The spider's web

Organized organisms

The internet has great potential as a repository of scientific data, but finding the information you need can be like trying to pick apart a frustrating knot; you can spend hours with nothing to show but the same knot and some broken fingernails.

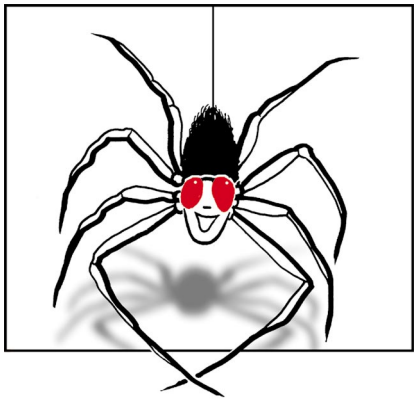
Fortunately, a few brave souls have taken on the task of organizing the tangled knot of biological information into web 'directory' links pages and online databases. If, for example, you need to find out about a particular experimental organism, web resources range from small-scale links pages maintained by researchers in their spare time, to enormous searchable databases maintained by full-time staff.

At one end of this spectrum lies the *Drosophila* FlyBase. This remarkable 'one-stop shop' brings together a wealth of information about the organism in a condensed, searchable form (though some users find the format confusing). The FlyBase Consortium, which manages the database, was founded in 1992 by five senior drosophilists and is funded by the US National Institutes of Health (NIH) and the British Medical Research Council. It takes 13 full-time employees to read the primary literature and condense the findings into a searchable format. The information is divided into various categories, including genes, aberrations, clones, genetic maps, vectors and stocks. The site has an 'Encyclopaedia of *Drosophila*', which integrates information from FlyBase and the Berkeley *Drosophila* Genome Project. FlyBase also allows for direct submission of some data that have not been published elsewhere.

Other fields have begun to follow the drosophilists' impressive lead. The Zebrafish Database Project — led by a group appointed at the 1994 Cold Spring Harbor meeting on zebrafish genetics and development — has recently obtained funding from the National Science Foundation (NSF) and the W.M. Keck Foundation for an expansive web database. The project now employs nine people, who are working on a database of neuronal, genetic, and developmental data. The database is growing out of a University of Oregon server, and will be up and running shortly. For now, there is The Fish Net, a relatively comprehensive site that includes searchable data from the Zebrafish Genome Project, references and news.

The Fish Net also does a great service to researchers working on other organisms — it includes a links page for other genome databases, from mouse to sheep to dog, and so on. Unfortunately, some of these links, though perfectly valid, connect to databases that are structured in difficult or archaic formats.

It may not be easy to obtain funding for such projects, but it is possible. The funding for the Zebrafish Database Project came initially from a new program within the NSF, launched at the beginning of this year. So far, 38 grants have been awarded for the design of computer databases, from organism-specific data, as in the case of the zebrafish, to data on nucleic acid structure and population genetics. Although the program is for computer databases, more than half of the grants awarded are for databases with an internet interface.



But large-scale web databases for scientific fields are a relatively recent phenomenon. Most organizing of biological information on the web is done on a spare-time, small-scale basis. One of the best of such sites for experimental organisms is the *Caenorhabditis elegans* Server run by Leon Avery from the University of Texas Southwestern Medical Center. The site, which is heavily used by worm people, contains thousands of abstracts as well as numerous links to other relevant web pages. It relies for much of its data on the gopher for the *Caenorhabditis* Genetics Center at the NIH. Maintaining the page, however, requires at least a half an hour of Avery's time each day, sometimes much more, and still, he admits, it is overdue for a major, time-consuming reorganization. He hopes to get some funding in the near future.

If they are maintained only through generosity and spare-time, web pages can easily fall into disrepair. The Mouse and Rat Research Page has not been updated in 18 months, though, incredibly, at one point it was among the top 5% of pages visited on the web.

Address: spider@cursci.co.uk

This month's URLs

FlyBase	http://flybase.bio.indiana.edu/
The Fish Net	http://zfish.uoregon.edu
Zebrafish Database Project	http://zfish.uoregon.edu/zf_info/dbase/db.html
List of Other Genome Databases	http:/zfish.uoregon.edu/zf_info/genome.html
National Science Foundation	http://www.nsf.gov
List of Grants Awarded for the Database Activities Program (go to NSF Award Search then search for Database Activities)	http://www.fastlane.nsf.gov
Program announcement NSF Database	http://www.nsf.gov/bio/pubs/bir_dba/dba9606.htm
Caenorhabditis elegans WWW Server	http://eatworms.swmed.edu
Mouse and Rat Research Page	http://www.cco.caltech.edu/~mercer/htmls/rodent_page.html