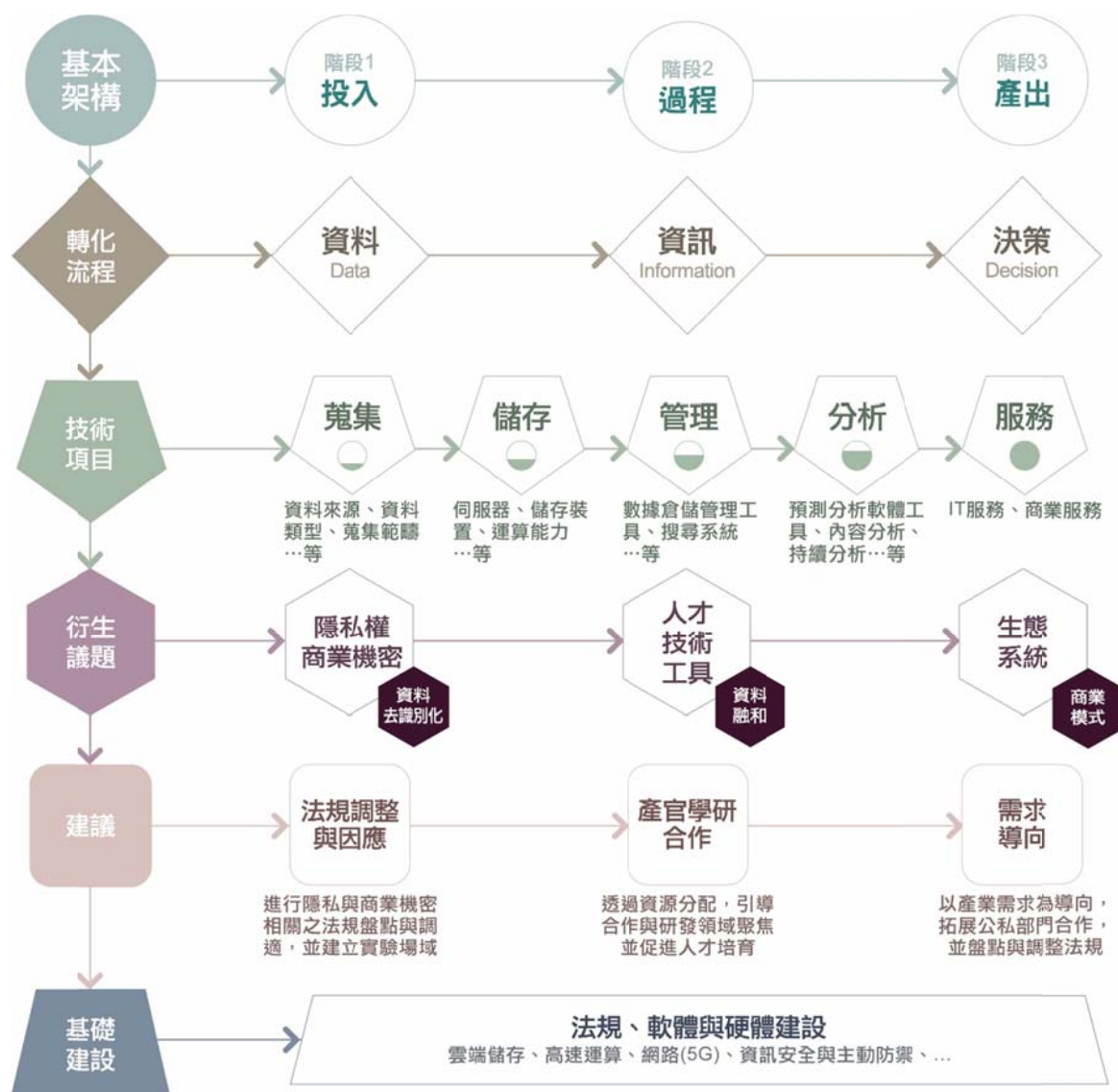


## 大數據分析的未來發展

李國安 / 2017-06-22

大數據分析可視為一種由資料蒐集、整合與分析進而萃取真知灼見(Insight)的轉化過程，主要可分為三個階段，其中所涵蓋的技術可分為五個領域，分別是蒐集、儲存、管理、分析、服務等。每一個領域都有其對應的技術項目，以蒐集為例，其技術項目為如何從不同的來源蒐集不同類型的資料，以及確認蒐集資料的範疇等，而必須面對的議題則是資料中的隱私權、商業機密洩漏之疑慮等，雖然目前的對應解決方案為資料去識別化，但是未來應可進一步透過法規調整與建立實驗場域因應上述的法規議題。而政府做為科技創新發展的領頭羊，除了聚焦於促進相關技術發展外，亦可透過強化基礎建設以達到事半功倍的效果。大數據分析之整體架構如下圖一所示。



