Nobel Symposium (NS 120)
"Virtual Museums and Public Understanding of Science and Culture"
May 26-29, 2002, Stockholm, Sweden

All About Electronic Scientific Publication

Hector Rubinstein
Stockholm University, Stockholm

Why is it worth discussing?

As early as in Greek antiquity scientists wrote books. Later, scientists exchanged information by letters and even later results were presented at meetings of the recently created academies. Scientific journals have been published since the second half of the 17th century.

It took time until the peer review system was established. At first the communications were accepted and presented by leading academicians. Later, a more formalized procedure developed which then evolved into the present structure. Proliferation of scientific research, already significant in the middle of the 19th century, accelerated enormously after the First World War, and even further after the Second World War.

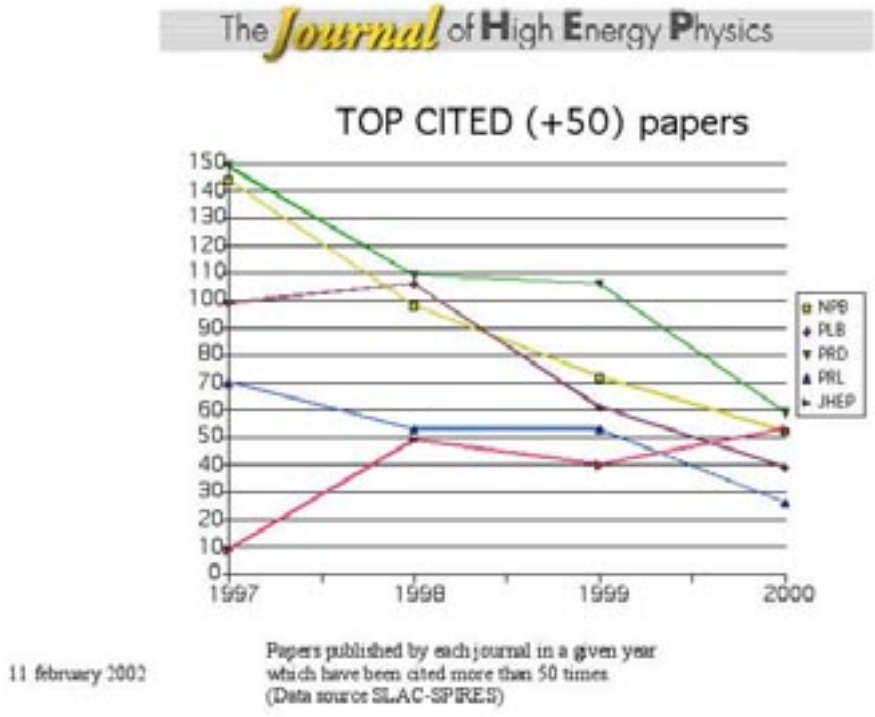
It is around this time that book publishers entered the field of journal publication, until then controlled by the learned societies. The rapid growth of research and the abundant resources in the exact sciences gave great impulse to libraries. As a consequence of continuously increasing budgets libraries expanded vigorously. They were often run with economic largesse reflecting, the aftermath of Sputnik. Private publishers proliferated after the fifties.

However, as in other economic branches, the eighties saw the rapid appearance of conglomerates and a concentration of journal publication in a few large companies. An example of this is *Elsevier* Science who first bought up *North Holland*, then *Pergamon Press.*

In 1999, *Elsevier* Science had a turnover of one billion euros in basic science journals alone. They publish more than 1200 scientific journals. Their relentless growth continues. Recently, they also acquired *Academic Press.*

Some intended takeovers or mergers (*Kruver*) were not allowed by governmental action. The merger with *Reed* was, however, approved and *Reed-Elsevier* are now a most formidable conglomerate, of which *Elsevier Science* is a branch.

I have used *Elsevier* as an example since I worked with them for many years. Let us now study the economic results of the dominance of private publishers. *Elsevier*, *Springer* and others - all have similar costs per page. Others, which claimed to be user friendly are not different. No private publisher I know holds to a reasonable price level.This is a generic situation with private publishers for whom dividends for shareholders is the driving force. If we add the revenue of these companies to that of the non profit Societies we will reach library expenditures, just for journals, going beyond 2 billion euros per year. This sum does not of course, include the cost of the infrastructure necessary for handling and housing these thousands of journals, which need an estimated one hundred metres of added shelving per year.



