INFORMETRICS 89/90
L. Egghe and R. Rousseau (Editors)

THE UNIFICATION OF ADDRESSES IN SCIENTIFIC PUBLICATIONS

R.E. de BRUIN and H.F. MOED

Center for Science and Technology Studies (LISBON),
University of Leiden, P.O. Box 9555, RB Leiden,
The Netherlands

Abstract

Addresses in scientific publications contain a large number of variations. This phenomenon has serious consequences for the availability of information and for evaluative bibliometrics. Organizations are difficult to identify and become 'scattered' in rankings. To solve this problem, we developed a strategy of unification, i.e. classification of all variations by one name. To begin with, we sorted on the first part of the address. Next, we unified these addresses by the name of an overall organization (such as a university). We did so by analyzing the publication data. For some countries we performed further research in handbooks and encyclopedias. We solved the problem of scattering of organizations over several suburbs by bringing these city names under the denominator of the conurbation like Paris, London, etc. We unified variations in country names by the official name or its abbreviation.

1. INTRODUCTION

Scientific publications usually contain addresses of contributing authors. Such addresses consist of organizational and geographic names : the institutes where the authors are working and their locations (cities and countries). However, variations may occur in the names of publishing institutes, cities and countries. As a result, an organization may appear under different names (Moed, 1988). This has several serious consequences.

(a) Readers of articles may not be able to identify an organization, if a rather unusual name is used.

(b) For the scientific communication network (e.g. the ordering of reprints) the addresses are not always adequate.

(c) Scientific publications are indexed and included in scientific literature databases. In alphabetic indexes of publishing organizations, such as the Corporate Index, a part of the Science Citation Index (SCI), produced by the Institute for Scientific Information (ISI) at Philadelphia (USA), organizations are scattered, i.e. listed under different entries in the index.

(d) In evaluative bibliometrics (1), analyses of output or impact of research institutes are made (Carpenter et al., 1988; Moed & van Raan, 1988). In using addresses from literature databases, one can easily make errors and compile incomplete data. To a lesser extent this problem exists on country level as well. Variations in country names can hamper the assessment of national shares in scientific output, which is an important

part of science studies (Persson, 1988). For the problems resulting from variations in country names we can refer to a recent dispute between Leydesdorff (1988) and Anderson et al. (1988).

(e) The assessment of cooperation and communication between organizations is hampered as well. Variations in addresses make it much more difficult to trace inter-organizational networks in scientific research. Network analysis is becoming more and more important in the study of science and technology (Shrum and Mullins, 1988). Basically, the same is true for countries (Moed, 1988a; Zwaan, De Bruin and Moed, 1988). For assessing the internationality of publications and international scientific contacts incomplete data on country names can be quite a problem.

It appears to be important to analyze this problem of variations more closely. This paper represents an exploration of the problem. It is based on corporate source addresses in SCISEARCH, the on-line version of the SCI. A systematic comparison between what is included in SCISEARCH and what is printed in the original text, will be made in a follow-up study. Our preliminary conclusion is that ISI unifies organizational names to a certain extent, but many variations remain.

This paper deals with the following aspects of the variation problem.

(i) First, we describe the phenomenon of variations in some detail. We examine what types of variations occur and how frequently. Variations in institute names, city names and country names will be discussed separately. We present a 'phenomenology' of variations.

(ii) Next, we discuss the kinds of problems encountered in attempting to classify addresses under one denominator (unification). The three main types are unification problems with organizations, cities and countries. For many problems considered, we suggest possible solutions.

(iii) At present, we are creating a database on addresses in scientific publications. In this database, we have stored both variations and unified names of publishing organizations, as well as those of cities and countries in which they are located. We describe briefly the structure and content of this database, and evaluate our achievements on the unification problem.

The structure of the paper is as follows. In section 2 we give a description of the data that were analyzed and of technical aspects of data handling. In section 3 we give a 'phenomenology' of variations. Section 4 contains a description of the unification process : the problems and possible solutions. In section 5 we describe structure and content of our database. Moreover, we give an evaluation of our achievements and present some results. In the conclusion (section 6) we indicate applications and implications of the unification method.

2. THE DATA

We obtained our data from SCISEARCH through the host computer of the Deutsches Institut für Medizinische Dokumentation und Information (DIMDI) at Cologne (Federal Republic of Germany). We extracted data on articles, published in some 75 scientific journals from the sub-fields of chemistry, physics, biomedical sciences and applied sciences. We reformatted data and created an 'off-line' database. The methods we applied are outlined by Moed, 1988. Thus far we have processed about 50,000 publications from 75 journals. These publications contain over 85,000 addresses, since publications may contain more than one address. In SCISEARCH the addresses of all contributing authors are processed.

