

# RDA研究資料聯盟第二次大會出國報告

李士傑 · CODATA Taiwan 國際科學與技術資料委員會 · 中華民國委員會

## 摘要

本次代表 CODATA Taiwan 國際科學與技術資料委員會中華民國委員會，參加研究資料聯盟 ( Research Data Alliance ) 第二屆國際大會，目的是協助 CODATA 台灣、國科會國際合作處 ( 與國科會駐華盛頓科技組代表一同與會 ) 評估研究資料聯盟大會的重要性、對未來國際科學社群研究資料與科學政策的影響、各國代表的觀點，與台灣適宜參與的形式與規模。

透過追溯 RDA 發展歷史，作者釐清了 RDA 所涉及的美國、澳洲與歐盟研究資料相關計畫與資源，以及在國家政策發展中涉及的層級高度與決策重點。大會中的專題演講與工作小組等會內組織的資訊更新，讓與會科學家更進一步理解「分享研究資料」這個議題的難度與重點所在。最後作者闡述對台灣的意義，可以如何進行，幫助台灣透過關鍵資料參與國際科學社群。作者在結語提出了對國科會的建議：協助國科會建立研究資料共享的短中長期政策，鼓勵提出計畫參與 RDA 國際社群，發展國內的研究資料共享社群同時扶植國內的科學資料基礎建設，鼓勵大學與研究機構以具有國際知名度的研究資料來進行規劃，以善用國際舞台創造台灣的影响力。

## 背景資料

RDA ( Research Data Alliance, 研究資料聯盟 ) 是美國國科會網路基礎建設 ( cyberinfrastructure )、歐盟研究資料基礎建設國際合作計畫 ( iCordi ) 與澳洲國科會資料服務部門 ( ANDS, Australia National Data Service ) 2012年八月所共同發起的國際組織。RDA 的前身，DAITF ( Data Access and Interoperability Task Force, 資料存取與共通性工作小組，根據歐盟「科學資料高階專家小組」報告所成立 ) 與 DWF ( Data Web Forum, 資料網論壇 )，乃是 2011– 2012 年之間由 NSF、NIST、白宮總統辦公室與歐洲議會 ( European Commission ) 與澳洲政府研究與教育計畫等資源支持的各項計畫，共同匯聚而成的。( 包括：EUDAT 與 OpenAIRE，等計畫。可參考其關係示意圖 )

這些不同的計畫經過溝通、協調與整合之後，形成一個共通的框架，稱之為「研究資料聯盟」( Research Data Alliance )。目標搭建起聯繫的橋樑：

- 與未來的橋樑：資料保存 ( data preservation )

- 研究夥伴間的橋樑，
- 學科間的橋樑，
- 區域間的橋樑，
- 為了解決新的問題/挑戰，整合的橋樑
- 社群間的橋樑。

RDA 的角色是連結資料與人。其創始的資助單位包括：

- 澳洲政府，透過澳洲國家資料服務 ( ANDS ) 來參與，由「國家合作研究基礎策略計畫」、「教育投資基金 ( EIF )：超級科學計畫」資助
- 歐盟執委會，透過第七屆 Framework Program 的 iCORDI 計畫參與與資助
- 美國政府，透過 RDA/US 參與，由 NSF 國家科學基金會與 NIST 美國國家標準與科技機構資助

## 組織願景與運作原則

其組織願景是，全世界的研究者，無疆界地分享與使用研究資料。其目標界定在加速國際資料導向的創新與發現，透過下列策略來推動：協助研究資料分享與交換、運用與再利用、標準整合，與讓資料更容易被發現。重點的工作方向鎖定在基礎建設、政策、實踐與標準四個方向，與其他的計畫成果。

RDA 確立了核心的指導原則：開放 ( 成員、會議、決策過程與產出開放給社會大眾 )、共識 ( 透過共識解決異議 )、平衡 ( 平衡個別科學家、機構組織、資助者社群間的代表聲音 )、和諧 ( 標準、政策、科技、工具與其他資料基礎架構元素的調和與和諧 )、自發 ( RDA 非政府組織、也非管制單位，而是一個公共組織對其成員透明與負責 )、非營利 ( RDA 不設計、推廣、背書、販賣商業產品、科技與服務 )。

