GuestEditorial

PETER LYSTER, PhD

Infrastructure and Workforce Needs in Biomedical Informatics and Computational Biology

In science, there is a need to balance research in domain sciences and the infrastructure to support that research. Basic research mediated through peer review is understood to produce useful discoveries. But outcomes of infrastructure development and implementation—whether it is hardware, instrumentation, software repositories, databases, tools, or workforce—are more difficult to reliably assess and predict. NIH is actively addressing this issue and the investment will be an integral part of discovery, and, in the case of biomedical research, improved understanding of human health and treatment of disease.

In 1999, NIH Director Dr. Harold Varmus asked eminent scientists David Botstein and Larry Smarr to lead a working group to examine strategies for biomedical informatics and computational biology at NIH. The resulting Biomedical Information Science and Technology Initiative (BISTI) report (http://www.nih.gov/about/director/060399.htm) recommended increased resources for basic computational grants; workforce development; implementation of information storage, curation, analysis and retrieval (ISCAR); and improved access to cyberinfrastructure. This report laid the groundwork for establishing a number of initiatives and dialogues at NIH.

In response to the BISTI report, NIH developed new computational program announcements (PAR 06 410/411) and set up several review study sections focused on computing to foster basic biocomputational research and infrastructure. Through the NIH Roadmap for Medical Research, NIH established seven National Centers for Biomedical Computing (NCBCs, http://www.bisti.nih.gov/ncbc/) with the goal of creating "the networked national effort to build the computational infrastructure for biomedical computing for the nation." The NCBCs are required to balance the domain sciences and computing, as well as develop cores devoted to computational science, biomedical computing, and driving biological projects.

The Centers gathered in Bethesda, MD, in July 2006 for the first all hands meeting (http://www.bisti.nih.gov/ahm2006/). The output described an impressive array of new scientific results and generated new collaborations. Meeting participants helped develop a resource compendium of existing multi-agency programs, initiatives, and communities (http://www.bisti.nih.gov/ahm2006/Building%20Bridges%20Compendium.htm). We have built bridges between a number of programs represented in the Compendium and the NCBCs through a new postdoctoral program.

Efforts to coordinate across the NCBCs have taken place under the Software and Data Integration Working Group wiki (http://na-mic.org/Wiki/index.php/SDIWG:Software_and_Data_Integration_Working_Group).

The goals of the SDIWG, in concert with the NCBC program, are to advance the domain sciences, promote software interoperability and data exchange, capture the collective knowledge of software engineering and practices among the Centers, and publish this knowledge widely. In furthering this, three working groups are dealing with how to access and query resources, scientific ontologies and terminologies in biomedicine, and cross-Center domain science. The SDIWG has been a hotbed of discussion about the roles of infrastructure and workforce in biomedical computing. The working group dealing with querying resources has developed a set of attributes for describing resources that will be a foundation for web-based tools some have called the "resourceome." The SDIWG also promotes the importance of software repositories and modern software engineering practices across the seven NCBCs. For example, such infrastructure is used in the medical image processing toolkits ITK/VTK. Similar approaches have been adopted by other Centers such as the physics-based simulation toolkit Simtk.org and the LONI imaging pipeline.

Advances in infrastructure and workforce under the NCBCs have highlighted another challenge that was in BISTI report: What is the biomedical computing and computational biology community doing about ISCAR? The time is right to employ the engineering principles of project development, which proceeds from an inventory, requirements analysis, design, prototype, testing and production process. Different infrastructure requires that this linear process cycle back at different stages. Requirements analysis, for example, might need to be revisited several times over the lifecycle of a project. Indeed, it can be argued that any strategic effort contains all of the stages of project development in varying degrees. This back-to-basics approach can be seen in efforts such as the NIH Roadmap Inventory and Evaluation of Clinical Research Networks (IECRN), the Blueprint Clearinghouse for Neuroimaging Informatics Tools and Resources, the Blueprint Neuroscience Information Network, the Internet Analysis Tools Registry (IATR) and the yellow pages and resourceome working group of the NCBCs.

The importance of ISCAR was highlighted at the December Knowledge Environments for Biomedical Computing (KEBR) meeting in Bethesda, MD, where special sessions highlighted attributes of successful informatics projects that address key considerations for knowledge environments, including information representation, sociocultural issues, and, perhaps most important of all, how to ensure that the domain sciences community is an active participant in the development of infrastructure. □

DETAILS

Peter Lyster is a Program Director at the Center for Bioinformatics and Computational Biology, NIH/NIGMS