3. 'PHENOMENOLOGY' OF VARIATIONS

In scientific journals authors do not use a standard code for their addresses. ISI is bringing some order in addresses while entering them in the database. For instance, names of universities are put mostly (though by far not exclusively) in the beginning of the address, which is not the case in the headings of many articles. An address processed by ISI consists of several parts of items, separated by commas and semicolons. The last two parts of such an address indicate the city and the country where the organization is located. The preceding part(s) of the address relate(s) to the organizational name. The number of parts relating to this name can differ widely. Consequently, the raw publication data, acquired from DIMDI, still contain a large variation in institute names. University names can be followed by departments, hospitals, laboratories, etc., but there is no rule. As an example we give the addresses of some departments of the University of Illinois in Chicago.

UNIV ILLINOIS, HLTH SCI CTR, ABRAHAM LINCOLN SCH MED, COLL PHARM

UNIV ILLINOIS, HLTH SCI CTR, DEPT CHEM

UNIV ILLINOIS, DEPT CHEM

UNIV ILLINOIS, COLL MED, DEPT BIOL CHEM

UNIV ILLINOIS, HLTH SCI CTR, DEPT PHYSIOL & BIOPHYS, POB 6998

UNIV ILLINOIS, DEPT PHARMACOL, 901 S WOLCOTT

UNIV ILLINOIS, MED CTR, DEPT ANAT, POB 6998

UNIV ILLINOIS, DEPT MECH ENGN

The names of the universities themselves may also differ widely. Some universities have the name of the city where they are located (with the name of the city before or after the word university, such as UNIV CAMBRIDGE or CAMBRIDGE UNIV). In other cases they bear the name of a person. This is especially common in Germany and France. In France the structure of higher education was reorganized as after the student revolt of May 1968. At the end of the same year the Orientation of Higher Education Act (Loi d'orientation de l'enseignement supérieur) was passed, creating more student participation and enlarging university autonomy (Knowles, 1977; Soëtard, 1987). The Act resulted in founding 74 universities in the early 1970's, by replacing or splitting up existing universities (Knowles, 1977; *International Handbook of Universities*, 1986). Every major city nowadays has at least one university. The universities are named after these cities, and when there are more in one city, numbers are given : University of Grenoble I, II and III, University of Paris I to XIII etc. Some of the universities officially bear the names of famous French scientists or philosophers. So University Pierre & Marie Curie is University Paris VI and University Louis Pasteur is University Strasbourg I. In other cases the names indicate the specialty. For instance, the Université scientifique et médicale/Scientific and Medical University Grenoble is the University of Grenoble I.

Although in Germany the formal name of the university is often that of a person, for publications in international journals, German authors mostly (though not always) use the name of the city. So the University of Frankfurt appears in the database sometimes as Johann Wolfgang Goethe University or in variations like JW Goethe University. Specialized universities in medical and technological fields are called in German 'Hochschule'. In international publications the possible translation gives rise to several variations. In Germany a reconstruction of the higher education system was begun around 1980 (Altbach, 1985; Wissenschaftsrat, 1988). When organizations change their names, whether or not as a consequence of a general reform, problems can rise in the comparison of data from different years. On department level this phenomenon

occurs much more frequently.

There are also differences in spelling between the English name of the city and the name in the local language, such as UNIV MUNICH/UNIV MUNCHEN and UNIV NAPLES/UNIV NAPOLI. Sometimes there are evident typing errors like 'NIV SYDNEY' in a row of publications from UNIV SYDNEY. Another problem is the tendency of giving only the part of the university to which the author is attached. In those cases one may find only the college, faculty, department, hospital, etc. in the address, but no indication of the university itself.

The result of all this is, that universities can become completely 'scattered', with serious consequences for ranking lists based on citations or publications. Scattering of organizations can also be caused by the fact that (parts of) universities, or other institutions, give different cities in the address. Sometimes a university has residences in several cities, but especially in big cities like London and Paris the addresses sometimes contain names of suburbs, which are often separate municipalities. We found this phenomenon also in somewhat smaller cities like Lyon and Antwerp.

The degree of variation differs from country to country. Some countries have much more structured addresses than other. Germany and Switzerland are examples of relatively 'clean' countries with the university name mostly as the first separate part of the address. The exceptions are in most cases problems which are easily solved : the name variations like JOHANN WOLFGANG GOETHE UNIV/UNIV FRANKFURT and the addition HOSP. Furthermore, in the addresses of most US universities we found relatively few variations. There are however some notable exceptions like the City University of New York (CUNY) and Harvard University. Harvard has the added problem of being spread over Cambridge and neighboring Boston.

