

國際科學與技術資料委員會中華民國委員會
104 年度第一次全體委員會暨常務委員會會議紀錄

時 間：民國 104 年 05 月 25 日（星期一）晚間六時三十分

地 點：紅豆小館（台北市南港區忠孝東路七段 369 號 8 樓）

主 席：許聞廉主任委員

記 錄：林照玲

出 席：許聞廉、王新民、林誠謙、羅慧珍、賴進貴、吳金洌、楊盛行、邵廣昭、蕭傳鐙、符文美、李德財、何孟樵、劉陵崗、莊裕澤、孫志鴻、廖啟成、朱有花、陳鴻宜、張淑媛（代 吳俊忠委員）、郭敏玲（代 許稟寧委員）、莊庭瑞（特別邀請）。

請 假：李元希、洪慧芝、趙丰。

會議附件：1. 103 年 05 月 21 日 103 年度第一次全體委員會暨常務委員會會議記錄。
2. SciDataCon 2014,the next CODATA International Conference 報告。
3. CODATA task group on Data Citation Standards and Practice 規劃調查表。。

一、主席報告：

（一）確認 103 年度第一次全體委員會暨常務委員會會議記錄（附件一）。

（二）去年由許聞廉主任委員、李德財校長、王新民秘書長一同參加在印度舉行的 SciDataCon 2014 & CODATA General Assembly，請李德財校長（同時也是 CODATA international 執行委員）口頭報告會議情形（附件二）。

李德財 校長補充：

1) CODATA 所發行的 Data Science Journal 目前已 call for papers，建議大家可以多多投稿。

2) RDA ([Research Data Alliance](#)) 將於 2016/09/11~16 那一週於華盛頓舉行，希望 CODATA 中華民國委員會組織明年也能有相

關議題參加 RDA 國際會議。

二、討論及決議：

(一) 請各位委員分享各部門或學會(國內外)在 open data 上新的進展。

許聞廉委員：

中研院資訊所是研究單位，我們要的 data 多半與政府有關，譬如：smart city、smart home，這些資料目前 open 的方式都是在網站上可以提問，然後會有答案，但無法下載原始資料，這些不符合 open data。而我們 IASL 實驗室有做一個系統是可以打台灣地址或美國地址，這些都全部 open，你打幾個字就可以全部出來。但如果在網站上是比較困難的，事先沒 process 過是很難的。另外就是我們的同仁有跟經濟部詢問過能源方面，譬如台灣用電來源的百分比，但資料都是不公布的，這是很難理解的，現在這麼重要的階段，應該把資料分享出來讓學者研究，如果核電廠現在慢慢取消的話，它的 impact 會在那一年到什麼程度，這一定要 open 了才有可能做研究，不可能叫學者接一個計畫，接了其他人也不會相信，一定是要資料 open 出來所有有興趣的人去研究。現在國科會有個 policy，有一些 open 的 data 希望有人做加值的研究，但是範圍挺小，農業方面比較多，真正比較重要的 data 都沒有 open，像美國政府是主動去推廣，我們政府目前還沒意識到這重要性。

邵廣昭委員：(中研院/生物多樣性研究中心)

目前國科會在 call for proposal，用 big data 鼓勵大家寫計畫書，但感覺是真正 open 出來的 data 很少、也很有限，特別是我關心的生態環境的領域。所以事實上其實鼓勵大家開放 data 是有點難，但是我們在推生物多樣性資料 open 推了很久，國科會生科司生物多樣性學門從 102 年開始要求每個老師在提計畫書的時候不只交報告書而已，同時需同意原始資料要建檔繳交，這件事情已經算是往前走一步。Open

data 在生物多樣性學門一直都在努力，TaiBIF 是跟 GBIF 接軌，大概有 300 多萬筆的原始資料是完全公開的，但還有更多的資料很多部會還沒 release 出來，但是都有進步，這是好事。GBIF 在中研院有一個中華民國委員會，今年 6 月底將在台灣舉辦第六屆[亞洲節點管理者會議](#)，約有 10 個亞洲地區國家代表會來開會，這是在亞洲地區國家生物多樣性資料的合作與整合。另外，DIVERSITAS 其實跟 IGBP、IHDP 都是 ICSU 下面，它們去年就算結束，通通要整併到 Future Earth，院裡面應該也會成立一個委員會。本人將應 ICSU 下屬 World Data System (WDS) 邀請參加今年 7/6~7/11 在巴黎舉行的 [Our Common Future under Climate Change](#)，目前台灣魚類資料庫是 WDS 的 regular member。

李德財委員補充：資訊所陳昇瑋老師去年 8 月舉辦 Data Science conference，辦的滿成功的，如果今年 Data Science conference 還有繼續辦的話希望能和 CODATA 結合。Data citation、data publication 及 data 提供者的 credit 等，也是 CODATA 重視的議題。2016 年 9 月就是 CODATA 年會，希望在這之間把 CODATA Executive Director Dr. Simon 邀請來演講。

