

Open Data and Information for a Changing Planet



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

Scientific Domain: Earth Science (sessions: **D5** and **E5**)

Report prepared by: John Wang and Andrea Huang

The Earth Science related sessions during Day 3 of CODATA 2012 include [the role of e-science in disaster mitigation](#) and [data sharing and reuse in the field of material science](#). The session starts from the point that not only data-intensive computing makes detailed and quantitative scientific understandings for disaster mitigation possible, but also international collaboration, interdisciplinary research and bottom-up approach are [what e-Science can provide for the mass](#).

The experience of Academia Sinica Grid Computing Center (ASGC), which serves as the World-Wide Grid and e-science Asian Center, presents its overall picture in several international collaboration tasks, and discusses some challenges for big data and cloud computing, such as simple services are hard to create, commercial clouds are too expensive for Big Data, as well as scalability and security remain complex. One of the projects using ASGC as a center for central data storage and access is the installation of broadband seismic array at Indochina and the South China Sea for [monitoring earthquake and tsunami](#); the data collected by the seismic array were also used by other projects for tsunami simulation as part of the [early warning mechanism](#). The [Weather Research and Forecasting \(WRF\) project](#), which deals with climate simulation, also employs gLite-based web portal on ASGC for connection to global grid infrastructure.

The session of material science covers three main issues, i.e. interoperability and standardization, the complexity of material database (e.g. nano scale, the difference of physical and chemical properties, or [measurement uncertainty](#)), as well as ontology making for nano materials. The new working group for describing materials on the nano scale is formed by CODATA and [VAMAS](#) (The Versailles Project on Advanced Materials and Standards) in 2012. [Uniqueness and equivalency](#) are two main goals for Materials Description System. However, several challenges remain for new materials to be described and defined. Factors, just to name a few, include the size effect, the new complexity, poor knowledge of mechanisms of action, or multiple users in multidisciplinary scientific domains, etc. The approach to meeting these challenges is suggested by the collaboration between International Council of Science (ICSU) and CODATA in the agenda for a new working group.

To improve data integration and interoperability of nanomaterials, the other alternative method is semantic web technologies. A project for the [nanotechnology ontology](#) provides a collaborative environment (Ontolution platform) for highlighting emerging terms and relations, updating ontologies, as well as combing and mapping existing ontologies from different sub-disciplines. Similar challenges were faced by the inorganic chemistry field while building the Information Resources on Inorganic Chemistry (IRIC) database: the database was built using the [Metabase concept](#) to integrate heterogeneous systems into a single user interface. Another complex material field that [requires extensive database](#) to cover various compounds with vastly different physical and chemical properties is the study of rare earth (RE) and rare earth/transition metal (RE-TM) materials; a team from the Beijing University of Technology is current developing just such a database.

CODATA 2012 會議第三天與地球科學有關的議程包括了 e 化科學在災害和緩的角色定位，以及材料領域中的資料分享與再使用。會議中首先提供了有關巨量資料的密集計算觀點，這些研究不僅促使和緩災害影響的研究，能在精細與量化的資料與分析上有所幫助，同時另一方面，國際合作、跨領域研究、以及由下往上的研究方法，均是 E 化科學能協助學界與大眾知識領域的探索。

首先介紹的是參與「全球網格」(WWG) 以及作為 E 化科學的亞州中心的中央研究院的「中研院網格計算中心」(ASGC)所參與的國際合作相關研究，以及探討巨量資料與雲端計算所面臨的挑戰，例如簡易的服務卻難以建立、對於巨量資料而言，商業雲端服務價格昂貴、規模的可擴展性以及安全性複雜。其中的一項計畫使用 ASGC 作為中南半島以及南海地震與海嘯觀測網的一個寬頻地震儀網檔案存取中心，其他學者也可透過 ASGC 取得這些觀測資料以作為地震與海嘯電腦模擬與預警使用。而模擬氣候變化的氣象模式預測(WRF)系統也是使用 gLite 系統通過 ASGC 與全球網格基礎設施連結。

材料領域議程則包含相互操作性與標準化、材料資料庫的複雜度 (例如，奈米規模的大小、物理與化學的特性差異、或是觀測的不確定性等)、以及奈米的知識本體模型之建立。CODATA 與 VAMAS (高等材料與標準的凡爾賽計畫)在今年(2012)合作建立了一個以奈米規模為基礎的材料描述工作小組。而唯一性(uniqueness)與對等性(equivalency)是材料描述系統的二大重點。然而，面對新型的材料，其描述與定義具有許多挑戰，如材料大小的影響、新材料的複雜性、新材料反應機制的知識缺乏、多元學科領域中多學門使用者的複雜度等。這些挑戰將由 CODATA 與「國際科學理事會」(ICSU) 合作組成的工作小組進行規劃。

為了改進奈米材料的資料整合與相互操作性，另外一個方法選項是語意網技術。奈米技術知識本體計畫提供了一個協同合作的平台(Ontolution platform)，進行逐漸增加的語彙與關係的加強、提供知識本體的更新、以及整合與對照眾多次領域學門中既有的知識本體。面臨類似挑戰的還有非有機化學領域：在架設非有機化學資源 (IRIC) 資料庫時，此資料庫以 Metabase 概念為基礎，使用後設資料庫將來自異質系統的資料，整合進入單一使用界面。另一個需要紀錄大量不同的物理及化學特性的材料科學領域為稀土及稀土過度金屬 (RE-TM) 材料的分析，而北京工業大學目前有一組研究團隊正在架設此領域用之資料庫。