

NATIONAL ACADEMY OF CIVIL SERVICE

—Donald F. Kettl

國家文官學院

## 壹、前言

近年來，資料科學（data science）大行其道，而大數據（big data）和資料視覺化（data visualization）成為重要的推動力量，不僅讓分析人員或決策者可以透過蒐集和分析即時、多樣與大量的資料，更精準地掌握與解決問題相關的資訊，且生動的圖表也易讓人一目了然，快速掌握重點。不過，決策

\* 世新大學行政管理學系教授。

者或管理者若只看數據或圖表就足以理解問題的本質和做出妥適決策嗎？政府部門只用數據或圖表就能達到和民眾溝通政策的效果嗎？隱藏在數據和圖表背後的政策意涵該如何有效且合理地解讀和詮釋呢？這或許是公共管理者在追求大數據之分析與應用的同時，值得深思與探究的課題。

換言之，吾人現今身處於資訊科技高度發達的時代，通過各式各樣的感測、辨識技術，可無時無刻地將人們的一舉一動轉化成大量的資料，如瀏覽網頁、搜尋文章、購買電子票證、刷卡停車等；再通過各種傳輸技術，隨時隨地將這些數據傳送到資料儲存設備，並持續地更新與累積。事實上，這些即時且大量的資料，在未經系統化整理與分析之前，大多充滿雜訊，管理者不易直接從中取得有用資訊，需經過清理、整合、分析和判讀之後，才能作為決策之用。近年來，無論是預測分析、機器學習，乃至於人工智能，都與大數據的分析與應用密切相關。

另一方面，就資料分析結果的呈現而言，雖說數據本身並不具有好壞的意涵、一張圖表勝過千言萬語，但是，一旦這些數據被置於不同的脈絡或情境中，便不再是中立的客體，往往被賦與特定意涵或價值判斷（莊文忠，2009）；若是將這些數據進一步轉換成視覺化的資訊圖表（infographic），不同的人看同一張圖表時，其感受可能完全不同，甚至有截然不同的解讀或詮釋。因此，如何精準地傳達數據或圖表所呈現的意義，實有賴於資料的分析人員和決策者，具有看數據或圖表說故事的能力，也應在理解問題情境和政策現

況下，對數據或圖表的意涵進行描述和解釋。

時至今日，在數位傳播與多元滙流之下，創造多種資料來源、多樣敘事方式的可能性，阮明淑（2019：98）對此一趨勢指出，「用資料說故事」（telling stories with data）乃是數位敘事跨域合作過程中之重要關鍵，有些研究關注在「敘事視覺化」（narrative visualization 或 visual storytelling），強調運用資訊視覺化技術對每一種資料片段進行組合，形成一組故事或一張資訊互動圖表；也有研究關注「資料驅動之敘事」（data-driven storytelling），強調資料是分析、視覺化及敘事之基礎，透過深入挖掘各種資料內容與特定資訊、發現資料間之隱含關連性，向閱聽眾闡述資料中之故事。是以，面對不同閱聽眾對資訊圖表的故事內容理解與解讀不一致之狀況，實有賴於分析人員系統性整理故事內容元素，使其可以映射「資料」與「故事」間之關係。

由此可知，在資料科學崛起與普及化之後，大數據分析或許不再是組織創造競爭優勢的利器，但卻是輔助決策和檢驗成效的有力證據之一；此外，就資訊科技的發展趨勢來看，資料分析的最大障礙應不在於技術的進階發展，而在於資料意義的擷取與詮釋，進而連結到決策過程。因之，本文的目的並非探討伴隨大數據分析所衍生的個人隱私、公平正義、資料濫用或資訊安全等議題，而是從資料敘事分析的觀點切入，析論大數據與政策敘事分析（policy narrative analysis）的基本框架，並以中央選舉委員會所建置之「選舉人投票資料檔」為例，進行初淺的展演與說明，期能發揮拋磚引玉之效果，

有更多研究投入此一議題的探究，開展「循證」(evidence-based)的政策分析途徑。

## 貳、大數據與敘事分析

人類社會隨著資訊科技的日新月異，擁有愈來愈多可以產生更多數據與儲存更多資料的技術與方法，在各種行業、組織或場所中，常可見個人行為與集體活動被量化、計算，並加以記錄、分析和應用。不過，到目前為止，我們利用這些資料進行統計分析，大多只能告訴我們「是什麼」(what)，而不是「為什麼」(why)和「如何做」(how)。舉例來說，分析人員可以從報表上看到現象 A 和現象 B 之間的相關係數高達 0.8 以上，但是，為什麼會有這麼高的關連性？兩者之間有無因果連結？其連結機制又是如何形成的？如何連結到政策規劃？從數據本身並無法得知。因此，資料分析人員若只具有分析大數據的能力是有所不足的，還需要具備敘事分析的能力。

誠如論者所言，身處於大數據時代中，在凡事關注資料之量化表現之餘，更要在意資料間之各種細節與故事，就如同「厚數據」(thick data)強調要讓資料變得有情感，應針對小量的樣本，利用質性研究方法深入分析與詮釋「不同資料背後」之隱含意義、脈絡或故事，而非僅限於資料視覺化後的故事展現(阮明淑，2019：99)。換言之，相較於大數據強調標準化的資料格式和依

賴機器大量運算，厚數據更強調通過非結構化的文字資料，深入挖掘和闡釋數據或文字所富有的意義，其目的在於找出事件的因果關連或特定人群的需求，解析故事背後的深刻意義、故事和情感。

在概念上，「敘事」是幫助個人理解、思考、組織和溝通數據的基本方法，將數據置於情境和脈絡之中，它便不再處於孤立狀態，且被賦與特定的意涵。因之，好的敘事不僅是表達清晰、邏輯嚴謹、自然生動，更重要的是，能與生活經驗相互連結，有助於掌握事件發生的前因後果。它在表達形式上，可以用客觀、理性的論述方式呈現，也可以用主觀、感性的抒情方式分享。阮明淑（2019：99）也指出，敘事就是在「說故事」，大多數人都喜歡聽故事，因為那是人類互動、獲取知識之基本方式，也是人類思考之主要模式，使用此種你所熟悉、自然自在之資訊接收與傳遞方式，用以感知世界、進行溝通、瞭解事物或彼此。由於說故事有著易吸引人、容易被處理及易被瞭解、可克服消費者對訊息內容之反抗性等優點，進而達到說服效果，因此，愈來愈多的人開始運用此一方式進行資訊傳播或產品行銷。

近年來，各學科領域盛行以視覺化資訊圖表來呈現大數據分析的結果，敘事分析因而增加可以輔助理解和溝通的利器，如 Kettl 即強調，光是挖掘資料仍然不夠，因為數據本身無法不言而喻；但如果表達方式正確，就能讓資料說出強大的故事，關鍵在於優秀的資料視覺化：將資料轉換為各式各樣的表現形式，包含圖、表、地圖，乃至其他資訊圖表（謝孟達譯，2019：42）。



