Research Evaluation, volume 3, number 1, April 1993, pages 55-62, Beech Tree Publishing, 10 Watford Close, Guildford, Surrey GU1 2EP, England.

Relational database

Constructing a relational database for bibliometric analysis

M T Fernández, A Cabrero, M A Zulueta and I Gómez

A relational database with information downloaded from different bibliographic databases is presented. Its structure aims to obtain bibliometric and scientometric indicators for science policy. All fields have to be adequately normalised. A semi-automatic system to standardise and codify the institutional corporate sources and the journals is described. Several applications to case studies in science policy are given.

M T Fernández, A Cabrero, M A Zulueta and I Gómez are at Centro de Información y Documentación Científica (CINDOC), CSIC, Joaquín Costa 22, 28002 Madrid, Spain.

THE DATA FOR BIBLIOMETRIC analysis can be extracted from different sources of information, but the most heavily used nowadays are the bibliographic databases. These are oriented towards document retrieval, as they have primarily been created for bibliographic purposes. In order to use them in bibliometric analysis several problems have to be faced, most of them due to the lack of standardisation of the information included in the different fields, and the lack of appropriate search software for bibliometric calculations (Moed *et al*, 1992; Bruin and Moed, 1990). The degree of standardisation of information of the different fields varies greatly in the databases and affects several indicators.

One general task faced in science policy is to analyse the national research performance of a specific country in a national and international context. To do this, the journal coverage and topics of the database used have to be taken into account, as well as whether the database offers the possibility of searching on the country of the authors, which is not always standardised, or in a separate field, originating noise in the retrieval (Gómez and Galbán, 1986). Besides, most databases introduce just one author's address, making impossible the retrieval of a country's production when it is the second country in multi-national papers.

The research institution of the authors is written in many different ways in the 'corporate source' field and has to be standardised in order to quantify the scientific performance of a certain research institute or faculty. Even in a database such as the *Science Citation Index* (SCI), in which the annual production of journal citation reports (JCR) dem-

0958-2029/93/010055-8 US$08.00

In three fields of the downloaded information there was a special need for standardisation in order to be able to apply computerised calculations: the author's institutional address, author's name and publication journal

onstrates the special importance attached to bibliometric analysis, this field is not homogeneous, which makes it difficult to obtain bibliometric indicators through any simple counting.

The difficulty is greater when trying to merge the information of several databases into a single data file. In this case, even a relatively standardised field such as the publication journal is abbreviated in different ways, which makes it necessary to previously merge the titles.

Obtaining an author's scientific output is another problem: all the different possible ways in which the author's name can be written have to be recognised and standardised, particularly in Spain where two family names are currently used.

All this points to the need to create a structured database with unified information in each of its fields and appropriate software to enable bibliometric studies.

In our Centre during the last decade we have analysed the scientific production of Spain both from a general perspective, using multidisciplinary databases such as SCI, SSCI (Social Science Citation Index), ICYT (Indice Español de Ciencia y Tecnología — Spanish Index of Science and Technology), and focusing on certain scientific disciplines, using *Physics Brief*, INSPEC, *Chemical Abstracts*, BIOSIS, *Medline*, *Excerpta Medica* and other specialised databases. We also have experience in creating a bibliographic database (ICYT), as well as the appropriate software for bibliometric analysis written in the Cobol language for the minicomputer Secoinsa 40/2.

There are several logical structures for designing databases, among which the relational model has been broadly used. It has recently been applied for bibliographic databases (Winterhager, 1991; Blair, 1988). Due to the structure of our data, we chose the relational model according to the Boys-Codd normal forms (Date, 1981) to satisfy the requirement to produce a unique database that offers the possibility of comparing and retrieving information for bibliometric studies.

Particular items and their attributes were selected and grouped into relations through tables and fields. Several key attributes are related to each other in order to construct bibliometric indicators capable of answering scientific policy questions, such as: which topics are covered by a particular research institute, or what is the geographical distribution of research activity in Spain in a certain field and which centres are involved.

The database management system DBase IV.2 was used to develop our specific database. The system runs in a stand-alone IBM compatible PC486-66 Mh; 8 Mb RAM; 520 SCSI-HD.

There follows a short description of the data used, the standardisation process and master files needed, and the database structure.

