Cyberinfrastructure: Enabling the Chemical Sciences

To set the context on the enabling role of cyberinfrastructure for the chemical sciences, it may be most appropriate to start with the definitions found in *Wikipedia*, the (cyber)encyclopedia. Cheminformatics, the instantiation of modern advances in information technology (IT) to the world of chemical sciences, is defined as “the use of computer and informational techniques, applied to a range of problems in the field of chemistry”. A relatively newer addition to our lexicon, cyberinfrastructure, appears as “the coordinated aggregation of software, hardware and other technologies, as well as human expertise, required to support current and future discoveries in science and engineering. The challenge ... is to integrate relevant and often disparate resources to provide a useful, usable, and enabling framework ... characterized by broad access...”

Understanding the distinction and interplay between informatics (the field that is enabled) and cyberinfrastructure (the enabling platform) is not simply an exercise in semantics but a foundational element for high-level strategic planning for any large organization. From a tight-knit ecosystem of R&D scientists within a large firm to a loosely coupled community of chemical scientists sharing common research goals, the appropriate prioritization of infrastructure investments timed to match successive waves of technologies entering the “sweet spot” may spell the difference between success and failure in the modern, data-intensive, and data-driven era.

Infrastructure projects are extremely expensive, but true infrastructure is a shared platform that supports many activities. And it is this aggregate of the many benefits that provides an acceptable to exceptional rate of return at the appropriate societal scale. This observation offers some guidance going forward for the chemical sciences community. First, we must be nimble and adapt our tools to take advantage of seismic, trend-setting directions in the global/societal cyberinfrastructure (examples: from almost zero to over 1 billion tunes downloaded *legally* over the Internet per year and from almost zero to over 1 billion radio frequency identification—RFID—tags consumed per year to drive ultraefficient supply chains). The “next generation” tools in information integration, data analytics, and work-flow management will supersede current models in part because of their alignment to match greater expectations set by new infrastructure. Second, where current cyberinfrastructure is inadequate to meet the needs of the chemical sciences and a key factor in holding back progress, the requisite large-scale investments can be best raised by articulating the broader utility of the shared infrastructure. This is a subtle difference (but complementary) to the standard strategy of espousing the broader impact of the chemical sciences per se.

The articles in this e-science focus issue of JCIM present a compelling case for expanding the frontiers of chemical sciences leveraging the “shared investment” cyberframework. Data-centric collaboratory environments, grid computing, interoperability standards, and the semantic Web are all prime examples of emerging infrastructures with the potential for broad societal impact. This issue illustrates that the chemical sciences community is well-positioned from the perspective of contributing to and benefiting from the next cycle of societal cyberinfrastructure investments.

Sangtae Kim

Donald W. Feddersen Distinguished Professor, Purdue University

CI060100+

10.1021/ci060100+ CCC: $33.50
Published on Web 05/22/2006