INFORMETRICS 87/88
L. Egghe and R. Rousseau (Editors)

THE COVERAGE OF SCIENCE IN THE THIRD WORLD: THE "PHILADELPHIA PROGRAM"

Michael J. MORAVCSIK*

Science Studies Unit, Lisbon Institute, Stationsplein 242, 2312 AR Leiden The Netherlands

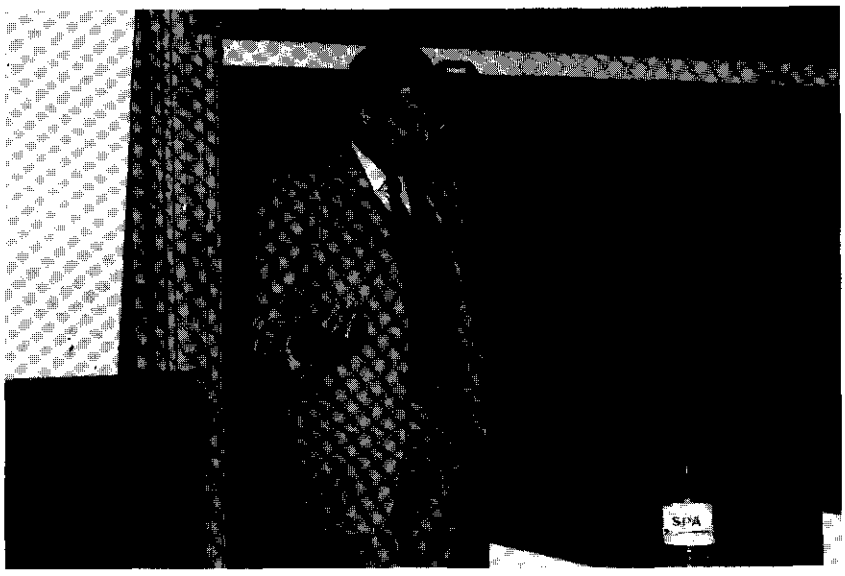
Abstract

A year-long study was conducted by a selected group of interested people, culminating in a three-day workshop in Philadelphia in July 1985, exploring the extent to which the coverage of science in bibliometric data bases is adequate, and what measures could be taken to eliminate the deficiencies in such coverage. The talk at the conference will summarize the main conclusions and recommendations of the final report of this effort. Since many of the recommendations can be implemented in a decentralized way, the aim of the presentation is to urge the individual members of the information science community to participate in this implementation.
The recommendations are in five groups: I. Studies. II. Steps requiring scientific consciousness raising but minimal resources. III. Steps requiring further organization but minimal resources. IV. Steps with small and catalytic costs. V. Steps with broadly distributed costs.
They pertain to editorial practices, peer reviews, technological aspects of publication, bibliographic control of published material, as well as to a worldwide cooperative program to increase the number of scientific publications produced in the Third World which are included in the ISI data base.

About five years ago I submitted, together with a coauthor, a paper for publication in the Latin American journal Interciencia. It dealt with a study of the scientific output of individual Third World countries, as a function of time. The indicator used for the output was the number of scientific authors, and the data were taken from the database of the Institute for Scientific Information (ISI) in Philadelphia, the institute which issues, for example, the Science Citation Index, and many other compilations.
The paper was duly referred by some Latin American referees and the verdict was that while the paper, in principle, provides interesting information and hence in principle deserves publication, it should not be published because the ISI database which provided the data on which the study and its conclusions rested was so prejudiced against developing countries that the results obtained from it are useless.
The paper was published in Scientometrics instead, but the apparent controversy regarding the ISI database sufficiently intrigued me to send a copy of the referee's report to Eugene Garfield, the founder and president of ISI, for comment. His reaction was that ISI has a systematic and biasfree methodology of deciding what journals to include among those surveyed for the ISI database, and

---

* Permanent address: Institute of Theoretical Science, University of Oregon, Eugene, Oregon 97403 USA.



Michael J. Moravcsik

hence the accusations were groundless. To continue the exploration of this controversy. I then sent Garfield's reply to Interciencia for their comments. Their reply was that the methodology used by ISI to select journals for inclusion was based on the citedness of the journal, and that journals from developing countries are not de facto available to scientists in most countries and hence they are unaware of the articles in such journals, unless the journal is included in the ISI database, in which case titles, abstracts, and citations by and of it are widely available within the worldwide scientific community. In other words, they claimed, being cited and being included in the ISI database form a vicious circle.