An example of a country with a large number of variations is France. The problem created by the tendency, referred to above, of using alternative names like UNIV LOUIS PASTEUR and UNIV STRASBOURG 1 (and the use of Roman and Arabic numerals) is a minor one. More serious is the absence of the university name itself in the address of publications by scientists of laboratories, institutes, hospitals, etc., belonging to a certain university or a research organization. This phenomenon is much more frequent in France than in many other countries. This is particularly true for Paris, where the problems are aggravated by the use of suburb names as city names. In the processed data we have found for the University Paris VI thus far 41 variations in the first part (before the first comma) of the address alone.

The problems for the United Kingdom are not as serious. Outside London we found in most university addresses the university name as the first part of the address. The University of London however is a major problem. A wide variation of colleges, hospitals and institutes is used, often without mentioning the University of London itself. When the name UNIV LONDON is used, it is mostly not a separate first part of the address, but a combination with college or hospital names. At the moment we have 20 combinations with UNIV LONDON in the first part of the address. Some typical examples are :

UNIV LONDON
UNIV LONDON BEDFORD COLL
UNIV LONDON KINGS COLL
UNIV LONDON KINGS COLL HOSP
UNIV LONDON LONDON HOSP
UNIV LONDON LONDON SCH HYG & T

Furthermore, there is some spreading over suburbs, but not to an extent comparable to Paris. Surprisingly, Oxford and Cambridge do not use college names very often. Authors from the University of Wales do use college names, while this university is spread over several cities.

In the Netherlands most university addresses have the name of the university in the first separate part of the address, but there are variations as a result of translating the Dutch name into English (or not translating it at all). An example is the University of Utrecht :

STATE UNIV UTRECHT
CENT UNIV UTRECHT
UNIV UTRECHT
UTRECHT UNIV
RIJKSUNIV UTRECHT
RIJKSUNIV

In addition, there are five variations of the combination with the Utrecht university hospital. Moreover, we found 14 names of laboratories, departments or institutes belonging to the university in addresses without the name of the university itself.

4. UNIFICATION : PROBLEMS AND SOLUTIONS

The solution to the problems described above is unification of the addresses, i.e. classification by one name. The aim of this unification is (if possible) to bring all publications from several parts of a university (or any other institution) under one denominator. To tackle this problem we developed a set of classification strategies. To bring some order in the chaotic structure of the addresses, we considered as a starting point only the first part of the address, i.e. the part before the first comma. We made this choice, because in SCISEARCH this part quite often contains the name of the overall organization, such as the university. In the sequence of the other items the hierarchical order is not as clear. Some addresses do not even have a second or third part. The number of items in the addresses differs widely. Sometimes a street (with or without number) is part of the address, but usually not. In a further study we will examine whether the unification can be expanded to a lower level of aggregation, e.g. university departments, by considering the other parts of the address as well.

After sorting on the first part we have at least the same structure in each address, but the problem of scattering still remains. When parts of an institution do not have the name of the overall organization in the first part of the address (or just differ in spelling), they cannot be recognized by the computer. It is our aim to bring all name and spelling variations, and all departments, sections, laboratories, hospitals, etc. (named in the first part of the address) under this denominator. To achieve this, important information can be drawn from the publication data. By comparing the addresses in a large set of publications, one can find out in some cases to which university a certain hospital or department belongs. A hospital or department mentioned in the first part of the address in publication A, can be named second or third after the university name in publication B. A good example of this phenomenon is that of some medical and pharmacological sections of the University of London :

UNIV LONDON, SCH PHARM

SCH PHARM, DEPT PHARMACOL, MRC, NEUROPHARMACOL RES GRP, 29-39 BRUNSWICK SQ

UNIV LONDON, ROYAL FREE HOSP, SCH MED, DEPT BIOCHEM & CHEM

ROYAL FREE HOSP, SCH MED, ACAD DEPT PHARMACOL

For the possible identification street names in the organizational part of the address and zip codes in the city or country part can be helpful. An addition to comparing publication data is the use of the Corporate Index of the SCI, which contains all addresses from publications processed by ISI in a certain year.

Although the number of variations decreased considerably, we still found the results unsatisfactory. Consultation of university handbooks and encyclopedias yielded additional information (2). Hospitals, departments, and colleges can be brought under one university; name variations like UNIV FRANKFURT and JW GOETHE UNIV can be compared and the results of the above described search operation can be checked.

The example of the University Paris VI, mentioned above, can be used to illustrate the method. The following list gives only a small sample of all the variations under which this university appeared in the addresses.

