

Jun 21st, 8:45 AM - 9:15 AM

## Remarks before the IATUL

Arden L. Bement

*Purdue University*, bement@purdue.edu

Follow this and additional works at: <http://docs.lib.purdue.edu/iatul2010>

---

Arden L. Bement, "Remarks before the IATUL" (June 21, 2010). *International Association of Scientific and Technological University Libraries, 31st Annual Conference*. Paper 7.

<http://docs.lib.purdue.edu/iatul2010/conf/day1/7>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact [epubs@purdue.edu](mailto:epubs@purdue.edu) for additional information.

## **Remarks Before the IATUL**

**Purdue University**

**June 21, 2010**

**Good Morning!**

**I am delighted to participate in this important conference and am grateful to Dean Mullins for inviting me to say a few words of welcome and comments.**

**I can't think of a more important yet daunting challenge than ensuring the utility and integrity of research data in a digital age, and I congratulate the report committee of the National Academy of Science for taking on this challenge. The report that this committee issued is serving as a valuable reference not only for the scientific and engineering research communities but also for government research funding agencies coping with all aspects of data management, evaluation and archiving in an open environment.**

**During my service as Director of the National Science Foundation during the past six years, I have witnessed a revolution unfolding in the generation, communication, retrieval, categorization, synthesis, storage, and archiving of massive streams of data. The increasing use of large sensor arrays, massive detectors of sub-atomic particles, autonomous astronomy telescopes on both Earth and in space, and teraflop computers has greatly increased the generation rate of data.**

**Furthermore, scientific discoveries are being made by data synthesis almost as frequently as by new data generation. Therefore, the intensity of community interests in open data sharing has increased greatly.**

**The dark side of this revolution has been a substantial increase of scientific misconduct. Advances in computer and software technology have facilitated the improper reporting and falsification of data and unabashed plagiarism. In some extreme cases these infractions have resulted in sanctions that can be career damaging to both faculty and graduates. It**

**is now essential for faculty reviewing thesis dissertations to pay as much attention to the literature review chapters as to the research procedures and results.**

**Among the many questions that have engaged the NSF in coping with new Congressional mandates requiring open access to data and publications produced from grants and contracts funded by the federal government, I will mention five:**

- 1. How should we differentiate the policies for open data access vis-à-vis access to metadata and publications?**
- 2. What constitutes data?**
- 3. How should a funding agency balance policies that appear to be in conflict with open access to the commons, such as: intellectual property rights, provisions for IP ownership by academic institutions under Baye-Dole, and requirements for peer review, assessment, and consensus building within the scientific and engineering research communities?**

**4. In the face of growing complexity, what allowances should be made for analyzing signals from noise and characterizing patterns in data before releasing it to the public? In the minds of some noise is in part signal that has yet to be detected.**

**5. With publically-supported data channeled into the commons and freely accessible, what signifies provenance; likewise attribution; likewise accountability?**

**I believe other issues in the open access of data need further deliberation. For example, what value can be placed on data derived from computer models and simulations that have not been validated by physical and mathematical models or working prototypes? Likewise, what value can be placed on measurement data that are not backed up by instrument calibration standards? The merging of such data with openly shared data bases without adequate evaluation or demonstrated reproducibility can erode the value and credibility of the database.**

**In this era of massive data generation, I like to think of data in three categories, using a mining metaphor: “raw ore”, “concentrate”, and “virgin metal”. The question is which data are worth saving and which throwing away? The increasing costs of data storage, evaluation, data base building, and archiving gives vital importance to this question. To my thinking it is only after raw data have been evaluated that it reaches the category of “concentrate”. It is only after evaluated data are assembled into metadata, databases, or functional relationships that they achieve the distinction of “virgin metal”.**

**I wish you success in this timely conference. Through your discussions and deliberations the U.S. can take a major step forward through the leadership of scientific and technological university libraries in serving the data needs of scientific communities at large in the digital age.**