基本上，資料視覺化乃是透過資料的圖像化再現（graphic representation）或符號式代表（symbolic presentation），以有效達到理解和溝通的效果，諸如美食地圖、捷運路線圖、空污顏色燈號等，分析人員透過將資料中的地理位置、連結關係、嚴重程度等資訊，以色彩鮮明圖形或對比強烈的符號呈現，縮短使用者閱讀文字的時間和增強訊息傳播的效果。要言之，資料視覺化的目的主要是呈現數據的趨勢化、結構化或空間化、以圖形、圖表或地圖理解數據之間的關係模式、規律通則或變化過程，此不僅具有檢查、驗證數據運算的正確性和合理性的功能，輔助資料使用者的探索、思考和判斷，也能加深對資料的認知和理解，促進有效溝通與傳播。

基本上，資料敘事分析乃是結合資料（Data）、敘事（Narrative）及視覺化（Visuals）3 個要素，將資料轉化成架構事實，放入敘事基模（Narrative Schema）中，以類故事之資訊型態傳遞資訊，並輔以視覺化圖表，透過資料與敘事之間之詮釋（Explain）、資料與視覺化間之啟發（Enlighten）、敘事與視覺化間之參與（Engage），提供讀者敘事感受，並在適當之安排下吸引注意（Dykes, 2016, 2017; 阮明淑，2019）。不過，學者們也提醒，資料敘事之核心概念應是以敘事為主，以資料及其視覺化技術為輔，在敘事分析的同時，加入一連串描述故事高潮迭起情節的佐證資料，並適度運用視覺化技術易受注意、易於理解及記憶等優點，將敘事主體進行重新解構與視覺化再製，使用符合使用者慣用之語言和資訊處理策略，吸引他們主動閱讀並能充分理

解，誘發他們關注原本未注意之主題、興趣，主動地、更深入地探究資料或發掘資料，進而達到有效溝通與傳播之目的（Kosara & Mackinlay, 2013；李長潔，2017；阮明淑，2019）。

由此可知，在大數據分析中，分析人員除要擅長蒐集及分析數據，設計和產製有內容的圖表，吸引讀者目光的能力外，也要具備良好的邏輯論述能力，把大量的資訊圖表串連成一個主題式的故事題材。若以公共政策為例，在資料敘事分析中，分析人員必須重新解構和建構問題，闡述問題和方案之間的因果關連，並結合相關資料的提示或佐證，選擇適當的文字語言表達意思，幫助決策者或一般民眾釐清問題的本質和解決方案的有效性，而資料視覺化不應只是吸引讀者注意力的手段，應思考如何運用合適之資料視覺化來增進理解，甚至提供同一資料不同觀點，幫助讀者掌握問題的不同視角，進而產生和不同立場者溝通政策的效果。

有鑑於故事不是說給自己聽，而是要讓標的對象能夠充分認知和理解，因此，在說故事時，必須先確認資料的使用者是誰，並以「使用者為中心」（user-centered），選取符合其特質的語言和場景，決定資料敘事分析的主題或軸線，若是在互動性的溝通場合中，例如會議、工作坊、公聽會等，還可以根據使用者的當下反應或參與程度進行調整或修正，以達到最佳的溝通效果。值得注意的是，雖然故事內容的戲劇性或新奇性可以吸引使用者的注意，但資料敘事分析畢竟不是個人意見的表達，而是以數據主要題材或是搭配圖

表的呈現，將其中所隱含的觀點和意義清楚地表達，應該避免過度詮釋、推論數據或圖表所無法佐證的論點，造成資料使用者的誤解，反而無助於問題的解決或方案的推動。

為達到溝通與傳播的效果，阮明淑（2019：106）提出資料敘事分析的概念性框架（如圖 1 所示），並將其區分為 3 個階段：

### 一、探索資料

先選擇不同資料來源，透過資料擷取、清洗及轉換等資料處理步驟，得到研究者所需之分析素材；再依據素材之特性（如文本、數值或其他型式），選擇合適之資料分析方法進行內容分析或主題分析，其目的在於梳理與解析不同資料間，對同一事件或議題之故事內容差異或邏輯。

### 二、建構故事

依據分析目的，選擇故事（或事件）內容元素並進行編排或重組，故事情境的連結應清楚且易為讀者所感知；再依照事件發生之時間序列進行故事情節安排；在角色的選擇上可安排在人口變項上與目標對象或意見領袖相近之素人證言說服，強調傳播者與接收者之共同性；同時，把故事情節連結至現實世界之發生情境，其目的在於表達分析者對故事內容之觀點或看法。



### 三、訴說故事

故事情節著重在真實性、可想像，並符合文化適當性之探討。當分析者完成故事內容編排後，可選擇合適之視覺化呈現型式（如圖文、影像或多媒體）或結合資料視覺化工具進行資訊圖表設計。其中，分析者需要思考何種視覺化呈現型式最能夠詮釋其所欲傳達之理念或意義，進而可以透過目標族群引發更大範圍之擴散效應，其目的在於讓目標族群能夠理解或感受故事內容之意義和價值。

