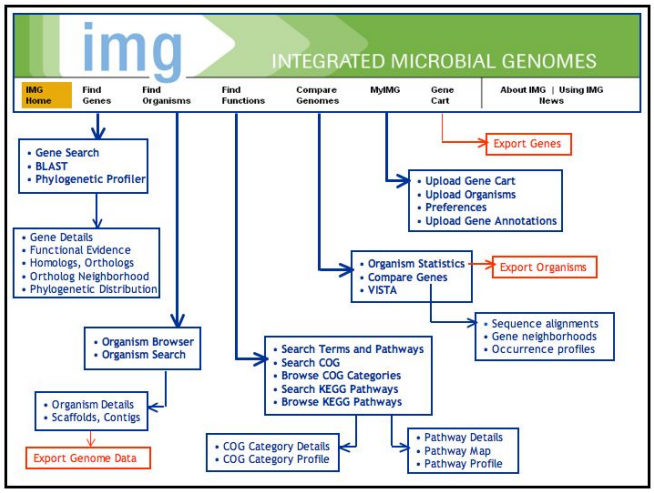
New Tools Make Biological Data More Manageable

In March 2005, DOE launched the Integrated Microbial Genomes (IMG) data management system, a community resource integrating the DOE Joint Genome Institute's (JGI) microbial genome data with publicly available microbial genome data to provide a powerful comparative context for microbial genome analysis.

IMG was jointly developed by the Biological Data Management and Technology Center (BDMTC) at Lawrence Berkeley National Laboratory and the Microbial Genome Analysis Program (MGAP) at JGI. Since its launch, IMG has been upgraded regularly and now contains 618 genomes (318 bacterial, 25 archaeal, 15 eukaryotic, 260 bacterial phage), 40 of which are finished and 80 of which are draft genomes sequenced by JGI.



The BDMTC, headed by Victor Markowitz, was established in January 2004 as a source of expertise in and provide support for data management and bioinformatics tool development projects at JGI, LBNL, the University of California and other organizations in the Bay Area. In the past, biological data management systems have typically been developed with an eye toward rapid development and low cost, often with minimal consideration for reliability, maintenance and evolution. BDMTC was created to bring industrial-grade software development practices to the increasingly important field of biological data management.

Scientists are already appreciating the systematic approach followed in developing IMG, from gathering and analysis of user requirements through the public release of the system.

"Although I was closely involved for years in the development of another microbial genome data system at Integrated Genomics, this is the first time I have experienced a well organized process in which requirements are documented, clarified, and continuously refined, and development follows a strict yet clear and predictable schedule" said Nikos Kyrpides, head of MGAP and IMG's scientific lead. "I fully appreciate the value of a disciplined development process which I now consider as critical to ensuring that the system addresses the needs of the scientific users."

Since its initial release, IMG has gained increasing recognition in the scientific community. "IMG is invaluable for comparative genomic analyses," said Mark A. Schell, professor in the Department of Microbiology and Plant Pathology at the University of Georgia. "My laboratory uses it all the time and I will use it in the classes I teach this fall." IMG is accessible to the public at <http://img.jgi.doe.gov/>.

In August, the value of BDMTC's approach was acknowledged with the announcement that the center will become a joint Biological Data Management Core Facility with the California Institute for Quantitative Biomedical Research (QB3), a partnership of three University of California campuses—San Francisco, Berkeley and Santa Cruz. "Leveraging the expertise of the two organizations will help provide new insights into preventing, treating and curing disease," said Regis Kelly, executive director of QB3. The main goal of the core facility will be to provide data management support for academic groups affiliated with QB3 help these groups to share data management experience, expertise, and technology.