圖一：大數據分析之基本架構

## 一、前言

每一天我們啟動手機、ipad、電腦或其他連網裝置的同時，也都創造了對應的資料流量，產生了各式各樣的資料類型，其中更隱含了點閱習慣與偏好、購買行為與決策等。在產品的製造過程中，各種機器設備的運作也都存在著大量的可蒐集數據，同樣也隱含了各式各樣的商業機密。意即一個產品從設計開始，經過生產與消費者使用，直至被摒棄不用，整個生命週期的軌跡都產生各式各樣的資料，也隱含著各種尚未被發掘的真相，並涉及商業機密與隱私權議題。大數據分析之目標即在於從上述的例子中找到真知灼見，而這也只是大數據的起始，過去就已經存在著各種資訊蒐集方式與分析模組，科學家也總希望藉由科學方法從過去的歷史中去預測未來，或者從中找到尚未被發現的獨特見解或事實真相，以作為決策制定的參考基礎。

目前大數據分析已經有相當多的成功案例與經驗，例如：亞馬遜(Amazon)正是依照客戶瀏覽的歷史，來比對「產品」和「產品」的關連性，開發出自動推薦系統，平均每三筆訂單，就有一筆是來自電腦推薦和客製化系統。另一個著名的例子是快遞公司優必速(UPS)，UPS 透過每台貨車的無線電設備和 GPS，精確的鎖定車輛位置和軌跡，並從累積下來無數筆的行車路徑，找出最佳行車路線。透過大數據分析，UPS 發現十字路口最易發生意外、紅綠燈最浪費時間，只要減少通過十字路口次數，就能省油、提高安全。靠著資料分析，UPS 一年送貨里程大幅減少 4,800 公里，等於省下 300 萬加侖的油料及減少 3 萬噸二氧化碳，安全性和效率也提高了。

麥肯錫全球研究所(Mckinsey Global Institute) 在 2011 年研究指出，僅 2010 年資料量即增加了 7,000 PB(Petabyte)，相當於美國國會圖書館典藏資料的 4,000 倍。隨著數位平台、感測器以及行動裝置產生的資訊，也讓全球資料量每 3 年翻倍成長。根據 IDC 的統計，2020 年時全球總資料量將到達 40 Zettabyte(ZB)，相當於約 43 兆 Gigabyte (GB)。而各種新興技術與相關基礎建設的發展加速了大數據分析的進展，但其中仍有若干議題有待進一步解決。

因此，大數據從過去探討的 3V(即 Volume，資料量；Velocity，資料輸入輸出的速度；Variety，多樣性)進入到分析和預測階段，意即要從資料中提煉出真實性(Veracity)和價值(Value)，進而形成 5V，也就是近期興起的資料科學，而目前的熱門領域多聚焦於機器學習與深度學習的發展。此外，根據 IDC 的研究指出，全球大數據與商業分析(Big Data Analytics, BDA)的市場規模在 2016 年已達到 1.3 千億美元，預估到了 2020 年將超過 2 千億美元，BDA 市場規模的年成長率為 11.7%。大數據分析的應用範疇，以及對產業的影響也都與日俱增，對日常生活的影響也益發明顯。

## 二、大數據的發展趨勢

Gartner 的技術成熟曲線報告提出資料科學的發展現況，並評估企業目前的能力與技術採用前景。在 Gartner 的研究中，大數據分析的相關技術多已被歸入資料科學中，在「2016 年資料科學的技術成熟曲線」(Hype Cycle for Data Science, 2016)的研究報告中，提出許多資料科學的相關技術，詳如下表一。其中，公民的資料科學(Citizen Data Science)是尚處於創新始動期(Innovation Trigger)的新興技術；機器學習則處於過度期望的高峰期(Peak of Inflated Expectations)，但前述兩項技術都具備顛覆性效益。整體學習(Ensemble Learning)目前處於穩定攀升的光明期(Slope of Enlightenment)階段，而 R 軟體則是步入實質生產高峰期的成熟技術，且兩者都具備高度效益。以下將分別描述這四種技術的發展與影響。