The situation appeared sufficiently intriguing for further action, and so, with the generous help of Garfield himself who offered the ISI premises free of charge, and with generous help by the National Science Foundation and the Rockefeller Foundation, I organized a year-long effort to explore this issue in more detail.

Some 25-30 individuals were selected from the world over, among those who were known for their interest and involvement in Third World science, in bibliographic indicators, and other related subjects. There are of course many more such people than 30, but we wanted to keep the effort small so it could be efficient.

The effort began, for a period of some 10 months, with postal correspondence among the members of the group, channeled through me, in which position papers were circulated, opinions were aired, data were provided, etc. During those 10 months some 4 or 5 rounds of correspondence took place. The postal rounds also aimed at shaping the agenda for a three-day meeting in Philadelphia at the end of this period at which all the members of the group met to discuss the issues. There were no set talks at the meeting at all, only informal, round-table discussions among all the participants. The sessions were monitored by rapporteurs chosen from the participants, and each rapporteur gave a 15 minute summary of the session at the end of the session.

After the meeting an overall report was prepared, summarizing the conclusions and listing the specific proposals. The report has been widely distributed and additional copies are available from my office. This publicity of the outcome of the meeting marks, we hope, a new phase of the effort, probably stretched over

several years, during which the members of the group, with the help of other interested people, will implement some of the proposals.
In this section of our discussion, I want to analyze some of the aspects that arose during this effort. The topic turned out to be a very interesting one, exhibiting a fascinating melee of conceptual, practical, logistic, and social factors.
Nobody disagrees that the ISI database is not complete, and was never designed to be complete. The total number of scientific journals in the world is hard to determine, since definitional questions arise immediately, but the number is somewhere around 50,000-70,000. The ISI database surveys less than 5,000 of these. Why is it not more complete?
Databases are constructed for three general purposes:

1) To provide information on research already performed in order to support ongoing research. Scientists involved in new research want to build on what has been done previously and hence need to have information on that.

2) To provide information on the scientific community for the management of science. Bibliometric indicators form an important (though not the only) output measure for scientific activity, and hence they are useful for various functions in the evolution, maintenance, and management of the scientific enterprize.

3) To provide information on the evolution of scientific knowledge and on the pursuit of science as a human activity. The scholars of the science utilize such information.

None of the above three aims necessarily demand that the database be complete. One can conceive of questions posed in any of the three contexts that would require a complete data base, but such questions generally do not form a sizeable or significant part of the entire range of functions.
Neither is completeness possible economically. Databases are constructed for users, and whether the undertaking happens to be in the private sector (as the ISI database is) or whether it is subscribed from public funds, the cost can not be ignored. Surveying the first few thousand large and frequently published journals which are heavily used and cited is, from this point of view, worth wile. ISI's selection of journals covers the overwhelming fraction of journals which are often cited. Extending such coverage to a much larger number of journals, each of which is more rarely issued with fewer articles in it, becomes quite costly without representing a substantial improvement in the eyes of most users.
Finally, it is not desireable either to make the database unnecessarily huge. The more compact the database is and still functional from the point of view of most users, the better it will serve those users.
Thus we see that in terms of the sides of the debate I mentioned at the beginning of this discussion, both sides are right in certain respects. From the point of view of most users of the ISI database, the present selection is justified. On the other hand, from the point of view of some functions within science generated in the Third World, the database is incomplete in a biased way.
The latter statement holds because jounals published in the Third World do tend to be, on the average, less often issued and thinner. There are, however, also other factors that play against such Third World journals.
One is language. The processing of articles by the ISI staff can also be done in languages other than English, but the list of these other languages that can be dealt with is understandably finite, even in the context of European languages. For example, the non-Indoeuropean language of Hungarian cannot be handled. In fact, some studies showed that even the Russian-language literature appears to be somewhat underrepresented in the ISI database, compared to other compilations. Whether this is the result of the citedness criterion or not is not clear. In any case, journals in local scripts (e.g. Thai) cannot be handled, unless the data which are needed for such handling (title, abstracts, references, etc.) are also given in English or at least in Latin characters. More about this point later.

There is also some bias built into the mode of selection which is related to the subject matter of the paper. Some applied scientific research in developing countries is on topics which pertain to the specific geographical, economic, or social circumstances of the country, and hence is unlikely to be cited outside the country in those journals already in the ISI database. We could not estimate the size of this particular effect. My own feeling is that it is not a very substantial one. Few topics are strictly country-specific. For example, research on tropical agriculture is pursued in many countries as well as in international research institutes, and the best journals serving these research needs are likely to satisfy the requirements of the ISI selection procedure.

