

Top-down standards will not serve systems biology

SIR — Marvin Cassman and his colleagues, in their Commentary “Barriers to progress in systems biology” (*Nature* **438**, 1079; 2005), discuss the development of standards in systems-biology research. We agree with the need for well-curated databases, software systems that can work together to analyse such data and integrated models that can deliver the fruits of systems-based research to laboratory biologists. But we have concerns about the proposed solution, which is presented as a ‘top-down’ approach that ignores many existing and emerging standards. It seems based on false assumptions about the research community and ignores the community it is intended to serve.

Cassman and his colleagues argue that standards are needed because much software developed in research settings is not reusable by other groups of working biologists, who are not appropriately trained. But the community has many excellent quantitative scientists and software developers — and with the advent of genomics, an increasing number of physicists, mathematicians, statisticians, computer scientists and engineers have joined the ranks of biologists.

It is not a lack of training that influences software design, but the realities of developing software in a research environment where developing a professional software system is not the primary goal. As fields mature and the methodologies used to generate the data become well known and established, it is both appropriate and valuable to have standardized, easy-to-use software. But standardized approaches are not always appropriate for developing software to support new research using novel methodologies in exciting new ways.

Our collective experience, gained through the Microarray Gene Expression Data Society and the BioConductor project, clearly demonstrates that flexible systems are needed and that most initial efforts are neither well documented nor widely used. But that is not a bad thing — as science charts a particular path, the appropriate tools, if given room to evolve, do emerge and rise to the top, becoming better documented and more robust.

Even with the relatively straightforward task of assembling and annotating genome-sequencing data, computationally elegant solutions to software interoperability (such as the common object request broker architecture, or CORBA) were ultimately abandoned in favour of FASTA-formatted sequence data and tab-delimited output from various analytical tools strung together using Perl. It wasn't elegant or pretty, but it delivered what was needed in a way that sophisticated users at various locations could

replicate and adapt to suit their needs. When combined with well-engineered databases and websites to provide access, the genome projects also delivered the fruits of their work to the broader community in a form that has been extremely useful and continues to evolve. Engineering this ahead of time, particularly when the field and the tools were evolving so rapidly, quite simply would have failed.

We believe that the centralized approach proposed by Cassman and colleagues would not fare well compared with more democratic, community-based approaches that understand and include research-driven development efforts. Creating a rigid standard before a field has matured can result in a failed and unused standard, in the best of circumstances, and, in the worst, can have the effect of stifling innovation.

John Quackenbush

Dana-Farber Cancer Institute and Harvard School of Public Health, Department of Biostatistics and Computational Biology, 44 Binney Street, Boston, Massachusetts 02115, USA

johnq@jimmy.harvard.edu

Other signatories of this letter:

Christian Stoeckert University of Pennsylvania, Philadelphia

Catherine Ball Stanford University, Stanford, California

Alvis Brazma European Bioinformatics Institute, EMBL, Cambridge

Robert Gentleman Fred Hutchinson Cancer Research Center, Seattle

Wolfgang Huber European Bioinformatics Institute, EMBL, Cambridge

Rafael Irizarry Johns Hopkins School of Public Health, Baltimore

Marc Salit US National Institute of Standards and Technology

Gavin Sherlock Stanford University, Stanford, California

Paul Spellman Lawrence Berkeley Laboratory, California

Neil Winegarden University Health Network, Toronto, Canada

Taxonomy on the fly in a European web project

SIR — Donat Agosti, in Correspondence, argues that restricted access to taxonomic information impedes many countries in their biodiversity monitoring efforts (“Biodiversity data are out of local taxonomists’ reach” *Nature* **439**, 392; 2006). This problem resulted in the establishment of the Global Biodiversity Information Facility in 2001, with free online access currently to almost 90 million records (see www.gbif.net).

The European Commission is about to sign into existence a network of excellence to create the European Distributed Institute of Taxonomy (EDIT; see www.mnhn.fr/edit), an initiative of the Consortium of European Taxonomic Facilities (www.cetaf.org).

A component within EDIT will provide a freely accessible web portal to integrate taxonomic revisions on the fly, allowing users to build a resource tailored to their needs. If data are properly structured, then identification keys, distributional checklists and other ready-to-use products can be built on demand.

There is, however, a potentially negative

impact on the career development of taxonomists. Currently, performance is measured using publication metrics that rely on the copyright restrictions to which Agosti refers. EDIT is mandated to examine those restrictions and to relieve them where possible, both through abstracting information from copyrighted publications for summary access (without copyright infringement) and by developing new metrics by which institutions can assess taxonomists’ productivity.

Readers specifically interested in making taxonomic material openly available are invited to contact the coordinator (D.R.) of this particular component of EDIT.

Simon Tillier*, Dave Roberts†

*Muséum National d’Histoire Naturelle, CP 43, 57 rue Cuvier, 75005 Paris Cedex 05, France

†The Natural History Museum, Cromwell Road, London SW7 5BD, UK

dmr@nhm.ac.uk

Taxonomy: programmes developing in the South too

SIR — According to Donat Agosti in Correspondence (*Nature* **439**, 392; 2006), biodiversity data are beyond the reach of many taxonomists in the developing world. But since 1999, the State of São Paulo Research Foundation in Brazil has supported a research programme called BIOTA/FAPESP on characterization, conservation and sustainable use of the state’s biodiversity (see www.biota.org.br).

In the past six years the programme, for which I am a member of the steering committee, has produced an atlas of the remnants of native vegetation, supported 75 research projects, trained about 250 post-graduate students, catalogued approximately 10,000 species and made data from 35 biological collections freely available.

In 2001, the programme launched an open-access electronic peer-reviewed journal, *Biota Neotropica* (www.biotaneotropica.org.br), for original research on biodiversity in the neotropical region. And in 2002 the programme began a venture called BIOprospecTA (www.bioprosecta.org.br), in order to search for new compounds of economic interest.

Similar initiatives are under way in Mexico (CONABIO; www.conabio.gob.mx), Costa Rica (INBIO; www.inbio.ac.cr) and Africa (BIOTA Africa; www.biota-africa.de), showing that many developing countries are aware of their responsibilities under the Convention on Biological Diversity.

Carlos A. Joly

Department of Botany, Biology Institute, State University of Campinas, CP 6109, CEP 13083-970, Campinas/SP, Brazil
cjoly@unicamp.br