

SPECIAL REPORT

Databases in peril

Life-sciences databases are in crisis, say their operators, as funders keen to support exciting new projects lose interest in maintaining existing services. *Nature* investigates the scale of the problem.

A lack of stable funding is threatening biology's core databases. Unless funding agencies set aside dedicated grants, the fear is that researchers will lose access to information vital to their work.

Several major international databases and research centres, including the European Bioinformatics Institute (EBI) at Hinxton near Cambridge, UK, face funding cuts. And the outlook for specialist databases is even worse: more than half of the operators contacted by *Nature* say their databases are updated sporadically or not at all because no funding was available after their original grants expired.

"There is a funding crisis right now," says Rolf Apweiler, a member of the EBI and head of the UniProt/Swiss-Prot protein-sequence database.

"It's a paradox," adds Lincoln Stein, a bioinformaticist at Cold Spring Harbour Laboratory in New York. "The funding system assumes that projects have a lifespan of three to five years. But if biological databases are to do their job, they need funding for a decade or so."

Life-sciences databases have proliferated over the past decade, driven by genome-sequencing efforts and easy Internet access.

For some scientists, resources such as UniProt are as much a part of the basic research infrastructure as reagents and test-tubes. The EBI's website, for example, recorded 2 million hits on a single day this April. Researchers use the site to access everything from molecular structures to nucleotide sequences. Hundreds of smaller databases, often maintained by individual labs, focus on molecules and genes associated with particular functions or species.

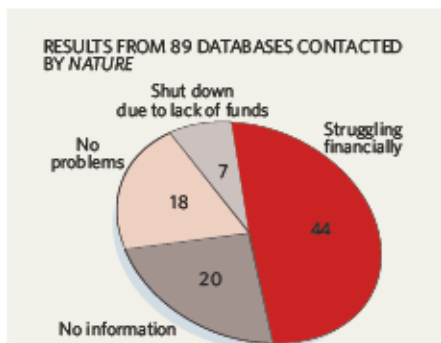
But as the number of databases mushrooms, many operators are finding that once their initial money has run out, funding agencies show little interest in helping maintain their service.

EBI director Janet Thornton is using the institute's reserves to support three databases whose funding has run out, including InterPro, an archive of data on protein families. "If we don't get new money we'll have to halve the number of staff on those projects," she says.

Thornton and others say the problem is particularly acute in Europe, where most grants are tied to original research. "Researchers feel like they have to invent new projects every three years to get money," says Thornton.

But databases in the United States are also feeling the pinch. The Alliance for Cellular Signaling, an ambitious ten-year attempt to amass data on the chemical signals inside cells, has scaled back its operations following a mid-project review. Funders at the National Institute of General Medical Sciences ruled last month that the project will receive less than half of the \$5 million a year it had asked for. The alliance says it will now have to shut five of the nine labs that are generating data from mouse-cell experiments.

Another key North American resource — the Biomolecular Interaction Network Database (BIND) — also faces cuts. Many journals, including *Nature*, routinely send their papers to BIND staff, who curate records on almost 180,000 molecular interactions (see page 1028). Last month BIND was forced to cut 33 jobs when a grant application to the Canadian government, which provided money to establish the database, fell through. Although BIND will continue to function thanks to money from the Singapore government, plans to integrate with other databases have been put on hold.



Survival of the fittest?

Nature contacted 89 databases listed in the Molecular Biology Database Collection (*Nud. Acids Res.* **28**, 1-7; 2000) to see how many still have funding five years on. Of these, 51 reported that they are struggling financially. Seven of these have closed; the rest are being updated sporadically in their owners' spare time.

Not all these databases necessarily deserve to receive continued funding, especially as technology advances and competing resources are set up. But bioinformaticists are concerned by the sheer number that cannot find the money to provide an ongoing service. **Z.M.**

"Canada is good at starting up projects like this, but there is no mechanism for continuing them," says Chris Hogue, principal investigator at the Blueprint Initiative, the Toronto-based organization that runs BIND.

Quest for novelty

The cutback at the Alliance for Cellular Signaling may be the result of a conscious change of heart by funders, but bioinformaticists say other cuts are part of a broader problem facing databases: agencies want to fund innovative and hypothesis-driven initiatives, rather than ongoing infrastructure projects.

"Long-term maintenance is expensive," says Carol Bult of the Jackson Laboratory in Bar Harbor, Maine, home of the Mouse Genome Database. She says it costs around US\$4 million a year to run. The resource is widely used and Bult is confident that funding will be renewed this year, but many other databases aren't so lucky. "We've faced this issue for a decade, but the funding agencies haven't caught up."

Smaller, cheaper databases are in even more trouble. *Nature* contacted 89 databases operating in 2000, and more than half said they are now struggling financially. Seven databases have folded, and many others are updated on an irregular basis as a labour of love by their owners (see 'Survival of the fittest').

"It is far more difficult for an individual researcher to obtain funding to maintain a database than it is to initiate a new project, even though constant updates are very important for the value of the database," says Ikuo Uchiyama of the National Institute for Basic Biology in Okazaki, Japan. He has had to temporarily halt updates to his Microbial Genome Database owing to lack of funds.

Students often end up filling the gaps. Will Ray of Ohio State University began working on PACRAT, which pre-processes genome data from the GenBank database, when he was a graduate student. "Development of PACRAT was squeezed out of my doctoral adviser's grant," he says. "But the hardware and system support were, and still are, donated to the university out of my own pocket."

Many warn that if the situation continues, future research could be severely compromised. "Every working biologist relies on these resources," says Thornton of the EBI's databases. "Over the next 20 or 30 years, the whole



Key services: databases are as much a part of scientists' basic equipment as test tubes and reagents

of biology will be built on protein structure information available through these databases."

The solution, say operators, is for funding organizations to set aside dedicated funding streams. Stein suggests that databases should be funded for longer than research projects, perhaps for five years, and that they should be judged separately. Bult adds that the system will need to be tailored so it can handle big projects such as the mouse genome database, as well as smaller, specialist archives that are used by just a few hundred researchers.

Any funding would also need to be vigorously peer-reviewed, and require databases to show they provide a service that is actually used. Bioinformaticists say that some databases lack community support and don't deserve continued funding. "You wouldn't want ten years of guaranteed support," says Stein. "That would encourage waste and drift."

Some agencies have already taken steps to provide earmarked funds. The Jackson Laboratory, for example, is applying to a panel of the National Human Genome Research Institute that focuses on large projects that need input from biologists and computer scientists. The Wellcome Trust, a London-based medical charity,

gives the EBI almost €2 million (US\$2.4 million) a year. And the next round of the European Union's Framework Programme will include a funding stream for infrastructure projects, including databases. This was originally set to distribute €3.5 billion, although that figure is likely to be cut substantially by the time the programme begins in 2007.

Operators say that a more general change in attitude is needed if biologists are to continue to enjoy the database services that they currently take for granted.

Apweiler suggests that databases should club together to make their voices heard. "They are small and not well-connected," he says. "They need to unite and demand community resources, and then perhaps the funding bodies will respond."

One success story cited by Apweiler is FlyBase, a resource partly funded by the US National Institutes of Health that links information on *Drosophila* from many databases.

But Apweiler warns that the money for the database comes at the expense of extra research funding in the field. "To increase these resources, the community needs to be willing to accept cuts in funding for individual projects," he says. ■

Zeeva Merali and Jim Giles

**IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS**