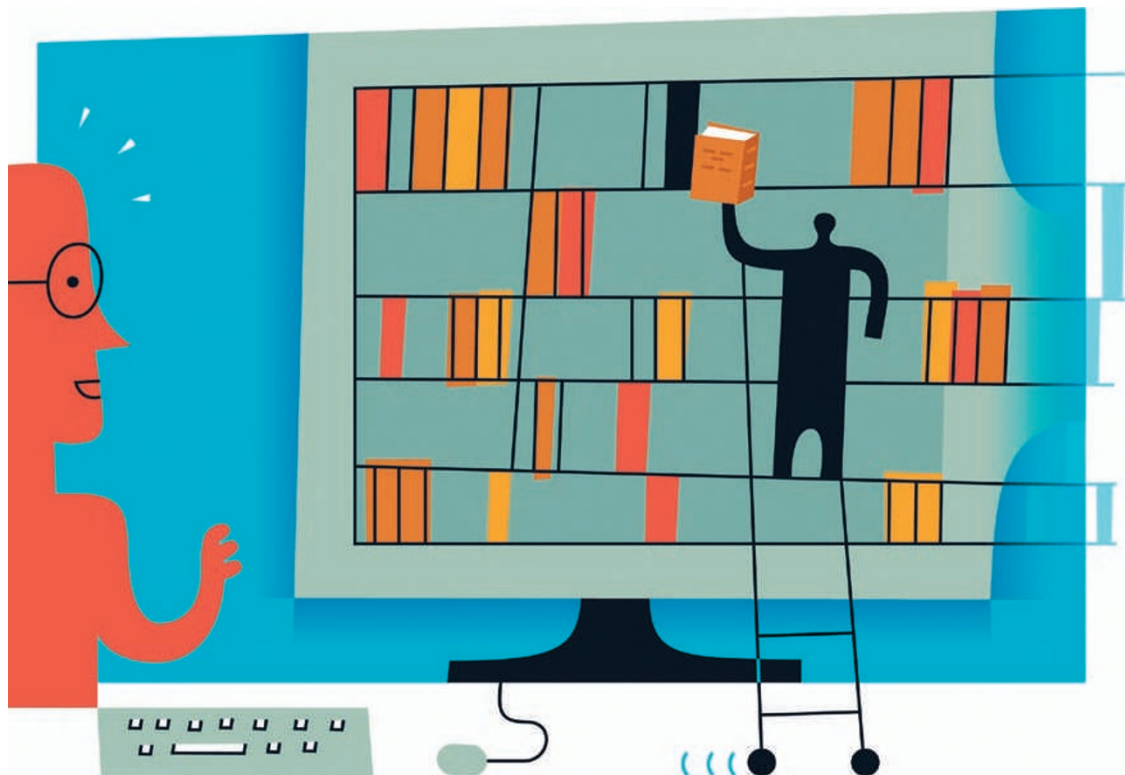


# CAREERS

**TURNING POINT** Turkish academic post fuels US molecular biologist's career **p.297**

**GRANTS** US National Science Foundation requires grantee data-sharing plan **p.297**

**NATUREJOBS** For the latest career listings and advice [www.naturejobs.com](http://www.naturejobs.com)



## BIOINFORMATICS

# Curation generation

*With biological databases growing in size and number, curators are needed to update and correct their contents. For those who prefer computers to pipettes, there are opportunities.*

BY KATHARINE SANDERSON

Biologist and self-confessed bookworm Klemens Pichler thinks that he has found his ideal vocation. Pichler is a biocurator at the European Bioinformatics Institute (EBI) in Hinxton, UK, working on the Universal Protein Resource (UniProt) database. Some scientists would find it onerous to spend their days reading papers and sifting through and cross-referencing data. Pichler sees it as satisfying detective work, with a well-organized database as the result.

Biocurators are an unusual type of biologist. Their job is to make sure that the data such as

gene or protein sequences entered into large biological databases are standardized and annotated so that other biologists can understand them. "Once you have generated a sequence and identified a gene, there is an enormous amount of pre-existing data that you search that gene against. You need an expert to refine that information and make it usable," says Owen White, a bioinformatician at the University of Maryland School of Medicine in Baltimore. White developed the first genome-annotation software in 1995, and has been involved in several high-profile genome-sequencing projects.

At present, the number of biocurators is small — the International Society of Biocuration,

founded in late 2008, has just 300 members who work at some 100 organizations. But the number is likely to increase as sequencing becomes easier and biological data continue to roll in. By July 2008, more than 18 million articles had been indexed in the PubMed biomedical database, and nucleotide sequences from more than 260,000 organisms had been submitted to the GenBank database (see *Nature* **455**, 47–50; 2008). Started in 2008, the 1000 Genomes project has added to the data influx.

Pichler started work at UniProt after completing a fairly typical early academic career path: a degree in biology at the University of Vienna; postgraduate lab experience at Harvard ▶

IMAGES.COM/CORBIS

► University in Cambridge, Massachusetts; and a PhD in virology at the University of Erlangen-Nürnberg in Germany, followed by a brief postdoc position there. It was during his postdoc that Pichler realized that he was on the wrong track. “I had grown tired of the frustrations of lab work,” he says. He read around and discovered biocuration; this was the change he had been looking for. “I’ve always been fond of computers but I never got round to integrating that into my career,” he says. Biocuration, Pichler found, was a way to make use of his training and move towards bioinformatics.