The business of distributing and accumulating information in Science alone uses the order of 4 billion euros a year. If we include other professional subjects like legal and some forms of business journals,

computer business journals and other fields, the figures run even much higher.

Another way to compute these costs is to use a few medium sized university libraries and average their costs. Indeed, each one has a running budget of a few million euros a year. These costs to the libraries are very disturbing. The prices charged by these private publishing houses have increased tremendously. Some of the commercial journals have increased their cost per page by a factor of 5, while inflation has given less than a factor 2. This is just since 1984, no connection to Orwell.

First, to avoid going into gritty details it is enough to point out that *Reed-Elsevier* reported in 1999 a whooping profit of 351 million euros on a turnover of 1000 million euros. This does not include the handsome salaries paid to the executives. Including perks, they total 1,000,000 euros a year. One can ask at this stage, what is the contribution of the scientists and publishers, respectively, to the product?

The scientific work leading to the papers is wholly the work of scientists, the peer review is also done by scientists. The authors receive no economic compensation and referees have been given (sometimes) handouts entirely out of proportion to their effort. Remember that book authors receive 7 to 15 % of the price.

In the development of electronic publishing, it is again the scientists who have done the work and had the ideas. Inspired by ARPANET, CERN scientists developed the web; people like Ginsparg developed the automatic databases, I started similar things at *Elsevier*, Stanford University created *Qspires*, which is a citation database, and so on… The publishers, until recently at least, have only produced the paper version and more recently an electronic counterpart, and provided the necessary secretarial office. These companies own the copyright to the work of the scientists.

Establishing the exact cost of a publication produced by a company with multiple activities is difficult. It is reassuring that the price of journals published by learned societies is lower. Moreover, even if the societies do make a profit, it is not distributed as dividends but returns to science. I will come back to this point.

Before leaving the economic aspects, it is possible to have an idea of cost to the community as a whole by defining an index that I presented last

year at a UNESCO conference on scientific publications. I called it the Ginsparg index in honor of the true inventor of electronic databases. If one considers the total revenue of a journal and divides it by the total number of submitted papers or published ones (rejection rates in good journals is about 25%), one gets a cost per paper to the community.

It is in this atmosphere of concern due to skyrocketing prices that a hard look was taken vis-a-vis the process of gathering information. Articles in leading newspapers such as *Le Monde*, the *New York Times*, *The Herald Tribune* and scientific journals like *Nature* have protested against the prices and the monopolistic nature of the business, and even boycotts have been organized.

Finally, it is worth pointing out the necessary inefficiency of scientific publication. Even in the best journals, only a few articles are read widely. Most articles discuss a technical point, sometimes very important but understood or valuable only to a handful of people. To print these huge numbers of copies is inefficient and bad for trees.

The future clearly belongs to the purely electronic journals.

The first stage of the acceleration in the demand of knowledge was the appearance of preprints. Authors first printed their results themselves and eventually many appeared in journals. Then came the Ginsparg database, which I will describe here very briefly. First, it is accessible to anyone for free. As a consequence the very existence of paper journals seems threatened. In fact, are they necessary? This leads us to the basic questions of my presentation:

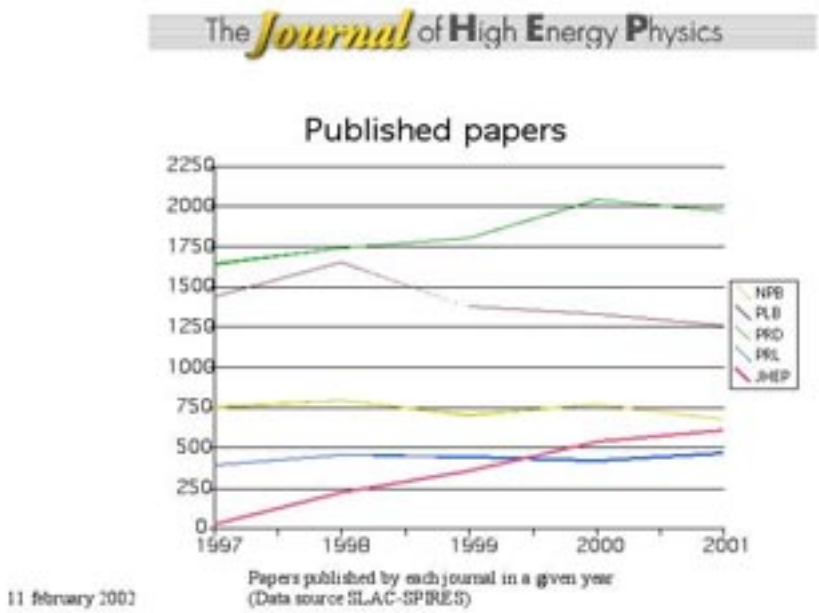
a) Are journals needed?
b) What should be the cost?
c) Who should control them?
d) Is electronic archiving safe?
e) What is the relation between journals and databases?
f) How does a modern journal work?
g) What is the future?