公民的資料科學是新世代使用者分析體驗的基礎，讓不具備專業分析技巧的使用者也能執行複雜的分析，預計將帶來重大的商業影響。此種資料科學讓先進分析技術所提供的遠見更容易在企業內存取與普及，創造新的商業價值並強化競爭力，這項資料科學將是未來十年促進分析技術採用的重要驅動力，並降低企業對資料科學家的需求。

機器學習是一種技術學科，目的於從一系列的觀測值中，萃取出知識/模式。根據觀測值的類型，可區分出三種型式，第一種是監督式學習，其次是非監督式學習，第三是增強式學習(Reinforcement Learning)。機器學習所具備的效益是指當問題變得更加複雜，即使是最聰明的工程師也可能無法有效掌控時，而機器學習卻可在各種商業和社會情境中，改善商業問題並提出新的解決方案。

整體學習是機器學習演算法的一種，透過建立一組預測模式，並將所有的預測結果整合成單一的整體結果。每個預測分析的案例和多數的機器學習任務都能從整體學習的技術中獲益，而成功的應用案例將持續促進整體學習的發展，並提高預測的精確度。此外，當商業套裝軟體價格過高的時候，R 軟體不僅是一個替代選項，越來越多先進的分析平台(例如：SAS、IBM、SPSS、SAP 與 KNIME)，以及商業分析平台(例如：Tableau Software、TIBCO Spotfire 與 Teradata)也都運用 R 來提升平台的靈活性。

從大數據的發展歷程來看，大數據分析已從過去的資料分析進階到發掘隱含的模式、未知的相關性、市場趨勢、顧客偏好和其他的商業資訊。這些分析的結果可以產生更有效的行銷、新的商機、更好的顧客服務、改善營運效率、競爭優勢等等。而根據 IDC 的研究，大數據分析主要可分為三部份，分別是硬體、軟體與服務等，硬體主要是管理網路資源的伺服器 and 用以存取資訊的儲存

裝置，軟體的部份則多以分析相關的工具與技術為主，服務是指 IT 服務與商業服務等兩類，IT 服務涵蓋 IT 諮詢、系統和網路的執行、IT 外包等，商業服務則是指與大數據分析之軟硬體相關的商業諮詢和外包等。詳如下表二。

表一：資料科學重點技術概觀

效益/預計採用時間	2年	2-5年	5-10年
顛覆性效益		<ul style="list-style-type: none"> <li>◆ 公民資料科學(Citizen Data Science)</li> <li>◆ 機器學習</li> </ul>	<ul style="list-style-type: none"> <li>◆ 演算法市場</li> <li>◆ 認知運算</li> <li>◆ 深度神經(DNN)</li> <li>◆ 深度強化學習</li> <li>◆ 事件串流處理</li> <li>◆ 機率管理</li> <li>◆ 智慧資料探勘</li> </ul>
高度效益	<ul style="list-style-type: none"> <li>◆ 整體學習(Ensemble Learning)</li> <li>◆ R軟體</li> </ul>	<ul style="list-style-type: none"> <li>◆ 自然語言問答</li> <li>◆ 預測分析</li> <li>◆ 自我服務式數據準備(Self-service Data Preparation)</li> <li>◆ 文本分析</li> </ul>	<ul style="list-style-type: none"> <li>◆ 邊際分析(Edge Analytics)</li> <li>◆ 圖表分析</li> <li>◆ 最佳化</li> <li>◆ 規定性分析(Prescriptive Analytics)</li> <li>◆ Spark</li> <li>◆ 語意分析</li> </ul>
中度效益	◆ 影音/影像分析		