There is of course also the claim that the distribution of functions of a scientist among research, interaction with technologists, maintenance of the scientific infrastructure, education, etc. is different in a Third World country than it is in the scientifically advanced countries, and such a difference works against the scientist in the developing country when it comes to being represented in ISI-type databases. That is, of course, a valid point, but one must remember that bibliometric indicators do not, and are not designed to, measure all aspects of scientific activity. For this reason the assessment of the performance of a scientist (or of groups of scientists) can very rarely be done adequately by using only bibliometric indicators by themselves. Bibliometric indicators at best measure the output of basic and applied scientific research. This issue arises because there are some who claim that using such bibliometric indicators at all in the context of developing countries leads such a country astray in its scientific development by placing undue emphasis on research. I strongly disagree. The preponderance of forces acting on a scientist in a developing country are against his performing good scientific research, and so there should be some indicators which measure this neglected dimension.

Then there are logistic factors that work against the inclusion of some journals published in the developing countries. These appear trivial details, but play an important part in practice. Some journals appear very irregularly, with the nominal date as much as two years behind the calendar date, or issue numbers are skipped in order to catch up. Some are difficult to acquire from Philadelphia, because the logistics of shipping is cumbersome or unreliable.

Then there is the question of refereeing articles submitted to journals. Refereeing within a small scientific community is difficult because there are very few people in the same specialty and because in such small communities personal animosities and affinities play a greater role. On the other hand, there are numerous problems with resorting to international refereeing. The cost is greater, the editor may not be able to select referees who are professionally appropriate, who can judge the article in the context of the work that is feasible in a developing country, or even who are willing to serve as referees. There is also reluctance because of the false impression that using international referees is a reflection on the independence of the country. As a result articles which basically are sound and useful but would need some reworking before they are in an optimally readable form do not get this extra attention and this affects their citedness.

Having considered these and other issues, the Philadelphia meeting addressed the question of what could be done to improve the situation.

There is no question that for some purposes of interest to the scientific community in the developing countries, a much more complete database of the science generated in those countries would be useful. It is also clear that the large international all-purpose databases, whether ISI's or others, can not be counted on providing that service. So perhaps the developing countries, individually or regionally, or even worldwide but perhaps subdivided by disciplines, should create their own databases?

Such an undertaking or undertakings would be quite demanding in financial and other resources. There are some countries which have begun some compilations of scientific work in the given country. Such systems are generally incomplete, not sufficiently computerized, and do not serve the aims of regional or international collaboration since they are country-specific. One could imagine, for example, a Latin American database, involving a dozen or so countries which all use only

one of two languages. The resources to do that would have to be considerable, and would probably not be recovered from user's payments without a huge amount of subsidy. Whether this use, of the resources earmarked for science and technology would be preferable to many other conceivable uses is open to debate. In any case, the group did not make specific recommendations with regard to such national or regional databases, beyond supplying some conceptual clarification of the potential status of such databases. the rest is up to the Third World.

The main result of the effort has been a set of eleven proposals which I will now discuss. They utilize latent opportunities and evident needs to build a small, catalytic bridge between the two. Thus the programs proposed are simple, mostly inexpensive, and implementable in a decentralized mode.

The initiatives outlined here are grouped into five categories. The first contains studies, or the acquisition of information on aspects of our problem area which would greatly facilitate solving them. Although one of the oldest "tricks" for procrastinating or for killing a project is to recommend "further studies", we believe that our list of studies is sufficiently specific to run parallel with (and not as a prerequisite for) other action.

The second category includes steps that are practically free (i.e. cost a negligible amount) but involve being aware of certain factors and willingness to introduce slight modifications in present practices. Creating such an awareness and the will to act accordingly, although not primarily a question of finances, is often not easy.

The third category includes action that is also practically free but involves a certain amount of organizational effort.

The fourth category contains small and catalytic initiatives wich are relatively inexpensive.

Finally, the fifth category features situations where the cost of the initiative, although not massive, is not small either, but where the cost can be shared among many participants so that the cost per participant is not excessive.

I. Studies

A. An analysis is needed of the nature and usage of the science and technology literature from developing countries (4.500 journals) that is acquired by the British Library Lending Division (25% of the use of the collection comes from foreign countries). Such information would be important for (a) assessing the characteristics of the literature, (b) assessing the relative use made of the literature by developing and developed countries (c) contructing scientometric indicators which include such literature (d) creating better and enhanced dissemination and use of such literature. Such a study can be carried out in the British Library, at a cost in the order of $ 12,500. Steps are being taken for executing such a study.