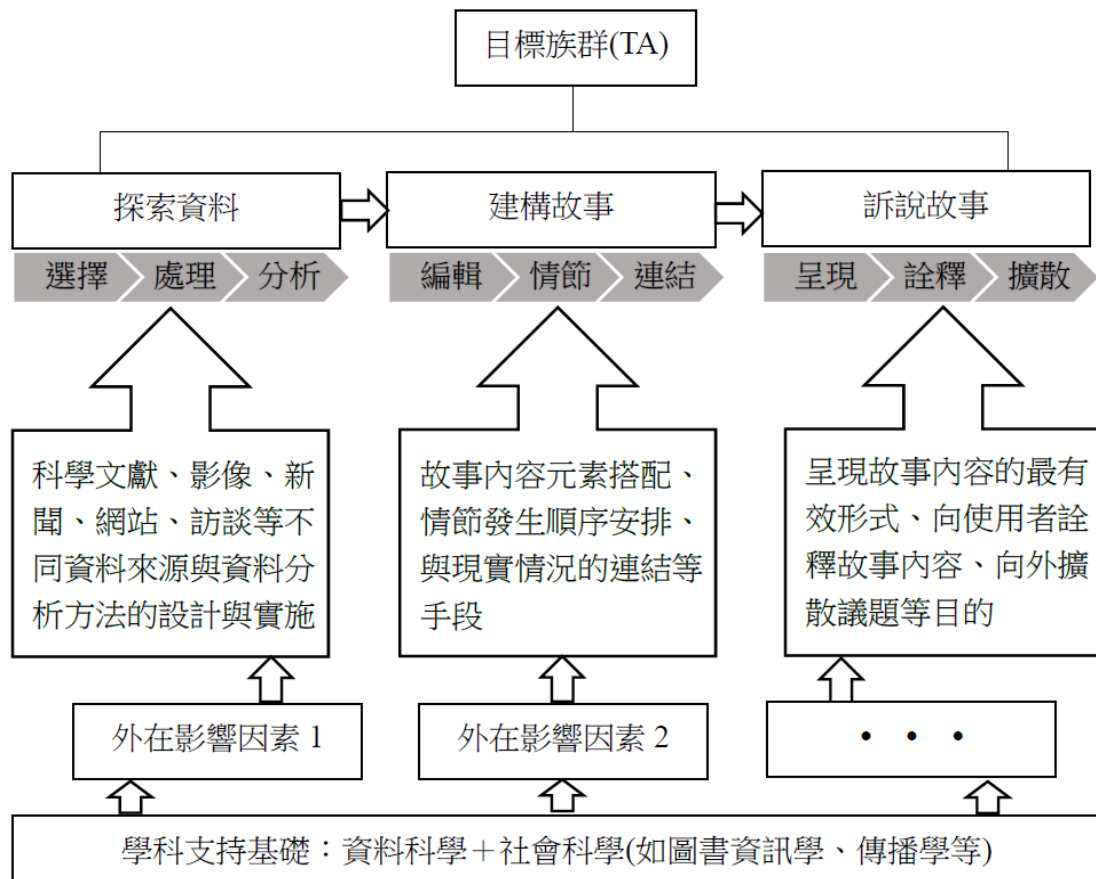


圖 1 資料敘事分析之概念性框架

資料來源：阮明淑（2019），頁 106。

綜上所述，吾人透過大數據進行敘事分析時，應注意幾個要點：(一) 以數據為本，但不囿限於數據本身，宜結合問題取向、背景因素和脈絡情境加以解讀，以避免陷入「數字迷思」，讓分析和論述更具有情節性和故事性，既見樹又見林，也更容易理解資料所映射之社會意涵。(二) 敘事本身不僅是反映客觀事實，也能透過分析重新還原或建構事實，另外，分析者或決策者的價值偏好也可能透過資料的選取和結果的詮釋，影響讀者的判斷和指引政策方向，多元資料來源與資料交叉檢證可以減少政策盲點。(三) 無論是橫斷面或縱時性的資料，大多是屬於觀察性的資料，並非是以隨機控制實驗 (Randomized Control Trial, RCT) 所取得資料，難以憑藉資料本身提出因果論述，若應用在政策方案的選擇或政策成效的評估上，宜有更為嚴謹與充分的邏輯論證過程。(四) 分析人員利用大數據進行敘事分析時，除了說明資料分析方法外，也應該清楚交待資料來源、資料性質、資料範圍、蒐集時間和處理過程等資訊，以提高資料敘事分析的透明度和可信度。

### 參、大數據與政策敘事分析

過去，由於社會科學領域所蒐集的數據都是屬於觀察性資料而非實驗性資料，因之，統計學所強調的基本常識是「相關不等於因果關係」(Correlation is not causation)，早已提醒分析人員不要誤將兩個具有高度相關的社會現象 (或事件)，直接解釋為因果關係，必須釐清這兩個現象之間的時間先後順

序，以及其背後是否還有其他可能的共同影響因素。換言之，觀察性的資料只適合用於描述現象之間的關連性，但並不足以充分解釋或推論現象之間是否具有因果關係，而公共政策期待透過行政作為或政策工具的「介入」(Intervention)，達到可欲的或預期的政策效果，實驗性的資料或許更為符合此一需求。此外，我們還是需要注意到大數據本身所存在一項隱性的偏差(bias)，即能夠進入到資料庫的個案，和還沒有進入到資料庫的個案，兩者可能是不同特徵的樣本，要用這個大數據來推論所有的個案，謹慎為之，方為上策。

然而，這些警告似乎並沒有阻止我們利用觀察性資料追求現象(事件)與現象(事件)之間因果關係的企圖心，其中一個重要的原因即是，每一個人都需要資訊做理性決策，如果可以確認導致某一現象發生的原因是什麼，便有助於我們針對原因採取正確的行動，以促成(或阻止)該現象的發生。個人的決策有此需求，公共政策的制定更是如此。每一項公共政策所影響的層面甚廣、所需耗費的資源甚多，若無法釐清公共問題發生的原因、或政策方案介入的結果，我們便無法評估某一政策的成敗，進而要求決策者負責。

不可否認地，公共政策不如自然科學之處在於，受到道德、倫理、平等及法律等諸多因素的限制，許多的社會問題難以透過嚴謹的實驗設計來控制可能的影響因素，進而確認政策介入的社會效果為何。因此，學術上有關公共政策因果機制的討論一直很缺乏完整的論述，在實務上自然也就導致不少

的誤用或濫用的情形。不過，這幾年來，由於資訊科技的快速發展，無論是資料的蒐集、儲存、轉換或分析，在數量與效率上呈現倍數增長，大數據分析浪潮席捲各個領域，也為政策過程因果機制的探討與檢證燃起希望。

申言之，分析人員不僅可以透過大量數據分析找到有相關的現象，擺脫抽樣調查或樣本數過少所產生的誤差問題；而且，也因為資料記錄的時間單位越來越精細，透過分析比對現象發生的時間軌跡，甚至可以研判現象的發生先後，確認彼此之間的時間順序；此外，「跨域」資料的整合分析也可以同時納入多元的原因變數（自變數），稍微彌補公共政策難以進行實驗設計的缺憾，此亦有助於問題與對策之間由相關論述向因果論述逼近。雖然，這離自然科學的因果推論還是有一段距離，但大數據分析如運用得當，不僅可以幫助企業行銷人員從資料中發掘潛在的消費族群、更精準地投放廣告，還可以幫助政策制定者提高決策的品質，減少錯誤決策的風險和成本。

由前文討論可知，政府或企業早已透過運用各種資訊技術將人類生活和社會現象「數據化」，姑且不論其對侵犯個人隱私的法律議題，大數據實已提供各種活動或現象的時間和空間的詳細紀錄，讓決策者找到一個利用龐大資料預測人類行為和理解社會問題的新方法。繼此接踵而至的問題是，如何研判哪些資料和問題的解決有關？如何詮釋資料分析結果背後的因果關係？如何理解其背後所隱含的社會意義或政策意涵？如此，大數據分析才不會只是停留在「資料本位」的探勘階段，真正進入「問題本位」或「政策本位」的

階段，由資料敘事分析導入政策敘事分析。

如前文所言，敘事分析是一種說故事的敘述邏輯方法，可以將冰冷、無趣的數字，轉換成有感、生動的語言，讓一般人可以理解，不可否認地，在大數據時代，運用說故事方式呈現資料將愈來愈普遍（Stolper, Lee, Riche, & Stasko, 2016; 阮明淑，2019）。事實上，大數據的資料敘事分析與政策敘事分析並非截然不同，兩者的共通性大於相異性，兩者的結合應用並無違和之處，反而是相得益彰。

要言之，政策敘事分析是以說故事的方式陳述問題界定到問題解決的過程，故事的內容主要包含：問題本身的陳述、問題與後果之間的關連性、重要利害關係人的指認、問題與方案之間的因果關係等，數據報表和統計模型同樣是最常見的資料呈現方式，而資料的視覺化則是加強故事的動態性和可理解度。如同資料敘事分析的框架，以大數據為基礎的政策敘事分析的元素至少包括：資料(Data)、敘事手法(Narrative)及政策意涵(Policy implication)，而視覺化(Visuals)此一要素的加入，則讓前三者的結合更為生動，因之，資料與政策透過敘事方式連結在一起，輔以資料視覺化的啟發，提高政策論述、溝通及傳播的效果。