RDA 藉由「工作小組」 ( Working Group ) 機制的設計，鼓勵科學家社群在資料分享實踐與基礎建設上的發展，要求工作小組能提出「短期」 ( 12– 18個月 ) 的工作目標，聚焦在跟特定社群對話、吸引國際參與、以開放自發與共識的方式運作，最終成果能夠被該社群採用。這些設計所帶來的影響，是被採用的標準、基礎建設被建置、政策被採用與執行，在實作上建立資料共享的成功案例。

## 如何參與 RDA ?

RDA 鼓勵提出「工作小組」的申請，必須聚焦在 12– 18 個月內，對資料分享與交換上能夠有可量化的成果。「興趣小組」則是對某個主題提出「個案陳述」 ( case statement )，交由社群與 RDA 大會審議。「社群成員」可以透過網站討論來加入 RDA 組織。「組織會員」則提供經費支援，或者贊助 2 年以上的人力資源來支持核心的 RDA 運作結構。

在今年三月在瑞典哥德堡召開第一次會議之前，RDA 已經透過參與的各國機構，與一系列國際研討會、計畫的推動，讓國際科學社群知悉與認同了他們的存在。在三月之前已經有 12 個「興趣小組」成立。這些先前各個計畫成果的累積與延續，是 RDA 在短短六個月內再次能夠動員科學社群參與的原因。在今年三月份在瑞典召開第一次大會之後，九月在華盛頓特區召開第二次大會。

## RDA 成果資料

### • 組織狀況：

- 2012年八月由美國、歐盟與澳洲的資助機構組成國際指導小組，開始運作。
- 主要運作由 RDA Council ( 委員會，目前有七名委員分別來自美國、法國、微軟、波札那 /IFLA/INASP/ICSU、德國、澳洲、英國 ) 與 Secretariat ( 秘書處，三名工作人員，分別來自美國、德國、澳洲各一位與IT代表 )、RDA Steering Group ( 指導委員會，七名委員分別來自美國 2、英國、芬蘭、澳洲 2、德國 )。
- 工作小組與興趣小組：目前有 8 個工作小組 ( WG, Working Group )，18 個興趣小組 ( IG, Interest Group ) 在運作中，其中 6 個工作小組、13 個興趣小組已經正式通過獲得大會認可 ( recognized & endorsed )
- 協調組織 ( Coordination Groups )：目前有 8 個組織，包括 RDA Council ( 委員會 )、RDA Plenary( 大會 )、RDA Management( 管理小組 )、RDA Organisational Advisory Board ( OAB，組織諮詢委員會 )、RDA Outputs and IP Task Force ( 產出與智財權任務編組 )、RDA Technical Advisory Board ( TAB，技術諮詢委員會 )、RDA Web Platform ( 網站平台 )、Working Group Chairs Group ( WGCG，工作小組召集人群組 )

### • 第一次大會成果

- 2013.3.18– 20 於瑞典哥德堡舉行
- 與會代表 250 人，來自 45 個國家

### • 第二次大會成果

- 2013.9.16– 18 於美國華盛頓特區，國家科學院 ( NAS ) 舉行
- 與會代表 369 人，各大洲比例 ( Australia/NZ 1.6%，Europe 29.0%，North America 65.0%，Asia 3.0%，Africa 1.4% )
- 來自領域的比例：Academia/Research 60.6%，Government/Policy 25.5%，Consultancy and Industry 6.3%，Other 7.6% )

## 第二次大會內容

RDA 第二次大會在華盛頓特區國家科學院 ( NAS, National Academy of Sciences ) 舉行。第一天開幕致詞與專題演講，定調主辦國家在開放資料領域的立場與願景、國家層級的計畫投入，同時也讓最具有前瞻性的研究者提出願景與思辨。

## 定調開放研究資料：重要性、意義，與操作原則

在白宮科學與科技政策辦公室幕僚 Deputy Director for Technology and Innovation [Tom Kaili](#) 的引言，介紹了兩個令人興奮的新計畫資訊：加州大學 [Cancer Genome Warehouse](#) 與紐約大學新成立都市資訊學研究的中心 [CUSP, Center for Urban Science and Progress](#)。Tom Kaili 接著介紹前 Science Commons 執行長 [John Wilbanks](#) 的專題演講：*Open and Machine Readable: Living in The New Default World*。特別值得一提的是，白宮科學與科技政策辦公室( OSTP )今年二月正式發表聲明回應 John Wilbanks 帶頭髮起的請願，將[延伸公共取用涵括聯邦資助的研究計畫](#)。

