

INFORMATION MANAGEMENT FOR RESEARCH MONITORING

Michel J. MENU
Consultant in Information Management Systems
Gentilly, France

ABSTRACT

Bibliometric techniques remain an essential basis for the analysis of scientific activities and the construction of meaningful indicators. Although the latter also requires that a suitable reference framework, or conceptual model is available, it is in first place dependant from the availability of data on publications, that is to say ultimately machine-readable records. Using data available on existing publications, a wealth of interesting work has been carried out on the publications patterns of scientists in the Less Developed Countries (LDCs), eg. international versus local publication, type of publication, language of publication, etc. Much less attention has been paid in our view to the basic conditions which command the very existence of publications and of usable records. This paper, which essentially refers to the situation in subsaharan Africa, will attempt to highlight the limitations affecting the communication cycle, the need to establish information management as an integral component of scientific research and its management and possible practical alternatives for ensuring the provision of suitable records. Since there is, to the best of our knowledge, a dearth of empirical studies on these problems, the paper is to be regarded as a personal point of view for discussion, based upon our experience in establishing information systems in these countries.

RESUME

Les techniques bibliométriques sont essentielles pour l'analyse de l'activité scientifique et la construction d'indicateurs pertinents. Bien que cette dernière suppose aussi l'existence d'un cadre de référence ou d'un modèle conceptuel adapté, elle est au premier chef dépendante de la disponibilité de données sur les publications, c'est à dire en fin de compte de références lisibles par ordinateur. Nombre de travaux intéressants ont pu être réalisés sur les modes de production des chercheurs des pays en développement à partir des données disponibles sur les publications existantes; ils concernaient par exemple les rapports entre publication internationale et locale, les types ou les langues de publication, etc .. Il nous semble que les conditions de base qui permettent l'existence même de publications et de références exploitables ont bien moins retenu l'attention. Cette communication, qui se réfère essentiellement à la situation en Afrique subsaharienne, cherche à montrer les limitations qui affectent le cycle de la communication, la nécessité de faire de la gestion de l'information une partie intégrante de la recherche et quelques approches pratiques qui permettraient d'obtenir des références utilisables. Ces problèmes n'ont à notre connaissance fait l'objet que de très rares études empiriques; cette communication expose un point de vue personnel pour contribuer à la

discussion en se fondant sur une longue expérience dans l'établissement de systèmes d'information dans les pays considérés.

1.LDCs LIMITATIONS IN THE INTERNATIONAL SCIENTIFIC COMMUNICATION PROCESS

In order for a scientific publication to be mentioned in other publications or major databases, it is necessary that:

1. a formal paper be prepared,
2. the paper be accepted by a recognized journal or conference,
3. the source document (journal or proceedings) be considered by a data base, and the paper be eventually selected by the publisher of the data base,
4. the reference be indexed under meaningful terms,
5. the document be obtained in time by the person who want to use it.

Prevailing conditions in the LDCs and practices in the international scientific communication system present many obstacles for the publications of LDCs scientist to successfully go through these steps.

The major ones lie probably with the first step. In order for a formal paper to be prepared, there ought to be:

- a researcher trained in the writing of scientific articles, aware of the standards and style conventions of the major publishers (which by the way differ noticeably);
- a researcher able to write and communicate in a major communication language if his/her mother tongue is not one, or is not the predominant one in his/her discipline;
- publishable results, what in turn is dependant from well designed projects and a steady provision of the resources required for their implementation, on the one hand, and availability of relevant and up to date literature on the other hand;
- a provision for publication in the research project programme and budget (time, funds for document production and distribution and/or for participation in conferences);
- a policy of publication in the researcher's institution, even if it is primarily concerned with the immediate transfer of results to the users community (eg. in the case of adaptive agricultural research);
- incentives for the research staff to publish (eg. recognition of publications in individual evaluation, promotion, etc.);
- equipment and supplies for the typing, editing, duplication of the paper;
- eventually clerical staff to prepare the document;
- a researcher having personal connections with fellow scientists who could review the paper and provide an authoritative feedback;
- communication facilities which allow for the latter interaction to take place in due time.