Kettl 也強調，要縮小公共政策在證據方面的「供給—需求」落差，讓證據在決策過程中說話，應重視 5 項原則和面對不同挑戰（謝孟達譯，2019：



9-11)：

**原則 1：客戶想要且用得上的證據，才有用處。**因此，除了須提供決策者想要且需要的佐證外，決策者也須創造讓證據產生有用性的需求。這就是平衡供給與需求的挑戰。

**原則 2：掌握故事很重要，而且必須正確掌握。**不論是依據「黃金標準」所做的政策分析、隨機對照試驗、或是來自街頭的印象，證據往往以各種形式現身。好的證據必須有效、可靠且及時。這就是資料分析的挑戰。

**原則 3：說故事時，要用各種方式傳達證據真正所言（不偏離分析專家提供的所知證據），且要用清楚的語言傳達（滿足需要該證據的政策制定者所須知）。**這是資料視覺化的挑戰。

**原則 4：推銷故事時，要用各種方式讓證據具有說服力。**分析人士有時假定，錯綜複雜的研究問題抽絲剝繭之後，成果將不言而喻。事實上卻從來不是如此，因為成果永遠伴隨著不確定性，而且總是會有多方論點爭取決策者的青睞。這是政策說服的挑戰。

**原則 5：讓證據凌駕喧囂。**在資訊爆炸與訊息洶湧的環境下，社群媒體隨時能使一切改觀，甚至讓分析專家多年來想要探究的現象出現更迭。為了做得更好，必須先有更透澈的認知；而要有更透澈的認知，則

須先克服最後這項挑戰：即時透明化。

論者指出，在資料敘事分析中，故事結構（即事件情節時序）可採三段式：以「開始」點出問題，以「過程」鋪陳發現解決問題方法之過程，以「結尾」為最初問題提供解決方法（Kreuter et al., 2007；阮明淑，2019）。本文從政策管理的觀點指出，政策敘事分析的主要結構可分為 5 個階段，如圖 2 所示，各個階段的重要元素如下所述：

### 一、陳述問題（Clarifying problem）

分析人員需要先釐清問題的背景與現況，或是透過資料的探索性分析，發掘出與問題相關連的重要因素，進而提出正式分析或大規模分析的需求。

### 二、分析資料（Analyzing data）

由分析人員就目前所蒐集和整理的資料，運用適當的統計技術進行分析，以檢證問題相關因素之間的關係，在分析過程中，如需要其他資料的佐證或延伸分析，便要再蒐集或串接其他的資料。

### 三、將資料視覺化（Visualizing data）

分析人員除了利用描述性統計或統計模型，說明數字的高低、分布、關連和變化等，基於方便認知和理解，可以選擇或搭配合適的視覺化工

具，將分析數據轉化成圖表、地圖等形式。

#### 四、述說故事 (Narrating a story)

分析人員利用各種分析方法所得到的數據作為主要素材，結合角色、情節、時間、地點等基本內容元素，構思和闡述問題的框架與內容，包括先後順序、因果聯結、推論邏輯、發展脈絡與可能結局等，據以指出各種可能的解決方案，其在整體情境中之意義。

#### 五、提出方案 (Proposing alternatives)

根據問題發展的過程與可能結果，思考政府可以採取什麼行動來解決問題或滿足需求、行動和問題成因之間的連結機制、行動和方案成效之間的連結機制等因果關係的論述，以提出相對應的解決方案，供決策者選擇。



圖 2 政策敘事分析的基本元素

資料來源：本研究整理繪製。

綜言之，大數據的政策敘事分析是以說故事的方式，說明問題與方案之間的關係，賦與人性化的情節發展，結合資料視覺化技術的呈現方式，提高

政策的可理解性。值得注意的是，面對同一個大數據，不同的分析專家及不同類型的分析，往往帶來不同的篩選機制，造成某些問題得到重視，其他問題則遭到冷落（謝孟達譯，2019：42），因此，如果要讓政策制定者或民眾對問題獲得充分的認知，就要讓資料與故事之間有正確的連結，透過完整的故事述說和嚴謹的因果論證，強化政策主張的說服力。

## 肆、政策敘事分析的案例淺析

簡單來說，以大數據為循證基礎的政策敘事分析乃是以問題解決為取向、以資料分析為佐證、以資訊圖表為輔助、以因果論述為焦點、以方案設計為回應及以政策溝通為目的，在此一基本的敘事架構下，分析者同時結合量化與質化的方法，清晰地陳述問題的情節與政策的內容。此種政策敘事模式傳遞了理性分析與感性訴求，不但更具有說服力，也往往能獲得政策利害關係人的認同，強化政策溝通與傳播的效果。因此，本文除在上一節提出政策敘事分析的基本框架外，本節嘗試以政治學中的選民投票參與為例，粗淺說明如何透過大數據進行政策敘事分析。

### 一、問題陳述

民主國家在選舉權普及化之後，幾乎所有的公民都享有投票的權利，以選票決定執政團隊和政策方向。因此，眾人最關心的問題莫過於：

誰去投票？尤其是政黨和政治人物，最為關心是誰決定執政權的歸屬，以及這些實際上會去投票的公民具有什麼特徵？另一方面，西方的政治學相關實證研究指出，民主國家的投票率持續下滑、男性的政治興趣高於女性、年輕人的投票率偏低等，我國的選民是否也有類似的現象？為了提高人民的政治參與，各國除將公民權的年齡下修，也推動通訊投票、不在籍投票、電子投票等方式，我國是否也有推動相關措施的政策需求？

## 二、資料來源與分析

雖然，上述的問題，透過選前或選後的民意調查可以得知梗概，學術界大多也是利用此一個體資料加以分析和推論，研判選民個人屬性變數與投票行為之間的關連性，然而，民意調查在樣本數規模、母體涵蓋率、抽樣誤差、樣本無反應等限制下，實難以窺其全貌。至於總體層次的投票資料，則受限於選舉應採取秘密投票原則的規範，僅能利用中央選舉委員會（以下簡稱「中選會」）所公布以投票所或行政區（如村里、鄉鎮）為單位的投票率和投票人性別統計等資料得知，但是卻難以掌握個別選民的投票參與情形，以致於「誰去投票？」這個政治參與行為中最基本的問題，一直無法獲得有效的答案（莊文忠，2016）。

有鑑於運用總體或個體資料進行選舉投票行為的分析各有其優缺



點，中選會開創各國之先例，以委外研究方式，利用 2016 年總統與立法委員選舉的選舉人名冊，進行個體資料建置與相關統計分析<sup>1</sup>。此一計畫所建置完成的「選舉人投票資料檔」共有 201,257 筆選舉人資料，變數項目包含：選舉人性別、出生年月日、不同選舉種類投票、戶籍地所屬直轄市、縣（市）別、鄰等資料，資料檔在去識別化後上網公開釋出給社會各界運用，有助於瞭解不同地理區域與個人特徵之選舉人的投票行為（莊文忠，2016）。此一高達 20 萬筆的資料檔，包含選民的性別、年齡及地區等資料，可據以回答「誰去投票？」的問題，並透過敘事分析的方式，提出「推動不在籍投票」的政策主張。基於篇幅的考量，此一階段的描述性分析結果直接以視覺化的資訊圖表呈現。

