Building a Biomedical Data Ecosystem

This issue of the Biomedical Computation Review features the Centers of Excellence for Big Data Computing. These 12 Centers, funded by the NIH’s Big Data to Knowledge Initiative (BD2K), have been established on the principle that we must be united in our efforts to accelerate the translational impact of big data on human health.

The Centers will become hubs in an emerging worldwide biomedical data ecosystem. The foundations for this ecosystem are already being built. Collaborations between international groups, federal agencies and the biomedical science community are forging the way forward with pilot projects and initiatives. These projects are designed to influence advancement through a set of central drivers: to inform policy decisions, to build infrastructure, and to expand the biomedical data science community.

Each of the Centers will investigate a different programmatic theme. However, most will also face similar challenges. Some of these challenges will be dealt with by consortium-wide consensus, while we envision that others will be addressed by each of the Centers in different ways. The individuality of each Center and the collaborations between them will allow us to identify best practices, effective strategies, and program models for solving common biomedical data science problems. These findings will be the foundation for future biomedical data science policy decisions.

The BD2K Centers consortium is also a pivotal element in NIH efforts to develop common infrastructure that supports data and software sharing and cloud computing efforts within the biomedical data science community. This infrastructure, which we call the Commons, will be piloted by the Centers and guided, in part, by the outcome of another BD2K funded project, the Data Discovery Index Coordination Consortium (DDICC). The DDICC is a community-based effort to establish core principles for finding, accessing, and citing digital research objects (data, software, narrative etc.). The results of the BD2K Center cloud pilots and the DDICC efforts will provide a basis for widespread application of the Commons.

The Centers, the DDICC, and the Commons are part of an emerging ecosystem that fosters collaboration and sharing. This environment will facilitate the expansion of data science beyond the Centers and the DDICC to their collaborators, their colleagues, and their students. A major focus of BD2K efforts is on supporting training to help today’s biomedical scientists incorporate data science into their research and to produce the next generation of data-centric biomedical researchers. Each Center has a training plan to support this mission and several of the Centers are planning to collaborate in their training efforts. The network of collaboration, sharing, and training provided by the Centers has great potential to accelerate the growth of a supportive and united biomedical data science community.

The Centers cover a broad swath of experimental data and metadata issues, and their research will provide use cases for essential problems to the biomedical community. Center directors and investigators showed great enthusiasm at a recent kick-off meeting, providing an early indication of the energy and commitment of this consortium. Sustainable growth of biomedical knowledge requires sharing. Through a united ecosystem we hope to improve productivity through faster discovery at reduced cost. Success will not only be measured by the Center’s individual projects, but through other laboratories becoming part of the ecosystem and sharing their digital research objects.

Welcome to the New Biomedical Computation Review

For nearly ten years, this magazine has been published by Simbios (under principal investigator [PI] Russ Altman) as part of the National Institutes of Health’s National Center for Biomedical Computing (NCBC) program. With the end of that program last summer, the magazine faced an uncertain future. But it has gained new life with the support of the Mobilize Center (under PI Scott Delp) as part of BD2K.

Expect to see similar wide-ranging stories that cover the gamut of biomedical computing. But also watch for some new big data–focused stories and columns as well.

Happy Reading.