In many LDCs these conditions are not met in most instances. As a matter of fact, a noticeable portion of the LDCs research results is not recorded in formal documents or is recorded in internal reports. Non conventional documents seem to be the dominant form of formal communication (1, p. 71).

In order for the paper to be accepted by a recognized journal:

- the researcher should be aware of those journals which are more likely to be interested in his/her topic, or informed in time of the conference;
- there should be at least one journal by assumption, but this is precisely not the rule, since journals, especially high status ones, are commercial ventures from organizations in the industrialized countries which naturally aim at satisfying the needs of their majority clientele, scientists in the industrialized countries; few journals with an international circulation are interested in LDCs related topics, thus making the competition more acute among the potential authors;
- the researcher or his/her institution should preferably be known from a member of the editorial board, or organizing committee, or someone having connections with it;
- communication facilities should allow for mailing the document, receiving the comments and sending back the revised version in due time;
- funds should be available for paying the page charge or conference fees and travel if appropriate.

Again it is difficult for these conditions to be met in LDCs. Assuming the document has been published, in order to be entered into a major data base related to its subject it should:

- appear in a publication which is regularly or specifically screened by the data base producer;
- be in a language and script which the data base could handle;
- in most instances, unless published in a "cover-to-cover" journal (because of the notoriety or special relevance of the journal all articles are systematically entered into the data base), be selected for inclusion.

An article on an LDCs related subject is not likely to appear in a publication which is regularly screened or to be selected, if it does. Publications issued in the LDCs, even of high quality, may be omitted simply because of a lack of awareness of the data base publisher, language barrier, delays in mail delivery, irregularity in the publication of the journals. Again, database production is a commercial venture which has to respond to the demand of the major segment of the market, the scientific community in the industrialized countries. In addition journal articles represent the predominant source of records while most of the literature produced in LDCs for many disciplines remain, especially in its initial life, in the form of non-conventional documents. The proportion of the documents cited in databases which have been produced in LDCs can be roughly estimated at around 10 % (2, p. 71); even in international cooperative databases which strive for an extensive coverage of this literature, such as AGRIS, the LDCs share remains around 25%.

Because databases are designed for a public in the industrialized countries, they use classification and indexing schemes which are best suited for this audience on the one hand, and are organized according to the size of the related literature, on the other hand. A topic which generates a very small percentual of publications will not be indexed under a specific entry. The particular terminology used in the LDCs in order to describe local objects, phenomena and concepts and geographic terms are therefore eventually grouped under broader entries which do not necessarily make apparent the connection with those terms or have a reference to them. Some fields of investigation may even be represented by so few entries or believed so marginal that they are practically omitted, as was the case for instance in the first version of AGROVOC (3) with water resources in agriculture or range management. This phenomenon is of course less sensible in basic sciences than applied or social ones. In any case, data base searches on LDCs related topics are slightly more complicated and may result in higher rates of misses.

Once the reference has been identified, the interested person has to obtain a copy of the document within reasonable delays. If the document has been published in a major source and he/she works in an industrialized country, this may go without much problem while, if he/she is located in an LDC, it is much likely that no nearby library will have the document, either because of shortage of acquisition funds or because of delays or losses in the delivery. If the document has appeared into a special or local source, it is not likely to be available anywhere except in a few organizations of the considered country. Obtaining copies supposes that the correct mailing address of the institution where the document could be found or from which it originates is known, the request is received and properly processed, the requester is in a position to pay, when not prepay, the copy, the institution is in a position to make a copy and to mail it back (i.e has envelopes and stamps) so that it will be received within reasonable delays. Even in industrialized countries, such trivial steps do not go without problems, at least for what concerns the effectiveness of mail services and the interference of customs offices, whose interpretations of the international convention on the free flow of scientific documents is often erratic. In LDCs, the outcome of each of the above mentioned steps in the process is unpredictable.

Data concerning research infrastructures and programmes (institutions, research staff, facilities, funding, etc.) are not compiled at the international level, except in a few cases such as UNESCO's science and technology statistics, agricultural research (ISNAR, CARIS, SPAAR) and programmes with international support (IDRIS) (4). They may not be available either at the national or institutional level in a form which would be both reliable and suitable for secondary analysis.