NATIONAL ACADEMY OF CIVIL SERVICE

### 三、資料視覺化

由圖 3 可看到，我國自總統選舉開放公民直選以來，選民人數由 1,430 萬人左右逐年成長至 1,931 萬人左右，近 25 年來增長 500 萬人左右，不過，歷屆選舉的投票人數並未呈現相同的增長趨勢，除 1996 年選舉的投票人數不到 1,100 萬人外，自 2000 年至 2016 年的選舉，投票人數大約介於 1,250 萬人到 1,350 萬人之間，呈現相對穩定的情形，2020 年的投

<sup>1</sup> 該計畫採取分層多階段集群抽樣的研究設計，以縣市為分層變數，並選擇「選舉人數」作為集群分析的變數，將各縣市的村里分為 2 至 4 個集群，以等比例原則分配各集群的樣本數；第一階段抽樣依據「抽取率和單位大小成比例」(Probability Proportional to size, PPS)原則，除連江縣外，每個縣市抽出 10 至 12 個中選村里；第二階段抽樣是整理中選村里的選舉人名冊，依「等距抽樣」(systematic sampling) 抽出中選選民，建立個體層次的選民投票資料檔，在個人資料去識別化上網公開釋出，供各界下載與使用(莊文忠，2016)。

票人數則比 2016 年選舉增加 200 萬人，若與 2012 年選舉相比，則是增加 100 萬人左右，可以說是參與投票人數有較為明顯成長的一次選舉；投票率則自 2000 年後呈現持續下滑的趨勢，顯示在扣除人口自然增減的因素後，這幾屆總統選舉實際參與投票的選民並無明顯的成長；不去投票的選民人數則由近 300 萬人增加到 650 萬人。至於是投票人口的特徵與變化為何，由於此一資料屬於總體層次的分析，無法直接進行個體層次的推論。

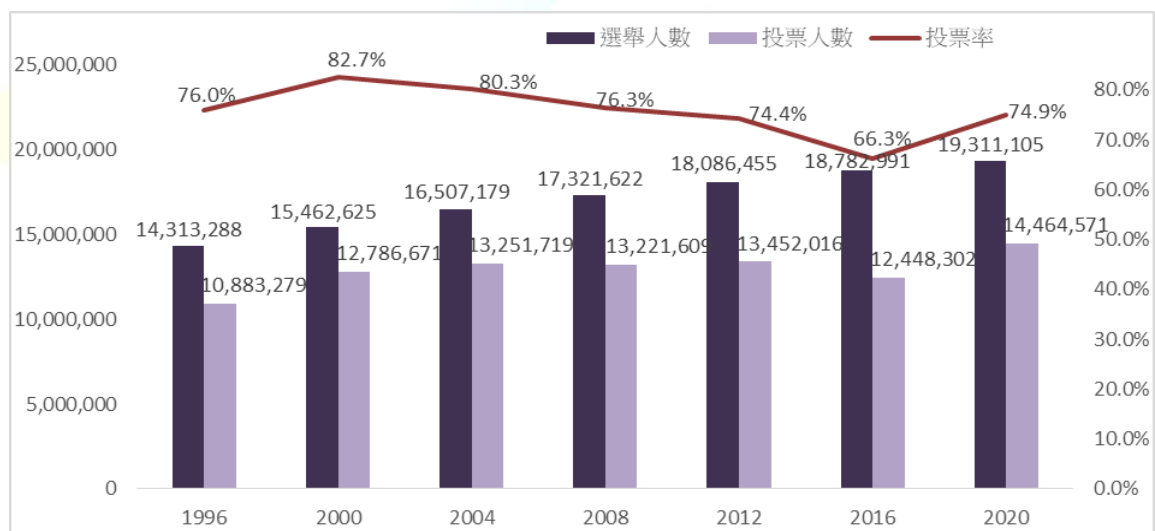


圖 3 歷屆總統選舉的選民人數與投票率

資料來源：本研究整理自中選會選舉資料庫網站。

本文認為可以利用 2016 年總統與立法委員選舉的「選舉人投票資料檔」進一步探討「誰去投票？」這個問題，其分析結果可以看到在 2016 年選舉中的選民投票參與情形有幾個特點：

- (一) 就性別而言，圖 4 顯示在 65 歲之前，女性選民的投票率普遍高於男性，65 歲至 74 歲之間的選民，男性與女性選民的投票率相當接近，不過，大約在 74 歲之後，女性選民投票率的下滑幅度明顯高於男性選民。
- (二) 就年齡而言，由圖 4 可以看到，剛滿 20 歲的選民有超過 6 成的投票率，21 至 39 歲的年輕族群投票率不到 6 成，40 至 49 歲選民的投票率介於 6 成至 7 成之間，51 歲至 74 歲之間的選民，其投票率則隨著年齡由 7 成上升至 8 成左右，75 歲之後的選民，其投票率呈現急劇下降的情形。
- (三) 就各個年齡和性別來看（如圖 4），20 至 29 歲之間的選民，男性與女性的平均投票率差距約為 3.5%，30 至 39 歲之間的選民，男性與女性的平均投票率差距不到 2.5%，40 歲至 49 歲之間的選民，男性與女性的平均投票率差距則是最大，為 5.5%，50 歲至 59 歲之間的選民，男性與女性的平均投票率差距則是縮小至 3.7%，60 歲至 73 歲之間的選民，男性與女性的平均投票率差距則是不到 1%，其後年齡層的兩性平均投票率差距則是逐拉大到超過 10% 以上。

(四) 就各地區的性別投票統計來看(如圖 5)，各縣市的整體投票率高低有別，其中，屬於離島的澎湖縣的投票率僅有 5 成左右，連江縣的投票率不到 4 成 5，金門縣的投票率更是不到 3 成 5；各縣市的性別投票比率的差異亦有所不同，其中，僅有苗栗縣、嘉義縣和金門縣的男性投票率略高於女性，其他直轄市、縣市均為女性高於男性，尤其是臺東縣、花蓮縣和臺中市的女性投票率比男性高出 5% 以上。

(五) 就各地區的年齡投票統計來看(如圖 6)，各縣市在各個年齡層的投票率呈現相同現象，選民的年齡層愈長，其投票率愈高；除金門縣、花蓮縣、臺東縣、澎湖縣和嘉義縣外，其他縣市的 50 至 59 歲和 60 歲以上選民的投票率差距不大；除南投縣、花蓮縣、高雄市、臺東縣、澎湖縣和連江縣外，其他縣市的 20 至 29 歲和 30 至 39 歲選民的投票率差距亦不大。