白宮科學與科技政策辦公室主任 [John Holdren](#) 指示只要有超過 100 萬美元以上研發支出的聯邦機構，必須規劃出版聯邦資助研究成果，免費讓一般民眾得以取用；出版一年內要求研究者將聯邦資助的科學研究之數位資料，提供公眾使用與管理。

Wilbanks 首先標示出最核心的問題：「為何要開放資料，而且是開放機器可判讀資料？」因為就像職棒選手的打擊率一樣，便宜資料 ( cheap data ) 的大量出現改變了知識論 ( epistemology )。開放資料則企圖反映這種知識論的變化。我們所感受到的巨量資料：面對「巨量資料」的浪潮，今日的我們就像是「威力旺卡的巧克力工廠」裡面的小孩一樣，對著巧克力喊著：「給我！給我！我現在就要！」然而資料的開放讓傳統廣告產業崩毀，也讓 Google 藉由廣告基礎建設轉換方向，重新建構起一個新的產業。

如果我們要分享資料，我們必須要規劃「資料集分享的優先性問題」：第一，是計算自己能承諾投入的資源；以 [Amazon Web Services](#) 為例，它們提供了公共資料集 ( public data set ) 不收取資料的額外費用，提供大眾使用。第二，這些資料要考慮到，能夠讓人們適當「引用」( [DataCite](#) )。第三，這些資料分享要付出什麼樣的代價。

所謂開放資料的價值，很重要的意義，乃是：

「... 從廣大與各式各樣的觀眾中，透過沒有過濾的各種貢獻，帶來製造非預期的改變之能力。」

推動開放的工作時，該用什麼樣的模式，來設計網路連結形式，才能一直永續演化下去？三種網絡形式：集中式、( 有序 ) 分散式，以及去中心化的有機網絡。

如果我們一定要決定優先順序，那麼我們該建立起決策的矩陣。[Is It Open Data\(beta\)?](#) 這個網站幫助人們

對資料「有系統地」提出探詢與疑問，尋求資料擁有者做出回應。簡言之，資料被隱私、政策、技術與智財權四個方向規範之。Wilbanks 用四個軸向討論資料開放的向度：accessibility vs. leverage（Y軸），以及 adaptability vs. ease of mastery（X軸）。之前白宮科學與科技政策辦公室發佈歐巴馬總統的「執行命令」（Executive Order），要求政府開放資訊、並且以機器可以判讀的形式開放，就是希望資料能夠容易被取用，同時也能進一步運用（leverage）。以 #NREL 的例子。即便連天氣預測的資料，Wilbanks 所說的「好資料」，都躲在 Google 搜索結果第三頁的 pdf 檔案中，很不容易被找到。機器可判讀的資料能夠急遽地降低使用門檻，增加使用的便利性。

當一切都變成資料的時候，這也意味著「文化將變成資料（Culture turned into data.）」關鍵不在於表面的價值，而是這些資料的「潛在價值」。例如每個人的個人醫療資料的開放。

這些開放過程的合作，意味著新的治理形式（new forms of government）必須要發展出來。同意的方式，以及這些同意資料將如何被管理，有爭議時該怎麼解決？John Wilbanks 所領導的 Portal Legal Consent 提供了線上的「同意研究」[Consent to Research](#) 法律資料服務。

## 跨領域理解開放資料：具體實例與深入分析

第二位專題演講講者是來自人文社會科學領域、伊利諾香檳大學圖館與資訊科學研究所的 Carole L. Palmer 教授，講題是 *Fuelling and Transforming the Evidential Cultures of Research*。

Carole 引述啟發性的種籽想法與實踐：[Understanding Infrastructure: Dynamics, Tensions, and Design — Report of a Workshop on “History & Theory of Infrastructure: Lessons for New Scientific Cyberinfrastructures”](#)，認為「資料作為一種基本上集體的共享資源，對科學研究有非常大的影響。」

學科的依存關係與差異性，在「證據文化」（Evidential cultures）上有所不同。例如有些同樣的地震波研究，在美國與義大利，由於他們在證據上集體主義與個人主義的差異，創造出了開放或封閉的文化，也決定了「何謂值得出版的成果」。在實驗室、或者國際研究社群中，誰負責驗證科學資料的效度與意義？

