

nature genetics

volume 16 no. 3 july 1997

The internet strikes back

SDI no longer refers to President Ronald Reagan's star wars project (properly called the Strategic Defense Initiative) but to Selective Dissemination of Information, which is meant to defend people against information overload and frustration with what is increasingly being referred to as the World Wide Wait. The idea is not new, but commercial interests in the electronic media have led to the recent explosion of so-called 'push technology' which may revolutionize the way we use the web over the next few years.

With more than two million different Uniform Resource Locators (URLs), it is getting harder and slower to 'pull', that is to search for, find and download, desired information. Imagine the convenience, if someone knew exactly what you wanted, and could 'push' it to your desktop. You would simply define specific areas of interest and where to find relevant information, and, on the basis of this customized profile, information would be regularly delivered to your computer.

This is not science fiction. 'Push' has been announced as the next big thing and is expected to generate an internet for couch potatoes. 'Webcasting' on personalized 'channels' is beginning to turn active surfers into passive viewers. Commercial information providers have realized that information pushed at customers is more likely to reach them than information they must go and pull out. In a potential market of several billion dollars, competition is fierce and developments are fast. And whereas, at the moment, leading 'pushers' like Pointcast (with an estimated 1.7 million customers) concentrate on general-interest areas such as weather, sports and business news, the technology clearly has potential to help researchers keep up with increasing amounts of scientific information.

Scientific alerting services, in fact, have been around for a long time. Many of them have made the switch from mailing journal clippings and tables of contents to delivery via e-mail, and like the commercial web giants, they have noticed and adapted to the increased need for customized information. 'Personal Alert' is a service introduced last year by the Institute for Scientific Information (ISI), the publisher of *Current Contents* and the *Scientific Citation Index*. For an annual flat fee of \$395, customers can choose a daily or weekly email update on papers published in their particular area of interest. Personalized profiles are based on keywords, author names and citation data.

In order to cover anything published in the field of diabetes genetics, for instance, a profile would include a number of keywords. Those could be grouped into 'diabetes' and 'insulin' on the one hand, and 'genetic', 'chromosome', 'link-

age, 'positional cloning' and 'hereditary' on the other, while searching only for papers that match both groups. In addition, one could ask to be notified of all papers citing oneself, or those that are published by one's fiercest competitors. Abstracts of the articles are included, and profiles can be adjusted as often as necessary to meet changing needs.

How many researchers would find such a push-deal attractive? According to ISI, 450 customers have signed up so far—not a great deal of business over the course of a year. Is it that scientists just don't like to be pushed? That may be true, but on closer inspection there are a number of other possible reasons why this and similar services have not yet revolutionized the way scientists are coping with ever more papers in ever more journals.

One of them is the problem of timeliness. Most scientists are news-junkies, and whereas it is now possible to look at the table of contents of the latest issue of most journals on the web, alerting services take more than a week after reception of a print issue before they inform their customers of its contents. Also, the service is not cheap. Would the money be better spent for subscriptions to journals in the field? After all, while a few papers of interest might appear in more obscure publications, the vast majority are published in a few familiar ones. And with web sites displaying current or even future table of contents, scanning the familiar ones is easy and library trips can be minimized.

Our hope (and that of other scientific publishers), is that there is another important reason. It may just be that scientists actually like to browse. That they have realized that even the most sophisticated filtering system will blindfold them, and that whereas pulling for information becomes harder and harder, the

real price for push technology is losing the pleasure of serendipitously finding a paper that tickles one's fancy for no good reason, or for a reason that will become apparent only with hindsight.

This is not to say that push technology does not have its merits; biological databases, in fact, have started to make use of it already—SwissProt from the University of Geneva, for example, and XREFdb from the US Center for Biotechnology Information both offer an e-mail alerting service where

subscribers get updated information on their specific searches. There is no doubt that developments in commercial push technology will improve the distribution of scientific information, and they might turn out to be crucial to ensure adequate dissemination of the exponentially increasing mapping and sequencing data generated by the various genome projects.

At the same time, let's not forget that scientific breakthroughs often connect facts that previously seemed to have little or nothing to do with one another. Looking only at information pre-defined as relevant is not the best way to push creativity.