2. THE NEED FOR INFORMATION MANAGEMENT

Scientific research is not only heavily dependant from information activities as input to and output from the projects but can be regarded as one of the information industries (5). In any case the acquisition and use of information, the production of new information, its formalization into documents, their dissemination and various additional forms of communication occupy a central place in the research process. The research cycle, including research management, and the communication cycle are intrinsically interwoven and the former would not exist without the latter. In this respect, it is sad to note that if few studies of the communication cycle such as those of Garvey (6) or King (7) were attempted for the industrialized countries, they have no equivalent for the LDCs.

Even though the availability and visibility of scientific literature from LDCs are subject to influences from the industrialized countries, they also have in first place deep roots in the LDCs themselves. A number of socio-cultural factors play an inhibiting role (8). A cultural bias is not less apparent in the assessment of LDCs publications by the scientific community of the industrialized countries. To the extent scientific research is carried out in organizations, either universities or research institutes, which for their vast majority belong to the public sector in the LDCs, one should also pay attention to the institutional factors which influence the overall communication or information cycle. There appear to be many drastic limitations for what concerns the manpower resources, the organizational set up and the provision of financial and material resources. Other noticeable constraints result from the overall socio-economic conditions, especially in that they do not allow for a regular support of research activities and for the effective operation of basic public services such as mail, transport and telecommunications, what not only results in additional costs and production losses (9) but also in constant burdens which have a demobilizing effect.

If one considers the series of roles and functions in the communication cycle as they were adapted by Borko and Menou (10), a realm of problems may be identified, of which the list below may give some idea, although it is neither systematic nor exhaustive. The information users have to specify their information needs with regard to subject, methods, challenges and plan of the projects, resources and strategies of the institutions. The major limitations associated with this function are: insufficient definition of socioeconomic constraints for the usability of research results; deficiencies in research planning; changing leadership, organization and objectives; inability to specify needs; unawareness of national and international past and current research. The users have further to obtain information where they face such limitations as: unawareness of sources; disorganized files and collections; inefficiency of information services; inability to submit clear requests; unreliable communication and postal services; prevalence of bureaucratic procedures; unavailability of funds especially hard currencies; delays and insecurity of delivery. Users have finally to

exploit information where they face such limitations as: lack of training in information analysis and evaluation; language barrier; lack of equipment and supplies; loose planning, monitoring and evaluation procedures.

The information producers have to outline their papers, record data, elaborate, organize, write, type, file their papers and related data. The major limitations faced for the fulfillment of this role are: lack of training in scientific and technical writing; lack of equipment and supplies; insufficient skills in communication; language barrier; low qualification of support staff; lack of training in records management; bureaucratic procedures in reporting.

The publishers have to prepare galleys, edit, print, distribute, file the publications and manage editorial boards. The major limitations faced for LDCs research institutions to fulfil this role are: lack of equipment and supplies; low qualification of support staff; difficulties of communication with authors; lack of standing publication budget; irreliability or unavailability of contract printers: unreliable postal services; extra cost of dispatching because of geographic location and dispersion; deficiencies in publication policies; lack of training in publications management.

The information services have to obtain information, process items, shelve documents, store, retrieve and disseminate information, answer queries, repackage, consolidate, print and distribute information products, provide copies of primary documents. The major limitations faced in this role are: lack of acquisition budgets; loose connections with producers; shortage of skilled manpower; inefficient procedures; duplication of work imposed by participation in incompatible documentation systems; inadequate premises; lack of equipment and supplies; lack of budget for current expenses; lack of reliable procurement and maintenance agents on site as well as abroad; inconsistent dissemination policies; loose connections with and remoteness from the users.

The research team leaders and managers of research institutions have to organize information functions and flows, establish procedures, allocate resources, supervise information activities, maintain internal and external linkages, evaluate the information system. The major limitations faced in this role are: lack of training in information management; persistence of obsolete formal procedures; excessive bureaucratization; irregular and insufficient funding; dependance from foreign assistance; deficiencies in publication policies; low level of interpersonal and inter-institutional interaction.