例如在黃石國家公園的科學研究，包含了地質學家、地質化學家、微生物學家等等，他們提出了從地球生命的源頭到其他星球上的生命形式等等研究問題，這些都依賴科學顯著場址的田野資料。Carole 發展出了「場所中心的資料策展」（site-based data curation）這個觀點，來看待從資料所創造出來、從以往證據個人主義轉向集體主義的轉向。這包括策展過程與政策的改變。

Carole 釐清資料共享跟資料生產者有關，而資料的再利用，則是跟資料消費者有關。各自關心的問題不

一樣：資料共享/生產涉及到複雜資料集與可用的資料部分；而資料重用/消費涉及到所願意釋出的版本，以及最適合重用的版本。

具體的例子如 NASA 美國太空總署的衛星資料，以及海洋模型與雨林的資料。重用資料甚至可以帶來價值：透過重用海洋模型資料來驗證田野資料，或者透過重用雨林資料，運用不同的設備來校正更為精確的雨林氣候溫度資料。

Carole 認為在三種資料實務 ( practice ) 當中，可以重新融入社會文化的面向：在研發當中，建立「快速回報框架」可以讓關鍵的研究成果更快地回饋到基礎架構團隊手上；在工作團隊中，建立 DPCVocab ( 資料實踐與策展詞彙 ) 可以發揮效率與溝通的作用；在資料策展教育上，資料導師 ( data mentors ) 與科學導師 ( science mentors ) 共同影響年輕科學家的實習與養成過程。

Carole 的演講則將人文社科在研究資料實踐與常規科學的「證據文化」 ( evidential cultures ) 上，不容置疑的重要性提出來：唯有深入了解當代 e-Science 的脈絡、實際實踐的細節，才有辦法重新型塑科學研究。

Carole Palmer 的研究架構在幾個合作研究團隊之上，包括資料保存-資料實踐 ( 伊利諾香檳大學，國科會 )、場所中心的資料策展 ( 伊利諾大學、約翰霍普金斯大學、國家公園管理處 )、DCERC 研究中心資料策展教育 ( NCAR，UTK )、策展檔案 ( 普渡大學 )，裡面所基於的問題意識與觀察經驗，是我們的人文社會科學學門很缺乏的。深入的了解也才能夠充分發揮資料在跨學門的影響力與效益。

RDA 藉由專題演講來替三天的研討會定調的努力相當成功。對於來自各地的科學研究者，他們可能各自關心不同的領域主題，但是透過專題演講來建立相互溝通的共同立場、語言與角度，讓後面幾天的合作與互動建立起成功的開始。

相較於三月瑞典的第一次大會，其專題演講勾勒出過去與未來，這次的專題演講更為鎖定當下，實際而具體可行。

## 大會報告：RDA 的各方面進展

接下來的 RDA Business 議程介紹 RDA Council 的新消息、召開第一次會員大會 ( Organisational Assembly )。介紹新成員，以及爭取主辦 Plenary 4 的地區團隊。已經確定第三次大會在由澳洲 ANDS 在愛爾蘭都柏林舉行，第四屆的大會由荷蘭代表組團爭取。透過爭取的過程，大會讓所有與會者都清楚看到提案團隊豐富的資源、對資料分享的整體視野，以及各個機構之間的合作關係。

WG 工作小組更新資訊報告。包括語言、資料範疇資訊進入工作小組，其他如 DFT (資料基礎與詞彙)、資料型態註冊、後設資料標準目錄、後設資料協定、長久識別子資訊型態、實際政策工作小組等都有更新的資訊提供大會報告。

在全體大會報告中，Juan 介紹組織會員 (organizational membership)，並且鼓勵組織加入 RDA。目前年費將從2014年一月開始計算，組織成員在50人以下的，每年要繳 1,000 美元年費；組織成員在 50-250 人範圍的，每年要繳納 2,000 美元年費；而組織成員超過 250 人的大型組織，每年需要繳納 10,000 美元的年費。在大會的集體討論中，許多科學家問了各種角度的問題，而主辦單位則逐一回應，獲得了大家的信任。目前將會推動 RDA 成為英國的一個立案公益組織，在英國正式註冊。這些報名經費、會員費用也都將送到英國的 RDA Council。