“It’s a wonderful career,” says Judy Blake, a bioinformatician at the Jackson Laboratory in Bar Harbor, Maine. Blake is a principal investigator on the Mouse Genome Informatics project, which employs 31 biocurators across multiple sites. She says that biocuration provides access to intellectual science without the stresses and responsibilities of finding funding and producing publishable results. Some researchers-turned-biocurators also relish the opportunity to be more of a generalist after academic careers that had a narrow scope.

#### PRACTICAL UNDERSTANDING

Although a PhD is not required, prospective biocurators need to be well trained in biology, with at least an undergraduate degree in a biological science and some related lab work. “Lab experience is important,” says Sandra Orchard, a senior scientific database curator at the EBI. “You can teach people curation but you can’t go back and teach them ten years at the bench.” Such experience helps biocurators to understand the data that they’re curating and how those data were generated.



**“You have to like reading and delving into matters, rummaging around looking for clues.”**

Klemens Pichler

Some universities offer specialist degree courses in biological information and the more software-design oriented bioinformatics, but none has a formal curation degree course specific to biological data. General data-curation programmes are available at the University of Illinois at Urbana-Champaign and the Digital Curation Centre in Edinburgh, UK, which offers short courses.

At UniProt, which employs almost 70 curators in Britain, Switzerland and the United States, Pichler spends half his time digging around to find out more about the protein sequences — the order of amino acids in a given protein — that are sent to the project from researchers around the world. He takes all the information he receives with each sequence

and compares it with existing entries in the database. He also does a thorough literature search. “You have to be a bit of a bookworm; you have to like reading and delving into matters and rummaging around and looking for clues,” says Pichler. He routinely scours the literature to find, for example, germane bits of information about the structure and function of a protein sequence. Next, he organizes and standardizes that information so others can interpret and understand it. “I concoct a new database entry, which then undergoes several rounds of quality control before it ends up being publicly available,” he says.

The other half of Pichler’s job is more technical, veering towards bioinformatics and software. He writes ‘rules’ so that computer programmes can annotate sequences with the structure and function of the genes or proteins. Researchers can then use these rules on their computers to predict protein function and structure from sequence data. Similar tasks are required for other databases, from those focused on gene-sequencing, such as Blake’s mouse-genome project, to efforts such as the Gene Ontology project, which aims to standardize gene representation across species.

The extent of curation depends on the database — the needs of a simple repository for information will differ from those of a comprehensive catalogue that combines information from direct submissions and published literature. Dealing directly with the scientists who produce the data — and can explain and modify the information on request — is easier than having to sift through the literature, says Orchard. “When working from a paper, you are dependent on it being well written in the first place and the data being complete and fully described. This is often not the case,” she says.

#### INTERNATIONAL COMMUNITY

Most large databases, and consequently curation jobs, are based in Europe and the United States, but that is changing, says Tadashi Imanishi, leader of the integrated-database and systems-biology team at the Biomedical Information Research Center in Tokyo, part of the National Institute of Advanced Industrial Science and Technology. The International Society of Biocuration has helped curators in Japan and other countries be part of the community. “By joining the society, they have the chance to communicate with curators in many other databases in the world,” says Imanishi, noting that Japan now has some 100 biocurators working on projects such as the DNA Database of Japan, which employs about 20 biocurators, and the H-Invitational, an international effort to catalogue all human genes.

At the moment, most jobs are at universities. But industry is beginning to offer biocuration services. For example, Ingenuity Systems in Redwood City, California, founded in 1998 by Stanford University graduate students, employs biocurators in its offices in Germany,

Switzerland, France, Britain and Japan. They look after the Ingenuity Knowledge Base, which the company claims is the world’s largest curated database of biological networks, documenting the relationships between proteins, genes, complexes, cells, tissue, drugs, disease and biological pathways.

Because of the skew towards academia, one of the biggest challenges to the growing field is its dependence on grant money. “Right now there is poor recognition for the value of curation,” says White. Funding agencies should



**“Lab experience is important. You can teach curation but you can’t teach ten years at the bench.”**

Sandra Orchard

factor the cost of curation into grants, he says, although this can be difficult given tight budgets and the field’s relative infancy. “We’re in a very, very competitive market and have to work hard to justify curation to agencies,” he says. Yet, he adds, “this kind of librarianship is critical”. Sequencing may be increasingly cheap and sequenced genomes plentiful, but without curation the data mean little.

Although long-term funding can be elusive, jobs can be lucrative. US biocurators in their first positions earn around \$65,000, says Blake — more than a postdoctoral researcher. In Britain, salaries start at around £31,000 (US\$48,000). And there is scope for advancement, says Orchard — a biocurator could end up running a database or training users. Curation also could be a doorway to computer programming and bioinformatics. Biocurators need not have any software-engineering expertise, but they do work closely with the people who write the programmes they use, and anyone interested in software design could move in that direction.

Blake says those considering a career in biocuration should know that it will move them away from the lab, which could pose a problem for those wishing to re-establish independent research, build a publication record or find grant funding. “None of these aspects is an integral part of the duties or outcomes of a biocurator position,” she says.

“There’s no doubt it’s a desk job,” Pichler concedes. But many don’t mind. They like the continued focus on science, as well as the occasional opportunity to attend conferences, give a talk or write an academic paper about their database, says Blake. “Curators,” she says, “do novel work that is required by everyone doing science.” ■

*Katharine Sanderson is a freelance writer based in Toulouse, France.*