

**Time & Location:** October 29, 30, 31 2012 @ Academia Sinica, Taipei, Taiwan

**Keynotes :** [Sálvano Briceño](#), [Der-Tsai Lee](#), [Hesheng Chen](#), [Geoffrey Boulton](#),  
[Ovid Tzeng](#), and [David Carlson](#)

**Report prepared by:** [Nancy Chen](#), [Pohan Chang](#), [Ming-Huang Wang](#) and [Andrea Huang](#)

---

**Why does open data and open information for a changing planet matter?** The science should be accessible, and the argument (the data and concept) should be presented together. This is the two fundamental principles of [Open Science](#) that [Geoffrey Boulton](#) has proposed to the CODATA 2012 community.

Data in a changing planet context faces several [challenges in large scale](#) scientific research as [Hesheng Chen](#) has indicated. Among the large scale research, the disaster management concerns the big data processing. [Sálvano Briceño](#) makes the point further clear that understanding [components of risk](#) (hazard and vulnerability) is the key to assist policy making for disaster mitigation and for sustainable development. On the other hand, [Ovid Tzeng](#) takes a different perspective to exam the boundary of freedom of data expression, which might influences policy directions in a digital age.

To further enhance our recognition with the importance of open data, Boulton provides us eight reasons to explain why open data is an urgent issue. Regarding the issue of reliable knowledge, they are:

(1) Closing the concept-data gap. (2) Maintaining the credibility of science. (3) Combating scientific fraud. Regarding the issue of big data challenges, the reasons include: (4) Exploiting the data deluge and computational potential. (5) Addressing planetary challenges. (6) Restraining to the “Database State”. And, regarding the issue of society needs, they are (7) Supporting citizen science. (8) Responding to citizen’s demands for transparency.

**What have been practiced, and who have been participated in the open data and information movement?** [Der-Tsai Lee](#) introduces the Taiwan e-Learning and Digital Archives Program ([TELDAP](#)), and shares the experience in digitalizing cultural, socio- and bio-diversity data. TELDAP, coordinate by Academia Sinica, involves many government institutions, such as National Palace Museum and National Taiwan University. In addition, this project involves the general public, as well as many content experts and software engineers. In TELDAP, a variety of databases, metadata standards for various types of digital content have been developed. Many techniques and researches are completed in TELDAP, including information retrieval, image processing, and so on. During this speech, Lee demonstrated the project and showed how this project brings our cultural heritage from institutes to the general public, and integrates it into daily lives.

The other project is the [collaboration between IRDR and CODATA](#). IRDR discovered a lot of disaster damage is due to human and social vulnerability. People died in natural disaster are often due to bad infrastructure and lack of awareness. It is important to

identify, monitor, collect and organize data on the hazards as well as on social and human vulnerability, which could be used in developing policies to facilitate building resilience to these hazards. CODATA and the IRDR collaborate closely and use the DATA Working Group for promoting and facilitating the exchange of information among institutions and developing standards, common approaches and terminology, and specific activities on the 2015 timeframe to achieve the sustainable development goals.

What if you organized an international project on a free and open data access basis, and everyone joined? The project of the [International Polar Year](#) lead by [David Carlson](#) is a sound illustration for the use of open data to advance polar knowledge through mass collaboration. More than 50,000 participants from over 63 countries have joined the programme and contributed to the primary strengths of IPY, widespread international participation and vast multi-disciplinary contributions. Most of the IPY funding (around 1.8 billion Euro) is used for data collection and analysis, that Carlson describes the practice as “a big project done by small hands”. The IPY’s Polar Information Commons (PIC) is a free and open data access test bed for innovation and change, for national, international, disciplinary, and interdisciplinary identification, and for sharing and preservation of data. However, while advancing knowledge, inspiring next generations and attracting public interest perform well in the IPY experience, enhancing infrastructure and integrating data become main obstacles to this open data movement. Thus, **how can we make open data and open information a reality?**

Boulton proposes a concept “intelligent openness” with four criteria, eight principles and eight tools and process for making data intelligently open.

The open science policy is for effective communication, replication and re-purposing. The fundamental demand is the open access of both data and meta-data, and they should not only be accessible, but also intelligible, assessable, and re-usable. Only when these four criteria are fulfilled are data properly open.

Among the eight key principles that regarding **norm changing** includes:

1. A shift away from a research culture where data is viewed as a private preserve.
2. The cost of intelligent openness is an integral part of the cost of doing science.
3. Replication is by far the best guarantee of preservation.

Regarding **data management methods**:

4. The data evidence for a published argument **MUST** be intelligently open at the time of publication.
5. Data management should be embedded in the community producing and using the data.
6. Science data should be as easy to “remix” as music is to a DJ.
7. Give credit for useful data communication and novel ways of collaborating
8. Common standards for communicating data

Eight essential enabling tools and process:

1. Data integration
2. Supporting dynamic data
3. Providing provenances
4. Annotation
5. Metadata generation
6. Citation
7. Access to data scientists
8. Changing the library

COATA 2012 專題演講內容針對此次會議，提供了最佳探討三大問題點的回應。主要包括，對於一個變動中的星球而言，開放資料與資訊為何重要？在開放資料與資訊的工作上，有哪些成員組織、以及哪些工作案例已經實際進行？在方法上我們如何落實開放資訊？