會員權益：

- 正式參與 RDA 所有工作小組的成果
- 在 RDA 網站上有正式露出，在 RDA 會議中主導世界資料共通性議題
- 在你自己所屬領域、市場與地理區域的資料共通性議題上代表 RDA 發聲
- 在你自己所屬領域與面臨到資料交換的真實問題時，能夠在 RDA 有建議與顧問的聲音
- 在新進發展的資料共享標準與協定上，有機會扮演 beta 網站的角色
- 能定期收到 RDA 工作進展的簡報

RDA Colloquium 是以贊助機構為主形成的組織，他們有獨立的身分參與 RDA 的決策，並且透過這些次級組織之間的平衡來充分反映在科學資料共享上國際科學社群的生態關係。大會邀請他們發言：主持者為 NIST Acting Director, Cyber Physical Systems and Smart Grid Program Office, Engineering Laboratory 的 Chris Greer。澳洲代表是 Clare McLaughlin, Counsellor, Education and Science, Australian Embassy to Belgium and Luxembourg and Mission to the European Union and NATO in Brussels；歐盟代表是 Kostas Glinos 參與 European Commission, 領導 e-Infrastructure。美國方面由 NSF 的 Mark Suskin Deputy Director, Division of Advanced Cyberinfrastructure (CISE/ACI), US National Science Foundation 代表發言。

這些演講都有線上影音檔案，可以讓線上與會者了解這些支持者與推手的意見。主要的 RDA 參與者檔案在 [Who's Involve?](#)

最後的議程是 TAB (Technical Advisory Board) 介紹。TAB 對 RDA 委員會提供技術專業與建議，也協助發展與評估 RDA 工作小組來提升其影響與效率。這個小組負責 RDA Technical Roadmap 這份文件的發展、維護與演進。其組成方式是分兩階段選出 12 名委員，初始的 6 位委員由 RDA 委員會提名，

接著尋找其他 6 位委員會，在第二次大會時選舉產生。許多傑出的 PI 在這裡扮演重要角色。包括：

- Francoise Genova, Director, Strasboug Astronomical Data Centre CDs, France
- Bill Michener, e-Science
- Beth Plale, Director, Data to Insight Center, Managing Director, Pervasive Technology Institute (PTI), Professor, School of Informatics and Computing, Indiana University
- Andrew Treloar, Director of Technology, Australian National Data Service, Australia
- Peter Wittenburg, The Language Archive, Max Planck Institute for Psycholinguistics, Netherlands

現在在進行中的是[第二階段的 TAB 委員選舉](#)。

## 分組討論：蓬勃發展的工作小組會議

第二天在大會第二場址舉辦各工作小組的 Breakout Session 分組討論議程。各工作小組的會議內容請參考[大會議程](#)。值得特別說明的，大會安排每個會議都有線上的會議服務，遠距的與會者可以透過網路連線參與工作小組工作的進展。

除了工作小組之外，興趣小組往往有新的議題或觀點拋出，透過 case statement 的討論，小組召集人邀請關鍵的科學家報告他們的看法。以歐洲代表為例，值得一提的是邀請與會者在 RDA Europe / ICORDI 所舉辦的研討會、討論活動中，已經有所初步討論，也獲得了初步的回饋與參與的建議，所以到大會的分組討論階段時，這些演講與討論都已經有一些成熟的內容，也非常能夠激發與會者回饋想法。

這是第二次參與工作小組會議的進行，我也更加了解 RDA 的參與團體與組織，如何藉著 RDA 的平台，將自己原本關心的議題在這個平台上展現與研究。與會者相當積極的參與，產出的成果也對於自己原本研究領域、社群有實質的貢獻。研究者自身的資源，例如原本預定 2014 要舉行的研討會，也一同被連結在工作小組的資源中，更有效地讓學術的對話成為永續進行的動力。

我這次參與的是 [Legal Interoperability](#)、[Meta Data](#) 與 [Long Tail of Research Data IG](#)。最後一天參與的是 [Digital Practices in History and Ethnography](#)。完整的工作小組與興趣小組清單請參考：[Working and Interest Groups](#)。

## 台灣可以怎麼參與？

台灣由於以往數位典藏與數位學習國家型科技計畫的成果，對於 [Data Citation](#) 與 [Digital Practices in](#)