孫志鴻委員（國土資訊系統）

政府的資料裡面有一大塊是國土資訊系統，有 9 大資料庫是政府各個部會的，國發會現在正在推動。在過去幾年已經有跨出一大步，政府各部門之間拿這個資料是免費的，現在也開發很多 API，但對學術單位現在還沒有，可能是下個階段可以從這邊行文到國發會，因為學術單位如果是科技部的計畫也算政府的資料，像氣象局、水利署這些大量資料對學術研究都很有用，這部分可以去爭取。國土資訊系統盡量朝開放方向發展，目前並未對民間完全開放，只有 release 部分給民間使用，下個階段會朝向完全開放發展，目前仍努力中。

羅慧珍委員（台灣營養學會）

在 open data 這部分主要是全國的營養調查數據，由中研院潘文涵教授召集國內學者一起執行，上禮拜五下午潘教授在台灣營養學會 41 屆年會裡有個場次，找了有使用 data bank 裡面數據的一些教授們，就做出來的研究，在那個地方跟所有會員發表/報告。所以 open data 是給需要做營養相關研究的人，看全國營養的相關問題，可以跟國衛院，也跟潘文涵教授申請使用權限，這部分還結合國健署的健保資料庫，可以供大家使用。

許聞廉委員補充：

內人（潘文涵教授）去開會的時候發現學術界的老師可能認為把研究做完，這研究的產品就會自然而然的發揚光大，沒有想到怎麼 promote，其實 open data 是一個很好的機會，像營養資料她已經統計了 20 年，都是全部 open，全台灣約有 100 多本碩博士論文是 base on 該資料庫，所以影響力非常大，國外現在也可以使用。其實我們做研究都沒有注意這一點，多數學者做完資料，學生走了就沒有了，沒有一個機制讓資料永久留下來，現在很多 open access journal 強調你給多少錢就可以一輩子幫你 archive 起來，這是很好的機制。

陳鴻宜委員（中華民國物理學會）

open data 在物理界發展的狀況：去年 11 月才開始有第一次的 open data，物理界在歐洲（法國瑞士）地底下開發了一個加速器（可以打出黑洞的機器），他們把所做的資料公布在網路上，但目前這些資料都屬於舊資料，資料公布後，史丹佛大學的評論是物理界的動作跟整個世界潮流來比是較慢的，因為放出來的 data 只適合高中生做統計的訓練，可以發現 1980 年諾貝爾得獎主丁肇中博士發現的那個東西。以目前整個物理學界來說，台灣物理界要推動這塊其實還有很長的路，目前我們做比較多的是 open source，可能會寫一些程式作分析資料，這些大多很樂意公開在網路上，但自己的實驗數據因為牽涉到

credit 的問題，大家就會保護的很好，算是比較保守的現象。

林誠謙委員補充：

高能物理的 open data 還沒有解決，雖然壓力夠了，必須 open，但當初實驗設計的原則是兩組必須完全獨立，完全都不知道對方結果，要確認兩個找到同樣的東西，不過 open 可能不是最大的問題，data open 出來怎麼用它才是問題。另外，台灣太少去解釋 rationale，為什麼要 open，open 對社會、科學有什麼好處。芬蘭有政府白皮書，但我們沒有 policy。

莊裕澤委員補充：我們今年要做報告，再請林委員提供各國相關資料。

朱有花委員（國際天文聯合會）

天文在 open data 這塊是挺先進的，我們已經進行好幾十年了，美國太空總署所有的 data 完全公開，我們寫 proposal 申請到的時間，拿到 data 只能用一年，一年之後完全公開。如果你需要很多時間，像哈伯太空望遠鏡或其他望遠鏡，需要很多時間，申請 proposal 時勾選完全不要公開就可以了。拿到 data 馬上公開，人家當然才會給你時間。不但公開，天文台還把資料提供給業餘人士去做精美圖片去比賽，做 outreach education，甚至經過基礎訓練之後參與星系分類。

何孟樵委員（生物物理學會）

生物物理與天文算是幸運的一群。生物物理有結構生物學，做結構全部的 data 你要 publish，我們看的 journal 中，reviewer 第一關，你 submit 到 PDB 沒有，PDB 審核會給你一個 code，他會做全盤的整合：你要的 information、你要的資料、你的 quality 全部都在裡面。journal 要求一定要 submit 到 PDB，reviewer 看到有 PDB 就會認為他的 data 是 safe and qualified。它是 reduced data，還沒有 database，true data 是 image，reduce 成數字化，把 model 放進去，reviewer 可以下載來直接

