

B PB 1654
1988

NLD

DON TEX

Jointly published by Elsevier Science Ltd, Oxford
and Akadémiai Kiadó, Budapest

Scientometrics,
Vol. 35, No. 2 (1996) 247-255

EXPERIENCES WITH THE NATIONAL CITATION REPORTS
DATABASE FOR MEASURING NATIONAL PERFORMANCE:
THE CASE OF MEXICO

R. ARVANITIS,* J. M. RUSSELL,** A. MA. ROSAS**

* ORSTOM-Mexique, Calle Cicerón 609, Col. Los Morales, 11530 México, DF (México)

bibliographic searches. However, sharing experiences on the pitfalls encountered in the creation of value-added datafiles for bibliometric purposes is an important step towards achieving greater reliability and replication of results.

Present study

In 1993 the Institute for Scientific Information (ISI) launched as a tool for

impact on results. Nonetheless, policy decisions are likely to be made on the basis of indicators regardless of the magnitude of the figures.

There is no doubt that access to NCR provides important additional information to that obtained by downloading records from the Citation Indexes CD-ROMs. The wider coverage of NCR with the addition of 1,300 from the Current Contents series and the provision of annual citation counts makes NCR a more comprehensive evaluation tool. Taking journals in Chemistry and related fields for the period 1981-1993 as an example of the added coverage of NCR, Fig. 1 shows the relative distribution of journals between SCI and CC in the NCR datafiles as a whole, and in the journals containing Mexican papers, together with the total numbers of Mexican papers in both ISI products. Although the majority of journals are covered by both SCI and CC, the addition of CC titles increased coverage by between 22-28%.

The main restraint is the lack of citation data for the additional titles included in Current Contents making calculations of national production and citation rates impossible for instance in different disciplines, using a simple point and click method using software, such as Microsoft ACCESS. The ability to easily separate those records which include citation data (i.e. the journals included in Science Citation Index) from those without citation data (the additional journals in the different series

are apparently not covered either by the ISI citation indexes or by CC. One of these, *Veterinaria-México*, we know is indexed in only one ISI product, namely Focus On: Veterinary Science and Medicine. The number of journal titles without any coding increased considerably from 1991 onwards.

Our ability to correctly analyze trends in the production and citation rates of Mexican institutes would be considerably improved with access to general statistics on the ISI products, such as number of journals indexed annually, new journal titles and ceased titles, for example. Although this type of information is published periodically by ISI in the SCI printed version and through Journal Citation Reports, providing the NCR customers with these data would be extremely useful for users, such as ourselves, wishing to carry out in-depth analyses of the productivity and visibility of Mexican scientific research.

Identification of institutions

One of the main difficulties when attempting to identify documents published by a particular institution, using a commercial bibliographic database is the lack of normalization of institutional names. This is a particularly acute problem in the case of institutions from non-English speaking countries, as names are found in both the original language and in their translated form. Taking the UNAM as an example, we found 19 different combinations of its name, in both Spanish and English, in the institutional level of the corporate source in National Citation Reports for 1992 only (Table 2). These corresponded to 98% of the 1118 records pertaining to UNAM studies, the remaining 2% of UNAM documents were identified as a result of a knowledge of the individual institutes, faculties, etc., belonging to the UNAM, or by recognizing a certain element, such as a PO Box number, which had found its way into the institutional field. Identification of the three most common forms of the UNAM name; the abbreviated name in Spanish, in English, and its acronym, accounted for 94.5% of all occurrences. A preliminary study of the complete database (1981 to 1993) suggests that the number of variations is more in the order of 160.

As institutional coding is both time-consuming and tedious we have been looking into the possibility of an expert system which would do this job for us, as least as far as the main Mexican institutions are concerned. Considering that just four Mexican research institutions are responsible for around 70% of all Mexican papers published at international level, automatic indexing of these corporate sources would save a considerable amount of time, although it would not offset the cost of an expert system unless this could be shared by other institutions.

Table 2
 Frequency of different elements occurring in the organization field of NCR by which records pertaining to the UNAM were retrieved for 1992

<i>1) Variations and misspellings of the institutional name</i>	
AUTONOMOUS NATL UNIV MEXICO	1
NACL UNIV MEXICO	1
NAT UNIV MEXICO	2
NATL UNIV MEXICO IZTACALA	1
NATL AUTONOMOUS UNIV MEXICO	707
NATL UNIV AUTONOMOUS MEXICO	2
UNAM	66
UNIV NAC AUTONOMA MEXICO	1
UNIV NACL AUTONOMA MADRID	6
UNIV NACL AUTONOMA MEXICO	283
UNIV NACL MEXICO	5
UNIV NATL AUTONOMA MEXICO	1
UNIV AUTONOMA MEXICO	11
UNIV NATL AUTONOMA MEXICANA	1
UNIV AUTONOMA NACL MEXICO	2
UNIV AUTONOMA MEXICO CITY	1
UNIV MEXICO	3
UNIV NATL AUTONOME MEXICO	1
UNIV NATL AUTONOMA MEXICO CIRC	1
Subtotal	1096 (98%)
<i>2) By name of institutes, faculties, centres, etc.</i>	18
<i>3) By elements of the address</i>	4
Total	1118

Our coding of Mexican institutions for the chemistry study was carried out by assigning each institution a unique ten letter code. We further identified departments, faculties, and institutes wherever possible for the three major institutions, UNAM, UAM (Universidad Autonoma Metropolitana) and IPN (Instituto Politecnico Nacional) included in our analysis.

However this additional, more specific coding was generally feasible only for the larger institutions, like the UNAM, and could be guaranteed only at faculty or institutional level. For instance, we could identify CINVESTAV (Centro de Investigación y de Estudios Avanzados), a large research institute belonging to the IPN, but not the departments making up the CINVESTAV. In the case of the UAM, we could rarely identify the specific departments involved and in many instances were

not able to assign papers to one or more of the three main campuses of UAM due to insufficient information being provided by the database records. Great care was taken to correctly assign records to the UNAM or the UAM, a task complicated by the presence of the word "Autonomous" in both university titles. In many cases correct coding was the result of the authors' intimate knowledge of the three institutions and was sometimes achieved through familiar elements in the address field, such as a postal code.

Field lengths

One of the difficulties we had with NCR files were their XBase format with fixed length fields. Many addresses, complete in the main ISI files, are truncated in the

particular institutional address in the address file. This facility would also make it possible to carry out in-depth studies of co-authorship patterns and institutional collaborations both at national and international level.

We have not attempted yet to measure the performance of individual Mexican scientists. However, due to the lack of any kind of standardization with respect to author names either by journal editors, and even by the authors themselves, this is a task which would require an considerable amount of time and effort even when aided by some kind of standardized list of researchers. This situation is made more difficult in developing countries by the fact that few scientists, particularly in applied areas, stay very long in research activities⁷ and in the Spanish-speaking world by the custom of using both parents surnames.⁸ For this reason reliable results could be expected only when evaluating the performance of a small elite of scientists easily identifiable by the constant use of a particular name form, or by other characteristics, such as a well-defined research interest, co-authorships, affiliations, etc..

Journal classification

NCR uses different classifications of journals according to which of the different

particular aspect of a country's scientific impact. However, the absence of information on the citing papers prevents it from acting as a dual purpose database which could also be used for carrying out citation analyses required by individual researchers, or as a research tool with respect to informetric studies such as co-citation analysis. However, this facility would add considerably to the size and cost of acquiring the NCR datafiles even in the case of small countries, such as Mexico.

We are of the opinion that these datafiles should be handled by specialists with an