**History and Ethnography** 等工作小組都可以有相當實質的貢獻。目前這些工作小組與興趣小組對於多語言、CJVK 等雙位元語言國家的研究資料著墨也有限，如果有這方面的專家投入，會很有影響。

我在會議中與 NIST 美國標準局的研究員 Wo L. Chang 討論到關於 video 標準的問題；目前 NIST 有跟德國的公共電視合作，規劃視訊影音的國際標準。此外他也關心影音資料的開放連結資料議題 ( LOD, Linked Open Data )。數位典藏當年曾經邀請過紐約大學的 Howard Besser 教授來台灣演講影音保存的議題，我們也知道影音典藏 ( video archive ) 其實在資料運作上、在許多層次上都還有很多發展空間。這些當年在數位典藏所開拓出來的領域，國內的研究者如何可以利用這些機會在國際場合中串連更多同好，將相關議題作領先的突破，也許真的對台灣國內可以有更多正面的影響。

台灣迫切需要這樣的跨領域資料分享、重用的討論會議，讓這些潛藏在個別研究者、各個研究領域的資料問題與困難，變成可以被互助合作、共同解決的挑戰。

## 總結討論與分享：回到研究社群

第三天的綜合討論，歐盟執委會 Carlos Morals-Pires 主持所有工作小組/興趣小組的綜合報告。其中包括農業資料、巨量資料分析 ( 透過大氣科學、天文學、生物資訊與語言學來建立使用案例 )、媒合問題IG ( Brokering, 透過情境來串接研究社群 )、資料註冊、溝通能力模型IG ( 建立分析模版工具 )、出版學術與研究資料、資料引用 ( 展示範例, LNEC 無線感測器網絡資料 )、DFT、資料型態註冊 ( 聯盟組織 )、資料密集科學領域之教育與技能發展、深度參與、法律互通性 ( 2014年將於美國國家科學院舉辦工作坊 )、LOD 開放連結資料 ( 建立管理長尾研究資料的最佳範例, 連結到圖書館、文化遺產研究領域, 館藏的使用案例研究 )、長尾研究資料、毒物基因體學共通性與化學安全等工作小組與興趣小組。

在最後的社群討論議程，大會主席 Mark Parsons 成功地整理與會者的共同問題/議題 ( 詳細紀錄有多少人表達對這些問題的關切 )，然後拋出來與大會討論。這些問題對於所有關心科學資料共享、以及國際合作的人來說，都是相當重要的議題；假使有一個國際社群，能夠邀集各國不同領域的優秀科學家關心這些議題、並且讓大家為此做出貢獻，在國際上的影響力將是難以估量的。

1. 如何估算 RDA 的產出？制定標準的數量？成功案例？討論？認證？
  2. 改變學院文化，重視資料引用與其他的非傳統產出
  3. 在不同組織間，類似的工作小組如何互通合作？建立共通性的測試平台，或者以資源、成功案例註冊的方式來為共通性預作準備
  4. 我們該如何透過 RDA 培養未來的資料領導者？
- 應該有早期或生涯中途研究者的 BoF 議程，討論如創新作法嗎？  
( 例如推動 RDA 學生與會或議程，耕耘下世代領袖，在工作小組中加入 hackerthon 黑客松等作法。 )

5. RDA WGs 與 IGs 怎麼看待 IP問題？
6. 如何從社群中找出更多的 input？
7. 失敗的臨床資料，與其他有價值的失敗經驗？
8. 性別平衡議題
9. 區域性平衡，如何讓 RDA 更加全球化？
10. 邀請生物醫學資料分享的計畫加入 RDA
11. 輸入與輸出檔案與相連結資料集的保存 / 典藏議題
12. 如何避免被科技拖著走？
13. High consequence 研究資料與網路安全
14. 如何避免在錯誤的工作小組中討論事情

在大會的社群討論中，英國的 DCC ( Digital Culture Curation ) 獲得了很多與會者的肯定。他們的 roadshow 經驗也被認為可以應用到 RDA 的推廣活動中。主辦者也跟與會者希望有更多的區域 RDA 活動可以進行，在不同層級上促進資料共享的深化與發展。

與會者在此希望藉由 NAS 等知名機構的影響力，團結起來為一些重要的根本議題來作發聲，希望能夠改變學術社群的方向 ( 例如更加重視 Data Citation )。這些文化的改變，例如在論文中包含進去資料、比沒有包含資料，給於更多的肯定，創造一些容易被衡量的回饋方式，讓改變得以發生。