看 paper 沒有辦法 describe 全部 pictures, quality 全部都在 PDB。PDB 是由好幾個機構共同 run, 他們也研發很多 tools, 全部都是 open, 除了學者上傳之外, 美國還有好幾個機構會將解出的結構上傳, 讓研究者實驗。User 上傳之後只可以鎖一年, 不開放, 一年之後若決定不開放 data 就 destroy 掉, 重新再來。對研究者而言, 這也是一種保障, 上傳等於 claim I am the first, 對 User 有很多好處, 也是 requirement, 就是 push 那個 database 一定會成形, 因為要 publish 論文一定要上傳, 學者要做研究只要到 database 去抓就可以, 裡面有很多 data, 所以可以整合出很多 modules 給大家使用, 所以生物物理界在這方面是很幸運。

朱有花委員補充：

天文物理與天文期刊的作法是出了 paper 以後, 他要你把 Data 與 paper 同時公開, 從 journal 可以把 data download 下來。

賴進貴委員 (中國地理學會 & ICU 會員)

在地理學界和 open data 比較相關的是

1) 目前政府部門有很多資料是開放的, 譬如說農委會農情調查報告, 全台灣的幾百種作物每年在各鄉鎮的產量與種植面積完全 open, 只是這些資料在網路上以表格方式呈現, 無法下載, 僅能複製貼上, 使用上較為不便。若有機會的話, 或許透過各種管道請政府想辦法讓資料提供更有效率。

2) 中國地理學會很多事情會和 GIS 綁在一起, 我們看到越來越多資料背後有其空間意涵 (譬如說他應該要有座標), 包括剛剛聽到的全台灣營養調查資料, 很好奇它背後有沒有 Geo-reference, 如果有的話就可以做區域的差異 (其他委員補充: 資料應該是有的)。現在政府部門有很多這種資料, 但是理論上它是可以變成有座標 GIS 在裡面, 但是原始的調查者可能沒有提供或是沒有附上, 或是甚至沒有把空間座標提上來, 這是蠻可惜的。因為現在我們所謂的 spatially integrated

social science 只要把座標加上就可以探討很多區域差異或人文社會意涵。所以我們建議政府部門，將來這些所有調查，只要有座標的都應該把它的空間意涵納進來。

3) 在 GIS 界（在地理學界）我們比較沒有那麼多政府部門相關資料庫，但 open data 對我們另外一個很重要的層面是現在在網路的世界裡，有越來越多的一般人稱為 crowdsourcing，我們稱為 VGI (volunteered geographical information)，這些資料也蠻龐大的，譬如說大家通報哪裡塞車，這些資料來自於使用者的智慧型手機，本身也是非常重要的資料來源，這樣的議題目前在地理學界很多人在探討，如何去收集這些資料，如何確保它的品質是可以用的，這背後可能需要資訊的幫忙，確保資料品質。

4) 107 年的課綱目前正在修訂，我知道在 107 的課綱裡面資訊教育又被拿回來兩個時數，在九年一貫的時候被拿掉，這次又放回來了，我想他們在裡面會強調公民素養，公民素養包括如何找到有效資料去正確使用。所以我們資訊界的朋友若有機會參與課綱的編撰的話，或許【資訊素養及如何使用這些公開資料】也是蠻適合放到課綱裡讓學生去了解概念。

許聞廉委員補充：

很多人不知道他的資料可以被別人加值/使用，如果把這些好處慢慢讓大家知道，讓提供資料的人有一些 credit（這個其實蠻重要的），他可能就更願意把它 release 出來，這個要靠大家在觀念上多溝通。

廖啟成委員（財團法人 食品工業發展研究所）

在新竹的食品工業發展研究所裡有一個單位，早期叫做菌種保存及研究中心，現在叫生物資源保存及研究中心，主要保存的有微生物、細胞及基因等三大資源，目前可以在我們的網站上查詢到三萬多筆微生物及細胞株之資料，同時我們在國際上有參加兩個系統，一個是 GBIF，透過 TaiBIF 系統，有一萬多株微生物資源的資料對外公開，

另外，國際上有一個 WFCC(國際菌種聯盟)，我們在這個系統裡也提供了一萬多株的微生物資源資料。

許聞廉委員補充：中研院的 Biobank 幾乎每一段時間就把收集到的資料公開，大家可以馬上做研究，沒有保護期。這是負責人沈志陽教授個人的決定，在國家立場，還是要有更好的立法推動會比較容易。

郭敏玲教授（中華民國免疫學會）

在氣喘過敏方面一直我們有 ISAAC 整合氣喘調查，我們後續再評估看看是不是可以結合營養、地理等資料公開，其他一般醫療相關，像是免疫、過敏方面，我們會員有 genetics 方面的研究，應該也有機會公開。