Data

The data for our bibliometric studies were obtained from different bibliographic databases and other sources (annual reports, curricula, and so on) according to the goals, scope and subject to be analysed. Specialised databases were used, together with SCI and ICYT multidisciplinary ones. The latter enable co-operation studies, as they record the corporate sources of all the authors, which is the simplest way of studying co-operation networks between institutions or countries.

The information from bibliographic databases was downloaded directly into ASCII files through tailor-made software, specific for each database or source used, and was split into three related files (Braun *et al*, 1985; Winterhager, 1991):

- documents
- authors
- institutions

Each document has a 'document number' as principal key that relates it to its authors and institutions.

These bibliographic data have been complemented by other information as will be shown in the master files.

Standardisation and codification

In three fields of the downloaded information there was a special need for standardisation in order to be able to apply computerised calculations: the author's institutional address, author's name and publication journal. These data present many variations, not only due to misspelling, but also to the different ways in which they are structured and expressed.

The institutional address of the authors is expressed differently in each of the databases used, both as to the number of fields in which it appears and their internal organisation. In general, the country is more or less standardised, although the United Kingdom and both Germanies present different problems over time. The city with its postal code is frequently immediately before the country.

The name of the research centre, faculty or hospital department where the author is working has many variations. In the SCI the institution is usually first, followed by the name of the research institute or department, postal code, city and country. Sometimes part of this information is missing. In *Chemical Abstracts* through STN International (The Scientific and Technical Network), the city, postal code and country are in a separate field, while the institution and research centre are in another field, which facilitates search accuracy and automatic treatment of the data.

In both BIOSIS and INSPEC one field contains the department, faculty, university, city and country. In *Physics Brief* all this information comes together with the names of the authors, making any automatic separation much more difficult, especially as this database registers the institutional addresses of all the authors.

The problem is not only the order in which the different items of information appear, but the different ways in which it is expressed and the possible errors introduced by the databases. As a consequence, no bibliometric indicators that imply the use of this field can be directly calculated.

The journal field also varies from one database to another, rendering it difficult to merge data from different sources. The ICYT database records full journal titles and ISSN (International Standard Serial Number); *Chemical Abstracts* records abbreviated title (following standard ISO 833-74), CODEN (another abbreviation of the title) and ISSN; BIOSIS gives an abbreviated title separated by hyphens, and CODEN; *Physics Brief* includes abbreviated title, ISSN and CODEN in a single field; INSPEC has full title, CODEN, ISSN and country of publication in the same field; SCI, depending on the version used, offers full journal title, abbreviated eleven digit title or short title following approximately the international standard. This database in its on-line version includes a subject classification of the journals.

Attempts have been made by different authors to use on-line bibliographic databases for science studies and the difficulties have been pointed out (Moed, 1988; 1991; Tijssen, 1991). In our Centre we solved these problems for the case of Spain by creating a tailored relational system and master files with standardised information related to research centres, cities, countries, journals and disciplines, which allowed us to validate, unify and codify the different fields.

The aim is to obtain a common language with no ambiguities that will enable automatic treatment of the data. It has to be exact, easy to use, meaningful and applicable to new items. The codification was planned according to the bibliometric indicators considered interesting for science policy studies. The structure used allowed us to introduce meta data with added value proceeding from different non-bibliometric sources. The information of these master files is updated and corrected as a result of the research projects developed in the Centre.

Table 1. Structure of the database

Item	Attributes
Data files	
Documents	Document number*; title; number of authors; number of institutions; document type; language; 11-character journal name; volume; number; page; year
Authors	Document number*; author ordinal; author name
Institutions	Document number*; institution ordinal; institution name; address; city; country; institution code*
Master files	
Research centres	Institution code*; standardised institution name; city; country; previous centre code
Spanish cities	Standardised city name; other forms of city name; postal code*; autonomous community
Foreign countries	Standardised English name; standardised Spanish name; other forms of country name; ISO country code*; regional code
Journals	Full journal name; 11-character journal name*; CODEN; ISSN; CODE (first letter of the non-empty words); edition country; basic-clinical level
Journal impact factor	11-character journal name*; year; year impact factor
Journal disciplines	11-character journal name*; SCI discipline
Discipline correspondence	SCI discipline*; SPRU area

Structure

The relational structure, shown in Table 1, is built up using the master files and the three files — documents, authors and institutions — obtained from the data to be studied, once it has been validated and standardised with the help of the master files.