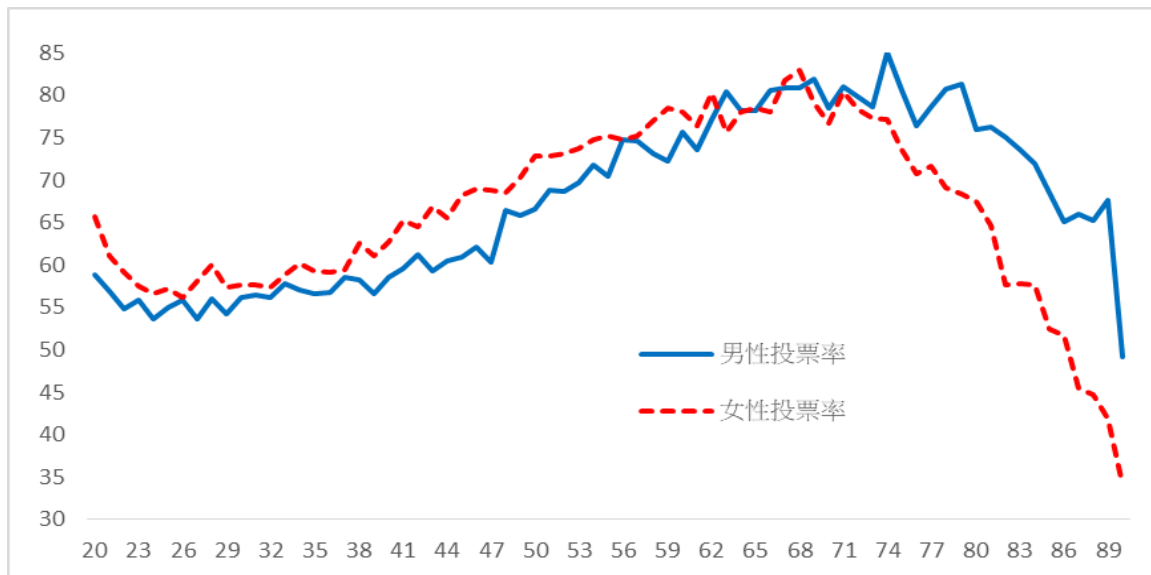


圖4 2016年總統選舉之性別與年齡投票統計 註：x軸為年齡，y軸投票率  
資料來源：莊文忠（2016）。

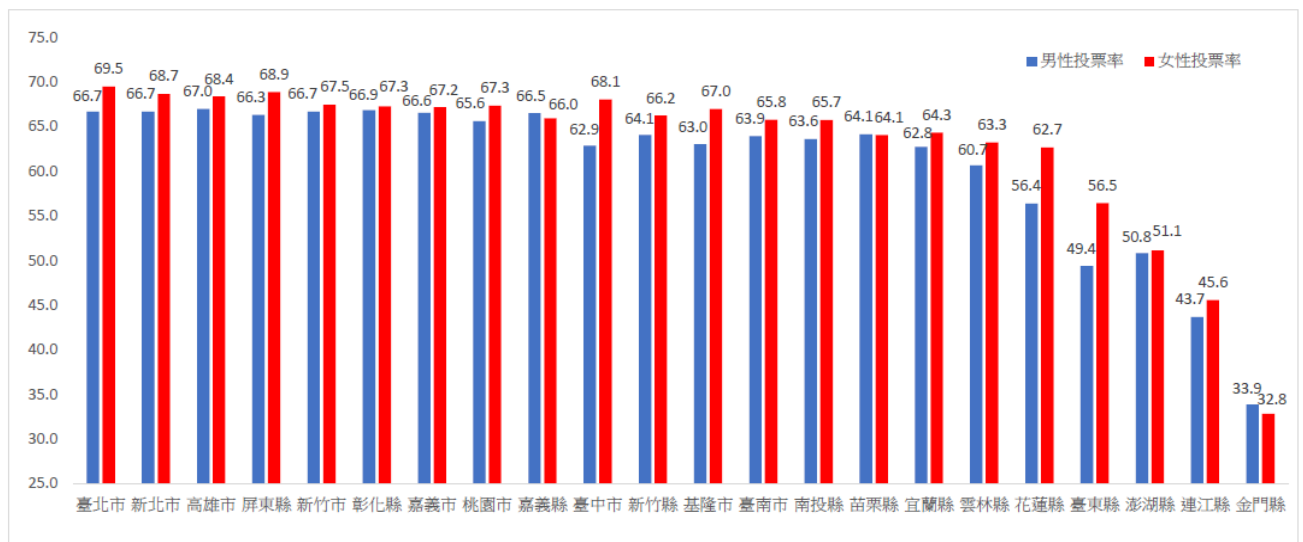


圖5 2016年總統選舉各縣市之性別投票統計  
資料來源：莊文忠（2016）。



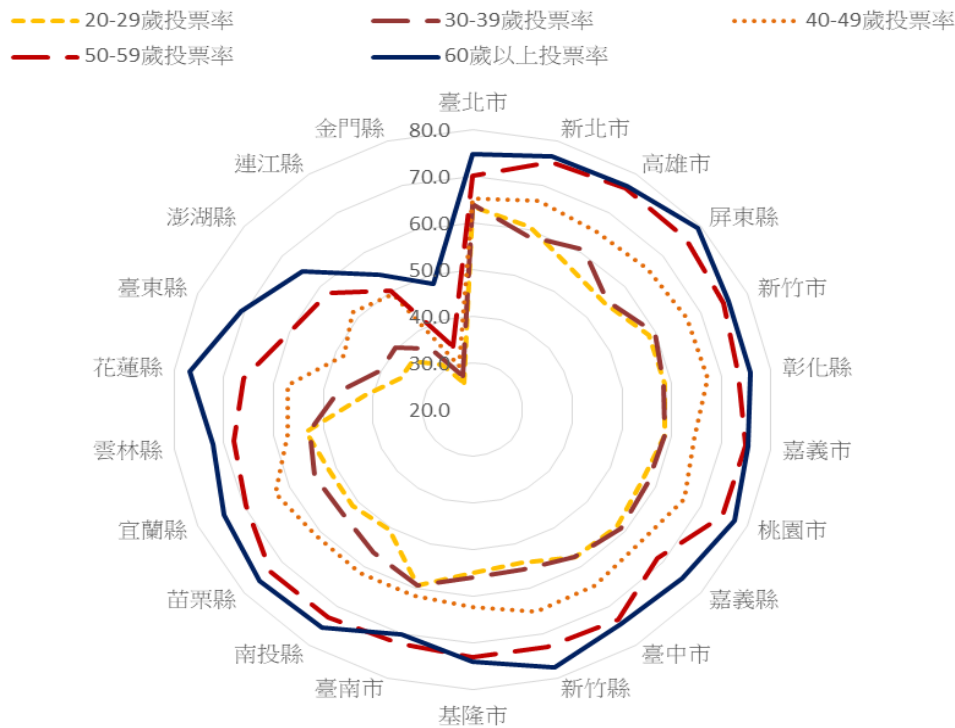


圖 6 2016 年總統選舉各縣市之年齡投票統計

資料來源：本研究整理繪製。

NATIONAL ACADEMY OF CIVIL SERVICE

#### 四、敘事分析

民主政治強調政治平等的原則，在選舉中反映在「一人一票，票票等值」上，然而，從上述的資料分析結果顯示，選民的投票參與在性別、年齡及地區等人口特徵上均出現差異。

首先，就性別的觀點來看，根據「臺灣選舉與主化調查」(TEDS) 在 2016 年總統選後的面訪調查顯示(見表 1)，就性別而言，男性選民的政治興趣高於女性，理論上，男性選民的投票參與應該會高於女性，然而，實際的投票情形並不完全一致，65 歲以下的選民，女性的投票參與比例高於男性，而 74 歲以上的選民，其結果正好相反，如果男性與女