UNIV PARIS 06, REGULAT METAB & DIFFERENCIAT PLASTES, 4 PL JUSSIEU

UNIV PARIS 06, CYTOL LAB

UNIV PARIS 6, LAB PHYS THEOR & HAUTES ENERGIES PARIS

UNIV PIERRE & MARIE CURIE, ELEFTROCHIM LAB, 4 PL JUSSIEU, BAT F

UNIV PIERRE & MARIE, SYNTH MACROMOLEC LAB 24

UNIV P & M CURIE, ELECTRODYNAM GAZ IONISES LAB

HOTEL DIEU, SERV MED INTERNE & ONCOL

CHU ST ANTOINE, SERV MED NUCL

HOP ST ANTOINE, SERV MED INTERNE CANCEROL, 184 RUE FG ST ANTOINE

FAC MED ST ANTOINE, SERV BIOPHYS, CNRS, LAB 163

FAC MED ST ANTOINE, SERV BIOPHYS, INSERM, U113

HOP PITIE SALPETIERRE, SERV HEMATOCANCEROL

HOP PITIE, ER LIPOPROT, INSERM, U9

HOP SALPETIERRE, INSERM, U289

The hundreds of variations are reduced to 41 by sorting on the part before the first comma. In handbooks and encyclopedias we found that the University Pierre & Marie Curie is the same as University Paris VI (06), and that Fac Med St Antoine and Hop Pitié Salpétierre are components of it. Hotel Dieu could be identified as part of the University Paris 06 through comparing the publication data. The rest can be qualified as name variations, which can easily be dealt with.

In one of the addresses "CNRS' is mentioned as third part. This Centre National de la Recherche Scientifique is the French national research organization, originally founded in the 1930's, and in its present shape established by the Orientation of Higher Education Act of 1968, to advise the Minister of Education in matters of higher education and scientific research (Knowles, 1977; van Raan, 1985). Furthermore, it controls more than 1500 laboratories and research centers. CNRS institutes are often located on a university

campus, but still have an autonomous status. In some cases they are not connected to a university. In our publication data 'CNRS' can be part of a university address, as in the example of the University Paris 06. Quite often however, it constitutes the first part of the address. Then it is the question whether or not a CNRS institute forms part of a university, and if so, of which one. Comparing addresses in publication data, and consulting lists with CNRS institutes can largely solve this problem (3). Research organizations exist in many countries, for instance in Italy (CNR) and in Spain (CSIC). Organizations like these (often with links to a university), as well as non-academic research centers tend to become more important in science (Altbach, 1985). Central institutions for research are dominant in Eastern Europe (Clark, 1983; Kluczynski, 1985).

Looking up variations in handbooks and encyclopedias is very time consuming. Thus far, we have applied this method to the countries of Western Europe, Canada, and Czechoslovakia. We plan to extent, the number of countries. Since the US addresses are relatively well unified by ISI and this country is dominant in most journals, we already have fairly good results at the moment. More detailed information can be derived from guides, annual research reports, and address books published by universities. We have performed this kind of research only for the Netherlands. The same is true for consulting specialists in the field. Recently we started with this method for Switzerland too. People who are familiar with particular organizations can give important information. One final possibility is to compare the authors of publications with staff member lists of universities. A problem with this method is that scientists quite often 'move' from one institute to an other. Staff lists for the year of publication must therefore be used. This research, which will be even more time consuming, has still to be carried out.

We have tackled more completely the problem of the scattering of institutions because of differences in city names. The first step was to eliminate spelling errors, abbreviations and variations resulting from translation problems. In case of language variations we opted for the English name. So ROMA becomes ROME and BRUXELLES, BRUSSELS. Next, we used atlases to check whether a city name relates to a part of a conurbation or to a separate city. In order to bring a city name under the denominator of a big city, we retained the official conurbation border. So in the case of Paris all city names within the departments Paris, Hauts-de-Seine, Seine-St. Denis and Val-de-Marne are reduced to PARIS. For London the boundary of Greater London is kept up. This method proved to be especially successful for Australia, Canada, the United Kingdom and France. In the last mentioned country, many universities use the name of the university centers outside the city as city name, for instance St. Martin d'Heres in the case of the University of Grenoble I. Furthermore, 46 city names are brought under the denominator Paris. In this way the problem of geographical scattering is largely solved. However, there is still the difficulty that organizations may have residences outside the formal conurbation border or in other cities.

The 'cleaning' of country names is relatively simple. It is mainly a matter of spelling errors, the adding of zip codes to the country names or different abbreviations. However, in databases like SCISEARCH sometimes separate parts of the country are used to indicate the country of origin. As for the United Kingdom, not only England, Scotland, Wales, and Northern Ireland are used, but also countries like Lancashire and Midlothian, whether or not in combination with England, Scotland, etc. In the USSR the separate Socialist Soviet Republics like the LISSR (Lithuania) appear as a country name. Recently we have found combinations like Lithuania USSR. Addresses of Australian, Canadian, and Indian publications mostly contain the name of the state or province and the country as a whole (such as Ontario Canada or Uttar Pradesh India). We unified all these combinations and variations by the official country name (Australia, Canada, India, United Kingdom, USSR). The

abbreviation of the state names, used in the ISI data for American publications, are all transformed to USA.