Most of the information functions and the problems they encounter are usually dealt with as separate areas. Research planning, monitoring and evaluation, computer facilities, telecommunications, printing facilities, editing, distribution and sales of documents, extension activities, special libraries or documentation centers, data collections, technical archives, etc. will be subject to individual decisions, attached to different departments or set up as particular units. They are in addition considered as ancillary services whose resources are eventually obtained from the provisions for overhead expenses. If it is true that information has become in the modern society the most important strategic resource, this

should be even more true for scientific institutions. The question is not to secure an acceptable level of resources to all these functions nor even to assemble them into a coherent set forming a purposeful corporate memory. A memory connected to a deficient brain can hardly be used. What is required is a composite, organization-wide system integrating the various specialized information sub-systems, which would retain the appropriate level of autonomy and decentralization, on the one hand. On the other hand, there should be a single and high level authority for the management of the entire system as an integral part of the overall management of the organization and collective mechanisms for establishing standards, planning and monitoring the operation of the system and adjusting its activities to the national and institutional strategies.

Information Management, as summarily described above, is a requisite for the research institutions, like any others, in LDCs to increase productivity. Without it, the availability of scientific literature from LDCs is not likely to increase noticeably. It is also perhaps the basic condition for the data to be effectively collected, which are required for strategic planning of research, project planning, monitoring and evaluation by scientists and managers in LDCs and for scientometric studies. Information management should thus deserve the highest priority from both national authorities and technical assistance agencies. One can not however avoid some fundamental questioning when noting for instance that a basic text on agricultural research policy and management in LDCs (11) does not devote to information activities more than a few pages out of several hundreds, which provide little more than general comments on the importance of having access to international literature, or that a report on the assessment of a LDC national agricultural research system does not devote more than one paragraph to information management (12, p. 13).

In order for information management to become a standard feature in the scientific research organizations of LDCs, the following actions may be considered. All categories of staff should be adequately trained in the use of information technology, implementation of the information activities which they have to fulfill under the prevailing conditions and in information management. To the extent a sufficient number of qualified information specialists staff can not be provided, the various tasks it normally performs should be taken care of by other categories of staff which should be trained accordingly. Mechanisms, both national and international, should be established in order to secure the permanence of a minimum nucleus of trained information staff within each research institution (eg. by buying the information products they generate rather than acquiring them on a donation or exchange basis). Simple rules should be established within the national research systems for the generation, recording, distribution, storage and dissemination of all types of information. Standard software packages should be selected and adapted or developed for the integrated implementation of the various information tasks, taking into account the specific requirements of the research institutions in LDCs. The acquisition of information technology and information management capabilities should be included in a

systematic and coordinated fashion into the international cooperation projects, as opposed to the current specialized and project and function specific approach (13).

3. ALTERNATIVE WAYS TO IMPROVE THE AVAILABILITY OF DATA

Publications may be regarded as the most straightforward medium for the communication of most data which are required for the analysis of scientific activities since they could indicate:

- a) original title of the publication;
- b) language of the publication;
- c) full name of all members of the research team, which is necessary in order to discriminate homonyms;
- d) the permanent affiliation of each member of the team;
- e) the nationality of each member, which may be added to the standard entries by using the ISO code of countries;
- f) the principal investigator;
- g) the institution(s) where the research was carried out;
- h) the source(s) of funding for the research;
- i) the programme(s) to which the particular project whose findings are reported is related;
- j) the amount of funding allocated to the project;
- k) the total duration of the project and the phase to which the published findings correspond;
- l) the subject of the research through a significant title and its eventual enhancement, and an informative abstract;
- m) the methodology of the research;
- n) the other publications and communications resulting from the research, through the citations;
- o) the previous related work and its above characteristics, through the citations.

The primary publications do not presently include all these data. Some are omitted in most cases, like nationality or permanent affiliation, others are not systematically recorded, like those related to funding, programmes and project or other publications. Most are not presented in a standardized fashion. It would thus be advisable that the LDCs research institutions establish standards by which their publications could carry all the information required for further monitoring and analysis of research activities. These standards should preferably be unified at the national, regional and international level. The inclusion of the data which are not commonly mentioned (eg. c, d, h, i, j, k above) and the comprehensive recording of the others would not impose much additional effort upon the authors. A standard cover page with all the entries required for the recording of the above indicated information could for instance be pre-printed, distributed to

the authors and used as a basic source for the monitoring of publications. Stocktaking and distribution data may eventually be added in order to further integrate the management of information.