莊庭瑞教授（前 CODATA-TAIWAN 執行秘書）

附件三中的 Data Citation Outreach 裡面提到很多 open data 的原則，也是一個實務的工作，資料發佈（data publication）之後希望可以被人使用（在論文裡面被引用，或是自己的資料集在發佈時如果有加入別人資料該如何引用的實務）。現在資料發佈有很多種方式，一個是自己研究室就可以發佈資料、另一個是研究機關或學校有 repository，學校會要求教授資料發佈時放在 repository 裡，像很多論文的 preprint 在這個 institutional repository，或是大的實驗室/大的出版公司他們也會經營這個 data repository，因為發表論文的時候希望伴隨論文的資料可以在一個固定地方被取用和被重複使用。

我和 Paul Uhler 先生（前美國科學院/科學資料與資訊委員會主任委員）比較熟，他過去幾年在 CODATA data citation standard practice group 的 co-chair，去年我去印度開會時 Paul 有提起 data citation 的想法，因為 data citation 是一個剛要起來的做法，希望各地可以去做推廣，當然他也尋求很多人的意見，我也表達台灣很願意配合做這樣的事情，今天（開會）也是一個很適合的場合來跟大家分享有這麼一個

工作，預計是今年或明年來做，因為需要各地 CODATA 委員會的經費、時程等各種因數來配合，據我知道，美國 CODATA 委員會今年沒有經費，可能是明年要做，我設想要做的話會邀請 1-2 位國外這方面比較有實務經驗的講者，譬如附件三中第 9 位的 Jan Brase (President of the International Council for Scientific and Technical Information (ICSTI)←CODATA 相對應的組織) 來台灣。台灣有在做資料寄存的單位，像中研院的人社中心有個調查研究專題計畫有寄存很多社會經濟資料集，資料如何被引用的實務，以及剛剛各位教授提到的各個學科資料被發佈及應用的實務，希望可以做一個分享。這樣的活動一定是 CODATA 台灣委員會或中研院科學資料的議題來一起辦理，我很樂意來做一些聯繫。

林誠謙委員問：可否簡單講一下它的 principle，data citation 和傳統的 citation 有什麼不一樣？

莊庭瑞教授答：citation 是一個 attribution，通常我們論文發表的時候通常是作者/論文名稱/發表期刊，在 data citation 裡有一個跟論文 citation 比較不一樣的是以前論文發表的發表人一定有一個 affiliation，這 affiliation 通常在論文裡面會看得到，data citation 有時候可能希望著重資料整理單位，希望可以去 attribute 做 data collection 的這個組織。data citation 還有一些比較實務的問題，資料在發佈的時候有時會有不同的版本，譬如人口調查資料，每年都有一個版本出來，你 cite 台灣人口統計資料就要講的很清楚是那一年的版本。資料發佈但是如果不能被公開取得的話，citation 要怎麼 cite，因為通常 citation 的好處是說我 cite，我有一個連結，這個連結之後我就可以去 download。所以 citation 和 open data 的關係到底如何有時候也不是很清楚，他們是希望透過這些實務的工作把它釐清。有一些原則，我們希望去遵守。如果這樣發佈資料的話，人家來引用就會比較方便。

三、會務規劃

許聞廉主任委員：

我們最主要是大家彼此交換意見，促進國內在 open data 作法上的新想法。像我們剛剛開會就會發現在不同的領域要推展的方向和 open data 的方式都需要有特別的 format。像我在做質譜儀資料分析，美國 NIH 也規定說只要你拿它的錢就一定要 open，但是那 open 沒有 format，你到上面完全不知道資料怎麼辦，其實你要知道這個質譜的 peak 是屬於什麼 protein，這些做實驗完是有資料的，但是他不寫上去，給你一堆頻譜，那沒有用，所以很多的 open data 要達到那個效果一定要有固定的 format 格式，一定要有管理 organization 單位其實非常重要。

另外，中研院因為國際科學總會的關係，下面大概有五、六十個相關的會議都是會員，造成一個現象就是我們每一年交的會費一直上漲，去年我們 CODATA 是 40 幾萬台幣（會費），不是個人、是團體的，60 個 member 各位可以算一下多少錢，每一個 member 大概 1 或 2 年要去開會員大會，像去年 CODATA General Assembly 大會有三位參加：李德財教授（CODATA Executive Committee Member）、許聞廉教授（主任委員）、王新民教授（執行秘書），但研究院說一年最多 support 一個人（其他你們自己研究所想辦法），像這類的負擔其實很重，不曉得國研院能否在這方面跟科技部反應，有一些其實是關係到整個國家的 policy，可以分擔支出。希望在各個不同領域能夠有一個方式一起推動，而不是只有中研院去參加這個會，其實很多學會我想應該是有很多類似這種機會。所以目前的規劃是希望能夠慢慢調查什麼樣的學會可以有什麼做法，讓台灣能在這些領域在國際上被看得到。

四、臨時動議：

無

五、散會（晚間九點）