One can also perform the identification of countries by using the Corporate Country (CCO) field in SCISEARCH. ISI claims that country names in this field are unified. However, the unification in the CCO field leaves some problems unsolved. The separate states of the US, several Socialist Soviet Republics in the USSR and the four main parts of the United Kingdom (England, Scotland, Wales, and Northern Ireland) are not combined to one single country. This has serious consequences for the comparison of countries.

The phenomenology of variations described above, is based on our research on data from SCISEARCH, implemented in the host computer of DIMDI. Experiences with other databases and other host computers, support our assumption that analysis of addresses can be even more problematic.

5. TOWARDS A DATABASE ON ADDRESSES IN SCIENTIFIC PUBLICATIONS

All the data concerning the unification process are stored in a database, consisting of three datasets - one for countries, one for cities and one for institutes - the so-called 'master files'. First, we describe the structure of the dataset on organizational names. This dataset contains all variations of institute names we have found thus far, as well as their corresponding unified names. Moreover, the dataset contains variables representing the city and country in which the organizations are located. Not surprisingly, 'problematic' organizations like the French universities and the University of London have a large number of records. So in the dataset there are 89 records with UNIV LONDON as unified name and 41 with UNIV PARIS 06.

In the dataset of country names we keep variations and standard names of sovereign states, unifying variations to the standard name. Where possible we use the English name as standard name, sometimes as an abbreviation, like PEOPLES R CHINA (Peoples Republic of China) and FED REP GER (Federal Republic of Germany). As far as possible we maintain the ISI spelling and abbreviations in our unified names. For each variation there is a record in the dataset, unifying all the variations to the standard name. Thus SWITSERLAND and SWITZERLAND relate to SWITZERLAND as a standard name and PEOPLES R CHI, PEOPLES R CHIN and PEOPLES R CHINA to PEOPLES R CHINA. The same is true for all the parts of the USSR. For countries with no variations there is only one record.

The same procedure is followed for cities, only the number of variations is much larger. For instance, 46 records in the city dataset have the unified name PARIS. One of them has the city name PARIS, all the other are suburbs of the French capital. Spelling and language variations are stored in the dataset. In both country and city dataset, zip codes are omitted. Storing these as well would make the number of combinations almost endless.

Each of the three datasets contains a separate variable ('check'), indicating the status of knowledge on specific combinations of variations and unified names. This field has three possibilities : Y, ? and blank. Y(es) means that a combination is confirmed by research in handbooks and encyclopedias. Thus far this applies only to the Western European countries, Canada and Czechoslovakia. We use the check '?' when we are not sure about a specific combination. A blank field means that we found a certain institute in the publication data (and possibly its unification to an overall organization through comparing the data), but not confirmed (yet) by further research.

Each time new publication data are downloaded from DIMDI, the addresses are matched with the master files. Variations 'known' to the master files are

changed to their corresponding unified name. Names of institutes, cities, or countries new to the masterfile are marked and analyzed separately. Afterwards these data are added to the master files. For the countries studied more detailed these new data are checked in the literature.

We use these master files for evaluative bibliometric studies and the role of journals in the dissemination of scientific knowledge. As part of this project we have now worked on the unification of addresses in scientific journals from the beginning of 1988 onwards. We collected thousands of data relating to countries, cities and institutes. At the moment we have 255 'country names' (including parts of countries), 3117 city names and 10,940 institute names in our database. They relate to 99 unified country names, 2625 unified city names and 7681 unified institute names. As for the institutes the reduction resulting from the unification is 30 %. With respect to the countries for which we have done more research, the reduction goes further : 44 %. In the Netherlands we even have a reduction of 61 %; our most complete research was carried out for Dutch organizations.

Since we focused our research on universities, reaching the unification largely in this type of addresses. The masterfile now contains much information about universities. Table 1 gives an impression of the growth in knowledge about eight universities from several parts of the world. The dates relate to the points of time when the master file was printed out for updating and correction; the numbers to the number of combinations present in the master file.

