

nature genetics

volume 31 no. 3

july 2002

When public goes private

On 1 June, researchers using *Caenorhabditis elegans* and *Saccharomyces cerevisiae* were cut off from a valuable resource—unless they were willing to start paying for it. A couple of months ago, word started to spread that the proteome databases owned by Incyte Genomics (Palo Alto, CA)—which include the Yeast Proteome Database (YPD), the Worm Proteome Database (WormPD) and the Pombe Proteome Database (PDB)—would become a subscription service. Reaction in the respective communities ranged from “I told you so,” to disappointment, to irritation. Many researchers used the databases regularly for their work, as some of their features are not currently provided by the major databases that serve model organisms. Incyte’s decision to start charging for access to its proteome databases was undoubtedly a financial one—after all, the curators who painstakingly read through the literature to annotate the information in the databases have to get paid. But did the company take into account how the decision would affect its relationship with the research community?

The information in Incyte’s proteome databases is largely derived from published studies and, to some extent, personal communications from researchers; Incyte repackages the information into a useful product. It is understandable that many researchers, whose work is part of the databases, are irritated at being asked to pay for access. In addition to contributing data, researchers have helped to develop the database by reporting glitches or errors, under the assumption that the information would be publicly available.

Incyte’s databases, collectively referred to as BioKnowledge Library, contain comprehensive information about all characterized proteins of *S. cerevisiae*, *C. elegans*, and *Schizosaccharomyces pombe*, as well as information on human, mouse and rat proteins. The information in the databases, gathered by combing through the scientific literature, includes descriptions of proteins and their main features, and lists of related and interacting proteins. Extensive annotations are sorted into topics by expert curators, citing primary source and providing hypertext links to abstracts in the PubMed database. A nice feature of the databases is that they are built using the same format, allowing for cross-referencing and comparisons between them.

The databases, starting with the yeast one, were originally founded in 1995 by James Garrels of Proteome Inc. (Beverly, MA), as a public resource for the community. In December 2000, Proteome Inc. was bought by Incyte Genomics. At first, the take-over did not affect the research community, but earlier this year the company decided it would start charging US\$2,000 per lab (defined as a principal investigator and 5–8 other members) for a one-year subscription. Collectively,

Incyte's proteome databases contain more than 640,000 annotations from over 50,000 literature references.

Being cut off from the proteome databases concerns many researchers who work on model organisms. In response to this concern, the main 'competing' public databases are refocusing their priorities to provide more extensive protein information and annotation. J. Michael Cherry of Stanford University, who together with David Botstein (also of Stanford Univ.) runs the *Saccharomyces* Genome Database, says the upcoming 2002 Yeast Genetics and Molecular Biology meeting will provide an opportunity for them to assess community needs. Similarly, WormBase, the *C. elegans* genome database founded about 18 months ago, will now make protein function annotation a priority, according to co-founder Paul Sternberg of Caltech. WormBase has already doubled the number of its annotators and is polling the worm community to see what users liked best about WormPD. This comes at a time when there is a renewed sense of urgency for more collaboration among the public model-systems databases, including sharing of software and tools, and more standardized formats across the board.

It is too early to tell how many researchers will end up subscribing to Incyte's databases. The decision will depend mainly on whether a lab can afford the hefty subscription fee and on the value of access for students and postdocs. It is also not clear, at this point, how quickly other databases will gear up to provide some of the needed resources to users.

But what may turn out to be more important to Incyte is not the final number of subscribers, but whether the company will lose out on community feedback. Database development and maintenance is built on an open relationship with the research community—the greater the number of people who use and critique the site, the better it will be. Incyte's management does not seem to have taken this relationship into account when making its decision to cut off some of its users. And, despite the fact that various groups tried to discuss possible compromises or alternatives to Incyte's decision, the company apparently did not show any interest in discussion.

The availability of databases that pool genetic and protein data from a variety of sources is critical to progress in the era of genomics. The impact is enormous if many researchers have access to data and if there is cross-talk between databases. Just think where genomics research would be without the freedom to use information from the GenBank or EMBL databases. The lesson from Incyte's decision to turn a public resource into a paid service is how important it is to give ample support to public database efforts from the start. The lesson for Incyte is not yet clear, but it may turn out that cutting off the research community does not pay off in the end.