B. A study is needed to explore, in depth, the barriers to scientific research in the developing countries due to deficiencies in access to databases, secondary sources, and other information sources. The study should pick a specific scientific field and document the findings in detail. A subgroup chosen from among the participants of the meeting is in the process of preparing a detailed proposal for this.

C. A study is needed on the editorial practices in international scientific journals vis-a-vis articles submitted to such journals by authors from the developing countries. Is there any evidence for discrimination in the refereeing of such articles ? The study may follow the general outlines of the Merton-Zuckerman study of biases in Physical Review (Minerva 9, 66 (1971)). The study could be carried out by taking two highly cited and refereed journals, and select a batch of articles submitted to them by authors from developing countries, with a second batch of other papers serving as a control group. Such a study should be designed to avoid methodological problems found in previous work on peer review (See John Bailar, NEJM 312 (10): 654 (1985). These papers could then be refereed again by a peer review group; the study, a fairly standard one except for its

particular subject matter, would require the usual budget for such projects, perhaps between $ 50,000 and $ 100,000.
Alternatively, it would be quite-feasible to choose world-class journals in the major scientific disciplines, identify submissions from DC's and trace those papers through the editorial reviewing system. Such a study could furnish preliminary data on biases against DC authors, and perhaps more significantly generate information that would enable DC authors to compete better for space in journals and generate guidelines for journal editors that would insure fairer treatment of DC authors. This also would cost about $ 50,000-$ 100,000 and would follow the methodology of the Merton-Zuckerman work referred to above.

II. Steps needing awareness but minimal resources

A. To facilitate inclusion into international databases, journals in the developing countries should provide the following features using the ISO transliteration system:

i) At the journal level
  a) An English title page
  b) An English table of contents
ii) At the article level
  c) a postal address for the authors in Latin characters
  d) a transliteration into Latin characters of the references, if these are written in another script.
  e) if possible, an English abstract.

These features are almost always a prerequisite for including a journal in databases, and hence would greatly improve the chances of journals' inclusion.

B. To facilitate the consideration of journals published in the developing countries for inclusion into databases, each country should select, with the help of a specially constituted committee consisting of working scientists, a modest list of journals published in that country that are of international quality. This list should then be submitted to ISI and to the managers of other international databases, with a sample copy of each of the journals included on the list. ISI will then study these lists in some detail and make decisions on inclusion.

III. Steps needing organization but minimal resources for operation

We need an initiative to define the best ways for the international scientific community to provide a peer review system for those journal editors and funding agency officials in the developing countries who wish to avail themselves of such opportunities. It is assumed that the operation of such a peer review system would be no more expensive than the system now used in the scientifically advanced countries. The initial formulation and organization of the system, perhaps first on a pilot-project basis in some particular scientific discipline, would require a grant of about $ 50,000.

IV. Steps of small and catalytic cost

Microform journals, books, reports, etc. are less expensive to produce and to ship air mail than hard copy. The use of micro products in the developing countries is inhibited by the unavailability of good quality but inexpensive microform readers which could be purchased without expending hard currency. Microform readers are simple gadgets which even in the industrial countries, are available for $ 150 or so, and could be manufactured in developing countries for a fraction of that cost. That they are not is due to the initial reluctance of manufacturers to enter a new product area, as well as to the uncertainty about the size of the

market. An initiative is needed to provide seed money (perhaps merely in the form of guarantees) for encouraging some manufacturers in some appropriately selected developing countries to produce such microform readers. An initial sum (perhaps $ 10,000) would support an exploratory project to identify possible locales and individuals to be involved in the pilot project.

V. Steps with broadly distributed cost

A. In cooperation with four other research funding organizations, the U.S.National Research Council has prepared a database of research projects funded by these organizations. The International Development Research Information System (IDRIS) was designed as a private, joint database to enhance cooperation among the founding agencies. The database describes research underway in developing countries, including the recipient organization and address, researcher name, geographical area under study, key words, abstract, and level and source of funding.
Currently IDRIS is limited to the information entered by one of the founding institutions and is accessible only through one of these organizations. The originating institutions are: The Canadian International Development Research Center, the Swedish Agency for Research Cooperation with Developing countries, the International Foundation for Science, the German Appropriate Technology Exchange, and the Board on Science and Technology for International Development of the U.S. National Research Council.
In order to make IDRIS more useful, additional donor agency projects need to be included. Access to the database also needs to be extended to Third World researchers. It is recommended that the founding institutions vigorously pursue the expansion of IDRIS and provide access to as wide an audience as possible. It is recognized by the workshop that the expansion of IDRIS may require additional work and financial resources from large donor agencies which are, however likely to be balanced by the benefits derived from the use of the system.