性的投票成本一樣，那麼，是什麼因素導致具有勞動力的人口中，男性的投票率低於女性？其次，就年齡而言，TEDS 的調查結果顯示，除 40 至 49 歲選民的政治興趣略低外，其他年齡層選民的政治興趣差距並不大；然而，實際的投票情形則是，75 歲以下的選民，無論是男性或女性，其投票率均隨著年齡層而增加，那麼，又是什麼因素導致年齡層愈高的選民，其投票率愈高？最後，就地區而言，TEDS 的調查結果顯示，北部地區的選民有較高的政治興趣，其次是高屏和花東外島地區的選民，而雲嘉南地區選民的政治興趣為最低，然而，實際的投票情形則是，花東外島地區選民的投票率低於其他縣市甚多，其影響因素又是為何？

表 1 不同人口特徵與政治興趣交叉分析

		政治興趣			
		有興趣	沒有興趣	無反應	樣本數
全體		42.8%	56.3%	0.9%	1690
性別	男性	48.7%	50.6%	0.7%	832
	女性	37.1%	61.8%	1.1%	856
年齡	20 至 29 歲	43.9%	55.7%	0.3%	287
	30 至 39 歲	42.5%	57.5%	0.0%	353
	40 至 49 歲	38.3%	61.3%	0.3%	326
	50 至 59 歲	44.4%	53.7%	1.9%	324
	60 歲及以上	44.5%	53.5%	2.0%	402
地理區域	北北基宜	47.2%	51.7%	1.1%	545
	桃竹苗	40.9%	58.3%	0.8%	254
	中彰投	40.8%	58.3%	0.9%	321
	雲嘉南	36.2%	63.4%	0.4%	246
	高屏澎	43.4%	55.1%	1.5%	272
	花東金馬	45.1%	54.9%	0.0%	51

資料來源：黃紀（2016）。

事實上，若不考慮選民個人的政治理念、政黨認同、對候選人的評價等個人層次的解釋因素，此一投票參與在性別、年齡及地區等面向所呈現的差異，在制度上至少可以找到一個共同的解釋因素：戶籍地投票制度。相較於世界上大多數國家，我國擁有行之有年且相當完善的戶籍登記與管理制度，此不僅是政府便於管理人民，不少的賦稅或福利政策，也是依據各縣市的戶籍登記人口進行資源分配。而在選舉方面，我國選民並非像美國等國家係採取事前登記制，幾乎所有的選舉均是以選民所屬的戶籍地為資格認定標準，如現行《總統副總統選舉罷免法》和《公職人員選舉罷免法》即規定「選舉人，除另有規定外，應於戶籍所在地投票」。

*NATIONAL ACADEMY OF CIVIL SERVICE*

在此一制度設計下，僅有特定選舉和少數特殊身份的選民可以申請「工作地投票」，絕大多數的選民必須回到戶籍所在地行使投票權，而根據陳明敏等人（2017：44-45；182）的研究報告指出：（1）臺灣地區戶籍地與現住地相同（即「籍在人在」）人口比率由民國 80 年的 94.6%逐年下降至 104 年的 76.3%，粗估目前約 540 萬人之戶籍地與現住地不同。（2）各縣市在民國 104 年的「籍在人在」比率最低前 5 名為宜蘭縣（67.3%）、臺東縣、雲林縣、苗栗縣及南投縣（比率皆低於 73%）；「籍在人在」比率最高者為屏東縣（80.9%）、臺北市（83.4%）。（3）青壯人口中「籍在人在」比率最低的縣市包括宜蘭縣（67.7%）、嘉義市（69.8%）、

臺東縣、雲林縣、苗栗縣、南投縣及澎湖縣（比率皆低於 72%）；「籍在人在」比率最高者為屏東縣（80.3%）、臺北市（82.8%）。

雖然在全國戶政系統電腦化後，民眾於國內之遷徙登記，無須先至遷出地戶政事務所辦理遷出登記，僅需向遷入地戶政事務所辦理遷入登記，但多數人除非有需要，否則在遷移時不一定會主動向戶政機關辦理戶籍異動登記，隨著交通的便捷化與都市化發展，造成「人」、「籍」不一致的比例愈來愈高。另外，前述報告也指出，「籍在人不在」的主要因素不外乎求學、就業等，而這正好是以青壯人口為主，也反映在前述農業縣市有較高比例的「籍在人不在」。

至於性別方面的差異，以臺灣目前的教育水平而言，因不同性別的升學差異所造成的「籍在人不在」或許較不明顯，但在就業方面，若鄉村地區的工作機會較少，必須前往都市地區工作，很少是整個家庭直接遷移，因為都市的生活費用較高，通常是具有勞動力的成年人先去找工作，其中可能又以男性居多，尤其是已婚女性，多數因為照顧小孩或家中長輩而難以離開原籍前往都市，男性在都市工作穩定或生活就緒後，才會將其他家庭成員接來同住，此時全家的戶籍也比較有可能會遷入都市，尤其是在小孩到了學齡階段。此外，前文的資料分析指出，40 歲至 59 歲之間的選民，男性與女性的平均投票率差距大於其他年齡層，應是佐證之一。

另外有 2 個可能造成不同性別與年齡之選民的投票率差異較大的因素：一是，根據行政院主計總處處運用戶籍檔、入出國檔、勞健保檔等公務登記資料進行推估，民國 105 年國人赴海外工作人數估計約 72 萬 8 千人，男性占 55.4%，女性占 44.6%；30 至 49 歲者占 48.3%，50 歲以上者居次，占 31.6%<sup>2</sup>。二是，根據內政部警政署的統計資料指出<sup>3</sup>，100 至 105 年的刑案嫌疑犯人數平均每年約 26 萬人，歷年平均約 8 成 2 為男性、以壯年（24 至 39 歲）的犯罪人口率為最高，其次是青年（18 至 23 歲）；由此可知，男性和中壯年是「籍在人不在」的主力，有些選民在投票日可能因為工作、交通、監禁等因素而無法返回戶籍地投票。

最後，就地區來看，花東和外島地區的投票率低於其他縣市甚多，尤其是金門縣和連江縣，推論其主要原因在於，不少在臺灣求學、工作的選民，在選舉期間回戶籍所在地投票的交通時間和經濟成本甚高，因而有較高比例的選民選擇放棄行使公民的權利。此外，由於金門縣的社會福利較好、以及政府提供離島補助，因而吸引不少臺灣民眾爭相將戶口遷到金門縣，這些選民在選舉期間絕大多數不會回到金門縣投票，導致金門縣的投票率遠低於其他縣市。

<sup>2</sup> 資料來源：<https://www.dgbas.gov.tw/ct.asp?xItem=42426&ctNode=5624&mp=1>（2020/03/30 查詢）

<sup>3</sup> 資料來源：<https://www.npa.gov.tw/NPAGip/wSite/lp?ctNode=12768&CtUnit=2561&BaseDSD=7>（2020/03/30 查詢）



