

Sharing science

The dawn of systems biology is apparent in many areas of cell biology and at its heart lie high-throughput datasets of increasing size and sophistication. These datasets contain a wealth of leads for hypothesis-driven research for the laboratory that generated them, but their true value is as a community resource. Although we will always require an exceptional level of biological insight beyond the listing of reams of factual information, part of the reason why *Nature* journals actively pursue the best of these studies is to ensure that the data and, importantly, the reagents, are publicly available. This is rarely a trivial issue, especially when the private sector is involved, and academic labs are also easily overwhelmed with requests for materials. When the small biotech company, Cellzome, published the first proteome analysis using the TAP-tag methodology (*Nature* **415**, 141–147; 2002), they were quickly overwhelmed with requests for reagents. Giulio Superti-Furga estimates that his group of less than 40 scientists were dealing with hundreds of requests for reagents, and consequently had to outsource reagent distribution to Euroscarf. It is unreasonable to expect individual laboratories to become the global distributors of new reagents or technologies they have developed, just as it is unreasonable to allow the over-commercialization of key published reagents. Our sharing-of-materials policy has been pragmatically adapted in light of these constraints.

A more systematic approach to community reagents is needed, especially given that smaller fields, such as the fission-yeast community, seem to be held back by the reliance on a piecemeal assembly of tools as a by-product of ongoing research. The establishment of an international non-profit service to archive, optimize, maintain and distribute key molecular tools in a manner analogous to the American Type Culture Collection (ATCC) would solve these problems.

This month we will publish online a study from one of the consortia formed to study systems level biology, the Alliance for Cellular Signaling (AfCS). A basic set of signalling parameters was monitored in a macrophage cell line in response to all pairwise combinations of 22 ligands, yielding a fairly global view of context-dependent signalling and pathway interactions. In this case, the raw data was made available, without delay, after validation, together with community resources, such as a database of antibodies used in the study. Projects like the AfCS or the Cell Migration Consortium (CMC) are sizeable inter-laboratory collaborations that were funded as part of the NIH 'glue grant' initiative. The aim of these large grants is to facilitate the systematic interrogation of complex biological questions — an important goal that is beginning to bear fruit. These projects generate large datasets and often ancillary databases. The data will generally be valuable long past the limited duration of the grant and it is less clear how this data, and indeed these databases, will be maintained to guarantee that the resources persist in a usable form. Traditional formats of publication are clearly not designed to preserve this information.

Indeed, databases are particularly at risk of falling victim to the constraints of traditional funding. They are often financed on the basis of research grants and assessed in competition with research projects. This is not constructive, as entirely different criteria have to be applied to evaluate the success of a database. For example, is it a unique and essential community resource? What are the access statistics? How is the data curated and validated? How comprehensive, detailed and interconnected is the data? Is data entered in a systematic, machine readable format? Databases that pass these criteria and emerge as authoritative and comprehensive community resources must remain open access and receive indefinite funding.

Financial pressures recently forced a number of databases to close, or to move behind commercial firewalls after being bailed out by commercial enterprises; included in this group are such significant resources such as the Biomolecular Interaction Network Database (BIND) and the Yeast Protein Database (YPD) (see *Nature* **435**, 1010–1011; 2005 and *Nature Biotechnol.* **24**, 115; 2006). The yeast community was rudely awakened when YPD, at the time the key resource for yeast scientists, suddenly charged several thousand dollars annually per user after being bought by Incyte (later Biobase). Since then other databases, such as Comprehensive Yeast Genome Database (CYGD) of the Munich Information Center for Protein Sequences (MIPS) and Stanford's *Saccharomyces* Genome Database (SGD), have taken over as *de facto* community repositories for the budding yeast community, while GeneDB, run from the Sanger Institute, is the accepted community portal for fission yeast. How perilous the future of core-information resources for a whole community can become is well illustrated by the current uncertainties about the future funding and hosting of GeneDB. The yeast databases also illustrate what the National Science Board, the policy advisory panel of the US National Science Foundation, has referred to as 'community-proxy functions' in a report published last September. This report correctly raises concern that the 'authority' to serve as a community resource is largely acquired informally and implicitly. We concur with the report in asserting that action is urgently required to select appropriate databases for funding on a stable, long-term basis. We also suggest that this selection should not be executed at the national level but rather in an international setting that reflects the origin of the research contained in the database. An international database panel should be set up collaboratively by the NSF, the new European Research Council (ERC), the Japan Science Foundation (JSF) and other Asian counterparts. This body should have the authority and resources to award indefinite funding to key community databases; it should set data standards and provide ongoing assessment and quality control; it should ensure that databases assimilate emerging technologies and that database cross-connectivity is maximized. Finally, and most importantly, it should ensure that databases remain open-access resources.

Further reading on <http://www.connotea.org/user/bpulver/tag/sharing%20science>