Address related master files

Research centres

This file comprises all Spanish research centres and a certain number of foreign ones. The principal data sources used for the creation of this master file are the Spanish science and technology journals indexed by the ICYT database and those foreign journals covered by international databases, in which Spain is publishing.

In this file the name of each research centre is standardised, and a five level institution code, which gives structured information on different aspects of each research centre, is introduced to produce bibliometric indicators (Fernández-Frial *et al*, 1990; Fernández *et al*, 1990). The Spanish research centres are codified at a deeper level than the foreign ones that co-operate with Spanish authors. This institution code has the following structure:

Level 1. (1 digit) Administrative dependency
- 0 Unknown dependency
- 1 University
- 2 National Research Council
- 3 Central administration
- 4 Autonomous administration
- 5 Local administration
- 6 Public enterprises
- 7 Foundations
- 8 Private enterprises
- 9 Foreign organisations

Level 2. (3 digits)
Type of organisation within each administrative dependency; for foreign institutions a 3-letter country code is introduced

Level 3. (6 digits)
Mnemotechnic acronym of research centre or institute code for the National Research Council

Level 4. (4 digits)
Disciplinary UNESCO code

Level 5. (4 digits)
Postal code for Spain; in the case of foreign centres, the city and state are codified.

In Table 2 an example of codification of several Spanish centres is shown.

Continuous effort is needed to update this master file, because new centres are added and names change (and all are collected and internally related) and administrative dependencies of existing centres alter. This is an open file, never completed. Numerous handbooks, university publications and annual reports have been consulted to decide the appropriate codes.

Spanish cities

This file was created to unify all variations and errors found in place names. It contains all variations of names of Spanish cities or suburbs with their postal codes that indicate the province and the autonomous community.

Foreign countries

The variations in country names found through the downloading of data from the different databases have been introduced, together with their correct Spanish and English names and a three character ISO (International Standards Organisation) code. The different parts of one country are aggregated (for instance, England, Scotland, Wales and Northern Ireland) and several countries are also aggregated into multinational regions (European Union, Latin America etc).

Table 2. Example of institution codes for Spanish centres

Research centre	Org	Acronym	UNESCO code	Postal code
Faculty of Medicine University of Navarra	1 PR	1ME	32	31
Estac Biol "Doñana" (CSIC) Sevilla	2	060401	24	41
Hospital "La Paz" (Insalud) Madrid	3 MSC	HINSA3	32	28
Laboratory Almirall S.A. Barcelona	8	ALMIRA	32	08

Coding addresses

When a new bibliometric study is started, the names of the institutional corporate sources must be unified, after the information from the bibliographic database has been downloaded into the separate files mentioned above. Each organisational heading line is matched with our master files. First, the country of origin is selected and coded using the country master file.

In the case of a Spanish organisation, the city and autonomous community are coded through the city master file before the corporate source line is compared against the research centres master file. The strategy is to obtain from the normalised research centre master file all names that match the new address within a predefined percentage. This is done by comparing the three first characters of all non-empty words linked with an 'and' operator. The position of the words does not influence the result, which is important because the order very frequently differs.

For each new address the system offers a set of possible candidates with their percentage of agreement (see Table 3). If one is manually accepted, the 18-digit significative 'institution code' is directly linked to the new address, together with its standardised name.

A temporary file is created with all centre names

For a new bibliometric study, the names of the institutional corporate sources must be unified, after the information from the bibliographic database has been downloaded into the separate files

Table 3. Example of semi-automatic standardisation of new addresses

Non-standardised centre name
*** Instituto de Bromatologia y Nutricion. CSIC. Facultad de Farmacia de la Universidad Complutense de Madrid

The system offers:

Standardised centre name	% of hits
Inst. Nutr. Bromatol. (CSIC) - Fac. Farm. Univ. Complut. Madrid	p:100
Inst. Farmacol. Toxico. (CSIC) - Fac. Farm. Univ. Complut. Madrid	p:077
Dep. Invest. Bromatol. Fac. Farm. Univ. Complut. Madrid	p:077
Inst. Bioqui. (CSIC) - Fac. Farm. Univ. Complut. Madrid	p:077
Inst. Geol. Econ. (CSIC) - Fac. Farm. Univ. Complut. Madrid	p:066
Inst. Astron. Geod. (CSIC) - Fac. Farm. Univ. Complut. Madrid	p:066
Dep. Bioqui. (CSIC) - Fac. Farm. Univ. Complut. Madrid	p:066

not found in the master file. They are coded manually off-line by consulting the necessary directories, and the new information is added to the master file.