We start with a).

Given free databases what is the added value of a Journal? Obviously an active scientist first thing in the morning connects herself/himself to xxx.lanl.gov. (I am now focusing on physics.) There she/he sees the 100 (about) new papers of the day from astrophysics to condensed matter.

The catch is that there is no quality control. First, sometimes extremely speculative papers, wrong papers or even unscientific papers appear in the lists. There is a danger that this type of contribution would saturate the system. Ginsparg does some preliminary filtering, but his is not a safe system.

If someone were to submit a paper from the Bern patent office, the system may reject it. Moreover, even if you are surprised to hear it, people find it reassuring that one leading scientist at least, the referee, reads the paper carefully and points out errors or praises it. Peer review gives this guarantee.



Second, refereeing improves papers. In reviewing a specialist can suggest changes or may look into some related point that could invalidate or improve the paper. In fact, in journals where I have been or am currently involved, 40 % of the papers are altered by refereeing.

Third, a good journal is nice to browse. You know that the material has been screened, and in all probability it is relevant.

Fourth, people trust good journals. Promotions and contracts are very much influenced by where your work is published. In many good universities, promotions and tenure committees are interdisciplinary and often the question is asked where the work of the candidate has appeared.

Thus, peer review entailing discussion with a knowledgeable referee or editor is a factor that would seem to justify the existence of journals.

Systems of open refereeing and chat sites have been proposed and even tried, not successfully. Secret appraisal by some authority still seems the best method.

This leads us to b).

b) What should be the cost?
Obviously, the added value discussed in a) is to "certify" the content of the database material and improve it somewhat via peer review. There are side bonuses, such as putting it together under a quality label, preserving it, and providing for sophisticated search engines to help the researcher find related papers. It is important that cost be proportionate to the improvements brought about by journals. More precisely, what is the added value?

Producing databases costs. There is a tendency for scientists to think that the xxx Ginsparg database is produced without cost since it is free to the user. In fact, Ginsparg estimates the cost to be 30 dollars per paper. He is financed by the National Science Foundation. I am not sure this covers long time archiving and research but it is a good number to start with. The cost of the commercial journals is not easy to determine. Since they are profit making commercial companies they do not disclose the number of subscribers to a given journal.

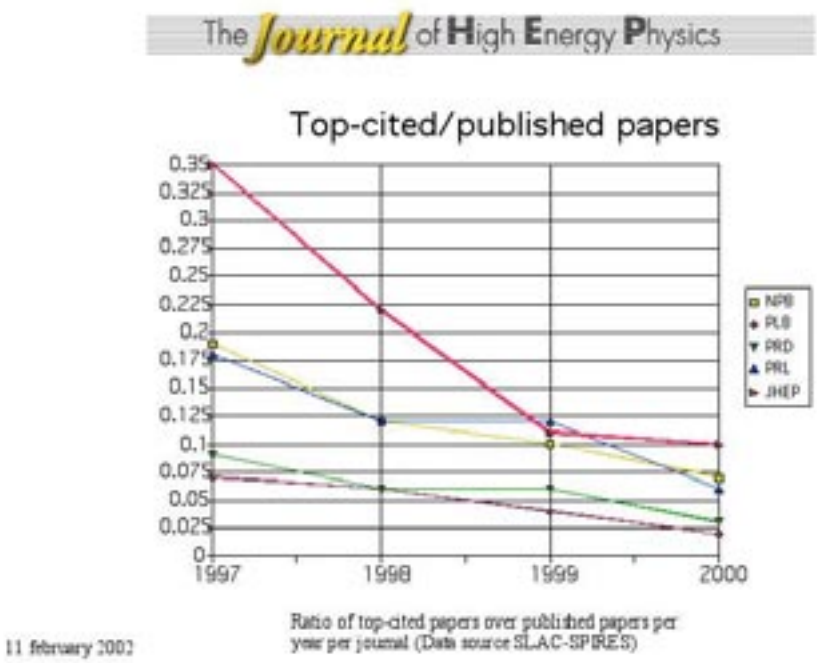
Forty years ago individuals used to subscribe to journals. Nowadays, a one year subscription to e.g. *Physics Letters B* costs 15.000 euros so, of course, the only possible subscribers are libraries.