### **(1) 對於一個變動中的星球而言，開放資料與資訊為何重要？**

科學必須能容易取用，而科學論點 (資料與概念) 也必須能一起呈現，並提供使用。這是 **Geoffrey Boulton** 在 **CODATA 2012** 會議中，對與會者所提出開放科學(Open Science)的 二大根本原則。呈如中國科學院陳和生院士所提出的研究指出，變動中的星球面臨許多大尺度科學研究的挑戰。而在大尺度的科學研究範疇下，巨量資料的處理是一大考量。**Sá Ivano Briceño** 亦清楚地論述了解災難風險的元素 (災害與弱點) 是能有效輔助災害應變政策的制定、以及永續發展的關鍵。在另一方面，曾志朗則提出在數位化時代中，我們必須考慮政策制定的方向，是可能受到資料的自由表達界限範圍而有所影響。

在 **Boulton** 所提出的八個理由中，則可幫助我們更加了解開放資料的重要性。在關於可信賴知識層次上，包括 (1) 縮短概念與實際資料之間的差距 (2) 維護科學的可信度 (3) 對抗科學的欺騙行為。在關於巨量資料的挑戰上，包括 (4) 防止資料的過度氾濫 (5) 因應星球層次的挑戰 (6) 對資料庫狀態加以控制。另外，對於社會的需求則包括 (7) 支持公民科學的發展 (8) 回應公民對於資料透明的需求。

### **(2) 在開放資料與資訊的工作上，有哪些成員組織、以及哪些工作案例已經實際進行？**

李德財院士在會議中和與會者分享台灣數位典藏與學習計畫，保存台灣的文化，社會與生態多樣性等資料的經驗。台灣數位典藏與數位學習計畫是由中研院協調，並有許多的政府單位與學術機構一同合作，包含國立故宮博物院與國立台灣大學。當然，還有數位內容專家，軟體工程師，與社會大眾也參與了這項計畫。在此計畫中，建立了多種類的資料庫與後設資料標準。許多的技術與研究也在此計畫中完成，包含資訊檢索，影像處理等。在此場演說中，李院士展示了這個計畫的成果，並分享如何透過此計畫，將我們的文化遺產離開展示櫃，讓大眾能夠體驗，進而融入於全民的日常生活中。

另外一個案例則是介紹 **IRDR** 與 **CODATA** 的合作計畫。**IRDR**(災害風險綜合研究計劃) 發現許多災害所產生的損害，其實是由於人性與社會的脆弱點所造成。人命的損失往往是由不良的基礎建設、或是缺乏對災害應變的認知所造成。因此，辨別，監看，蒐集災害資料以及社會和人性弱點並加以運用於制定具災害韌性的法律非常重要。**CODATA** 與 **IRDR** 緊密合作並使用資料工作群組的方式，來促進不同研究組織之間資料的交換、建立標準共同的方法，專業辭彙統一，以及特定的活動等，以期能使用 **2015** 的時間框架來達成永續發展的目標。

那麼，若是組織了一個以開放資料與資訊為基礎的計畫，而且許多人都參與其中，那又會是一個怎樣的情況？國際極地年(IPY)是一個由 **David Carlson** 帶領組織，且基於開放精神的增進極地知識計畫，而 **Carlson** 也以「一個有由許多小手打造的大計畫」來形容該計畫。由於開放的精神，吸引了來自超過 **63** 個國家超過 **5** 萬名的參與者參加，這同時造就了計畫的主要特點—眾多的跨國參與以及巨量的多元領域知識成果。大多數的 **IPY** 經費 (大約 **1.8** 萬億歐元)都是被用於資料蒐集與分析的相關領域。**IPY** 創訂的極地資訊共享空間 (**Polar Information Commons**) 的成果，或許也可以被用來檢視國內或國際間關於創新與改變的開放資料存取、單領域或跨領域的辨別、以及資料的分享與保存。然而，當該計

畫在開放資料的前提下促進知識發展、鼓勵年輕世代、或是吸引一般民眾的注意力等方面，表現亮眼，但同時卻難免面臨資料整合與取用性等問題。因此，在方法上我們如何落實開放資訊？

### (3) 在方法上我們如何落實開放資訊？

Boulton 提出了一個「明智的開放」(intelligent openness)概念，可能提供些許答案。其中包含四項條件、八項原則、以及八種工具與程序。

一個適切的開放科學政策可確保資料的流通效率、可複製性，以及便於重新調整用途。資料與後設資料的開放取得都是基本的需求，應該使其能夠被以正確方式取用 (accessible)、可瞭解(intelligible)、必須能獲得評價的資訊 (assessable)、以及再利用(re-usable)。真正的資料開放須滿足此四項條件。

八項原則包含在二個層面之中 (意識型態的改變、以及資料管理方法)。在意識型態的改變中主要原則為：

1. 改變將資料視為私有品的研究文化。
2. 「明智的開放」的成本應視為科學工作所需成本的其中之一。
3. 重製是目前最佳的資料保存方法。

在資料管理方法方面的原則是：

4. 對於出版一個論點時，其輔佐論點的資料證據必須在出版時，同時明智的開放。
5. 資料管理，應由資料本身的生產和使用社群來進行管理。
6. 科學資料的混合運用，應如同一個 DJ 在混合音樂 (remix)一樣容易。
7. 對於有效的資料溝通、與創新的合作給予鼓勵。
8. 提供資料溝通的公用標準。

明智的開放，主要的八種工具與程序則包含：資料整合、動態資料支援、資料來源的提供、標記、產生後設資料、資料引用、接近資料科學家的管道、圖書館改革。

簡而言之，這些在變動星球脈絡下的開放資料與資訊的願景、任務與急迫的需求，都在在提示我們需要更多的努力，然而，「我們將無法到達終點，但我們仍必須保持前進」。

Geoffrey Boulton 以此作為專題演講的結語，也同時可做為我們對開放資料與資訊研究與實務工作的期許。