Currently the master file contains 5,512 Spanish institutions and 1,263 foreign ones with standardised name and institution code. Its update is continuous due to the new centres added as a result of our research, and the changes in names and administrative dependency.

Journal related master files

The information related to journals has been structured in four files, as shown in Table 1.

Journals master-file

Although the journal field is not as complicated as the corporate source field, it also requires unification. The journal master file was created by entering in a single file all the information gathered from the databases studied. To enhance this file and the possibility of calculating scientometric indicators, other data were manually added by consulting several manuals, handbooks and classification systems.

This file has the full name of each journal in which a Spanish author has published, as well as an abbreviated 11-character journal name according to the SCI database, CODEN, ISSN, country of edition and basic-clinical level as introduced by Narin *et al* (1976). A code with the first letter of non-empty words was introduced for semi-automatic matching (Table 4).

Journal impact factor

This file contains the JCR impact factor for different years for each journal.

Journal discipline

This complements the previous file giving the subject classification of journals according to the SCI (maximum three disciplines per journal).

Discipline correspondence

This file relates the SCI classification in disciplines with the SPRU areas (Irvine *et al*, 1990).

Coding journals

In the same way as for the corporate source field, the journal field has to be standardised when working with different databases. The journal field is matched to the journal master file, if possible through the ISSN, CODEN or full name. If these fields are not available, the first letter of all non-empty words of the journal name (code) are com-

Table 4. Example of journal master file

Full name	11-char name	CODEN	ISSN	CODE	Country of edition	Basic clinical level
Acta Paediatrica Scandinavica	Act Paed Sc	APSVAM	0001-656X	APS	DNK	2
British Journal of Nutrition	Br J Nutr	BJNUA	0007-1145	BJN	GDR	2
Pharmacology	Pharmacol	PHMGB	0031-7012	P	CHE	3
Revista Española de Fisiología	Rev Esp Fis	REFIAS	0034-9402	REF	ESP	4

Table 5. Example of semi-automatic coding of journals

Non-standardised journal name:
*** European Journal Mechanics Solids

Code	Full title	11-char name
ejm	European Journal of Mineralogy	Eur J Miner
ejm	European Journal of Morphology	Eur J Morph
ejmc	European Journal of Medicinal Chemistry	Eur J Med C
ejmf	European Journal of Mechanics Fluids	Eur J Mec B
ejms	European Journal of Mechanics Solids	Eur J Mec A

pared. The system offers all the possible hits, and when one is manually selected the journal is automatically linked (Table 5).

If a journal is not found in the master file its attributes are searched for in other sources and added to the master file.

Bibliometric indicators

The files presented in Table 1 constitute a related system that enables us, through associative linking and searching, to manipulate the items and obtain valuable information and statistics to produce bibliometric indicators. For special cases the appropriate *ad hoc* queries are constructed; for repeat searches predefined subschema implemented in DBaseIV are used. We have started with basic topics in bibliometric analysis, but our intention is to broaden the scope by adding to the database schema new items and relations: citations, research budgets and manpower, grants and so on.

This structure has been used in the following research projects related to science policy issues, for which bibliometric indicators were considered as partial indicators in the analysis of research activity: scientific co-operation between Spain and Latin American countries (Sancho *et al*, 1992); Spanish scientific production in biomedicine by specialty (Camí *et al*, 1992); research activity of the Autonomous Community of Madrid (Gómez *et al*, 1993a) and of the institutes of the Spanish Research Council (Gómez *et al*, 1993b). Some of the results of these studies are described as an example of the information that can be obtained.

An interesting use of this relational database is the study of co-operation networks at different levels: at international level (with which foreign countries Spain is publishing); inter-institutional co-operation (for instance, between university and Research Council in Spain); inter-regional co-operation between the different cities or autonomous communities in Spain; intra-institutional co-operation between research centres, or departments of the same organisation.