資料來源:Gartner，科技發展觀測平台整理

表二：大數據分析相關技術分類

類別	技術項目
硬體	伺服器與儲存
軟體	先進與預測分析軟體工具
	認知軟體平台
	內容分析
	持續分析(Continuous analytics)
	顧客關係管理之分析應用程式
	資料倉儲生成工具
	終端使用者之詢問、報告與分析工具
	企業績效管理應用程式
	無相關的分析資料儲存(Non-relational analytic data store)
	生產規劃的分析應用
	相關的數據倉儲管理工具
	搜尋系統
	服務營運的分析應用
	空間資訊分析工具
	供應鏈分析的應用程式
	勞動力分析的應用程式
服務	IT 服務
	商業服務

資料來源:IDC，科技發展觀測平台整理

大數據分析的相關技術多以演算法為基礎，例如：目前被高度關注的機器學習、深度學習等。而大數據分析的實際應用範疇也相當廣泛，根據 IDC 的研究指出，大數據分析目前可被應用於金融保險、電信、零售、政府、健康照護、製造業等，大數據未來在各產業的應用情境與案例，詳如下表三所示。

表三：大數據分析之應用案例

 <p><b>金融/保險</b></p> <ul style="list-style-type: none"> <li>• 詐騙偵測與預防</li> <li>• 客戶全面檢視</li> <li>• 目標客戶行銷</li> <li>• 風險管理</li> <li>• 資料貨幣化</li> </ul>	 <p><b>電信</b></p> <ul style="list-style-type: none"> <li>• 儲值客戶的銷售和儲值</li> <li>• 客戶全面檢視</li> <li>• 目標客戶行銷</li> <li>• 資料貨幣化</li> </ul>	 <p><b>零售</b></p> <ul style="list-style-type: none"> <li>• 存貨管理</li> <li>• 目標客戶行銷</li> <li>• 下一個精選優惠</li> </ul>	 <p><b>政府</b></p> <ul style="list-style-type: none"> <li>• 交通壅塞</li> <li>• 即時災害回應</li> </ul>	 <p><b>健康照護</b></p> <ul style="list-style-type: none"> <li>• 病患管理</li> <li>• 醫藥研究與發掘</li> </ul>	 <p><b>製造</b></p> <ul style="list-style-type: none"> <li>• 供應鏈最佳化</li> <li>• 收益管理與生產品質</li> </ul>
--	--	--	---	--	--

資料來源：IDC，科技發展觀測平台整理

綜合上述可知，大數據分析的主要應用模式可分為三個階段，第一為蒐集現有的顧客、營運、設備運轉等相關資料，第二為運用演算法程式或模組進行資訊萃取或模式辨認，最後再由相關領域專家的判讀，提出真知灼見以作為調整企業營運的預測分析或決策輔助的主要策略，包含確認目標客戶、設定行銷策略、改善營運等等。

### 三、國際政策動向

運用大數據分析能讓產業獲取各種有利的效益與競爭優勢，並促進產業升級或開發新領域。因此，各主要國家紛紛推出相關的大數據創新政策，希冀能進一步支援或將相關的技術導入實際應用中，各國提出的大數據相關創新政策可略分為開放政府資料、人才培育、基礎建設等幾個方向，其目標多圍繞於改善公共服務、促進產業發展、科學研究或發展大數據分析之生態體系等，各主要國家之大數據政策與目標，詳如下表四所示。



表四：各主要國家之大數據政策方案

國家	政策方案	政策目標
美國	大數據研究和發展方案 (Big Data Research and Development Initiative, 2012)	強調美國需發展能有效分析並管理大量複雜資訊的技術，以促進科技的發展。欲藉由提升大數據工具及技術，以改善科學發明、環境、生物醫學研究、教育、國家安全的能力
	美國大數據研究與發展之策略計畫(The Federal Big Data Research and Development Strategic Plan, 2016)	希冀能透過七個策略的推動，以支持美國大數據創新生態系統的發展，包含提升分析、萃取資訊、制定決策和發掘的能力，以為美國提供新能量，並加速學研發現和創新進展、衍生新的調研領域、培育新世代科學家和工程師，進而促進新的經濟成長。
澳洲	澳洲公共服務之大數據策略(The Australian Public Service Big Data Strategy, 2013)	希望透過推動大數據分析以改善政府的公共服務品質，提升政府效率，並提供更好的政策發展
	澳洲公共服務的大數據實務指南(Australian Public Service Better Practice Guide for Big Data, 2015)	主要用於確認目前資訊管理和分析實務的發展，運用這些實務方法使政府機構能強化服務、改善政策制定，並加以確認澳洲政府利用資訊資產的服務和商機
英國	掌握數據的機會-英國的數據能量策略(Seizing the Data Opportunity – A strategy for UK data capability, 2012)	提出大數據的三個發展方針，分別是加強人才培養及處理資料的能力、強化基礎建設及硬體投資，並從政府部門推動雲端服務、化資料為力量。主要政策目標為：完善大數據政策、發展長期的大數據教育計畫、強化基礎建設
	英國社經研究之資料資源策略(UK Strategy for Data Resources for Social and Economic Research 2013-2018)	目標在於促進和協調資料資源在社會和經濟科學，以及其他領域的發展。確認資料研究之基礎建設的優先領域，藉以滿足未來資料研究的需求
日本	世界最先端 IT 國家創造宣言(2016)	日本於 2013-2020 年將以開放公共資料與大數據為核心，達到「世界最高水準的廣泛運用資訊產業技術之社會」的目標

