

INFORMATICS

Databases fight funding cuts

Online tools are becoming ever more important to biology, but financial support is unstable.

BY MONYA BAKER

In the era of 'big data', it is a bitter blow for scientists to lose access to the online tools they use to analyse and share terabytes of information. Yet funding cuts by the US National Library of Medicine (NLM) are threatening five widely used biological databases, and user communities are now rallying to save them. "The idea that this resource could just disappear is a serious problem for everyone who relies on it," says Mark Musen, a bioinformatician at Stanford University in California, and manager of Protégé, which provides open-source software to organize and interrelate biological data.

Protégé has 200,000 registered users, and the NLM, part of the National Institutes of Health (NIH) in Bethesda, Maryland, has contributed millions of dollars to maintain it. But in 2007, the NLM decided that it would stop supporting infrastructure grants and would redirect resources to informatics research, says Valerie Florance, director of extramural programmes at the library. Consequently, the NLM's support for Protégé and similar projects is not being renewed (see 'Endangered databases'). "It is not a reflection of the value of the resources to any of their users," says Florance. "It is part of our determination to put our funds into research and training."

The argument is playing out at other funding agencies, says David Botstein, a genomicist at Princeton University in New Jersey, and a member of the NIH Data and Informatics Working Group, which published a draft report on the issue in June. "The whole system is rigged against infrastructure of any kind," he says, predicting that "many, many resources" will face similar funding crises in the near future.

The Biological Magnetic Resonance Data Bank (BioMagResBank, or BMRB), for example, has been funded by the NLM since 1990 and holds more than 7,500 entries on biomolecules. Structural biologists use the nuclear magnetic resonance data to probe questions such as how proteins contort as they catalyse reactions.

More than 90 scientists have written letters to *Nature Structural and Molecular Biology* this month in support of the BMRB (J. Markley *et al.* *Nature Struct. Molec. Biol.* **19**, 854–860; 2012). Inês Chen, chief editor of the journal, says that losing the database would deprive researchers of access to crucial data. "As journals, we cannot host all the data that are part of the paper, and so if they disappear, it's a big deal."

John Markley, director of the BMRB and

ENDANGERED DATABASES

The US National Library of Medicine (NLM) is cutting resources that biologists say are vital to their research.

Resource	NLM-funded since	Function	Usage	Last NLM award
Protégé	1990	Creating tools to organize and analyse data	200,000 registered users	\$956,625
BioMagResBank	1990	Holds spectroscopy data for biomolecules	500–1,000 unique users per day	\$727,129
Repbase	1994	Identifying families of non-coding DNA across species	8,000 registered users	\$551,544
REBASE	1995	Finding where enzymes bind to and cut DNA	495,844 website hits per month	\$235,911
CASP	2001	Testing techniques to predict protein structure	More than 100 research groups participate	\$515,168

a structural biologist at the University of Wisconsin-Madison, hopes to attract other federal funders to support the database.

Another option is to charge users, but Musen calls that "absurd", arguing that it would discourage scientists from accessing sites and, in the case of Protégé, from contributing the code and plug-ins that make it a useful resource. Musen wants to win funding from the NIH to keep Protégé going as a key component of new research projects. In June 2011, he submitted a grant application with more than 100 letters of support from scientists; reviewers acknowledged the letters but said that they had nothing to do with the grant's specific research goals, and turned it down. Musen resubmitted the application, and should learn the results this month.

Other databases are putting their trust in commercial sponsors. REBASE, which holds data on where enzymes bind to and cut DNA, is partially supported by laboratory-reagent company New England Biolabs of Ipswich, Massachusetts. When federal money runs out in 2014, the company will take on the full costs, says Richard Roberts, chief scientific officer of New England Biolabs and founder of REBASE. But he acknowledges that this potentially leaves the database at the mercy of

shifting commercial priorities.

The least vulnerable databases are those directly run by government agencies, says Francis Ouellette, a bioinformatician at the Ontario Institute for Cancer Research in Toronto, Canada. Investigator-driven databases face more challenges because "they don't fit the research-based standard model" used to dispense grants. Cutting funding for poorly performing or obsolete databases is sensible, says Ouellette, but choking established sites that have significant user communities is "really short-sighted. If it's a good database it should be maintained."

Florance argues that the NLM should back innovation, which is difficult when its funds are tied up in infrastructure. "I don't think anyone would say that because they got a grant and built a database, they should get money forever."

One solution, says Musen, could be to wean successful projects off investigator-initiated grants and move them into the NIH's longer-term intramural programmes. But Botstein thinks that would require a philosophical change at the agency. "What's really required is an understanding of the larger problem," he says. "This is a big thing, and it will be a big thing for years to come." ■



**MORE
ONLINE**

Q&A



Astronomer Cui Xiangqun on China's plans for more telescopes and probes go.nature.com/zzhveu

NEWS

- The TB test you can do at home go.nature.com/wmtneq
- Controversy over widely used intravenous drips go.nature.com/eqlcj2
- Ancient humans interbred with Denisovans go.nature.com/jj8dxt