

StrainInfo 2.0: providing passports for cultured microbes

Peter Dawyndt, Department of Applied Mathematics and Computer Science, Krijgslaan 281 (S9), B-9000 Ghent, Belgium.
email: Peter.Dawyndt@UGent.be

Microbial research is inherently distributed in nature. With many researchers in different research institutes collecting and analyzing bits and pieces of data on microbial diversity, knowledge gained on the microbial world also has a tendency to become spread across the information jungle. As today's upscaling of research projects often is hampered by the burden to integrate data extracted from several autonomous and heterogeneous information systems, the famous Newton quote "*if I have seen a little further it is by standing on the shoulders of giants*" is endangered to become rephrased as "*If I can see no more, it is because giants were standing on my shoulders*". An immediate and important challenge that lies ahead of us is that of end-to-end scientific data management, from data acquisition and data integration, to data treatment, provenance and persistence. While advances in computing, and in particular scientific data management and application development environments for science will become important in the future, what is vitally more important and dramatic in its impact is the integration of new conceptual and technological tools from computer science into the life sciences. This integration is likely to accelerate key breakthroughs in science and benefits to society, from understanding biology and revolutionizing medicine and healthcare, and from understanding the universe to the origin of life, and understanding and helping to protect the life-support systems of Earth on which we all depend for our survival.

The StrainInfo.net bioportal (www.straininfo.net) was established to stimulate this movement towards using multi-perspective integrated information in a broadened biological and clinical context. In particular, it concentrates on establishing automated ways to collect and integrate all information that is available about the microorganisms that were deposited into a global network of biological resource centers. Fully understanding the biology of an organism (or a population of organisms) requires to take a holistic approach wherein the phenotypic and genotypic traits of the organism are inspected in light of its ecological role and geographical spread. Moreover, only by having high quality information on those organisms we already know, we will be able to further explore and understand those regions of microbial diversity that we don't know yet. In this presentation we will discuss some of the impediments against seamless data integration, propose possible ways of better collaboration and open the discussion on novel ways to exploit the vast array of legacy data that drives microbiology.