

many fascinating things, but people in this profession rarely say 'no,' political scientist Langdon Winner of Rensselaer Polytechnic Institute in New York state told the House science committee in April 2003. "Indeed, there is a tendency for career-conscious social scientists and humanists to become a little too cosy with researchers in science and engineering, telling them exactly what they want to hear."

After the completion of the human genome project, the ELSI programme at the NIH evolved to direct significant support to four centres of excellence. The National Human Genome Research Institute, which provides most of the funding, says it hopes that these centres will each develop a sufficiently strong identity to support the growth of the next generation of ELSI researchers.

Several of these bioethics centres of excellence say they are trying to establish closer links with more laboratory researchers (see page 1104), and that their bioethicists will make themselves more readily available to provide rapid advice to faculty members. They see this as a way of fostering more ethical thinking in the design and execution of research, long before they would otherwise think of bioethics.

Researchers have to seek ethical advice willingly and then be prepared to act on it. There is no enforcement mechanism for bioethics

or sanctions; and, unlike Institutional Review Boards, bioethics advice rarely carries any weight in law. But that should not dissuade more ethicists from getting involved with the practical applications of their expertise.

Moreover, this kind of advice may turn out to serve a useful promotional function for bioethics itself. One part of this is simply to increase awareness of ethics; it is telling that half of the enquiries during the pilot phase of the first of these projects, at Stanford University in California, simply required lab researchers to be reminded of existing ethical guidelines. But much of the advice is provided in confidence, raising issues of transparency in the event that research practices violate ethical norms.

Notwithstanding concerns about excessive cosiness with scientists, an academic bioethics community that is too aloof won't help science or ethics. As bioethicists strive to be heard beyond their own community, and for their advice to be valued, these models provide one useful way forwards. ■

"Several of these bioethics centres of excellence say they are trying to establish closer links with more laboratory researchers."

Machine readability

A publishing initiative seems ready to make text mining simpler.

For many years Tim Berners-Lee, the inventor of the World Wide Web, has dreamed of machines being able to help humans use his creation. This would enable not only sophisticated search tools to hunt for words or phrases, but also for other engines to hunt for meanings and patterns. This 'semantic web' is being pieced together gradually. The latest step forward brings users of the scientific literature closer to that dream by enhancing computer access to the full text of the scientific literature.

Many scientists are used to the idea of data mining: the ability to plunder all the available databases to search not only for relevant nuggets, but also for unexpected combinations of data that reveal — or at least hint at — relationships and mechanisms. They are not so used to the analogous function of mining texts.

But some researchers have made a start. Biologists, for example, have developed software that explores open 'text bases', especially the PubMed database. They scan many publications in order to discover relationships based on phrases or sentences that, when analysed in combination, cumulatively link one object (such as a disease) to another (such as a molecule). At the University of California, Berkeley, the BioText project is being used to explore apoptosis, for example (<http://biotext.berkeley.edu>). At the University of Illinois in Chicago, the Arrowsmith software explores the causes of disease (http://arrowsmith.psych.uic.edu/arrowsmith_uic/index.html). And at the European Bioinformatics Institute near Cambridge, UK, the EBIMed retrieval engine explores protein-protein interactions (<http://www.ebi.ac.uk/Rebholz-srv/ebimed/index.jsp>).

But publishers have yet to develop a standard annotation of their

content that allows computers access to the full text. Earlier this month, the Nature Publishing Group launched a preliminary proposal for such a standard. The proposal is not a commercial product but rather a potential service for the community. It is open for comment and is not intended to provide a competitive advantage to us: on the contrary, it will only succeed if adopted by other publishers.

The proposal is the Open Text Mining Interface (OTMI), which was first presented at the Life Sciences Conference and Expo in Boston earlier this month. A description and examples can be found at http://blogs.nature.com/wp/nascent/2006/04/open_text_mining_interface_1.html. The proposal would make coded text freely available to all. If all publishers were to adopt this or some similar standard, the entire literature would become accessible for mining.

How does this proposal relate to publishers' various business models? 'Author pays' publishers would be able to use this approach to machine readability and help users find their content more easily. 'Subscriber pays' publishers would follow the Nature Publishing Group in making this version of the full text freely explorable by machines but unreadable by humans. (Charging for machine access across diverse publishers' firewalls would effectively make machine text-mining impossible.) The OTMI approach to encryption is to jumble up sentences, retaining semantic relationships as far as possible.

Critics will point out that this limits the machine readability too; for example, some proximity searching becomes impossible. But the subscriber-pays model is strongly supported in the marketplace. OTMI represents a potential compromise between business needs and open access. *Nature* and its publishers welcome feedback about this initiative, which should be sent either to nature@nature.com or to the above-mentioned blog. ■

"If all publishers were to adopt this or some similar standard, the entire literature would become accessible for mining."