Table 1 : Increase of the number of variations in the Masterfile

	Feb 88	Apr	May	July	Nov	Jan 89
UNIV LONDON, London United Kingdom	45	49	63	63	76	89
UNIV ABERDEEN, Aberdeen, UK	2	2	2	3	6	8
UNIV PARIS 06, Paris, France	29	32	36	39	38	41
ERASMUS UNIV ROTTERDAM Rotterdam, Netherlands	2	6	13	13	13	19
UNIV FRANKFURT, Frankfurt, Fed Rep Ger	10	10	11	11	10	11
UNIV BERN, Bern, Switzerland	6	6	7	7	7	8
YESHIVA UNIV, New York, USA	1	4	6	6	7	8
UNIV COLORADO, Denver, USA	1	1	1	1	1	4
UNIV NACL CUYO, Bariloche, Argentina	5	5	7	7	7	7

The increase is the result of both adding data from new sets of journals, and of corrections as outcome of our research. For the Western European countries major changes occured between April and May 1988, and between July and November. Then most of our research in handbooks took place. This kind of research led us, in some cases, to removal of combinations from the masterfile, such as

UNIV PARIS 06 and UNIV FRANKFURT. Research based on comparing publication data is then corrected by further research. The statement that German, Swiss, and US addresses are in general relatively 'clean' is also illustrated in this table.
Between November 1988 and Januar 1989 we processed a large number of new publication data.

When new 'raw' publication data are processed, the master file recognizes the vast majority of the first parts of the addresses. When we processed a set of journals in December 1988 (14,000 addresses) almost 90 % of the organizational was recognized by the masterfile, of which 25 % confirmed by further research. In figures 1 and 2 we compare two stages in the processing, performed during 1988. Figure 1 contains all addresses, figure 2 relates to The Netherlands. As indicated above, we carried out most research for The Netherlands.

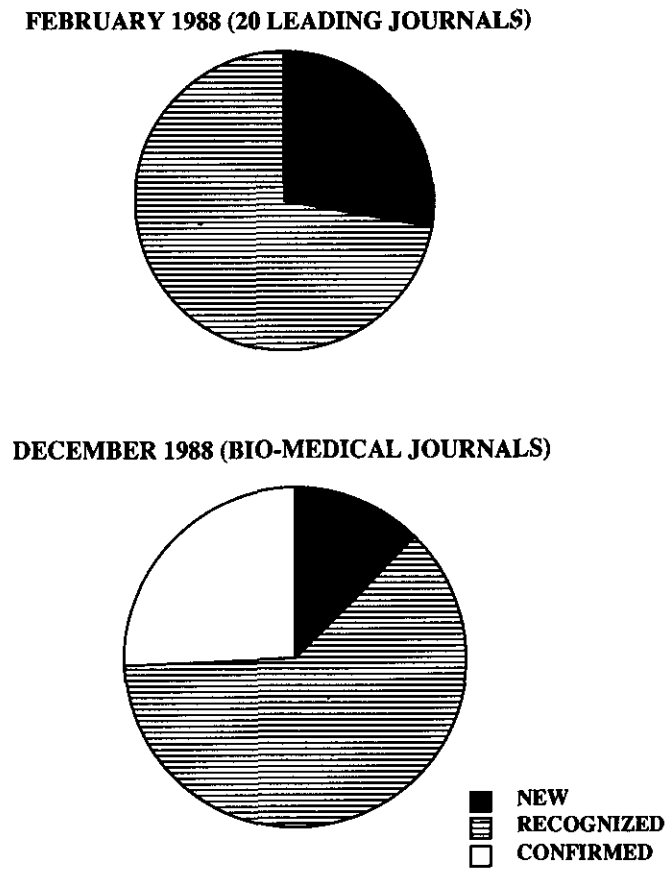

Figure 1 : Results of matching the masterfiles with new 'raw' addresses : Publications from all countries.

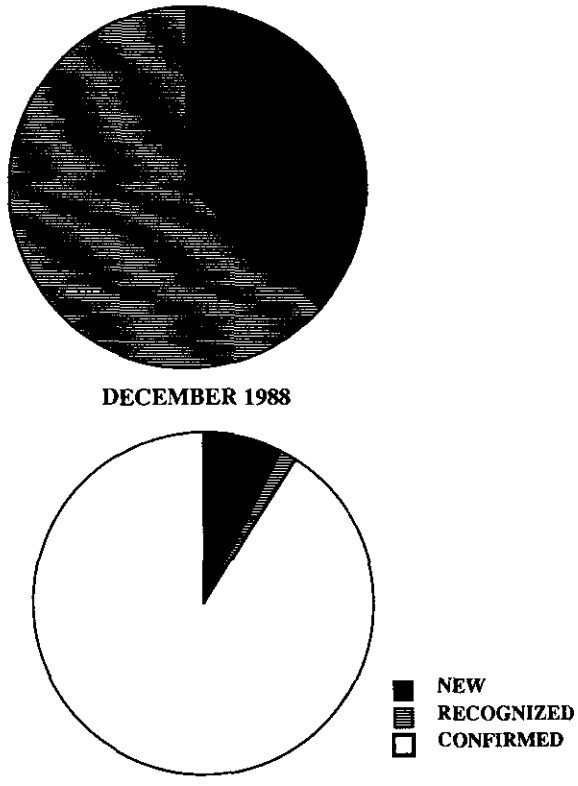

Figure 2 : Results of matching the masterfiles with new 'raw' addresses : The Netherlands