Along these lines, a subschema related to the collaboration matrix between countries has been implemented. The system gives options as to how many and which countries are to be related, which of them must appear in rows and which in columns; which single country has to be compared with all the others in time series to analyse their co-authored documents; whether all document types

Table 6. Collaboration between Latin American countries

Country	AR	BO	BR	CL	CO	CR	CU	DO	EC	GT	HN	MX	PA	PY	PE	UY	VE
Argentina	-	5	281	108	10	1	2	3	3	6	-	46	1	4	6	18	60
Bolivia	5	-	2	4	1	-	-	-	1	-	-	2	-	-	4	-	-
Brazil	281	2	-	159	43	17	6	24	6	4	2	99	5	3	26	19	53
Chile	108	4	159	-	14	3	3	14	3	1	1	52	-	-	17	5	28
Colombia	10	1	43	14	-	15	2	-	4	1	-	31	9	-	9	1	10
Costa Rica	1	-	17	3	15	-	3	3	2	6	5	21	16	2	2	-	6
Cuba	2	-	6	3	2	3	-	-	1	-	-	9	-	-	1	2	4
Dominican	3	-	24	14	-	3	-	-	1	2	-	1	2	-	2	-	4
Ecuador	3	1	6	3	4	2	1	1	-	1	-	3	1	-	3	-	1
Guatemala	6	-	4	1	1	6	-	2	1	-	1	8	2	-	1	-	1
Honduras	-	-	2	1	-	5	-	-	-	1	-	2	-	-	-	-	-
Mexico	46	2	99	52	31	21	9	1	3	8	2	-	9	2	7	11	37
Panama	1	-	5	-	9	16	-	2	1	2	-	9	-	-	-	-	2
Paraguay	4	-	3	-	-	2	-	-	-	-	-	2	-	-	-	--	-
Peru	6	4	26	17	9	2	1	2	3	1	-	7	-	-	-	-	11
Uruguay	18	-	19	5	1	-	2	-	-	-	-	11	-	-	-	-	4
Venezuela	60	-	53	28	10	6	4	4	1	1	-	37	2	-	11	4	-
Total col	554	19	750	412	150	105	33	58	31	35	11	340	47	11	89	60	222
Total doc	517	15	693	367	126	78	26	44	20	28	10	296	32	11	72	55	200

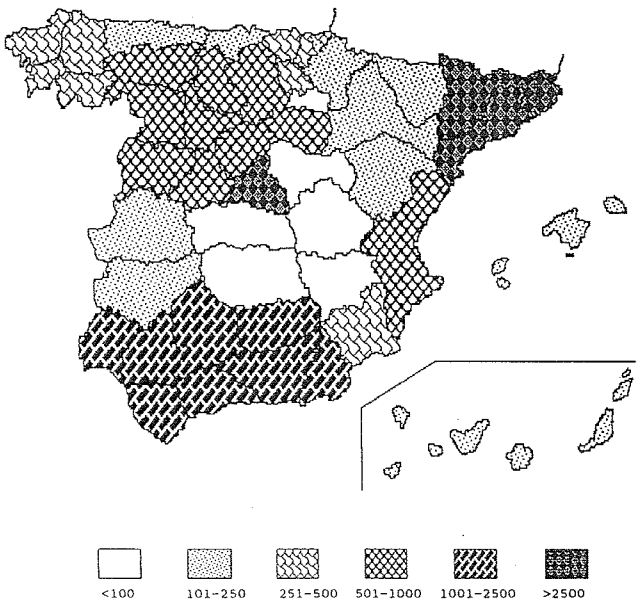

Figure 1. Distribution of biomedical articles by autonomous community

or only articles are to be selected; whether the information is displayed on screen or printer. The example in Table 6 is part of a research project on Latin America and its co-operation links.

The next example was obtained within a research project studying the international visibility of Spanish biomedical research. By coding the institutional addresses, the geographical distribution of research activity in Spain by autonomous communities was obtained. In Figure 1 the strong concentration of biomedical research activity in Madrid and Cataluña can be observed, which is important when new Health Ministry research grants are being distributed.

The scientific performance by institutions in Spain was studied at two levels. In the first place at a general level as productive sectors: university, hospital, Spanish Research Council (CSIC), joint centres CSIC-univ, industry and others. The second deeper level of analysis is as individual research centres, hospitals or university faculties, each having a unique institution code.

To analyse scientific production not only quantitatively but also qualitatively, some metadata introduced in the journal master file were used: subject classification of journals, impact factor, basic or clinical research level. The scientific production of a research centre in each of the different specialties was thus obtained, with their average impact factor (IF) and basic-clinical level.