大會在充分的討論與對話、達成共識的過程中，結束了這場豐富的社群會議。跟這麼多關心不同領域的科學家一起開會，要不是透過 CODATA、RDA 的年輕科學家、社群媒體 twitter 上優秀同事的摘要，很難捕捉到這樣的大型會議的面貌。

## 對台灣的啟發

### 共同面對的挑戰

資料的重要性毫無疑問已經是世界各國科學政策單位的共識。然而如何才能夠讓科學家社群 ( scientists community ) 更有效地運用資料、提升對資料的使用與再利用，甚至走向合作分享資料、發現科學研究成果的新典範？科學社群創造力的黑箱，恐怕是科學發展的瓶頸角色與關鍵要素。透過資料合作來促成科學研究合作，在重要的科學發展中不缺席而能夠有獨特的角色，這中間需要很多的實證研究與分析才能夠看清楚全貌。

美國政府在 Cyberinfrastructure 對科學研究的基礎建設研究下功夫不可謂之不深，誠如專題演講者 Carole Palmer 教授所引述，啟發的種籽可追溯至 2007 年的研究便指出基礎建設之諸多複雜張力、動態關係。歐盟則透過高層科學家的專家小組提出研究報告，一路上由歐盟議員領導追索這個問題的解決方向，

最終集結對開放資料的種種努力到 iCORDI / RDA Europe 計畫。澳洲的國家資料服務 ( ANDS ) 也是如此。三個政治實體各自透過最高層的政策方向、結合鼓勵社群的發展與投入資源。今日我們所見 RDA 的熱烈參與場面，只不過是這長期努力的延續而已。透過諸多計畫的匯聚資源、共識與溝通，RDA 計畫讓科學家社群的黑箱得以打開，藉由對共享的要素一一投入資源來推動，RDA 成為看起來在推動科學家沒有疆界分享研究成果的領頭羊，實質上則讓歐盟、美國與澳洲高階的資深科學家領導者，有機會共同形成彈性而有機的網路組織，透過社群、連結起具有創新能量的精彩前瞻研究社群。

## 台灣的可能性：透過關鍵資料參與國際科學社群

台灣年輕科學家應該要有機會可以參與這樣的國際會議；在這樣的會議中，找到對解決同樣資料困難與挑戰的同好。這種經驗是與單打獨鬥截然不同的一種情境；組織、資助機構、國家 扮演了參與與連結的角色，讓科學家在此發展出對研究社群有貢獻的基礎元素，並且加以應用之。我們可以從重新檢討自己的組織的國際角色，開始思考我們要如何面對 RDA 這樣的國際社群計畫。我們可以試想，這些國際組織可以怎麼幫助我們，發展科學計畫、共享與共用/重用研究資料；在這個過程中，可以加速原本研究的發展，並且回到我們自己最擅長的角色與更關注的角度。

舉例來說：台灣的海洋研究，與日本、亞洲鄰近各國的關係為何？空氣汙染的研究模型，對於開發中國家的環境污染研究是否有助益？災難科學研究資料的共享，在台灣長期的研究經驗上，是否可以在這樣的國際場合中變成一個主要的主導議題？福爾摩沙衛星影像是否有其他國家可以利用的資訊，甚至他們的回饋是否能夠幫助我們修正更好的觀測模組？這些資料共享的問題，在全球規模上如何快速更進一步，直接縮短我們的專業者養成過程中，資料專業者的養成時程與對國內/國際社會發揮貢獻的時間？台灣的資料中心如國家高速網路中心，是否有可能搭配領先的科學社群成果，透過 RDA 促成亞太地區的科學合作，發揮台灣的戰略優勢？

RDA 讓各國的與會者看到的，應該是這樣的一個整合的契機：有一筆定位清楚、方向明晰的資源，挹注在科學研究資料共享的主題上，探討如何「縫合」各自努力發展的研究社群，形成綜合的戰力。而且這個整合的努力，不僅跨越了國內研究社群的分野，同時也跨越了海洋與板塊，創造未來的「潛在價值」。附帶的成果，就是讓區域自己的科學家，站在一個未來世代資料科學合作的起點上。

## 結語

「為何要參加 RDA？」RDA 與 RDA 前身的各國國際合作計畫，是美國、歐盟與澳洲主動出擊、獲取重要領域參與合作方研究資料的計畫。透過設計一個橋樑與平台，關注資料保存、國際跨學科的新問題與挑