It is clear that a few years ago leading journals had about 700 subscribers per journal. These numbers have been declining for a few years. The publishing companies then invented the method of consortia subscribing, thereby forcing libraries to buy the whole set of journals in a given field, forbidding cancellations of single journals.

A journal like *Nuclear Physics B* has a revenue above 10 million euros a year and about 600 papers published. Its Ginsparg index is then 13.000 euros per paper.

The *Journal of High Energy Physics* (JHEP, soon to be described) already has more papers, and costs (at the moment it is free to the user

like a database) 350.000 euros a year. Its Ginsparg index is thus 500 euros per paper. A shocking difference, though not surprising in view of the profits described above. To add, JHEP, as seen below has the highest quality in the field measured by impact or citations.



Let us look briefly at the history and workings of JHEP. It was initiated at SISSA, the Italian Institute for Advanced Studies in Trieste with the dual aims of modernizing publishing and lowering costs. The debate centers on whether the cost of commercial journals is justified. At a fraction of the cost of a commercial journal, JHEP has become a leading if not the leading journal.

This will become clear in the second part of this presentation. Remember also that by going *purely* electronic, great savings occur at the libraries as well. Shelves, personnel, protective measures, and new buildings are very costly. There is sometimes the added inconvenience of having to go to another place on campus to look for a given journal. As an example, a typical high energy physics library requires today 100.000 euros to be fairly complete. A better service can be provided at a fraction of that sum.

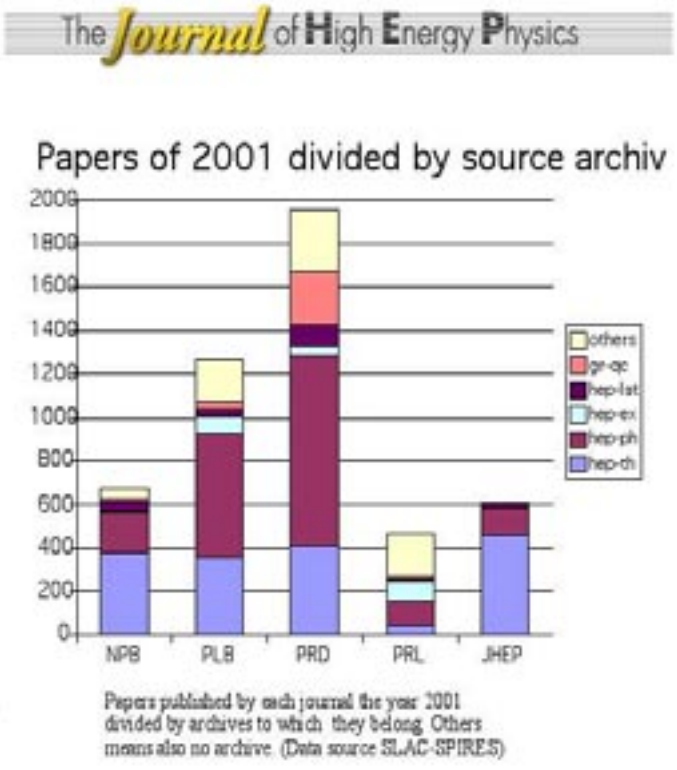
c) Who should control journals?

This point is a delicate one and not every one may share my views.

The point applies both to the learned societies and to the commercial publishers. The producers, reviewers, and users of the scientific materials being one and the same body, the question can be asked, “who should control the journals?”

I refer in particular to copyright, quality and price. One must remember that we have a monopolistic situation. If you don't like a Volvo, you buy an Audi and more or less you get the same product. If you need a paper that is in a copyright protected journal, access may be denied unless you buy a package. This is forcing the libraries to buy even what they do not want as well as not being able to do selective cancellations!

My answer is simple: control must be in the hands of active scientists and whenever controversies arise, leading scientists must resolve them, not publishers.



d) Is archival safe in the electronic era?