To detect 'centres of excellence', a relative impact factor (RIF) was introduced, using the average IF of a specialty at the national level for comparison (when RIF>1 the expected IF is higher than the Spanish average in the specialty). In Table 7 an example of a research centre of the Spanish Research Council is shown.

Another possible breakdown is to highlight the active research institutions and centres in a certain

Table 7. Production, basic/clinical level, IF and RIF of the Instituto de Ciencia de Materiales, Sede A (CSIC)

	Instituto de Ciencia de Materiales , Sede A				Spain		
Specialty	Docs	Level	IF	RIF	Docs	Level	IF
Physics, solid state	51	3.5	1.69	0.91	956	3.6	1.85
Physics, applied	36	3.0	1.77	0.95	491	3.1	1.85
Chemistry, physical	25	3.0	2.09	1.37	1117	3.4	1.52
Materials science	22	2.9	1.40	1.25	655	2.4	1.11
Chemistry, inorg nucl	17	4.0	1.93	1.23	845	3.9	1.56
Chemistry organic	12	4.0	1.92	1.10	1382	4.0	1.74
Physics	12	3.8	1.25	0.57	1041	3.8	2.18
Chemistry	11	4.0	1.13	1.15	1376	3.5	0.98
Engineering, electr	7	3.0	0.89	0.99	261	2.0	0.89
Mineralogy	6	2.0	0.98	0.98	74	2.5	1.00
Physics, atomic & molec	3	4.0	2.40	0.96	456	3.9	2.48
Materials sci ceram	3	3.0	1.11	1.09	92	2.6	1.02
Food sci & technol	1	3.0	0.32	0.39	447	2.5	0.83
Chemistry applied	1	3.0	0.32	0.42	165	2.8	0.77
Nutrition & dietetics	1	3.0	0.32	0.39	164	2.9	0.82
Microscopy	1	4.0	1.56	1.27	54	3.6	1.22
Instrumentation	1		0.62	0.70	206	2.8	0.89
Engineering, chemical	1	3.0	2.38	2.58	333	2.2	0.92
Geography	1	1.0	0.63	0.80	47	2.7	0.79
Physiology	1	4.0	2.05	1.17	603	3.9	1.76
Crystallography	1	4.0	1.08	1.18	282	3.5	0.91
Biochem & mol biol	1	1.0	0.49	0.17	2080	3.9	2.82
Total	215				13127		

Table 8. Most productive research centres in physical chemistry from the Autonomous Community of Madrid

Spain		Madrid		
No of documents: 1117		No of documents: 473		
Average IF: 1.53		Average IF: 1.60		
Average level: 3.5		Average level: 3.5		
Centres	**Docs**	**Level**	**IF**	**RIF**
Fac Cienc UAM	110	3.3	1.72	1.13
Fac Quim UCM	80	3.6	1.36	0.89
I Catal Petroleoqui	78	3.1	1.83	1.19
I Quim Fis "Rocaso"	42	3.6	1.20	0.78
Fac Cien Univ Alcala	32	3.7	1.45	0.95
I Estructura Materia	26	3.9	1.01	0.66
Fac Cienc Uned	25	3.5	1.41	0.92
I Cienc Mater Sede A	25	3.0	2.09	1.37
Fac Farm UCM	23	3.3	1.18	0.77
I Cienc Mater Sede B	19	3.1	2.24	1.46
I Cienc Mater Sede C	18	3.7	2.27	1.48
I Opti Daza Valdes	16	3.4	1.13	0.74
I Quimica Medica	16	4.0	1.19	0.77
I Quim Organica Gen	11	3.7	1.46	0.95
I Cienc Tec Polimer	10	4.0	1.82	1.19
UCM (varios)	7	4.0	1.71	1.12
Fac Fis UCM	6	3.6	0.99	0.65
I Cienc Mater Sede D	5	3.2	1.13	0.74
Etsi Montes UPM	5	4.0	1.83	1.20
I Fis Fundam UCM #	5	3.2	1.86	1.22
I Ceramica Y Vidrio	4	3.5	0.96	0.63
C Nacl Microelec (M)	4	3.0	2.51	1.64

discipline. For example, the institutional sectors implied in physical chemistry of the Autonomous Community of Madrid can be selected, or the most active individual research centres (Table 8). The data from Madrid are compared with those of the whole of Spain by number of documents, average IF (through the RIF) and basic/clinical level.