## 五、提出方案

在現代社會中，民眾因為種種因素而必須在不同的地區遷徙或移動，也因為交通運輸的便利性而擴大工作或生活的範圍，選舉時參與投票僅是偶而為之的一次性活動，因此，對公民責任感或政治興趣不高的公民而言，投票的便利性可能成為重要的考量因素之一，如根據 TEDS 在 2016 年總統選後面訪調查結果亦發現，受訪者回答沒有去投票的主要理由有二：「上班工作」（占 31%）和「沒空、有事」（占 19%）。結合基於上述的分析結果可知，我國在戶籍地投票制度未變革之前，選民的投票率可能隨著「籍在人不在」的比例增加而下降，其中當然也可能摻雜了選民的政治興趣或政治信任下降、民主趨於成熟或政黨趨同等因素，而導致其投票參與意願下滑，但這還需要其他相關資料的佐證。因此，基於保障公民的參政權和擴大政治參與，「不在籍投票」或「缺席投票」（Absentee Voting）便是一個值得推動的政策方向，雖然各國因應國情與選民接受度而實施不同的不在籍投票方式，但主要可分為通訊投票、代理投票、事先投票、設置特別投票所投票、移轉投票、電子投票等類型，讓選民可以用此一替代的投票方式行使公民權。

## 伍、結語

無論是實證主義或後實證主義的傳統，均強調研究者的客觀角色，透過文獻檢閱和運用演繹方法，提出與政策相關的命題或假設，經由適當的研究設計，蒐集相關的資料後，再利用統計分析技術進行假設的檢證，以回答理論或實務上的提問，並透過客觀的研究態度，確保研究結果的可複證性。另一方面，基於詮釋主義和批判主義的傳統，則是強調研究者的主體性地位，重視利用研究者個人所擁有的知識和經驗來詮釋和解釋政策相關的現象與行為，其目的不在於理論假設的建構和驗證，而是針對研究個案的深入探索與理解，進而有所反思與檢討，其所依賴的資料並不限於量化的數字，也可以是質性的文本，通過歸納方法找到證據和通則。大數據的政策敘事分析則可擷取兩者之所長，在本土的政策背景和特定的脈絡情境下，結合資料與故事，用以分析和解釋複雜的公共問題和政策現象，進而提出政策主張。

誠如 Kettl 所言，資料向來是闡述公共政策中重點故事所不可或缺，自 1980 年代起，由於網路、電腦及行動裝置的興起，闡述的管道變得五花八門。如今，資料不僅變得更多，也產生出新型態的資料，再加上新興的分析技術及視覺呈現工具，讓說故事時如虎添翼，也更能讓資料說話（謝孟達譯，2019：61）。就技術上來說，大數據分析乃是結合一系列新興資訊科技的應用，從基礎建設、資料蒐集、資料傳輸、資料儲存、資料處理到資料分析，隨時隨地

將人類行為、社會活動或自然現象加以記錄和分析；而就概念上來說，在大數據時代，運用說故事方式呈現資料的方式勢必愈來愈普遍，透過分析人員明確解說數字背後所隱含的重要故事，在解析和詮釋數據所呈現的意義，同時考慮到社會文化的脈絡情境，讓閱聽眾得以理解和感同身受，大數據的政策敘事分析可說是科技與人文的結合與應用。

政府部門面對此一趨勢，透過大量資料的蒐集、整合及分析，針對服務個案的特質與需求提供客製化與人性化的服務，也有助於組織發掘過去未被發現的問題或是調整修正政策的方向，提升解決問題的能力。然而，現今的資料分析與闡述即使出現新型的策略與技巧，背後依然存在基本的一串問題，諸如資料所言可信嗎？如何用別人能夠懂的方式，使其曉得資料的意義，進而知道該採取什麼行動？如何藉由資料來提升政府效能和民主品質？這些都是用資料說故事時的關鍵難題（謝孟達譯，2019：61）。因此，誠如 Shanahan 等人（2018：173）所言，敘事是政治的生命力（Narrative are the lifeblood of politics），政治人物、政治策略家和媒體記者等，都是直觀地理解問題，以致於故事的呈現方式與採取的行動，對政策成功與否和政治生命延續是一樣重要。要言之，大數據的政策敘事分析，不僅僅是發掘數據背後所隱藏的意義和價值，分析者更需要善用敘事分析的技巧，才能用資料說好故事，並據以提出相對應的政策主張，回應民意的期待和民主的課責。

## 參考文獻

### 一、中文部分

阮明淑 (2019)。資料敘事分析之概念性框架初探。《圖書館學與資訊科學》，45 (1)，97-125。

李長潔 (2017)。為什麼視覺化很重要，<http://shucidi.strikingly.com/blog/6b2ecd8b1b2> (2020/04/20 下載)

莊文忠 (2009)，**績效管理與衡量：指標設計的議題與實務**。初版，臺北市：智勝。

莊文忠 (2016)。選舉人性別投票統計改良之研究。中央選舉委員會委託研究案。

陳明敏、陳玟伶、李長明、賴湘如 (2017)。我國戶籍遷徙登記制度之研析——人籍不合一之問題與策進。內政部戶政司自行研究報告。

黃紀 (2016)。2012 年至 2016 年「選舉與民主化調查」四年期研究規劃 (4/4)：2016 年總統與立法委員選舉面訪案。科技部補助專題研究計畫，計畫編號：MOST 101-2420-H-004-034-MY4。

### 二、英文部分

Dykes, B. (2016). *Data Storytelling: The Essential Data Science Skill Everyone*

---

*Needs*. Retrieved from <https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/#79214a7952ad>

Dykes, B. (2017). *Adventures in Data Storytelling: Three Key Traps to Avoid*.

Retrieved from <https://www.forbes.com/sites/brentdykes/2017/04/25/adventures-in-data-storytelling-three-key-traps-to-avoid/#542684f32343>

Kosara, R., & Mackinlay, J. (2013). “Storytelling: The Next Step for

Visualization.” *Computer*, 46(5): 44-50. doi:10.1109/MC.2013.36

Kreuter, M. W., Green, M. C., Cappella, J. N., Slater, M. D., Wise, M. E., Storey,

D.,...Woolley, S. (2007). “Narrative Communication in Cancer Prevention and Control: A Framework to Guide Research and Application.” *Annals of Behavioral Medicine*, 33(3): 221-235. doi:10.1080/08836610701357922.

Lorenz, M. (2010). *Data-driven Journalism: What is there to learn?* Retrieved from <http://ij7.innovationjournalism.org/>

Simons, H. W., Berkowitz, N. N., & Moyer, R. J. (1970). “Similarity, credibility, and attitude change: A review and a theory.” *Psychological Bulletin*, 73(1): 1-16. doi:10.1037/h0028429

Stolper, C. D., Lee, B., Riche, N. H., & Stasko, J. (2016). *Emerging and Recurring*

*Data-driven Storytelling Techniques: Analysis of a Curated Collection of*

*Recent Stories*. Retrieved from <https://www.microsoft.com/en-us>

[research/publication/emerging-and-recurring-data-driven-storytelling-techniques-analysis-of-a-curated-collection-of-recent-stories/](https://www.microsoft.com/en-us/research/publication/emerging-and-recurring-data-driven-storytelling-techniques-analysis-of-a-curated-collection-of-recent-stories/)

Shanahn, E. A., M. D. Jones, M. K. McBeth, and C. M. Radaelli, (2018). “The



Narrative Policy Framework.” In Weible, C. M. and P. A. Sabatier (eds.),  
*Theories of the Policy Process*, 4<sup>th</sup> ed., London: Routledge. pp.173-213.



*NATIONAL ACADEMY OF CIVIL SERVICE*

國家文官學院