

Data hoarding is harming proteomics

To the editor:

A commentary in the April issue by Prince *et al.* (*Nat. Biotechnol.* 22, 471–472, 2004) cites the difficulty of sharing data in the field of protein mass spectrometry. I am a relative newcomer to the field, having worked primarily on the analysis of DNA and protein sequences and microarray gene expression data. I was therefore surprised to learn that protein mass spectrometry lacks a public repository or even an agreed-upon standard for representing data.

However, I suspect that this lack of standardization is a symptom, rather than a cause, of the corresponding lack of publicly available protein mass spectrometry data. Microarray researchers were sharing data on the web in the form of tab-delimited text files and Excel spreadsheets long before the development of the data exchange formats and online data repositories. In contrast, protein mass spectrometry researchers to whom I have spoken nearly uniformly agree to share data only in the context of a collaboration.



This policy extends even to published data. I recently contacted the authors of an article published in *Nature Biotechnology*, asking to receive a copy of the mass spectra used in their study. I was told that the data set described in their paper is not yet available because they are using it for

further studies.

I will not speculate about why the field of protein mass spectrometry is so competitive. But I am certain that the development of methods for analyzing biological sequences and microarray expression would have matured much more slowly without a culture of scientific openness. Scientists should freely post their published mass spectrometry data sets on the web, and funding agencies and journals should require such publication.

William Stafford Noble

Department of Genome Sciences, Department of Computer Science and Engineering, University of Washington, Health Sciences Center, Box 357730, 1705 NE Pacific Street, Seattle, Washington 98195, USA. e-mail: noble@gs.washington.edu

Nature Biotechnology responds: It is the policy of this journal, and of all *Nature* journals, that any supporting data sets for which there is no public repository should be made available to any interested reader on and after the publication date from the authors directly. As a condition of publication, authors are required to make all materials and methods used promptly available to researchers for their own use. They are also required to state in the Methods section of their published manuscript any conditions for use of materials and to provide full disclosure of the conditions to the editors at submission and on a freely accessible, identified web site after publication. Authors of published papers are allowed to charge a small amount for materials distributed to

other researchers to cover the authors' costs in producing the materials. Those who refuse to share data required to reproduce or build on the findings of a paper published in our pages are thus in breach of our guidelines and behave contrary to the spirit of scientific conduct. In this regard, we applaud the new rules introduced by the US National Institutes of Health (NIH, Bethesda, MD, USA) this month to require grant applications to outline a specific plan for sharing model organisms and materials that may result from the funded research (*Nature* 430, 953, 2004). Perhaps most importantly, this plan has teeth: an applicant's track record in sharing materials will be taken into consideration when their NIH grant is up for renewal.