Discussion

Due to the difficulty of obtaining bibliometric indicators directly from on-line bibliographic databases, we decided to create a tailor-made, dedicated, relational structure and the necessary software tools, to enable us to create, manipulate, update and handle metadata and produce information of interest to bibliometricians and science-policy makers. Questions at meso-and micro-levels can be addressable to the system focusing on institutions, regions, groups, disciplines, or national and international co-operation.

The time-consuming validation and standardisation process has been considerably reduced with the methodology established, and at the same time more accuracy is attained and inconsistencies and redundancy have been avoided.

As the information and inter-related keys contained in the relational tables grow, the response time in the PC system slows down. Our immediate project is to transfer to a mainframe system that will reduce the response time and will be able to integrate the data collected from our research projects on Spanish research activity.

References

D C Blair (1988), "An extended relational document retrieval model", *Information Processing and Management*, 24(3), pages 349-371.

T Braun, W Glänzel and A Schubert (1985), *Scientometric Indicators. A 32-Country Comparative Evaluation of Publishing Performance and Citation Impact* (edited by World Scientific, Singapore, Philadelphia).

R E Bruin and H F Moed (1990), "The unification of addresses in scientific publications", *Informetrics, 89/90* (Elsevier Science Publishers, Amsterdam-New York-Oxford-Tokyo) pages 65-78.

J Camí and I Gómez (1992), "La producción científica española en Biomedicina y Salud. Un estudio a través del Science Citation Index (1986-1989)", report.

C J Date (1981), *An Introduction to Database System, vol 1* (Addison-Wesley, 3rd edition).

M J Fernández-Frial, C Ortega, I Gómez, M V de Dios and M T Fernández (1990), "Sistema de codificación y normalización de instituciones", *Actas de las terceras jornadas españolas de documentación automatizada* (Palma de Mallorca).

M T Fernández, C Ortega and M V de Dios (1990), "Automated bibliometric use of databases: a case study", *Proceedings of 45th Conference and Congress of the International Federation for Information and Documentation (FID)*, La Habana.

I Goméz, M T Fernández, A Cabrero, M A Zulueta, G Buey and A Méndez (1993a), "La producción científica de la Comunidad Autónoma de Madrid en el Science Citation Index y en el Social Science Citation Index en el trienio 1990-1992. Un análisis bibliométrico", report, Madrid.

I Goméz, M T Fernández, A Cabrero, M A Zulueta and G Buey (1993b), "La investigación del CSIC en el Science Citation Index y en el Social Science Citation Index en el trienio 1990-1992. Un análisis bibliométrico", report, Madrid.

I Gómez and C Galbán (1986), "Lack of standardisation in the corporate source field of different databases", *10th International Online Information Meeting* (Learned Information Ltd, London).

J Irvine, B R Martin and P Isard (1990), *Investing in the Future. An International Comparison of Government Funding of Academic and Related Research* (Billing and Sons Ltd, Worcester, Vermont, and Edward Elgar Publishing Ltd).

H F Moed (1988a), "The use of on-line databases for bibliometric analysis", in L Egghe and R Rousseau (editors), *Informetrics 87/88* (Elsevier Science Publishers, Amsterdam) pages 145-158.

H F Moed (1991), "New developments in on-line bibliometric analysis", *Proceedings of the Joint EC-Leiden Conference on Science and Technology Indicators* Leiden (The Netherlands).

H F Moed, R E Bruin, A J Nederhof, A F J Van Raan and R J W Tijssen (1992), "State of the art bibliometric macro-indicators. An overview of demand and supply", *Research Evaluation EUR 14582 EN* (Monitor Spear, Commission of the European Communities).

F Narin, G Pinsky and H H Gee (1976), "Structure of the biomedical literature", *Journal of the American Society for Information Science*, 27(1), pages 25-45.

R Sancho, A Cabrero, F Delgado and M T Fernández (1992), "A study of scientific cooperation between South and Central American and Caribbean countries", *Proceedings of 46th Conference and Congress of the International Federation for Information and Documentation (FID)*, Madrid (in press).

R J W Tijssen and H F Moed (1991), "Literature-based statistical analyses of international scientific co-operation", *Proceedings of the Joint EC-Leiden Conference on Science and Technology Indicators*, Leiden (The Netherlands).

M Winterhager (1991), "Towards bibliometric objects: a relational view to ISI's Science Citation Index", *Conference on Science and Technology Indicators*, Leiden, The Netherlands.