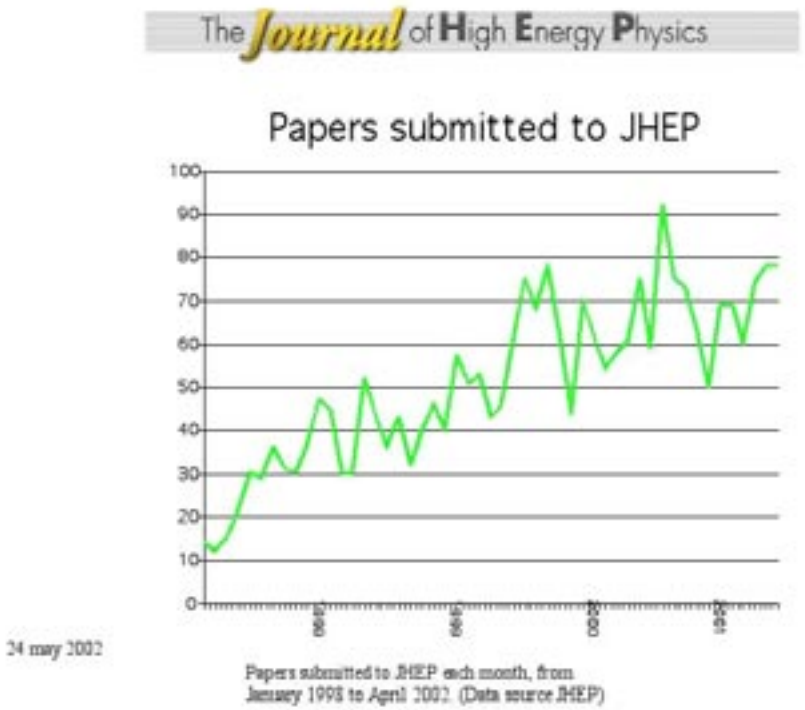
This is a point where much suspicion exists among librarians and other people. What will happen with these electronic papers? Will it be possible to read them in 100 years? This is an important point, and the answer is already with us. Pragmatically, we know that xxx

has been working for more than a decade without a hitch. We believe that present research, for example at SISSA, separates the documents from the formatting, thereby solving the problem completely. Also, it is worth mentioning that paper journals, specially the ones produced since the beginning of the 20th century are disappearing because they have been produced on acid paper. They are being optically read for electronic preservation! Not to speak of space savings. Retrieving an old article is frustrating. No search is possible on paper journals.

e) What is the relation between journals and databases?
The relations between xxx and the JHEP are immaculate. People can submit directly from xxx and we do not object if authors put back into the net their refereed version after peer review. One again asks, “so what are journals for?”

As explained, the uniformizing, reviewing and concentrating of the material in a journal has value. Even rejecting papers sends a signal. Also, I repeat, every scientist wants his work to be looked into by a leading person. We invite you to browse JHEP and use its retrieval facilities to confirm this point. Given the cost, we believe it is worthwhile.

f) How does a modern journal work?
It is worthwhile to describe JHEP in detail since its success shows it is a viable alternative. First and most important it is paperless. When a paper is submitted, a robot chooses an Editor from a set of keywords the author must provide. There is a reasonable overlap among Editors to make it formally impossible to know who is the Editor for a given paper.



The Editor is alerted by email that in his editorial page there is a new paper. The Editor can either assign a referee or take the paper on his own. In the first case, he sends the information to the referee, who is also told there is such a paper with him. He/she gets the manuscript in full and is given a few weeks to act. The robot will automatically remind him/her if he/she falls behind. Finally, the paper comes back to the Editor with a recommendation. The Editor may decide that he needs more information and acts. Also, if the author sees the paper rejected, he/she generally protests.

Most of the time these problems are solved at the Editorial level.
I am the director of JHEP, but I intervene in very few cases, mainly advising the Editor what to do in a thorny case. Acceptances are something like 72%. Once the paper is accepted after revisions it is automatically transferred to the office. The paper is “massaged” to look like the Journal and this is the end. A few seconds after that, the paper is on line.

There are some problems. The time taken by the referees is definitely increasing. Because of the database people feel no urgency, nor do they feel it is that important. This is a problem and we have not yet found an answer. The office is run by four people, including secretarial work and

emailing (big traffic), maintaining the system, as well as intervening in cases where I am consulted.

We have had only one case where the authors went to the Advisory Board whose members I consult in difficult cases. Loriano Bonora supervises the running of the office with Simona Cerrato. They also look into future projects from the technical point of view. The present cost of the office was revealed above when we wrote the Ginsparg index: about 300.000 euros p.a., including payments to the editors. The journal receives indirect subsidies from SISSA, like using their office space and network amounting to about another 50.000 euros p.a. I present below a few graphs with some spectacular data and explanations. It is a rapid overview of the Journal that can also be seen directly in its homepage (http://jhep.sissa.it).

g) What is the future?