'New' in these figures indicates (first parts of) addresses appearing for the first time in the processing (and therefore unknown to the masterfiles). 'Recognized' means that we found the name before, and that it was unified by analyzing publication data or that it was not changed. 'Confirmed' relates to addresses of which the unification is based on further research. The figures show the increasing usefulness of the masterfiles. The results for the Netherlands are especially striking. The vast majority of Dutch unified addresses was confirmed by further research. The results in the all countries figure are highly influenced by the US, which dominates the dataset. As we pointed out, we have not carried out further research for the US. However, not only the knowledge in the master files is decisive for the share of the categories. When processing a journal or set of journals with a completely different scope, we find many new addresses.

The definition of the types of organizations publishing in scientific organizations is the next step in our unification research. Recently we added a field of categories to the master files. Thus we can define easily the share of universities, research centers, companies, etc.

6. CONCLUSIONS

Based on our analysis of addresses in scientific publications, extracted from SCISEARCH, we have shown that many variations occur with respect to the names of organizations and their geographic locations. The method of unification we propose, has proved to be highly successful for purposes of evaluative bibliometrics. The scattering problem is largely solved and, in our opinion, rankings of institutes by publications and citations have become much more reliable. The same can be said for rankings by country.

An extension of the database is the addition of categories to the unified addresses : universities, research centers, companies, etc. This addition makes it possible to analyze what kind of organizations publish in different journals.

However, more applications of the unification method are possible. For assessing international cooperation connections, unified addresses relating to institutes, cities, and countries are extremely important. The same is true of citing patterns. In the field of science policy a dataset of unified addresses may be very useful for evaluation purposes. Evaluation of research performance can be simplified (or even made possible) by unification of addresses. For tracing developments in productivity, universities can now be compared. Furthermore, shifts in the scope of interest can be discovered more easily. The planned addition of categories is also helpful for these purposes.

ACKNOWLEDGEMENT

The research described in this paper was partly financed by Elsevier Science Publishers.

NOTES

1. Evaluative bibliometrics is characterized by the use of bibliometric data for evaluation purposes of 'qualitative' aspects of scientific activity, particularly of research performance. In this approach, bibliometric data are applied in the operationalization of 'qualitative' concepts (Moed, 1989).

2. We have made use of the following handbooks and encyclopedias :
   - *Commonwealth Universities Yearbook 1987. A Directory to the Universities of the Commonwealth and the Handbook of their Association* (4 vols.). Association of Commonwealth Universities, London, 1987.
   - *International Handbook of Universities and Other Institutes of Higher Education* (10th ed.). MacMillan - Stockton - International Association of Universities, New York - Paris, 1986.
   - *Internationales Universitäts Handbuch/World Guide to Universities* (2nd ed.; 4 vols.). Saur, Munich - New York, 1976-1977.
   - Knowles, S., ed., *The International Encyclopedia of Higher Education.* Jossey-Bass Publishers, San Francisco - Washington - London, 1977.
   - Lengenfelder, H., ed., *Handbuch der Universitäten und Fachhochschulen. Bundesrepublik Deutschland, Oesterreich, Schweiz* (3rd ed.), Saur, Munich - New York - London - Paris, 1985.

- *The World of Learning* (38th ed.). Europa Publications, London, 1988.
- *World Guide to Scientific Associations and Learned Societies/Internationaler Verzeichnis wissenschaftlicher Verbände und Gesellschaften* (4th ed.), Saur, Munich - New York - London - Paris, 1984.
- *World List of Universities/Liste mondiale des Universités* (17th ed.), MacMillan - Stockton - International Association of Universities, New York - Paris, 1988.

3. *The World of Learning* (1988) contains such a list of CNRS institutes.

REFERENCES

Altbach, P.G., Perspectives on Comparative Higher Education : A Survey of Research and Literature, In : P.G. Altbach and D.H. Kelly (eds.), *Higher Education in International Perspective. A Survey and Bibliography*. Mansell, London - New York, 1985.

Anderson, J., P.M.D. Collins, J. Irvine, P.A. Isard, B.R. Martin, F. Narin, and K. Stevens, On-line Approaches to Measuring National Scientific Output - A Cautionary Tale. *Science and Public Policy*, 15 (1988), 153-161.