B. Each country should ensure complete bibliographic control and complete availability of its own publications. Specifically, this means the progressive establishment of appropriately structured national information systems, wich would include a comprehensive national collection of all publications produced in the country, which can then be used for consultation or remote supply (by loan or photocopy). The system should also include a bibliographic database, compatible with international standards, and suitable provisions for easy access by local and other users. Such a goal requires a firm and standing commitment on behalf of the local authorities for the regular funding and support of such systems, as well as joint efforts on behalf of donor agencies in providing the assistance eventually needed for the development of such endeavors. There are already two international programs (in addition to many national ones) with such an objective, both originating with the International Federation of Library Associations and Institutions: Universal Availability of Publications (supported in part by UNESCO), and Universal Bibliographic Control.
As an example, steps toward such a goal in Thailand may include the following measures:
1) A Union List of periodicals prepared by the Thai National Documentation Center should be regularly updated.
2) The publication of Thai Abstracts by the Thai National Documentation Center should be strengthened.
3) The Thailand Information Center of Science and Technology (TICST) should form a network of information centers and libraries within the country and should work on bibliographic control as a major task.

C. There is need for monographic coverage of certain scientific fields which are of special interest to developing countries but which are not adequately covered in the regular databases for a variety of reasons, for example, because their literature is predominantly "gray", that is, in the form of reports. Tropical

fisheries is an example. Such monographs would also provide bibliographic coverage, explore the pattens of the research community, promote communication within such community, and promote the utilization of available resources. It would also contribute to a publication and citation database which is independent of the existing ones. Each field needs to organize its own effort.

D. This last initiative is the most specific and intensive of those considered by the meeting and addresses directly the expanded coverage of databases to include more of the scientific literature produced in the developing countries. As explained in the text, the aim is to add another 25,000-50,000 or so articles from the developing countries to the ISI database for SCI. This would involve covering 500-1,000 more journals. The additional cost would be $ 250,000 per year, and this amount would have to be acquired from sources ouside ISI for three years, after which it could be absorbed into the regular ISI budget. The first step would be that described in II-B above, namely the "nomination" of journals for inclusion. Eventual selection for SCI will be based upon acceptable standards, and coverage in the SCI would not necessarily guarantee coverage in Current Contents. The $ 250,000 needed would be acquired, for the first three years, through additional purchases of copies of SCI by the developing countries, at the special discount rate of $ 3,000 per year granted by ISI to developing countries. Approximately 100 such new subscriptions are needed, which would be pledged by developing countries according to their size and resources. For example, a large country like India would purchase 20 additional copies, medium level countries 5, etc.. To cover some of the smallest and least affluent countries, the Rockefeller Foundation is pledging an annual amount of $ 15-30,000 as part of this effort. The outcome of this initiative would be a substantial improvement of the representation of science by the developing countries in the ISI database as well as an enhanced access of scientists and policy makers in the developing countries to SCI. After a few years, further determinations could be made as to the adequacy of coverage of Third World science under these new circumstances.
This initiative requires the organization of II-B as well as the organization of pledges from developing countries for the additional purchases of the SCI. This task will be undertaken by the participants of the meeting personally and also through the wide distribution of this report. The initiative is an example of a concrete step to resolve a concrete problem, using a decentralized, unbureaucratic structure and widely dispersed cost distribution which needs to be carried by Third World countries only over a limited amount of time. The initiative involves the cooperation of individuals and organizations in both the advanced and developing countries, divides the cost among them, and includes in cooperation governmental agencies, private organizations, individuals from the scientific community, and others. If the effort is successful, it may serve as a precedent for solving other problems in science development. It is important, however, to ascertain that this initiative does not draw away resources from the regular aspects of scientific information services such as subscriptions to primary journals or to abstracting services.
Respondents who wish to place new orders for SCI and thereby participate in this initiative should deal directly with Dr. Eugene Garfield, ISI, 3501 Market Street, Philadelphia, Pa. 19104 USA. The coordinator of this report (whose address appears on the cover) should be kept informed so that progress in archieving goals can be monitored and future efforts targeted on areas of greatest need.

EPILOGUE

Because of the decentralized nature of most of the proposals, it is difficult to assess to what extent these recommendations have been implemented in the last two years since the Philadelphia meeting. It is possible that there has been more taking place than meets the eye. Nevertheless, judging from the visible evidence, it appears that hardly anything of the proposed programs has been converted into

reality so far.

The aim of publicizing this effort at this conference, therefore, is to give it renewed publicity in the hope that eventually some action will follow the words. Success depends entirely on personal initiatives.