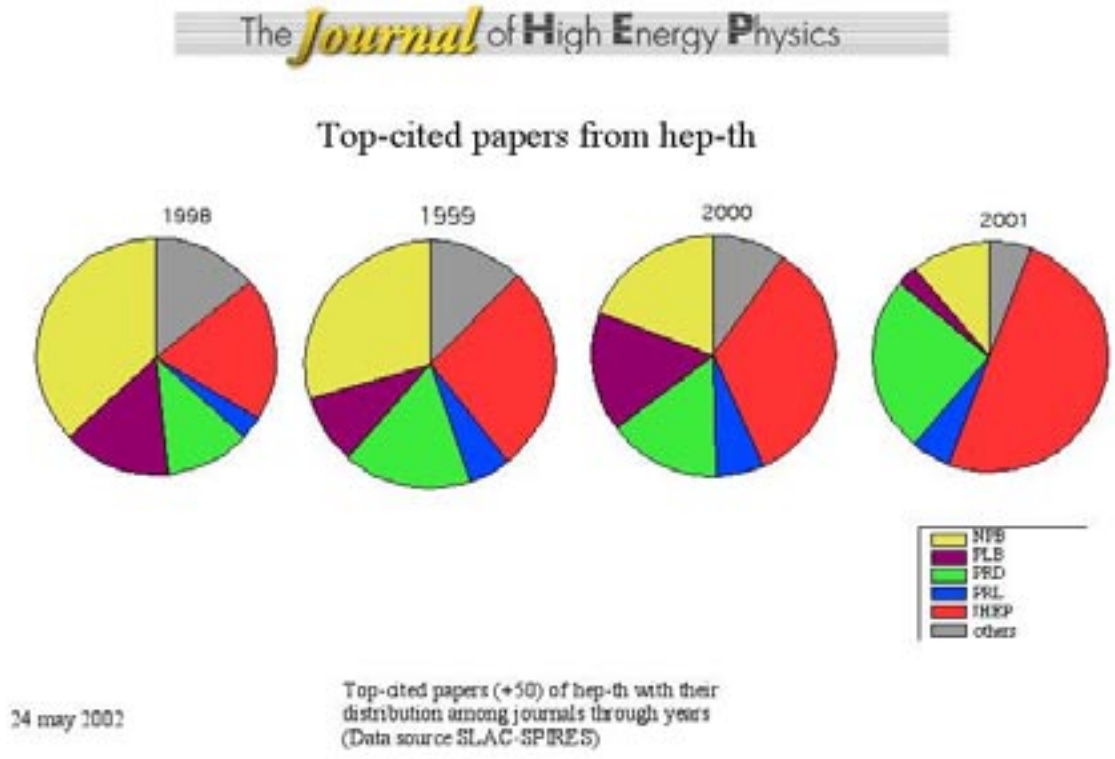
As the Americans say: there is no such thing as a free lunch. JHEP costs. Running, development, computers, personnel, Editors, space, creating new products. It is probably a reasonable estimate to double the cost presented if we include developing, marketing and other matters. This is still a fraction of the cost to the scientific community of the commercial journals.

If we get 500 subscriptions after a few years at 1000 euros a piece, we seem to be in a good position. Developing countries will continue to have free access. The model we have developed together with IOPP (the Institute of Physics Publishing) is that SISSA will keep ownership, copyright and scientific control. IOPP will run the technical aspects of distribution, like the web site, and market the product as well as helping on new developments. They will have a fair return for their investment at a price that is jointly determined by them and SISSA.

Moreover, we will jointly expand on new ventures of the same style, e.g. A journal on *Astroparticles and Cosmology*, which was launched yesterday; It will have modern technology, allowing movies and simulations to take place. A journal for conferences in which participation from remote places will be possible, a review journal on topics in physics and others are being considered.

Roughly speaking, a full library in a field like particle physics, cosmology and related subjects costs a library 100.000 euros a year. We claim that having three or four independent journals of each type like ours and *Physical Review*, etc will bring the price down to at most 20.000

euros p.a. and be of the same or better quality. This situation should propagate to other fields, like biology and mathematics. We are trying to help this happen and other scientists are working for similar goals.



As a consequence of what I described, many smaller journals should join the technology and maintain standards. The future looks bright!
Science will be available to all, in a controlled and affordable way.