Physicists Write Too Much

In 1992, over 160 000 articles appeared in respected physics journals — more than 400 a day. As a result, scientists are less and less able to digest their own production. Yet they need to keep up, to learn of possible relevant developments, and to make sure that they are ahead of their competitors.

The problem of the overproduction of scientific literature was the topic of a thought-provoking symposium organized on March 20 by the Foundation for Fundamental Research of Matter (FOM), a scientific funding organization in The Netherlands. The aim of the one-day symposium entitled *The information inundation of physics* was to discuss the difficulties arising from the growth of scientific literature. Speakers and audience represented the various interested parties: some 50 leading scientists, information brokers such as publishers and database producers, and policy makers attended. One way of tackling the information excess is to use literature databases, such as INSPEC in Britain or FIZ in Germany. After defining a set of keywords, the database produces a list of potentially relevant articles. Some scientists have very detailed keyword listings. For example, Pieter Kruit of University of Delft showed the audience his elaborate INSPEC instructions, which look like a computer program, defining amongst other things the characteristics of articles that he did not want to obtain, but had erroneously received during previous searches.

But databases still have significant drawbacks. One of these is demonstrated by Kruit's intricate instructions: the computer gives many references which prove to be irrelevant, after laborious trips to the library. A much more serious obstacle was defined by Jook Walraven from Amsterdam University, who felt that the information excess represented the greatest barrier to entering a new sub-field of physics. Walraven explained that scientists know what is happening in their own speciality because they receive preprints and visit conferences. An immediate judgement can be formed of new work in a known area, and even a surplus of references in computer searches is easily weeded out. However, there is no "gut feeling" for publications in other fields, and immediate judgements cannot be made. It is difficult enough to find out what the articles are about.

This issue became one of the main concerns formulated by scientists during the course of the symposium: the fraction of publications which one can even vaguely understand is rapidly decreasing. A member of the audience described the dilemma as follows: "My specialty is defined by the publications that I can understand. It is a frightening experience that this is a rapidly decreasing part of what I read."

A Bright Future?

After the scientists, the second group of speakers at the FOM symposium were publishers and librarians involved in developing the technology to access and retrieve information. The tone during this part of the symposium was one of excited anticipation: the future looks bright. Technically, we are not far removed from a personalised system which allows us to retrieve, combine and annotate information, all in the same (virtual) environment. Each scientist can have access to single documents from all possible sources, which as "experimental conditions" or "sample preparation". To be able to properly identify which independent units of information exist in a scientific article, Joost Kircz of Elsevier Science Publishers suggested performing a semantic analysis of scientific rhetoric. Eric van Herwijnen of CERN showed that the coding of such modules is possible today using the Standard General Mark-up Language.

Jook Walraven explains the scientist's view. In the foreground, Joost Kircz of Elsevier Science Publishers.

Once information is coded into modules, it can be distributed and accessed in a multitude of ways. A computer search can retrieve individual units; even references could be made to these bits of text. Arie de Ruiter from Elsevier Science Publishers showed how information units can be combined in a Hypertext environment, which can be built into a database of information units and annotated and linked by the person who will use the information.

At this point in the symposium, a scientist in the audience stood up and said: "this looks wonderful, and as soon as it becomes available and affordable, I'll be the first one to use it. But I would first of all just like to have the existing journals on-line in my office. And I'd like to be able to read them."

This sums up the two essential problems that scientists and publishers face. The first problem is that scientists need to reverse the trend of continuing specialisation. This cannot be solved by technology; without it, in fact, technology is virtually useless. The only remedy here is that physicists take the matter to heart, and spend time and energy to explain their work to non-specialists.

The second problem is the need for the development of efficient ways to access existing literature. Issues of economy and standardization that still stand in the way of developing an effective system of electronic publishing should be addressed.

Revive a IUPAP Commission

A resolution will be submitted by the Netherlands Liaison Committee at the next IUPAP General Assembly in Japan in September to address this last point. The Dutch propose to reconstitute IUPAP's Commission on Publications, originally established in 1931 under Robert Millikan but placed on stand-by in 1987. The Commission should coordinate and investigate not only developments related to publications on paper, but also electronic communications. The IUPAP could play a crucial rôle in standardizing methods of electronic publication and communication, and in developing standardized approaches to indexing and retrieval.

Anita de Waard
Elsevier Science Publishers