Databases can only record in each reference the data which can be clearly identified in the primary document. Should the above listed data be available, international bibliographic standards and their particular application in each data base, whether local, institutional, national or international, may however prevent their inclusion. For instance the original title may be mentioned only if it is in the carrier language of the database, the given name is often represented only by an initial, the number of authors cited may be limited, the affiliation may be indicated only for the first author, who is not necessarily the principal investigator, temporary attachment to another institution is treated as affiliation in place of the permanent one, the number of citations is not always recorded and the references are not recorded, links among the publications resulting from a single project are not always recorded, most of the other data are neglected, although, in principle, they could at least enter into a note field. Such limitations have clearly a cost-effectiveness rationale in addition to systems constraints. Bibliographic databases are meant in first place for providing references in response to subject oriented queries and not for the analysis of research activities. The major databases, with stocks over the million records, can hardly consider any significant change in their format and procedures unless they switch for an entirely new system.

The provision of such data on LDCs research output can only be the responsibility of LDCs research institutions themselves and possibly those institutions in the industrialized countries which are directly involved with scientific research in LDCs. The documentation services in the research institutions of LDCs face drastic limitations with regard to their manpower resources, material facilities, financial resources and support. The range, quality and effectiveness of the services they can offer to the users are thus far from satisfactory what in turn reduces their ability to liaise with the authors and collect their papers and encourages the latter to look for solutions of their own for both control of their own information and access to external one. These trends are reinforced by the over centralization of information services, on the one hand, which is in contradiction with the prevalent individual and institutional culture (14), and their use of inadequate documentary systems and procedures. The establishment within each organization of flexible documentation networks which could ensure an appropriate level of bibliographic control at all levels, from the individual researcher up to the central services and the progressive aggregation and enrichment of the records in cooperative databases is a first requirement. This however can only be achieved if a suitable format such as BABINAT (15), specially designed for this purpose, is available and suitably adapted.

The documentation networks in the LDCs research institution should further develop particular routines with a view to ensure a more comprehensive recording of the relevant data, as earlier mentioned. In particular, when processing papers produced by their organization, they should create records for

the quoted literature which would bear a special identification and be linked with the citing papers. This clearly implies an additional work, whose magnitude should however be tolerable (roughly for an institution producing 100 papers per year with an average of 50 citations, this would represent 3 man/months) but would yield invaluable benefits both for information access and for the monitoring of research activities. The fact that a noticeable portion of the scientific output is at best recorded in the form of non-conventional documents, often internal ones, and thus are not included in bibliographic files, will however remain. Such types of literature could be consolidated by first specifying a series of types of documents and their characteristics (eg. project or section notes, occasional papers, etc.), second establishing suitable mechanisms of review which could ensure the required quality level for each category, third creating bibliographic records for those documents, once at least one copy is stored in an identified location and remains available, fourth including those records into the input sent to international information systems. It should be noted that other forms of scientific output such as participation in education and training programmes, in extension programmes, in national seminars, etc. could also be subject to some form of summary recording and become accountable.

The implementation of the above suggestions can hardly be practical without the use of modern information technology. In this respect, it is fascinating to observe that the appropriateness of information technology in LDCs is still much questioned (16) on the basis of the many cultural, social, human, economic, logistic, technical, administrative and environmental obstacles it faces, which nevertheless apply as well for the more traditional techniques, and certainly for the operation of sophisticated scientific equipment which is nowadays very easily provided to these countries and not so well utilized (17, p. 68). To the extent information activities along the communication cycle involve a large proportion of successive exercises of data gathering, recording, organization, processing and dissemination which immobilize scarce qualified manpower in repetitive data manipulation, information technology, especially stand alone or networked microcomputers, could provide an unique answer.

Assuming that bibliographic records on LDCs scientific production would become more readily available as a result of the above suggested actions, the problem of their international dissemination will not be solved. With compatible electronic records, it is relatively easy to produce compilations such as institutional and national bibliographies and to circulate them. Specific products could also be developed in cooperation as demonstrated by SESAME (18) for instance. The optical disk technology is even likely to allow in the near future for compiling mixed databases, including the full text of papers, which may be more easily screened by the producers of large databases. The latter would also be in a much better position to reuse the data with reformatting routines. In the case of international cooperative information systems, such as AGRIS, the data elements which are presently missing as previously discussed, could easily be preserved if

moved into a notes field which will be saved in the international data base even if it is not available in the standard search routines and printed products.