國家	政策方案	政策目標
南韓	政府 3.0 總體計畫 (Government 3.0 Master Plan, 2013)	目的是收集龐大規模的數據資料，對未來進行預測與展望，進而以客觀科學的角度，擬出解決問題的合理方案與策略。期許應用大數據，提升科學的問題解決能力，並實現低支出、高效能等社會智慧化的理想
	大數據方案(National Big Data Initiative, 2017)	預計於 2017 年底提出新大數據方案，預期將透過設立國家大數據組織，進行管理和分析政府的大數據資料、發展下世代技術，並提供一般企業諮詢和服務，以加速形成韓國的大數據生態體系。
新加坡	資通訊與媒體總體規劃諮詢文件(InfoComm and Media Masterplan, 2014)	新加坡 2025 年資通訊媒體總體規劃的願景與目標，除了建設新加坡為智慧國家之外，還要運用創新的資通訊媒體解決方案提升新加坡人民的生活品質，以及要培養創新型的人才和企業，進而強化全球競爭力，達到永續與高品質的產業成長，以打造新加坡成為亞太地區的大數據分析樞紐中心
中國	促進大數據發展行動綱要(2015)	主要的發展目標有五項，(1)2017 年底前形成跨部門資料資源分享共用格局；(2)充分運用巨量資料，提升各領域資料資源的獲取和利用能力；(3)圍繞服務型政府建設，在各領域全面推廣巨量資料應用；(4) 2018 年底前建成國家政府資料統一開放平台；(5)推動巨量資料與雲端計算、物聯網、行動網路等新一代資訊技術融合發展。

#### 四、創新應用案例

依據 OECD 的研究指出，資料分析的價值創造機制可分為獲取真知灼見與自動化決策兩種，而獲取真知灼見屬於知識創造的範疇，例如透過資料分析除可了解個體的行為，並能進一步了解、評估和最佳化這些個體之行為的影響；自動化決策則是指透過資料分析，機器或系統可從過去的情況中學習經驗，並以此為基礎進而在資料分析後自動制定決策。

大數據分析的實際案例已在各主要產業中展開，而其應用的產業與範疇如表五所示，為進一步了解大數據分析中的機器學習演算法，本文以 Gartner 與 Frost & Sullivan 所提出的創新案例進行介紹，主要為深度學習與預測分析兩種技

術領域，深度學習又可分為認知與非認知兩種領域，認知領域中的技術類型包含影像、語音、機器翻譯等，非認知領域則以詐騙偵測和推薦系統兩種技術為主；而分析預測技術的創新應用則包含醫療保健、設備維護管理、混合雲基礎設備與行車安全等。