戰·加速「資料導向的創新與發現」。藉由 (1) 協助研究資料分享與交換、(2) 運用與再利用、(3) 標準整合，與 (4)讓資料更容易被發現等策略，鎖定工作方向在基礎建設、政策、實踐與標準四個向度，創造出短期 (12- 18 個月內) 就能快速發揮戰力的新型態科學社群。

RDA 同時也是連結科學家研究社群合作議題與解決方法的「社交場合」；各國的贊助機構對此投入資源來主辦系列活動，並且藉由組合這些組織、結構的影響，發展自己區域內的科學研究。這是一個以機構、組織為主的國際戰場。台灣可以考慮的參與方式，我建議有下列作法，國際科學與技術資料委員會中華民國委員會 (CODATA Taiwan) 可以就中研院的資源與立場，主動協助國科會來善用這個國際舞台：

1. 協助建立研究資料共享的政策：國科會應收集國內科學研究社群在研究資料分享上的實務經驗，據此建立研究資料共享的短、中、長期政策。而此政策應回答台灣是否要建立國家資料服務？台灣研究資料生產的獨特性？諸角色如資料科學 (data science) 與資料科學家 (data scientist) 是最主要的科學研究資料生產者，科學家、民間與業者作為資料使用者，資料儲存設施等基礎建設的資料營運者等各自角色與資源如何界定，如何彼此相互合作；資料如何維護、管理與永續經營等問題。
2. 投入資源建立國內研究資料共享社群：國科會提出「研究資料共享」的國際合作計畫框架，邀請各機構的研究者提案參與。這個計畫應該包含：核心計畫 (如何運作 RDA Taiwan，在國內舉辦活動，介紹 RDA，使大家有概念；推動台灣內部自己的研究資料共享議題，這樣參與國際大會才會有想法)，與內容計畫 (需指名研究領域、研究社群、國際合作單位、合作模式、研究成果、影響)
3. 邀請大學或研究機構以「組織會員」(organizational membership) 的角色參加，以具有國際能見度的研究資料 (如中研院有環變中心、地震中心、生物多樣性中心與數位典藏、調查研究專題中心、統計所等不同類型的重要研究資料叢集，中興大學近期在環境污染方面被 Nature 多次報導的知名資料等)，提出自己的「工作小組」參與計畫，並對國內的大學研究機構舉辦說明會，介紹 RDA 的成果與影響。
4. 對個別參與 RDA 後續會議 (第三次大會將於 2014 年 3 月在愛爾蘭，第四次大會將於 2014 年九月在荷蘭舉行) 的研究科學家，國科會可以用專案經費補助前往參加的科學家/機構代表，並要求與會研究者參與前項核心計畫/內容計畫成果發表。
5. 扶植國內研究資料基礎建設：國科會鼓勵國科會項下，關注研究資料共享議題的業務單位提出內容計畫共同參與 RDA Taiwan，並從前項的核心與其他內容計畫成果獲益。業務單位應該在單獨的申請計畫中，列舉清楚自己因為參與 RDA 計畫，獲得了什麼樣的成果，對業務內容的具體貢獻。

科學資料的越來越被重視，已經是所有科學機構與政策單位所認知的現實。甚至連 Nature 雜誌從今年十月起，也開始在正式徵集科學資料，在科學資料出版、引用、共享的領域跨出一大步。他們的行動被 RDA

社群密切地關注與影響。國際科學研究社群未來還會有非常大的變動面貌即將浮現。RDA 是一個快速形成、得以讓我們一瞥最新發展的窗口，讓台灣的研究成果能夠更有效率地與國際社群對話、整合。對內台灣未來科學研究的資料政策如何規劃發展，對外如何透過分享更為國際化，RDA 可以扮演一個相當重要的國際橋樑。如果我們不利用最有優勢的台灣科學研究成果，與 CODATA Taiwan 多年累積的國際社群人脈與影響力，規劃下一個世代資料科學 ( data science ) 的科學發展政策，屆時到國際發展局勢底定時，台灣的選擇與發展空間將大受侷限。

## 參考資料

- RDA Europe 簡報：[RDA & iCORDI](#)
- [Open Access](#) 開放取用

出差人簽名/簽章：