The implementation of such measures could not go without the allocation of sufficient funding by both national and international organizations, which will be harder than ever to obtain in the foreseeable future. The funding available is however not insignificant. The question is more of the share granted to information management, what ultimately points to the concepts and mentalities of those in charge of scientific development in LDCs. As long as information activities will "go without saying", thus without specific and/or adequate allocation of resources, little change will occur.

NOTES & REFERENCES

- (1) Chatelin, Yvon and Arvanitis, Rigas. *Stratégies scientifiques et développement; sols et agriculture des régions chaudes*. Paris, ORSTOM, 1988.
- (2) Whitney, Gretchen. *Language distribution in databases: An analysis and evaluation*. Metuchen NJ, Scarecrow, 1990.
- (3) AGROVOC. *Multilingual thesaurus for agricultural sciences and technology* (produced by FAO and the Commission of the European Communities). Rome, Apimondia, 1982.
- (4) ISNAR: International Service for National Agricultural Research, The Hague, Netherlands; CARIS: Current Agricultural Research Information System, FAO, Rome; SPAAR: Special Programme on African Agricultural Research, coordinated by the World Bank; IDRIS: International development research information system, coordinated by IDRC, Canada.
- (5) Porat, Marc Uri and Rubin, Michael. *The information economy*. Washington DC, U.S. Department of Commerce, 1977.
- (6) Garvey, W.D. and Griffith, B.C. *Communication and information processing within scientific disciplines: Empirical findings for psychology*. *Information storage and retrieval*, Vol 8, No. 3 (1972), p. 123-126.
- (7) King, D.W., McDonald, D.D., Roderer, N.K., and Wood, B.L. *Statistical indicators of scientific and technical communication (1960-1980)*. Washington DC, U.S. Government Printing Office, 1976.
- (8) Menou, Michel J. *Cultural barriers to the international transfer of information*. *Information Processing and Management*, Vol. 19, No. 3 (1983), p. 121-129.
- (9) World Bank. *World Development Report*. New York, Oxford University Press, 1988.

- (10) Borko, Harold; Menou, Michel J. Index of Information Utilization Potential (I.U.P.). Final report of the phase 2 of the I .U. P. pilot project . Paris, UNESCO, 1983 .
- (11) Ruttan, V.W. Agricultural research policy and development. Rome, FAO, 1987 .
- (12) Report of the CGIAR Documentation and Information Services Meeting, Patancheru (India), 16-20 January 1989. Patancheru (India), International Crops Research Institute for the Semi-Arid Tropics, 1989.
- (13) Menou, Michel J . Potential impact of information technology on the enhancement of agricultural research in the Less Developed Countries. In: Feeney, M. & Merry, K., eds.; Information technology and the research process. Proceedings of a conference held at the Cranfield Institute of Technology, UK, 18-21 July 1989. London, Bowker-Saur, 1990, p. 316-318.
- (14) Molino, Enzo; Guadarrama, Luis. Microcomputers for information and documentatlon activities in developing countries. In: Keren, C. & Perlmutter, L.; eds. The application of mini and microcomputers in information, documentation and libraries. Amsterdam, North Holland, 1983, p. 370-375.
- (15) Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD); Institut Français de Recherche Scientifique pour le Développement en Coopération (ORSTOM); Système d'Information sur les Pays en voie de développement (IBISCUS). BABINAT. Manuel de référence pour les bases de données bibliographiques nationales. Montpellier (France), CIRAD, 1989.
- (16) Garriott, G . Computopians and Nouveaumanics . Development Communi-cation Report, No. 67 (1989), p. 15-16.
- (17) Gaillard Jacques. Scientists in the Third World. The University Press of Kentucky. Lexington, USA, 1991.
- (18) SESAME. Bibliographic data base on CD-ROM of French language literature on agricultural research and rural development in developing countries. Distributed by CIDARC, CIRAD, Montpellier, France.