表五：大數據分析之創新應用

技術領域	應用	簡介
深度學習之認知領域 (Cognitive Domain)	影像辨識	2015至2016年間，Microsoft的ResNet(Deep Residual Networks)與Google的GoogLeNet(v4)帶來了傑出的影像辨識系統，在ImageNet的影像分類工作中超越了人類。
	語音辨識	百度的語音的文字轉換服務，現已在類似的工作中勝過人類。
	機器翻譯	Google已經發佈了神經機器翻譯(Google Neural Machine Translation, GNMT)技術，並聲稱相較於過去最先進的機器翻譯，這項新技術帶來顯著的改善。
深度學習之非認知領域 (Noncognitive Domain)	詐騙偵測	PayPal正運用深度學習技術做為阻擋詐騙支付的最先進方法。
	推薦系統	Amazon已將深度學習技術應用於最先進的產品推薦服務。
預測分析 (Predictive Analytics)	醫療保健	美國Quire公司的預測分析演算法可以解析大量的臨床資料，並為醫療服務供應方提供病患行為的預測模型。
	設備維護管理	英國Warwick Analytics公司的預測分析技術提供自動化的異常偵測功能，有助於判斷設備的狀態，並及早解決可能發生的問題。
	混合雲基礎設備	美國Perspica公司運用機器學習演算法，協助辨識混合雲(Hybrid Clouds)基礎設備內的應用程式異常。
	行車安全	美國Omnitracs公司研發的行車系統會針對駕駛人可能遭遇的意外狀況提供預測資訊，尤其是注意力失焦導致車輛失控的情形。

資料來源：Gartner、Frost & Sullivan，科技發展觀測平台整理

## 五、簡析與建議

大數據分析未來的可應用範疇將會涵蓋各主要產業，並成為一種不可或缺的基礎技術，而更進一步的發展即為人工智慧中有關機器學習、深度學習的一部份，並將在物聯網、工業 4.0 等系統中扮演大腦的角色。現階段的發展中，除了資料來源涉及隱私與商業機密，需要調整相關的法規以支持其創新發展外，另一個重點即在於速度，包含資料整合的速度、存取速度、開發相關工具的速度、運算速度等。

因為現今資料成長的速度比以往快上數倍，從資料來源、結構與類型變得更加多元便可知，從過去的電子信箱、瀏覽器、社交媒體等，進階到現在地理空間數據、商業交易數據、感測器等。而在瞬息萬變的商業環境中，即時資料擷取與分析功能的提供，將有助於企業占據優勢地位，而其主要的考驗將會是速度，能即時提供對應資訊與洞見者將獲得相對更大的利益。



在美國科學技術顧問委員會的大數據檢討報告中，提出六項政策建議，其中四項政策建議與修改法律相關。相關政策建議可作為我國推動大數據相關計畫之參考，分述如下：

- 修改消費者隱私保護法案，讓消費者能夠清楚知道個人資訊如何被使用。
- 制定國家資料洩漏法，避免資料外洩造成嚴重的影響。
- 將隱私保護法擴大適用於非美國人民。
- 確保所收集的學生資料只被用於教育目的，保護學生資料免於不當使用或共用。
- 運用技術專業，確保受保護團體不會因大數據分析造成歧視結果。
- 修改電子通訊隱私保護法，確保線上的數位資料有一致的保護標準。

綜上所述，政府推動大數據分析時，除了發展創新技術，同時也得兼顧相對應的法規問題，方能避免因法律障礙扼殺了技術的創新發展。此外，發展創新技術的原動力仍來自於人才，唯有培育與延攬相關人才並提供適合的環境與設備，才能進一步發展創新的大數據。再進一步運用資源分配以促進產學研合作，從產業需求端出發，整合現有的技術與研發能量，擴大產業參與比重，將研發成果導入產業的應用中，將有助於形成大數據分析的生態系統。

**參考資料：**

1. Frost & Sullivan (2016). Innovation in Predictive Analytics - IT, Computing and Communications.
2. Gartner (2013). Market Trends: Asia/Pacific Is Catching Up Fast in Adoption of Business Intelligence and Analytics Services.
3. Gartner (2016). Hype Cycle for Data Science, 2016.
4. Gartner (2016). Data and Analytics Programs Primer for 2017.
5. IDC (2017). Big Data & Analytics and Cognitive /AI.
6. IDC (2017). IDC's Worldwide Semiannual Big Data and Analytics Spending Guide Taxonomy, 2017.
7. IDC (2017). Worldwide Big Data Technology and Services Forecast, 2016-2020.
8. OECD (2014). Data-driven Innovation for Growth and Well-being.
9. McKinsey Global Institute (2011). Big data: The next frontier for innovation, competition, and productivity.
10. McKinsey Global Institute (2016). The Age of Analytics: Competing in a Data-Driven world.
11. The President's Council of Advisors on Science and Technology, USA (2014). Report to the President Big Data and Privacy: A Technology Perspective.