M.P. Carpenter, F. Gibb, M. Harris, J. Irvine, F. Narin, Bibliometric Profiles for British Academic Institutions : An Experiment to Develop Research Output Indicators. *Scientometrics* 14 (1988), 213-234.

Clark, Burton R., *The Higher Education System. Academic Organization in Cross-National Perspective*. University of California Press, Berkeley - Los Angeles - London, 1983.

Commonwealth Universities Yearbook 1987. A Directory to the Universities of the Commonwealth and the Handbook of their Association (4 vols.). Association of Commonwealth Universities, London, 1987.

International Handbook of Universities and Other Institutes of Higher Education (10th ed.), MacMillan - Stockton - International Association of Universities, New York - Paris, 1986.

Internationales Universitäts Handbuch/World Guide to Universities (2nd ed.; 4 vols.), Saur, Munich - New York, 1976-1977.

Kluczynsky, J., Research on Higher Education in European Socialist Countries, In : P.C. Altbach and D.H. Kelly (eds.), *Higher Education in International Perspective. A Survey and Bibliography*. Mansell, London - New York, 1985.

Knowles, S., ed., *The International Encyclopedia of Higher Education*. Jossey-Bass Publishers, San Francisco - Washington - London, 1977.

Lengenfelder, H., ed., *Handbuch der Universitäten und Fachhochschulen. Bundesrepublik Deutschland, Oesterreich, Schweiz* (3rd ed.), Saur, Munich - New York - London - Paris, 1985.

Leydesdorff, L., Problems with the 'Measurement' of National Scientific Performance. *Science and Public Policy* 15 (1988), 149-152.

Moed, H.F., The Use of On-line Bibliometric Analysis, In : L. Egghe and R. Rousseau (eds.), *Informetrics 87/88*. Elsevier Science Publishers, Amsterdam - New York - Oxford - Tokyo, 1988, 133-146.

Moed, H.F., *The Publication Activity and Internationality of Dutch Research in the Natural and Life Sciences during the Period 1984-1987. A Data Report prepared for the National Science Advisory Council (RAWB)*, Lisbon Institute, Leiden, 1988a.

Moed, H.F., *The Use of Bibliometric Indicators for the Assessment of Research Performance in the Natural and Life Sciences. Aspects of Data Collection, Reliability, Validity and Applicability.* DSWO Press, Leiden, 1989.

Moed, H.F. and A.F.J. van Raan, Indicators of Performance : Applications in University Research Policy, In : A.F.J. van Raan (ed.), *Handbook of Quantitative Studies of Science and Technology.* Elsevier Science Publishers, Amsterdam - New York - Oxford - Tokyo, 1988.

Persson, O., Measuring Scientific Output by Online Techniques, In : A.F.J. van Raan (ed.), *Handbook of Quantitative Studies of Science and Technology.* Elsevier Science Publishers, Amsterdam - New York - Oxford - Tokyo, 1988.

Shrum, W. and N. Mullins, Network Analysis in the Study of Science and Technology, In : A.F.J. van Raan (ed.), *Handbook of Quantitative Studies of Science and Technology.* Elsevier Science Publishers, Amsterdam - New York - Oxford - Tokyo, 1988.

Soëtard, Michel, Die grosse Umwälzungdes Universitätswesens in den 60er Jahren in Frankreich, In : H. Roehrs, ed., *Tradition und Reform der Universität unter internationalem Aspekt.* Verlag Peter Lang, Frankfurt am Main - Bern - New York - Paris, 1987.

Van Raan, A.F.J., *Wetenschapsbeleid : Schaduw van de wetenschap? College-serie over Wetenschap en samenleving.* Studium Generale, University of Leiden, 1985.

Wissenschaftsrat. *Empfehlungen des Wissenschaftsrates zu den Perspektiven der Hochschulen in den 90er Jahren.* Cologne, 1988.

World Guide to Scientific Associations and Learned Societies/Internationaler Verzeichnis wissenschaftlicher Verbände und Gesellschaften (4th ed.), Saur, Munich - New York - London - Paris, 1984.

The World of Learning (38th ed.). Europa Publications, London, 1988.

World List of Universities/Liste mondiale des Universités (17th ed.). MacMillan - Stockton - International Association of Universities, New York - Paris, 1988.

Zwaan, R.A., R.E. de Bruin, and H.F. Moed, *Publication Activity and Internationality of Dutch Research in the Social Sciences : A Methodological Discussion and some Preliminary Results of an On-line Bibliometric Study. Research Report to the National Science Advisory Council (RAWB).* Lisbon Institute, Leiden, 1988.